

UNIVERSITÉ PARIS 13  
« ÉQUIPE DE RECHERCHE EN ÉPIDÉMIOLOGIE NUTRITIONNELLE »  
U1153 Inserm / U1125 Inra / Cnam / Université Paris 13

Année 2019

N°

**THÈSE**

Pour obtenir le grade de

**DOCTEUR DE L'UNIVERSITÉ PARIS 13**

Discipline : Epidémiologie-Santé Publique

Présentée et soutenue publiquement le 30 septembre 2019 par

**Lucie Lécuyer**

Née le 14 avril 1992, à St Cyr l'Ecole

**Signatures métabolomiques associées au risque à long terme de  
cancers du sein et de la prostate et à l'alimentation dans la cohorte  
SU.VI.MAX**

Nouveaux horizons ouverts par la métabolomique appliquée à  
l'épidémiologie nutritionnelle

Thèse dirigée par :

**Madame la Docteure Mathilde Touvier**

**JURY :**

Madame la Docteure Alice Carrier  
Monsieur le Professeur Jean-Marie Bard  
Monsieur le Professeur Philippe Savarin  
Madame la Docteure Mathilde Touvier

Rapporteure  
Rapporteur  
Examineur  
Examinatrice/ Directrice de Thèse



## Remerciements

---

Je tiens tout d'abord à remercier la Docteure Mathilde Touvier, mon encadrante, de m'avoir donné l'occasion de réaliser cette thèse. Un grand merci de m'avoir confié ce beau projet qui correspondait tout à fait à mes attentes et d'avoir cru en mes capacités après quelques semaines de travail ensemble. Merci pour la confiance que tu as eue en moi et pour ton écoute. J'exprime également toute ma gratitude au Professeur Serge Hercberg et à Pilar Galan pour m'avoir accueilli chaleureusement au sein de leur Équipe de Recherche en Épidémiologie Nutritionnelle ainsi qu'au Professeur Dominique Ledoux au sein de l'école doctorale Galilée.

J'adresse ensuite mes sincères remerciements à la Docteure Alice Carrier et au Professeur Jean-Marie Bard d'avoir si gentiment accepté de prendre le temps de juger cette thèse en tant que rapporteurs ainsi qu'au Professeur Philippe Savarin d'avoir accepté chaleureusement d'évaluer ce travail. Je souhaite également remercier les membres de mon comité de suivi, Valentina Andreeva, Benjamin Alles et Pietro Ferrari, qui ont veillé au bon déroulement de ma thèse. Merci pour leur soutien, leur disponibilité et leurs encouragements.

Je remercie l'Institut National du Cancer d'avoir financé ma thèse ainsi que la Fondation de France, l'Institut National du Cancer et le l'Institut Fédératif de Recherche Biomédicale de l'Université Paris 13 d'avoir financé les projets sur lesquels portaient ma thèse. Merci également au réseau NACRe d'avoir labellisé ces projets et de m'avoir permis de présenter mes travaux lors des nombreux séminaires organisés.

Un immense merci à tous les partenaires des projets « Metabo-ks » et « Metabo-kp » et aux personnes de l'EREN pour leur contribution à ces travaux de recherche, qui a notamment permis leur valorisation sous forme d'articles scientifiques : Estelle Pujos-Guillot, Stéphanie Durand, Claudine Manach, Philippe Savarin, Marie-Paule Vasson, Céline Dalle, Agnès Victor Bala, Pierre Micheau, Adrien Rossary, Aicha Demidem, Mohamed Nawfal Triba, Nadia Bouchemal, Laurence Le Moyec, Mélanie Petera, Bernard Lyan, Marie Lagree, Delphine Centeno, Sophie Lefevre-Arbogast, Cécilia Samieri, Nada Assi, Pietro Ferrari, Vivian Viallon, Pilar Galan, Serge Hercberg, Paule Latino-Martel, Emmanuelle Kesse-Guyot, Nathalie Druesne-Pecollo, Valentin Partula, Mélanie Deschasaux, Bernard Srour, Thibault Fiolet, et bien sûr, Mathilde Touvier.

Un grand merci également à tous les membres de l'EREN et notamment aux chercheurs, à l'équipe informatique, à l'équipe médicale, aux data-managers, aux statisticiens, aux diététiciens, au personnel administratif et opérationnel ainsi qu'à tous les participants de l'étude SU.VI.MAX sans qui tous ces travaux n'auraient pas été possibles. Un remerciement particulier à Jagatjit pour son amitié, son immense disponibilité et pour avoir trouvé les solutions à tous les petits problèmes du quotidien et à Marie pour sa joie de vivre et ses bons petits plats.

Je tiens à remercier très chaleureusement tous les doctorants, anciens et nouveaux, et notamment Manon, Louise, Dan, Valentin, Eloi, Pauline, Bernard et Rolland, pour leur soutien, leurs conseils, leur esprit d'équipe, les rires et les débats (n'est-ce pas Dan !). Merci notamment à mes collègues de bureau qui m'ont supporté ! Un grand merci tout particulier à Manon, ma collègue des 400 coups depuis le premier jour de stage ! Merci pour ton soutien aussi bien professionnel que personnel, pour tous ces beaux moments, merci d'avoir répondu à toutes mes questions et de toujours avoir été là, notamment dans les moments difficiles. Je remercie également la team des « popins », Philippine, Claudia, Jagatjit, Fred,

Anouar et Camille, pour tous les supers moments que l'on a partagés ensemble et notamment des week-ends inoubliables.

Un immense merci à ma famille pour leur soutien inconditionnel et leurs encouragements, quelles que soient les circonstances, pour leur intérêt dans mon travail et leur bienveillance constante. Je remercie mes amis qui ont permis que ces 3 ans se déroulent au mieux en me faisant déconnecter du monde professionnel grâce à tous ces moments intenses, de joie, de rires, de danse. Merci pour leur enthousiasme, leur bonne humeur et leur affection. Mention spéciale à Lise qui, depuis le lycée, m'a toujours soutenu dans tous mes projets, a toujours été là, même à l'autre bout du monde. Merci pour tous ces moments de partage et de bonheur, merci de m'avoir guidée sur tous les chemins, sans toi je n'en serai pas là aussi bien sur le plan personnel que professionnel. Enfin, merci à Guillaume, pour tout... Merci de croire en moi quoi qu'il arrive, merci de me supporter au quotidien, d'être là à chaque moment, de m'accompagner à chaque étape et d'avoir réussi à gérer mon stress grâce à ta sagesse et ta zénitude ! C'est une chance inouïe de t'avoir à mes côtés.

## TABLE DES MATIERES

---

Remerciements .....	1
Liste des illustrations .....	7
Liste des tableaux .....	8
Liste des abbrévations .....	10
Liste des publications et communications .....	12
I. Publications originales dans des journaux à comité de lecture .....	12
A. Faisant l'objet du travail de thèse .....	12
B. Autres publications au cours de la thèse .....	12
II. Communications effectuées dans le cadre de congrès ou séminaires .....	13
A. Conférences invitées .....	13
B. Communications orales .....	14
C. Communications affichées (posters) .....	15
III. Autres productions ou communications scientifiques .....	17
A. Rapports .....	17
B. Communications grand public .....	17
Introduction générale .....	18
I. Contexte mondial et connaissances actuelles .....	18
A. Le cancer : une charge sociétale .....	18
B. La prévention primaire : un moyen efficace de lutter contre ce fardeau .....	20
C. La cancérogenèse .....	21
II. Nutrition et prévention des cancers du sein et de la prostate .....	23
A. Facteurs de risque et facteurs protecteurs des cancers du sein et de la prostate .....	23
B. Influence des facteurs alimentaires, de l'adiposité et de l'activité physique dans la cancérogenèse : aspects mécanistiques .....	28
III. Nouvelles perspectives pour l'épidémiologie nutritionnelle : l'exploration du métabolome humain	
30	
A. Intérêt des biomarqueurs pour compléter les études d'épidémiologie nutritionnelle .....	30
B. Le métabolome humain .....	31
C. Définitions et classifications des biomarqueurs .....	32
1. Les biomarqueurs d'expositions alimentaires .....	35
2. Les biomarqueurs d'effet ou « intermédiaires » .....	36
3. Les biomarqueurs de susceptibilité .....	37

D.	La métabolomique : un outil innovant pour l'identification de biomarqueurs à <i>grande échelle</i> .....	39
IV.	Etat de l'art des études ayant utilisé la métabolomique dans le champ de l'épidémiologie nutritionnelle.....	40
A.	Métabolomique et biomarqueurs nutritionnels .....	41
1.	Design des études.....	41
2.	Etat des connaissances.....	42
B.	Métabolomique et épidémiologie des cancers.....	51
C.	Analyses statistiques des données issues de la métabolomique .....	61
	Objectifs du travail de thèse.....	63
A.	Financements .....	63
B.	Objectifs.....	63
C.	Consortium multidisciplinaire .....	63
	Méthodologie générale .....	64
I.	Présentation et objectif général de l'étude SU.VI.MAX .....	64
II.	Collecte des données.....	64
A.	Caractéristiques sociodémographiques et de mode de vie .....	64
B.	Données alimentaires.....	65
C.	Données cliniques et biologiques.....	65
D.	Evénements de santé .....	65
III.	Sélection des populations d'études .....	66
A.	Etudes cas-témoin nichées (objectifs 1 et 3).....	66
1.	Pour le cancer du sein .....	66
2.	Pour le cancer de la prostate.....	67
B.	Etudes transversales (objectif 2) .....	68
1.	Chez les femmes.....	71
2.	Chez les hommes.....	71
IV.	Analyses métabolomiques des échantillons plasmatiques et prétraitements des données .....	72
A.	Principes généraux .....	72
1.	La RMN .....	72
2.	La spectrométrie de masse .....	73
B.	Analyses métabolomiques réalisées dans le cadre de cette thèse .....	74
1.	Analyses en RMN.....	74
2.	Analyses en spectrométrie de masse.....	75
V.	Estimation des expositions alimentaires.....	76

VI. Analyses statistiques .....	79
Objectif 1 - Identifier des signatures métabolomiques plasmatiques associées au risque à long terme de cancers du sein (objectif 1.I.) et de la prostate (objectif 1.II.) .....	84
I. Signature du risque de cancer du sein .....	84
A. Analyses en H-RMN .....	84
1. Méthodes .....	84
2. Résultats .....	85
B. Analyses en LC-MS.....	89
1. Méthodes .....	89
2. Résultats .....	90
II. Signature du risque de cancer de la prostate – analyses en H-RMN .....	94
A. Méthodes .....	94
B. Résultats .....	94
III. Discussion .....	98
Objectif 2: Identifier des signatures métabolomiques plasmatiques associées à l'alimentation dans un échantillon de femmes (objectif 2.I.) et d'hommes (objectif 2.II.) de la cohorte SU.VI.MAX.....	106
I. Signature de l'alimentation globale chez les femmes.....	106
A. Analyses en LC-MS.....	106
1. Méthodes .....	106
2. Résultats .....	108
3. Objectif secondaire pour les données de LC-MS : identifier de potentiels biomarqueurs de consommation d'aliments ou de groupes d'aliments spécifiques.....	115
B. Analyses en H-RMN .....	116
1. Méthodes .....	116
2. Résultats .....	116
II. Signature de l'alimentation globale chez les hommes – analyses en H-RMN .....	120
A. Méthodes .....	120
B. Résultats .....	121
III. Discussion .....	125
Objectif 3 - Identifier des signatures métabolomiques plasmatiques liées à la fois à l'alimentation et au risque à long terme de cancers du sein (objectif 3.I.) et de la prostate (objectif 3.II.) .....	133
I. Signature du risque de cancer du sein liée à l'alimentation .....	133
A. Analyses en LC-MS.....	133
1. Méthodes .....	133
2. Résultats .....	135

B.	Analyses en H-RMN .....	138
1.	Méthodes .....	138
2.	Résultats .....	138
II.	Signature du risque de cancer de la prostate liée à l'alimentation – analyses en H-RMN .....	138
A.	Méthodes .....	138
B.	Résultats .....	139
III.	Discussion .....	140
	Discussion générale .....	145
I.	Bilan des résultats .....	145
II.	Considérations méthodologiques .....	148
A.	Limites inhérentes aux analyses métabolomiques .....	148
B.	Limites inhérentes à la population et au design d'études .....	148
1.	Causalité .....	148
2.	Fiabilité des profils métabolomiques .....	149
3.	Aspects statistiques .....	150
4.	Potentiels biais épidémiologiques .....	150
III.	Perspectives de recherche .....	151
A.	Au sein du laboratoire .....	151
1.	Analyses supplémentaires dans le cadre du projet (données disponibles) .....	151
2.	Elargissement du projet .....	153
B.	Etudes complémentaires .....	154
C.	Impacts et applications futures .....	157
	Conclusion .....	159
	Références .....	160
	Annexes .....	186
	Résumé .....	427
	Abstract .....	428



## Liste des illustrations

---

Figure 1 Nombre estimé de nouveaux cas en 2018 sur l'ensemble de la population mondiale, tout cancer et tout âge confondus chez les femmes.....	18
Figure 2 Nombre estimé de nouveaux cas en 2018 sur l'ensemble de la population mondiale, tout cancer et tout âge confondus chez les hommes.....	18
Figure 3 Cartes mondiales présentant le type d'incidence du cancer le plus courant en 2018 dans chaque pays parmi les hommes et les femmes. ....	19
Figure 4 Le fardeau du cancer varie selon le niveau de l'Indice de développement humain. ....	20
Figure 5 Parts des cancers liées aux facteurs de risque nutritionnels. ....	21
Figure 6 Impacts des facteurs nutritionnels sur les mécanismes sous-jacents du cancer. ....	29
Figure 7 Composition et influences du métabolome humain.....	32
Figure 8 Diversité des interactions entre le système biologique et les variables intrinsèques du système et les variables environnementales environnantes.....	33
Figure 9 Voie causale de l'initiation à l'apparition de la maladie.....	37
Figure 10 Flow chart de l'étude cas-témoin nichée pour le cancer du sein.....	67
Figure 11 Flow chart de l'étude cas-témoin nichée pour le cancer de la prostate.....	68
Figure 12 Flow chart de l'étude transversale selon le score mPNNS-GS chez les femmes.....	71
Figure 13 Flow chart de l'étude transversale selon le score mPNNS-GS chez les hommes.....	72
Figure 14 Exemple de spectre H-RMN (séquence NOESY-1D) obtenu dans le cadre de cette thèse .....	73
Figure 15 Exemple de spectre de masse (validation par MS/MS) obtenu dans le cadre de cette thèse....	74
Figure 16 Procédure de la validation croisée simple pour le calcul de l'AUC .....	82
Figure 17 Procédure de la double validation croisée pour le calcul de l'AUC.....	83
Figure 18 Score plot de l'ACP réalisée par la plateforme RMN entre les cas de cancer du sein et les témoins .....	87
Figure 19 Score plot de l'ACP réalisée par la plateforme MS entre les cas de cancer du sein et les témoins pour les données en mode négatif.....	92
Figure 20 Hypothèses mécanistiques des variations métaboliques observées chez les femmes ayant développé un cancer du sein et les hommes ayant développé un cancer de la prostate au cours du suivi.. .....	105
Figure 21 Score plot de l'ACP réalisée par la plateforme MS entre les femmes du 1 <sup>er</sup> décile du mPNNS-GS et du 10 <sup>ème</sup> décile pour le protocole endogène en mode positif .....	110
Figure 22 Score plot de l'ACP réalisée par la plateforme RMN entre les femmes du 1 <sup>er</sup> décile du mPNNS-GS et du 10 <sup>ème</sup> décile .....	117
Figure 23 Score plot de l'ACP réalisée par la plateforme RMN entre les femmes du 1 <sup>er</sup> décile du mPNNS-GS et du 10 <sup>ème</sup> décile chez les femmes en surpoids.....	117
Figure 24 Fréquence de sélection $\geq 40\%$ des ions pour le risque de cancer du sein (% sur 1000 jeux de données ré-échantillonnés).....	137
Figure 25 Fréquence de sélection $\geq 40\%$ des signaux RMN pour le risque de cancer de la prostate (% sur 1000 jeux de données ré-échantillonnés) .....	140
Figure 26 Approche « Meeting in the Middle” .....	152
Figure 27 Synthèse des perspectives de recherche sur la thématique.....	158

## Liste des tableaux

---

Tableau 1 : Niveau de preuve concernant l'alimentation, la nutrition, l'activité physique et le risque de cancer du sein pré-ménopausique .....	25
Tableau 2 Niveau de preuve concernant l'alimentation, la nutrition, l'activité physique et le risque de cancer du sein post-ménopausique .....	26
Tableau 3 Niveau de preuve concernant l'alimentation, la nutrition, l'activité physique et le risque de cancer de la prostate .....	27
Tableau 4 Etudes épidémiologiques ayant investigué les profils métabolomiques associés à des régimes alimentaires.....	45
Tableau 5 Etudes épidémiologiques prospectives ayant investigué le lien entre profils métabolomiques pré-diagnostic et risque de cancer .....	54
Tableau 6 Composantes et système de notation pour le calcul du mPNNS-GS (recommandations de 2001) .....	70
Tableau 7 Liste des 74 groupes d'aliments .....	78
Tableau 8 Caractéristiques des cas de cancer du sein et des témoins pour les analyses en H-RMN – cohorte SU.VI.MAX, 1994-2007.....	86
Tableau 9 Métabolites associés au risque de cancer du sein issus des analyses H-RMN – cohorte SU.VI.MAX, 1994-2007 .....	88
Tableau 10 Associations entre les profils métabolomiques issus des analyses H-RMN obtenus par ACP et le risque de cancer du sein – cohorte SU.VI.MAX, 1994-2007.....	89
Tableau 11 Caractéristiques des cas de cancer du sein et des témoins pour les analyses en LC-MS – cohorte SU.VI.MAX, 1994-2007.....	91
Tableau 12 Métabolites identifiés et associés au risque de cancer du sein issus des analyses LC-MS – cohorte SU.VI.MAX, 1994-2007.....	92
Tableau 13 Associations entre les profils métabolomiques issus des analyses LC-MS obtenus par ACP et le risque de cancer du sein – cohorte SU.VI.MAX, 1994-2007.....	93
Tableau 14 Caractéristiques des cas de cancer de la prostate et des témoins – cohorte SU.VI.MAX, 1994-2007 .....	95
Tableau 15 Métabolites associés au risque de cancer de la prostate – cohorte SU.VI.MAX, 1994-2007... ..	96
Tableau 16 Associations entre les profils métabolomiques obtenus par ACP et le risque de cancer de la prostate – cohorte SU.VI.MAX, 1994-2007 .....	97
Tableau 17 Caractéristiques à l'inclusion des femmes du 1 <sup>er</sup> et du 10 <sup>ème</sup> déciles du mPNNS-GS – cohorte SU.VI.MAX, 1994-2007 .....	108
Tableau 18 Associations entre les ions et le niveau d'adéquation aux recommandations nutritionnelles mesuré par le mPNNS-GS chez les femmes – cohorte SU.VI.MAX, 1994-2007 .....	111
Tableau 19 Associations entre les ions et les composantes alimentaires du mPNNS-GS chez les femmes – cohorte SU.VI.MAX, 1994-2007.....	112
Tableau 20 Associations entre les ions et les profils alimentaires obtenus par ACP chez les femmes – cohorte SU.VI.MAX, 1994-2007.....	114
Tableau 21 Associations entre les ions identifiés et les groupes d'aliments spécifiques chez les femmes – cohorte SU.VI.MAX, 1994-2007.....	116
Tableau 22 Associations entre les signaux RMN et le niveau d'adéquation aux recommandations nutritionnelles mesuré par le mPNNS-GS chez les femmes – cohorte SU.VI.MAX, 1994-2007 .....	118

Tableau 23 Associations entre les signaux RMN et les composantes alimentaires du mPNNS-GS chez les femmes – cohorte SU.VI.MAX, 1994-2007.....	119
Tableau 24 Associations entre les signaux RMN et les profils alimentaires obtenus par ACP chez les femmes – cohorte SU.VI.MAX, 1994-2007.....	120
Tableau 25 Caractéristiques à l’inclusion des hommes du 1 <sup>e</sup> et du 10 <sup>ème</sup> déciles du mPNNS-GS – cohorte SU.VI.MAX, 1994-2007 .....	121
Tableau 26 Associations entre les signaux RMN et le niveau d’adéquation aux recommandations nutritionnelles mesuré par le mPNNS-GS chez les hommes – cohorte SU.VI.MAX, 1994-2007 .....	123
Tableau 27 Associations entre les signaux RMN et les composantes alimentaires du mPNNS-GS chez les hommes – cohorte SU.VI.MAX, 1994-2007.....	124
Tableau 28 Associations entre les signaux RMN et les profils alimentaires obtenus par ACP chez les hommes – cohorte SU.VI.MAX, 1994-2007.....	125
Tableau 29 Associations entre les 14 ions sélectionnés par Elastic Net et le risque de cancer du sein – cohorte SU.VI.MAX, 1994-2007.....	136
Tableau 30 Associations entre les 3 signaux RMN sélectionnés par Elastic Net et le risque de cancer de la prostate – cohorte SU.VI.MAX, 1994-2007 .....	139
Tableau 31 Résumé des résultats et éléments de discussion .....	146

## Liste des abbrévations

---

2DG : 2-deoxy-D-glucose  
AAA : Acides Aminés Aromatiques  
ACP : Analyse en Composante Principale  
AICR : “American Institute for Cancer Research”  
AIP : Acide 3-indolepropionique  
aMED : “alternate Mediterranean Diet Score”  
ANCOVA : Analyse de la Covariance  
ANOVA : Analyse de la Variance  
AP : Acide pipécolique  
ATBC : AcetylTributylCitrate  
AUC : “Area Under the Curve” (Aire sous la courbe)  
BCAA : Acides aminé ramifiés  
BSD : “Baltic Sea Diet”  
CPMG : “Car-Purcell-Meiboom-Gill”  
CUP : “Continuous Update Project”  
DHEA : Dehydroepiandrosterone  
DQI-I : “Diet Quality Index-International”  
ECMDB : “E. coli Metabolome Database”  
EFSA : “European Food Safety Authority”  
ER +/- : Récepteurs aux oestrogènes positifs/négatifs  
FDR : False Discovery Rate (Taux de fausses découvertes)  
FFQ : Questionnaire de fréquence alimentaire  
FoodDB : “Food Component Database”  
FWER : “Family Wise Error Rate”  
GC : Chromatographie en phase gazeuse  
GGT :  $\gamma$ -glutamyl-threonine  
GWAS : “Genome Wide Association Study”  
HCC : Carcinome hépatocellulaire  
HDI : “Healthy Diet Indicator”  
HDL : “High Density Lipoprotein”  
HEI : “Healthy Eating Index”  
HMDB : “Human Metabolome Database”  
IC : Intervalle de Confiance  
ICM : Intervention Court ou Moyen terme  
IDH : Indice de Développement Humain  
IFRB : Institut Fédératif de Recherche Biomédicale  
IGF-1 : “Insulin-like growth factor-1”  
IL : Intervention Long terme  
IMC : Indice de Masse Corporelle  
INCa : Institut National du Cancer  
LASSO : “Least Absolute Shrinkage and Selection Operator”  
LDL : “Low Density Lipoprotein”  
LysoPC : Lysophosphatidylcholine

mPNNS-GS : Modified Programme National Nutrition Santé- Guideline Score  
MS : Spectrométrie de masse  
MWAS : “Metabolome Wide Association Study”  
NOESY1D : “One-dimensional 1H nuclear overhauser effect spectroscopy”  
OMS : Organisation Mondiale de la Santé  
O-PLS-DA : Orthogonal Partial Least Square Discriminant Analysis  
OR : Odds Ratios  
OT : Etude d’Observation Transversale  
PC : Phosphatidylcholine  
PHI : “Prostate Health Index”  
PR +/- : Récepteurs à la progesterone positifs/négatifs  
PSA : Antigène prostatique spécifique de la prostate  
RMN : Résonance Magnétique Nucléaire  
SNP : “Single-nucleotide polymorphism” (polymorphismes d’un seul nucleotide)  
SU.VI.MAX : SUplémentation en Vitamines et Minéraux AntioXydants  
SXR : “Steroid and xenobiotic sensing nuclear receptor”  
T : Tertile  
TMAO : Oxyde de triméthylamine  
TOCSY : “Total Correlation Spectroscopy”  
UPLC : Chromatographie liquide à ultra-haute performance  
WCRF : “World Cancer Research Fund”  
YMDB : “Yeast Metabolome Database”

## Liste des publications et communications

---

### I. Publications originales dans des journaux à comité de lecture

#### A. Faisant l'objet du travail de thèse

**Lécuyer L**, Victor Bala A, Deschasaux M, Bouchemal N, Nawfal Triba M, Vasson MP, Rossary A, Demidem A, Galan P, Hercberg S, Partula V, Le Moyec L, Srour B, Fiolet T, Latino-Martel P, Kesse-Guyot E, Savarin P, Touvier M. NMR metabolomic signatures reveal predictive plasma metabolites associated with long-term risk of developing breast cancer. **Int J Epidemiol.** 2018 Apr 1;47(2):484-494. doi: 10.1093/ije/dyx271.

**Lécuyer L**, Dalle C, Lyan B, Demidem A, Rossary A, Vasson MP, Petera M, Lagree M, Ferreira T, Centeno D, Galan P, Hercberg S, Deschasaux M, Partula V, Srour B, Latino-Martel P, Kesse-Guyot E, Druesne-Pecollo N, Durand S, Pujos-Guillot E, Touvier M. Plasma metabolomic signatures associated with long-term breast cancer risk in the SU.VI.MAX prospective cohort. **Cancer Epidemiol Biomarkers Prev.** 2019 Jun 4. pii: cebp.0154.2019. doi: 10.1158/1055-9965.EPI-19-0154.

#### Publications soumises ou en préparation :

**Lécuyer L\***, Dalle C\*, Micheau P, Pétéra M, Centeno D, Lyan B, Lagree M, Galan P, Hercberg S, Rossary A, Demidem A, Vasson MP, Partula V, Deschasaux M, Srour B, Latino-Martel P, Druesne-Pecollo N, Kesse-Guyot E, Durand S, Pujos-Guillot E, Manach C#, Touvier M#. \*,#: contributions égales. Untargeted plasma metabolomic profiles associated with overall diet in women from the SU.VI.MAX cohort. **Soumise**

**Lécuyer L**, Dalle C, Lefevre-Arbogast S, Micheau P, Lyan B, Rossary A, Demidem A, Petera M, Lagree M, Centeno D, Galan P, Hercberg S, Samieri C, Assi N, Ferrari P, Viallon V, Deschasaux M, Partula V, Srour B, Latino-Martel P, Kesse-Guyot E, Druesne-Pecollo N, Vasson MP, Durand S, Pujos-Guillot E, Manach C, Touvier M. Diet-related metabolomic signature of long-term breast cancer risk using penalized regression: an exploratory study in the SU.VI.MAX cohort. **Soumise**

**Lécuyer L**, Victor Bala A, Demidem A, Rossary A, Bouchemal N, Nawfal Triba M, Galan P, Hercberg S, Deschasaux M, Partula V, Le Moyec L, Srour B, Latino-Martel P, Kesse-Guyot E, Druesne-Pecollo N, Vasson MP, Savarin P, Touvier M. Untargeted metabolomic profiles associated with long-term risk of prostate cancer risk. **En préparation**

Dalle C\*, **Lécuyer L\***, Micheau P, Pétéra M, Centeno D, Lyan B, Lagree M, Galan P, Hercberg S, Rossary A, Demidem A, Vasson MP, Partula V, Deschasaux M, Srour B, Latino-Martel P, Druesne-Pecollo N, Kesse-Guyot E, Durand S, Pujos-Guillot E, Touvier M#, Manach C#. \*,#: contributions égales. Metabolomics applied to nutritional epidemiology to identify plasma biomarkers of food intake in women from the SU.VI.MAX cohort. **En préparation**

#### B. Autres publications au cours de la thèse

Egnell M, Fassier P, **Lécuyer L**, Gonzalez R, Zelek L, Vasson MP, Hercberg S, Latino-Martel P, Galan P, Druesne-Pecollo N, Deschasaux M, Touvier M. Antioxidant intake from diet and supplements and risk of digestive cancers in middle-aged adults: results from the prospective NutriNet-Santé cohort. **Br J Nutr.** 2017 Oct;118(7):541-549. doi: 10.1017/S0007114517002392. Epub 2017 Sep 20.

Fassier P, Zelek L, **Lécuyer L**, Bachmann P, Touillaud M, Druesne-Pecollo N, Galan P, Cohen P, Hoarau H, Latino-Martel P, Kesse-Guyot E, Baudry J, Hercberg S, Deschasaux M, Touvier M. Modifications in dietary

and alcohol intakes between before and after cancer diagnosis: Results from the prospective population-based NutriNet-Santé cohort. *Int J Cancer*. 2017 Aug 1;141(3):457-470. doi: 10.1002/ijc.30704. Epub 2017 May 15.

Deschasaux M, Julia C, Kesse-Guyot E, **Lécuyer L**, Adriouch S, Méjean C, Ducrot P, Péneau S, Latino-Martel P, Fezeu LK, Fassier P, Hercberg S, Touvier M. Are self-reported unhealthy food choices associated with an increased risk of breast cancer? Prospective cohort study using the British Food Standards Agency nutrient profiling system. *BMJ Open*. 2017 Jun 8;7(6):e013718. doi: 10.1136/bmjopen-2016-013718.

Egnell M, Fassier P, **Lécuyer L**, Zelek L, Vasson MP, Hercberg S, Latino-Martel P, Galan P, Deschasaux M, Touvier M. B-Vitamin Intake from Diet and Supplements and Breast Cancer Risk in Middle-Aged Women: Results from the Prospective NutriNet-Santé Cohort. *Nutrients*. 2017 May 13;9(5). pii: E488. doi: 10.3390/nu9050488.

Deschasaux M, Souberbielle JC, Partula V, **Lécuyer L**, Gonzalez R, Srour B, Guinot C, Malvy D, Latino-Martel P, Druesne-Pecollo N, Galan P, Hercberg S, Kesse-Guyot E, Fassier P, Ezzedine K, Touvier M. What Do People Know and Believe about Vitamin D? *Nutrients*. 2016 Nov 11;8(11). pii: E718.

#### **Publications soumises ou en préparation :**

Partula V, Mondot S, Victor Bala A, Bouchemal N, **Lécuyer L**, Bodin-Dubigeon C, Torres M, Kesse-Guyot E, Deschasaux M, Assmann K, Latino-Martel P, Buscail C, Julia C, Galan P, Hercberg S, Quintana-Murci L, Albert ML, Duffy D, Lantz O, Savarin P, Nawfal Triba M, Touvier M, Milieu Intérieur Consortium. Associations between untargeted plasma metabolomic signatures and gut microbiota composition in the Milieu Intérieur population of healthy adults. **Soumise**

#### II. Communications effectuées dans le cadre de congrès ou séminaires

##### **A. Conférences invitées**

Touvier M, **Lécuyer L**.

*3rd edition of Cancer and Metabolism symposium* (Marseille April 1st-2nd 2019): "Plasma metabolomic signatures associated with long-term breast cancer risk in the SU.VI.MAX prospective cohort".

Srour B, **Lécuyer L**, Touvier M.

*Symposium CRNH Journées Francophones de Nutrition* (Nice 28-30 novembre 2018) ; «Signatures métabolomiques par spectrométrie de masse associées à des profils alimentaires spécifiques et au risque à long terme du cancer du sein dans la cohorte SU.VI.MAX (CRNH IdF) »

**Lécuyer L**, Victor Bala A, Deschasaux M, Bouchemal N, Nawfal Triba M, Vasson M-P, Rossary A, Demidem A, Lyan B, Pétéra M, Durand S, Lagrée M, Centeno D, Galan P, Hercberg S, Partula V, Le Moyec L, Srour B, Fiolet T, Latino-Martel P, Kesse-Guyot E, Manach C, Dalle C, Pujos-Guillot E, Savarin P, Touvier M.

*Journée NACRe « Partenariat de recherche nutrition et cancer »* (Paris, 17 mars 2017): « Découverte de nouveaux biomarqueurs prédictifs du risque de cancer du sein et de la prostate : Nouveaux horizons ouverts par la métabolomique appliquée à l'épidémiologie nutritionnelle »

Touvier M, **Lécuyer L**.

*Symposium CRNH Journées Francophones de Nutrition* (Montpellier 30 novembre – 2 décembre 2016) ; « Découverte de nouveaux biomarqueurs du risque de pathologies chroniques par une approche

métabolomique appliquée à l'épidémiologie nutritionnelle : exemples de projets menés au sein de l'Équipe de Recherche en Épidémiologie Nutritionnelle (CRNH IdF) »

## B. Communications orales

**Lécuyer L**, Touvier M, Groupe de travail Metabo-ks, Metabo-kp

*Journée de l'école doctorale Galilée* (Paris, Juin 2018) : Découverte de nouveaux biomarqueurs du risque à long terme du cancer du sein et de la prostate : Nouveaux horizons ouverts par la métabolomique appliquée à l'épidémiologie nutritionnelle

**Lécuyer L**, Touvier M, Groupe de travail Metabo-ks, Metabo-kp

*Journée des doctorants Centre de Recherche en Epidémiologie et Biostatistiques Sorbonne Paris Cité* (Paris, 7 Mai 2018): « Discovery of new long-term biomarkers of breast and prostate cancer risks : new insights from metabolomics applied to nutritional epidemiology »

**Lécuyer L**, Dalle C, Lyan B, Pétéra M, Lagree M, Rossary A, Demidem A, Ferreira T, Centeno D, Galan P, Hercberg S, Deschasaux M, Partula V, Srouf B, Latino-Martel P, Kesse-Guyot E, Manach C, Vasson MP, Durand S, Pujos-Guillot E, Touvier M

*Journées Francophones de Nutrition* (Nantes, 13-15 décembre 2017): « Signatures métabolomiques par spectrométrie de masse et risque de cancer du sein. » (meilleur poster en 90 sec)

**Lécuyer L**, Dalle C, Lyan B, Pétéra M, Lagree M, Rossary A, Demidem A, Ferreira T, Centeno D, Galan P, Hercberg S, Deschasaux M, Partula V, Srouf B, Latino-Martel P, Kesse-Guyot E, Manach C, Vasson MP, Durand S, Pujos-Guillot E, Touvier M

*Colloque NACRe "Cancer : Les actualités de la recherche en nutrition"* (Paris, 30 novembre et 1er décembre 2017) : « Signatures métabolomiques par spectrométrie de masse et risque de cancer du sein. »

**Lécuyer L**, Victor Bala A, Deschasaux M, Bouchemal N, Triba M, Vasson M-P, Rossary A, Demidem A, Galan P, Hercberg S, Partula V, Le Moyec L, Srouf B, Fiolet T, Latino-Martel P, Kesse-Guyot E, Savarin P, Touvier M

*IUNS 21st International Congress of Nutrition (ICN) 2017* (Buenos Aires, 15-20 octobre 2017): « NMR metabolomic signatures reveal predictive plasma metabolites associated with long-term risk of developing breast cancer»

**Lécuyer L**, Victor Bala A, Deschasaux M, Bouchemal N, Triba M, Vasson M-P, Rossary A, Demidem A, Galan P, Hercberg S, Partula V, Le Moyec L, Srouf B, Fiolet T, Latino-Martel P, Kesse-Guyot E, Savarin P, Touvier M, (on behalf of the Metabo-Ks Working Group)

*Congress of Translational Research in Human Nutrition 2017* (Clermont-Ferrand, 22-23 juin 2017): « NMR metabolomic signatures reveal predictive plasma metabolites associated with long-term risk of developing breast cancer»

**Lécuyer L**, Victor Bala A, Deschasaux M, Bouchemal N, Triba M, Hercberg S, Galan P, Kesse-Guyot E, Latino-Martel P, Fezeu LK, Savarin P, Touvier M.

*Journées Francophones de Nutrition* (Montpellier, 29 Novembre-2 Décembre 2016): « Découverte de métabolites prédictifs du risque de cancer du sein : approche métabolomique RMN appliquée à



l'épidémiologie nutritionnelle ». Abstract in Nutrition clinique et métabolisme, Volume 31, Issue 3, September 2017, Page 226

### C. Communications affichées (posters)

**Lécuyer L**, Dalle C, Lyan B, Demidem A, Rossary A, Vasson MP, Petera M, Lagree M, Ferreira T, Centeno D, Zelek L, Galan P, Hercberg S, Deschasaux M, Partula V, Srouf B, Latino-Martel P, Kesse-Guyot E, Druetne-Pecollo N, Manach C, Durand S, Pujos-Guillot E, Touvier M

San Antonio Breast Cancer Symposium (4-8 décembre 2018): "Plasma metabolomic signatures associated with long-term breast cancer risk in the SU.VI.MAX prospective cohort". AACR; Cancer Res 2019;79(4 Suppl) :Abstract nr P4-10-17.

**Lécuyer L\***, Victor Bala A, Bouchemal N, Nawfal Triba M, Demidem A, Rossary A, Galan P, Hercberg S, Partula V, Le Moyec L, Latino-Martel P, Kesse-Guyot E, Deschasaux M, Vasson MP, Savarin P, Touvier M  
*Journées Francophones de Nutrition* (Nice, 28-30 novembre 2018): « Association entre profils métabolomiques plasmatiques par RMN et risque à long terme de développer un cancer de la prostate. » Abstract in Nutrition Clinique et Métabolisme Vol 33, N° 1, mars 2019, P. 1-114

**Lécuyer L\***, Dalle C\*, Micheau P, Pétéra M, Centeno D, Lyan B, Morand C, Galan P, Hercberg S, Rossary A, Demidem A, Vasson MP, Partula V, Deschasaux M, Srouf B, Latino-Martel P, Kesse-Guyot E, Durand S, Pujos-Guillot E, Manach C#, Touvier M#. \* et # : contributions égales  
*Journées Francophones de Nutrition* (Nice, 28-30 novembre 2018): « Signatures métabolomiques associés à des profils alimentaires spécifiques dans la cohorte SU.VI.MAX ». Abstract in Nutrition Clinique et Métabolisme Vol 33, N° 1, mars 2019, P. 1-114

**Lécuyer L**, Touvier M, Groupe de travail Metabo-ks, Metabo-kp  
*Journée de l'école doctorale Galilée* (Paris, Juin 2018) : Découverte de nouveaux biomarqueurs du risque à long terme du cancer du sein et de la prostate : Nouveaux horizons ouverts par la métabolomique appliquée à l'épidémiologie nutritionnelle

**Lécuyer L\***, Victor Bala A\*, Bouchemal N, Triba M, Rossary A, Demidem A, Galan P, Hercberg S, Partula V, Le Moyec L, Srouf B, Latino-Martel P, Kesse-Guyot E, Deschasaux M, Vasson M-P, Savarin P#, Touvier M#. \* et # : contributions égales  
*11èmes journées scientifiques du RFMF* (Réseau Francophone de Métabolomique et Fluxomique) (Liège, 23-25 mai 2018) : "NMR metabolomic signatures reveal predictive plasma metabolites associated with long-term risk of developing prostate cancer"

**Lécuyer L\***, Dalle C\*, Petera M, Centeno D, Lyan B, Durand S, Pujos-Guillot E, Micheau P, Morand C, Galan P, Hercberg S, Partula V, Deschasaux M, Srouf B, Latino-Martel P, Kesse-Guyot E, Touvier M#, Manach C#. \* et # : contributions égales  
*Journées Francophones de Nutrition* (Nantes, 13-15 décembre 2017): « Identification de biomarqueurs nutritionnels par une approche métabolomique en spectrométrie de masse ». Abstract in Nutrition Clinique et Métabolisme, Volume 32, Issue 4, November 2018, Pages 335-336

**Lécuyer L**, Dalle C, Lyan B, Pétéra M, Lagree M, Rossary A, Demidem A, Ferreira T, Centeno D, Galan P, Hercberg S, Deschasaux M, Partula V, Srouf B, Latino-Martel P, Kesse-Guyot E, Manach C, Vasson MP, Durand S, Pujos-Guillot E, Touvier M

*Journées Francophones de Nutrition* (Nantes, 13-15 décembre 2017): « Signatures métabolomiques par spectrométrie de masse et risque de cancer du sein. » Abstract in *Nutrition Clinique et Métabolisme*, Volume 32, Issue 4, November 2018, Pages 325-326

**Lécuyer L**, Victor Bala A, Deschasaux M, Bouchemal N, Nawfal Triba M, Vasson M-P, Rossary A, Demidem A, Galan P, Hercberg S, Partula V, Le Moyec L, Srouf B, Fiolet T, Latino-Martel P, Kesse-Guyot E, Zelek L, Savarin P, Touvier M

*San Antonio Breast Cancer Symposium* (5-9 décembre 2017): “NMR metabolomic signatures reveal predictive plasma metabolites associated with long-term risk of developing breast cancer”. *AACR; Cancer Res* 2018;78(4 Suppl):Abstract nr P1-02-01.

**Lécuyer L**, Victor Bala A, Deschasaux M, Bouchemal N, Triba M, Vasson M-P, Rossary A, Demidem A, Galan P, Hercberg S, Partula V, Le Moyec L, Srouf B, Fiolet T, Latino-Martel P, Kesse-Guyot E, Savarin P, Touvier M

*International Union of Nutrition Societies (IUNS) 2017* (Buenos Aires, Argentina, 16-20 Octobre 2017): “NMR metabolomic signatures reveal predictive plasma metabolites associated with long-term risk of developing breast cancer”. Abstract in *Annals of Nutrition and Metabolism* 2017;71(suppl 2): P 1–1433 (<https://doi.org/10.1159/000480486>)

**Lécuyer L\***, Dalle C\*, Pétéra M, Centeno D, Lyan B, Durand S, Pujos-Guillot E, Micheau P, Morand C, Galan P, Hercberg S, Partula V, Deschasaux M, Srouf B, Latino-Martel P, Kesse-Guyot E, Touvier M#, Manach C#. \* and # : equal contribution

*Congrès SMMAP 2017 (Spectrométrie de Masse, Métabolomique et Analyse Protéomique)* (Marne-la-Vallée-Chessy, 3-5 octobre 2017) : “ Metabolomics applied to nutritional epidemiology to identify biomarkers of food intake in the framework of the Metabo-Breast cancer project, SU.VI.MAX cohort”

**Lécuyer L\***, Victor Bala A\*, Deschasaux M, Bouchemal N, Triba M, Vasson M-P, Rossary A, Demidem A, Galan P, Hercberg S, Partula V, Le Moyec L, Srouf B, Fiolet T, Latino-Martel P, Kesse-Guyot E, Touvier M#, Savarin P#

\* et # : contributions égales

*Congrès SMMAP 2017 (Spectrométrie de Masse, Métabolomique et Analyse Protéomique)* (Marne-la-Vallée-Chessy, 3-5 octobre 2017) : “ Etude sur l’identification de biomarqueurs prédictifs du risque du cancer du sein par RMN métabolomique. “

**Lécuyer L\***, Dalle C\*, Pétéra M, Centeno D, Lyan B, Durand S, Pujos-Guillot E, Micheau P, Morand C, Galan P, Hercberg S, Partula V, Deschasaux M, Srouf B, Latino-Martel P, Kesse-Guyot E, Touvier M#, Manach C#. \* and # : equal contribution

*Congress of Translational Research in Human Nutrition 2017* (Clermont-Ferrand, 22-23 juin 2017): “ Metabolomics applied to nutritional epidemiology to identify biomarkers of food intake in the framework of the Metabo-Breast cancer project, SU.VI.MAX cohort”

**Lécuyer L**, Victor Bala A, Deschasaux M, Savarin P, Touvier M, on behalf of the Metabo-Ks working group. *25th AICR Research Conference* (North Bethesda, 14-16 Novembre 2016): "NMR metabolomic analysis of breast cancer risk in the SU.VI.MAX prospective cohort study"

### III. Autres productions ou communications scientifiques

#### **A. Rapports**

- Projet INCa. Rapport scientifique d'activité final 2019 "Découverte de nouveaux biomarqueurs prédictifs du risque de cancer du sein : nouveaux horizons ouverts par la métabolomique appliquée à l'épidémiologie nutritionnelle" (INCA\_8085)
- Projet Fondation de France. Rapport scientifique d'activité intermédiaire 2018 "Découverte de nouveaux biomarqueurs prédictifs du risque de cancer de la prostate : nouveaux horizons ouverts par la métabolomique appliquée à l'épidémiologie nutritionnelle" (N°2015 00060743)
- Projet Fondation de France. Rapport scientifique d'activité intermédiaire 2017 "Découverte de nouveaux biomarqueurs prédictifs du risque de cancer de la prostate : nouveaux horizons ouverts par la métabolomique appliquée à l'épidémiologie nutritionnelle" (N°2015 00060743)
- Projet INCa. Rapport scientifique d'activité intermédiaire 2017 "Découverte de nouveaux biomarqueurs prédictifs du risque de cancer du sein : nouveaux horizons ouverts par la métabolomique appliquée à l'épidémiologie nutritionnelle" (INCA\_8085)

#### **B. Communications grand public**

Conférence inter-masters (innovation marketing) 2017 " Que mangerons-nous demain ? ". Table ronde sur les tendances des régimes végétariens. Université Paris Sud, campus de Sceaux 2017.

## Introduction générale

### I. Contexte mondial et connaissances actuelles

#### A. Le cancer : une charge sociétale

Etant l'une des principales causes de mortalité et morbidité dans le monde, le cancer est devenu l'une des principales maladies chroniques et un véritable fardeau mondial (1). En 2018, 18.1 millions de nouveaux cas de cancer étaient diagnostiqués dans le monde, et d'ici 2040 ce chiffre devrait atteindre près de 29.5 millions (2). Environ 38.4% des hommes et des femmes auront un cancer diagnostiqué au cours de leur vie (selon les données de 2013–2015) (3). En termes de nombre de nouveaux cas de cancer diagnostiqués en 2018 (incidence), sur l'ensemble de la population mondiale, tous âges et tous cancers confondus, le cancer du sein est le premier chez les femmes, avec 2 088 849 nouveaux cas et le celui de la prostate est le deuxième chez les hommes avec 1 276 106 nouveaux cas (Figure 1 et Figure 2 respectivement).

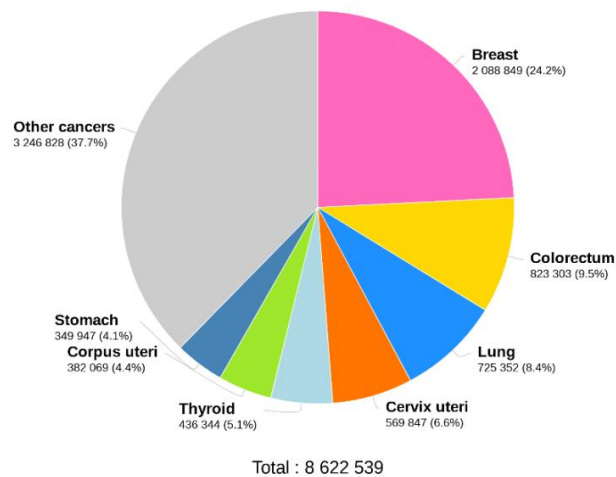


Figure 1 Nombre estimé de nouveaux cas en 2018 sur l'ensemble de la population mondiale, tous cancers et tous âges confondus chez les femmes.

Source : GLOBOCAN 2018 (2)

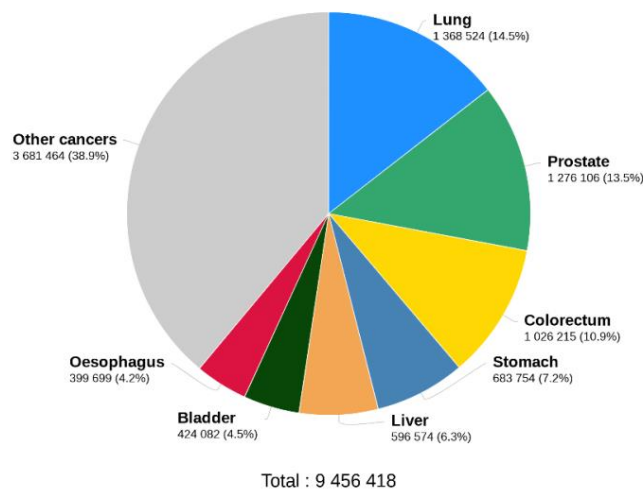


Figure 2 Nombre estimé de nouveaux cas en 2018 sur l'ensemble de la population mondiale, tous cancers et tous âges confondus chez les hommes.

Source : GLOBOCAN 2018 (2)

La Figure 3 montre les cancers les plus souvent diagnostiqués dans le monde par pays en 2018, chez l'homme (Figure 3.A) et chez la femme (Figure 3.B). Chez les hommes, le cancer de la prostate est le plus fréquemment diagnostiqué dans 105 pays, suivi du cancer du poumon (37 pays) et du cancer du foie (13 pays). Chez les femmes, le cancer du sein est de loin le plus souvent diagnostiqué dans la majorité des pays (154 pays) (2).

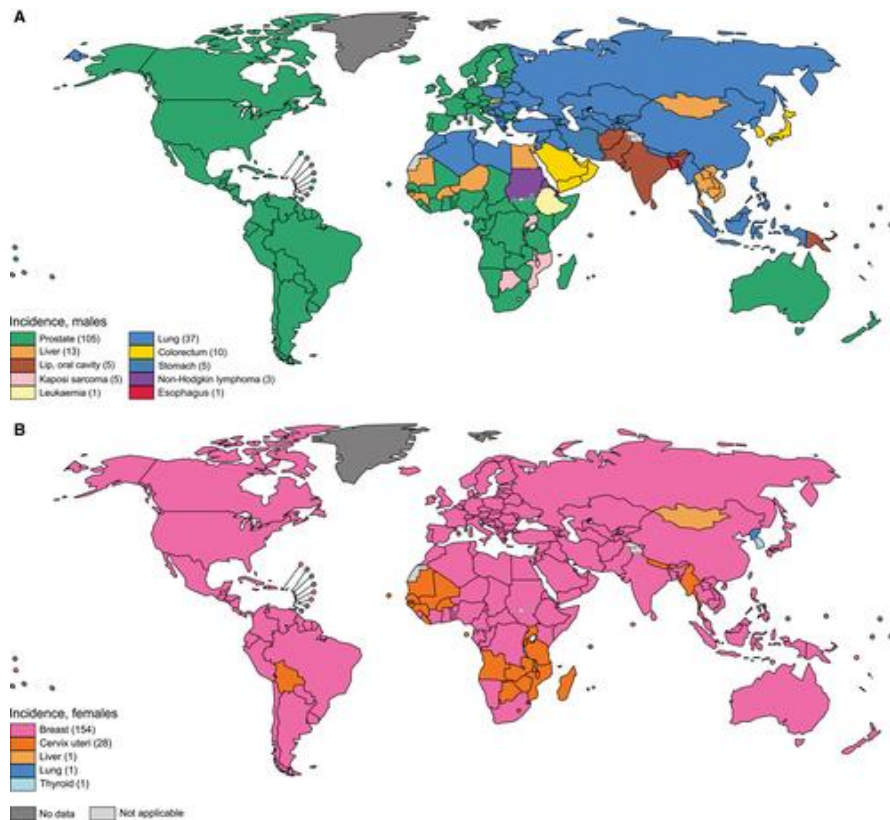


Figure 3 Cartes mondiales présentant le type d'incidence du cancer le plus courant en 2018 dans chaque pays parmi (A) les hommes et (B) les femmes. Le nombre de pays représentés dans chaque groupe est inclus dans la légende. Source: GLOBOCAN 2018 (2).

En examinant la répartition géographique du fardeau mondial du cancer, la diversité est assez frappante (Figure 4). Les cancers les plus fréquents dans les pays à revenu faible et intermédiaire sont liés en particulier à des infections, tels que le cancer du col de l'utérus et d'autres cancers liés au papillomavirus humain, le cancer de l'estomac et le cancer du foie (1). Dans les pays ayant un indice de développement humain (IDH) élevé ou très élevé, les cancers les plus répandus sont le sein, la prostate, le poumon, l'estomac et le côlon-rectum.

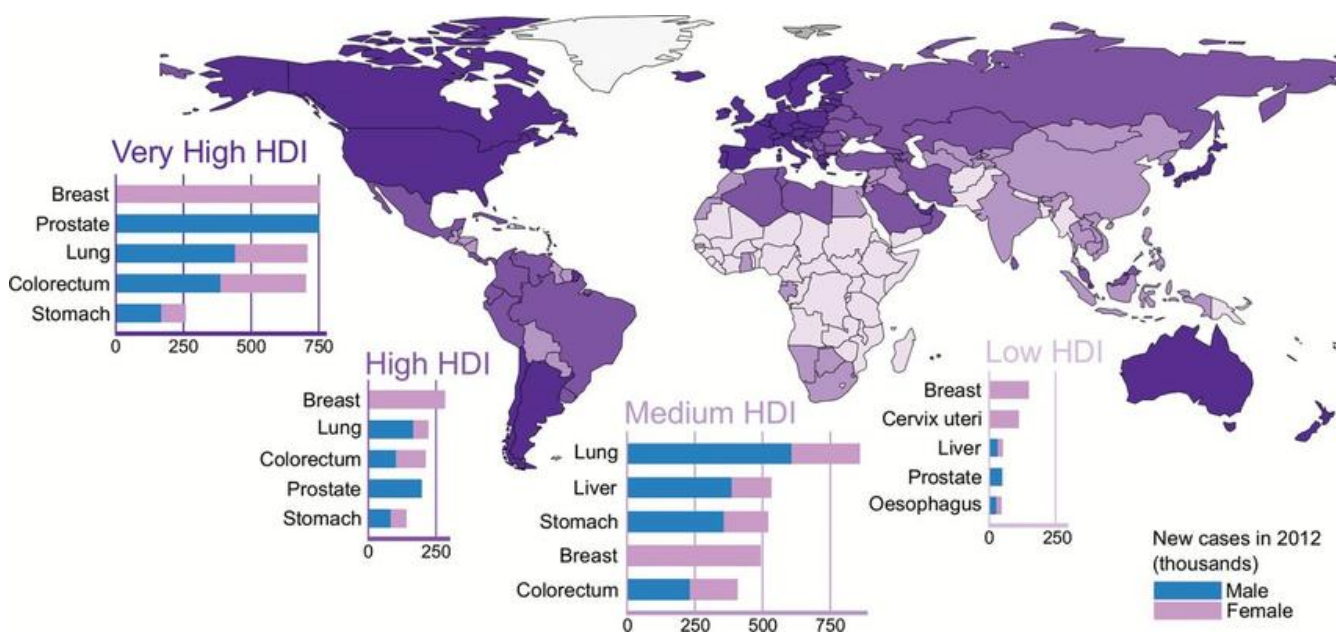


Figure 4 Le fardeau du cancer varie selon le niveau de l'Indice de développement humain.

Source: Cancer Atlas, American Cancer Society, Atlanta, 2014 (4). HDI = Human Development Index (Indice de développement humain).

## B. La prévention primaire : un moyen efficace de lutter contre ce fardeau

Ces variations géographiques indiquent à quel point les changements sociétaux, économiques et de mode de vie impactent le profil de ce groupe de maladies extrêmement complexe. La culture occidentale basée sur la consommation d'alcool, d'aliments ultra-transformés, une faible activité physique et une augmentation de la sédentarité, très marquée dans les pays ayant un IDH élevé et très élevé se répand désormais également dans les pays à revenu faible et intermédiaire.

Le cancer est une maladie multifactorielle (5) et bien qu'une grande partie des cancers soit expliquée par des facteurs intrinsèques comme des erreurs de réplication de l'ADN, entre 30 et 50% des cancers pourraient être évités dans le monde si les connaissances accumulées sur les causes étaient traduites en prévention primaire efficace (6). Plusieurs études (1,4,7) et plusieurs agences (dont l'Organisation Mondiale de la Santé (8), le Centre de Recherche International contre le Cancer (9), le World Cancer Research Fund (WCRF), et l'American Institute for Cancer Research (AICR) (10)) montrent que la prévention primaire serait un moyen particulièrement efficace de prévenir un large spectre de cancers. La prévention primaire du cancer se traduit par une exposition réduite aux substances cancérigènes reconnues et/ou une augmentation de la promotion de la santé, tandis que la prévention secondaire est illustrée par le dépistage, la détection plus précoce et la chimioprévention.

Le Centre International de Recherche sur le Cancer a réalisé une estimation de la proportion des cancers liés aux principaux facteurs de risque en France métropolitaine en 2015 (11). Pour un cancer donné, la part attribuable est calculée à partir de la proportion de la population exposée durant un temps défini au facteur de risque, de l'estimation du risque de cancer associé au facteur de risque et du nombre de cas observés durant un temps défini dans la population, en considérant un temps de latence entre l'exposition et la survenue de la maladie (12). Au total, 13 facteurs considérés comme (probablement ou certainement) cancérigènes pour l'homme ont été étudiés : le tabagisme, l'alcool, l'alimentation, le surpoids et l'obésité, une activité physique insuffisante, l'utilisation d'hormones exogènes, les infections, les radiations

ionisantes, la pollution atmosphérique, le rayonnement solaire, les expositions professionnelles, une durée d’allaitement de moins de 6 mois, et l’exposition aux substances chimiques de la population. Les conclusions de ce rapport ont montré qu’en France, 40 % des nouveaux cas de cancers observés en 2015 étaient liés au mode de vie et à l’environnement. Parmi les quatre principales causes de cancer évitables, trois concernaient des facteurs nutritionnels : le tabac (en tête avec 20 % des cancers attribuables), puis la consommation d’alcool avec 8 %, et une alimentation déséquilibrée (notamment une faible consommation de fruit et de fibres alimentaires) (13) ainsi que le surpoids et l’obésité avec 5,4 % chacun (voir Figure 5). Au total en France, 16 % des nouveaux cas de cancer chez les hommes et 20 % chez les femmes étaient attribuables aux facteurs nutritionnels dans leur ensemble (alimentation, surpoids et obésité, activité physique et alcool). En outre, environ 3000 nouveaux cas de cancer de la prostate en 2015 étaient attribuables à une consommation élevée de produits laitiers (plus de 2 portions/jours) et plus de 4000 nouveaux cas de cancer du sein étaient attribuables à une consommation faible de fibres et produits laitiers (moins de 25g/jour et de 2 portions/jour respectivement) et une consommation élevée de viande rouge (plus de 42,9g/jour) (13).

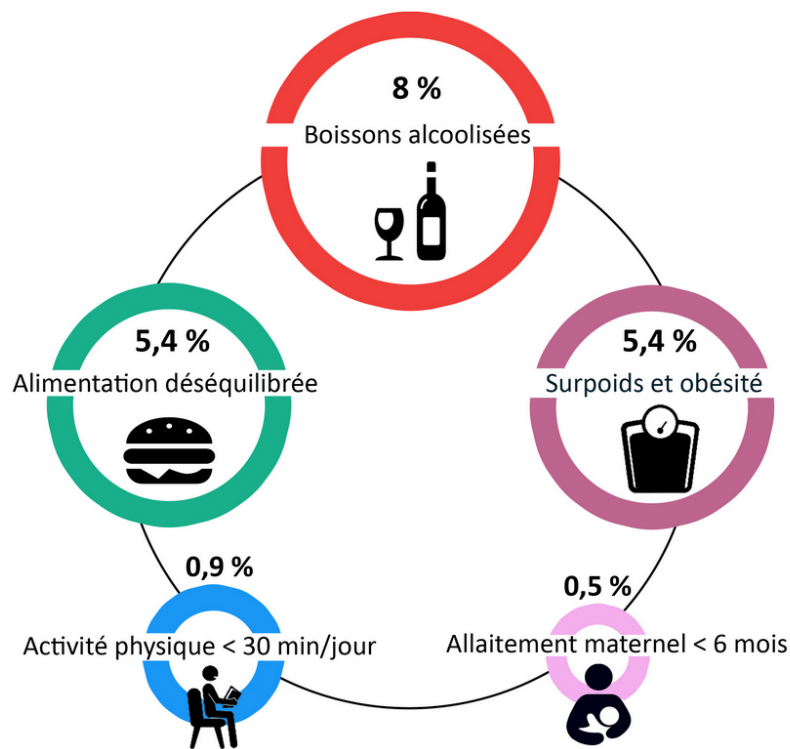


Figure 5 Parts des cancers liées aux facteurs de risque nutritionnels.

Source : <https://www6.inra.fr/nacre/Prevention-primaire/Vous-informer-sur/Cancers-attribuables-aux-facteurs-nutritionnels>

Il est également important de noter que de nombreuses interventions pour la prévention du cancer, telles que la lutte contre le tabac et l'alcool, la promotion d'un poids sain, une alimentation saine et l'activité physique apportent des avantages supplémentaires dans le traitement de nombreuses autres maladies chroniques (14).

### C. La cancérogenèse

La cancérogenèse est un processus complexe et multifactoriel. De plus, il existe une grande hétérogénéité des cancers avec souvent, au sein d’une même tumeur, différents clones néoplasiques (15). Ce qui caractérise la cancérogenèse est la transformation de cellules normales en cellules néoplasiques ayant des

comportements cellulaires anormaux tels qu'une division cellulaire accélérée et un déséquilibre entre prolifération et mort cellulaire programmée (apoptose). La cancérogenèse se divise en 3 grandes phases : l'initiation (lésions de l'ADN, mutations, altérations de l'expression de gènes, de la réparation de l'ADN ou de la détoxification) induisant la transformation cellulaire ; la promotion (multiplication des cellules anormales, diminution de l'apoptose) ; et la progression (accumulation de nouvelles altérations génétiques, angiogenèse, migration vers d'autres tissus avec formation de métastases) (16).

Le cancer se développe lorsque les processus de défense qui maintiennent l'intégrité des composants cellulaires (notamment de l'ADN) sont dépassés ce qui résulte en l'accumulation de dommages génétiques dans les cellules au fil du temps (15). Toutes les cellules cancéreuses possèdent des caractéristiques communes, notamment une signalisation induisant une prolifération soutenue, une résistance à l'apoptose, l'échappement aux régulations par les gènes suppresseurs de tumeurs, la multiplication cellulaire illimitée (immortalité), l'induction de l'angiogenèse, la dérégulation de l'homéostasie énergétique, l'activation de la capacité invasive et métastatique. L'acquisition de ces caractéristiques est rendue possible par le développement d'une instabilité génomique dans les cellules cancéreuses, générant des mutations aléatoires, notamment des réarrangements chromosomiques (17).

Les variations héréditaires les plus courantes de la séquence nucléotidique de l'ADN (qui se produisent chez plus d'une personne sur 100) sont appelées polymorphismes d'un seul nucléotide (SNPs, single-nucleotide polymorphism) et correspondent à la variation d'une seule paire de bases du génome, entre individus d'une même espèce (15). Ces variations dans la séquence d'ADN sont à l'origine de nombreuses différences génétiques entre les individus et impactent la susceptibilité à la maladie. Les études d'association pangénomique, évaluant simultanément des milliers de variations du génome dans les populations, ont permis de découvrir des centaines de variations liées au cancer. Néanmoins, il reste beaucoup à découvrir en ce qui concerne la contribution des polymorphismes génétiques à la cancérogenèse, ainsi que l'interaction entre ces variants génétiques et les expositions environnementales (tous les facteurs exogènes auxquels l'organisme est exposé) (15).

La régulation épigénétique de l'expression des gènes est régie par plusieurs phénomènes biochimiques distincts comme la méthylation de l'ADN. Dans les cellules tumorales, plusieurs gènes peuvent être hyperméthylés, notamment des gènes impliqués dans la régulation du cycle cellulaire ou dans la réparation de l'ADN. Certains gènes peuvent également être hypométhylés ce qui induit une instabilité génomique favorisant le développement du cancer (15). La modification de la structure des histones par méthylation ou acétylation ou encore la fixation des microARN sur les ARN messagers affectant leur stabilité sont d'autres modifications épigénétiques pouvant être impliquées dans les voies de la cancérogenèse.

Les gènes associés à la cancérogenèse sont catégorisés en gènes dits oncogènes ou suppresseurs de tumeurs. Les oncogènes sont des activateurs puissants des signaux de croissance, de réplication et de survie et leur expression est hautement régulée (15,16). Des modifications génétiques ou épigénétiques au niveau des oncogènes (mutations, surexpression) peuvent activer et provoquer des altérations du fonctionnement cellulaire. Les gènes suppresseurs permettent la réparation de l'ADN muté ou empêchent le développement des cellules potentiellement cancéreuses, en bloquant la prolifération ou en induisant l'apoptose (15). Ils peuvent eux aussi subir des mutations qui altèrent leur fonction (16).

Il est possible d'être prédisposé génétiquement au risque de cancer ; par exemple, des mutations héréditaires des gènes BRCA1 et BRCA2 causent 5 à 10% des cancers du sein. Par ailleurs, presque toutes



les cellules de l'organisme sont vulnérables aux dommages causés à leur ADN notamment durant certaines étapes du cycle cellulaire, un processus hautement régulé. Les cellules sont capables, grâce à divers mécanismes, de prévenir l'accumulation de dommages à l'ADN, en éliminant ou détoxifiant les agents externes pouvant causer ces dommages, en réparant les altérations causées à l'ADN afin qu'elles ne soient pas transmises aux cellules-filles ou en s'assurant que les cellules dont l'ADN est endommagé ne survivent pas (15). La protéine suppresseur de tumeur p53 joue un rôle majeur dans la protection des cellules contre la transformation néoplasique et le gène TP53 codant pour cette protéine est le plus souvent muté dans les cancers humains (15). Ces mécanismes de protection peuvent parfois échouer en présence de multiples facteurs de risque de cancer cumulés.

L'exposition à des agents « cancérogènes » de l'environnement extérieur peut causer des mutations, des lésions de l'ADN ou altérer l'expression de certains gènes en modifiant la structure de l'ADN et de la chromatine (changements épigénétiques réversibles). D'autres facteurs résultant de processus internes (endogènes) peuvent augmenter le risque de cancer, comme par exemple, l'accumulation de dérivés réactifs de l'oxygène tels que les radicaux libres entraînant un stress oxydatif, l'inflammation chronique, un système immunitaire affaibli ou encore des dérégulations des voies de signalisation cellulaire impliquant notamment des cytokines, des facteurs de croissance et des hormones (15). Le vieillissement entraîne l'accumulation de lésions de l'ADN et s'accompagne souvent d'une capacité de défense et de réparation réduite, augmentant ainsi le risque néoplasique (15).

La cancérogenèse résulte donc le plus souvent d'une interaction entre les gènes d'un individu, son métabolisme ainsi que les facteurs environnementaux (au sens large) auxquels il est exposé. Ce mécanisme évolue en plusieurs phases qui peuvent se dérouler sur de nombreuses années et met en jeu un ensemble de mécanismes activateurs ou protecteurs. Certains mécanismes sont communs à plusieurs sites ou types histologiques de cancers et d'autres sont plus spécifiques (16). Pour la plupart des cancers, il est impossible d'en déterminer la cause initiale. L'épidémiologie peut notamment aider à identifier un ou plusieurs facteurs associés au risque de cancer grâce à des études de cohorte prospectives en comparant l'exposition des personnes qui développent un cancer au cours du suivi avec des personnes qui n'en développent pas.

## II. Nutrition et prévention des cancers du sein et de la prostate

### A. Facteurs de risque et facteurs protecteurs des cancers du sein et de la prostate

Bien que la nutrition ne soit pas le seul facteur déterminant des maladies multifactorielles telles que les cancers du sein et de la prostate, il s'agit d'un facteur de risque modifiable sur lequel il est possible d'agir par l'intermédiaire d'interventions individuelles et collectives.

Les facteurs de risque non modifiables du cancer de la prostate sont l'âge, les antécédents familiaux de cancer de la prostate et l'ethnicité (18). Des concentrations sanguines élevées de facteur de croissance analogue à l'insuline (IGF-1) ont été suggérées comme facteur de risque modifiable de ce cancer (19,20).

Concernant le cancer du sein, certaines mutations héréditaires, en particulier des gènes BRCA1, BRAC2 et p53 et les antécédents familiaux de cancer du sein sont associés à une augmentation du risque (21). Par ailleurs, des premières règles précoces, une ménopause naturelle tardive, le fait de ne pas avoir d'enfants et la première grossesse à plus de 30 ans augmentent l'exposition aux œstrogènes et à la progestérone et le risque de cancer du sein au cours de la vie (21). L'exposition aux rayonnements ionisants provenant de traitements médicaux tels que les rayons X, en particulier pendant la puberté, augmente également le

risque de cancer du sein (22,23). Par ailleurs, la prise de contraceptifs oraux contenant à la fois des œstrogènes et de la progestérone augmenterait le risque de cancer du sein chez les jeunes femmes (parmi les utilisatrices actuelles et récentes) (24), ainsi que certains traitements hormonaux de la ménopause (traitement par hormonothérapie substitutive THS) selon le type histologique de ce cancer (25). De plus, une méta-analyse a montré que parmi ces derniers, l'utilisation d'œstrogènes et de progestérone était associée à une augmentation probable de la mortalité par cancer du sein, alors que l'utilisation d'œstrogènes ne l'était pas (26).

Concernant les facteurs nutritionnels, le groupe d'experts du WCRF / AICR tient régulièrement à jour les données sur le rôle des facteurs nutritionnels au sens large dans la prévention des cancers grâce au Continuous Update Project (CUP), en collaboration avec une équipe de l'Imperial College de Londres en charge des revues systématiques et des méta-analyses de données épidémiologiques. La base de données du CUP contient tous les articles pertinents issus d'essais contrôlés randomisés et d'études de cohortes publiées pour 17 cancers et survivants du cancer du sein (environ 10 000 publications sur ces cancers à ce jour). Les rapports du CUP les plus récents concernant le cancer du sein et le cancer de la prostate ont été publiés en 2017 (21) et en 2014 (18) respectivement, puis révisés en 2018. Selon ces rapports, des preuves robustes montrent que le risque de cancer du sein augmente avec la taille à l'âge adulte et la consommation d'alcool alors que l'allaitement diminue le risque de cancer du sein (Tableau 1 et Tableau 2). De plus, un poids élevé à la naissance, une plus faible activité physique vigoureuse et une plus faible adiposité augmenteraient le risque de cancer du sein préménopausique alors qu'une plus forte adiposité en général, la prise de poids chez l'adulte, une plus faible activité physique au global et une plus faible adiposité chez les jeunes adultes augmenteraient le risque de cancer du sein postménopausique.

Tableau 1 : Niveau de preuve concernant l'alimentation, la nutrition, l'activité physique et le risque de cancer du sein préménopausique

ALIMENTATION, NUTRITION, ACTIVITE PHYSIQUE ET CANCER DU SEIN PREMENOPAUSIQUE			
		RISQUE DIMINUE	RISQUE AUGMENTE
FORT NIVEAU DE PREUVE	Convainquant		Taille adulte <sup>1</sup>
	Probable	Activité physique vigoureuse, adiposité <sup>2</sup> , allaitement <sup>3</sup>	Consommation d'alcool <sup>4</sup> , plus haut poids à la naissance <sup>5</sup>
NIVEAU DE PREUVE LIMITE	Limité - suggéré	Légumes (hors féculent, ER-) <sup>6</sup> , produits laitiers, aliments contenant des caroténoïdes <sup>7</sup> , régimes riches en calcium, activité physique <sup>8</sup>	
	Limité - pas de conclusion	Céréales et produits céréaliers, fibres alimentaires, pomme de terre, légumes (hors féculent, ER+), fruits, légumineuses, soja et produits à base de soja, viande rouge et charcuterie, volaille, poisson, œufs, matière grasse et huiles, matière grasse totale, matière grasse végétale, composition en acides gras, acides gras saturés, acides gras mono-insaturés, acides gras polyinsaturés, acides gras trans, cholestérol, sucre (saccharose), autres sucres, produits et boissons sucrés, café, thé, glucides, féculents, index glycémique, charge glycémique, protéine, vitamine A, riboflavine, vitamine B6, folate, vitamine B12, vitamine C, vitamine D, vitamine E, suppléments en calcium, fer, sélénium, phytoestrogènes, isoflavones, dichlorodiphényldichloroéthylène, dichlorodiphényltrichloroéthane, dieldrine, hexachlorobenzène, hexachlorocyclohexane, trans-nonachlor, polychlorobiphényle, acrylamide, habitudes alimentaires, régimes alimentaires culturels, comportement sédentaire, prise de poids chez l'adulte, apport énergétique	
FORT NIVEAU DE PREUVE	Effet substantiel sur le risque peu probable		

SOURCE : WCRF/AIRC, CUP 2017

ER-, récepteurs d'œstrogènes négatifs ; ER+, récepteurs d'œstrogènes positifs

<sup>1</sup> Il est peu probable que la taille adulte influence directement le risque de cancer. C'est un marqueur des facteurs génétiques, environnementaux, hormonaux, mais aussi nutritionnels affectant la croissance durant la période allant de la préconception à la fin de la croissance linéaire.

<sup>2</sup> L'adiposité est marquée par l'indice de masse corporelle (IMC), le tour de taille et le rapport taille / hanches. Cela comprend également des preuves concernant les jeunes femmes âgées de 18 à 30 ans environ. La graisse corporelle chez les jeunes adultes est estimée par l'IMC.

<sup>3</sup> La conclusion du groupe d'experts relate les éléments de preuve concernant le cancer du sein au global (non spécifié). Les preuves relatives aux cancers du sein préménopausiques et postménopausiques étaient moins concluantes, mais cohérentes avec les résultats globaux.

<sup>4</sup> Aucun seuil n'a été identifié.

<sup>5</sup> Le poids à la naissance est un marqueur à la fois de la croissance prénatale, reflétant la nutrition fœtale, et est un facteur prédictif de la croissance et de la maturation ultérieures - par exemple, l'âge à la ménarche - qui sont également des déterminants du risque de cancer du sein.

<sup>6</sup> La conclusion du groupe d'experts relate les éléments de preuve concernant le cancer du sein au global (non spécifié). L'association observée concernait uniquement le cancer du sein avec récepteurs d'œstrogènes négatifs (ER-).

<sup>7</sup> La conclusion du groupe d'experts relate les éléments de preuve concernant le cancer du sein au global (non spécifié). L'association observée était plus forte pour le cancer du sein avec récepteurs d'œstrogènes négatifs (ER-). Cela comprend les aliments qui contiennent naturellement des caroténoïdes et les aliments auxquels des caroténoïdes sont ajoutés.

<sup>8</sup> Activité physique, y compris activité professionnelle, de loisir, de marche et à la maison. Il y avait suffisamment de preuves pour que le groupe d'experts puisse rendre un jugement séparé pour une activité physique vigoureuse.

Tableau 2 Niveau de preuve concernant l'alimentation, la nutrition, l'activité physique et le risque de cancer du sein postménopausique

ALIMENTATION, NUTRITION, ACTIVITE PHYSIQUE ET CANCER DU SEIN POSTMENOPAUSIQUE			
		RISQUE DIMINUE	RISQUE AUGMENTE
FORT NIVEAU DE PREUVE	Convainquant		Consommation d'alcool <sup>1</sup> , Adiposité <sup>2</sup> , prise de poids chez l'adulte, taille adulte <sup>3</sup>
	Probable	Activité physique <sup>4</sup> , adiposité chez les jeunes adultes <sup>5</sup> , allaitement <sup>6</sup>	
NIVEAU DE PREUVE LIMITE	Limité – suggéré	Légumes (hors féculent, ER-) <sup>7</sup> , aliments contenant des caroténoïdes <sup>8</sup> , régimes riches en calcium	
	Limité -pas de conclusion	Céréales et produits céréaliers, fibres alimentaires, pomme de terre, légumes (hors féculent, ER+), fruits, légumineuses, soja et produits à base de soja, viande rouge et charcuterie, volaille, poisson, œufs, produits laitiers, matière grasse et huiles, matière grasse totale, matière grasse végétale, composition en acides gras, acides gras saturés, acides gras mono-insaturés, acides gras polyinsaturés, acides gras trans, cholestérol, sucre (saccharose), autres sucres, produits et boissons sucrés, café, thé, glucides, féculents, index glycémique, charge glycémique, protéine, vitamine A, riboflavine, vitamine B6, folate, vitamine B12, vitamine C, vitamine D, vitamine E, suppléments en calcium, fer, sélénium, phytoestrogènes, isoflavones, dichlorodiphényldichloroéthylène, dichlorodiphényltrichloroéthane, dieldrine, hexachlorobenzène, hexachlorocyclohexane, trans-nonachlor, polychlorobiphényle, acrylamide, habitudes alimentaires, régimes alimentaires culturels, comportement sédentaire, apport énergétique	
FORT NIVEAU DE PREUVE	Effet substantiel sur le risque peu probable		

SOURCE : WCRF/AIRC, CUP 2017

ER-, récepteurs d'œstrogènes négatifs ; ER+, récepteurs d'œstrogènes positifs

<sup>1</sup> Aucun seuil n'a été identifié.

<sup>2</sup> L'adiposité à l'âge adulte, marquée par l'indice de masse corporelle (IMC), le tour de taille et le rapport taille / hanche

<sup>3</sup> Il est peu probable que la taille adulte influence directement le risque de cancer. C'est un marqueur des facteurs génétiques, environnementaux, hormonaux mais aussi nutritionnels affectant la croissance durant la période allant de la préconception à la fin de la croissance linéaire.

<sup>4</sup> Activité physique, y compris vigoureuse, activité professionnelle, de loisir, de marche et à la maison.

<sup>5</sup> Jeunes femmes âgées de 18 à 30 ans environ. L'adiposité chez les jeunes adultes est estimée par l'IMC.

<sup>6</sup> La conclusion du groupe d'experts relate les éléments de preuve concernant le cancer du sein au global (non spécifié). Les preuves relatives aux cancers du sein préménopausique et postménopausique étaient moins concluantes, mais cohérentes avec les résultats globaux.

<sup>7</sup> La conclusion du groupe d'experts relate les éléments de preuve concernant le cancer du sein au global (non spécifié). L'association observée concernait uniquement le cancer du sein avec récepteurs d'œstrogènes négatifs (ER-).

<sup>8</sup> La conclusion du groupe d'experts relate les éléments de preuve concernant le cancer du sein au global (non spécifié). L'association observée était plus forte pour le cancer du sein avec récepteurs d'œstrogènes négatifs (ER-). Cela comprend les aliments qui contiennent naturellement des caroténoïdes et les aliments auxquels des caroténoïdes sont ajoutés.

Concernant le cancer de la prostate, des preuves robustes montrent que l'adiposité (pour les cancers de la prostate avancés) et la taille adulte sont associées à une augmentation du risque (Tableau 3).

Tableau 3 Niveau de preuve concernant l'alimentation, la nutrition, l'activité physique et le risque de cancer de la prostate

ALIMENTATION, NUTRITION, ACTIVITE PHYSIQUE ET CANCER DE LA PROSTATE			
		RISQUE DIMINUE	RISQUE AUGMENTE
FORT NIVEAU DE PREUVE	Convainquant		
	Probable		<b>Adiposité (cancer de la prostate avancé)<sup>1,2</sup>, taille adulte<sup>3</sup></b>
NIVEAU DE PREUVE LIMITE	Limité - suggéré		Produits laitiers, régimes riches en calcium, concentration plasmatique faible en alpha-tocophérol, concentration plasmatique faible en sélénium
	Limité - pas de conclusion	Céréales et produits céréaliers, fibres alimentaires, pomme de terre, légumes (hors féculent), fruits, légumineuses, viande rouge et charcuterie, volaille, poisson, œufs, matière grasse totale, acides gras saturés, acides gras mono-insaturés, acides gras polyinsaturés, matière grasse végétale, sucre (saccharose), produits et boissons sucrés, café, thé, alcool, glucides, protéine, vitamine A, rétinol, alpha carotène, lycopène, folate, thiamine, riboflavine, niacine, vitamine C, vitamine D, suppléments en vitamine E, gamma tocophérols, multivitamines, suppléments en sélénium, fer, phosphore, suppléments en calcium, zinc, activité physique, dépense énergétique, régimes végétariens, régimes adventistes du septième jour, habitudes alimentaires individuelles, adiposité (cancer de la prostate non avancé), poids à la naissance, apport énergétique	
FORT NIVEAU DE PREUVE	Effet substantiel sur le risque peu probable	Beta-carotène <sup>4,5</sup>	

SOURCE : WCRF/AIRC, CUP 2014

<sup>1</sup> L'adiposité est marquée par l'indice de masse corporelle (IMC), le tour de taille et le rapport taille / hanches. L'effet a été observé pour les cancers de la prostate avancés seulement.

<sup>2</sup> Avancé dans ce rapport inclus avancé, haut grade et cancers de la prostate mortels.

<sup>3</sup> Il est peu probable que la taille adulte influence directement le risque de cancer. C'est un marqueur des facteurs génétiques, environnementaux, hormonaux mais aussi nutritionnels affectant la croissance durant la période allant de la préconception à la fin de la croissance linéaire.

<sup>4</sup> Comprend à la fois les aliments contenant naturellement le constituant et les aliments auxquels le constituant est ajouté.

<sup>5</sup> Les preuves incluent les études utilisant des suppléments ayant des doses de 20, 30 et 50 mg/jour.

Plusieurs autres facteurs nutritionnels, ont été suggérés comme facteurs de risque de cancers du sein et de la prostate. Un rôle protecteur dans la cancérogenèse mammaire a notamment été suggéré pour les légumes (hors féculents, chez les cancers du sein ER-), les produits laitiers (chez les femmes préménopausées), les aliments contenant des caroténoïdes et les aliments riches en calcium (21). D'autre part, la consommation de produits laitiers, des régimes riches en calcium et des concentrations plasmatiques faible en alpha-tocophérol et sélénium ont été suggérés comme facteurs augmentant le risque de cancer de la prostate (18). Par ailleurs, deux méta-analyses et une revue de la littérature ont mis en évidence une association positive entre l'index inflammatoire de l'alimentation (27) et le risque de cancer du sein (28,29) et de la prostate (30).

## B. Influence des facteurs alimentaires, de l'adiposité et de l'activité physique dans la cancérogenèse : aspects mécanistiques

Un apport adéquat en énergie et nutriments permet à l'organisme une mise en réserve fonctionnelle, permettant de faire face aux défis inévitables que tout organisme est amené à relever. Les substrats et les cofacteurs nécessaires au métabolisme (appelés nutriments) proviennent de l'alimentation. L'énergie dérive des macronutriments ; lipides et glucides principalement, et des protéines et de l'alcool (15). Le régime alimentaire apporte également de nombreuses substances qui ne sont pas des nutriments mais qui peuvent néanmoins influencer les processus métaboliques comme par exemple, la caféine ou des composés phytochimiques comme le lycopène des tomates et les isoflavones du soja.

L'alimentation, le statut nutritionnel et l'activité physique peuvent influencer le risque de cancer de différentes façons. L'un des principaux mécanismes par lesquels les composés alimentaires pourraient influencer la cancérogenèse est la modulation des voies par lesquelles les agents cancérogènes sont métabolisés (15). Certains aliments et boissons peuvent être des vecteurs de substances spécifiques qui agissent comme cancérogènes sur des cibles particulières. Par exemple, l'éthanol contenu dans les boissons alcoolisées est métabolisé en acétaldéhyde qui peut augmenter l'inflammation et l'instabilité du génome (15). Le fer héminique présent dans la viande rouge pourrait avoir un effet catalytique sur la formation endogène de composés N-nitrosés cancérogènes et d'aldéhydes cytotoxiques et génotoxiques par peroxydation lipidique (31).

A l'inverse, des facteurs nutritionnels pourraient influencer les mécanismes de réparation de l'ADN et moduler l'activité des enzymes du métabolisme des xénobiotiques (comme par exemple, les isothiocyanates du brocoli). Des substances à activité biologique potentiellement protectrice sont présentes dans les fruits et légumes (vitamines A, B, C, E, polyphénols, caroténoïdes, butyrate, sélénium) et peuvent intervenir dans les différentes étapes de la cancérogenèse (initiation, promotion, progression) (32). Les réponses physiologiques aux lésions de l'ADN peuvent être modifiées par l'action d'hormones susceptibles d'être impactées par l'alimentation, l'activité physique et l'excès de graisse corporelle (15).

Les facteurs environnementaux tels que la nutrition peuvent également influencer la cancérogenèse *via* des altérations épigénétiques qui sont réversibles contrairement aux mutations. Par exemple, les folates alimentaires et d'autres donneurs de méthyle comme la choline, la méthionine et la bétaine sont essentiels à la synthèse de l'ADN et à la régulation épigénétique de celui-ci (15). Par ailleurs, d'autres facteurs comme l'obésité et le mode de vie sédentaire peuvent altérer les régulations métaboliques de l'organisme et générer un microenvironnement cellulaire favorisant le développement du cancer (15). Par exemple, l'obésité est associée à des médiateurs inflammatoires et à des anomalies endocriniennes qui favorisent la croissance cellulaire et impactent le système immunitaire (33). L'alimentation déséquilibrée et/ou insuffisante peut en effet entraîner une immunodéficience et une susceptibilité accrue aux infections (15). La Figure 6 résume le large éventail de facteurs nutritionnels au sens large pouvant influencer les processus de la cancérogenèse. Cependant, les relations entre nutrition et cancer sont très complexes et nombreuses, il ne s'agit donc que de quelques exemples.

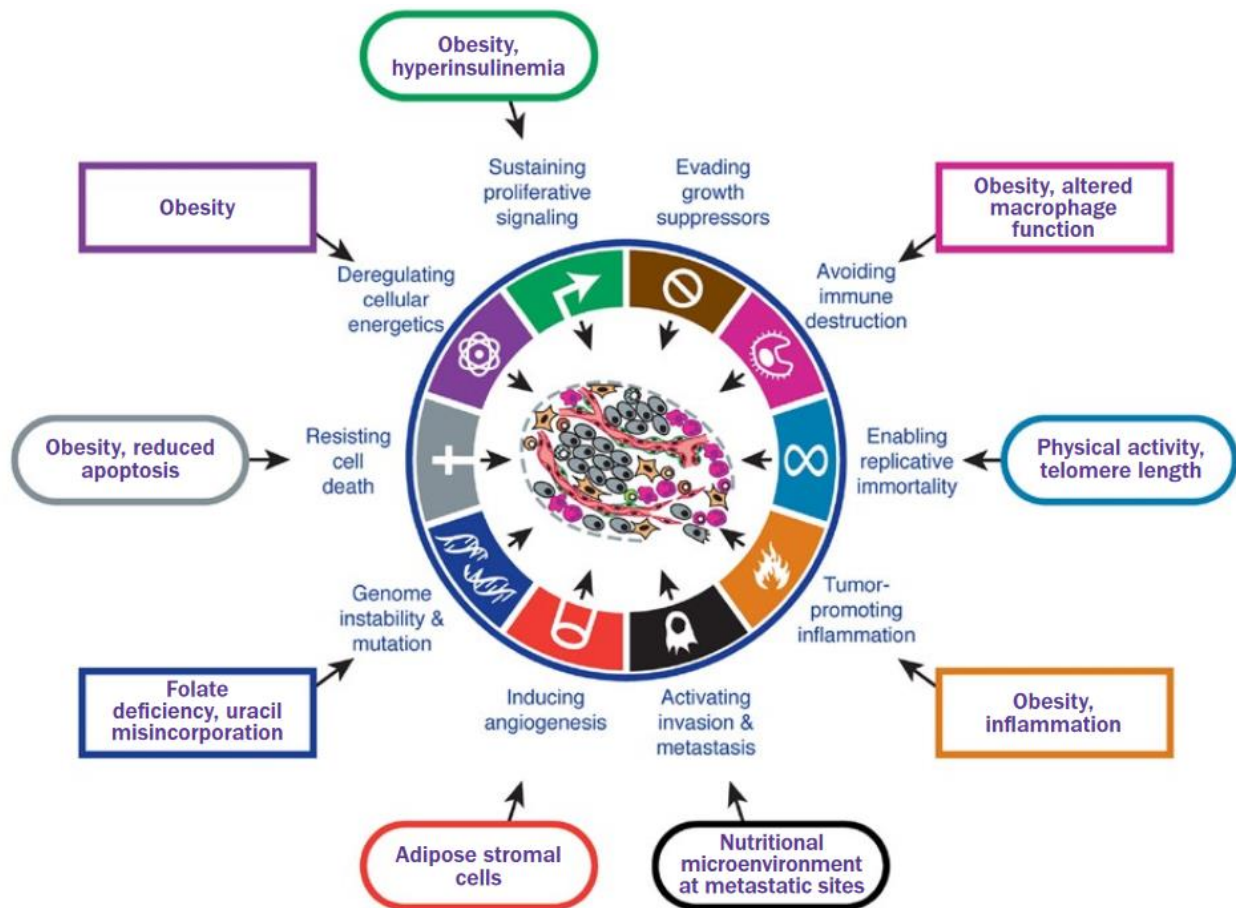


Figure 6 Impacts des facteurs nutritionnels sur les mécanismes sous-jacents du cancer.  
 Source : World Cancer Research Fund/American Institute for Cancer Research, 2018 (15)

Les facteurs de risque de cancers du sein et de la prostate exposés plus haut ont été identifiés par le WCRF qui a, pour ce faire, tenu compte à la fois des résultats des méta-analyses des études épidémiologiques, mais également de leur plausibilité biologique. Par exemple, l'activité physique, associée à une réduction du risque de cancer du sein, diminuerait les taux d'œstrogènes et d'androgènes chez les femmes post-ménopausées, réduirait l'adiposité et induirait un effet immunomodulateur (21). L'adiposité, en plus d'avoir une activité inflammatoire, affecte directement les concentrations de nombreuses hormones circulantes telles que l'insuline, les œstrogènes, la testostérone, les adipokines et les facteurs de croissance, qui favorisent le développement de cancer notamment du sein et de la prostate (18,21). La taille à l'âge adulte est liée au taux d'hormones de croissance durant l'enfance et à l'imprégnation hormonale qui peut avoir des effets à court et à long terme sur la production de facteurs de croissance, d'insuline et autres médiateurs endocriniens ou tissulaires spécifiques qui peuvent influencer le risque néoplasique mammaire et prostatique (18,21).

Certains mécanismes expliquant les liens entre consommation de boissons alcoolisées et augmentation du risque de cancer du sein ont été avancés, mais ils ne sont pas encore totalement élucidés notamment du fait de leur complexité (21). L'alcool a un effet médiateur sur le métabolisme des lipides bioactifs, y compris la production de prostaglandines, la peroxydation des lipides et la production d'espèces réactives à l'oxygène. Il modulerait également les régulations endocriniennes, notamment les taux d'hormones stéroïdiennes et de facteurs de croissance (21).

Le calcium des produits laitiers peut réguler la formation de 1,25-dihydroxy vitamine D3, augmentant ainsi la prolifération cellulaire dans la prostate (18). Les fruits et légumes sont riches en composés phytochimiques qui réduiraient l'expression du récepteur au facteur de croissance épidermique et donc le risque de développer un cancer du sein ER- ainsi qu'en caroténoïdes ayant des propriétés antioxydantes offrant une protection contre les dommages de l'ADN (21).

La combinaison des études en biologie fondamentale et en épidémiologie permet de fournir des arguments convaincants en faveur d'un lien causal entre la nutrition et le risque de plusieurs cancers. De nombreux mécanismes restent à élucider ainsi que probablement de nouveaux liens avec d'autres agents cancérigènes (15).

III. Nouvelles perspectives pour l'épidémiologie nutritionnelle : l'exploration du métabolome humain

#### **A. Intérêt des biomarqueurs pour compléter les études d'épidémiologie nutritionnelle**

L'objectif des méthodes d'évaluation des consommations alimentaires dans le contexte d'études de cohortes est d'estimer précisément les apports en aliments, nutriments et composés bioactifs, dans le but d'étudier leurs associations avec l'état de santé et/ou d'assurer un suivi de l'état nutritionnel de la population. Dans ce type d'études, les sujets sont en conditions réelles et les apports nutritionnels sont complexes à mesurer, en particulier du fait de la large gamme d'aliments consommés et de la grande hétérogénéité et variabilité des choix alimentaires. Des méthodes performantes ont été développées et validées pour évaluer les consommations alimentaires. Il s'agit notamment des carnets de consommation, des rappels ou des enregistrements de 24h, ou des questionnaires de fréquence (FFQ). Les rappels et enregistrements de 24h répétés ont prouvé leur utilité et leur fiabilité, toutefois, ils se révèlent très coûteux et donc parfois difficiles à implémenter dans les grandes cohortes. Le FFQ reste donc la méthode la plus « coût-efficace » et donc la plus répandue dans les cohortes (34). Cependant, sa fiabilité a été questionnée du fait de plusieurs publications observant des associations plus robustes entre facteurs nutritionnels et pathologies lorsque des enregistrements ou rappels de 24h répétés ou des biomarqueurs étaient utilisés au lieu du FFQ (35–37). Les limites des FFQ incluent les erreurs de mémoire et déclarations erronées inhérentes à toutes les méthodes d'autodéclaration, les rendant moins appropriés pour certaines populations (personnes âgées, enfants et personnes en situation d'obésité par exemple), la difficulté à capturer le large éventail de la diversité alimentaire du fait du regroupement des aliments, et la difficulté d'évaluer les tailles de portion (38). De plus en plus d'éléments suggèrent que les lacunes méthodologiques liées à la mesure des consommations alimentaires sont en grande partie responsables de certaines incohérences dans les résultats des études épidémiologiques portant sur les relations nutrition-cancer ou d'absence d'association observée (35). La découverte de nouveaux biomarqueurs d'exposition nutritionnelle apparaît donc comme une voie stratégique pour compléter les méthodes de recueil des consommations alimentaires.

Par ailleurs, les modèles prédictifs estimant la probabilité qu'une personne en bonne santé à un instant t présentant des facteurs de risque spécifiques développe une maladie dans le futur pourraient renseigner sur la charge de morbidité future, les politiques de santé et les décisions individuelles en termes de prévention. Les modèles contenant des facteurs de risque modifiables, tels que la consommation d'alcool, l'utilisation de traitements hormonaux, l'alimentation ou encore l'activité physique sont particulièrement intéressantes dans cette optique et en vue de servir de base à l'élaboration des politiques de santé



publique pour réduire les taux d'incidence dans la population. De nos jours, de tels modèles concernant le risque de cancer sont peu nombreux, manquent de validation externe et de précision et restent peu performants (39,40). Les recherches futures devraient donc être axées sur la validation externe ainsi que l'identification de nouveaux prédicteurs ou biomarqueurs ayant une grande capacité prédictive.

Les études épidémiologiques ont permis d'identifier de nombreux facteurs de risque, notamment environnementaux, qui augmentent ou réduisent le risque de cancers spécifiques, conduisant à des stratégies de prévention du cancer. Dans les critères de Bradford-Hill (41) permettant de juger de la causalité d'une association, la plausibilité biologique constitue un critère important. Bien que de nombreuses preuves s'accumulent en faveur d'un rôle fondamental de l'environnement et notamment de la nutrition dans la promotion et la progression des cancers, les mécanismes précis par lesquels ces expositions, individuelles ou en combinaisons, interagissent tout au long de la vie pour créer un microenvironnement cellulaire propice au cancer ne sont pas encore complètement élucidés (15,42). Dans ce contexte, le terme d'«épidémiologie moléculaire» a fait son apparition dans la littérature au début des années 1980. Conçu à l'origine comme une extension de la recherche épidémiologique « traditionnelle », il incorporait des biomarqueurs (biochimiques et moléculaires) à des données de questionnaires afin de mieux comprendre les mécanismes de la carcinogenèse et les événements tout au long du continuum entre exposition et développement du cancer (43).

## **B. Le métabolome humain**

Comme nous l'avons vu précédemment, un grand nombre de facteurs, notamment environnementaux, peuvent favoriser l'apparition des cancers ainsi que de nombreuses autres maladies chroniques. Les expositions à ces facteurs sont variables selon les individus. Le concept d'« exposome » a été proposé en 2005 et comprend toutes les expositions environnementales qu'une personne subit depuis sa conception et tout au long de sa vie (44,45). L'exposome complète donc le génome en fournissant une description globale de l'historique de l'exposition tout au long de la vie d'un individu (45). L'utilisation de biomarqueurs, tels que des métabolites, pourrait constituer une mesure objective pouvant améliorer la compréhension des mécanismes reflétant l'interaction des expositions environnementales avec la susceptibilité de l'hôte à développer une maladie, ces notions étant le fondement de l'épidémiologie moléculaire (46). Les biomarqueurs ne représentent bien sûr qu'une partie de la solution et devront être couplés à des mesures par questionnaires, captant les comportements et modes de vie de plus en plus finement. Le métabolome humain est composé de toutes les petites molécules – métabolites - présentes dans les biofluides ou tissus (47) et est une partie complexe, sensible et dynamique de l'exposome, d'une complexité comparable à celle du génome humain (48). Il contient au moins 50 000 composants détectables et ce chiffre devrait augmenter avec le progrès des technologies (49,50). Le métabolome comporte et est influencé par plusieurs entités en interaction (Figure 7) : l'environnement, l'alimentation, le métabolisme endogène (tous les métabolites de l'hôte) et celui du microbiote – ayant chacun leur propre « métabolome » - ainsi que par d'autres facteurs tels que l'âge, la génétique, le sexe, le mode de vie et l'état de santé (47,51).

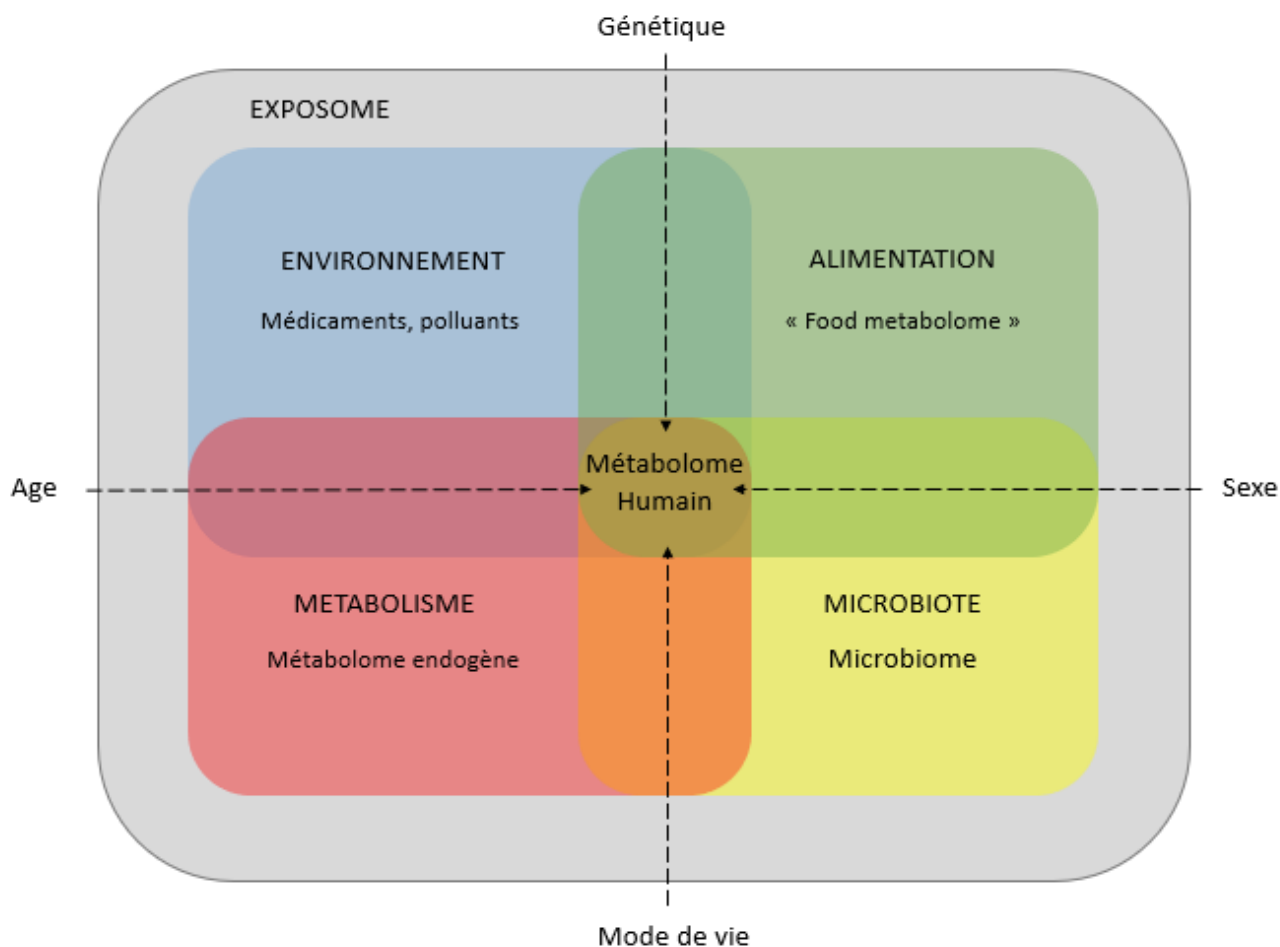


Figure 7 Composition et influences du métabolome humain  
Source : Athersuch, Mutagenis, 2015 (47)

L'alimentation impacte le métabolome humain via le métabolisme endogène et *via* le « food metabolome » (51), celui-ci ayant été défini pour la première fois en 2008 comme l'ensemble des métabolites directement dérivés de la digestion des aliments, de leur absorption dans l'intestin et de la biotransformation par les tissus de l'hôte et du microbiote (52). Bien que certains métabolites microbiens soient spécifiques (notamment des vitamines, certains acides aminés essentiels et quelques acides gras), la grande majorité provient de la biotransformation du métabolome endogène et du « food metabolome » et fait donc partie intégrante de ces 2 métabolomes. Ils comprennent les acides gras à chaîne courte, les acides biliaires secondaires, des métabolites dérivés de protéines et d'acides aminés, ainsi que les métabolites de polyphénol de plante (53). Il est particulièrement ardu d'analyser le « food metabolome » car il possède une diversité chimique beaucoup plus grande que toutes les autres parties du métabolome, une vaste gamme de concentrations (allant des concentrations picomolaires ou nanomolaires pour certains contaminants ou métabolites phytochimiques aux concentrations millimolaires pour les nutriments tels que les sucres) et de nombreux composants du « food metabolome » restent inconnus (51).

### C. Définitions et classifications des biomarqueurs

Il existe un grand nombre de définitions des biomarqueurs, dépendantes du domaine dans lequel ils sont utilisés et de leurs applications, résumées dans la revue de Gao et al. (54). De manière générale, les

biomarqueurs sont des mesures objectives permettant de caractériser l'état actuel d'un système biologique (54). Dans le domaine de la nutrition, une déclaration de consensus lors de la 23<sup>ème</sup> conférence de Hohenheim a abouti à une définition des biomarqueurs en tant que «résultats de tests liés à l'exposition, à la susceptibilité ou à l'effet biologique» (55). L'exposition reflète le niveau de facteurs externes auxquels un individu est exposé, ils peuvent être bien visibles comme les aliments, les médicaments et le tabagisme, mais également moins tangibles comme le stress physique ou psychologique et les facteurs sociétaux (54). La susceptibilité représente la résilience individuelle à une exposition en prédisant l'intensité de son effet sur l'individu. La susceptibilité peut être considérée comme « l'état de santé général », c'est-à-dire la somme des facteurs intrinsèques expliquant les risques sanitaires individuels (55). Les biomarqueurs d'effet désignent la réponse fonctionnelle du corps humain à une exposition, définis en fonction du temps qui s'écoule avant qu'ils n'atteignent leur état d'équilibre comme par exemple, la réponse glycémique plasmatique et la tension artérielle (56). La Figure 8 présente la diversité des interactions entre le système biologique et les variables intrinsèques du système et les variables environnementales environnantes.

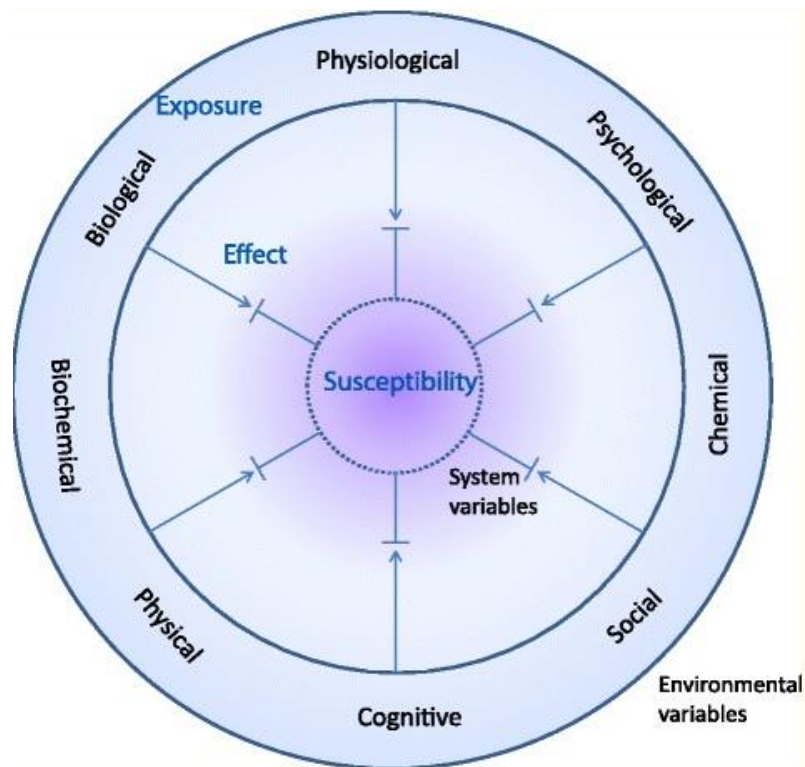


Figure 8 Diversité des interactions entre le système biologique et les variables intrinsèques du système et les variables environnementales environnantes.

Source : Gao et al. 2017 (54)

Ces trois classes de biomarqueurs se recoupent puisque la même mesure de biomarqueurs peut se retrouver dans plusieurs classes en fonction de leur utilisation dans différentes études. Par exemple, les mesures de la glycémie peuvent refléter directement un apport récent en glucose (exposition), la réponse cinétique du glucose (effet) ou la capacité de faire face à un défi glycémique (sensibilité) (54). Gao et al. ont proposé des sous-classes de ces catégories afin d'établir une classification des biomarqueurs nutritionnels déterminée uniquement par le but de l'utilisation du biomarqueur et non par le dosage en tant que tel (54) :

- Biomarqueurs d'exposition
  - Biomarqueurs de l'apport en composés alimentaires (nutritifs ou non nutritifs) représentant des apports spécifiques pendant une période bien définie (ponctuellement ou à long terme)
  - Biomarqueurs de l'apport de groupes d'aliments, d'aliments ou de composants alimentaires spécifiques (comme les ingrédients) pouvant être utilisés pour estimer les apports récents ou moyens de ces entités et donc fournir une évaluation objective de l'apport alimentaire dans la recherche en nutrition. Cette classe peut être représentée par des biomarqueurs individuels ou en combinaison. Idéalement, ce biomarqueur ne sera pas retrouvé lorsque l'aliment n'est pas ingéré et présentera des réactions distinctes selon la dose et le moment de consommation, de manière indépendante des facteurs individuels de l'hôte.
  - Biomarqueurs des habitudes alimentaires reflétant le régime alimentaire moyen d'un individu. En général, ces biomarqueurs représentent les aliments les plus abondants dans un certain régime alimentaire, par exemple les biomarqueurs de l'huile d'olive, des agrumes, des légumes, des noix et des poissons ainsi que l'acide alpha-linoléique et les polyphénols représentant les régimes de type méditerranéen ; ces biomarqueurs pouvant être utilisés en combinaison pour produire un score d'évaluation du régime. Ils peuvent être utiles pour évaluer la qualité globale de l'alimentation et leurs effets sur la santé dans le cadre d'études à long terme puisqu'il existe des interactions entre les différentes expositions alimentaires. Pour ce type de biomarqueurs, il est préférable de procéder à des échantillonnages de biofluides répétés afin d'éliminer la variabilité externe (comme le moment de prélèvement) et la variabilité individuelle.
- Biomarqueurs d'effet ou « intermédiaires »
 

Ils permettent de mesurer la réponse à une exposition, celle-ci peut être soit purement fonctionnelle pour indiquer un certain mécanisme d'action soit avoir une conséquence sur un risque de pathologie.

  - Les biomarqueurs de réponse fonctionnelle peuvent être corrélés à un risque de pathologie mais ne sont pas reliés de manière causale.
  - Les biomarqueurs d'effet sur le risque sont aussi classiquement utilisés comme biomarqueurs de risque mais dans ce cas on cherche à mesurer leur variation suite à une intervention (nutritionnelle ou traitement par exemple). Ce sont donc des biomarqueurs modifiables qui par exemple pourraient permettre de diminuer un risque suite à une intervention.
- Biomarqueurs de susceptibilité
  - Biomarqueurs du statut en composé alimentaire pouvant être des nutriments activement absorbés et retenus ou des composés non nutritifs accumulés dans le corps.
  - Biomarqueurs physiologiques ou d'état de santé pouvant être subdivisés en biomarqueurs du facteur hôte et biomarqueurs du risque. Les biomarqueurs du facteur hôte - considéré comme une qualité intrinsèque ou la variabilité individuelle ou un trait de caractère influençant la susceptibilité à la maladie ou la résilience - ne sont pas utilisés pour prédire le risque ; ce sont par exemple, l'immunité ou la résistance à un poison. A l'inverse, les biomarqueurs de risque sont normalement utilisés pour prédire des aspects spécifiques du risque de maladie d'une personne. Ils servent à caractériser le risque à l'inclusion dans une population ou de manière personnalisée pour déterminer si un traitement doit être instauré ou modifié. Ils reflètent le risque actuel ou la susceptibilité à développer la maladie.

Les biomarqueurs peuvent donc permettre d'améliorer l'évaluation de l'exposition notamment nutritionnelle, d'apporter des informations sur les changements précoces précédant la maladie et d'identifier les sous-groupes de la population les plus vulnérables à la maladie, augmentant ainsi la capacité des études épidémiologiques à identifier les causes et les mécanismes des pathologies (57). Les trois grandes catégories de biomarqueurs peuvent alors être recherchées en épidémiologie moléculaire des cancers : les biomarqueurs d'exposition pour l'identification de nouveaux carcinogènes par exemple, les biomarqueurs d'effet permettant d'identifier les événements biologiques qui se produisent dans le continuum entre l'exposition et le développement du cancer (58) et donc démontrer la plausibilité biologique d'une exposition proposée comme étant étiologiquement pertinente (57) et les biomarqueurs de susceptibilité contribuant à l'identification des sous-groupes de la population à haut risque (57).

### *1. Les biomarqueurs d'expositions alimentaires*

En épidémiologie, les informations d'exposition sont collectées grâce à des questionnaires et autres outils de collecte des modes de vie. Les expositions alimentaires sont généralement mesurées avec des méthodes validées de type rappels ou enregistrements de 24h répétés ou fréquentiels alimentaires. L'utilisation de biomarqueurs alimentaires comme mesures objectives de l'exposition en épidémiologie nutritionnelle apparaît pertinente pour compléter des mesures d'exposition (59–63). Une étude a notamment montré que la combinaison d'un biomarqueur de l'exposition alimentaire avec des données alimentaires autodéclarées permettrait d'augmenter la puissance statistique pour détecter une association entre l'alimentation et la maladie (64).

Les composés alimentaires peuvent être métabolisés de 3 façons différentes : 1) ils peuvent être digérés dans la bouche, l'estomac et l'intestin grêle en nutriments simples qui peuvent être absorbés par la barrière intestinale ; 2) ils peuvent être transformés par les tissus hôtes, en particulier le foie et les reins ; ou 3) ils peuvent être traités par le microbiote intestinal du gros intestin (51). La majorité des composés de la première catégorie sont communs à tous les organismes vivants et appartiennent aux métabolites endogènes humains, ne pouvant donc généralement pas être utilisés comme biomarqueurs d'exposition alimentaire car il est impossible de retracer leur origine alimentaire à l'exception des acides aminés et acides gras essentiels, ainsi que la plupart des vitamines et minéraux qui ne peuvent être produits par l'homme et qui doivent provenir d'une source alimentaire externe (51). Les composés de la deuxième catégorie ne correspondant pas aux métabolites endogènes sont traités comme "étrangers" ou comme xénobiotiques ; ces métabolites dérivés exogènes peuvent être très utiles comme biomarqueurs d'expositions alimentaires puisqu'ils sont spécifiques et non synthétisés par l'organisme (51). Enfin, certains métabolites microbiens peuvent être utiles comme biomarqueurs alimentaires cependant il existe une relation complexe entre la source alimentaire, l'espèce microbienne intestinale prédominante, et les métabolites alimentaires qui en résultent (65) induisant des corrélations plus faibles comparativement aux composés alimentaires non transformés et aux métabolites transformés par l'hôte (66) ; ces métabolites microbiens doivent donc être traités avec une certaine prudence lorsqu'ils sont utilisés comme biomarqueurs alimentaires.

Un nombre encore très limité de biomarqueurs d'exposition ont été correctement validés et appliqués en épidémiologie nutritionnelle (59,67). La base de données Exposome-Explorer (68) recense divers biomarqueurs d'expositions alimentaires présumés, identifiés à l'aide d'analyses de corrélations qui ont été mesurés dans des études épidémiologiques, notamment pour les fruits et légumes (polyphénols, caroténoïdes, et vitamine C), les céréales complètes/fibres (alkylresorcinols), le soja (isoflavones), la viande (acides aminés et acides gras), les produits laitiers (acides gras), le poisson (acides gras) et le thé et le vin

(polyphénols) (51). Cependant, un bon nombre d'entre eux n'ont pas été validés dans des études d'intervention dose-réponse et leurs sensibilité et spécificité n'ont pas été examinées. Un examen complet de ces biomarqueurs présumés et de leur validité est en cours par le consortium FoodBall (<http://foodmetabolome.org/>).

Les biomarqueurs d'expositions alimentaires peuvent être classés selon leur temporalité : les biomarqueurs à court terme (reflétant l'apport des heures/jours passés), à moyen terme (reflétant l'apport des semaines/mois) et à long terme (reflétant l'apport des mois/années) (69).

## 2. Les biomarqueurs d'effet ou « intermédiaires »

Ces mêmes biomarqueurs peuvent être également identifiés dans les deux autres catégories (biomarqueurs d'exposition ou de susceptibilité) mais sont identifiés ici dans le but de mesurer la réponse biologique à une exposition. Cette réponse peut être ponctuelle suite à l'exposition ou peut générer une perturbation du métabolisme à long terme et induire la mise en place de mécanismes tels que la cancérogénèse. Obtenir de bonnes estimations de l'exposition à long terme, notamment de l'apport alimentaire, représente un défi majeur ; cela nécessite également de tenir compte de la contribution potentielle du microbiote intestinal (70). Depuis quelques années, l'estimation des habitudes alimentaires par opposition à des aliments spécifiques est de plus en plus étudiée vis-à-vis de son impact sur la santé (95–97), cette évaluation pouvant permettre de classer les individus selon leur régime alimentaire ou leur adéquation à des scores prédéfinis (70). En effet, la mesure des habitudes alimentaires permet de saisir les interconnexions complexes dérivées des différentes combinaisons d'aliments que les individus consomment habituellement suivant certains régimes et qui peuvent agir en synergie pour améliorer la santé ou inversement, pour causer des maladies, dont les cancers (74,75).

Les études prospectives sont très pertinentes pour l'identification des biomarqueurs d'effet, puisque l'identification des événements précoces conduisant à la pathologie pourrait avoir un impact sur les stratégies de prévention (58). En effet, si l'on regarde la Figure 9 représentant un schéma de la voie causale de l'initiation à l'apparition de la maladie, pour obtenir le plus grand effet de levier dans le temps et offrir ensuite le plus grand potentiel d'intervention préventive, les biomarqueurs devraient témoigner d'un événement biologique survenant aux premiers stades de la maladie, bien que le « prix à payer » en termes de force d'association soit une réduction de l'association entre le biomarqueur et le résultat clinique au vu de la longue période avant la mise en place de la maladie (76). L'une des principales difficultés dans l'identification de ces biomarqueurs d'effet est de vérifier qu'ils appartiennent à la voie causale entre l'exposition et la maladie, ou s'ils sont simplement un effet secondaire de l'exposition ou de la maladie, ou que leur mesure est faussée par une autre exposition (57). Les analyses de médiations permettent d'obtenir de premiers indices sur cet effet causal puisqu'elles s'intéressent aux mécanismes qui sous-tendent une relation observée entre une variable d'exposition et une variable à expliquer en examinant comment ils se rapportent à une troisième variable intermédiaire, le médiateur (77,78). Cependant, le meilleur moyen de vérifier la causalité est de mener des études d'intervention, mais qui ne sont malheureusement pas toujours possibles d'un point de vue éthique et pratique chez l'homme (79).

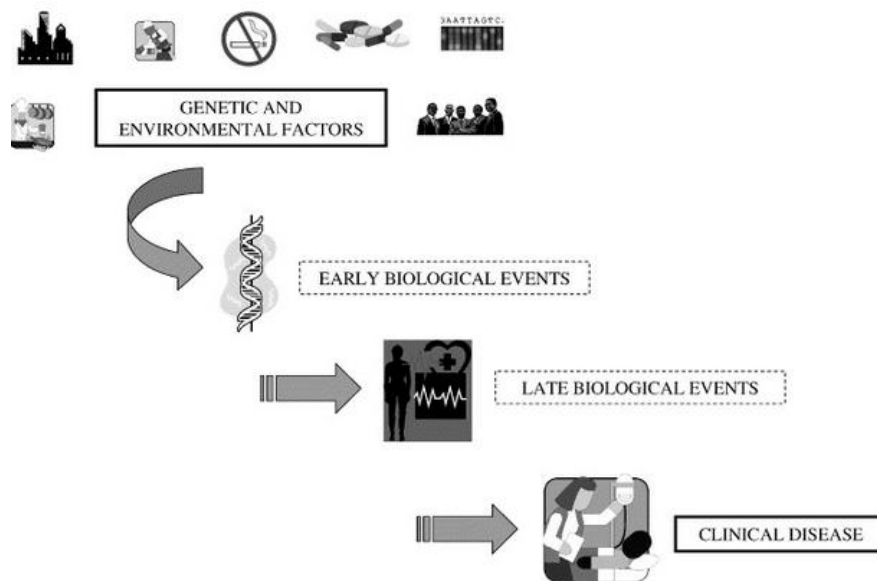


Figure 9 Voie causale de l'initiation à l'apparition de la maladie.  
Source : Bonassi et al. 2002 (80)

### 3. Les biomarqueurs de susceptibilité

Les biomarqueurs de susceptibilité sont liés à différents phénotypes intermédiaires d'une maladie ou à la sévérité de celle-ci (81). Ils peuvent contribuer à l'identification des sous-groupes à risque élevé de la population lorsqu'ils sont utilisés très en amont de la pathologie, mais également au diagnostic de la pathologie ou à son pronostic, notamment dans des contextes de prévention personnalisée. La plupart des auteurs s'entendent pour dire que le concept de causalité n'est pas strictement nécessaire pour considérer ces biomarqueurs comme des prédicteurs (82). Ils peuvent cependant contribuer à la compréhension de l'étiologie en matière d'identification de l'exposition et des mécanismes (58). De tels biomarqueurs peuvent permettre d'améliorer les modèles de prédiction de risque, notamment en combinaison (83). Le dépistage précoce est l'une des tâches les plus difficiles de la prise en charge du cancer ; cependant, si les tumeurs pouvaient être détectées et diagnostiquées avant la dissémination des cellules cancéreuses, le taux de mortalité diminuerait considérablement (84).

Les biomarqueurs génétiques et génomiques peuvent permettre d'identifier des individus dont la déclaration de la maladie sur le plan clinique n'est pas évidente ; les tests de biomarqueurs génomiques peuvent détecter des variations moléculaires dans des gènes isolés, des groupes de gènes ou des génomes entiers (85). Des gènes de la régulation du cycle cellulaire, comme la cycline A et CDK5, peuvent être utilisés pour la détection de multiples cancers et l'évaluation du risque (86). L'estimation de la méthylation de l'ADN mesurée au moment du diagnostic peut s'avérer utile pour les choix de traitement et pour prédire la survie globale (prévention tertiaire) ; par ailleurs, déterminer le rôle des biomarqueurs épigénétiques dans la prévention primaire peut aider à identifier des voies modifiables pour cibler les interventions et réduire l'incidence des maladies (87). Le dépistage des tumeurs au niveau moléculaire à l'aide de biomarqueurs serait très pertinent pour la détection précoce du cancer du sein à l'avenir notamment à l'aide de techniques non invasives, par exemple dans les liquides organiques des patients (84). Quelques exemples de biomarqueurs liés aux cancers du sein et de la prostate sont détaillés ci-dessous.

Les biomarqueurs génomiques du cancer du sein se composent de mutations de gènes rares, très pénétrantes telles que BRCA1 ou BRCA2, de mutations de gènes modérément pénétrantes telles que CHEK2, TP53, STK11/LKB1, CDH1, ainsi que de variantes génomiques plus courantes, dont les polymorphismes d'un seul nucléotide, associées à des effets de faible ampleur (85,88). Les MicroARNs sont stables dans le sérum et présentent des profils d'expression distincts chez les cas de cancers et chez les individus normaux (89), faisant d'eux des biomarqueurs pertinents pour la détection ou le pronostic des cancers. Une revue de la littérature s'intéressant aux biomarqueurs du stress oxydatif, mesurés avant diagnostic, a montré que le 8-hydroxy-2'-deoxyguanosine, un biomarqueur de l'oxygénation de l'ADN, était associé à une augmentation de risque de cancers du sein chez les femmes post-ménopausées (90). Des taux plus élevés d'œstrogènes et d'androgènes endogènes et des taux plus faibles de globuline liant les hormones sexuelles (SHBG) sont liés à un risque accru de cancer du sein post-ménopausique (91). De plus, les associations entre le risque de cancer du sein post-ménopausique et l'hormonothérapie substitutive (92) et l'utilisation efficace d'anti œstrogènes pour prévenir le cancer du sein (93) appuient fermement un rôle causal pour les œstrogènes. L'insuline chroniquement élevée peut augmenter la bioactivité des œstrogènes et promouvoir les activités des adipokines liées au cancer du sein (94) et de l'IGF-1 (95). Une récente méta-analyse a suggéré que la leptine pourrait être un biomarqueur potentiel du risque de cancer du sein chez les femmes, en particulier chez les femmes en surpoids, obèses ou postménopausées (96).

Des données épidémiologiques moléculaires confirment l'existence d'une association entre plusieurs variantes alléliques de certains gènes androgènes-métaboliques et la prédisposition au cancer de la prostate et son évolution (97). Depuis plusieurs années, le taux d'antigène prostatique spécifique de la prostate (PSA) est la mesure standard pour le diagnostic du cancer de la prostate et d'autres maladies prostatiques et des données récentes indiquent que le PSA peut également être utilisé pour déterminer le risque de développer un cancer de la prostate (98). Une revue publiée en 2004 présente les développements récents des biomarqueurs moléculaires et immunohistochimiques utilisés pour la stratification du risque de cancer de la prostate et qui pourraient permettre une approche personnalisée pour les patients atteints de ce cancer (99). De nouveaux biomarqueurs sanguins utilisés pour le diagnostic du cancer de la prostate, comme le Prostate Health Index (PHI) et le 4Kscore, montrent une capacité similaire de prédiction (100). L'antigène 3 du cancer de la prostate (PCA3) est un biomarqueur urinaire dont la prédiction est inférieure à celle du PHI, mais qui peut être associé à d'autres marqueurs comme TMPRSS2-ERG pour améliorer ses performances (100). La Glutathion S-Transférase a été montré comme surexprimée dans les tissus cancéreux du colon et de la prostate (86). De faibles taux d'adiponectine ont été reportés comme un facteur de risque de cancer lié à l'obésité tels que le cancer colorectal et le cancer de la prostate (101) et une récente méta-analyse a révélé des rôles importants de la leptine, de l'adiponectine et de leurs récepteurs dans le développement et la progression du cancer de la prostate (102). Plusieurs facteurs de croissance peptidiques ont également été associés au développement de ce cancer (103).

Les profils génétiques, protéiques et métaboliques peuvent refléter à la fois le génotype et les antécédents d'exposition, y compris des expositions telles que l'alimentation et l'activité physique au cours des mois ou années précédents (104). Ces dernières décennies, la recherche de biomarqueurs dans le contexte d'épidémiologie moléculaire du cancer était très centrée sur l'ADN et les mutations. Des recherches récentes ont mis en évidence plusieurs étapes intermédiaires supplémentaires entre le génotype et le phénotype et l'importance de l'expression/modulation des gènes pour la cancérogenèse, axant donc les



nouvelles recherches sur des biomarqueurs de ces étapes intermédiaires et sur les altérations épigénétiques qui affectent l'expression des gènes et des protéines (57).

Par ailleurs, l'intégration de biomarqueurs issus de données omiques dans l'épidémiologie moléculaire du cancer présente un intérêt certain car ces sciences permettent des investigations massives non basées sur des hypothèses *a priori*, contrairement aux méthodes traditionnelles (sans utilisation de biomarqueurs), pouvant être utilisées à des fins étiologiques (pour étudier les causes et les mécanismes d'apparition de la maladie) ou cliniques (diagnostic précoce, pronostic et suivi) (57). Les omiques et notamment, la protéomique, la lipidomique et la métabolomique, des méthodes de profilages non invasives, sont considérées comme une grande innovation pour la découverte de nouveaux biomarqueurs d'exposition, d'effet et de susceptibilité (81) - les trois grandes catégories de biomarqueurs citées précédemment. Le profilage métabolomique de l'urine et du sang offre notamment des possibilités immenses pour l'identification de biomarqueurs nutritionnels supplémentaires (104).

#### **D. La métabolomique : un outil innovant pour l'identification de biomarqueurs à grande échelle**

Les approches de profilage métabolique, comme la métabolomique, permettent une caractérisation détaillée et efficace des échantillons biologiques humains. L'application aux études de population est très prometteuse pour découvrir de nouvelles associations et des relations causales entre les facteurs environnementaux et les maladies chroniques (47). Durant la dernière décennie, la métabolomique a bénéficié de formidables avancées en matière d'outils analytiques à haut débit telles que la spectroscopie de résonance magnétique nucléaire (RMN) et la spectrométrie de masse (MS), mais également en matière de chimiométrie et de bioinformatique. Cette discipline a le pouvoir de mesurer simultanément plusieurs centaines, voire milliers, de macro- et micro-molécules (les métabolites endogènes et exogènes présents dans les cellules, les tissus et les biofluides) et donc de surveiller le fonctionnement de multiples voies métaboliques (105–107). Les métabolites sont les produits finaux des processus de régulation cellulaire et leurs niveaux peuvent être considérés comme la réponse ultime des systèmes biologiques aux changements génétiques ou environnementaux ; la métabolomique permet de faire le lien entre génotypes et phénotypes (108).

Les technologies omiques à haut débit permettent d'intégrer un large éventail d'expositions à partir d'un petit nombre de matrices biologiques voire même dans une seule mesure (109). Les méthodes utilisées pour la découverte de biomarqueurs peuvent être divisées en deux grandes catégories : fondées sur des hypothèses ou fondées sur les données. Dans les deux cas, la métabolomique peut être appliquée. Dans l'approche fondée sur des hypothèses, on utilise des connaissances préalables sur le biomarqueur ou une série de biomarqueurs candidats, comme par exemple, des données sur la composition des aliments pour les biomarqueurs nutritionnels, puis des méthodes sont développées pour mesurer ces biomarqueurs candidats (51), comme par exemple, la métabolomique « ciblée ». Jusqu'à présent, pour les biomarqueurs nutritionnels, cette approche a essentiellement été appliquée à des familles spécifiques de constituants alimentaires tels que les acides gras ou les caroténoïdes (110,111). Dans l'approche axée sur les données, il n'y a pas de connaissance préalable des biomarqueurs et un grand nombre de métabolites est mesuré, les principales limites étant la capacité de l'instrument analytique à les détecter. Cette approche « non ciblée » ne fait donc pas d'hypothèse *a priori* concernant les sources d'exposition qui sont causales pour une maladie particulière par exemple, mais s'appuie en général sur des comparaisons de profils pour trouver des biomarqueurs discriminants qui, une fois identifiés *a posteriori*, seront analysés dans d'autres études pour en déterminer leurs sources (112). Cette approche est très utile pour découvrir de nouveaux

biomarqueurs exogènes (alimentation, pollutions, médicaments) mais aussi ceux liés aux facteurs endogènes, au microbiote et à la progression des maladies (51). Cette approche a été utilisée pour découvrir de nouveaux biomarqueurs pour un certain nombre d'aliments, de nutriments ou de régimes alimentaires (51).

Bien que les approches classiques fondées sur des hypothèses aient pu permettre de nombreuses découvertes, notamment les nutriments essentiels, elles sont moins adaptées pour comprendre le rôle de composés non essentiels très divers dans les aliments ni pour caractériser les régimes alimentaires plus globalement (51). Les approches fondées sur les données offrent la possibilité d'explorer les interactions complexes entre l'alimentation et l'organisme humain et de distinguer les individus ou les populations exposés à différents environnements, modes de vie ou régimes alimentaires (113). Traditionnellement, des biomarqueurs uniques ont été en général utilisés pour caractériser des expositions alimentaires complexes, cependant des combinaisons de biomarqueurs peuvent fournir des mesures plus précises de l'exposition (51) et également augmenter la sensibilité et spécificité pour évaluer la susceptibilité à une maladie par exemple. Toutefois, il existe très peu d'exemples où des combinaisons de biomarqueurs ont été utilisées pour améliorer la spécificité des mesures de l'exposition par voie alimentaire (51). Certains biomarqueurs peuvent être communs à plusieurs aliments ou caractéristiques d'un groupe alimentaire entier ; par exemple, la vitamine C et un certain nombre de caroténoïdes et de flavonoïdes sont communs à de nombreux fruits et légumes (51) et ont été utilisés comme biomarqueurs génériques de l'apport en fruits et légumes (110,114).

Plusieurs bases de données sont en cours d'élaboration permettant de répertorier l'ensemble de ces biomarqueurs identifiés par des analyses métabolomiques. La base HMDB (Human Metabolome Database) (115) est une base de données en ligne qui ne cesse de s'enrichir et contenant actuellement plus de 114 000 métabolites, dont des composés endogènes, microbiens, biotransformés et exogènes/xénobiotiques. La base FoodDB (Food Component Database) (116) est une base de données de plus de 28 000 constituants alimentaires, y compris les additifs alimentaires artificiels. Une grande partie des données de FoodDB est maintenant dans la HMDB, mais FoodDB fournit des informations supplémentaires sur les sources alimentaires et les concentrations alimentaires. D'autres bases de données existent telles que Phenol-Explorer (117), PhytoHub (118), ECMDB (E. coli Metabolome Database) (119) et YMDB (Yeast Metabolome Database) (120); exposées dans la revue de Scalbert et al. (51).

#### IV. Etat de l'art des études ayant utilisé la métabolomique dans le champ de l'épidémiologie nutritionnelle

De nombreuses études d'association pangénomique (GWAS : Genome Wide Association Study) analysant de nombreuses variations génétiques dans de grandes cohortes en population humaine ont été menées dans un contexte de révolution dans le génotypage à haut débit, des approches très efficaces pour la détection de mutations associées aux maladies ou phénotypes intermédiaires (121). Le concept parallèle de ce type d'étude a été établi en métabolomique (MWAS : Metabolome Wide Association Study) afin de prendre en compte les interactions gènes-environnement agissant aux niveaux individuel et populationnel (121). La première preuve de principe de cette approche a été démontrée pour l'hypertension artérielle (121) mais les MWAS sont encore peu nombreuses dans le domaine de la prévention primaire malgré le fait qu'il existe déjà un énorme potentiel inexploité pour ce type d'études. En effet, des dizaines d'études épidémiologiques longitudinales ont été réalisées au fil des décennies, couvrant un éventail de maladies avec des échantillons de biofluides stockés (121). De plus, une étude méthodologique a montré que la

précision analytique et la variabilité interindividuelle des profils métabolomiques étaient adaptées pour permettre l'utilisation de ce type de marqueurs en épidémiologie (122).

La spectrométrie de masse (MS) couplée à la chromatographie en phase gazeuse ou liquide et la Résonance Magnétique Nucléaire (RMN) sont actuellement les technologies les plus utilisées pour l'analyse du métabolome et constituent des méthodes complémentaires (51). La MS est la technique la plus sensible et notamment capable de couvrir les métabolites non nutritifs présents à de faibles concentrations dans les échantillons biologiques. La RMN est une technique robuste et non destructive, mais possède une sensibilité relativement faible réduisant sa couverture du métabolome alimentaire aux nutriments, sucres et métabolites microbiens prédominants (51). Le couplage de la MS avec la chromatographie en phase gazeuse a été utilisé pour analyser des constituants tels que les acides phénoliques ou les acides gras (111,123,124). La chromatographie liquide couplée à la MS reste à ce jour la technique la plus utilisée dans les études sur le métabolome alimentaire (51). Cependant, aucune méthode chromatographique ne peut à elle seule couvrir le large éventail de polarité existant pour les composés du métabolome alimentaire (51).

## **A. Métabolomique et biomarqueurs nutritionnels**

### *1. Design des études*

Les échantillons biologiques à analyser peuvent provenir soit d'essais d'interventions alimentaires contrôlées, soit d'études d'observation. Le choix du design de l'étude dépend de l'hypothèse scientifique de recherche, ainsi que du temps disponible, des questions éthiques et des ressources économiques (125).

Les interventions peuvent concerner un apport d'un nutriment particulier, d'un aliment entier, d'un groupe alimentaire, d'un régime complet ou encore d'un composé bioactif et/ou d'un placebo. Ces éléments sont administrés aux participants à des occasions uniques (études « aiguës ») ou à domicile à intervalles réguliers pendant une certaine période (de quelques jours : court terme, à quelques mois ou années : moyen/long terme). Dans les essais d'intervention aiguë, les échantillons sont en général collectés de manière postprandiale dans un délai de 24h après consommation du composé d'intérêt. Les biomarqueurs identifiés dans les études d'intervention aiguë doivent idéalement être validés par des études d'intervention plus longues pour s'assurer de l'aspect dose-réponse et être reproduits dans différentes populations (51,126). Dans les interventions à intervalles réguliers, les échantillons sont recueillis à la fin de la période d'intervention et sont souvent comparés entre les sujets ayant consommé le ou les composés et ceux que ne l'ont pas consommé. Les échantillons peuvent également être collectés avant et après l'intervention et chaque participant exerce son propre contrôle (51). Les essais contrôlés randomisés en double aveugle sont considérés comme le « gold-standard » des études nutritionnelles. Dans ce type d'étude, les sujets sont assignés au hasard à des groupes de traitement et cette attribution est dissimulée aux participants et aux chercheurs. Cependant, ce type de design est souvent impossible à mettre en place dans les études d'intervention alimentaire, en particulier, lorsque des aliments ou des repas sont testés (125). Les études d'intervention sont souvent difficiles à élaborer sur le long terme car les participants ne sont pas en conditions réelles et peuvent finir par se lasser (127) et leur propre régime alimentaire peut interférer avec l'intervention (51). En outre, ils sont éthiquement impossibles à mettre en place sur le long terme dès lors qu'un effet délétère est suspecté.

Les études d'observation, où les individus sont en conditions réelles, jouent un rôle important dans la découverte de biomarqueurs puisqu'elles permettent de fournir des informations sur les habitudes alimentaires réelles (126). Ces études permettent de générer des hypothèses et d'explorer les associations

entre l'alimentation et la santé mais elles ne peuvent toutefois pas prouver catégoriquement l'existence d'un lien causal (125). Les consommations alimentaires sont en général recueillies à l'aide de fréquentiels alimentaires ou de rappels ou enregistrements de 24h répétés. Elles permettent de comparer des individus qui consomment faiblement ou fortement des nutriments/aliments ou encore des régimes différents puis de caractériser des biomarqueurs objectifs qui reflètent ces apports habituels à condition qu'ils aient une demi-vie suffisante dans l'organisme ou que les aliments soient consommés régulièrement (51,126). Il est important de noter que de nombreux aliments consommés sont fortement corrélés et qu'il existe un risque d'identification de biomarqueurs qui ne sont pas spécifiques de l'aliment en question à moins que leur identité et leur présence spécifique dans les aliments considérés soient établies (51). Un certain nombre d'études observationnelles sont transversales, cependant il est aussi possible d'étudier les variations longitudinales des concentrations de métabolites et leurs associations avec l'alimentation et la santé (51).

## 2. *Etat des connaissances*

De nombreuses revues de la littérature ont recensé la majorité des biomarqueurs alimentaires présumés identifiés à l'aide de la métabolomique dans des études d'intervention ou d'observation (51,126,128,129). La majorité de ces métabolites ont été associés à la consommation de nutriments tels que les fibres (130,131), les protéines (132) ou des isolats de protéines de lactosérum (133) ou d'aliments/groupes d'aliments comme les agrumes (134–137), les crucifères (138,139), la poudre de cacao (140,141), les noix (142), le café (143,144), le vin rouge et le raisin (123,145), le thé et notamment le thé noir et le thé vert (134,146–148) ou encore le poisson (144,149). Certains métabolites ont été retrouvés dans plusieurs études comme l'acide hydroxynicotinique pour la poudre de cacao, le *N*-Méthylpyridinium et la trigonelline pour le café, l'acide hippurique pour le thé ou la méthylhistidine pour le poisson. La proline bêtaïne est un bon exemple de biomarqueur de la consommation d'agrumes. Ce marqueur a été identifié dans des études d'intervention aiguë à petite échelle et validé chez des sujets en conditions réelles dans 2 études indépendantes (136,137) puis a été observé dans une étude transversale ayant utilisé la métabolomique non ciblée (135). En revanche, un certain nombre de biomarqueurs présumés de groupe d'aliments spécifiques n'ont pas été validés dans suffisamment d'études indépendantes ayant des design différents (51). D'autres biomarqueurs peuvent s'avérer moins robustes en raison de l'existence possible d'une variété de précurseurs, comme l'oxyde de triméthylamine (TMAO) étant un biomarqueur présumé de la consommation de viandes (150–152) mais qui a été identifié par d'autres études (149,153) comme biomarqueur de la consommation de poisson avec une réponse plus sensible que celle de la consommation de viande (138). De plus, plusieurs précurseurs alimentaires du TMAO tels que la choline ou la carnitine ont été découverts (154) nécessitant donc d'interpréter avec prudence l'origine des variations circulantes de TMAO dans les populations (51). Il est également important de souligner l'identification d'un bon nombre des biomarqueurs présumés n'a pas été entièrement validée par des standards chimiques en raison de l'indisponibilité sur le marché. Enfin, une majorité de ces études a utilisé des échantillons d'urine plutôt que des échantillons sanguins en raison des concentrations plus élevées de métabolites d'origine alimentaire dans l'urine que dans le sang. Ceci contraste avec l'utilisation privilégiée des prélèvements sanguins pour mesurer les biomarqueurs de l'état nutritionnel dans les études épidémiologiques (51).

Au cours des dernières décennies, les régimes alimentaires et la qualité de l'alimentation globale ont suscité de plus en plus d'attention, notamment en raison des effets cumulatifs de multiples composants alimentaires pouvant agir en synergie et avoir un impact sur la santé (74,75). En raison de la nature holistique de la métabolomique, elle a naturellement été appliquée depuis les deux dernières décennies

pour rechercher des associations entre les profils métaboliques urinaires, plasmatiques ou sériques, avec l'alimentation globale (155). L'ensemble de ces études sont présentées dans le Tableau 4.

Quelques études d'interventions alimentaires à court ou moyen terme ont analysé les changements dans les profils métaboliques après la consommation de régimes tels que des régimes à base de viande (150), des régimes végétariens (150), le "New Nordic diet" (riche en fruits, légumes, produits riches en fibres et poissons) et le "Average Danish diet" (riche en produits importés et transformés) (134,156,157), des régimes « sains » ou riches en fibres (158,159), des régimes basés sur les recommandations de l'OMS (160), le régime méditerranéen (161) et d'autres régimes plus divers (162,163). De manière individuelle, ces études ont pu mettre en évidence des métabolites caractéristiques de ces régimes. Malgré la diversité de ces derniers entre les études, certains semblent être associés de manière cohérente avec des régimes comportant des similarités, ils sont cependant très peu nombreux. Par exemple, l'hippurate a été associé à des régimes basés sur aux recommandations nutritionnelles de l'OMS et de l'American Heart Association (160,161), à un régime « sain » (159), au régime « New Nordic » (134,156), ou encore à un régime ayant un index glycémique faible (163). Une étude d'intervention à long terme (3 ans) sur des individus ayant un haut risque cardiovasculaire a étudié l'impact d'un régime méditerranéen supplémenté en huile d'olive extra vierge ou en noix par rapport à un régime pauvre en graisse sur le profil métabolique urinaire (164). Elle a notamment montré que le régime méditerranéen augmentait le taux de certains métabolites issus du métabolisme microbien (la phénylacetylglutamine et le p-cresol) et que le régime pauvre en graisse était associé notamment à l'hippurate et le TMAO. Ces résultats sont très pertinents pour démontrer l'impact des régimes alimentaires sur le métabolisme de l'hôte et du microbiote ; cependant, ce sont des interventions alimentaires très spécifiques, en général dans des conditions expérimentales qui ne reflètent pas les habitudes alimentaires des individus et la durée est encore trop limitée pour évaluer un potentiel impact sur le risque de développer une maladie chronique.

Dans ce contexte, il est donc difficile de conclure sur l'impact à long terme de ces régimes et de prendre en compte les facteurs endogènes et de susceptibilité. Plus récemment, quelques études d'observation transversales ont comparé les profils métaboliques majoritairement plasmatiques et sériques aux habitudes alimentaires globales, réelles et complexes et à la qualité de l'alimentation des individus. Ces résultats peuvent donc refléter directement l'exposition alimentaire mais également son effet sur le métabolisme ou le statut en composés alimentaires. Ces études d'observation ont notamment comparé des régimes lactovégétarien et omnivore (165), des régimes « sains » (166–168) et « occidental » (166–169), des consommateurs de viandes ou de poissons (170), des régimes végétariens (169,170) ou végétariens (170), des régimes basés sur l'adéquation à divers scores (Healthy Eating Index 2010, Alternate Mediterranean Diet Score, WHO Healthy Diet Indicator, Baltic Sea Diet) (171), le régime méditerranéen (172) et d'autres régimes divers (132,162,166,173–177). Playdon et al. ont montré que les métabolites étaient plus fortement corrélés aux aliments qu'aux macro- ou micronutriments qui sont moins spécifiques (171). Malgré le fait que l'on puisse observer quelques tendances comme par exemple, une augmentation des acides aminés branchés circulants associés à des régimes plutôt « occidentaux » ou la phénylacetylglutamine associée à des régimes plutôt « sains », il reste très difficile de comparer ces études, en premier lieu à cause de la diversité des régimes étudiés et notamment la diversité de leur méthode d'estimation (cf. Tableau 4). En effet, certaines études sont basées sur des approches *a priori* en utilisant des scores d'adéquation ou des régimes alimentaires prédéfinis et d'autres sur des approches *a posteriori* (habitudes alimentaires observées dans la population d'étude avec des méthodes basées sur les données) comme par exemple, l'analyse en composante principale ou encore le partitionnement en k-moyennes. De plus certaines méthodes comme la Reduced Rank Regression qui permet de construire des

combinaisons des expositions nutritionnelles et des métabolites en maximisant la variabilité des expositions nutritionnelles rendent les résultats difficiles à interpréter en termes de régimes habituels. Floegel et al. ont utilisé cette méthode d'analyse et les résultats obtenus semblaient plutôt refléter des processus endogènes que l'exposition alimentaire (174).

De manière générale, ces études, encore peu nombreuses, qui s'intéressent aux régimes alimentaires plutôt qu'à des aliments spécifiques montrent des résultats assez mitigés ou non reproduits et, à ce jour, il ne semble pas être possible d'établir une signature robuste des régimes alimentaires habituels. Cet état des connaissances dans la littérature peut s'expliquer par différents aspects. La variété des méthodes analytiques et de prétraitement des données non normalisées en RMN ou en spectrométrie de masse - deux méthodes de profilage déjà très distinctes - influençant également la couverture des métabolites étudiés qui diffère selon l'approche (ciblée/non ciblée) ou le protocole (par exemple LC-MS versus GC-MS) pourrait expliquer certaines incohérences entre les études (178). La diversité des populations étudiées dans le monde ayant des habitudes alimentaires très variables impactant directement leur méthode d'estimation est également à prendre en compte. Une standardisation des approches serait donc nécessaire afin de pouvoir mener de nombreuses autres études plus reproductibles permettant de discriminer les individus selon leurs habitudes alimentaires générales grâce à leurs profils métaboliques dans le but de mieux comprendre l'impact de l'alimentation sur le métabolisme et les liens avec la santé et compléter les mesures d'expositions nutritionnelles.

Tableau 4 Etudes épidémiologiques ayant investigué les profils métaboliques associés à des régimes alimentaires

Régime alimentaire	Type d'étude	Effectif	Estimation des régimes	Biofluide	Technique analytique	Méthodes statistiques pour comparaison des régimes	Métabolites	Référence
- Régime faible en viande - Régime à base de viande - Régime végétarien	ICM Hommes blancs, non-fumeurs, de poids normal, de 25 à 74 ans	12	Adhérence au régime	Urine	H-RMN	ACP, O-PLS-DA	Taurine, carnitine, acetylcarnitine, 1-methylhistidine, 3-methylhistidine, TMAO plus élevés après une forte consommation de viande. p-Hydroxyphenylacetate plus élevé après un régime végétarien par rapport à un régime à base de viande.	Stella et al., J Proteome Res, 2006 (150)
- Hommes lactovégétariens - Femmes lactovégétariennes - Hommes omnivores - Femmes omnivores	OT Hommes et femmes en bonne santé de 18 à 55 ans	161	Adhérence au régime	Urine	H-RMN	ACP	TMAO, diméthylamine, phenylalanine, methylhistidine caractéristiques du régime omnivore. Citrate caractéristique du régime lactovégétarien.	Xu et al., Anal Bioanal Chem, 2010 (152)
- Cluster 1: régime "sain" riche en légumes, pain complet et poisson - Cluster 2: régime riche en fruits - Cluster 3: régime "occidental" à base de viandes, pain blanc, beurre avec une consommation faible en fruits et légumes et pain complet	OT Hommes et femmes en bonne santé de 18 à 63 ans. Participants issus d'une intervention en double aveugle, randomisée et contrôlée par placebo sur la vitamine D et les probiotiques	160	Partitionnement en k-moyennes	Plasma et urine	H-RMN	ACP, O-PLS-DA	Phenylacetylglutamine marqueur de la consommation de légumes. O-acetylcarnitine marqueur de la consommation de viande rouge. Glycine, phenylacetylglutamine, acetoacetate plus élevés dans le cluster 1. TMAO, O-acetylcarnitine et N,N-diméthylglycine plus élevés dans le cluster 3.	O'Sullivan et al., Am J Clin Nutr, 2011 (166)
- Composante 1: consommation élevée en volaille, poisson, riz, légumes, fruits, chocolat, flocons d'avoine, lait caillé, fromage, lait et faible en viande et saucisses - Composante 4: régime traditionnel Bavarois : consommation élevée en viande, volaille, poisson, snacks et oeufs et faible en pomme de terre, pâtes, pain complet et flocons d'avoine - Composante 6: consommation élevée en pain complet, oeuf, lait et faible en poisson, fruits frais et gâteaux	OT Hommes en bonne santé de 55 à 79 ans	284	ACP	Plasma	MS/MS ciblée	Régression linéaire	Composante 1 associée à diacyl-phosphatidylcholines avec 4 et 5 liaisons et acyl-alkyl- phosphatidylcholines avec 1 ou 2 doubles liaisons, et à une diminution dans la saturation des acides gras des glycerol-phosphatidylcholines. Composantes 4 et 6 associées à plusieurs phospholipides.	Altmaier et al., Eur J Epidemiol, 2011 (173)

- Régime riche en composés phytochimiques (crucifères, agrumes, soja) - Régime sans fruits et légumes	ICM, OT Hommes et femmes en bonne santé, non-fumeurs, de 21 à 45 ans	ICM=10, OT=93	Adhérence au régime et niveau de consommation	Urine	LS-MS/MS	ACP, modèle linéaire - analyse d'expression différentielle	Proline betaine marqueur de la consommation d'agrumes, régime riche en composés phytochimiques plus élevés en sulforaphane, acide hippuric, genistein, daidzein, equol, glycitein, O-desmethylangolensin, enterolactone, trigonelline	May et al., Br J Nutr, 2013 (162)
- Régime "sain" (riche en légumes, fruits, produits complets) - Régime "occidental" (riche en produits céréaliers, desserts, bonbons, charcuterie)	OT Participants en bonne santé de 18 à 50 ans	210	ACP	Plasma	MS ciblée	ACP, corrélations de Pearson	Régime "occidental" associé à acylcarnitines à courtes-chaînes, leucine, phenylalanine. Régime "sain" associé acylcarnitines à moyennes et longues chaînes.	Bouchard-Mercier et al., Nutr J, 2013 (167)
- Régime riche en beurre et faible en margarine - Régime riche en viande rouge et poisson et faible en pain complet et thé - Régime riche en pomme de terre, produits laitiers et cornflakes	OT Participants en bonne santé de 35 à 65 ans	2380	Reduced Rank Regression	Sérum	MS ciblée	Reduced Rank Regression	Acylcarnitines, acides aminés, hexoses, diacyl-, acyl-alkyl- lysophosphatidylcholines, et sphingomyelines associés au régime riche en beurre et faible en margarine. Hexoses et phosphatidylcholines associés au régime riche en viande rouge et poisson et faible en pain complet et thé. Méthionine et acides aminés branchés associés au régime riche en pomme de terre, produits laitiers et cornflakes.	Floegel et al., Eur J Clin Nutr, 2013 (174)
Selon des scores nutritionnels : - Consommation de fruits et légumes - Consommation d'alcool - Régime faible en viande - Régime hypocalorique - Régime traditionnel Anglais	OT Femmes en bonne santé ayant une moyenne d'âge de 59.5 ans	1003	ACP	Sérum	MS ciblée	Corrélations	Fruits et légumes associés aux glycerophospholipides (phosphatidylcholine-dyacyl C38:6 et Sphingomyeline C26:1)	Menni et al., Metabolomics, 2013 (132)
- "New Nordic diet" riche en choux, racines, baies et légumes - "Average Danish diet" diffère en consommation de fruits et légumes et notamment plus pauvres en ces aliments	ICM Participants avec un tour de taille élevé et au moins 1 facteur de risque du syndrome métabolique de 18 à 65 ans	107	Adhérence au régime	Urine	UPLC-qTOF-MS	T-test appariés	TMAO associé à la consommation de produits de la mer. « Nordic diet » associé au TMAO, hydroquinone-glucuronide, acide hippurique, 3,4,5,6-tetrahydrohippurate	Andersen et al., Anal Bioanal Chem, 2014 (134)



<ul style="list-style-type: none"> <li>- "New Nordic diet" (NND) riche en choux, racines, baies et légumes</li> <li>- "Average Danish diet" (ADD) diffère en consommation de fruits et légumes et notamment plus pauvres en ces aliments</li> </ul>	ICM Participants avec un tour de taille élevé et au moins 1 facteur de risque du syndrome métabolique de 18 à 65 ans	107	Adhérence au régime	Urine	UPLC-qTOF-MS	PLS-DA	ADD associé à pyrrolidine, theobromine, acide 7-méthylurique, 6-amino-5-[N-méthylformylamino]-1- méthyluracil, acide 3,7-diméthylurique, 7-méthylxanthine, pyroglutamyl proline, p-menth-1-ène-6,8,9-triol, perillic acid-8,9-diol-glucuronide, limonene-8,9-diol-glucuronid, perillic acid glucuronide, dihydroperillic acid glucuronide, limonene-1,2-diol glucuronide, octanoyl-glucuronide, acide 3-indoleacétique glucuronide. NND associé hydroquinone glucuronide, TMAO, acide hippurique, (2-oxo-2,3-dihydro-1H-indol-3-yl) acétique acide, 3,4,5,6-tetrahydrohippurate	Andersen et al. J.Proteome.Res. 2014 (156)
<ul style="list-style-type: none"> <li>- Régime «sain» ; fibres, poissons gras, myrtilles</li> <li>- Produits riches en fibres</li> <li>- Témoin : pain blanc</li> </ul>	ICM Participants ayant un métabolisme du glucose diminué et des facteurs de risque du syndrome métabolique de 40 à 70 ans	106	Adhérence au régime	Sérum	H-RMN	Modèle mixte et ANCOVA	Régime "sain" augmente la poly-insaturation des acides gras plasmatiques, augmente le taux d'acides gras n-3 et diminue le taux d'acides gras n-6 et n-7. Consommation de poisson associée aux composantes HDL.	Lankinen et al., PloS One, 2014 (158)
<ul style="list-style-type: none"> <li>- Régime «sain» ; fibres, poissons gras, myrtilles</li> <li>- Produits riche en fibres</li> <li>- Témoin : pain blanc</li> </ul>	ICM Participants ayant un métabolisme du glucose diminué et des facteurs de risque du syndrome métabolique de 40 à 70 ans	106	Adhérence au régime	Plasma	LC-qTOF-MS	Corrélations de Spearman et ANCOVA	Glucuronidated alk(en)ylresorcinols plus élevé dans les deux groupes d'intervention et corrélé aux produits riches en fibres. Régime "sain" augmente les acides gras furaniques, l'acide hippurique et diverses espèces de lipide comme les acides gras polyinsaturés. Acides gras furaniques associés à la consommation de poisson.	Hanhineva et al., J Nutr, 2015 (159)
<ul style="list-style-type: none"> <li>- Consommateurs de viande</li> <li>- Consommateurs de poissons</li> <li>- Végétariens</li> <li>- Végans</li> </ul>	OT Hommes en bonne santé de plus de 20 ans	379	Adhérence au régime	Plasma	LC-MS/MS ciblée	ANOVA	6 acylcarnitines, 61 glycérophospholipides, et 12 sphingolipides plus élevés chez les consommateurs de viande et moins élevés chez les végétariens. Acides aminés et amines biogènes plus élevés chez les consommateurs de poisson et végétariens	Schmidt et al., Am J Clin Nutr, 2015 (170)

<ul style="list-style-type: none"> <li>- Régime Méditerranéen supplémenté en huile d'olive extra vierge</li> <li>- Régime Méditerranéen supplémenté en noix</li> <li>- Témoin : régime pauvre en graisse</li> </ul>	IL participants non diabétiques ayant un haut risque cardiovasculaire de 55 à 80 ans	98	Adhérence au régime	Urine	H-RMN	O-PLS-DA et classification ascendante hiérarchique	Régimes méditerranéens : glucides, 3-hydroxybutyrate, citrate, cisaconitate, créatine, créatinine, acides aminés (proline, N-acetylglutamine, glycine, acides aminés branchés et métabolites dérivés), lipides (acides oléique et subérique), et métabolites du microbiote (phenylacetylglutamine et p-cresol). Régime témoin faible en graisse : hippurate, TMAO, histidine, et dérivés (methylhistidines, carnosine, et anserine), et xanthosine.	Vázquez-Fresno R et al., J Proteome Res, 2015 (164)
4 régimes selon les recommandations de l'OMS (moins de sucre, de sel, de graisse totale et plus de fibres, fruits, légumes, produits complets), régime 1 étant le plus concordant avec les recommandations et le 4 le moins concordant	ICM Participants en bonne santé de 21 à 65 ans	19	Adhérence au régime	Urine	H-RMN	PLS-DA	Régime 1: plus élevé en hippurate (marqueur des fruits et légumes), N-acetyl-S-methyl-cysteine-sulfoxide (marqueur des crucifères), diméthylamine, et TMAO (poisson) et 1-methylhistidine et 3-methylhistidine (poisson gras et volaille), et créatine. Régime 4: plus élevé en O-acetylcarnitine, carnitine (viande rouge) et en glucose	Garcia-Perez I et al., Lancet Diabetes Endocrinol, 2017 (160)
2 régimes faible en énergie : <ul style="list-style-type: none"> <li>- Régime méditerranéen</li> <li>- Régime témoin (basé sur les recommandations de l'American Heart Association)</li> </ul>	ICM Participants ayant des facteurs de risque du syndrome métabolique	72	Adhérence au régime	Plasma	LC-QTOF/MS	T-test appariés	Régime méditerranéen : plasmalogène phosphatidylcholine (P-18:1/20:3) après 2 mois d'intervention, et acide palmitique après 6 mois d'intervention. Pphosphatidyl-choline (P-18:1/20:3) augmenté chez le régime méditerranéen et diminué dans le groupe témoin après 2 mois d'intervention	Bondia-Pons et al., Mol Nutr Food Res, 2015 (161)
Selon les scores de 4 index : <ul style="list-style-type: none"> <li>- Healthy Eating Index (HEI) 2010</li> <li>- Alternate Mediterranean Diet Score (aMED)</li> <li>- WHO Healthy Diet Indicator (HDI)</li> <li>- Baltic Sea Diet (BSD)</li> </ul>	OT Hommes fumeurs (≥5 cigarettes/jours) de 50 à 69 ans issus d'études cas témoins nichés	1336	Adéquation aux recommandations	Sérum	UPLC-LC-MS/MS	Corrélations de Spearman	HEI-2010 et HDI associés à des acides aminés, des cofacteurs ou vitamines, des lipides, et des xenobiotiques exogènes. aMED associé à des acides aminés, un glucide, des cofacteurs ou vitamines, des lipides, des xenobiotiques exogènes. BSD associé à des acides aminés, un glucide, des cofacteurs ou vitamines, des lipides. La voie des lysolipide contenait le plus de métabolites associés à la qualité de l'alimentation.	Playdon et al., Am J Clin Nutr, 2017 (171)

<ul style="list-style-type: none"> <li>- Composante 1 énergétique</li> <li>- Composante 2 aliments végétaux versus aliments animaux</li> <li>- Composante 3 régime moderne avec plus de sucre versus régime traditionnel Danois avec plus de viandes, poissons, œuf, alcool, café</li> <li>- Composante 4 régime moderne avec plus de volaille, huile végétale, bonbons versus régime traditionnel Danois avec des pommes de terre, des fibres, des produits laitiers</li> <li>- Composante 5 produits à base de lait écrémé versus produits à base de lait entier</li> </ul>	OT Participants de 18 à 67 ans	969	ACP	Plasma	H-RMN	ACP, PLS-DA	<p>Composante 1 : lipides associés à une consommation énergétique plus élevée et phosphatidylcholine associée à un moindre apport énergétique.</p> <p>Composante 2 : produits animaux associés à des acides aminés, citrate et polyol.</p> <p>Composante 3 : produits animaux associés à des acides aminés et phosphatidylcholine associée au régime traditionnel.</p> <p>Composante 4 : acides aminés et phosphatidylcholine associés au régime traditionnel.</p> <p>Composante 5: tyrosine et histidine.</p>	Peré-Trepat Chemom Intell Lab Syst. 2010 (175)
Régime riche en viande, sauce, saucisse, œuf, glace et faible en noix, céréales, champignon et légumineuses	OT Hommes et femmes en bonne santé de 18 à 80 ans	298	Reduced Rank Regression	Plasma	MS ciblée	Reduced Rank Regression	Régime associé à la lysine, alanine, phénylalanine, tryptophane, tyrosine, isoleucine, leucine, valine.	Merz B Nutrients. 2018 (176)
<ul style="list-style-type: none"> <li>- "New Nordic diet" (NND) riche en fruits, légumes, produits riches en fibres et poisson</li> <li>- "Average Danish diet" (ADD) riche en produits importés et transformés</li> </ul>	ICM Participants avec un tour de taille élevé et au moins 1 facteur de risque du syndrome métabolique de 18 à 65 ans	146	Adhérence au régime	Plasma	UPLC-qTOF-MS	PLS-DA avec sélection de variable	<p>NND : acide pipécolique, betaine, TMAO, prolyl hydroxyproline, polyunsaturated phosphatidylcholines.</p> <p>ADD : théobromine et proline, betaine, acide indolelactique, hydroxy-3-méthylbutyrate et butyryl carnitine.</p>	Acar Mol Nutr Food Res 2019 (157)
<ul style="list-style-type: none"> <li>- Cluster 1, régime "sain" : céréales du petit déjeuner, lait écrémé et allégé, pommes de terre, fruits, poissons et produits de la mer</li> <li>- Cluster 2, régime "occidental"; chips, pommes de terre transformées, produits à base de viande, produits apéritifs, boissons énergétiques</li> </ul>	OT Hommes et femmes de plus de 18 ans	567	Mann-Whitney U test	Urine	H-RMN	Analyse de Cluster, PLS-DA, T-test	<p>Cluster 1 : hippurate, betaine, N-phenylacetylglutamine, anserine, citrate, 3-hydroxybutyrate, 2-aminoadipate.</p> <p>Cluster 2 : créatinine, glycylproline, théophylline, N-acetylglutamate, tryptophane.</p>	Gibbons Mol Nutr Food Res. 2017 (168)

- Forte adhérence au régime méditerranéen (H-MDA) - Faible adhérence au régime méditerranéen (L-MDA)	OT Hommes de 55 à 80 ans et femmes de 60 à 80 ans ayant des antécédents cardiovasculaires et des facteurs de risque cardiovasculaires. Participants issus d'une intervention, randomisée et contrôlée	119	Adhérence au régime	Urine	H-RMN	OSC-PLS-DA, T-test	H-MDA : plus élevé en anserine, carnosine, creatine, creatinine, guanidoacetate, histidine, N-acetylglutamine, proline betaine, 4-hydroxyhippurate, 4-hydroxyphenylacetate, dimethylsulfone, HPPHA, p-Cresol, phenylacetylglutamine, dimethylamine, TMAO, scyllo-inositol, N-methylnicotinamide, urée, methylguanidine, malonate et moins élevé en 3-methylhistidine, alanine, glycine, lysine, 3-indoxyl sulfate, hippurate, isobutyrate, beta-glucose, lactate, succinate, betaine, isopropanol, xanthosine	Almanza J Nutr Biochem. 2017 (172)
- Régime non végétarien / "occidental" - Régime végétarien	OT Participants de 45 à 79 ans	145	ACP	Sérum	MS ciblée	ACP, corrélations de Pearson	Régime « occidental » associé aux acides aminés branchés et aromatiques et aux acylcarnitines à courte chaîne (C2, C5)	Bhupathiraju J Nutr. 2018 (169)
- Régime à base de végétaux (consommation de fibres totales, $\beta$ -carotène, potassium, glucides totaux, saccharose) - Régime à base de viande (consommation de protéines totales, vitamine B12, protéines de viande, cholestérol, phosphore)	OT Hommes et femmes de plus de 18 ans	89	ACP, Nonnegative matrix factorization	Plasma et urine	GC-MS et LC-MS	Correlations, Wilcoxon test, Sparse generalized canonical correlation analysis	Régime à base de viande : augmentation de 2-hydroxybutyrate, pyroglutamate, acide deoxycholique tauro- et glyco-conjugués, acide lithocholique sulfaté, acide 23-nordeoxycholique, acide L-Aspartique, acide caproïde; diminution de l'acide chenodeoxycholique et acide hyocholique. Régime à base de végétaux : augmentation de phenylacetate; diminution de $\gamma$ -aminobutyrate, 3-methyl-2-oxo-valerate et acide arachidique.	Wei Mol Nutr Food Res. 2018 (177)
- Régime faible en graisse - Régime ayant un index glycémique bas - Régime très pauvre en glucides	ICM Participants ayant un IMC élevé de 18 à 40 ans	21	Adhérence au régime	Plasma	LC-MS	ANOVA, t-test	152 métabolites avaient une concentration qui différait dans au moins un des régimes, dont des diacylglycérols et triacylglycérols, des acides aminés branchés et des marqueurs du statut métabolique	Esko AJCN 2017 (163)

ICM, Intervention Court ou Moyen terme ; IL, Intervention Long terme ; OT, étude d'Observation Transversale ; ACP, Analyse en Composante Principale ; O-PLS-DA, Orthogonal Partial Least-Squares Discriminant Analysis ; ANCOVA, ANalyse de la COVariance ; ANOVA, ANalyse de la VAriance

Source : Guasch-Ferré, Clin Chem. 2018 (126); Scalbert, Am J Clin Nutr. 2014 (51), et revue additionnelle de la littérature.

## B. Métabolomique et épidémiologie des cancers

L'utilisation de la métabolomique dans le domaine de l'épidémiologie des cancers est vaste. Elle peut notamment contribuer à l'élucidation des voies mécanistiques liées à des expositions spécifiques, permettre l'identification de marqueurs circulants qui indiquent un risque de développement ou progression du cancer ou encore permettre de stratifier les cas de cancer par sous-type moléculaire (179). Comme suggéré plus haut, une application des plus pertinentes est l'identification d'un état métabolique qui annoncerait l'évolution vers le développement d'un cancer mais à un stade encore réversible et pour lequel des interventions pourraient permettre de diminuer ce risque ou au moins arrêter la progression vers la maladie. De tels progrès réduiraient l'écart actuel entre les approches de dépistage du cancer plutôt coûteuses et potentiellement nocives (détection précoce des lésions néoplasiques ou pré néoplasiques pouvant nécessiter une intervention médicale ou chirurgicale) et les approches de dépistage plus efficaces des maladies cardiovasculaires (détection et traitement médical des conditions prédisposantes comme l'hypertension et l'hypercholestérolémie) (179). Ces approches préventives couplées à la métabolomique peuvent permettre d'identifier des biomarqueurs d'origine exogène ayant un lien causal avec la cancérogénèse, des biomarqueurs reflétant l'impact de l'exposome ou un statut métabolique perturbé présageant un futur cancer ou encore des biomarqueurs purement prédictifs du risque de cancer (sans nécessairement de lien causal).

Une revue de la littérature publiée en 2013 a montré qu'à cette date l'application principale de la métabolomique dans le domaine de la recherche sur le cancer consistait en l'identification de biomarqueurs de diagnostic en comparant le profil métabolique de cas et de témoins (73% des 160 études incluses) dans le cadre de la prévention secondaire (outil de diagnostic) (180). La plupart de ces études étaient basées sur des tailles d'échantillons similaires à des études pilotes. Dans une moindre mesure, la métabolomique a été utilisée pour améliorer la classification des tumeurs, identifier de nouveaux biomarqueurs pronostic, découvrir de nouvelles cibles pour des interventions thérapeutiques, ou encore améliorer le suivi de la réponse aux traitements anticancéreux (181–187). Notamment, le projet européen FP7 METAcancer a récemment montré que la métabolomique pouvait apporter des informations sur le stade, le grade et le sous-type des tumeurs du sein et fournir des pistes mécanistiques (188). Depuis, quelques études ont été réalisées en prévention primaires de manière prospective sur des prélèvements avant diagnostic du cancer (présentées dans le Tableau 5).

Ces études se basent majoritairement sur des études cas-témoins nichés ; c'est-à-dire que des cas de cancer incidents et des témoins appariés, selon des caractéristiques similaires, sont sélectionnés de manière rétrospective. Les prélèvements sanguins de ces individus sont effectués à l'inclusion, et les cas de cancers diagnostiqués entre ce prélèvement et la fin du suivi sont sélectionnés. Les témoins sont en général sélectionnés selon la méthode « density sampling » (189) c'est-à-dire que le témoin n'a pas encore de cancer diagnostiqué au moment du diagnostic de son cas apparié. Il est également possible d'utiliser un design de type cas-cohorte, ici les cas ne sont pas appariés à des témoins mais sont comparés à une « sous-cohorte » d'individus n'ayant pas de cancer prévalent ou incident. Ce design permet d'augmenter la puissance mais est malheureusement rarement possible en métabolomique au vu de son coût. Ces études ont concerné le cancer du pancréas (190,191), les carcinomes hépatocellulaires et cancers des voies biliaires (192–198), le cancer de poumon et autres cancers dérivés (199,200), les tumeurs du cerveau (201,202), le cancer du sein (203–207), le cancer colorectal (206–212) et le cancer de la prostate (206,213–216).

On constate qu'au global et par localisation les résultats ne sont en général pas reproduits. Cela peut s'expliquer par le fait que la carcinogénèse est un processus multifactoriel et multiprocessus, et qu'au sein d'une même localisation, les types de cancer peuvent être différents (par exemple selon l'agressivité du cancer ; post/pré-ménopausique, ER+/-, PR+/- pour le cancer du sein). La durée de suivi est également très variable selon les études ce qui peut impacter les résultats puisque selon le délai entre le prélèvement et le diagnostic, l'avancement de l'état métabolique pré-néoplasique va différer. Par exemple, dans un délai inférieur à 2 ans, le profil métabolique peut refléter un état induit par une tumeur déjà présente mais encore non diagnostiquée. De plus, malgré le fait que la majorité de ces études utilise la régression logistique conditionnelle pour estimer l'intensité de la relation entre le métabolite et le risque de cancer (une méthode adaptée au design « cas-témoin niché »), certaines études ont une approche plutôt tournée vers la prédiction comme par exemple l'étude de Bro et al. sur le cancer du sein. Cette variété d'approches statistiques influence également la variabilité des résultats que l'on peut observer pour une même localisation. De plus, comme pour l'identification de biomarqueurs nutritionnels, la variété des méthodes analytiques utilisées influence les familles de métabolites détectées. Notamment, pour ces quelques études la plupart a utilisé la métabolomique ciblée, allant de quelques métabolites mesurés à plusieurs centaines. En revanche, très peu de ces études ont analysé des composés encore inconnus dans les bases de données, et se sont limités aux métabolites déjà identifiés. Par ailleurs, plusieurs de ces études se sont intéressées à des expositions particulières pouvant avoir un lien avec les profils métabolomiques, notamment le risque de carcinome hépatocellulaire et le mode de vie, dont l'alimentation (192,194,196), la fumée de cigarette et le risque de cancer du poumon (199) ou colorectal (210), le risque de cancer du sein et l'alimentation (205), l'IMC (204) ou le mode de vie dont l'alimentation (203,207) ou encore le risque de cancer colorectal et la consommation de café (211), d'alcool (212) ou le mode de vie dont l'alimentation (207).

Peu d'études ont donc utilisé la « métabolomique nutritionnelle » pour identifier des biomarqueurs de risque de cancer en prévention primaire malgré l'importance de la nutrition dans la prévention de ces maladies, comme expliqué plus haut. Seulement quelques-unes d'entre elles ont montré des associations entre expositions nutritionnelles (notamment consommation de café, d'alcool, de fibres, de vitamine E et d'aliments frits, IMC, activité physique) et risque de cancer (192,194,196,204,205,207,211) et deux de ces études ont prouvé un effet médiateur malgré le fait que les expositions nutritionnelles ne soient pas les seules prises en compte (192,196). Ce manque d'association pourrait être dû notamment à des biais de déclaration, à un manque de puissance ou encore à un manque de sensibilité dans la détection des métabolites notamment en RMN (217). En revanche, de nombreuses études ont utilisé la métabolomique afin de relier l'exposition nutritionnelle et les biomarqueurs des cancers prévalents (diagnostic) (217).

Concernant les profils métabolomiques pré-diagnostic du risque de cancer du sein, une première étude pilote a été réalisée en 2011 (207) mais n'a pas pu mettre en évidence de métabolites associés, probablement à cause d'un manque de puissance (19 cas, 20 témoins). En 2015, Bro et al. ont utilisé des échantillons de plasma pré-diagnostic pour établir un « Biocontour » consistant en une combinaison d'informations biologiques et phénotypiques grâce à la RMN (203). Ce Biocontour permettait de prédire un risque accru de cancer du sein dans les 2 à 5 ans suivant le prélèvement avec une sensibilité et une spécificité supérieures à 80 % alors que celles de la mammographie sont autour de 75%. Ce modèle a été validé grâce à un jeu de données indépendant. Cependant, très peu de métabolites ont pu être identifiés au sein de cette signature, ne donnant que peu d'information sur la spécificité de ces métabolites sur le cancer du sein et n'apportant pas d'information étiologique. De plus, la durée de suivi étant limitée à 5 ans, cette approche permettra plutôt d'améliorer une détection précoce qu'une application en prévention

primaire. En 2016, une étude ciblée sur une centaine de métabolites plasmatiques a montré que plusieurs lysophosphatidylcholines étaient associées à une diminution du risque de cancer du sein (206). En 2017 et 2018 deux études ont recherché des métabolites associés d'une part à l'alimentation (205) ou à l'IMC (204) et d'autre part au risque de cancer du sein chez des femmes post-ménopausées avec une durée de suivi plus conséquente (médiane de 6,7 ans). La première étude a montré 3 métabolites sériques (caprate (10:0),  $\gamma$ -carboxyethyl hydrochroman, et 4-androsten-3 $\beta$ ,17 $\beta$ -diol-monosulfate) associés à au moins une exposition alimentaire et à un risque global de cancer du sein. Une vingtaine de métabolites ont été associés à un cancer du sein ER+, dont une dizaine de métabolites associés à l'alcool (dont 7 androgènes et  $\alpha$ -hydroxyisovalerate), trois dérivés de la vitamine E, le caprate (10:0) associé au beurre et le 2-hydroxyoctanoate associé aux aliments frits. La deuxième étude a observé deux métabolites liés à l'IMC - le 16 $\alpha$ -hydroxy-DHEA-3-sulfate et la 3-méthylglutarylcarnitine - associés positivement au risque de cancer du sein et qui diminuaient l'effet de l'IMC sur le risque de ce cancer. Ces deux métabolites ainsi que l'allo-isoleucine et la 2-méthylbutyrylcarnitine étaient associés au risque de cancer du sein ER+. Ces études n'ont été réalisées que chez les femmes post-ménopausées ce qui, malgré des résultats très prometteurs, limite la généralisation de ces résultats chez les femmes pré-ménopausées, notamment pour l'IMC qui est un facteur de risque protecteur du risque de cancer du sein pré-ménopausique. De plus, ces deux études n'ont analysé que les métabolites déjà identifiés grâce aux standards de Metabolon ce qui, malgré le nombre déjà conséquent de métabolites identifiés, limite le potentiel de découverte de nouveaux biomarqueurs. A notre connaissance, aucune étude n'avait donc investigué les profils métabolomiques pré-diagnostic de manière complètement non ciblée avec une durée de suivi et un effectif conséquent.

Parallèlement, concernant le cancer de la prostate, deux études ont été réalisées chez des hommes fumeurs avec des durées de suivi conséquentes et ont montré que le 1-stearoylglycerol et le glycérol étaient associés inversement au risque (213) (mais non répliqués dans la deuxième étude (214)), tout comme l'alpha-ketoglutarate (associé lui, dans les deux études). Des associations inverses entre les métabolites énergétiques et lipidiques (notamment l'inositol-1-phosphate et les glycérophospholipides) et le citrate et le risque de cancer de la prostate agressif ont également été montrées ainsi que des associations positives avec la thyroxine et l'oxyde de triméthylamine (214). En 2016 et 2017, deux études ciblées sur une centaine de métabolites plasmatiques ont montré que l'acylcarnitine C18:1, la citrulline (216) et certaines lysophosphatidylcholines (206) étaient associées à une diminution du risque de cancer de la prostate alors que la trans-4-hydroxyproline, plusieurs glycérophospholipides, un sphingolipide (216) et la phosphatidylcholine PC ae C30:0 (206) étaient associés à une augmentation de risque. Enfin, Huang et al. ont montré qu'une trentaine de métabolites sériques étaient associés au risque de cancer de la prostate et notamment la pyroglutamine, la gamma-glutamylphénylalanine, le phénylpyruvate, la N-acétylcitrulline, et la stéaroylcarnitine (associations inverses pour les 4 premiers, et directe pour le dernier) (215). Toutes ces études sur le risque de cancer de la prostate ont été menées de manière ciblée sur une centaine de métabolites ou de manière semi-ciblée sur quelques centaines ; mais aucune n'a analysé les composés non identifiés, ce qui limite là aussi le potentiel de découverte de nouveaux biomarqueurs. De plus, très peu de résultats ont été répliqués.

Tableau 5 Etudes épidémiologiques prospectives ayant investigué le lien entre profils métabolomiques pré-diagnostic et risque de cancer

Référence	Titre article	Localisation	Expositions spécifiques recherchées	Design de l'étude	Temps de suivi	Population/efficatif	Biofluide	Méthode analytique	Méthodes d'analyse statistiques	Résultats principaux
Shu, Int J Cancer. 2018 (190)	Prospective metabolomics study identifies potential novel blood metabolites associated with pancreatic cancer risk.	Pancréas		Cas-témoin niché	11 à 17 ans	Femmes de 40 à 70 ans, hommes de 40 à 74 ans 226 cas / 226 témoins	Plasma	UPLC-QTOFMS et GC-TOFMS	Régression logistique conditionnelle	Six glycérophospholipides associés à une diminution du risque et un à une augmentation. Coumarine et acide picolinique associés à une augmentation de risque et acide tétradécanoïque associé à une diminution. 4 métabolites significatifs après correction de test multiple.
Nakagawa, Cancer Sci. 2018 (191)	Metabolome analysis for pancreatic cancer risk in nested case-control study: Japan Public Health Center-based prospective Study.	Pancréas		Cas-témoin niché	Médiane de 16,4 ans	Femmes et hommes de 40 à 69 ans 170 cas / 340 témoins	Plasma	GC-MS ciblée	Régression logistique conditionnelle	1,5-anhydro-d-glucitol (1,5-AG) associé à une diminution du risque, méthionine associée à une augmentation du risque. En comparant les sujets diagnostiqués dans les 6 premières années : 1,5-AG, asparagine, tyrosine et acide urique associés à une diminution du risque.
Assi, Cancer Epidemiol Biomarkers Prev. 2018 (192)	Are metabolic signatures mediating the relationship between lifestyle factors and hepatocellular carcinoma risk? Results from a nested case-control study in EPIC.	Carcinome hépatocellulaire	Modes de vie (7 variables du Healthy Lifestyle Index)	Cas-témoin niché	17 ans	Femmes et hommes de 25 à 85 ans 147 cas / 147 témoins	Sérum	LC-MS ciblée	Approche Meeting in the Middle; PLS (mode de vie), analyse de médiation, régression logistique conditionnelle	IMC associé positivement à l'acide glutamique, tyrosine, PC aaC38:3, et score du fonctionnement du foie et négativement à lysoPC aC17:0 et aC18:2. Exposition à l'alcool négativement associée aux sphingomyélines (SM C16:1, C18:1, SM(OH) C14:1, C16:1 et C22:2). IMC et alcool associés à une augmentation du risque. Effet médiateur de ces deux signatures.
Zhang Sci Rep 2015 (193)	A weighted relative difference accumulation algorithm for dynamic metabolomics data: long-term elevated bile acids are risk factors for hepatocellular carcinoma.	Carcinome hépatocellulaire		Cas-témoin niché	3,5 ans	Hommes et femmes de 44 à 66 ans 11 cas / 12 témoins	Sérum	LC-MS	Algorithm wRDA (weighted relative difference accumulation)	Une élévation à long terme des acides biliaires sériques identifiée comme facteur de risque du développement HCC.



Assi, Am J Clin Nutr. 2018 (194)	Metabolic signature of healthy lifestyle and its relation with risk of hepatocellular carcinoma in a large European cohort	Carcinome hépatocellulaire	Modes de vie (7 variables du Healthy Lifestyle Index)	Cas-témoin niché	17 ans	Femmes et hommes de 25 à 85 ans 147 cas / 147 témoins	Sérum	UPLC-MS ciblée	PLS (mode de vie), régression logistique conditionnelle	Composante de mode de vie négativement associée avec l'exposition à l'alcool, l'IMC, le statut tabagique, le diabète et positivement associée à l'activité physique. Composante métabolique associée positivement sphingomyelines (OH) C14:1, C16:1, et C22:2, et négativement glutamate, hexoses, et diacyl-phosphatidylcholine PC aaC32:1. Ces 2 composantes étaient inversement associées au risque.
Fages, BMC Medicine. 2015 (195)	Metabolomic profiles of hepatocellular carcinoma in a European prospective cohort.	Carcinome hépatocellulaire		Cas-témoin niché	Médiane de 4,8 ans	Femmes et hommes de 35 à 85 ans 114 cas / 222 témoins	Sérum	H-RMN	Régression logistique conditionnelle	Signature métabolique associée au risque comprenant des perturbations dans l'oxydation des acides gras et du métabolisme des acides aminés, des lipides et des glucides. 16 métabolites endogènes et exogènes associés au risque.
Assi, Mutagenesis. 2015 (196)	A statistical framework to model the meeting-in-the-middle principle using metabolomic data: application to hepatocellular carcinoma in the EPIC study	Carcinome hépatocellulaire	21 variables de mode de vie	Cas-témoin niché	14 ans	Femmes et hommes de 35 à 85 ans 114 cas / 222 témoins	Sérum	H-RMN	Approche Meeting in the Middle ; PLS (mode de vie), analyse de médiation, régression logistique conditionnelle	2 <sup>ème</sup> score : composante métabolique positivement associée à tyrosine et glucose, et composante de mode de vie associée à une augmentation de risque. 3 <sup>ème</sup> score : composante de mode de vie associée positivement à l'exposition à l'alcool, hépatite et statut tabagique et négativement à la consommation de légumes et composante métabolique associée positivement à l'éthanol, glutamate et phénylalanine. Ces composantes étaient associées positivement au risque. Effets médiateurs des 2 composantes métaboliques
Stepien Int J Cancer. 2016 (197)	Alteration of amino acid and biogenic amine metabolism in hepatobiliary cancers: Findings from a prospective cohort study.	Carcinome hépatocellulaire (HCC), cancers intrahépatiques des voies biliaires (IHBC) et de la vésicule biliaire et extra hépatique du tractus biliaire (GBTC)		Cas-témoin niché	Médiane de 9,6 ans	Femmes et hommes de 35 à 70 ans Cas HCC=147, IHBC=43, GBTC=134 / témoins=327	Sérum	LC-MS ciblée	Régression logistique conditionnelle, PLS-DA	14 métabolites associés au risque HCC. Leucine, lysine, glutamine et le ratio acides aminés branchés sur aromatiques associés inversement au risque de HCC et phénylalanine, tyrosine et leur ratio, glutamate, ratio glutamate/glutamine, kynurenine et son ratio avec le tryptophane associés positivement au risque de HCC. Effet du statut hépatique et du taux d'enzymes hépatiques. Pas d'associations pour les autres cancers

Mayers, Nat Med. 2014 (198)	Elevated circulating branched chain amino acids are an early event in pancreatic adenocarcinoma development	Adénocarcinome hépatique		Cas-témoin niché	Médiane de 8,7 ans	Hommes et femmes de 30 à 84 ans 453 cas / 898 témoins	Plasma	LC-MS et GC-MS ciblée	Régression logistique conditionnelle	Acides aminés branchés associés à une augmentation du risque avec la plus forte association chez les sujets avec des échantillons collectés entre 2 et 5 ans.
Yuan, Cancer Res. 2011 (199)	Urinary levels of cigarette smoke constituent metabolites are prospectively associated with lung cancer development in smokers	Poumon	Constituants de la cigarette	Cas-témoin niché	Jusqu'à 20 ans	Hommes fumeurs de 45 et 64 ans 476 cas / 476 témoins	Urine	GC-MS ciblée, quantitative	Régression logistique conditionnelle	R-1,t-2,3,c-4-tetrahydroxy-1,2,3,4-tetrahydrophenanthrene, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol et ses glucuronides, cotinine et ses glucuronides associés à une augmentation de risque.
Wikoff, J Clin Oncol. 2015 (200)	Diacetylspermine Is a Novel Prediagnostic Serum Biomarker for Non-Small-Cell Lung Cancer and Has Additive Performance With Pro-Surfactant Protein B.	Cancer bronchique non à petites cellules		Cas-témoin niché	1 an	Hommes et femmes 100 cas / 199 témoins	Sérum	MS ciblée	AUC, Mann-Whitney U test or Kruskal-Wallis test, régression logistique	N(1),N(12)-diacetylspermine (DAS) augmenté de 1,9 fois, démontrant une sensibilité et spécificité significative dans le set de découverte où les échantillons étaient collectés jusqu'à 6 mois avant le diagnostic. DAS complète la performance de pro-surfactant protéine B dans les sets de validation et de découverte (AUC= 0.808)
Huang, Oncotarget 2017 (201)	A prospective study of serum metabolites and glioma risk.	Gliome		Cas-témoin niché	Médiane de 9 ans	Hommes fumeurs de 50 à 69 ans 64 cas / 64 témoins	Sérum	LC/MS-MS	Régression logistique conditionnelle	43 métabolites associés au risque. 2-Oxoarginine, cystéine, alpha-ketoglutarate, chenodeoxycholate et arginate étaient les plus inversement associés. 7 métabolites de la xanthine liés au métabolisme de la caféine associés à une augmentation de risque. Résultats similaires pour les cancers de haut-grade, métabolites les plus inversement associés : acides biliaires secondaires glycocholate sulfate et acide 3β-hydroxy-5-cholanoïque, xénobiotique méthyl 4-hydroxybenzoate sulfate, hormone sexuelle stéroïdienne 5alpha-pregnan-3beta, 20beta-diol-monosulfate, et cofacteur/vitamine oxalate
Björkblom, Oncotarget 2016 (202)	Metabolomic screening of pre-diagnostic serum samples identifies association between α- and γ-tocopherols and glioblastoma risk.	Glioblastome		Cas-témoin niché	Moyenne de 12,6 ans	Hommes et femmes 110 cas / 110 témoins	Sérum	GC-MS	OPLS-EP, régression logistique conditionnelle	9 métabolites (γ-tocopherol, α-tocopherol, erythritol, acide erythronique, myo-inositol, cystine, acide 2-keto-L-gluconique, hypoxanthine et xanthine) associés au risque et impliqués dans le métabolisme des antioxydants. α-tocopherol et γ-tocopherol associés à une augmentation de risque.

Moore, J Nat'l Cancer Inst. 2018 (204)	A Metabolomics Analysis of Body Mass Index and Postmenopausal Breast Cancer Risk.	Sein	IMC	Cas-témoin niché	Médiane de 6,7 ans	Femmes post- ménopausées de 55 à 74 ans 621 cas / 621 témoins	Sérum	MS Metabolon	Corrélations de Pearson (IMC) et régression logistique conditionnelle	Sur les 67 métabolites associés à l'IMC, 16a-hydroxy-DHEA-3-sulfate et 3- methylglutaryl carnitine associés positivement au risque. 4 associés positivement aux cancers du sein ER+ : 16a-hydroxy-DHEA-3-sulfate, 3- methylglutaryl carnitine, allo-isoleucine et 2-methylbutyryl carnitine. 16a-hydroxy- DHEA-3-sulfate et 3- methylglutaryl carnitine diminuaient l'association entre IMC et risque de cancer du sein.
Playdon, Am J Clin Nutr. 2017 (205)	Nutritional metabolomics and breast cancer risk in a prospective study.	Sein	Alimentation	Cas-témoin niché	Médiane de 6,7 ans	Femmes post- ménopausées de 55 à 74 ans 621 cas / 621 témoins	Sérum	MS Metabolon	Corrélations de Pearson (alimentation) et régression logistique conditionnelle	Sur les 113 métabolites liés à l'alimentation, 3 étaient associés au risque : caprate (10:0), $\gamma$ -carboxyethyl hydrochroman ( $\gamma$ -CEHC), et 4-androsten- 3 $\beta$ ,17 $\beta$ -diol-monosulfate. 19 métabolites associés au risque de cancer du sein ER+ : 12 métabolites liés à l'alcool dont 7 androgènes et $\alpha$ -hydroxyisovalérate, 3 dérivés de la vitamine E; le caprate lié au beurre; et le 2-hydroxyoctanoate aux produits frits.
Bro, Metabolom ics 2015 (203)	Forecasting individual breast cancer risk using plasma metabolomics and biocontours	Sein	Mode de vie et phénotype	Cas-témoin niché	5 ans	419 cas / 419 témoins	Plasma	H-RMN	PLS-DA avec sélection de variable "forward"	Création d'un biocontour (combinaison d'informations biologiques et phénotypiques) permettant de prédire le risque plusieurs années avant le diagnostic (spécificité bien au-delà de 80 %). Le modèle a été construit à partir de données obtenues en 1993-1996 et testé sur des personnes échantillonnées un an plus tard, en 1997.
Shu, Int J Cancer. 2018 (208)	Prospective study of blood metabolites associated with colorectal cancer risk.	Colorectal		Cas-témoin niché	11 à 17 ans	Femmes de 40 à 70 ans, hommes de 40 à 74 ans 250 cas / 250 témoins	Plasma	UPLC- QTOFMS et GC-TOFMS	Régression logistique conditionnelle	35 métabolites associés au risque après correction de test multiple. 12 glycérophospholipides dont 9 associés à une diminution de risque et 3 à une augmentation. 9 autres lipides, 7 composés aromatiques, 5 acides organiques et 4 autres composés organiques.

Cross, Cancer 2014 (209)	A prospective study of serum metabolites and colorectal cancer risk.	Colorectal		Cas-témoin niché	Moyenne de 8 ans	Hommes et femmes de 55 à 74 ans 254 cas / 254 témoins	Sérum	LC-MS et GC-MS	Régression logistique conditionnelle	Pas d'associations entre métabolites du sérum et risque après correction de test multiple, quelques associations suggérées (leucyl-leucine inversement associée). Glycochenodeoxycholate positivement associé au risque chez les femmes.
Cross, Carcinogenesis 2014 (210)	Metabolites of tobacco smoking and colorectal cancer risk	Colorectal	Tabac	Cas-témoin niché	Moyenne de 7,8 ans	Hommes et femmes de 55 à 74 ans 255 cas / 254 témoins	Sérum	LC-MS et GC-MS	Régression linéaire (tabac) et régression logistique conditionnelle	Statut fumeur associé à cotinine, O-cresol sulfate et hydroxycotinine. Statut fumeur de tabac associé à une augmentation du risque. Individus ayant des taux détectables d'hydroxycotinine avaient une augmentation de risque. Augmentation de risque suggéré pour les taux détectable de cotinine.
Guertin, Am J Clin Nutr 2015 (211)	Serum biomarkers of habitual coffee consumption may provide insight into the mechanism underlying the association between coffee consumption and colorectal cancer	Colorectal	Café	Cas-témoin niché	18 ans	Hommes et femmes de 55 à 74 ans 251 cas / 247 témoins	Sérum	MS Metabolon	Régression linéaire et corrélations de Pearson (café) et régression logistique conditionnelle	29 métabolites associés positivement à la consommation de café après correction de test multiple (parmi les plus associés : trigonelline (N'-methylnicotinate) et quinate). Parmi ces métabolites, 8 directement liés au métabolisme du café, et 3 de ceux-là associés négativement au risque (théophylline, caféine, paraxanthine)
Troche, Plos One 2016 (212)	Alcohol Consumption-Related Metabolites in Relation to Colorectal Cancer and Adenoma: Two Case-Control Studies Using Serum Biomarkers.	Colorectal	Consommation d'alcool	Cas-témoin niché	Jusqu'à 16 ans	Hommes et femmes de 55 à 74 ans 252 cas / 250 témoins	Sérum	MS Metabolon	Régression logistique	Pas d'association entre métabolites liés à l'alcool et le risque de cancer colorectal
Chadeau-Hyam, Biomarkers 2011 (207)	Meeting-in-the-middle using metabolic profiling - a strategy for the identification of intermediate biomarkers in cohort studies.	Sein et colon	Anthropométrie, mode de vie, alimentation	Cas-témoin niché	Moyenne de 7 ans	Hommes et femmes Colon : 24 cas / 23 témoins; sein : 19 cas / 20 témoins	Plasma	H-RMN	Approche meeting-in-the-middle ; ANOVA et corrélations de Spearman (exposition) et O2-PLS-Bootstrap (cancers)	Association la plus forte liée négativement au risque de cancer du côlon et liée à la consommation de fibres. Thiamine associée négativement à trois régions spectrales et au cancer du colon

Mondul, Metabolomics. 2014 (213)	1-stearoylglycerol is associated with risk of prostate cancer: results from serum metabolomic profiling.	Prostate		Cas-témoin niché	23 ans	Hommes fumeurs de 50 à 69 ans 74 cas / 74 témoins	Sérum	MS Metabolon	Régression logistique	1-stearoylglycerol associé inversement au risque après correction de test multiple. L'intensité de la relation ne dépendait pas de l'agressivité du cancer et était observée chez les cas diagnostiqués jusqu'à 23 ans après le prélèvement. Associations similaires mais moindres pour le glycérol et alpha-ketoglutarate.
Schmidt, BMC Med. 2017 (216)	Pre-diagnostic metabolite concentrations and prostate cancer risk in 1077 cases and 1077 matched controls in the European Prospective Investigation into Cancer and Nutrition	Prostate		Cas-témoin niché	Jusqu'à 16 ans	Hommes de 40 à 77 ans 1077 cas / 1077 témoins	Plasma	MS ciblée	Régression logistique conditionnelle	Acylcarnitine C18:1, citrulline, trans-4-hydroxyproline, glycérophospholipides PC aa C28:1, PC ae C30:0 et PC ae C30:2, et sphingolipide SM (OH) C14:1 associés au risque, mais aucun après correction de test multiple. Citrulline associée à une diminution de risque dans les 5 premières années de suivi après correction de test multiple. 12 glycérophospholipides associés à une diminution de risque d'un cancer avancé après correction de test multiple. Mortalité due au cancer de la prostate associée positivement à acylcarnitine C3, methionine, trans-4-hydroxyproline, amine biogène ADMA, hexose et sphingolipide SM (OH) C14:1 et négativement au glycérophospholipide PC aa C42:4.
Mondul, Int J Cancer. 2015 (214)	Metabolomic analysis of prostate cancer risk in a prospective cohort: The alpha-tocopherol, beta-carotene cancer prevention (ATBC) study.	Prostate		Cas-témoin niché	Jusqu'à 20 ans	Hommes fumeurs de 50 à 69 ans 200 cas / 200 témoins	Sérum	MS Metabolon	Régression logistique conditionnelle	Associations inverses entre les métabolites énergétiques et lipidiques et le risque de cancer de prostate agressif. Inositol-1-phosphate plus forte association et glycérophospholipides et acides gras étaient fortement représentés (oleoyl-linoleoyl-glycérophosphoinositol, 1-stéaroylglycérophosphoglycérol, stéarate et docosadienoate. Alpha-ketoglutarate, citrate associés à une diminution du risque agressif et thyroxine et oxyde de triméthylamine à une augmentation. L'ajustement sur le taux de PSA sérique n'altérerait pas les résultats.

Huang Br J Cancer. 2016 (215)	Serum metabolomic profiling of prostate cancer risk in the prostate, lung, colorectal, and ovarian cancer screening trial.	Prostate		Cas-témoin niché	17 ans	Hommes de 55 à 74 ans 380 cas / 380 témoins	Sérum	MS Metabolon	ACP et Régression logistique conditionnelle	27 métabolites associés au risque. Pyroglutamine, gamma-glutamylphenylalanine, phenylpyruvate, N-acetylcitrulline, et stearylcarntine les plus associés (négativement pour les 4 premiers, positivement pour le dernier). Résultats similaires pour les cancers agressifs.
Kühn, BMC Med 2016 (206)	Higher plasma levels of lysophosphatidylcholine 18:0 are related to a lower risk of common cancers in a prospective metabolomics study.	Sein, prostate et colorectal		Cas-cohorte	Médiane de 8,3 ans pour les non-cas et 6,5 ans pour les cas incidents	Hommes et femmes de 35 à 65 ans Cas sein=362, prostate=310, colorectal=163 / 774 témoins (sous-cohorte)	Plasma	MS ciblée	Régression de Cox	Lysophosphatidylcholines (notamment lysoPC a C18:0) associées à une diminution du risque de cancers du sein, de la prostate et colorectal. Phosphatidylcholine PC ae C30:0 associée à augmentation du risque. Pas d'hétérogénéité dans les associations observées par le temps de latence entre le prélèvement et le diagnostic.

### C. Analyses statistiques des données issues de la métabolomique

La métabolomique est une méthode d'analyse à haut débit qui génère un volume important de données et s'inscrit donc dans le domaine des données de grande dimension (communément appelé « big data »). Il est possible d'analyser ces données de plusieurs manières. Les trois grandes familles d'analyses les plus utilisées sont : les approches univariées et les procédures de correction de test multiple associées, les techniques de réduction de dimension et enfin les approches de sélection de variables. L'analyse statistique de données métabolomiques non ciblées est souvent une première étape du processus de découverte de biomarqueurs qu'il ne faut pas confondre avec la vérification d'hypothèses, car il n'existe aucune hypothèse *a priori*.

L'analyse univariée est supervisée, c'est-à-dire qu'une hypothèse concernant une différence entre les groupes est implicite. Dans ce type d'approches, les métabolites sont analysés un à un avec des ajustements potentiels pour limiter les effets de confusions. Les modèles de régressions linéaires ou logistiques font parties de ces méthodes. Quel que soit le modèle statistique, l'utilisation d'approches univariées dans ce contexte génère des problèmes de multiplicité des tests non négligeables. Pour caractériser le nombre d'associations qui sont faussement déclarées comme statistiquement significatives dans tous les tests effectués, deux grands types de correction existent : le « Family Wise Error Rate » (FWER) (la plus connue étant la correction de Bonferroni (218)) et le « False Discovery Rate » (FDR). En science exploratoire, les corrections de type FDR sont souvent préférées car moins strictes et permettent de limiter le taux de faux négatifs (risque de ne pas détecter des associations qui existent réellement). Cependant, il est important d'établir un seuil raisonnable pour les taux de faux positifs (FDR) (219,220). Dans un contexte exploratoire, il n'y a pas de règle fixe pour le niveau d'acceptabilité du FDR, et tout niveau de 5 % à 50 % peut être utile, selon les étapes ultérieures pour confirmer ou non l'association (51). Dans tous les cas, les marqueurs identifiés doivent être validés lors d'études indépendantes ultérieures.

Par nature, les méthodes univariées ne tiennent pas compte des combinaisons possibles de prédicteurs qui peuvent agir conjointement comme médiateurs de leur effet sur les réponses (221). Cette limitation a donné lieu à l'élaboration d'approches multivariées pour tenir compte de la corrélation au sein des variables prédictives et pour identifier des sous-ensembles explicatifs de variables. Les analyses multivariées sont les plus utilisées dans le domaine de profilage métabolomique en exploratoire (222). Notamment, les techniques de réduction de dimensions permettent de résumer l'information contenue dans de grands ensembles de données en quelques variables synthétiques (les composantes principales) qui saisissent la structure latente des données. Ces méthodes de réduction de dimensions recherchent des combinaisons de variables qui optimisent une certaine mesure de la diversité entre les observations, comme l'Analyse en composantes principales (ACP) (223,224). La plupart des méthodes impose une contrainte d'orthogonalité aux composantes principales pour assurer l'absence de redondance. Les méthodes de partitionnement ou « clustering » sont également largement utilisées pour subdiviser et ordonner un ensemble de données en groupes de données présentant un degré élevé de similarité et des caractéristiques voisines, fournissant des informations biologiques utiles (51,225). Toutefois, ces méthodes peuvent sélectionner des caractéristiques n'étant pas pertinentes pour un paramètre d'intérêt. Les méthodes supervisées comme la PLS (Partial Least Square analysis) permettent de surmonter cette limitation en définissant des structures latentes qui maximisent la covariance de ce paramètre (221). Elles ont cependant tendance à « sur-ajuster » les données ; une validation à l'aide de techniques telles que les tests de permutation et la validation croisée est donc nécessaire (51), ainsi, dans l'idéal, qu'une validation sur population indépendante. Il existe un grand nombre de méthodes supervisées (222,226), l'analyse la

plus couramment utilisée étant l'analyse discriminante par moindres carrés partiels (PLS-DA) afin de comparer deux groupes (227,228) comme par exemple des cas de cancer et des témoins. Dans ces méthodes multivariées, aucune garantie n'existe quant à l'interprétabilité des variables latentes, notamment leur signification biologique peut être impossible à caractériser (221).

Pour surmonter ce problème, plusieurs approches ont été proposées pour limiter le nombre de prédicteurs entrant dans la construction de chaque variable latente. Dans un contexte de régression, l'utilisation de la pénalisation tend à réduire la valeur des coefficients de régression vers zéro, ce qui s'avère utile pour élaborer des modèles prédictifs. Les deux principales méthodes de pénalisation utilisées jusqu'à maintenant sont le « Lasso » (Least Absolute Shrinkage and Selection Operator) (229) et la régression Ridge (230). De nombreuses variantes de ces méthodes ont été proposées, notamment Elastic Net (231) qui est une combinaison des 2 méthodes. Ces trois méthodes seront détaillées plus loin. La régression pénalisée a notamment été appliquée avec succès aux données omiques (221).

Dans le contexte de l'exposome et donc du métabolome, on s'attend à ce que des combinaisons complexes de marqueurs puissent être liées à des mélanges complexes d'expositions, influençant à son tour le risque de certaines maladies. Par conséquent, il semble que les approches univariées puissent convenir dans une phase exploratoire, tandis qu'une étude approfondie des processus complexes impliquera des approches multivariées telles que les techniques de réduction de dimensions ou l'utilisation de méthodes de sélection de variables définissant des ensembles de prédicteurs prédisant conjointement un paramètre, comme par exemple le risque de cancer (221).



## Objectifs du travail de thèse

---

### A. Financements

Cette thèse s'inscrit dans le cadre de deux projets nationaux financés par l'INCa et l'IFRB de l'université Paris 13, chez les femmes, et la Fondation de France et l'IFRB de l'université Paris 13, chez les hommes. Des données de métabolomique non ciblée ont été générées dans le cadre de ces projets et notamment en LC-MS (spectrométrie de masse couplée à la chromatographie liquide) chez les femmes et en H-RMN (résonance magnétique nucléaire du proton) chez les femmes et les hommes.

### B. Objectifs

Les objectifs des projets et donc de cette thèse se divisent en trois objectifs spécifiques :

**Objectif 1 : Identifier des signatures métabolomiques plasmatiques associées au risque à long terme de cancers du sein (objectif 1.I.) et de la prostate (objectif 1.II.)**

**Objectif 2 : Identifier des signatures métabolomiques plasmatiques associées à l'alimentation dans un échantillon de femmes (objectif 2.I.) et d'hommes (objectif 2.II.) de la cohorte SU.VI.MAX**

**Objectif 3 : Identifier des signatures métabolomiques plasmatiques liées à la fois à l'alimentation et au risque à long terme de cancers du sein (objectif 3.I.) et de la prostate (objectif 3.II.)**

Les études menées dans le cadre de ces projets sont des études exploratoires et visent à identifier de potentiels biomarqueurs par la métabolomique non ciblée. Elles se placent donc dans les premières étapes du processus de validation de biomarqueurs. De nombreuses études, de diverses natures, seront nécessaires pour valider ou invalider ces résultats. De potentiels résultats redondants avec la littérature viendront apporter des preuves supplémentaires. Les choix méthodologiques adoptés dans ces projets se placent donc dans ce contexte d'analyses exploratoires et les résultats ne seront pas directement applicables en clinique ou en santé publique.

### C. Consortium multidisciplinaire

Ces projets sont multidisciplinaires et bénéficient des expertises de plusieurs équipes françaises dans divers domaines. L'Equipe de Recherche en Epidémiologie Nutritionnelle a coordonné les projets et était en charge de la sélection des populations d'étude, des analyses statistiques, de l'interprétation épidémiologique et de la valorisation des données et des résultats. La Plate-Forme d'Exploration du Métabolisme (PFEM) de l'Unité du Nutrition Humaine de l'INRA de Clermont-Ferrand (UMR 1019) a généré et prétraité les données de LC-MS. L'équipe Nanomédecine Biomarqueurs Détection (NBD) de l'unité de « Chimie Structures et Propriétés de Biomatériaux et d'Agents Thérapeutiques » (CSPBAT, UMR 7244) a généré et prétraité les données de H-RMN. Les projets ont bénéficié de l'expertise du centre Bordeaux Population Health (Inserm, UMR 1219) et du groupe « Nutritional Methodology and Biostatistics » du Centre International de Recherche sur le Cancer à Lyon en biostatistiques et analyse de données de grande dimension. L'équipe Nutrivasc de l'Unité du Nutrition Humaine de l'INRA de Clermont-Ferrand (UMR 1019) a également été impliquée pour son expertise dans le métabolome alimentaire et l'identification de biomarqueurs d'expositions nutritionnelles. Enfin, l'équipe ECREIN (Micro-Environnement Cellulaire, Immunomodulation et Nutrition) de l'INRA de Clermont-Ferrand (UMR 1019) a apporté son expertise concernant les interprétations biologiques et mécanistiques des résultats observés.

### I. Présentation et objectif général de l'étude SU.VI.MAX

L'étude SU.VI.MAX (Supplémentation en vitamines et minéraux antioxydants) (clinicaltrials.gov ; NCT00272428) est un essai d'intervention contrôlé randomisé en double aveugle qui avait pour but d'évaluer l'effet d'une supplémentation quotidienne en antioxydants à dose nutritionnelle sur l'incidence des maladies cardiovasculaires et des cancers et sur la mortalité (232,233). 13 017 participants (5141 hommes et 7876 femmes) âgés entre 35 et 60 ans parmi la population générale française ont été inclus entre 1994 et 1995. La supplémentation était constituée d'une combinaison de 120 mg de vitamine C, 30 mg de vitamine E, 6 mg de bêta-carotène, 100 µg de sélénium (sous forme de levures enrichies en sélénium) et 20 mg de zinc (sous forme de gluconate) ou d'un placebo dans une capsule quotidienne. Cette intervention a duré jusqu'en 2002 (7,5 ans) et le suivi des participants a duré jusqu'en 2007 afin d'étudier les relations entre les comportements alimentaires globaux (typologies alimentaires) et/ou certains facteurs nutritionnels spécifiques sur une large population d'adultes d'âge mûr, et la qualité globale du vieillissement de cette population, évaluée 10 à 13 années plus tard.

A l'issue de l'essai d'intervention (SU.VI.MAX), aucune différence significative n'a été observée entre le groupe supplémenté en antioxydants et le groupe placebo chez les femmes pour l'ensemble des effets de santé étudiés. Une diminution du risque de cancer et de la mortalité a été mise en évidence chez les hommes supplémentés. Par ailleurs, la cohorte SU.VI.MAX a permis de réaliser de nombreuses études d'observation descriptives ou étiologiques sur les relations entre la nutrition et la santé.

Tous les participants ont fourni un consentement écrit. Les études ont été réalisées conformément aux recommandations de la Déclaration d'Helsinki et ont été approuvées par le Comité d'éthique pour les études avec des sujets humains de l'hôpital Paris-Cochin (CCPPRB 706/2364) et la Commission Nationale de l'Informatique et des Libertés (CNIL 334641/907094).

### II. Collecte des données

#### A. Caractéristiques sociodémographiques et de mode de vie

Les données sur les caractéristiques sociodémographiques, l'usage du tabac, la consommation de médicaments, l'activité physique, l'état de santé et les antécédents familiaux de cancer ont été recueillies au moment de l'inclusion dans l'étude, à l'aide de questionnaires autoadministrés.

Des questionnaires de suivi mensuels ont été envoyés tout au long de l'intervention pour évaluer le respect des protocoles de l'essai et suivre l'état de santé des participants. L'âge de la ménopause a été collecté grâce à ces questionnaires.

Deux types d'information sur l'activité physique ont été recueillis : une question dans le questionnaire d'inclusion demandait si les participants avaient une activité physique régulière et, si oui, si cette activité correspondait à plus d'une heure de marche par jour. De plus, en 1998, la version française validée du questionnaire sur l'activité modifiable a été administrée aux participants de l'étude SU.VI.MAX (234–236).

## B. Données alimentaires

Au cours de l'intervention, tous les participants ont été invités à remplir tous les deux mois un enregistrement alimentaire de 24 heures *via* le Minitel. Ces questionnaires couvraient tous les jours de la semaine et toutes les saisons de l'année de manière aléatoire permettant de prendre en compte la variabilité intra-individuelle des consommations alimentaires. Pour indiquer la taille des portions consommées au cours de la journée, les participants ont bénéficié de l'aide d'un manuel d'instructions validé avec des photographies de plus de 250 produits alimentaires génériques correspondant à 990 aliments spécifiques (237). Une table de composition des aliments (publiée) a été utilisée pour calculer les apports en nutriments (238). Les habitudes alimentaires des participants au moment du prélèvement ont été estimées en faisant la moyenne des apports de tous les enregistrements alimentaires de 24h recueillis au cours des deux premières années de participation à l'étude SU.VI.MAX. Dans le but d'obtenir des estimations assez précises, les habitudes alimentaires ont été calculées uniquement chez les participants ayant fourni plus de 3 enregistrements alimentaires de 24h. De plus, les enregistrements caractérisés par un apport énergétique global inférieur à 100 kcal par jour ou supérieur à 6000 kcal par jour ont été exclus, ainsi que les hommes déclarant moins de 800 kcal par jour et les femmes déclarant moins de 500 kcal par jour dans au moins un tiers de leurs enregistrements. Cette procédure est généralement appliquée pour les études basées sur les données de l'étude SU.VI.MAX (239,240).

Pour l'alcool et le poisson, des aliments consommés avec une fréquence très variable, ou rarement par de larges groupes de la population, des données ont été collectées à partir de questionnaires de fréquence en plus des informations fournies par les enregistrements alimentaires de 24h. La consommation habituelle d'alcool a également été évaluée à l'aide d'un questionnaire semi-quantitatif validé (241), et la consommation habituelle de fruits de mer par une question spécifique lors d'un questionnaire général (239).

## C. Données cliniques et biologiques

Des mesures anthropométriques (taille au centimètre près et poids au 0.1 kg près) ainsi qu'une prise de sang (après une période de jeûne de 12 heures) ont été effectuées par les infirmières et les médecins de l'étude lors d'un examen clinique à l'inclusion. Tous les prélèvements ont été réalisés entre octobre et mai. 35 ml de sang veineux ont été prélevés dans des tubes Vacutainer d'héparine sodique (Becton Dickinson, Rungis, France) et immédiatement centrifugés. Des aliquotes plasmatiques ont été préparées moins d'une heure après la centrifugation à +4°C, puis conservées dans des paillettes de 500µL, congelées à -20°C pendant moins de deux jours avant d'être transférées dans l'azote liquide.

Tous les ans au cours du suivi, les sujets étaient invités à participer à des visites médicales. Au cours de ces visites, étaient réalisés en alternance soit un prélèvement sanguin (inclusion puis tous les 2 ans : mesure de la cholestérolémie, de la glycémie et des concentrations sanguines en antioxydants inclus dans la capsule d'intervention) soit un examen clinique (1 an après l'inclusion puis tous les 2 ans : mesures anthropométriques, mesure de la tension artérielle, électrocardiogramme, etc.).

## D. Événements de santé

Les participants ont déclaré leurs événements de santé tels que des cancers et des maladies cardiovasculaires qui se sont produits pendant le suivi (de 1994 à 2007) *via* des questionnaires de suivi mensuels ou lors des visites médicales annuelles. Suite à ces déclarations, les informations médicales et les rapports pathologiques étaient collectés auprès des participants, des médecins ou des hôpitaux puis ils

étaient examinés par un comité indépendant de médecins experts. Les cancers ont été classés selon la classification internationale des maladies chroniques (ICD-10) (242).

### III. Sélection des populations d'études

#### A. Etudes cas-témoin nichées (objectifs 1 et 3)

Pour les objectifs 1 et 3, des études observationnelles prospectives de type cas-témoins nichés ont été réalisées. Les individus ont été sélectionnés de manière rétrospective ; parmi les individus n'ayant pas de cancers du sein ou de la prostate diagnostiqués à l'inclusion, ceux ayant déclaré un de ces cancers pendant le suivi (13 ans) ont été sélectionnés comme « cas » et des participants possédant des caractéristiques similaires à chacun des cas mais n'ayant pas déclaré un de ces cancers au moment du diagnostic du cas ont été sélectionnés comme « témoins ». Cette méthode est appelée « density sampling » (189) et permet de limiter les biais de confusion au moment des analyses statistiques en comparant « ce qui est comparable ».

##### 1. Pour le cancer du sein

Parmi les femmes incluses dans l'étude SU.VI.MAX (7876), tous les cas de cancer prévalents, les cas de cancer du sein prévalents *in situ* et les cas de décès liés à l'appareil génital ont été exclus de la population éligible (N=156). De même, les femmes n'ayant pas d'échantillon plasmatique disponible et ayant des données manquantes sur les potentiels facteurs de confusion (comme par exemple, l'âge, l'IMC, le statut tabagique) ont été exclues (N=569). 236 premiers cas de cancer du sein invasif ainsi que 6915 potentiels témoins ont donc constitué la population éligible pour ce cas-témoins niché. Une sélection aléatoire a été réalisée parmi cette population éligible afin « d'apparier » 2 témoins par cas pour les analyses en H-RMN et 1 témoin par cas pour les analyses en LC-MS (plus coûteuses). Chaque paire ou triplet devait posséder des caractéristiques similaires au moment du prélèvement sur des classes d'âge (35-39 ans/40-44 ans/45-49 ans/45-49 ans/50-54 ans/55-59 ans/>60 ans), sur le statut ménopausique (au départ et au diagnostic du cas), sur des classes d'IMC (sous poids : <18.5kg/m<sup>2</sup> , poids normal : ≥18.5 kg/m<sup>2</sup> et <25 kg/m<sup>2</sup> et surpoids/obèses : ≥25 kg/m<sup>2</sup>), sur le groupe d'intervention SU.VI.MAX (placebo ou supplément), sur le statut tabagique (fumeuses ou non/anciennes fumeuses) et sur la saison de prélèvement sanguin (octobre/novembre, décembre/janvier/février ou mars/avril/mai). Cette étape a permis d'apparier 228 cas de cancer du sein incidents à 456 témoins pour la H-RMN et 215 cas à 215 témoins pour la LC-MS (les échantillons de tous les individus analysés en LC-MS étaient également analysés en H-RMN). Le flow chart de cette sélection est présenté en Figure 10.

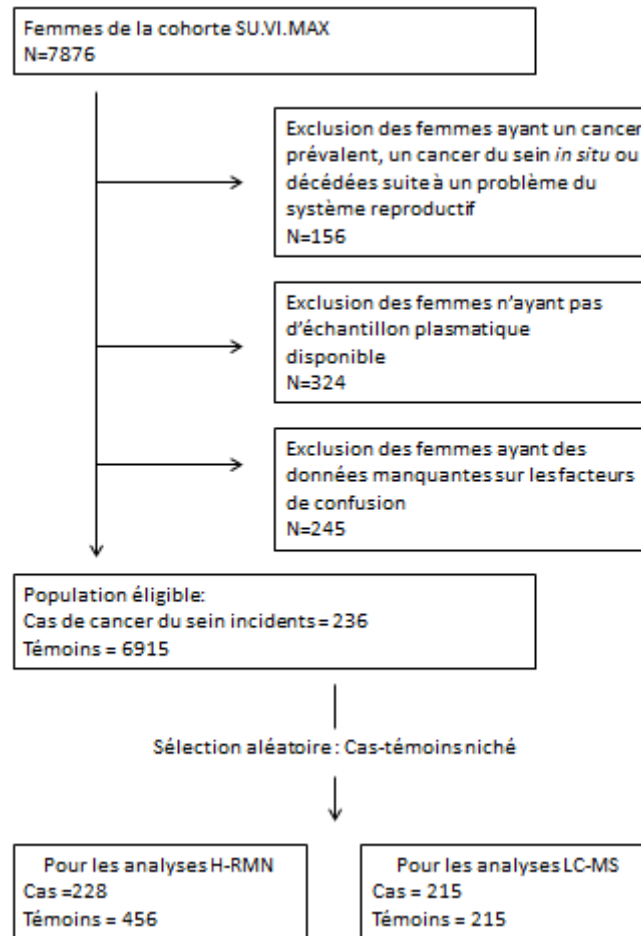


Figure 10 Flow chart de l'étude cas-témoin nichée pour le cancer du sein

## 2. Pour le cancer de la prostate

Parmi les hommes inclus dans l'étude (5141), tous les cas de cancer prévalents ont été exclus de la population éligible (N=41). Les hommes n'ayant pas d'échantillon plasmatique disponible et ayant des données manquantes sur les potentiels facteurs de confusion ont également été exclus (N=360). 181 premiers cas de cancer de la prostate ainsi que 4559 potentiels témoins ont donc constitué la population éligible pour ce cas-témoins niché. Comme précédemment, une sélection aléatoire a été réalisée parmi cette population éligible afin d'apparier 1 témoin par cas (analyses en H-RMN seulement), chaque paire possédant des caractéristiques similaires au moment du prélèvement sur des classes d'âge (<45 ans/45-49 ans/50-54 ans/55-59 ans/≥60 ans), sur des classes d'IMC (sous poids : <18.5kg/m<sup>2</sup>, poids normal : ≥18.5 kg/m<sup>2</sup> et <25 kg/m<sup>2</sup> et surpoids/obèses : ≥25 kg/m<sup>2</sup>), sur le groupe d'intervention SU.VI.MAX (placebo ou supplément), sur le statut tabagique (fumeurs actuels ou non/anciens fumeurs) et sur la saison de prélèvement sanguin (octobre/novembre, décembre/janvier/février ou mars/avril/mai). Cette étape a permis d'apparier 171 cas de cancer de la prostate incidents à 171 témoins. Le flow chart de cette sélection est présenté en Figure 11.

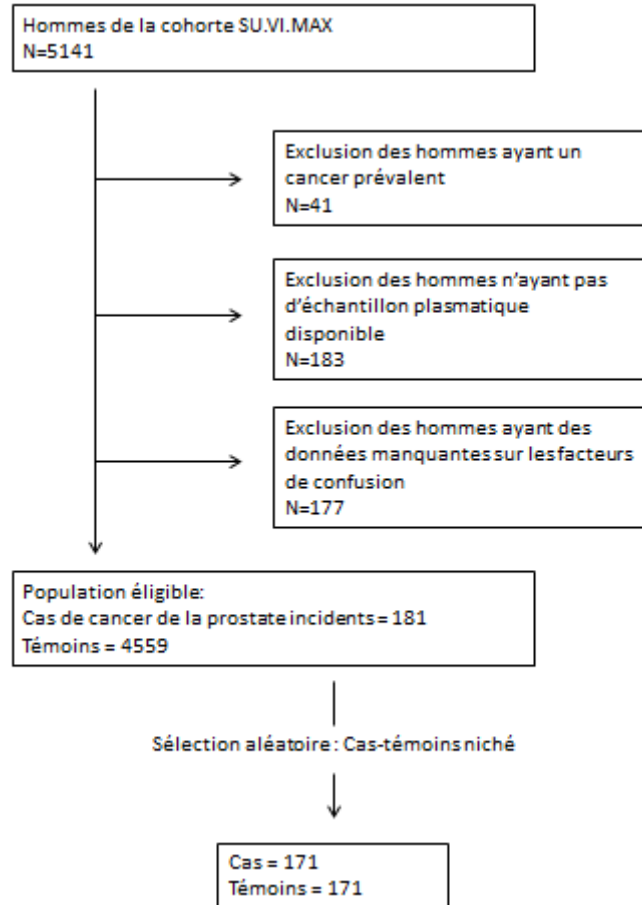


Figure 11 Flow chart de l'étude cas-témoin nichée pour le cancer de la prostate

## B. Etudes transversales (objectif 2)

Pour l'objectif 2, des études d'observation transversales ont été réalisées selon les habitudes alimentaires de la population au cours des deux premières années de participation à l'étude SU.VI.MAX. Ces deux premières années étaient considérées comme représentatives de leurs habitudes alimentaires au moment du prélèvement (inclusion). Pour ces études, nous nous sommes intéressés à l'adéquation des participants au score PNNS-GS (Programme National Nutrition Santé-Guideline Score) (243) reflétant leur adéquation aux recommandations alimentaires nationales françaises de 2001. Ce score a été validé (243) et associé à un moindre risque de maladies chroniques (239,240,244,245). Le PNNS-GS, allant de 0 à 15 points (15 étant la plus haute adéquation et 0 la plus basse), comprend 13 composantes : consommation de fruits et légumes ; pain, céréales, pommes de terre et légumineuses ; aliments à grains entiers ; lait et produits laitiers ; viande et volaille, fruits de mer et œufs ; fruits de mer ; matière grasse ajoutée végétale; matière grasse ajoutée totale; aliments sucrés ; boissons sucrées ; sel ; alcool ; et une composante représentant l'adéquation aux recommandations sur l'activité physique (le détail des composantes alimentaires est fourni dans le Tableau 6). Une pénalité de surconsommation était appliquée lorsque les apports énergétiques étaient supérieurs de 5 % à l'estimation de la dépense énergétique calculée sur le métabolisme basal et le niveau d'activité physique. Des scores plus élevés indiquent une meilleure adéquation aux recommandations nutritionnelles nationales françaises et une meilleure qualité de l'alimentation globale. Plusieurs études (246,247) ont utilisé un score modifié (« mPNNS-GS ») comprenant uniquement les 12 composantes alimentaires (valeur maximale du score à 13,5) afin d'étudier

spécifiquement la qualité de l'alimentation et non une combinaison de la qualité alimentaire et de l'activité physique. Dans le cadre de cette thèse, nous avons utilisé ce score modifié. Afin de maximiser le potentiel de découverte de biomarqueurs associés à une moindre qualité ou à une meilleure qualité de l'alimentation, les participants des études transversales ont été sélectionnés soit parmi le 1<sup>e</sup> décile de la distribution du score mPNNS-GS dans la population (individus ayant une très mauvaise adéquation aux recommandations) soit parmi le 10<sup>e</sup> décile du score (individus ayant une très bonne adéquation aux recommandations). Par ailleurs, ici encore, chaque individu au sein de chaque groupe (sélection au sein du 1<sup>e</sup> décile ou sélection au sein du 10<sup>e</sup> décile) partageait des caractéristiques similaires avec un individu de l'autre groupe.

Tableau 6 Composantes et système de notation pour le calcul du mPNNS-GS (recommandations de 2001)

	Recommandations	Système de notation	Score
1. Fruits et légumes	5 portions/jour	[0-3,5[	0
		[3,5-5[	0,5
		[5-7,5[	1
		≥7,5	2
2. Pain, céréales, pommes de terre et légumineuses	A chaque repas selon l'appétit	[0-1[	0
		[1-3[	0,5
		[3-6[	1
		≥6	0,5
3. Produits céréaliers complets	Favoriser les aliments céréaliers complets et le pain complet	[0-1/3[	0
		[1/3-2/3[	0,5
		≥2/3	1
4. Lait et produits laitiers	3 portions/jour (individus ≥55 ans: 3 à 4 portions/jour)	[0-1[	0
		[1-2,5[	0,5
		[2,5-3,5] (≥55 ans: [2,5-4,5])	1
		>3,5 (≥55 ans : >4,5)	0
5. Viande, volaille, produits de la mer et oeufs	1 à 2 portions/jour	0	0
		]0-1[	0,5
		[1-2]	1
		>2	0
6. Produits de la mer	Au moins 2 portions/semaine	< 2/semaine	0
		≥ 2/semaine	1
7. Matière grasse ajoutée totale	limiter la consommation	Lipides des matières grasses ajoutées animales et végétales >16% AE/jour	0
		Lipides des matières grasses ajoutées animales et végétales ≤16% AE/jour	1
8. Matière grasse ajoutée	Favoriser les matières grasses végétales	Pas de consommation de matière grasse ajoutée végétale ou ratio matières grasses ajoutées végétale/ totale ≤0,5	0
		Pas de consommation de matière grasse ajoutée totale ou ratio matières grasses ajoutées végétale / totale >0,5	1
9. Produits sucrés	limiter la consommation	Sucre ajouté issu des produits sucrés ≥15% AE/jour	-0,5
		Sucre ajouté issu des produits sucrés : 10 à 15% AE/jour	0
		Sucre ajouté issu des produits sucrés <10% AE/jour	1
10. Boissons sucrées	limiter la consommation à <1 verre/jour	<1l eau et >250 ml soda/jour	0
		≥1l eau et >250 ml soda/ jour	0,5
		<1l eau et ≤250 ml soda/ jour	0,75
		≥1l eau et ≤250 ml soda/ jour	1
11. Alcool	≤2 verres de vin/jour pour les femmes et ≤3 verres de vin/jour pour les hommes	Ethanol >20 g/jour pour les femmes et >30g/jour pour les hommes	0
		Ethanol ≤20 g/jour pour les femmes et ≤30g/jour pour les hommes	0,8
		Non-consommateurs ou consommateurs irréguliers (<1 verre de vin/semaine)	1
12. Sel	limiter la consommation	>12g/jour	-0,5
		]10-12] g/jour	0
		]8-10] g/jour	0,5
		]6-8] g/jour	1
		≤6 g/jour	1,5

AE, apport énergétique total sans l'alcool; mPNNS-GS, Programme National Nutrition Santé Guideline Score modifié.



## 1. Chez les femmes

Parmi les femmes incluses dans l'étude (7876), toutes celles ayant fourni moins de 10 enregistrements alimentaires de 24h ont été exclues afin d'obtenir des estimations des consommations alimentaires suffisamment précises (N=5262). Les femmes n'ayant pas d'échantillon plasmatique disponible et ayant des données manquantes sur les potentiels facteurs de confusion ont été exclues (N=126). 2488 participantes ont donc constitué la population éligible pour cette étude transversale chez les femmes. Une sélection aléatoire a été réalisée parmi cette population éligible : 80 femmes au sein du 1<sup>e</sup> décile du score mPNNS-GS et 80 femmes au sein du 10<sup>e</sup> décile pour les analyses H-RMN et LC-MS. Chaque femme de chaque groupe a été appariée avec une femme de l'autre groupe ayant des caractéristiques similaires au moment du prélèvement sur des classes d'âge (35-39 ans/40-44 ans/45-49 ans/45-49 ans/50-54 ans/55-59 ans/>60 ans), sur le statut ménopausique, sur des classes d'IMC (sous poids : <18.5kg/m<sup>2</sup> , poids normal : ≥18.5 kg/m<sup>2</sup> et <25 kg/m<sup>2</sup> et surpoids/obèses : ≥25 kg/m<sup>2</sup>), sur le statut tabagique (fumeuses ou non/anciennes fumeuses) et sur la saison de prélèvement sanguin (octobre/novembre, décembre/janvier/février ou mars/avril). Le flow chart de cette sélection est présenté en Figure 12.

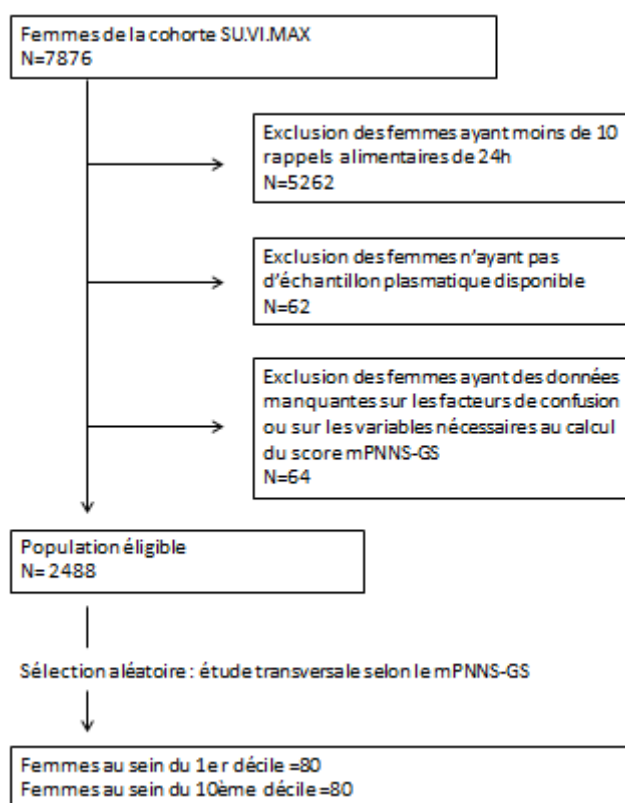


Figure 12 Flow chart de l'étude transversale selon le score mPNNS-GS chez les femmes

## 2. Chez les hommes

Parmi les hommes inclus dans l'étude (5141), tous ceux ayant fourni moins de 10 enregistrements alimentaires de 24h ont été exclus (N=3108). Les hommes n'ayant pas d'échantillon plasmatique disponible et ayant des données manquantes sur les potentiels facteurs de confusion ont été exclus (N=117). 1016 participants ont donc constitué la population éligible pour l'étude transversale chez les hommes. Une sélection aléatoire a été réalisée parmi cette population éligible : 70 hommes au sein du 1<sup>er</sup> décile du score mPNNS-GS et 70 hommes au sein du 10<sup>ème</sup> décile (analyses en H-RMN seulement). Chaque homme de chaque groupe a été apparié avec un homme de l'autre groupe ayant des caractéristiques

similaires au moment du prélèvement sur des classes d'âge (<45 ans/45-49 ans/50-54 ans/55-59 ans/≥60 ans), sur des classes d'IMC (sous poids : <18.5kg/m<sup>2</sup> , poids normal : ≥18.5 kg/m<sup>2</sup> et <25 kg/m<sup>2</sup> et surpoids/obèses : ≥25 kg/m<sup>2</sup>), sur le statut tabagique (fumeurs ou non/anciens fumeurs) et sur la saison de prélèvement sanguin (octobre/novembre, décembre/janvier/février ou mars/avril). Le flow chart de cette sélection est présenté en Figure 13.

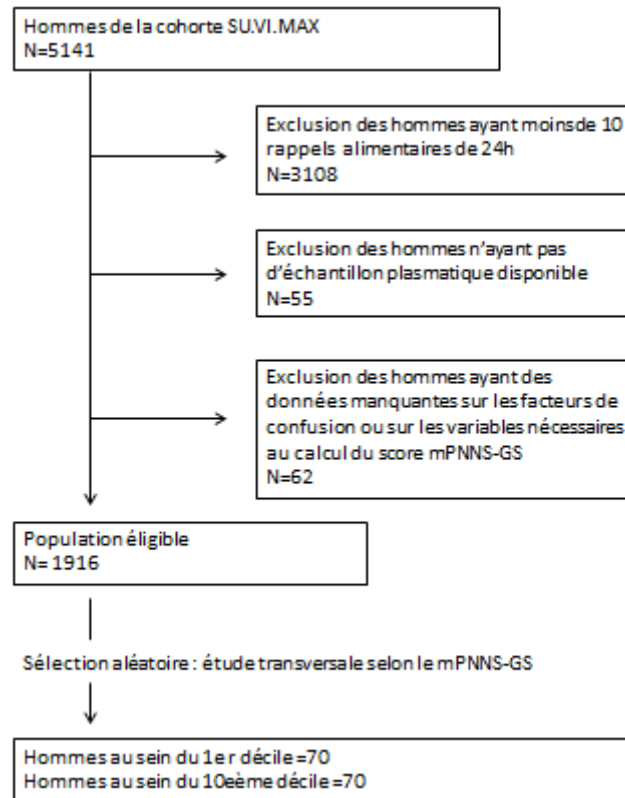


Figure 13 Flow chart de l'étude transversale selon le score mPNNS-GS chez les hommes

#### IV. Analyses métabolomiques des échantillons plasmatiques et prétraitements des données

La spectrométrie de masse (MS) et la Résonance Magnétique Nucléaire (RMN) sont actuellement les technologies les plus utilisées en métabolomique (51). Elles permettent de fournir des informations spectroscopiques et structurales d'un large éventail de classes de composés simultanément, avec une grande précision analytique (248). Dans le cadre de cette thèse, elles ont toutes les deux été utilisées pour les études chez les femmes, et seule la RMN a été utilisée chez les hommes.

##### A. Principes généraux

###### 1. La RMN

La spectroscopie RMN repose sur la détection du phénomène de résonance magnétique nucléaire qui se produit lorsque des noyaux atomiques de spin non nuls sont placés dans un champ magnétique et qu'ils sont excités par un rayonnement radiofréquence accordé sur les différences d'énergie entre les différents états possibles du spin nucléaire. A la fin de l'impulsion d'excitation, les spins retournent à leur état d'équilibre selon une fréquence de Larmor, propre à chaque atome. Ce signal est transformé en spectre (exemple en Figure 14) à l'aide de la transformation de Fourier (fonction mathématique). Cette technique

peut être applicable par exemple sur le  $^1\text{H}$ , le  $^{13}\text{C}$ , le  $^{15}\text{N}$  ou encore le  $^{31}\text{P}$ . Dans le cadre de cette thèse, nous avons utilisé la RMN du proton (H-RMN). Le signal des noyaux d'hydrogène est caractérisé par son déplacement chimique (exprimé en ppm de la valeur du champ magnétique ; en général de 0 à 10 ppm), en abscisse sur le spectre, et reflète directement la fréquence de résonance décrite précédemment. Le déplacement chimique dépend donc de la nature de l'atome et de son environnement. La présence de substituants comme des  $-\text{OH}$ ,  $=\text{O}$ , ou celle de liaisons insaturées ( $\text{C}=\text{C}$ ) affectent de façon caractéristique la valeur du déplacement chimique. Par ailleurs, les noyaux d'hydrogène reliés de manière covalente à un même atome ou à des atomes adjacents vont présenter des couplages, qui vont se traduire par une multiplicité du signal. Le spectre d'une molécule peut donc être constitué de plusieurs pics selon le nombre d'hydrogènes qu'elle porte. De plus, des protons ayant des environnements chimiques assez similaires peuvent se superposer sur le spectre, rendant l'interprétation plus difficile. L'interprétation des spectres se fait notamment par intégration de l'aire sous la courbe des pics (intensité du signal représenté en ordonnée) qui est proportionnelle au nombre de protons qui résonnent par molécule et à la concentration de la molécule dans l'échantillon.

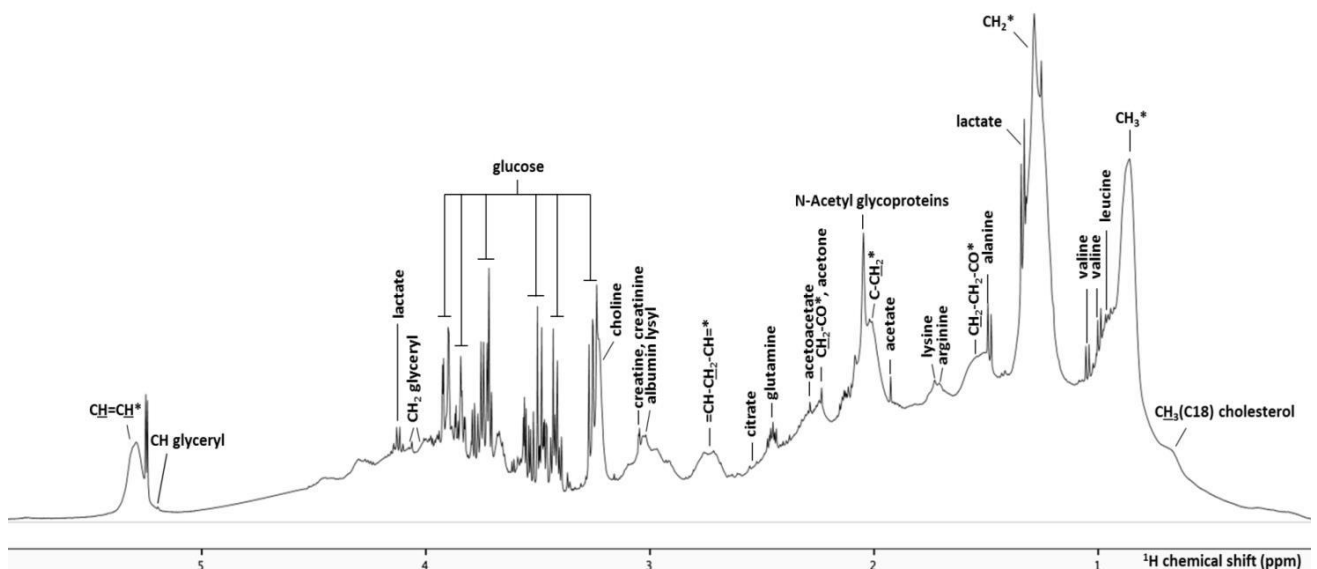


Figure 14 Exemple de spectre H-RMN (séquence NOESY-1D) obtenu dans le cadre de cette thèse

## 2. La spectrométrie de masse

Le principe de la spectrométrie de masse (MS) repose sur la séparation de molécules chargées (ions) en fonction de leur rapport masse/charge ( $m/z$ ). L'analyse en spectrométrie de masse se compose de 3 grandes étapes. La première étape consiste à vaporiser les molécules et les ioniser, pour produire des ions à partir d'un échantillon et grâce à une source d'ionisation. L'échantillon peut être introduit directement dans la source, sous forme gazeuse ou liquide (infusion directe) ou encore par couplage en amont à une méthode séparative selon les propriétés physico-chimiques des composés, la chromatographie liquide (LC-MS) ou en phase gazeuse (GC-MS). Lors de ce processus d'ionisation, les molécules se fragmentent et peuvent ainsi produire plusieurs ions. Plusieurs types de sources existent et sont utilisés en fonction du couplage en amont, du résultat recherché, et des molécules à analyser. Lorsque la séparation se fait par chromatographie liquide, la source généralement utilisée est l'électrospray. Cette combinaison est optimale pour l'analyse des fluides biologiques. La deuxième étape consiste à séparer et analyser les ions produits selon leur rapport  $m/z$ . Plusieurs manières de mesurer ce rapport existent, notamment selon

l'énergie cinétique des ions, leur vitesse ou encore leur mouvement périodique. Enfin, les signaux sont détectés grâce à la conversion du courant ionique en courant électrique. Les données sont représentées dans un spectre de masse (exemple en Figure 15) afin de pouvoir traiter le signal. Le spectre représente en abscisse le rapport  $m/z$  et en ordonnée l'abondance relative (ou intensité) des ions étant proportionnelle à la concentration des molécules correspondantes dans l'échantillon.

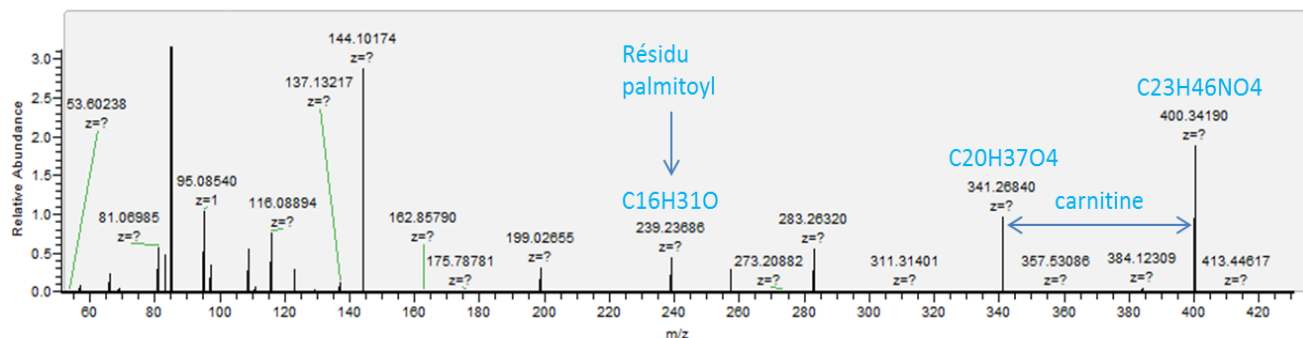


Figure 15 Exemple de spectre de masse (validation par MS/MS) obtenu dans le cadre de cette thèse

Par ailleurs, la détection des composés séparés est représentée par un chromatogramme caractérisé par un ensemble de pics gaussiens. Ces pics sont caractérisés par leur temps de rétention et leur intensité. Ainsi, les données LC/MS utilisées en métabolomique sont tridimensionnelles : chaque ion est caractérisé par son temps de rétention, sa masse et son intensité.

## B. Analyses métabolomiques réalisées dans le cadre de cette thèse

Aucune décongélation des échantillons plasmatiques n'a été réalisée avant ces études. Les analyses métabolomiques RMN et MS ont été réalisées de manière semi-quantitative (intensité relative) et non ciblée. Tous les signaux ont donc été analysés statistiquement et l'identification des métabolites d'intérêt s'est faite *a posteriori* (après les analyses statistiques).

### 1. Analyses en RMN

L'équipe Nanomédecine Biomarqueurs Détection (NBD) de l'unité de « Chimie Structures et Propriétés de Biomatériaux et d'Agents Thérapeutiques » (CSPBAT, UMR 7244) a généré et prétraité les données de H-RMN.

#### Préparation des échantillons et acquisition spectrale

Les échantillons de plasma ont été analysés en H-RMN. Tous les échantillons ont été traités de la même façon, à l'aveugle sur le type de groupe auquel appartenait le sujet (cas/témoin, 1<sup>er</sup> décile/10<sup>ème</sup> décile) et ont été extraits le même jour par le même technicien. La préparation des échantillons a été la même pour toutes les analyses RMN réalisée dans cette thèse. Un volume de 0,35 mL de plasma décongelé a été placé dans un tube RMN de 5 mm de diamètre avec une solution deutérium de PBS (0,25 mL) et NaN<sub>3</sub>. Une randomisation a été effectuée pour le positionnement en lots des échantillons. L'acquisition des spectres s'est faite à l'aide de séquences d'impulsion. La première séquence ayant été réalisée est la séquence NOESY1D (One-dimensional 1H Nuclear Overhauser Effect Spectroscopy) (249), très utilisée en RMN et permettant une meilleure suppression du signal de l'eau en empêchant le phénomène de résonance des protons de l'eau. Cette séquence ne nécessite que peu d'optimisation des paramètres et c'est une séquence robuste et reproductible (250). La deuxième séquence utilisée est la CPMG (Carr-Purcell-

Meiboom-Gill) (251), permettant de mettre en évidence les signaux de petites molécules comme les acides aminés en limitant les signaux des macromolécules telles que les protéines et les lipides. Les expériences NOESY1D et CPMG ont été utilisées avec 128 acquisitions du signal. Les spectres ont été acquis, pour chaque échantillon de plasma, à 500 MHz sur un spectromètre Bruker AVANCE III avec changeur automatique à 300 K. Les spectres ont été enregistrés automatiquement. La fenêtre spectrale était de 10 kHz. Le signal de l'eau a été supprimé par une séquence de présaturation utilisant une irradiation de faible puissance sur la fréquence du signal de l'eau pendant le délai de relaxation. 39 échantillons de plasma de sujets de la cohorte SU.VI.MAX mélangés ont été utilisés comme contrôles qualité ; les spectres ayant été enregistrés tous les 20 échantillons pour assurer la reproductibilité de l'acquisition des données RMN.

### *Prétraitements*

La transformation de Fourier a été appliquée aux données RMN avec un fenêtrage exponentiel pour produire un élargissement de ligne de 0,3 Hz. Les spectres ont été phasés et une correction de la ligne de base a été appliquée avec 10 points (de 0 à 10 ppm). Tous les spectres ont été alignés sur le doublet de glucose à 5,23 ppm. Les spectres ont été normalisés sur la somme des intensités pour supprimer l'effet de dilution. Chaque spectre a été découpé automatiquement tous les 0,001 ppm et l'intégrale de l'aire sous la courbe de chacune de ces zones a été calculée. Tous ces prétraitements ont été réalisés à l'aide du logiciel NMRPipe (252). La procédure "intelligent bucketing" (253) a ensuite été appliquée afin de reconstituer les pics. Les entités obtenues ont été standardisées pour chaque spectre sur la somme totale des intégrales des aires sous les pics, puis utilisées dans les modèles statistiques (appelées par la suite "signaux RMN").

### *Identification des métabolites*

L'assignation la plus probable des signaux de RMN a été obtenue en collectant des informations provenant de séquences 2D supplémentaires (1H-1H TOCSY et J-resolved), acquises sur un ensemble d'échantillons représentatifs. La TOCSY (Total Correlation Spectroscopy) fournit les couplages spin-spin donnant des informations sur la proximité des hydrogènes en termes de liaison chimique. La J-resolved donne des informations sur la multiplicité et les couplages scalaires aidant à l'identification des molécules (254). Le logiciel CHENOMX, la base de données HMDB et la littérature ont été utilisés pour confirmer les assignations.

## *2. Analyses en spectrométrie de masse*

La Plate-Forme d'Exploration du Métabolisme (PFEM) de l'Unité du Nutrition Humaine de l'INRA de Clermont-Ferrand (UMR 1019) a généré et prétraité les données de LC-MS.

### *Préparation des échantillons et acquisition*

Les échantillons de plasma ont été analysés par chromatographie liquide à ultra haute performance (UPLC, gradient de 26 min) couplée à un spectromètre de masse à haut débit et à haute résolution (quadripôle couplé à un temps de vol QToF Impact HDII, Bruker) ou à ultra haute résolution (LTQ-Orbitrap) équipé d'une source d'ionisation par électrospray (ESI). La préparation des échantillons était différente selon les objectifs de cette thèse. Elle sera détaillée plus loin. La séparation chromatographique a été réalisée à l'aide d'une colonne WatersAcquity HSS T3 C18 (150 x 2,1 mm), avec une taille de particules de 1,8 µm. Les échantillons ont été conservés à 4°C pendant les analyses. Les profils métabolomiques ont été acquis

en mode positif et négatif avec une plage de balayage pour les rapports masse/charge ( $m/z$ ) de 50 à 1000. Les échantillons ont été randomisés au sein de la séquence analytique selon le carré Latin de Williams Latin Square définie en fonction du statut du participant (cas/témoin, 1<sup>er</sup> ou 10<sup>ème</sup> décile). La stabilité du système d'analyse a été contrôlée à l'aide d'échantillons mélangés utilisés comme contrôles qualité, injectés une fois au début de chaque séquence, puis après chaque série de 12 échantillons.

#### *Extraction des données*

Les données brutes ont été traitées par la plate-forme sur internet Galaxy (255) à l'aide du module XCMS (détection des pics, regroupement et alignement chromatographique) et une correction de la dérive du signal a été réalisée selon l'algorithme décrit par Van der Kloet et al (256) pour obtenir une matrice contenant les temps de rétention et les masses des ions ainsi que les intensités des pics ayant été corrigées sur les effets de lots. Après correction de la dérive du signal, seuls les ions ayant une variabilité des intensités dans les échantillons inférieure à celle des contrôles qualité et ne correspondant pas à des isotopes ont été sélectionnés. En raison du large éventail d'extraction des ions, ceux ayant une intensité nulle dans plus de 30% des contrôles de qualité par lot analytique ont également été rejetés.

#### *Identification des métabolites*

Les ions d'intérêt (c'est-à-dire les ions discriminants ressortant des analyses statistiques) ont été annotés grâce à plusieurs étapes. Dans un premier temps, une inspection manuelle des spectres QToF MS (Bruker Compass DataAnalysis version 4.3) a été réalisée pour les ions parents des molécules, les fragments et les adduits. Pour déterminer la formule élémentaire des ions parents, un scan MS complet ( $m/z$  : 70 - 1 000) à haute puissance de résolution ( $R = 100\ 000$  à  $m/z = 800$ ) a été obtenu avec un spectromètre de masse hybride ThermoScientific LTQ Orbitrap Velos (Thermo Fisher Scientific, San José, CA, USA). Des hypothèses d'identification ont été obtenues en interrogeant des bases de données en ligne grâce à la formule élémentaire ou la masse monoisotopique (fenêtre de masse 0,005 Da) : Metlin (<https://metlin.scripps.edu>), HMDB (<http://www.hmdb.ca>), FooDB (<http://foodb.ca>), Dictionary of Food Compounds (<http://dfc.chemnetbase.com>), PhytoHub (<http://phytohub.eu>), ChEBI (<https://www.ebi.ac.uk/chebi/>), KNApSACK (<http://kanaya.naist.jp/KNApSACK>). La fragmentation des spectres (MS/MS) a été obtenue sur les spectromètres de masse LTQ Orbitrap (techniques « Higher-energy C-trap Dissociation ») et/ou QToF (énergies de collision 12eV, 20eV, 30eV et 40eV ; amplitude  $m/z$  de 7 Da) et comparée aux spectres de référence dans les bibliothèques spectrales internes à la plateforme (environ 1000 composés), en ligne (mzcloud : <https://www.mzcloud.org>, Metlin : <https://metlin.scripps.edu>, HMDB : <http://www.hmdb.ca>), à la littérature, et aux voies de fragmentation *in silico* prédites, générées avec MassFrontier (Thermo-Scientific). Les identifications ont été confirmées à l'aide de standards commerciaux disponibles. Les niveaux de preuve dans les identifications de la classification Sumner et al (257) ont été utilisés : le niveau 1 se réfère aux composés identifiés (masse monoisotopique, schéma de fragmentation MS/MS et temps de rétention correspondants à ceux des standards) ; le niveau 2 se réfère aux composés annotés de manière présumée (masse monoisotopique et schéma de fragmentation MS/MS correspondants aux spectres publiés dans les bibliothèques/publications/en ligne) ; le niveau 3 se rapporte aux classes de composés caractérisées de manière présumée (similarité spectrale avec une classe chimique) ; le niveau 4 se rapporte aux composés inconnus.

#### V. Estimation des expositions alimentaires

Dans le cadre de cette thèse basée sur des études d'observation, plusieurs types d'expositions alimentaires ont été estimés et utilisés selon les objectifs des études.

Tout d'abord, deux scores *a priori*, ayant des constructions différentes ont été utilisés afin d'estimer l'adéquation des participants aux recommandations et donc la qualité globale de leur alimentation. Le premier, le score mPNNS-GS, a été détaillé plus haut (Partie Méthodologie générale III.B.) et est basé sur les recommandations nutritionnelles nationales françaises concernant la consommation de groupes d'aliments. Le deuxième est le DQI-I (Diet Quality Index-International) (258), développé pour des comparaisons internationales. Le DQI-I est basé sur quatre aspects majeurs d'une alimentation saine et de haute qualité, à savoir la variété, l'adéquation, la modération et l'équilibre global, couvrant les préoccupations nutritionnelles des pays développés et en développement. Le DQI-I intègre des estimations sur l'apport à la fois en aliments et en nutriments (souvent moins hétérogène que l'apport en aliments à l'échelle mondiale) permettant de mieux décrire la diversité de la consommation d'un pays à l'autre. Le score varie de 0 à 100 points ; 100 étant la plus haute adéquation aux recommandations et 0 la plus basse. La composante variété du régime alimentaire est évaluée de deux façons : la variété globale (0-15 points ; viande/volaille/poisson/œufs, produits laitiers/haricots, céréales, fruits, légumes) et la variété des sources protéiques (0-5 points ; viande, volaille, poisson, produits laitiers, haricots, œufs) afin d'évaluer si les apports sont variables entre ou au sein des groupes d'aliments. La composante adéquation évalue l'apport en composés alimentaires devant être fourni en quantité suffisante pour garantir une alimentation saine et prévenir la dénutrition (légumes, fruits, céréales, fibres, protéines, fer, calcium et vitamine C ; 0-5 points chacun). La composante modération permet d'évaluer la consommation d'aliments et de nutriments liés à des maladies chroniques et qui peuvent nécessiter des restrictions (matière grasse totale, matière grasse saturée, cholestérol, sodium et aliments à calorie vide ; 0-6 points chacun). La dernière catégorie examine l'équilibre global de l'alimentation en termes de proportionnalité des sources d'énergie et de la composition en acides gras (rapport en macronutriments glucides-protéines-graisses, 0-6 points ; rapport en acides gras AGPI- AGMI- AGS, 0-4 points). Dans le cadre de cette thèse, le système de score pour les composantes d'adéquation et de modération correspondait aux recommandations nutritionnelles françaises (259). Dans d'autres indices alimentaires, les composantes d'adéquation et de modération sont combinées en une seule composante, rendant plus difficile l'interprétation d'un faible score de qualité alimentaire (dû à un déficit ou à un excès alimentaire).

Une autre manière d'estimer l'alimentation globale des individus est d'utiliser des méthodes *a posteriori* : habitudes alimentaires observées dans la population d'étude sans *a priori* avec des techniques non supervisées. Dans le cadre de cette thèse, nous avons donc également réalisé une Analyse en Composante Principale (ACP) afin de mettre en évidence des profils alimentaires (« patterns ») expliquant le maximum de variabilité dans les données alimentaires. Notre population d'étude étant la cohorte SU.VI.MAX nous avons réalisé cette ACP sur l'ensemble de cette population ayant des données alimentaires disponibles avec une rotation varimax qui consiste à associer chacune des variables à un nombre réduit de facteurs et à représenter chaque facteur par un nombre limité de variables. La sélection du nombre de facteurs à interpréter a été réalisée grâce au diagramme des valeurs propres de la variation totale expliquée (Cattell's scree plot) (260). Cette ACP a été réalisée sur 29 groupes d'aliments détaillés juste après.

Des groupes d'aliments ont également été estimés afin d'étudier des aspects plus spécifiques de l'alimentation. Deux catégories de groupes d'aliments ont été définies. La première, utilisée pour l'ACP, consistait à mettre en évidence des groupes d'aliments (N=29) assez larges correspondant à des comportements alimentaires ou des familles d'aliments globales (groupes généraux). La deuxième

consistait à mettre en évidence des groupes d'aliments plus spécifiques ou des aliments individuels (N=58) pour mettre en évidence des potentiels biomarqueurs spécifiques. Certains groupes d'aliments sont communs aux deux catégories (N=13). La liste de l'ensemble des groupes d'aliments est présentée dans le Tableau 7.

Tableau 7 Liste des 74 groupes d'aliments<sup>1</sup>

---

*Abats*  
*Agrumes*  
*Aliments à base de soja*  
*Allium*  
*Apéritifs (alcool)*  
*Autres alcools (cidre, champagne, etc.)*  
*Autres féculents (châtaigne, tapioca, plantain)*  
*Autres légumes (avocat, asperge, artichaut, champignon, céleri)*  
**Autres types d'aliments (condiments, produits énergétiques, à base de soja etc)**  
*Baies*  
*Banane*  
*Bières*  
**Boissons alcoolisées (totales)**  
**Boissons non sucrées (totales)**  
**Boissons sucrées**  
*Café*  
**Céréales pour petit-déjeuner**  
**Charcuterie**  
*Chocolat*  
*Choux*  
*Cola*  
*Condiments et aides culinaires*  
*Crustacés et fruits de mer*  
*Desserts lactés*  
*Fromage à pâte pressée cuite*  
*Fromages (sauf à pâte pressée cuite)*  
**Fromages (totaux)**  
**Fruits (totaux)**  
*Fruits à noyau*  
*Fruits exotiques*  
**Fruits secs (hors fruits à coque)**  
*Gâteaux et biscuits*  
*Haricots verts*  
*Infusion, chicorée et boissons sucrées artificiellement*  
**Jus de fruits**  
**Lait**  
**Légumes (totaux)**  
*Légumes-feuilles*  
*Légumes-fruits*  
**Légumineuses**  
**Matière grasse animale**  
*Matière grasse végétale (sauce huile d'olive et soja)*  
**Matière grasse végétale (totale)**  
**Noix/fruits à coque**  
**Œufs**  
*Olive*  
**Pain**  
*Pâtes et céréales*  
**Pâtes, riz et céréales (total)**  
**Pâtisseries et aliments sucrés à base de céréales (total)**  
*Poissons gras*  
*Poissons maigres*  
*Pomme et poire*



*Pommes de terre*

**Pommes de terre, racines et tubercules**

**Produits à base de céréales complètes**

*Produits à base de sucre/sucreries (miel, bonbons, confitures, etc.)*

**Produits de la mer (totaux)**

**Produits salés (totaux)**

*Produits salés apéritifs : oeufs de poisson, chips, biscuits salés et amuse-gueule à la viande*

**Produits sucrés (totaux)**

*Racines et tubercules*

*Raisin*

*Rhubarbe*

*Riz*

**Sauces**

*Thé*

*Tomates*

**Viande blanche**

*Viande rouge*

**Viandes rouges et abats**

*Vin rouge*

*Vins blancs et rosés*

**Yaourts**

<sup>1</sup> Les groupes en italique sont les groupes spécifiques et les groupes en gras sont les groupes généraux (en gras et italique : commun aux deux groupes).

## VI. Analyses statistiques

Différents modèles statistiques ont été utilisés dans le cadre de cette thèse selon les objectifs et les variables à expliquer. Ces analyses seront davantage détaillées de manière spécifique au sein de chaque objectif excepté les modèles utilisés par les plateformes de métabolomique, réalisés en routine sur les données sortantes, qui sont détaillés ci-dessous.

De manière générale, les plateformes de métabolomique réalisent des analyses statistiques multivariées sur les données obtenues soit comme étape de prétraitement, soit comme analyses statistiques préliminaires, soit comme analyses statistiques principales. Dans le cadre de cette thèse en épidémiologie, elles ont été réalisées pour les deux premières raisons.

Une analyse en composante principale a été réalisée au sein de chaque plateforme pour chaque objectif comme étape de prétraitement afin de vérifier qu'il n'y avait pas de valeurs aberrantes, de dérive liée à l'appareil de mesure ou de problèmes lors de la préparation des échantillons grâce aux contrôles qualité. Dans un deuxième temps l'ACP, utilisée comme analyse préliminaire permettait de voir si le statut des participants (cas/témoin ou appartenance au 1<sup>er</sup> décile/10<sup>e</sup> décile du mPNNs-GS) était la principale source de variabilité dans les données. L'ACP étant une analyse non supervisée, le statut était regardé *a posteriori* sur les score-plot obtenus, en colorant les points des individus selon leur statut. Cela permettait de voir si l'ACP distinguait les individus selon leur statut (deux groupes distincts).

Par la suite, des analyses supervisées PLS (Partial Least Squares) ont été réalisées au sein de chaque plateforme pour chaque objectif comme analyse préliminaire. Ce modèle permet de différencier les individus selon un ou plusieurs critères (Y) à partir des données observées (X). Lorsque le critère est binaire, le modèle utilisé est la PLS-DA (Discriminant Analysis). De manière générale, le critère testé était le statut du participant (avec des PLS-DA). La PLS construit des composantes correspondantes à des combinaisons linéaires des variables explicatives (X) d'une part et des variables à expliquer (Y) d'autre part, permettant d'expliquer au maximum la variabilité des X et des Y. Les critères de validité du modèle sont le R<sup>2</sup> permettant de voir si le modèle permet de bien expliquer les données (valeur devant être au plus

proche de 1) et le Q2 permettant de voir si le modèle permet de bien prédire les données grâce notamment à des méthodes de validation croisée (détaillée plus loin) (valeur devant être supérieure à 0.5). Des tests de permutation sont réalisés afin de valider le modèle en attribuant aléatoirement le critère et en vérifiant que les modèles obtenus avec ces distributions aléatoires sont moins bons que celui correspondant à la distribution réelle. Si le modèle est valide, le score-plot illustre la population d'étude séparée en deux pour les variables binaires. De manière générale, les variables X permettant de discriminer les individus selon le critère Y sont sélectionnées avec le critère VIP (Variable Importance for the Projection) résumant l'importance des variables dans l'explication et la prédiction du modèle. L'O-PLS-DA (ajout d'une variable orthogonale) a également été réalisée afin de mieux discriminer les groupes et de faciliter l'interprétation. L'O-PLS permet d'éliminer la variation systématique des données d'entrée X qui n'est pas corrélée au jeu de réponse Y, en d'autres mots cela revient à supprimer la variabilité dans X qui est orthogonale à Y.

Cependant, ces modèles multivariés ne prenaient pas en compte les potentiels facteurs de confusion pouvant amener à des résultats biaisés. Une approche épidémiologique a donc été appliquée par la suite pour l'ensemble des analyses statistiques réalisées sur les données de métabolomique.

Comme expliqué plus haut (Introduction générale IV.C), les données de grande dimension que sont les données de métabolomique peuvent être analysées par 3 grandes familles d'analyse et applicable en épidémiologie : les approches univariées et les procédures de correction de test multiple associées, les techniques de réduction de dimension et les approches de sélection de variables. Ces trois types d'analyses ont été utilisés dans le cadre de cette thèse dans une optique d'analyse exploratoire sans aucune hypothèse *a priori*. Les modèles utilisés pour ces analyses seront explicités plus loin, cependant une brève description est mentionnée ci-dessous notamment pour les aspects communs aux objectifs ainsi qu'une brève présentation des approches de pénalisation pouvant permettre la sélection de variables.

Dans un premier temps, les populations d'étude ont été décrites sur l'ensemble des caractéristiques des individus (socio-démographique, anthropométrique etc) grâce à des modèles permettant de prendre en compte l'appariement entre les deux groupes (cas/témoin ou appartenance au 1<sup>e</sup> décile/10<sup>e</sup> décile du mPNNS-GS). La régression logistique est utilisée pour ce type de variable réponse binaire permettant de mettre en évidence des différences sur certains paramètres (variables X explicatives) entre les deux groupes (variables Y à expliquer) et la régression logistique conditionnelle permet de prendre en compte l'appariement en comparant les individus au sein de chaque paire/triplet (témoin comparé à son cas apparié par exemple).

Ces modèles de régression logistique conditionnelle multi-ajustée ont également été utilisés pour comparer un à un l'intensité relative des métabolites issus des analyses métabolomiques entre les groupes appariés en prenant en compte l'ensemble des facteurs de confusion potentiels. Une correction de test multiple de Benjamini Hochberg de type FDR (False Discovery Rate) (261) a été systématiquement appliquée suite à ces analyses univariées. En revanche, dû notamment à l'approche non ciblée et exploratoire, nous n'avons pas fixé de seuil strict de significativité afin de limiter le risque de ne pas détecter des associations qui existent vraiment (faux négatifs). La plausibilité biologique a permis de discuter les résultats mais ils seront dans tous les cas à répliquer et valider dans des études indépendantes.

Des ACP (analyses non supervisées) ont été réalisées sur les métabolites discriminants dans les analyses univariées ci-dessus afin de mettre en évidence des profils de métabolites pouvant par exemple correspondre à des voies métaboliques spécifiques. Ces modèles d'ACP ont également permis de mettre

en évidence des profils alimentaires dans la population SU.VI.MAX comme expliquer précédemment (partie Méthodologie générale V.).

Des corrélations de Spearman ont été réalisées pour étudier les corrélations entre les métabolites 2 à 2 et également des corrélations partielles de Spearman entre les métabolites et les variables d'expositions nutritionnelles. Ce type de corrélation est une mesure de la dépendance statistique non paramétrique entre deux variables qui peuvent avoir une relation non linéaire, comme c'est possiblement le cas en biologie (effet seuil par exemple). Elle permet d'estimer un coefficient de corrélation, non pas entre les valeurs prises par les deux variables mais entre les rangs de ces valeurs. Les analyses de corrélations partielles permettent de prendre en compte l'ensemble des facteurs de confusion potentiels.

Une technique de sélection de variables a également été implémentée afin de sélectionner un ensemble de prédicteurs permettant de prédire conjointement un paramètre. Les techniques de régression pénalisée ont été appliquées dans la littérature pour contrôler la variance des estimations (qui augmente en présence de nombreux prédicteurs ou de multicollinéarité) en réduisant les coefficients vers zéro (229). Les deux grands types de pénalisation sont la pénalisation L1 correspondant à la méthode du LASSO (Least Absolute Shrinkage and Selection Operator)(229) et la pénalisation L2 correspondant à la régression Ridge (230). La pénalisation L1 permet de faire de la sélection automatique de variables en réduisant les coefficients des prédicteurs non pertinents directement à zéro. Cependant elle est instable en présence de variables très corrélées. La pénalisation L2 limite l'amplitude de valeurs des coefficients des paramètres et donc limite la variance et permet de gérer la multicollinéarité (variables très corrélées entre elles). Cependant, elle ne permet pas de faire de la sélection de variables. La méthode Elastic Net (231) a été développée afin de combiner les avantages des pénalisations L1 et L2, permettant donc de faire de la sélection de variables et de gérer la multicollinéarité. Dans le cadre de cette thèse, nous avons choisi d'implémenter la méthode Elastic Net, comme nous traitons des données biologiques (souvent très corrélées par exemple lorsqu'elles appartiennent à la même voie métabolique) issues de la métabolomique non ciblée (production de nombreux signaux dont certains appartiennent à la même molécule).

Ce modèle de régression prenait en compte tous les métabolites simultanément (intensité centrée-réduite) et les facteurs de confusion potentiels étaient inclus comme variables explicatives non pénalisées (forcés dans le modèle). Il a été exécuté sur le package R *glmnet* (262). Les paramètres de pénalisation d'Elastic Net  $\alpha$  et  $\lambda$  ont été optimisés à l'aide d'une validation croisée « 5-fold » (263) répétée 100 fois, pour tenir compte de la variance de la validation croisée. Cette validation croisée consistait à diviser la population en 5 sous-échantillons ; 4 des sous-échantillons étaient utilisés pour construire le modèle et le dernier pour calculer sa performance, chacun des sous-échantillons étant utilisé à son tour pour la construction et le calcul de la performance. Les paires d'individus appariés étaient réparties ensemble dans les sous-échantillons. Le paramètre  $\alpha$  a été optimisé sur une grille de 0,5 à 0,9 ( $\alpha=0$  n'implique aucune sélection de variables ;  $\alpha=1$  équivaut à une régression LASSO) afin d'obtenir suffisamment de sélection tout en contrôlant la multicollinéarité. Le paramètre  $\lambda$ , définissant la force de la pénalisation, a été optimisé en minimisant la déviation moyenne. Afin de déterminer la stabilité des métabolites sélectionnés par cette méthode, nous avons utilisé une méthode de ré-échantillonnage par bootstrap (264) répétant le processus de sélection 1000 fois sur des jeux de données ré-échantillonnés, afin d'enregistrer le pourcentage de fois où chaque métabolite a été sélectionné.

Dans le cadre de cette thèse, une estimation optimale de la performance prédictive des modèles n'a pas pu être établie puisque nous ne disposons pas de jeu de données (population) indépendant avec des

profils métabolomiques disponibles. Cependant, des approches alternatives ont tout de même été testées. La première méthode était d'estimer l'AUC (Area Under the receiver operating characteristic Curve) grâce à une validation croisée simple sur le modèle obtenu par Elastic Net. Un modèle de régression logistique non pénalisée contenant les métabolites sélectionnés (procédure décrite dans le paragraphe ci-dessus) ainsi que les facteurs de confusion était construit sur 4/5 de l'échantillon et la performance prédictive de ce modèle était calculée sur le 1/5 restant ; chaque sous-échantillon servant à son tour pour construire et calculer la performance. Ce procédé était répété 100 fois pour tenir compte de la variance de la validation croisée puis la moyenne de l'AUC du modèle et son intervalle de confiance ont été calculés sur 1000 jeux de données ré-échantillonnés (bootstrap).

La deuxième méthode était la double validation croisée. Pour cela, la procédure de sélection de métabolites par Elastic Net a été réitérée en laissant 1/5 du jeu de données à part à chaque fois. Une sélection de métabolites était donc réalisée 5 fois (validation croisée 5-fold) et un modèle de régression logistique non pénalisée pour chacune de ces sélections a permis de calculer une AUC sur les 1/5 du jeu de données non utilisés pour la sélection. Cette procédure, beaucoup plus longue en temps de calcul, a été répétée 10 fois et la moyenne et l'intervalle de confiance de l'AUC du modèle ont été estimés. Ces deux méthodes sont résumées en Figure 16 et Figure 17.

La première méthode permet de ne s'intéresser qu'aux métabolites sélectionnés en première intention sur le jeu de données initial. Cependant, elle ne prend pas en compte le fait que toutes les données ont été « vues » pendant ce processus de sélection et donc l'estimation de la performance prédictive finale est surestimée. La deuxième méthode, laissant de côté à chaque fois un sous-échantillon (1/5), qui n'est jamais utilisé pendant le processus de sélection, permet de surmonter cette limite. Cependant, la procédure Elastic Net étant réoptimisée pour tous les 4/5 d'échantillon, le calcul de performance se base sur des sélections de métabolites différentes à chaque fois, pouvant ne pas correspondre à celle initialement étudiée. Cette procédure est plutôt utilisée pour calculer la performance prédictive de la méthode que de la signature métabolomique d'intérêt.

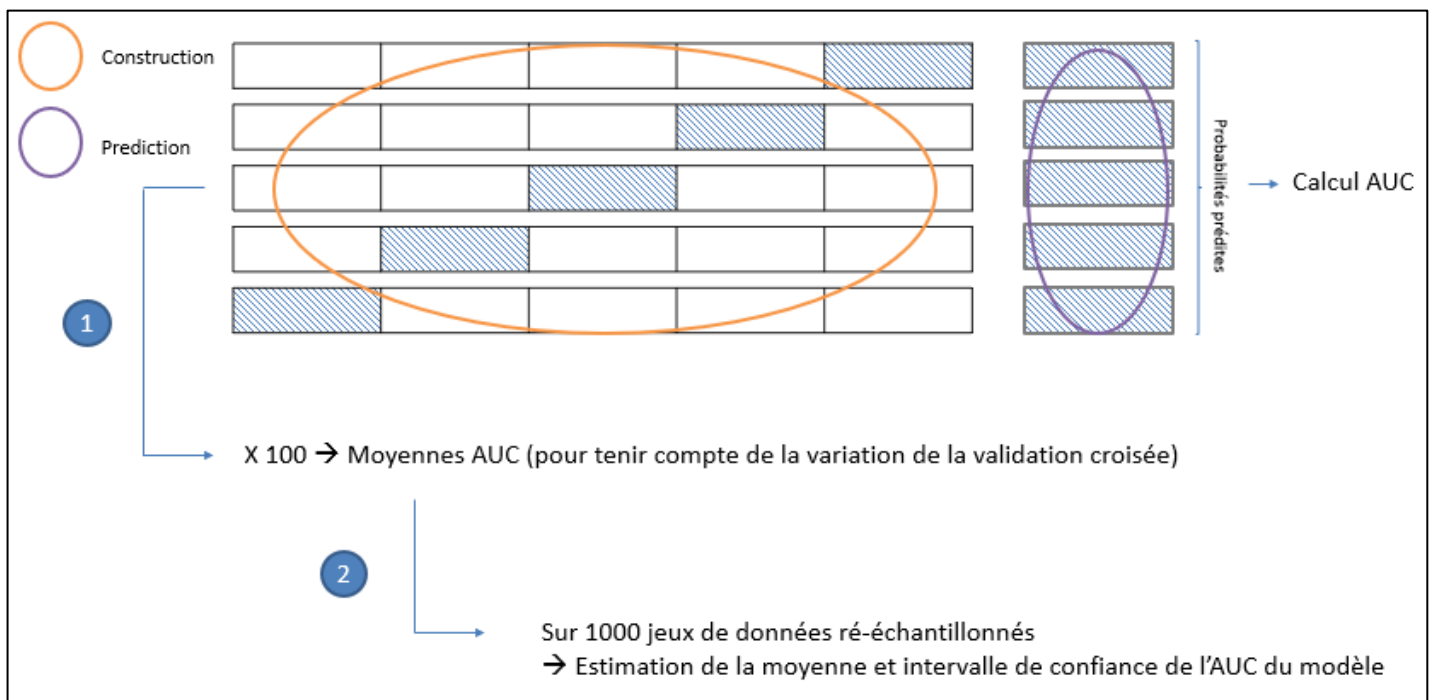


Figure 16 Procédure de la validation croisée simple pour le calcul de l'AUC



## Objectif 1 - Identifier des signatures métabolomiques plasmatiques associées au risque à long terme de cancers du sein (objectif 1.I.) et de la prostate (objectif 1.II.)

---

Ce premier objectif visait à explorer les données de métabolomique à partir de prélèvements plasmatiques d'individus *a priori* en bonne santé afin de mettre en évidence de potentiels biomarqueurs associés au risque de développer un cancer du sein ou de la prostate dans la décennie qui suit. Ces potentiels biomarqueurs peuvent appartenir aux trois grandes classes de biomarqueurs : exposition, effet et susceptibilité ; être sur le chemin causal menant vers le cancer ou être tout simplement associés au risque sans relation de causalité. Ces résultats pourraient apporter des éléments pour la compréhension des mécanismes liant facteurs endogènes et exogènes et mise en place de ces cancers ainsi que de nouvelles pistes de composés carcinogènes.

### I. Signature du risque de cancer du sein

#### A. Analyses en H-RMN

Cette étude a fait l'objet d'un article publié dans l'*International Journal of Epidemiology*, présenté en Annexe 1.

Lécuyer L, Victor Bala A, Deschasaux M, Bouchemal N, Nawfal Triba M, Vasson MP, Rossary A, Demidem A, Galan P, Herberg S, Partula V, Le Moyec L, Srour B, Fiolet T, Latino-Martel P, Kesse-Guyot E, Savarin P, Touvier M. **NMR metabolomic signatures reveal predictive plasma metabolites associated with long-term risk of developing breast cancer.** *Int J Epidemiol.* 2018 Apr 1;47(2):484-494. doi: 10.1093/ije/dyx271.

#### 1. Méthodes

La sélection de la population d'étude pour ces analyses a été détaillée dans la partie Méthodologie générale (III.A.1.). Il s'agit d'une étude cas-témoins nichée dans la population SU.VI.MAX avec 228 cas de cancer du sein incidents et 2 témoins appariés par cas (N=456).

Les échantillons de plasma des individus prélevés à l'inclusion ont été analysés en H-RMN avec une séquence NOESY1D et une séquence CPMG. Le détail de ces analyses est présenté dans la partie Méthodologie générale (IV.B.1.). Au total, 33 sujets (30 témoins et 3 cas) - ont été exclus en raison de problèmes d'acquisition de spectres et/ou de préparation des échantillons. La procédure "intelligent bucketing" (253) a été appliquée sur les spectres de 0 à 9 ppm afin de reconstituer les pics. Les régions du spectre comprises entre 5,5 et 9,0 ppm ont ensuite été supprimées car elles ne montraient que des signaux très faibles proches du niveau du bruit. Après prétraitements des données, 114 signaux RMN et 96 ont été obtenus pour la séquence NOESY1D et CPMG respectivement. La plateforme de métabolomique RMN a réalisé des analyses multivariées (détaillées de la partie Méthodologie générale (6.)) sur les signaux RMN obtenus.

La description de la population d'étude a été réalisée grâce à des modèles de régression logistique conditionnelle. Les associations entre chaque signal RMN et le risque de cancer du sein ont été estimées grâce à des modèles de régressions logistiques conditionnelles permettant de fournir des Odds Ratio (OR) exprimant le degré relatif de dépendance entre les variables (statut cas/témoin et intensité du signal) et les intervalles de confiance à 95% (IC95%). Les signaux RMN ont été analysés en continu et divisés par

l'écart type (OR correspondant à l'incrémentation d'un écart type) et codés en tertile (division de la distribution ordonnée en trois parties égales), le tertile le plus faible étant utilisé comme référence (OR correspondant donc à l'incrémentation d'une classe : du tertile 1 au tertile 2 ou du tertile 2 au tertile 3). Les modèles de régression logistique conditionnelle ont été ajustés sur les principaux facteurs de confusion potentiels disponibles, à savoir les caractéristiques suivantes à l'inclusion : âge (en continu), IMC (en continu), statut ménopausique, groupe d'intervention SU.VI.MAX, statut tabagique (fumeuses, anciennes fumeuses, non-fumeuses), saison de prélèvement sanguin (octobre/novembre, décembre/janvier/février ou mars/avril/mai), taille, consommation d'alcool, activité physique (irrégulière, <1h/jour de marche ou équivalent ou ≥1h/jour de marche ou équivalent), niveau d'éducation (primaire, secondaire, supérieur), antécédents familiaux de cancer du sein, nombre d'enfants et utilisation de la thérapie de substitution hormonale de la ménopause. Les signaux RMN associés avec le risque de cancer du sein et correspondant à la même molécule ont été combinés en une seule variable en sommant l'intensité des signaux. La correction de test multiple de Benjamini Hochberg générant des FDR a été appliquée. Les exclusions d'individus durant les analyses métabolomiques ont mené à l'exclusion d'autres sujets pour les analyses statistiques conditionnelles ci-dessus (comparaison au sein des triplets 1 cas + 2 témoins) nécessitant au minimum 1 cas et 1 témoin avec des données de métabolomique disponibles par triplet. Les données de 220 cas et 423 témoins pour la séquence NOESY1D et 216 cas et 402 témoins pour la séquence CPMG ont donc été utilisées pour les analyses statistiques.

Une matrice de corrélation 2 à 2 de Spearman des signaux d'intérêt a été réalisée. Des analyses de sensibilité ont été effectuées en excluant les cas (et leurs témoins appariés) diagnostiqués dans la première ou dans les deux premières années de suivi afin de contrôler un potentiel effet de causalité inverse (cancer déjà présent mais pas encore diagnostiqué qui impacterait le métabolisme). Des analyses stratifiées ont été réalisées sur le statut ménopausique au diagnostic, sur la durée de suivi entre le prélèvement sanguin et la survenue du cancer afin d'investiguer des associations avec le risque de cancer du sein à court ou à long terme, ainsi que sur les types de cancer du sein hormono-sensibles (ER+/ER- et PR+/PR-).

Une ACP normalisée avec rotation varimax a été réalisée sur les métabolites associés au risque de cancer du sein mis en évidence par la régression logistique conditionnelle. Les deux premières composantes (combinaisons linéaires/profils de métabolites) ont été sélectionnées puis analysées en régression logistique conditionnelle ajustée sur les facteurs de confusion cités plus haut afin de mettre en évidence des associations entre ces profils et le risque de cancer du sein.

## *2. Résultats*

Le Tableau 8 résume les caractéristiques des cas de cancer du sein et des témoins. Comparés aux témoins, les cas avaient tendance à être plus grandes et à avoir plus d'antécédents familiaux de cancer du sein.

Tableau 8 Caractéristiques des cas de cancer du sein et des témoins pour les analyses en H-RMN<sup>1</sup> – cohorte SU.VI.MAX, 1994-2007

	Cas de cancer du sein (N=228)	Témoins (N=456)	P <sup>2</sup>
Age (années)	49,1±6,0	49,1±6,1	0,5
IMC à l'inclusion (kg/m <sup>2</sup> )	23,1±3,9	23,5±3,9	0,06
			Non applicable
<18.5 (sous-poids)	10 (4,4)	11 (2,4)	
≥18.5 to <25 (poids normal)	161 (70,6)	331 (72,6)	
≥25 (surpoids)	57 (25,0)	114 (25)	
Taille à l'inclusion (cm)	162,9±6,1	161,0±5,9	0,0002
Groupe d'intervention			Non applicable
Placebo	110 (48,2)	220 (48,2)	
Supplémentation en antioxydants	118 (51,8)	236 (51,8)	
Statut tabagique à l'inclusion			Non applicable
Anciennes ou non fumeuses	179 (78,5)	358 (78,5)	
Fumeuses	49 (21,5)	98 (21,5)	
Activité physique à l'inclusion			0,5
Irrégulière	71 (31,1)	121 (26,5)	
< 1h/jour de marche ou équivalent	76 (33,3)	163 (35,7)	
≥ 1h/jour de marche ou équivalent	81 (35,5)	172 (37,7)	
Niveau d'éducation à l'inclusion			0,2
Primaire	43 (18,9)	95 (20,8)	
Secondaire	88 (38,6)	201 (44,1)	
Supérieur	97 (42,5)	160 (35,1)	
Nombre d'enfants biologiques à l'inclusion	1,9±1,2	2,0±1,1	0,2
Prise de la thérapie hormonale substitutive de la ménopause à l'inclusion (oui)	84 (36,8)	172 (37,7)	0,8
Statut ménopausique à l'inclusion			Non applicable
Préménopause	137 (60,1)	274 (60,1)	
Postménopause	91 (39,9)	182 (39,9)	
Statut ménopausique au diagnostic			Non applicable
Préménopause	81 (35,5)	Non applicable	
Postménopause	147 (64,5)	Non applicable	
Antécédents familiaux de cancer du sein <sup>3</sup> (oui)	36 (15,8)	48 (10,5)	0,048
Consommation d'alcool à l'inclusion (g/j)	11,4±11,9	11,3±13	1
Mois de prélèvement sanguin			Non applicable
Mars-Avril-Mai	84 (36,8)	168 (36,8)	
Octobre-Novembre	34 (14,9)	70 (15,4)	
Décembre-Janvier-Février	110 (48,2)	218 (47,8)	
Récepteurs aux œstrogènes			Non applicable
Positif	144 (63,16)	Non applicable	
Négatif	38 (16,67)	Non applicable	
Donnée non disponible	46 (20,17)	Non applicable	
Récepteurs à la progestérone			Non applicable
Positif	117 (51,32)	Non applicable	
Négatif	46 (20,18)	Non applicable	
Donnée non disponible	65 (28,50)	Non applicable	

<sup>1</sup> Les valeurs présentées sont les moyennes ± écart type ou l'effectif (%)

<sup>2</sup> P de tendance linéaire pour la comparaison entre les cas de cancer du sein et les témoins grâce aux modèles de régression logistique conditionnelle. Non applicable pour les variables d'appariement sauf si variables plus précises (ex. âge et IMC).

<sup>3</sup> Parmi le premier degré de parenté

Les résultats des analyses multivariées réalisées au sein de la plateforme de métabolomique n'ont pas permis de distinguer le statut des participants grâce aux signaux RMN analysés (ni par ACP, ni par PLS-DA). Un exemple graphique de l'ACP est présenté en Figure 18.



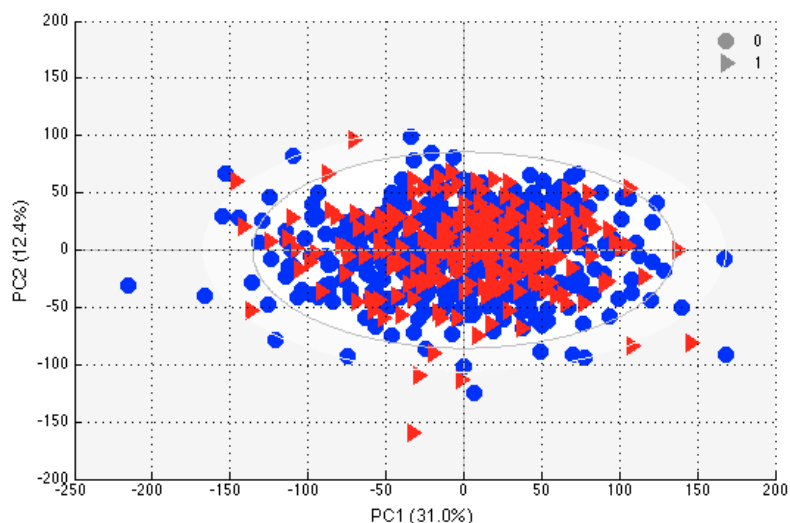


Figure 18 Score plot de l'ACP réalisée par la plateforme RMN entre les cas de cancer du sein (en rouge) et les témoins (en bleu)

Concernant les régressions logistiques conditionnelles, les résultats des associations ( $p < 0.05$ ) entre les signaux RMN et le risque de cancer du sein, présentés dans le Tableau 9, sont les résultats après exclusion des cas diagnostiqués lors de la première année de suivi et leurs témoins appariés et pour les signaux RMN codés en tertiles qui ont été valorisés dans l'article en Annexe 1. L'ensemble des associations (114 pour la séquence NOESY-1D et 96 pour la séquence CPMG) est présenté en Annexe 1.

Les femmes ayant des taux plasmatiques plus faibles en lipoprotéines, lipides, glycoprotéines, acétone, composés dérivés du glycérol comme les glycérides et les phospholipides et en acides gras insaturés et des taux plus élevés en valine, lysine, arginine, glutamine, créatine, créatinine et glucose avaient plus de risque de développer un cancer du sein au cours du suivi. Ces métabolites étaient globalement très corrélés entre eux (voir Annexe 1). Les associations observées pour les lipides, l'acétone et les composés dérivés du glycérol atteignaient un  $FDR < 0.05$ . Les résultats étaient sensiblement similaires sans exclusion des cas diagnostiqués lors de la première année de suivi, excepté pour les lipides insaturés qui n'étaient plus associés. Ces métabolites étaient toujours associés au risque de cancer du sein après exclusion des cas diagnostiqués pendant les deux premières années de suivi (et leurs témoins) sauf la glutamine (voir Annexe 1). En continu et sans exclusion, seuls la lysine, la créatine, la créatinine et le glucose étaient associés au risque de cancer du sein. Les métabolites lipidiques et la glutamine semblaient être plutôt associés au risque à court terme alors que les acides aminés et dérivés semblaient plutôt être associés au risque à long terme (voir Annexe 1). Certains métabolites semblaient être plutôt associés au risque de cancer du sein préménopausique alors que d'autres plutôt au risque de cancer du sein postménopausique sans que les tendances soient très marquées. Les métabolites lipidiques (sauf les lipides insaturés) et l'acétone étaient spécifiquement associés au risque de cancers du sein ER+ et PR+. Les glycoprotéines et les lipides insaturés étaient également associés au risque de cancer du sein PR+. Aucune association n'a été détectée pour le risque de cancers du sein ER- et PR-.

Tableau 9 Métabolites associés au risque de cancer du sein<sup>1</sup> issus des analyses H-RMN – cohorte SU.VI.MAX, 1994-2007

Métabolites mis en évidence avec la séquence NOESY1D								
Assignations	Déplacement chimique (ppm)	T1 (ref)		T2		T3		P
		cas/témoins	OR	cas/témoins	OR [IC95%]	cas/témoins	OR [IC95%]	
Lipoprotéines (CH <sub>3</sub> )	0,86	84/114	1,00	60/141	0,52 [0,33-0,82]	62/141	0,56 [0,36-0,88]	0,01
Lipoprotéines (CH <sub>2</sub> )	1,27	84/117	1,00	67/133	0,69 [0,44-1,07]	55/146	0,52 [0,33-0,82]	0,004
Lipides (CH <sub>2</sub> CH <sub>2</sub> CO)	1,56	84/117	1,00	72/128	0,7 [0,45-1,09]	50/151	0,43 [0,27-0,68]	0,0004
Pic non défini	1,83-1,85	78/120	1,00	67/133	0,72 [0,46-1,12]	61/143	0,55 [0,35-0,89]	0,01
Lipides (CCH <sub>2</sub> ) et glycoprotéines (N-acetyl) <sup>2</sup>	2,001 & 2,035	79/121	1,00	72/124	0,86 [0,55-1,33]	55/151	0,55 [0,35-0,87]	0,01
Lipides (CH <sub>2</sub> CO) & Acétone <sup>2</sup>	2,22	86/113	1,00	69/135	0,62 [0,4-0,95]	51/148	0,43 [0,27-0,69]	0,0004
Composés dérivés du glycérol (CH <sub>2</sub> )	4,04 & 4,25	90/113	1,00	60/141	0,54 [0,35-0,83]	56/142	0,49 [0,32-0,77]	0,001
Composés dérivés du glycérol (CH)	5,18	88/111	1,00	68/132	0,6 [0,38-0,94]	50/153	0,37 [0,23-0,61]	0,00007
Lipides insaturés (CH=CH)	5,28	80/122	1,00	68/129	0,79 [0,51-1,22]	58/145	0,60 [0,38-0,94]	0,03
Métabolites mis en évidence avec la séquence CPMG								
Assignations	Déplacement chimique (ppm)	T1 (ref)		T2		T3		P
		cas/témoins	OR	cas/témoins	OR [IC95%]	cas/témoins	OR [IC95%]	
Valine	0,98 & 1,03	53/134	1,00	69/129	1,35 [0,86-2,11]	81/115	1,83 [1,15-2,92]	0,01
Lipoprotéines (CH <sub>2</sub> )	1,26	78/114	1,00	70/128	0,83 [0,53-1,29]	55/136	0,58 [0,36-0,92]	0,02
Lysine et arginine <sup>2</sup>	1,70	57/139	1,00	63/121	1,19 [0,77-1,85]	83/118	1,62 [1,05-2,48]	0,03
Lipides (CH <sub>2</sub> CO) et acétone <sup>2</sup>	2,22	72/123	1,00	80/110	1,16 [0,76-1,78]	51/145	0,58 [0,36-0,93]	0,03
Non identifié	2,33	54/140	1,00	72/118	1,58 [1-2,51]	77/120	1,67 [1,08-2,6]	0,02
Glutamine	2,45	59/131	1,00	64/129	1,17 [0,74-1,84]	80/118	1,61 [1,02-2,55]	0,04
Pic non défini	2,55-2,57	52/140	1,00	77/115	1,74 [1,11-2,74]	74/123	1,63 [1,04-2,55]	0,04
Pic non défini	2,57-2,61	56/131	1,00	65/133	1,11 [0,71-1,76]	82/114	1,79 [1,15-2,8]	0,009
Lysine, Créatine & Créatinine <sup>2</sup>	3,01 & 3,03	52/141	1,00	70/120	1,64 [1,02-2,63]	81/117	1,84 [1,19-2,85]	0,007
Non identifié	3,34	56/140	1,00	72/115	1,72 [1,1-2,69]	75/123	1,59 [1,01-2,5]	0,04
Glucose	3,36-3,86	57/137	1,00	66/124	1,17 [0,73-1,86]	80/117	1,61 [1,04-2,5]	0,03
Non identifié	5,18	81/112	1,00	64/133	0,6 [0,38-0,94]	58/133	0,61 [0,39-0,97]	0,04
Pic non défini	5,38-5,43	81/115	1,00	61/129	0,68 [0,43-1,07]	61/134	0,63 [0,41-0,98]	0,046

T, tertile ; ref, référence ; OR, odds ratio ; IC, intervalle de confiance

<sup>1</sup> Modèles de régression logistique conditionnelle sur les signaux RMN codés en tertiles ajustés sur l'âge, l'IMC, le statut ménopausique, le groupe d'intervention SU.VI.MAX, le statut tabagique, la saison de prélèvement sanguin, la taille, la consommation d'alcool, l'activité physique, le niveau d'éducation, les antécédents familiaux de cancer du sein, le nombre d'enfants et l'utilisation de la thérapie de substitution hormonale de la ménopause. Des tests de tendance linéaire ont été réalisés.

<sup>2</sup> Assignations multiples pour le même signal RMN dû à la superposition spectrale

Les deux premières composantes de l'ACP permettaient d'expliquer 78% de la variabilité totale. La première composante était caractérisée par des taux plasmatiques élevés en lipoprotéines, lipides, acétone, lipides insaturés, composés dérivés du glycérol et glycoprotéines et des taux plasmatiques plus faibles en glutamine, lysine, créatine et créatinine. La deuxième composante était caractérisée par des

taux plasmatiques plus élevés en valine, lipoprotéines, lysine, arginine, lipides, acétone, glutamine, créatine, créatinine et glucose. Les associations entre ces composantes et le risque de cancer du sein issues des régressions logistiques conditionnelles ajustées sont présentées dans le Tableau 10. Seule la première composante était inversement associée au risque de cancer du sein.

Tableau 10 Associations entre les profils métabolomiques issus des analyses H-RMN obtenus par ACP et le risque de cancer du sein<sup>1</sup> – cohorte SU.VI.MAX, 1994-2007

Composantes <sup>2</sup>	T1 (ref)		T2		T3		P
	cas/témoins <sup>3</sup>	OR	cas/témoins <sup>3</sup>	OR [IC95%]	cas/témoins <sup>3</sup>	OR [IC95%]	
Composante 1	82/111	1,00	74/120	0,77 [0,48-1,22]	47/146	0,45 [0,28-0,72]	0,0008
Composante 2	67/126	1,00	65/129	0,93 [0,61-1,42]	71/122	1,07 [0,69-1,66]	0,8

T, tertile ; ref, référence ; OR, odds ratio ; IC, intervalle de confiance

<sup>1</sup> Modèles de régression logistique conditionnelle sur les signaux RMN codés en tertiles ajustés sur l'âge, l'IMC, le statut ménopausique, le groupe d'intervention SU.VI.MAX, le statut tabagique, la saison de prélèvement sanguin, la taille, la consommation d'alcool, l'activité physique, le niveau d'éducation, les antécédents familiaux de cancer du sein, le nombre d'enfants et l'utilisation de la thérapie de substitution hormonale de la ménopause. Des tests de tendance linéaire ont été réalisés.

<sup>2</sup> La première composante était caractérisée par des taux plasmatiques élevés en lipoprotéines, lipides, acétone, lipides insaturés, composés dérivés du glycérol et glycoprotéines (poids des variables  $\geq 0.15$ ) et des taux plasmatiques plus faibles en glutamine, lysine, créatine et créatinine (poids des variables  $\leq -0.15$ ). La deuxième composante était caractérisée par des taux plasmatiques plus élevés en valine, lipoprotéines, lysine, arginine, lipides, acétone, glutamine, créatine, créatinine et glucose (poids des variables  $\geq 0.15$ ).

<sup>4</sup> Seules les paires d'individus ayant des données métabolomiques disponibles dans les deux modes d'ionisation ont pu être prises en compte dans cette analyse. Résultats après exclusion des cas diagnostiqués lors de la première année de suivi et leurs témoins appariés.

## B. Analyses en LC-MS

Cette étude a fait l'objet d'un article publié dans le journal *Cancer, Epidemiology, Biomarker and Prevention* présenté en Annexe 2.

Lécuyer L, Dalle C, Lyan B, Demidem A, Rossary A, Vasson MP, Petera M, Lagree M, Ferreira T, Centeno D, Galan P, Herberg S, Deschasaux M, Partula V, Srouf B, Latino-Martel P, Kesse-Guyot E, Druetne-Pecollo N, Durand S, Pujos-Guillot E, Touvier M. **Plasma metabolomic signatures associated with long-term breast cancer risk in the SU.VI.MAX prospective cohort.** *Cancer Epidemiol Biomarkers Prev.* 2019 Jun 4. pii: cebp.0154.2019. doi: 10.1158/1055-9965.EPI-19-0154.

### 1. Méthodes

La sélection de la population d'étude pour ces analyses a été détaillée dans la partie Méthodologie générale (III.A.1.). Il s'agit d'une étude cas-témoins nichée dans la population SU.VI.MAX avec 215 cas de cancer du sein incidents et 1 témoin apparié par cas (N=215).

Les échantillons de plasma des individus prélevés à l'inclusion ont été analysés en UPLC-MS en mode positif et négatif. Le détail de ces analyses est présenté dans la partie Méthodologie générale (IV.B.2.). Les méthodes spécifiques à cet objectif sont détaillées ici. La préparation des échantillons a été réalisée selon la procédure légèrement modifiée décrite par Pereira et al (265). En résumé, 200  $\mu$ L d'échantillons de plasma ont été déprotéinés à l'aide de méthanol froid ; après centrifugation et évaporation sous azote, les résidus secs ont été redissous dans une préparation 50/50 (v/v) d'acétonitrile/eau contenant 0,1% d'acide formique. 6 $\mu$ L d'échantillon ont été injectés dans un système LC (Argilent 1200 series) couplé à un

spectromètre de masse microQTOF. Le débit a été réglé à 0,2 mL/min à 30°C. Les phases mobiles A et B étaient l'eau milliQ et l'acétonitrile respectivement, contenant chacune 0,1 % d'acide formique. Le gradient d'élution a été réalisé comme suit : 100% A (0-1 min) ; gradient linéaire à 100% B (1-13 min) ; 100% B (13-20 min) ; 100% A (20-26 min pour rééquilibrage). Les contrôles qualité ont été préparés de la même façon en mélangeant 20 µL de plasma de chaque échantillon extrait. Après un rapide aperçu de tous les chromatogrammes, 4 paires d'individus en mode positif et 12 paires en mode négatif ont été exclues en raison de problèmes lors de la préparation des échantillons, d'un excès d'ions ayant une intensité nulle (>50%) ou de pollution sur les chromatogrammes. Les données de 211 cas et 211 témoins en mode positif et de 203 cas et 203 témoins en mode négatif ont été utilisées pour les analyses statistiques. 1780 ions et 1785 ont été extraits en mode positif et négatif respectivement. Afin de tenir compte de la redondance inhérente à l'approche non ciblée (ions parents, fragments, isotopes, etc. correspondants à la même molécule) et de ne conserver que les ions pertinents (comme par exemple, les ions ayant une intensité suffisamment élevée), seuls les ions présents dans plus de 90% de la population des cas ou des témoins, ou ayant une intensité  $\geq 1,500$  (unité arbitraire) pour le 90<sup>e</sup> percentile ont été conservés et les corrélations analytiques ont été filtrées à l'aide de l'outil " Metabolites Correlation Analysis " (Galaxy Version 1.0.1) avec un seuil de coefficient de corrélation de 0,9. Ces étapes de prétraitements ont permis d'éliminer les ions considérés comme étant du bruit, caractérisé par une variabilité trop élevée, ainsi que de limiter la redondance entre les signaux. Après ces prétraitements, 530 et 691 ions (en mode positif ou négatif respectivement) ont été utilisés dans les analyses statistiques.

La description de la population d'étude et les analyses statistiques réalisées étaient sensiblement les mêmes que pour les analyses H-RMN (Partie Objectif 1 I.A.1.). En revanche, la combinaison des signaux associés avec le risque de cancer du sein et correspondants à la même molécule n'était pas applicable ici. Concernant l'ACP, les quatre premières composantes ont été sélectionnées pour cette étude.

## 2. Résultats

Le Tableau 11 résume les caractéristiques des cas de cancer du sein et des témoins. Comparés aux témoins, les cas avaient tendance à être plus grandes et à avoir plus d'antécédents familiaux de cancer du sein.

Les résultats des analyses multivariées réalisées au sein de la plateforme de métabolomique n'ont pas permis de distinguer le statut des participants grâce aux ions analysés (ni par ACP, ni par PLS-DA). Un exemple graphique de l'ACP est présenté en Figure 19.

Tableau 11 Caractéristiques des cas de cancer du sein et des témoins pour les analyses en LC-MS<sup>1</sup> – cohorte SU.VI.MAX, 1994-2007

	Cas de cancer du sein (N=215)	Témoins (N=215)	P <sup>2</sup>
Age (années)	49,0±6,0	48,9±6,0	0,4
IMC à l'inclusion (kg/m <sup>2</sup> )	23,1±3,9	23,6±4,2	0,09
			Non applicable
<18.5 (sous-poids)	9 (4,2)	6 (2,8)	
≥18.5 to <25 (poids normal)	152 (70,7)	155 (72,1)	
≥25 (surpoids)	54 (25,1)	54 (25,1)	
Taille à l'inclusion (cm)	163,0±6,1	161,0±6,0	0,0007
Groupe d'intervention			Non applicable
Placebo	105 (48,8)	105 (48,8)	
Supplémentation en antioxydants	110 (51,2)	110 (51,2)	
Statut tabagique à l'inclusion			Non applicable
Anciennes ou non fumeuses	168 (78,1)	168 (78,1)	
Fumeuses	47 (21,9)	47 (21,9)	
Activité physique à l'inclusion			0,8
Irrégulière	67 (31,2)	61 (28,4)	
< 1h/jour de marche ou équivalent	73 (34,0)	74 (34,4)	
≥ 1h/jour de marche ou équivalent	75 (34,9)	80 (37,2)	
Niveau d'éducation à l'inclusion			0,3
Primaire	38 (17,7)	44 (20,5)	
Secondaire	84 (39,1)	94 (43,7)	
Supérieur	93 (43,3)	77 (35,8)	
Nombre d'enfants biologiques à l'inclusion	1,9±1,2	2,1±1,2	0,1
Prise de la thérapie hormonale substitutive de la ménopause à l'inclusion (oui)	77 (35,8)	77 (35,8)	1
Statut ménopausique à l'inclusion			Non applicable
Préménopause	131 (60,9)	131 (60,9)	
Postménopause	84 (39,1)	84 (39,1)	
Statut ménopausique au diagnostic			Non applicable
Préménopause	79 (36,7)	Non applicable	
Postménopause	136 (63,3)	Non applicable	
Antécédents familiaux de cancer du sein <sup>3</sup> (oui)	36 (16,7)	21 (9,8)	0,03
Consommation d'alcool à l'inclusion (g/j)	11,2±11,9	11,5±13,2	0,8
Mois de prélèvement sanguin			Non applicable
Mars-Avril-Mai	79 (36,7)	79 (36,7)	
Octobre-Novembre	32 (14,9)	33 (15,3)	
Décembre-Janvier-Février	104 (48,4)	103 (47,9)	
Récepteurs aux œstrogènes			Non applicable
Positif	135 (62,79)	Non applicable	
Négatif	36 (16,74)	Non applicable	
Donnée non disponible	44 (20,47)	Non applicable	
Récepteurs à la progestérone			Non applicable
Positif	109 (50,70)	Non applicable	
Négatif	43 (20,00)	Non applicable	
Donnée non disponible	63 (29,30)	Non applicable	

<sup>1</sup> Les valeurs présentées sont les moyennes ± écart type ou l'effectif (%)

<sup>2</sup> P de tendance linéaire pour la comparaison entre les cas de cancer du sein et les témoins grâce aux modèles de régression logistique conditionnelle. Non applicable pour les variables d'appariement sauf si variables plus précises (ex. âge et IMC).

<sup>3</sup> Parmi le premier degré de parenté

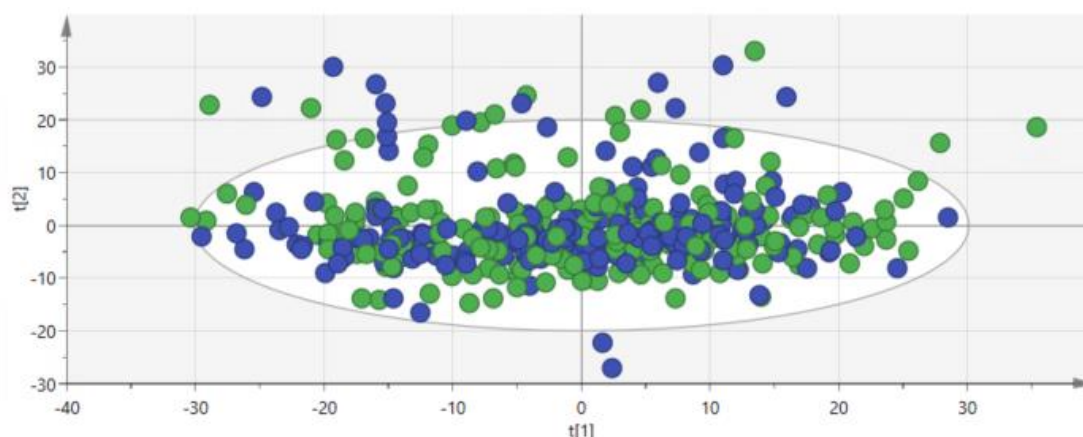


Figure 19 Score plot de l'ACP réalisée par la plateforme MS entre les cas de cancer du sein (en rouge) et les témoins (en bleu) pour les données en mode négatif.

Concernant les régressions logistiques conditionnelles, 53 ions étaient positivement associés au risque de cancer du sein et 20 étaient négativement associés au risque (pour une augmentation d'un écart type ;  $p \leq 0.01$ ,  $FDR \leq 0.2$ ). Parmi ces ions, 9 ont pu être identifiés et sont présentés dans le Tableau 12. Les résultats de l'ensemble de ces associations sont présentés dans l'Annexe 2.

Tableau 12 Métabolites identifiés et associés au risque de cancer du sein<sup>1</sup> issus des analyses LC-MS – cohorte SU.VI.MAX, 1994-2007

Métabolite (Niveau de preuve dans l'identification <sup>2</sup> )	Annotation	Masse	Temps de rétention (min)	Mode	Cas/ témoins	OR [IC95%]	P
L-phenylalanine (1)	[M+H] <sup>+</sup>	166,08618	10,11	ESI +	211/211	1,43 [1,14-1,78]	0,002
L-valine/norvaline (1)	[M+H-CH <sub>2</sub> O <sub>2</sub> ] <sup>+</sup>	72,08089	3,08	ESI +	211/211	1,45 [1,15-1,83]	0,002
Acide 5-aminovalérique (1)	[M+Na] <sup>+</sup>	140,06816	2,29	ESI +	211/211	1,46 [1,14-1,87]	0,003
O-succinyl-L-homoserine (1)	[M+Na-2H] <sup>-</sup>	240,04755	2,86	ESI -	203/203	0,70 [0,55-0,89]	0,004
ATBC (acetyl tributyl citrate) (1)	[M+Na] <sup>+</sup>	425,21694	19,31	ESI +	211/211	1,41 [1,10-1,79]	0,006
L-tryptophane (1)	[M+Na] <sup>+</sup>	227,07887	11,09	ESI +	211/211	1,40 [1,10-1,79]	0,007
L-γ-glutamyl-L-thréonine (1)	[M+H-H <sub>2</sub> O] <sup>+</sup>	231,1451	2,52	ESI +	211/211	1,39 [1,09-1,77]	0,008
L-glutamine / L-isoglutamine (1)	[M+H-C <sub>4</sub> H <sub>8</sub> O] <sup>+</sup>	90,0551	2,32	ESI +	211/211	1,33 [1,07-1,66]	0,009
Pregnene-triol sulfate (2)	C <sub>21</sub> H <sub>33</sub> O <sub>6</sub> S	413,19771	16,11	ESI -	203/203	1,38 [1,08-1,77]	0,01

OR, odds ratio ; IC, intervalle de confiance

<sup>1</sup> Modèles de régression logistique conditionnelle sur les ions en continu divisés par l'écart type ajustés sur l'âge, l'IMC, le statut ménopausique, le groupe d'intervention SU.VI.MAX, le statut tabagique, la saison de prélèvement sanguin, la taille, la consommation d'alcool, l'activité physique, le niveau d'éducation, les antécédents familiaux de cancer du sein, le nombre d'enfants et l'utilisation de la thérapie de substitution hormonale de la ménopause. Des tests de tendance linéaire ont été réalisés.

<sup>2</sup> Niveaux de preuve dans l'identification: (1) identifié, (2) annoté de manière présumée

Un taux plasmatique plus faible en O-succinyl-homoserine et des taux plasmatiques plus élevés en valine, glutamine, acide 5-aminovalérique, phénylalanine, tryptophane, γ-glutamyl-thréonine (GGT), acetyl tributyl citrate (ATBC) et en pregnene-triol sulfate étaient associés à une augmentation du risque de cancer du sein. Les associations concernant la phénylalanine et la valine atteignaient un FDR de 0.1 alors que les associations concernant les autres métabolites atteignaient un FDR de 0.2. Des tendances non linéaires ont été également observées pour le tryptophane, la valine, la phénylalanine et la glutamine (voir Annexe 2). Des corrélations positives ( $p < 0.05$ ) ont été observées pour le pregnene-triol sulfate avec l'acide 5-

aminovalérique, le tryptophane et la GGT; pour la phénylalanine avec la valine, le tryptophane, l'acide 5-aminovalérique et la glutamine; pour la valine avec l'acide 5-aminovalérique, le tryptophane et la glutamine; pour l'acide 5-aminovalérique avec la GGT; pour le tryptophane avec la glutamine; et pour la GGT avec la glutamine. Des corrélations négatives ont été observées pour l'O-succinyl-homoserine avec l'acide 5-aminovalérique et l'ATBC; et pour l'acide 5-aminovalérique et l'ATBC.

Ces métabolites étaient toujours associés au risque de cancer du sein après exclusion des cas diagnostiqués pendant la première année et pendant les deux premières années de suivi sauf la glutamine (voir Annexe 2). Aucune interaction n'a été observée entre le risque de cancer du sein et la durée de suivi. Un taux plasmatique plus élevé en glutamine était associé à une augmentation de risque de cancer du sein préménopausique. Tous les métabolites identifiés, sauf la glutamine, étaient associés au risque de cancers du sein ER+ et PR+. Ils n'étaient cependant pas associés au risque de cancers du sein ER- et PR- sauf la phénylalanine (ER- et PR-), l'ATBC et le pregnene-triol sulfate (PR-).

Les quatre premières composantes de l'ACP permettaient d'expliquer 36% de la variabilité totale. La première composante était caractérisée par des taux plasmatiques élevés en valine et acide 5-aminovalérique et la deuxième par des taux élevés en phénylalanine, valine, acide 5-aminovalérique et tryptophane. La troisième composante était caractérisée par un taux plus faible en pregnene-triol sulfate et la quatrième par des taux plus faibles en phénylalanine, valine, acide 5-aminovalérique, GGT et glutamine. De manière cohérente avec les résultats des régressions logistiques, les profils d'ions correspondants aux composantes 1, 2 et 4 étaient associés à une augmentation du risque de cancer du sein alors que le profil correspondant à la composante 3 était associé à une diminution du risque (résultats présentés dans le Tableau 13).

Tableau 13 Associations entre les profils métabolomiques issus des analyses LC-MS obtenus par ACP et le risque de cancer du sein<sup>1</sup> – cohorte SU.VI.MAX, 1994-2007

Composantes <sup>2</sup>	cas/témoins <sup>3</sup>	OR [95% CI]	P
Composante 1	200/200	1,38 [1,11-1,73]	0,004
Composante 2	200/200	1,47 [1,16-1,85]	0,001
Composante 3	200/200	0,49 [0,37-0,65]	<0,0001
Composante 4	200/200	1,49 [1,17-1,89]	0,001

OR, odds ratio ; IC, intervalle de confiance

<sup>1</sup> Modèles de régression logistique conditionnelle sur les ions en continu divisés par l'écart type ajustés sur l'âge, l'IMC, le statut ménopausique, le groupe d'intervention SU.VI.MAX, le statut tabagique, la saison de prélèvement sanguin, la taille, la consommation d'alcool, l'activité physique, le niveau d'éducation, les antécédents familiaux de cancer du sein, le nombre d'enfants et l'utilisation de la thérapie de substitution hormonale de la ménopause. Des tests de tendance linéaire ont été réalisés.

<sup>2</sup> La première composante était caractérisée par des taux plasmatiques élevés en valine et acide 5-aminovalérique, la deuxième par des taux élevés en phénylalanine, valine, acide 5-aminovalérique et tryptophane. La troisième composante était caractérisée par un taux plus faible en pregnene-triol sulfate et la quatrième par des taux plus faibles en phénylalanine, valine, acide 5-aminovalérique, GGT et glutamine. D'autres ions non identifiés caractérisaient ces composantes. Poids des variables pour l'interprétation  $\geq |0.20|$ .

<sup>3</sup> Seules les paires d'individus ayant des données métabolomiques disponibles dans les deux modes d'ionisation ont pu être prises en compte dans cette analyse.

## II. Signature du risque de cancer de la prostate – analyses en H-RMN

Cette étude fait l'objet d'un article en cours de rédaction.

*Lécuyer L, Victor Bala A, Demidem A, Rossary A, Bouchemal N, Nawfal Triba M, Galan P, Hercberg S, Deschasaux M, Partula V, Le Moyec L, Srouf B, Latino-Martel P, Kesse-Guyot E, Druesne-Pecollo N, Vasson MP, Savarin P, Touvier M. **Untargeted metabolomics profiles associated with long-term risk of prostate cancer risk.** En préparation*

### A. Méthodes

La sélection de la population d'étude pour ces analyses a été détaillée dans la partie Méthodologie générale (III.A.2.). Il s'agit d'une étude cas-témoins nichée dans la population SU.VI.MAX avec 171 cas de cancer de la prostate incidents et 1 témoin apparié par cas (N=171).

Les échantillons de plasma des individus prélevés à l'inclusion ont été analysés en H-RMN avec une séquence NOESY1D et une séquence CPMG. Le détail de ces analyses est présenté dans la partie Méthodologie générale (IV.B.1.). 9 paires et 12 paires d'individus ont été exclues en raison de problèmes d'acquisition de spectres et/ou de préparation des échantillons pour les séquences NOESY1D et CPMG respectivement. Les données de 162 cas et 162 témoins pour la séquence NOESY1D et de 159 cas et 159 témoins pour la séquence CPMG ont été utilisées pour les analyses statistiques. La procédure "intelligent bucketing" (253) a été appliquée sur les spectres de 0 à 8 ppm afin de reconstituer les pics. Après prétraitements des données, 159 signaux RMN et 178 ont été obtenus pour les séquences NOESY1D et CPMG respectivement. La plateforme de métabolomique RMN a réalisé des analyses multivariées sur les signaux RMN obtenus, détaillées de la partie Méthodologie générale (6.).

La description de la population d'étude et les analyses statistiques réalisées étaient sensiblement les mêmes que pour les analyses sur le cancer du sein (Partie Objectif 1 I.A.1.). Tous les modèles de régression logistique conditionnelle ont été ajustés sur l'âge (en continu), l'IMC (en continu), le statut tabagique (non fumeurs, arrêt de la cigarette depuis plus de 16ans, arrêt de la cigarette depuis moins de 16 ans et plus de 6, arrêt de la cigarette depuis moins de 6 ans, fumeurs), taille, consommation d'alcool, activité physique ((irrégulière, <1h/jour de marche ou équivalent ou ≥1h/jour de marche ou équivalent), niveau d'éducation (primaire, secondaire, supérieur), antécédents familiaux de cancer de la prostate. Seuls les cas diagnostiqués durant la première année de suivi (et non pendant la deuxième) ont été exclus lors des analyses de sensibilité en raison de la puissance limitée. Une autre analyse de sensibilité a été réalisée en excluant les cas ayant un dysfonctionnement prostatique à l'inclusion (taux de PSA ≥4µg/l). Des analyses stratifiées selon le grade de la tumeur (Gleason <7 ou ≥7). Concernant l'ACP, les trois premières composantes ont été sélectionnées.

### B. Résultats

Le Tableau 14 résume les caractéristiques des cas de cancer de la prostate et des témoins. Comparés aux témoins, les cas avaient tendance à être plus âgés, à avoir plus d'antécédents familiaux de cancer de la prostate et avoir un taux supérieur de PSA (prostate-specific antigen) à l'inclusion.



Tableau 14 Caractéristiques des cas de cancer de la prostate et des témoins<sup>1</sup> – cohorte SU.VI.MAX, 1994-2007

	Cas de cancer de la prostate (N=171)	Témoins (N=171)	p <sup>2</sup>
Age (années)	54,5±4,7	54,3±4,4	0,049
IMC à l'inclusion (kg/m <sup>2</sup> )	25,6±3,2	25,4±3,5	0,5
			Non applicable
<18.5 (sous-poids)	1 (0,6)	1 (0,6)	
≥18.5 to <25 (poids normal)	84 (49,1)	84 (49,1)	
≥25 (surpoids)	86 (50,3)	86 (50,3)	
Taille à l'inclusion (cm)	173,48±6,4	174,5±6,2	0,1
Groupe d'intervention			Non applicable
Placebo	94 (55,0)	94 (55,0)	
Supplémentation en antioxydants	77 (45,0)	77 (45,0)	
Statut tabagique à l'inclusion			Non applicable
Anciens ou non fumeurs	153 (89,6)	153 (89,6)	
Fumeurs	18 (10,5)	18 (10,5)	
Activité physique à l'inclusion			0,7
Irrégulière	36 (21,1)	42 (24,6)	
< 1h/jour de marche ou équivalent	44 (25,7)	39 (22,8)	
≥ 1h/jour de marche ou équivalent	91 (53,2)	90 (52,6)	
Niveau d'éducation à l'inclusion			0,3
Primaire	36 (21,1)	45 (26,3)	
Secondaire	63 (36,8)	67 (39,2)	
Supérieur	72 (42,1)	59 (34,5)	
Antécédents familiaux de cancer de la prostate <sup>3</sup> (oui)	21 (12,3)	8 (4,7)	0,02
Consommation d'alcool à l'inclusion (g/j)	29,1±21,6	25,4±22,5	0,7
Mois de prélèvement sanguin			Non applicable
Mars-Avril-Mai	67 (39,2)	67 (39,2)	
Octobre-Novembre	21(12,3)	20 (11,7)	
Décembre-Janvier-Février	83 (48,5)	84 (49,1)	
Grade de la tumeur			Non applicable
Haut grade (Gleason ≥7)	72 (42,11)	Non applicable	
Bas grade (Gleason <7)	88 (51,46)	Non applicable	
Donnée non disponible	11 (6,43)	Non applicable	
Prostate-specific antigen (PSA) total à l'inclusion (µg/l)	3,22±3,27	1,26±1,20	<0,0001
			<0,0001
PSA≥4µg/l	40 (23,39)	5 (2,92)	
PSA<4µg/l	131 (76,61)	166 (97,08)	

<sup>1</sup> Les valeurs présentées sont les moyennes ± écart type ou l'effectif (%)

<sup>2</sup> P de tendance linéaire pour la comparaison entre les cas de cancer de la prostate et les témoins grâce aux modèles de régression logistique conditionnelle. Non applicable pour les variables d'appariement sauf si variables plus précises (ex. âge et IMC).

<sup>3</sup> Parmi le premier degré de parenté

Les résultats des analyses multivariées réalisées au sein de la plateforme de métabolomique n'ont pas permis de distinguer le statut des participants grâce aux signaux RMN analysés (ni par ACP, ni par PLS-DA).

Concernant les régressions logistiques conditionnelles, 26 signaux RMN étaient positivement associés au risque de cancer de la prostate alors que 13 signaux étaient associés négativement ( $p < 0,05$ , voir Tableau 15). Les résultats des associations de l'ensemble des signaux sont présentés en Annexe 3.

Tableau 15 Métabolites associés au risque de cancer de la prostate<sup>1</sup> – cohorte SU.VI.MAX, 1994-2007

<b>Métabolites mis en évidence avec la séquence NOESY1D (N=162 cas/162 témoins)</b>			
<b>Assignations</b>	<b>Déplacement chimique (ppm)</b>	<b>OR [IC95%]</b>	<b>P</b>
Valine	0,97, 1,02	1,37 [1,07-1,76]	0,01
Pic non défini	1,65	1,37 [1,05-1,8]	0,02
Pic non défini	2,37	0,73 [0,57-0,95]	0,02
Glutamine	2,41	1,30 [1,00-1,70]	0,047
Pic non défini	2,58	1,42 [1,07-1,88]	0,02
Albumin lysil	2,89	1,48 [1,12-1,95]	0,006
Non identifié	2,93	1,39 [1,06-1,82]	0,02
Albumine lysil & créatine <sup>2</sup>	2,98	1,37 [1,04-1,80]	0,02
Urée	5,69, 5,73, 5,83	0,70 [0,54-0,92]	0,009
Pic non défini	5,84	0,71 [0,52-0,96]	0,02
Pic non défini	5,86	0,67 [0,49-0,91]	0,01
Pic non défini	6,07	0,63 [0,45-0,89]	0,008
Pic non défini	6,53	1,32 [1,03-1,69]	0,03
Pic non défini	6,96	1,36 [1,05-1,76]	0,02
Tyrosine	7,15, 7,18	1,33 [1,03-1,71]	0,03
Pic non défini	7,21	1,33 [1,02-1,73]	0,03
Histidine	7,69	1,34 [1,03-1,73]	0,03
<b>Métabolites mis en évidence avec la séquence CPMG (N=159 cas/159 témoins)</b>			
<b>Assignations</b>	<b>Déplacement chimique (ppm)</b>	<b>OR [IC95%]</b>	<b>P</b>
Valine	1,02	1,29 [1,02-1,64]	0,04
Pic non défini	2,15	0,77 [0,60-0,98]	0,04
Albumine	2,87	1,51 [1,13-2,02]	0,005
Urée	5,69, 5,73	0,75 [0,59-0,97]	0,03
Pic non défini	5,01	1,32 [1,01-1,72]	0,04
Pic non défini	5,08	1,34 [1,02-1,75]	0,04
Pic non défini	5,86	0,72 [0,55-0,94]	0,01
Pic non défini	6,53	1,34 [1,02-1,77]	0,03
Pic non défini	6,77	1,35 [1,03-1,77]	0,03
Tyrosine	6,86	1,40 [1,06-1,85]	0,02
Pic non défini	6,94	1,35 [1,06-1,72]	0,02
Pic non défini	6,97	1,45 [1,10-1,92]	0,01
Pic non défini	7,21	1,76 [1,25-2,48]	0,001
Pic non défini	7,27	1,58 [1,18-2,12]	0,002
Phénylalanine	7,28	1,39 [1,08-1,79]	0,01
3-methylhistidine	7,54	1,37 [1,05-1,80]	0,02
Histidine	7,69	1,46 [1,12-1,88]	0,004
Pic non défini	7,83	0,73 [0,56-0,95]	0,02
Pic non défini	7,87	0,70 [0,54-0,92]	0,01
Pic non défini	7,91	0,68 [0,51-0,90]	0,007
Pic non défini	7,92	0,76 [0,59-0,98]	0,03
Pic non défini	7,98	0,66 [0,50-0,86]	0,002

OR, odds ratio ; IC, intervalle de confiance

<sup>1</sup> Modèles de régression logistique conditionnelle sur les signaux RMN en continu divisés par l'écart type ajustés sur l'âge, l'IMC, le statut tabagique, la taille, la consommation d'alcool, l'activité physique, le niveau d'éducation, les antécédents familiaux de cancer de la prostate. Des tests de tendance linéaire ont été réalisés.

<sup>2</sup> Assignations multiples pour le même signal RMN dû à la superposition spectrale

Les hommes ayant des taux plasmatiques plus élevés en valine, glutamine, tyrosine, créatine, albumine lysyl, phénylalanine, histidine, albumine, 3-méthylhistidine et un taux plus faible en urée avaient un plus haut risque de développer un cancer de la prostate au cours du suivi. Parmi ces ions, l'albumine, l'histidine et la phénylalanine atteignaient un FDR de 0.2 ; 0.3 pour les autres, sauf la glutamine (0.4). Les corrélations entre les signaux et les résultats des régressions logistiques pour les signaux codés en tertiles sont présentés en Annexe 4 et Annexe 5 respectivement. Concernant ces derniers, deux associations négatives supplémentaires ont été observées : des taux plasmatiques plus faibles en citrate et lipoprotéines (groupement CH<sub>2</sub>) étaient associés à une augmentation de risque de cancer de la prostate. Les acides aminés et dérivés étaient positivement corrélés entre eux, ainsi qu'avec l'albumine. Le taux d'urée était très corrélé négativement avec ceux de la glutamine et de l'histidine ( $r < -0.35$ ).

Concernant les analyses de sensibilité, parmi les signaux identifiés, tous étaient toujours associés au risque de cancer de la prostate sauf la glutamine après exclusion des cas diagnostiqués pendant la première année de suivi. Tous les métabolites identifiés toujours associés ( $p < 0.05$ ) au risque de cancer de la prostate après exclusion des cas ayant un dysfonctionnement prostatique à l'inclusion (taux de PSA  $\geq 4 \mu\text{g/l}$ ) sauf l'urée, la glutamine et la tyrosine ( $p = 0.05$  pour ce dernier). Par ailleurs, l'albumine-lysyl, un signal regroupant à la fois la créatine et l'albumine-lysyl (superposition de signaux), l'urée et plusieurs régions non identifiées étaient associés spécifiquement au cancer de la prostate de bas grade (Gleason  $< 7$ ) et une région non identifiée (à 7.26 en CPMG) était associée spécifiquement au cancer de la prostate de haut grade (Gleason  $\geq 7$ ). Aucune interaction n'a été observée entre le risque de cancer de la prostate et la durée de suivi.

Les trois premières composantes de l'ACP permettaient d'expliquer 60% de la variabilité totale. La première composante était caractérisée par des taux plasmatiques élevés en albumine, phénylalanine, 3-méthylhistidine, histidine, tyrosine, créatine et albumine lysyl et un taux plus faible en urée. La deuxième composante était caractérisée par un taux plus élevé en urée et des taux plus faibles en tyrosine, histidine et glutamine. La troisième composante était caractérisée par des taux plus élevés en valine, albumine, tyrosine, phénylalanine, 3-méthylhistidine et histidine. Deux de ces profils de métabolites (composante 1 et 3) étaient associés à une augmentation du risque de développer un cancer de la prostate et la composante 2 à une diminution du risque (résultats présentés dans le Tableau 16).

Tableau 16 Associations entre les profils métabolomiques obtenus par ACP et le risque de cancer de la prostate<sup>1</sup> – cohorte SU.VI.MAX, 1994-2007

Composantes <sup>2</sup>	cas/témoins <sup>3</sup>	OR [IC95%]	P
Composante 1	157/157	1,33 [1,01-1,75]	0,04
Composante 2	157/157	0,74 [0,56-0,96]	0,02
Composante 3	157/157	1,51 [1,14-1,98]	0,003

OR, odds ratio ; IC, intervalle de confiance

<sup>1</sup> Modèles de régression logistique conditionnelle sur les signaux RMN en continu divisés par l'écart type ajustés sur l'âge, l'IMC, le statut tabagique, la taille, la consommation d'alcool, l'activité physique, le niveau d'éducation, les antécédents familiaux de cancer de la prostate. Des tests de tendance linéaire ont été réalisés

<sup>2</sup> La première composante était caractérisée par des taux plasmatiques élevés en albumine, phénylalanine, 3-méthylhistidine, histidine, tyrosine, créatine et albumin lysyl et un taux plus faible en urée. La deuxième composante était caractérisée par un taux

plus élevé en urée et des taux plus faibles en tyrosine, histidine et glutamine. La troisième composante était caractérisée par des taux plus élevés en valine, albumine, tyrosine, phénylalanine, 3-méthylhistidine et histidine (poids des variables  $\geq |0.25|$ ). De plus, plusieurs signaux non identifiés caractérisaient ces composantes.

<sup>3</sup> Seules les paires d'individus ayant des données métabolomiques disponibles dans les deux séquences ont pu être prises en compte dans cette analyse.

### III. Discussion

Dans cette première partie, les analyses en H-RMN ont permis de mettre en évidence un large éventail de métabolites endogènes (synthétisables par l'organisme) comme certains acides aminés et dérivés, le glucose, des lipides ou entre le citrate, associés au risque de cancer du sein ou de la prostate. Les analyses en LC-MS ont mis en évidence des acides aminés et dérivés associés au risque de cancer du sein dont deux communs avec la H-RMN (la valine et la glutamine) mais aussi, grâce à sensibilité élevée, un métabolite provenant de l'environnement et non synthétisable par l'organisme (l'ATBC), un métabolite de la famille des hormones stéroïdiennes (le pregnene-triol sulfate) ainsi que des métabolites impliqués dans le métabolisme du microbiote (l'O-succinyl-homosérine et l'acide 5-amino-valérique). La RMN, ayant une sensibilité parfois limitée, n'a pas permis d'identifier de manière précise les métabolites lipidiques rendant les interprétations et comparaisons avec la littérature difficile. De plus, la superposition de certains signaux (lipides et glycoprotéines ; lipides et acétone ; lysine et arginine ; lysine, créatine et créatinine ; albumine lysyl et créatine) ne permet pas de distinguer à quel(s) métabolite(s) l'association observée doit être attribuée, si ces métabolites agissent en synergie ou si les signaux de certains atténuent les associations des autres. Ces résultats doivent donc être interprétés avec précaution et être validés dans d'autres études permettant de distinguer les signaux de ces métabolites. Cependant, il faut noter que la RMN est une technique robuste, stable et reproductible (266).

Nos travaux ont permis de mettre en évidence plusieurs métabolites plasmatiques associés positivement au risque de cancer du sein (valine, lysine, arginine, glutamine, créatine, créatinine, acide 5-aminovalérique, phénylalanine, tryptophane, GGT, ATBC et pregnene-triol sulfate) ou associés négativement au risque (O-succinyl-homoserine, lipoprotéines, lipides dont insaturés, glycoprotéines, acétone, composés dérivés du glycérol). De plus, la plupart des métabolites (surtout en LC-MS) étaient associés au risque de cancer du sein hormono-sensible (ER+ ou PR+). Cependant, très peu d'associations ont été détectées pour le risque de cancers du sein ER- et PR-. Toutefois, étant donné le faible effectif de ces derniers dans notre population, il est possible que l'on ait manqué de puissance pour les détecter.

Nous avons également observé que les hommes ayant des taux plasmatiques plus élevés en valine, glutamine, créatine, albumine lysyl, tyrosine, phénylalanine, histidine, albumine, 3-méthylhistidine et des taux plus faibles en urée, citrate et lipoprotéines avaient plus de risque de développer un cancer de la prostate au cours du suivi. De plus, quelques métabolites (albumine lysyl, créatine et urée) étaient associés au risque de cancer de la prostate de bas grade alors que seul un signal qui n'a pas pu être identifié était associé au risque de haut grade. Il est, là encore, possible que nous ayons manqué de puissance.

Certaines associations n'étaient plus observées lorsque l'on excluait les cas de cancer du sein ou de la prostate diagnostiqués lors des premières années de suivi ou encore lorsque l'on excluait les cas de cancer de la prostate qui avaient un dysfonctionnement prostatique à l'inclusion (taux de PSA  $\geq 4\mu\text{g/l}$ ). Pour les variations des quelques métabolites concernés, ceci pourrait suggérer un effet de causalité inverse où par exemple, une tumeur déjà présente induirait les variations métaboliques observées.

Par ailleurs, on remarque dans notre étude que la créatine, la valine, la glutamine et la phénylalanine étaient associées positivement à la fois au risque de cancer du sein et de la prostate. Ceci suggère que ces métabolites pourraient ne pas être spécifiques du risque de l'un ou l'autre de ces cancers, mais au contraire être impliqués dans des processus communs à la cancérogenèse pour les deux localisations. De la même manière, certains des métabolites que nous avons observés comme associés au risque de cancer du sein ou de la prostate dans notre étude ont également été associés positivement ou négativement au risque d'autres cancers grâce à la métabolomique. L'acétone a été identifiée dans deux analyses RMN comme associée négativement au risque de carcinome hépatocellulaire (HCC) (195,196). D'autres analyses métabolomiques ont montré que des augmentations de la valine (195), de la glutamine (192,195,197), de la lysine (197) étaient associées à une diminution du risque de HCC ou de cancer des voies biliaires alors que la phénylalanine était associée à une augmentation du risque de HCC de manière assez robuste dans 4 études (192,195–197). Le signal N-acetyl des glycoprotéines a également été inversement associé au risque de carcinome hépatocellulaire (HCC) en RMN (195). La lysophosphatidylcholine (18:0) et la phosphatidylcholine (30:0) (deux composés dérivés du glycérol) ont été associées positivement et négativement respectivement, avec le risque de cancer du sein, de la prostate et colorectal dans une étude utilisant la spectrométrie de masse ciblée (206). La tyrosine et la glutamine ont été associées à une diminution du risque de cancer pancréatique (191) et hépatocellulaire (197) respectivement. La tyrosine a également été associée positivement au risque de HCC (197).

Les profils de métabolites (ou combinaisons de plusieurs métabolites) permettent d'améliorer la spécificité des signatures du risque à un seul type de cancer. Dans les analyses détaillées au-dessus les profils de métabolites ont été réalisés avec des ACP sur les métabolites qui étaient associés de manière individuelle (résultats des régressions logistiques). Les analyses multivariées réalisées sur l'ensemble des métabolites par les plateformes de métabolomique n'ont pas permis d'identifier des profils permettant de distinguer le statut (cas/témoin) des participants. L'approche non ciblée apporte une très grande quantité d'information dans laquelle il peut parfois être difficile de mettre en évidence les informations pertinentes pour les paramètres étudiés. Dans notre étude, les profils de métabolites étaient associés avec le risque de cancers du sein et de la prostate dans le même sens que lorsqu'ils étaient analysés de manière individuelle.

Les résultats que nous avons obtenus peuvent être comparés aux autres études de la littérature qui avaient un objectif similaire. Concernant le cancer du sein, dans l'étude de Bro et al. (203), la signature métabolomique identifiée en lien avec le risque de cancer du sein incluait le glycéryl et le cholestérol, qui étaient inversement associés au risque, en cohérence avec nos résultats. Le HDL-cholestérol dosé avec des techniques non métabolomiques a été associé à une diminution du risque de cancer du sein dans plusieurs études (267,268) et notamment chez les femmes préménopausées (269). Nous avons réalisé des investigations supplémentaires sur les dosages biologiques des lipoprotéines à l'inclusion, qui ont montré que le signal des lipoprotéines en CPMG (à 1.258 ppm) était très corrélé positivement au dosage des triglycérides sanguins ( $r=0.79$ ) et dans une moindre mesure aux taux sanguins de cholestérol ( $r=0.24$ ), de l'apolipoprotéine B ( $r=0.20$ ) puis de l'apolipoprotéine A1 ( $r=0.16$ ) mais pas de la lipoprotéine(a) ( $r=0$ ). Les signaux des lipoprotéines en NOESY1D (à 0.86 et 1.27 ppm) étaient très corrélés positivement aux taux sanguins des triglycérides ( $r=0.52$  et  $0.73$  respectivement), du cholestérol ( $r=0.51$  et  $0.34$  respectivement), de l'apolipoprotéine A1 ( $r=0.45$  et  $0.28$  respectivement) puis de l'apolipoprotéine B ( $r=0.34$  et  $0.25$  respectivement) mais pas de la lipoprotéine(a) ( $r=0.03$  et  $0$  respectivement). Par ailleurs, dans une étude similaire, la valine a été associée positivement à l'IMC et au risque de cancer du sein ER+ chez les femmes postménopausées en spectrométrie de masse (204). La créatine et la créatinine faisaient également partie

de la signature prédictive du risque de cancer du sein de l'étude de Bro et al. (203). Dans cette dernière, la superposition de signaux de la créatine et la créatinine était faiblement associée de manière inverse alors que la créatine seule était fortement associée de manière directe. Dans notre étude, les signaux de la lysine, de la créatine et de la créatinine se superposaient, il n'était donc pas possible de distinguer les effets. Des augmentations du taux sérique de trois  $\gamma$ -glutamyl dipeptides ont été associées positivement à l'IMC et au risque de cancer du sein ER+ chez les femmes postménopausées (204). Deux études similaires ont montré des associations positives entre des hormones stéroïdiennes (la famille de composés à laquelle appartient le pregnene-triol sulfate) et le risque de cancer du sein chez les femmes postménopausées (204,205). Une méta-analyse étudiant les associations entre le taux de glucose sanguin chez des femmes non diabétiques et le risque de cancer du sein hors métabolomique a montré une élévation modérée du risque associée à une augmentation du taux de glucose (270).

Concernant le cancer de la prostate, à notre connaissance, 4 études (213–216) ont investigué les associations entre métabolites issus d'analyses métabolomiques et risque de cancer de la prostate. De manière cohérente avec nos résultats, le citrate (214) et le cycle de l'urée (215) ont été associés à une diminution de risque de la prostate alors que la 3-méthylhistidine et la phénylalanine ont été associées à une augmentation de risque (213,214). Cependant, le citrate a également été associé à une augmentation de risque dans une autre étude (215). Hors métabolomique, une étude cas-cohorte (554 cas de cancer de la prostate et 2739 témoins) s'intéressant au taux plasmatique d'albumine et au risque de cancer de la prostate n'a pas montré d'association significative (271). Les associations entre lipoprotéines circulantes et risque de cancer de la prostate ont été plus étudiées dans la littérature. Dans une étude cas-cohorte, la lipoprotéine A circulante a été associée à une augmentation de risque de cancer de la prostate (272) et une revue de la littérature a montré une association positive avec le cholestérol sérique total (273). Dans une population de fumeurs, un taux élevé de cholestérol sérique total était associé à un risque plus élevé de cancer avancé de la prostate (412 cas), et un taux élevé de cholestérol HDL était associé à une réduction du risque de cancer de la prostate (2041 cas) (274). Une autre étude sur 2008 cas de cancer de la prostate a montré qu'un taux sanguin plus faible en HDL et apolipoprotéine A diminuerait le risque (275). En revanche, une méta-analyse publiée en 2015 n'a montré aucune association entre taux sanguins de HDL et LDL et risque de cancer de la prostate (276). Les résultats sont donc encore mitigés. Dans notre étude, des investigations supplémentaires sur les dosages biologiques à l'inclusion ont montré que le signal des lipoprotéines était très corrélé positivement aux taux sanguins des triglycérides ( $r=0.93$ ) et de l'apolipoprotéine B ( $r=0.39$ ) et dans une moindre mesure au taux de cholestérol total ( $r=0.24$ ). Il était également corrélé négativement au taux sanguin d'apolipoprotéine A1 ( $r=-0.31$ ) mais n'était pas corrélé au taux de lipoprotéine(a).

Il faut noter que certaines associations observées dans les autres études ayant un design similaire n'ont pas été répliquées dans la nôtre. Plusieurs raisons peuvent expliquer cela, comme par exemple, la différence de population étudiée et notamment les modes de vie, la variabilité des techniques analytiques et couverture des classes de métabolites détectées ou encore un possible manque de puissance. Etant donné le nombre restreint d'études ayant utilisé la métabolomique pour identifier des métabolites associés au risque de cancer du sein ou de la prostate, la comparaison avec nos résultats est limitée. Au-delà même de la métabolomique, le dosage de métabolites n'a pas encore été si largement utilisé pour étudier ce type d'associations. D'autres études sont donc encore nécessaires pour la validation de nos résultats.

Les lipides, les acides aminés et le glucose sont des métabolites qui peuvent être apportés par l'alimentation. Cependant, les variations observées ne proviennent probablement pas d'une prise alimentaire directe, les individus étant à jeun lors du prélèvement. Notamment, la prise alimentaire impacte le taux d'acides aminés circulants dans les quelques heures seulement qui suivent la prise alimentaire. Les variations observées dans notre étude proviennent plus probablement d'une modification métabolique dont l'origine ne peut être élucidée ici. Un impact de l'alimentation à plus long terme pourrait expliquer ces perturbations, mais d'autres origines sont possibles. Le tryptophane, la lysine, la phénylalanine, la valine et l'histidine sont des acides aminés essentiels ne pouvant être apportés que par l'alimentation. La lysine est strictement indispensable car elle n'est pas soumise à la transamination, ce qui rend sa désamination irréversible. La tyrosine un acide aminé semi-essentiel qui peut être synthétisé par les humains mais seulement à partir de la phénylalanine. Par ailleurs, la 3-méthyl-histidine est un biomarqueur candidat de la consommation de produits animaux (51).

La plupart des métabolites concernés par les variations observées ne sont pas spécifiques d'une voie métabolique et sont impliqués dans de nombreux processus endogènes. Des facteurs exogènes ou des processus endogènes pourraient expliquer ces perturbations. L'interprétation mécanistique de ces résultats issus d'études d'observation comme les nôtres est donc délicate et de nombreuses hypothèses sont possibles. Quelques-unes de ces hypothèses sont exposées ci-dessous et sont résumées sous forme de schéma en Figure 20.

Plusieurs résultats sembleraient suggérer une perturbation du métabolisme énergétique chez les femmes et chez les hommes qui ont développé un cancer du sein ou de la prostate au cours du suivi. Une diminution des taux de lipides, lipoprotéines et de composés dérivés du glycérol, plutôt associée au risque de cancer du sein que de la prostate dans notre étude, pourraient faire penser à une augmentation de la  $\beta$ -oxydation. Parallèlement, une augmentation du taux de lysine (un acide aminé cétogène) pourrait être liée une augmentation de la cétogenèse. Les lipides et les corps cétoniques sont des substrats énergétiques alternatifs au glucose. Une amplification de la  $\beta$ -oxydation et de la cétogenèse pourrait entraîner une accumulation du glucose dans le sang et refléter un schéma similaire à l'insulino-résistance. De plus, la glutamine et la valine (un des acides aminés à chaîne ramifiée ; BCAAs) sont deux acides aminés glycoformateurs. Les BCAA ont été associés au risque de maladies cardiovasculaires (277), d'insulino-résistance (278,279) et de diabète (280). Les acides aminés aromatiques (AAA, phénylalanine, tyrosine, tryptophane, histidine) ont également été associés au développement de l'insulino-résistance chez les jeunes et les adultes ayant une glycémie normale avec des associations plus prononcées chez les hommes (279). Les voies liées à la résistance à l'insuline sont un des nombreux mécanismes pouvant expliquer les relations entre obésité et risque de cancer (15). Dans nos études, les femmes étaient majoritairement de poids normal alors que les hommes étaient plutôt de poids normal ou en surpoids. Par ailleurs, le citrate était associé à une diminution de risque de cancer de la prostate dans notre étude. La quantité de citrate est présente de manière significative dans une grande variété de fruits et de légumes, notamment dans les agrumes. Le citrate est souvent utilisé comme exhausteur de goût et régulateur alimentaire de pH. C'est également un métabolite endogène, intermédiaire du cycle de Krebs et un acteur clé du métabolisme prostatique (281). En effet, le citrate se trouve concentré dans l'épithélium glandulaire de la zone périphérique normale de la prostate du fait de l'inhibition en présence d'ions zinc de l'aconitase mitochondriale (conversion du citrate en isocitrate). Le citrate est sécrété en forte concentration dans le fluide prostatique, probablement pour maintenir l'acidité ou pour fournir une source d'énergie pour la motilité du sperme lors de la fécondation des ovocytes (281). Dans les tissus impactés par le cancer de la prostate, la concentration en citrate est diminuée de manière drastique probablement due à une

amplification de l'oxydation du citrate dans le cycle Krebs ayant pour conséquence une production accrue d'ATP permettant l'anabolisme et la prolifération dans les cellules tumorales (281). De plus, une étude a montré une inhibition de la croissance tumorale par le citrate dans divers types de tumeurs et selon différents mécanismes comme l'inhibition de la glycolyse et du cycle de Krebs (282).

Les résultats de nos études pourraient suggérer une perturbation du métabolisme des acides aminés et protéique, notamment chez les hommes ayant développé un cancer de la prostate au cours du suivi. La créatinine est un catabolite résultant de la dégradation de la créatine, éliminée en majeure partie par voie urinaire. Le taux sanguin de créatinine dépend de la capacité d'élimination du rein et de la masse musculaire ; son évaluation ainsi que celle de l'urée donne une indication de la capacité de filtration glomérulaire (283). Par ailleurs, la 3-méthylhistidine serait un marqueur urinaire du catabolisme musculaire à condition que l'apport alimentaire d'origine carnée soit limité (284). La glutamine plasmatique (une molécule d'ammoniac NH<sub>3</sub> fixée au glutamate) permet, entre autres, le transport de l'azote produit en périphérie vers le foie pour être éliminé par le cycle de l'urée. Dans notre étude chez les hommes, le taux de glutamine plasmatique était corrélé négativement aux taux d'urée. L'identification des taux d'ornithine et de citrulline - deux métabolites clés dans le cycle de l'urée - pourrait renseigner sur la fonctionnalité de ce cycle. Cependant, ces résultats semblent suggérer un dysfonctionnement de celui-ci entraînant une diminution de la production d'urée et une accumulation de glutamine en amont. L'albumine est la protéine sanguine la plus abondante chez les mammifères. Elle est produite par le foie chez ces êtres vivants et se trouve dans le lait et les muscles. L'albumine est un acteur majeur de la régulation de la pression oncotique. La présence d'albumine-lysyl pourrait suggérer une dérégulation de la dégradation de l'albumine et/ou une augmentation globale du turn-over de l'albumine impliquant à la fois sa synthèse et sa dégradation.

Par ailleurs, certaines variations observées pourraient suggérer une mise en place de conditions favorables à la transformation et prolifération cellulaires chez les femmes et chez les hommes ayant développé un cancer du sein ou de la prostate pendant le suivi. La valine et la glutamine, associées positivement avec le risque de cancers du sein et de la prostate dans notre étude, sont des régulateurs de nombreuses voies de signalisation cellulaire, y compris des activateurs de la voie mTOR, l'une des principales voies de prolifération cellulaire (280). En outre, la glutamine n'était plus associée au risque de cancer de la prostate lorsque l'on excluait les cas ayant un dysfonctionnement prostatique au moment du prélèvement, elle pourrait donc être le reflet d'une activité tumorale émergente ou d'une perturbation de son métabolisme suite à ce dysfonctionnement. L'ATBC, associé à une augmentation de risque de cancer du sein dans nos travaux, est un plastifiant alternatif aux phtalates (285) largement utilisé dans des produits comme les emballages alimentaires, les jouets en vinyle et les excipients pharmaceutiques. Bien que ce contaminant ne semble pas s'accumuler dans les tissus (286), des données récentes ont révélé une activité biologique potentielle sur la croissance tissulaire (287). Certains composés alternatifs aux phtalates pourraient interagir avec la signalisation PPAR impliquée dans la prolifération cellulaire et l'inflammation (288). La créatine, associée positivement avec le risque de cancers du sein et de la prostate dans notre étude et considérée comme un donneur de méthyle, peut jouer un rôle dans la méthylation de l'oncoprotéine PP2A qui active la prolifération cellulaire et les voies de transformation néoplasiques (289). La glutamine est un donneur d'azote (NH<sub>4</sub><sup>+</sup>), son élévation associée à une diminution de l'urée pourrait suggérer une redistribution de l'azote à des fins de synthèse. Le pregnene-triol sulfate est une hormone sexuelle stéroïdienne appartenant à la famille des progestatifs et était associé positivement au risque de cancer du sein dans notre étude. Des analyses complémentaires ont montré un taux diminué de ce stéroïde chez les personnes prenant des contraceptifs oraux à l'inclusion dans l'étude. Un ajustement supplémentaire dans



les modèles sur cette variable diminuait légèrement la force de l'association du pregnene-triol sulfate avec le risque de cancer du sein mais il restait associé. On pourrait donc supposer qu'une partie de cette variation observée pourrait être influencée par une imprégnation hormonale à l'inclusion (association inverse), cependant elle pourrait également provenir d'un processus endogène ou d'un autre facteur exogène non mesuré dans cette étude. Dans la littérature, les hormones stéroïdiennes circulantes sont associées à une augmentation du risque de cancer du sein chez les femmes post-ménopausées (91). La progestérone et le récepteur à progestérone (RP) agissent avec les protéines kinases mitogènes et les médiateurs du cycle cellulaire pour induire la régulation des gènes cibles des RP et des réponses prolifératives (290). Par ailleurs, il a été montré que l'ATBC activerait le pregnane X receptor (PXR), également connu sous le nom de « steroid and xenobiotic sensing nuclear receptor » (SXR) chez l'humain et le rat et pourrait modifier le métabolisme des hormones stéroïdes endogènes (285).

Il a été montré que dans les cellules mononuclées du sang périphérique (PBMC) cultivées, l'ajout de 10 mmol/L de BCAA augmentait la production d'espèces réactives de l'oxygène via la NADPH oxydase et les mitochondries, et activait la signalisation Akt-mTOR (291). Les BCAA stimulaient également l'activation du facteur de transcription sensible au potentiel redox NF- $\kappa$ B, entraînant la libération de molécules pro-inflammatoires comme l'interleukine-6, le facteur de nécrose tumorale- $\alpha$ , la molécule d'adhésion intracellulaire-1 ou CD40L. Les conclusions de cette étude étaient que des concentrations élevées de BCAA pourraient exercer des effets délétères sur les cellules sanguines et contribuer ainsi à l'état pro-inflammatoire et oxydatif. Chez la souris, des taux urinaires plus élevés d'acide 5-aminovalérique ont été associés à une inflammation intestinale (292). Des données sur le rapport kynurénine/tryptophane seraient utiles pour étayer l'hypothèse d'un état pro-inflammatoire, car il reflète l'expression du gène IDO1, augmentant dans les situations d'inflammation chronique de bas grade. Le métabolisme humain peut produire de la cystéine (un composant du glutathion, molécule antioxydante) en utilisant l'O-succinyl-homosérine au lieu de l'homocystéine (293). Les  $\gamma$ -glutamyl-dipeptides peuvent également être impliqués dans la synthèse de glutathion. Ces données de la littérature pourraient suggérer un processus inflammatoire et un état de stress oxydatif chez les femmes ayant développé un cancer du sein pendant le suivi.

Enfin, nos résultats pourraient également refléter une influence du microbiote intestinal dans les perturbations observées chez les femmes ayant déclaré un cancer du sein au cours du suivi, puisque la synthèse de l'O-succinyl-homosérine est catalysée par l'homosérine O-succinyltransferase, une classe d'enzyme non présente chez l'humain mais essentielle à la survie des bactéries, des plantes et des champignons (294). D'autre part, l'acide 5-amino-valérique, un composé dérivé des acides gras à courte chaîne a été suggéré comme métabolite du co-métabolisme microbiotique (295).

Les résultats de nos études d'observation ne permettent pas de tirer de conclusions fermes en termes d'interprétations mécanistes. L'élucidation de celles-ci nécessiterait davantage d'études, notamment *in vivo* et *in vitro*. La plupart des métabolites concernés par les variations observées sont au croisement de nombreuses voies métaboliques. Sans quantification absolue et d'information sur les ratios de certains métabolites, il est difficile d'interpréter ces observations qui reflètent des flux de métabolites en constante variation. Néanmoins, ces résultats semblent suggérer des perturbations métaboliques précédant le diagnostic des cancers du sein et de la prostate jusqu'à 13 ans avant. Les hypothèses évoquées ci-dessus sembleraient donc refléter une perturbation du métabolisme énergétique et des acides aminés, une mise en place de conditions favorisant la prolifération et croissance cellulaire ainsi qu'un processus inflammatoire et de stress oxydatif chez les femmes ou les hommes ayant développé un cancer du sein ou

de la prostate au cours du suivi. Nos observations suggèrent également un rôle du microbiote et de l'exposition environnementale dans ces perturbations métaboliques pour le cancer du sein (voir Figure 20). De nombreuses autres hypothèses peuvent être envisagées et des variations de métabolites plus spécifiques de ces voies métaboliques nous permettraient de valider ou non ces hypothèses.

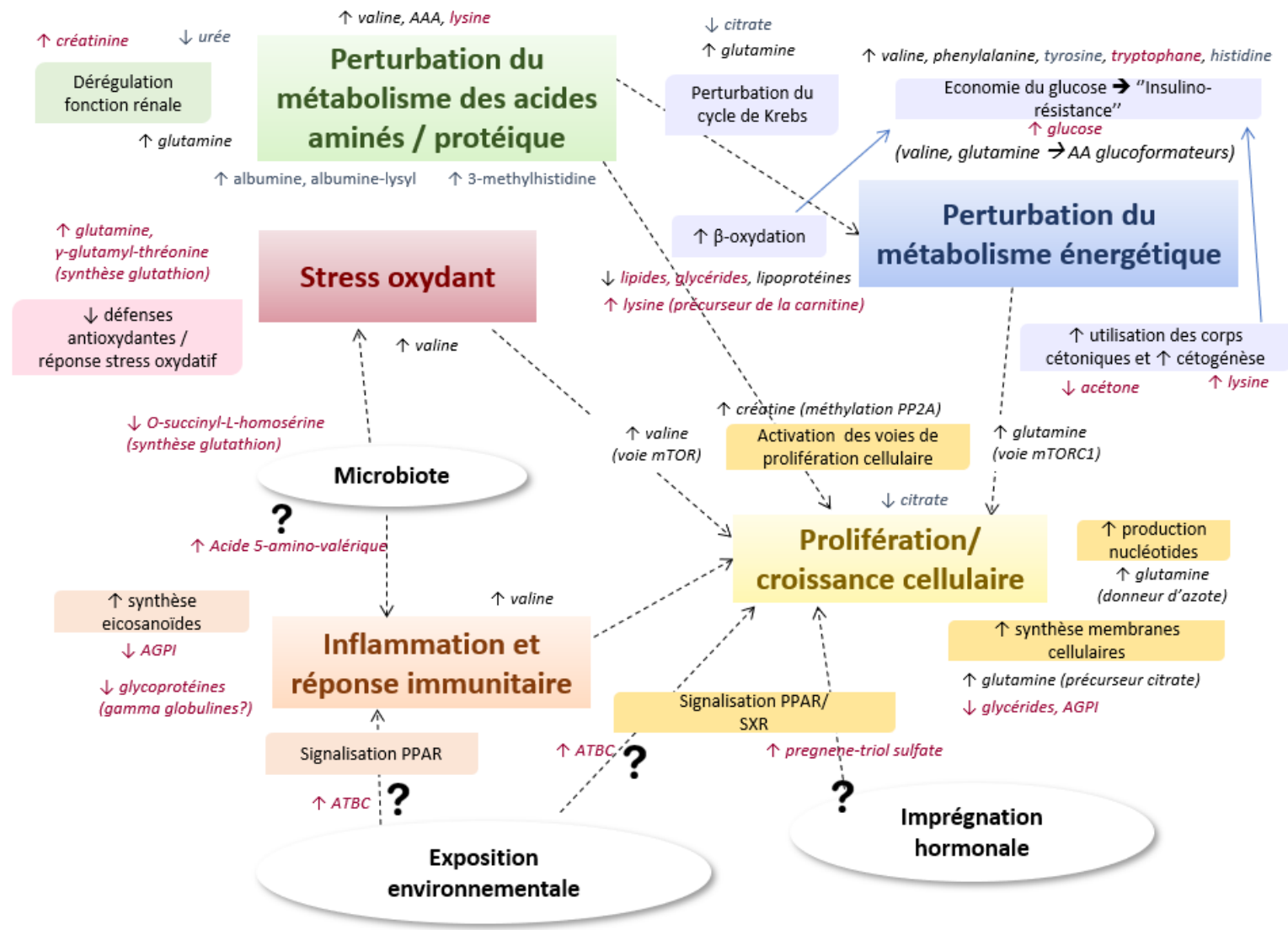


Figure 20 Hypothèses mécanistiques des variations métaboliques observées chez les femmes ayant développé un cancer du sein (métabolites en rouge) et les hommes ayant développé un cancer de la prostate (métabolites en bleu) au cours du suivi. Métabolites en noir : variations observées dans les deux cas.

## Objectif 2: Identifier des signatures métabolomiques plasmatiques associées à l'alimentation dans un échantillon de femmes (objectif 2.I.) et d'hommes (objectif 2.II.) de la cohorte SU.VI.MAX

Ce deuxième objectif visait à explorer les données de métabolomique à partir de prélèvements plasmatiques afin de mettre en évidence de potentiels biomarqueurs de l'exposition alimentaire ou reflétant l'impact de l'alimentation usuelle globale sur le métabolisme. Là encore, ces potentiels biomarqueurs peuvent être des biomarqueurs d'exposition (notamment nutritionnelle), d'effet (impact de la nutrition sur le métabolisme par exemple) ou de susceptibilité. Ce type de métabolites pourrait permettre de compléter l'évaluation de l'exposition nutritionnelle dans les études épidémiologiques et également permettre de mieux comprendre les relations nutrition/santé.

Afin de pouvoir répondre à ces deux types de problématique, nous nous sommes intéressés aux habitudes alimentaires globales (régimes/profils alimentaires) puisque les multiples des composants alimentaires peuvent agir en synergie et avoir un impact sur la santé (74,75). Dans un but secondaire, nous avons également recherché de potentiels biomarqueurs de consommation d'aliments ou de groupes d'aliments plus spécifiques. Cet objectif secondaire a été réalisé uniquement en LC-MS, une méthode plus sensible (donc uniquement chez les femmes).

### I. Signature de l'alimentation globale chez les femmes

#### A. Analyses en LC-MS

Cette étude a fait l'objet d'un article actuellement soumis pour publication, présenté en Annexe 6.

*Lécuyer L\*, Dalle C\*, Micheau P, Pétéra M, Centeno D, Lyan B, Lagree M, Galan P, Herberg S, Rossary A, Demidem A, Vasson MP, Partula V, Deschasaux M, Srour B, Latino-Martel P, Druetne-Pecollo N, Kesse-Guyot E, Durand S, Pujos-Guillot E, Manach C#, Touvier M#. \*,#: contributions égales. **Untargeted plasma metabolomic profiles associated with overall diet in women from the SU.VI.MAX cohort.***

#### 1. Méthodes

La sélection de la population d'étude pour ces analyses a été détaillée dans la partie Méthologie générale (III.B.1.). Il s'agit d'une étude transversale dans la population SU.VI.MAX incluant 80 femmes du 10<sup>ème</sup> décile du mPNS-GS (ayant une bonne adéquation aux recommandations) et 80 femmes du 1<sup>er</sup> décile appariées (ayant une faible adéquation aux recommandations). L'adéquation aux recommandations nutritionnelles permet d'estimer la qualité globale de l'alimentation puisque ces recommandations couvrent diverses composantes de l'alimentation. Ce design a permis de sélectionner deux groupes de femmes ayant des habitudes de consommation très différentes.

Les échantillons de plasma des individus prélevés à l'inclusion ont été analysés en UPLC-MS en mode positif et négatif. Le détail de ces analyses est présenté dans la partie Méthodologie générale (IV.B.2.). Les méthodes spécifiques à cet objectif sont détaillées ici. Deux protocoles complémentaires de préparation des échantillons plasmatiques ont été réalisés. Le premier (le même que celui utilisé pour l'objectif 1) a été réalisé selon la procédure modifiée décrite par Pereira et al (265) et est plutôt adapté à la détection de métabolites endogènes (appelé « protocole endogène » par la suite par souci de simplicité). Le deuxième a été développé afin de favoriser la mise en évidence de métabolites du « food metabolome »

en précipitant les protéines avec du méthanol acidifié (appelé « protocole exogène » par la suite). Pour le premier protocole, 100 µL d'échantillons de plasma ont été déprotéinés à l'aide de méthanol froid. Après centrifugation et évaporation, les résidus déshydratés ont été redissous dans une préparation 50/50 (v/v) d'acétonitrile/eau contenant 0,1% d'acide formique. Pour le deuxième protocole, 100 µL d'échantillons de plasma ont été déprotéinés à l'aide de méthanol contenant 0.1% d'acide formique pour casser les liaisons hydrogènes entre les métabolites et les protéines plasmatiques. La préparation a ensuite été centrifugée et 250 µL d'acétonitrile ont été ajoutés pour collecter le surnageant. 5µL d'échantillon ont été injectés dans un système U3000 UHPLC (Thermo Scientific, France). Le débit a été réglé à 0,4 mL/min et 30°C. Les phases mobiles A et B étaient l'eau milliQ et l'acétonitrile respectivement, contenant chacune 0,1 % d'acide formique. Le gradient d'élution a été réalisé comme suit : 100% A (0-2 min) ; gradient linéaire à 100% B (2-15 min) ; 100% B (15-22 min) ; 100 % à 0 % B (22.1 min) ; 100% A (22.1-26 min pour rééquilibrage). Des contrôles qualité ont été préparés en mixant 10 µL de tous les échantillons de plasma extraits. Les données de quatre échantillons issus du protocole exogène en mode d'ionisation positif ont été exclues en raison de la faible intensité anormale de tous les ions. Pour cette étude, les ions présents dans au moins 40 % des échantillons du 1<sup>er</sup> ou du 10<sup>ème</sup> décile ont été conservés permettant de traiter la variabilité interindividuelle des habitudes alimentaires (seuil plus élevé pour le premier objectif). Les corrélations analytiques ont été filtrées à l'aide de l'outil " Metabolites Correlation Analysis " (Galaxy Version 1.0.1) avec un seuil de coefficient de corrélation de 0.8. Ces étapes de prétraitements ont permis d'éliminer les ions considérés comme étant du bruit, caractérisés par une variabilité trop élevée, ainsi que d'éliminer la redondance entre les signaux (isotopes, adduits, fragments). Suite à ces prétraitements, 570 et 197 ions (en mode positif et négatif) et 969 et 375 ions (en mode positif et négatif) pour les protocoles endogène et exogène respectivement ont été utilisés pour les analyses statistiques. La plateforme de métabolomique a réalisé des analyses multivariées sur les ions obtenus, détaillées de la partie Méthologie générale (6.).

La population d'étude a été décrite grâce à des modèles de régression logistique conditionnelle. En première intention nous avons estimé les associations entre l'alimentation globale et les profils métabolomiques grâce à des régressions logistiques conditionnelles sur le statut des participantes (appartenance au 1<sup>er</sup> décile vs 10<sup>ème</sup> décile du mPNNS-GS). Cette approche est donc une approche *a priori* puisqu'elle se base sur un score d'adéquation en amont. Des OR pour une augmentation d'un écart type de chaque ion et leur intervalle de confiance à 95% ont été calculés. Les modèles ont été ajustés sur l'âge (en continu), l'IMC (en continu), le statut tabagique (fumeuses/anciennes fumeuses et non fumeuses), la moyenne de l'apport énergétique et le nombre d'enregistrements de 24h validés pendant les 2 ans suivant le prélèvement sanguin. Afin d'étudier les associations avec des composantes de l'alimentation plus précises que l'alimentation globale nous avons également investigué les associations entre les ions et la consommation des aliments représentant les 12 composantes alimentaires du mPNNS-GS (détaillées dans le Tableau 6 dans la partie Méthodologie générale III.B.). Ces associations ont été mises en évidence par des analyses de corrélation de Spearman ajustées sur l'âge, l'IMC, le statut ménopausique, le statut tabagique, la saison de prélèvement sanguin, la moyenne de l'apport énergétique et le nombre d'enregistrements de 24h validés pendant les 2 ans suivant le prélèvement sanguin.

De manière complémentaire, nous avons étudié l'alimentation globale par une approche *a posteriori* en mettant en évidence des profils alimentaires (« patterns ») caractérisant notre population d'étude. Afin de permettre une meilleure réplique de nos résultats comme notre échantillon d'étude était spécifique (80 femmes qui mangent « très bien » et 80 qui mangent « très mal ») nous avons réalisé une ACP avec rotation varimax sur la population générale de SU.VI.MAX ayant des données alimentaires très détaillées pendant les 2 ans suivant l'inclusion (10 enregistrements de 24h ou plus, N=4645). Les trois premières

composantes ont été sélectionnées et des analyses de corrélation de Spearman ont été réalisées pour chaque composante et pour chaque ion, ajustées sur les mêmes facteurs de confusion que précédemment.

La correction de test multiple de Benjamini Hochberg générant les FDR a été appliquée sur l'ensemble de ces analyses.

## 2. Résultats

Le Tableau 17 résume la comparaison des caractéristiques des deux groupes de participantes incluses dans notre étude. Les femmes du 10<sup>e</sup> décile avaient de meilleurs scores d'adéquation aux recommandations pour les composantes fruits et légumes ; produits céréaliers complets ; lait et produits laitiers ; viande, volaille, produits de la mer et œufs ; produits de la mer ; matière grasse totale ajoutée, matière grasse végétale ajoutée ; produits sucrés ; alcool et sel. Les composantes pain, céréales, pommes de terre et légumineuses et boissons sucrées ne permettaient pas de discriminer les deux groupes. Les femmes du 1<sup>er</sup> décile avaient une moins bonne adéquation aux recommandations concernant l'activité physique et avaient une dépense énergétique plus faible et un apport énergétique plus élevé et donc un ratio apport/dépense énergétique plus élevé.

Tableau 17 Caractéristiques<sup>1</sup> à l'inclusion des femmes du 1<sup>er</sup> et du 10<sup>ème</sup> déciles du mPNNS-GS<sup>2</sup> – cohorte SU.VI.MAX, 1994-2007

	Femmes du 1 <sup>er</sup> décile (N=80)	Femmes du 10 <sup>ème</sup> décile (N=80)	p <sup>3</sup>
mPNNS-GS	4,4±0,8	10,1±0,7	Non applicable
Age	48,3±6,7	48±6,4	0,2
IMC (kg/m <sup>2</sup> )	22,2±3,1	22,4±2,7	0,5
IMC (catégories)			Non applicable
<18.5 kg/m <sup>2</sup> (sous-poids)	1 (1,3)	1 (1,3)	
≥18.5 to <25 kg/m <sup>2</sup> (poids normal)	67 (83,8)	67 (83,8)	
≥25 kg/m <sup>2</sup> (surpoids)	12 (15,0)	12 (15,0)	
Statut tabagique			Non applicable
Non fumeuses	58 (72,5)	53 (66,3)	
Anciennes fumeuses	21 (26,3)	26 (32,5)	
Fumeuses	1 (1,3)	1 (1,3)	
Mois de prélèvement sanguin			Non applicable
Mars-Avril	31 (38,8)	31 (38,8)	
Octobre-Novembre	12 (15,0)	12 (15,0)	
Décembre-Janvier-Février	37 (46,3)	37 (46,3)	
Statut ménopausique			Non applicable
Préménopausées	52 (65,0)	52 (65,0)	
Postménopausées	28 (35,0)	28 (35,0)	
Nombre d'enregistrements de 24h	12,1±1,1	12,1±1,2	0,8
Moyenne de l'apport énergétique quotidien (kcal)	2369,5±506,1	1771,2±272,7	<0,0001
<b>Fruits et légumes</b>	0,7 ±0,6	1,5 ±0,5	<0,0001
[0-3.5[ = 0	20 (25,0)	0 (0,0)	0,0005
[3.5-5[ = 0.5	19 (23,8)	5 (6,3)	
[5-7.5[ = 1	33 (41,3)	34 (42,5)	
≥7.5 = 2	8 (10,0)	41 (51,3)	
<b>Pain, céréales, pommes de terre et légumineuses</b>	0,8 ±0,3	0,7 ±0,3	0,5
[0-1[ = 0	1 (1,3)	2 (2,5)	0,7
[1-3[ or ≥6 = 0.5	35 (43,8)	38 (47,5)	
[3-6[ = 1	44 (55,0)	40 (50,0)	
<b>Produits céréaliers complets</b>	0,1 ±0,3	0,4 ±0,4	<0,0001
[0-1/3[ = 0	73 (91,3)	37 (46,3)	0,0003
[1/3-2/3[ = 0.5	2 (2,5)	20 (25,0)	
≥2/3 = 1	5 (6,3)	23 (28,8)	
<b>Lait et produits laitiers</b>	0,3 ±0,4	0,7 ±0,4	<0,0001

[0-1[ ou >3.5 (>4.5 si ≥55 ans) = 0	43 (53,8)	18 (22,5)	<0,0001
[1-2.5[ = 0.5	19 (23,8)	11 (13,8)	
[2.5-3.5] ([2.5-4.5] si ≥55 ans) = 1	18 (22,5)	51 (63,8)	
<b>Viande, volaille, produits de la mer et œufs</b>	0,7 ±0,4	0,9 ±0,2	0,0004
0 ou >2 = 0	20 (25,0)	2 (2,5)	0,007
]0-1[ = 0.5	12 (15,0)	8 (10,0)	
[1-2] = 1	48 (60,0)	70 (87,5)	
<b>Produits de la mer</b>	0,1 ±0,3	0,7 ±0,4	<0,0001
<2 portions/semaine = 0	71 (88,8)	22 (27,5)	<0,0001
≥2 portions/semaine = 1	9 (11,3)	58 (72,5)	
<b>Matière grasse ajoutée totale</b>	0,5 ±0,5	0,9 ±0,4	0,0002
Lipides des matières grasses animales et végétales >16% AE/j = 0	38 (47,5)	12 (15,0)	0,0002
Lipides des matières grasses animales et végétales ≤16% AE/j = 1	42 (52,5)	68 (85,0)	
<b>Matière grasse ajoutée végétale</b>	0,4 ±0,5	0,9 ±0,2	0,0002
Pas de consommation de matière grasse ajoutée végétale ou ratio matière grasse ajoutée végétale / totale ≤0.5 = 0	48 (60,0)	5 (6,3)	0,0002
Pas de consommation de matière grasse ajoutée totale ou ratio matière grasse ajoutée végétale / totale >0.5 = 1	32 (40,0)	75 (93,8)	
<b>Produits sucrés</b>	0,5 ±0,5	1,0 ±0,2	0,0005
Sucre ajouté issu des produits sucrés ≥15% AE/jour = -0.5	3 (3,8)	0 (0,0)	0,003
Sucre ajouté issu des produits sucrés : 10 à 15% AE/jour = 0	34 (42,5)	2 (2,5)	
Sucre ajouté issu des produits sucrés <10% AE/jour	43 (53,8)	78 (97,5)	
<b>Boissons sucrées</b>	0,9 ±0,1	0,9 ±0,1	0,06
<1l eau et >250 ml soda/jour = 0	0 (0,0)	0 (0,0)	0,06
≥1l eau et >250 ml soda/ jour = 0.5	0 (0,0)	0 (0,0)	
<1l eau et ≤250 ml soda/ jour = 0.75	33 (41,3)	21 (26,3)	
≥1l eau et ≤250 ml soda/ jour = 1	47 (58,8)	59 (73,8)	
<b>Alcool</b>	0,5 ±0,4	0,8 ±0,2	<0,0001
Ethanol >20 g/jour pour les femmes et >30g/jour pour les hommes = 0	34 (42,5)	5 (6,3)	0,0005
Ethanol ≤20 g/jour pour les femmes et ≤30g/jour pour les hommes = 0.8	41 (51,3)	66 (82,5)	
Non-consommateurs ou consommateurs irréguliers (<1 verre de vin/semaine) = 1	5 (6,3)	9 (11,3)	
<b>Sel</b>	-0,1 ±0,5	0,6 ±0,5	<0,0001
>12 g/j = -0.5	43 (53,8)	3 (3,8)	<0,0001
]10-12] g/j = 0	16 (20,0)	13 (16,3)	
]8-10] g/j = 0.5	14 (17,5)	28 (35,0)	
]6-8] g/j = 1	5 (6,3)	30 (37,5)	
≤6 g/j = 1.5	2 (2,5)	6 (7,5)	
<b>Activité physique</b>	0,3 ±0,5	0,6 ±0,6	0,002
[0-30[ min/j = 0	62 (77,5)	41 (51,3)	0,007
[30-60[ min/j = 1	10 (12,5)	23 (28,8)	
≥60 min/j = 1.5	8 (10,0)	16 (20,0)	
<b>Métabolisme basal</b>	1314,2 ±80,3	1326,2 ±65,0	0,2
<b>Dépense énergétique (kcal)</b>	2097,9 ±172,8	2167,9 ±163,1	0,007
<b>Ratio apport/dépense énergétique</b>	1,1 ±0,2	0,8 ±0,1	<0,0001

AE, apport énergétique

<sup>1</sup> Les valeurs présentées sont les moyennes ± écart type ou l'effectif (%)

<sup>2</sup> Les composantes du mPNNS-GS sont identifiées en gras

<sup>3</sup> P de tendance linéaire pour la comparaison entre les femmes du 1e décile et du 10e décile du mPNNS-GS grâce aux modèles de régression logistique conditionnelle. Non applicable pour les variables d'appariement sauf si variables plus précises (ex. âge et IMC).

Les résultats des analyses multivariées réalisées au sein de la plateforme de métabolomique n'ont pas permis de distinguer le statut des participants grâce aux ions analysés (ni par ACP, ni par PLS-DA). Un exemple graphique de l'ACP est présenté en Figure 21.

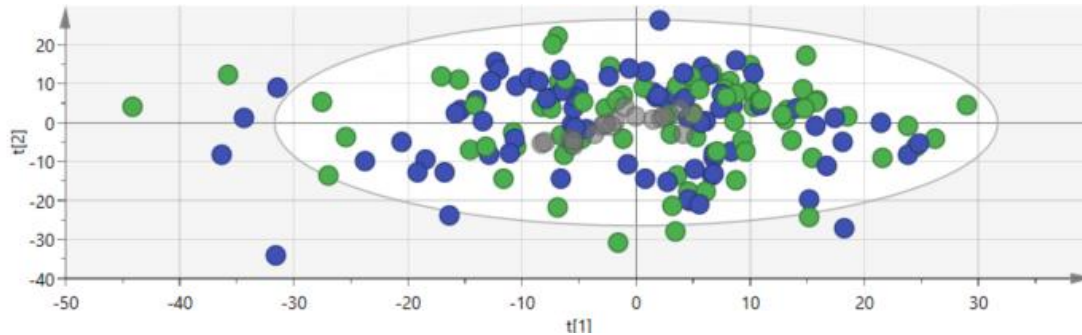


Figure 21 Score plot de l'ACP réalisée par la plateforme MS entre les femmes du 1<sup>er</sup> décile du mPNNS-GS (en vert) et du 10<sup>ème</sup> décile (en bleu) pour le protocole endogène en mode positif (les contrôles qualité sont représentés en gris)

Les 15 plus fortes associations entre les ions et le statut mPNNS-GS des participantes (1<sup>er</sup> vs 10<sup>ème</sup> décile) issues des régressions logistiques conditionnelles sont présentées dans le Tableau 18 et la liste de toutes les associations ( $p < 0.05$ ,  $N = 102$  associations) est présentée dans l'Annexe 6. Aucune de ces associations n'atteignait un  $FDR < 0.05$ . L'acide pipécolique (AP) et l'acide 3-indolepropionique (AIP) étaient supérieurs dans le plasma des femmes ayant une plus forte adéquation aux recommandations nutritionnelles alors que la lysophosphatidylcholine 17 :1 (LysoPC(17 :1)), les acylcarnitines C11 :1 et C9 :1 et le 2-déoxyglucose (2DG) étaient augmentés dans le plasma des femmes ayant une plus faible adéquation aux recommandations. L'acylcarnitine C9 :1 correspondrait soit à la 2,6-diméthylheptanoyl carnitine soit à la nonanoyl-L-carnitine et l'acylcarnitine C11 :1 correspondrait soit à la 4,8-diméthylnonanoyl carnitine soit à l'undecanoylcarnitine.



Tableau 18 Associations entre les ions et le niveau d'adéquation aux recommandations nutritionnelles mesuré par le mPNNS-GS  
<sup>1</sup> chez les femmes – cohorte SU.VI.MAX, 1994-2007

Ions Masse (M). Temps de rétention (T)	Identification <sup>2</sup>	Protocole/ mode d'ionisation <sup>3</sup>	OR <sup>4</sup> [IC95%]	P
M131.0895T1.16	<b>Acide pipécolique</b> , [M+H] <sup>+</sup> <sup>13</sup> C HMDB0000070 (niveau 1)	Endogène ESI+	13,12 [2,58- 66,60]	0,002
M130.0861T1.16	<b>Acide pipécolique</b> , [M+H] <sup>+</sup> HMDB0000070 (niveau 1)	Endogène ESI+	11,39 [2,39- 54,32]	0,002
M190.0859T10.97	<b>Acide 3-indolepropionique</b> , [M+H] <sup>+</sup> HMDB0002302 (niveau 1)	Exogène ESI+	4,39 [1,69- 11,42]	0,002
M84.0808T1.17	<b>Acide pipécolique</b> , [M+H-(HCOOH)] <sup>+</sup> HMDB0000070 (niveau 1)	Endogène ESI+	10,43 [2,23- 48,73]	0,003
M508.3395T15.27	(niveau 2: <b>LysoPC (17:1)</b> , [M+H] <sup>+</sup> ) LMGP01050002	Exogène ESI+	0,23 [0,09-0,63]	0,004
M330.2639T11.6	(niveau 3 : <b>4,8 dimethylnonanoyl carnitine</b> , [M+H] <sup>+</sup> HMDB0006202/ <b>undecanoylcarnitine</b> , [M+H] <sup>+</sup> HMDB0013321; <b>acylcarnitine C11:1</b> )	Exogène ESI+	0,37 [0,18-0,75]	0,006
M510.4363T17.5	(niveau 4: non identifié)	Endogène ESI+	0,33 [0,15-0,74]	0,006
M303.2357T10.49	(niveau 3: <b>2,6 dimethylheptanoyl carnitine</b> , [M+H] <sup>+</sup> <sup>13</sup> C HMDB0006320/ <b>nonanoyl-l-carnitine</b> , [M+H] <sup>+</sup> <sup>13</sup> C HMDB0013288; <b>acylcarnitine C9:1</b> )	Exogène ESI+	0,28 [0,11-0,70]	0,007
M76.0393T1.17	(niveau 4: non identifié)	Endogène ESI+	3,18 [1,37-7,35]	0,007
M189.0784T10.98	(niveau 4: non identifié)	Exogène ESI+	3,49 [1,40-8,67]	0,007
M737.8604T0.77	NaCOOH adduit (niveau 4)	Exogène ESI+	3,58 [1,40-9,17]	0,008
M252.0874T18.1	(niveau 4 : non identifié)	Exogène ESI+	3,43 [1,38-8,53]	0,008
M187.0577T0.97	<b>2-deoxy-D-glucose</b> , [M+Na] <sup>+</sup> HMDB0062477 (niveau 1)	Endogène ESI+	0,36 [0,17-0,77]	0,009
M302.2325T10.49	(niveau 3: <b>2,6 dimethylheptanoyl carnitine</b> , [M+H] <sup>+</sup> HMDB0006320/ <b>nonanoyl-l-carnitine</b> , [M+H] <sup>+</sup> HMDB0013288; <b>acylcarnitine C9:1</b> )	Exogène ESI+	0,35 [0,16-0,78]	0,01
M325.1129T0.87	(niveau 3 : classe des glucides C <sub>12</sub> H <sub>21</sub> O <sub>10</sub> )	Endogène ESI+	0,29 [0,11-0,75]	0,01

OR, odds ratio ; IC, intervalle de confiance

<sup>1</sup> Les modèles de régression logistique conditionnelle ont été ajustés sur les caractéristiques suivantes : âge, IMC, statut tabagique, nombre d'enregistrements alimentaires sur 24 heures et moyenne de l'apport énergétique quotidien pendant les 2 années suivant le prélèvement sanguin. Des tests de tendance linéaire ont été effectués sur les variables continues.

<sup>2</sup> Niveaux de confiance de l'identification : niveau 1 = composé identifié et validé par standard, niveau 2 = composé identifié de manière présumée, niveau 3 = classe du composé annotée de manière présumée et niveau 4 = composé inconnu.

<sup>3</sup> 160 individus ont été analysés pour le protocole endogène et le protocole exogène en mode d'ionisation négative et 152 individus pour le protocole exogène en mode ionisation positive (en raison des analyses conditionnelles, les paires correspondant aux échantillons exclus lors du prétraitement (N=4) n'ont pas été analysées).

<sup>4</sup> OR<1 : associé à une faible adéquation aux recommandations nutritionnelles françaises

OR>1 : associé à une forte adéquation aux recommandations nutritionnelles françaises

L'ensemble des corrélations (p<0.05, N=1522 associations) entre les composantes alimentaires du mPNNS-GS et les ions est présenté en Annexe 6 et les plus fortes sont présentées dans le Tableau 19. Les associations positives observées entre la consommation d'alcool et l'ethyl-β-D-glucopyranoside et entre la consommation de produits de la mer et des ions M157.0118T23.44 et M810.5654T24.77 ainsi que les

associations négatives entre la consommation de produits de la mer et les ions M136.8774T0.94, M318.1622T15.23 et M241.1031T0.89 atteignaient un FDR<0.05 (les FDR sont présentés en Annexe 6). Par ailleurs, des corrélations positives ont été observées entre l'acide pipécolique et la consommation de fruits et légumes et des corrélations négatives ont été observées entre la consommation de fruits et légumes et la phénylalanine et l'acylcarnitine C9 :1, entre la consommation de lait et produits laitiers et le dipeptide phénylalanine-phénylalanine, entre l'acide cholique et la consommation de produits sucrés et entre la LysoPC(17 :1) et la consommation de matière grasse végétale ajoutée.

Tableau 19 Associations entre les ions et les composantes alimentaires du mPNNs-GS<sup>1</sup> chez les femmes – cohorte SU.VI.MAX, 1994-2007

Ions Masse (M). Temps de rétention (T)	Identification <sup>2</sup>	Protocole/ mode d'ionisation <sup>3</sup>	R	P	Composantes mPNNs-GS
M231.0839T1.16	(niveau 2 : <b>Ethyl-β-D-glucopyranoside</b> , [M+Na] <sup>+</sup> HMDB0029968	Exogène ESI+	0,37	0,000003	
M136.8774T0.94	(niveau 4: non identifié)	Endogène ESI-	-0,32	0,00006	Alcool (g/jour)
M138.8751T0.92	(niveau 4: non identifié)	Exogène ESI-	-0,30	0,0001	
M131.0895T1.16	<b>Acide pipécolique</b> , [M+H] <sup>+</sup> <sub>-13</sub> C HMDB0000070 (niveau 1)	Endogène ESI +	0,31	0,00009	
M303.2357T10.49	(niveau 3: <b>2,6 dimethylheptanoyl carnitine</b> , [M+H] <sup>+</sup> <sub>-13</sub> C, HMDB0006320 / <b>nonanoyl-L-carnitine</b> , [M+H] <sup>+</sup> <sub>-13</sub> C HMDB0013288; <b>acylcarnitine C9:1</b> )	Exogène ESI +	-0,31	0,0001	Fruits et légumes (nombre de portions/jour)
M149.0596T5.81	<b>Phénylalanine</b> , [(M+H)-NH <sub>3</sub> ] <sup>+</sup> HMDB0000159 (niveau 1)	Exogène ESI +	-0,30	0,0002	
M342.2637T11.93	(niveau 4: non identifié)	Exogène ESI +	-0,31	0,0001	Viande, volaille, produits de la mer, oeufs (nombre de portions/jour)
M313.1546T8.83	<b>Phénylalanine-phénylalanine</b> , [M+H] <sup>+</sup> HMDB0013302 (niveau 1)	Exogène ESI +	-0,31	0,0002	Lait et produits laitiers (nombre de portions/jour)
M312.2261T15.13	(niveau 4: non identifié)	Exogène ESI -	0,30	0,0002	
M318.1622T15.23	(niveau 4: non identifié)	Exogène ESI +	-0,35	0,00001	
M241.1031T0.89	(niveau 4: non identifié)	Exogène ESI +	-0,32	0,00007	Produits de la mer (fréquence de consommation/ semaine)
M157.0118T23.44	(niveau 4: non identifié)	Exogène ESI -	0,31	0,00009	
M810.5654T24.77	(niveau 4: non identifié)	Exogène ESI -	0,31	0,0001	
M373.2736T12.81	<b>Acide cholique</b> , [(M+H)-2(H <sub>2</sub> O)] <sup>+</sup> HMDB00619 (niveau 1)	Exogène ESI +	-0,31	0,00009	Produits sucrés (% de l'apport énergétique sans l'alcool/jour)
M508.3395T15.27	(niveau 2: <b>LysoPC (17:1)</b> , [M+H] <sup>+</sup> LMGP01050002)	Exogène ESI +	-0,30	0,0002	Matière grasse végétale ajoutée (% d'huile végétale/matière grasse totale)

R, coefficient de corrélation

<sup>1</sup> Les corrélations de Spearman ont été ajustées sur les caractéristiques suivantes : âge, IMC, statut tabagique, statut ménopausique, saison de prélèvement sanguin, nombre d'enregistrements alimentaires sur 24 heures et moyenne de l'apport énergétique quotidien pendant les 2 années suivant le prélèvement sanguin.

<sup>2</sup> Niveaux de confiance de l'identification : niveau 1 = composé identifié et validé par standard, niveau 2 = composé identifié de manière présumée, niveau 3 = classe du composé annotée de manière présumée et niveau 4 = composé inconnu.

<sup>3</sup> 160 individus ont été analysés pour le protocole endogène et le protocole exogène en mode d'ionisation négative et 156 individus pour le protocole exogène en mode ionisation positive.

Les trois premières composantes de l'ACP permettaient d'expliquer 23% de la variabilité totale des données alimentaires de la population générale de SU.VI.MAX. La première composante était caractérisée par une consommation plus élevée de boissons alcoolisées, de matière grasse animale, de pain, de fromage, de pâtes, de riz et de céréales, de pommes de terre et de tubercules, de charcuterie, de légumineuses, de viande rouge et d'abats, de produits salés et sucrés, de sauces, de matière grasse végétale et de volaille. Elle a donc été considérée comme la représentation d'un profil alimentaire « occidental ». La deuxième composante était caractérisée par une consommation plus élevée de fruits, de légumes, de fruits secs, de poisson et de fruits de mer, de céréales pour petit-déjeuner, de matière grasse végétale, de céréales complètes et de yaourt – représentant un profil alimentaire « sain ». La troisième composante était caractérisée par une consommation plus élevée de gâteaux et de biscuits, de fruits secs, de jus de fruits, de matière grasse animale, de lait, de noix et de produits sucrés. Par conséquent, cette composante a été considérée comme la représentation d'un « régime à base de produits sucrés ». L'ensemble des corrélations de Spearman ( $p > 0.05$ ,  $N = 336$ ) entre ces 3 profils alimentaires et les ions est présenté en Annexe 6, et les plus fortes dans le Tableau 20. Les associations de l'acide pipécolique, positive avec le profil alimentaire « sain » et négative avec le profil alimentaire « occidental » atteignaient un  $FDR < 0.05$ . Par ailleurs, des corrélations positives entre la pipérine et le profil alimentaire « occidental », et entre le dihydro-4-mercapto-3(2H) furanone et le profil alimentaire « sain », ainsi qu'une corrélation négative entre l'acylcarnitine C9:1 et un profil alimentaire « sain » ont été observées.

Tableau 20 Associations entre les ions et les profils alimentaires obtenus par ACP<sup>1</sup> chez les femmes – cohorte SU.VI.MAX, 1994-2007

Ions Masse (M). Temps de rétention (T)	Identification <sup>2</sup>	Protocole/ mode d'ionisation <sup>3</sup>	R	P	Profils alimentaires obtenus par ACP
M131.0895T1.16	<b>Acide pipécolique</b> , [M+H] <sup>+</sup> <sub>-13</sub> C HMDB0000070 (niveau 1)	Endogène ESI+	-0,36	0,000006	
M130.0861T1.16	<b>Acide pipécolique</b> , [M+H] <sup>+</sup> HMDB0000070 (niveau 1)	Endogène ESI+	-0,35	0,000008	
M84.0808T1.17	<b>Acide pipécolique</b> , [M+H-(HCOOH)] <sup>+</sup> HMDB0000070 (niveau 1)	Endogène ESI+	-0,32	0,000005	Composante 1 Profil alimentaire "occidental"
M246.9051T14.25	(niveau 4: C <sub>2</sub> H <sub>2</sub> O <sub>7</sub> N <sub>2</sub> Br)	Exogène ESI-	-0,30	0,0002	
M286.1437T13.64	<b>Pipérine</b> , [M+H] <sup>+</sup> HMDB0029377 (niveau 1)	Exogène ESI+	0,26	0,001	
M131.0895T1.16	<b>Acide pipécolique</b> , [M+H] <sup>+</sup> <sub>-13</sub> C HMDB0000070 (niveau 1)	Endogène ESI +	0,36	0,000004	
M130.0861T1.16	<b>Acide pipécolique</b> , [M+H] <sup>+</sup> HMDB0000070 (niveau 1)	Endogène ESI +	0,34	0,000002	
M84.0808T1.17	<b>Acide pipécolique</b> , [M+H-(HCOOH)] <sup>+</sup> HMDB0000070 (niveau 1)	Endogène ESI +	0,32	0,000005	
M303.2357T10.49	(niveau 3: <b>2,6 diméthylheptanoyl carnitine</b> , [M+H] <sup>+</sup> <sub>-13</sub> C HMDB0006320/ <b>nonanoyl-l-carnitine</b> , [M+H] <sup>+</sup> <sub>-13</sub> C HMDB0013288; <b>acylcarnitine C9:1</b> )	Exogène ESI +	-0,30	0,0002	Composante 2 Profil alimentaire "sain"
M119.016T1.03	(niveau 2: <b>Dihydro4mercapto-3(2H) furanone</b> , [M+H] <sup>+</sup> ) HMDB0039786	Endogène ESI +	0,28	0,0004	
M73.0284T1.02	(niveau 4 : non identifié)	Endogène ESI +	0,28	0,0006	
M918.3358T10.11	(niveau 4 : non identifié)	Endogène ESI +	-0,27	0,0009	
M224.0998T11.87	(niveau 4 : formule brute C <sub>12</sub> H <sub>15</sub> O <sub>4</sub> )	Endogène ESI +	-0,30	0,0002	
M340.0238T11.87	(niveau 4 : non identifié)	Endogène ESI +	-0,26	0,001	Composante 3 Régime à base de produits sucrés
M498.9301T19.76	(niveau 4 : non identifié)	Exogène ESI -	-0,26	0,001	

R, coefficient de corrélation

<sup>1</sup> Les corrélations de Spearman ont été ajustées sur les caractéristiques suivantes : âge, IMC, statut tabagique, statut ménopausique, saison de prélèvement sanguin, nombre d'enregistrements alimentaires sur 24 heures et moyenne de l'apport énergétique quotidien pendant les 2 années suivant le prélèvement sanguin.

<sup>2</sup> Niveaux de confiance de l'identification : niveau 1 = composé identifié et validé par standard, niveau 2 = composé identifié de manière présumée, niveau 3 = classe du composé annotée de manière présumée et niveau 4 = composé inconnu.

<sup>3</sup> 160 individus ont été analysés pour le protocole endogène et le protocole exogène en mode d'ionisation négative et 156 individus pour le protocole exogène en mode ionisation positive.

3. *Objectif secondaire pour les données de LC-MS : identifier de potentiels biomarqueurs de consommation d'aliments ou de groupes d'aliments spécifiques*

Cette étude fait l'objet d'un article en cours de rédaction.

*Dalle C\*, Lécuyer L\*, Micheau P, Pétéra M, Centeno D, Lyan B, Lagree M, Galan P, Hercberg S, Rossary A, Demidem A, Vasson MP, Partula V, Deschasaux M, Srour B, Latino-Martel P, Druésne-Pecollo N, Kesse-Guyot E, Durand S, Pujos-Guillot E, Touvier M#, Manach C#. \*,#: contributions égales. **Metabolomics applied to nutritional epidemiology to identify plasma biomarkers of food intake in women from the SU.VI.MAX cohort.** En préparation*

▪ Méthodes

La même population et la même méthode analytique que celles décrites pour l'objectif principal ont été utilisées. Pour cet objectif secondaire, les expositions alimentaires étudiées étaient les 58 aliments ou groupes d'aliments spécifiques décrits dans le Tableau 7 (en italique) dans la partie Méthodologie générale V. Des corrélations de Spearman ajustées sur les mêmes facteurs que ceux décrits juste avant ont été testées. Seules les corrélations positives ont été sélectionnées.

• Résultats

Les plus fortes corrélations ( $r > 0.2$ ,  $p < 0.05$ ,  $FDR < 0.1$ ) entre les métabolites et les groupes d'aliments spécifiques sont présentés en Annexe 7. Parmi celles-ci, celles correspondantes à des composés identifiés sont présentées dans le Tableau 21. La trigonelline, la paraxanthine, le cyclo(isoleucyl-prolyl), l'atractyligenin glucuronide étaient associés à la consommation de café ; la proline bétaine et l'hydroxyproline bétaine étaient associées à la consommation d'agrumes ; la glucokinase était associée à la consommation de lait ; la N6,N6,N6-Triméthyl-L-lysine était associée à la charcuterie. Dans une moindre mesure, nous avons trouvé que l'acide pipécolique (représenté par 3 ions) était associé à la consommation de pommes et de poires (coefficients de corrélation  $r$  entre 0.22 et 0.28 et  $p$  entre 0.0006 et 0.003), de yaourt ( $r$  entre 0.17 et 0.24 et  $p$  entre 0.003 et 0.04) et de fibres ( $r$  entre 0.26 et 0.28 et  $p$  entre 0.0004 et 0.0009), et que la pipérine était associée à la charcuterie ( $r = 0.27$  et  $p = 0.0009$ ).

Tableau 21 Associations entre les ions identifiés et les groupes d'aliments spécifiques chez les femmes – cohorte SU.VI.MAX, 1994-2007

Ions Masse (M). Temps de rétention (T)	Identification <sup>2</sup>	Protocole/ mode d'ionisation <sup>3</sup>	R	P	Groupe d'aliments
M144.1018T0.96	Proline bêtaïne (niveau 1)	Exogène ESI+	0,53	<0,0001	Agrumes
M144.1018T0.99	Proline bêtaïne (niveau 1)	Endogène ESI+	0,54	<0,0001	Agrumes
M160.0968T0.88	Hydroxyproline bêtaïne (niveau 1)	Exogène ESI+	0,37	<0,0001	Agrumes
M138.0542T0.92	Trigonelline (niveau 1)	Exogène ESI+	0,47	<0,0001	Café
M138.0545T0.94	Trigonelline (niveau 1)	Endogène ESI+	0,51	<0,0001	Café
M181.0727T7.41	Paraxanthine (niveau 1)	Endogène ESI+	0,40	<0,0001	Café
M181.0727T7.44	Paraxanthine (niveau 1)	Exogène ESI+	0,36	<0,0001	Café
M211.144T8.79	Cyclo(isoleucyl-prolyl) (niveau 1)	Exogène ESI+	0,39	<0,0001	Café
M497.238T10.12	Atractyligenin glucuronide (niveau 1)	Exogène ESI+	0,56	<0,0001	Café
M189.1597T0.79	N6,N6,N6-Trimethyl-L-lysine (niveau 1)	Exogène ESI+	0,34	<0,0001	Charcuterie
M130.0862T0.72	Glycokinase (niveau 1)	Endogène ESI+	0,34	<0,0001	Lait

<sup>1</sup> Les corrélations de Spearman ont été ajustées sur les caractéristiques suivantes : âge, IMC, statut tabagique, statut ménopausique, saison de prélèvement sanguin, nombre d'enregistrements alimentaires sur 24 heures et moyenne de l'apport énergétique quotidien pendant les 2 années suivant le prélèvement sanguin. Associations présentées :  $r > 0.2$ , P-corrigée  $< 0.1$ , pour les ions identifiés.

<sup>2</sup> Niveaux de confiance de l'identification : niveau 1 = composé identifié et validé par standard, niveau 2 = composé identifié de manière présumée, niveau 3 = classe du composé annotée de manière présumée et niveau 4 = composé inconnu.

<sup>3</sup> 160 individus ont été analysés pour le protocole endogène et le protocole exogène en mode d'ionisation négative et 156 individus pour le protocole exogène en mode ionisation positive

## B. Analyses en H-RMN

### 1. Méthodes

La population d'étude était la même que pour les analyses en LC-MS. Les échantillons de plasma ont été analysés en H-RMN avec une séquence NOESY1D et une séquence CPMG. Le détail de ces analyses est présenté dans la partie Méthodologie générale (IV.B.1.). Les données d'une participante ont été exclues en raison de problème pendant la préparation des échantillons. La procédure "intelligent bucketing" (253) a été appliquée sur les spectres de 0 à 10 ppm afin de reconstituer les pics. Après prétraitements des données 233 signaux RMN et 214 ont été obtenus pour la séquence NOESY1D et CPMG respectivement.

Les analyses statistiques réalisées étaient les mêmes que pour les données de LC-MS (Partie Objectif 2 I.A.1.).

### 2. Résultats

Les caractéristiques de la population d'étude sont présentées dans le Tableau 17 précédent. Les résultats des analyses multivariées réalisées au sein de la plateforme de métabolomique n'ont pas permis de distinguer le statut des participants grâce aux signaux RMN analysés (ni par ACP, ni par PLS-DA). Un exemple graphique de l'ACP est présenté en Figure 22. Il a cependant été possible de discriminer les

participantes du 1<sup>er</sup> décile de celles du 10<sup>ème</sup> décile du mPNNS-GS chez les personnes en surpoids ou obèses (Figure 23).

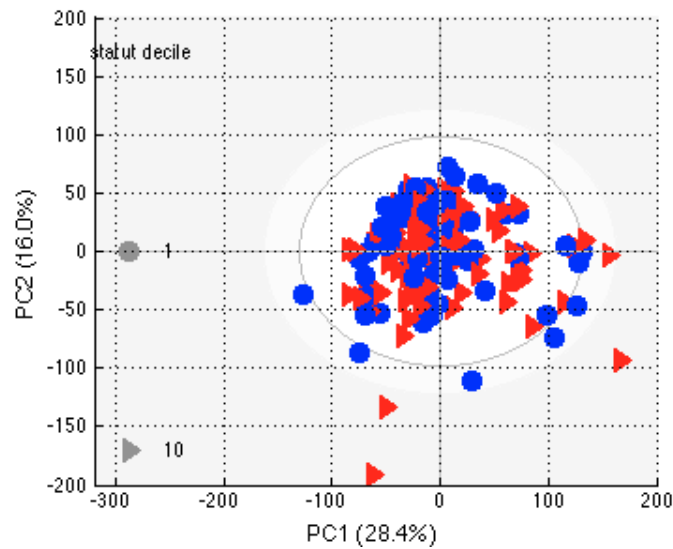


Figure 22 Score plot de l'ACP réalisée par la plateforme RMN entre les femmes du 1<sup>er</sup> décile du mPNNS-GS (en bleu) et du 10<sup>ème</sup> décile (en rouge)

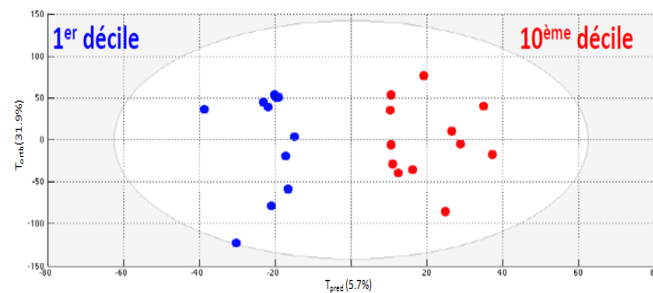


Figure 23 Score plot de l'ACP réalisée par la plateforme RMN entre les femmes du 1<sup>er</sup> décile du mPNNS-GS (en bleu) et du 10<sup>ème</sup> décile (en rouge) chez les femmes en surpoids.

Les associations ( $p < 0.05$ ) entre les signaux RMN et le statut mPNNS-GS des participantes (1<sup>er</sup> vs 10<sup>ème</sup> décile) issues des régressions logistiques conditionnelles sont présentées dans le Tableau 22. Aucune de ces associations n'atteignait un  $FDR < 0.05$ . Des taux plasmatiques plus élevés en leucine et créatine et des taux plus faibles en glucose, lipoprotéines, lipides insaturés, lactate et thréonine ont été observés chez les femmes ayant une meilleure adéquation aux recommandations nutritionnelles.

Tableau 22 Associations entre les signaux RMN et le niveau d'adéquation aux recommandations nutritionnelles mesuré par le mPNNS-GS<sup>1</sup> chez les femmes – cohorte SU.VI.MAX, 1994-2007

Assignations	Déplacement chimique (ppm)	Séquence <sup>2</sup>	OR [IC95%] <sup>3</sup>	P
Leucine	0,9103	CPMG	2,7 [1,29-5,65]	0,009
Leucine	0,9100	NOESY1D	2,72 [1,29-5,75]	0,009
Pic non défini	0,2463	NOESY1D	2,97 [1,26-6,97]	0,01
Pic non défini	0,1263	NOESY1D	2,55 [1,19-5,44]	0,02
Non identifié	1,4323	CPMG	0,31 [0,11-0,82]	0,02
Glucose	3,3322	NOESY1D	0,45 [0,23-0,88]	0,02
Créatine	3,9200	CPMG	2,38 [1,15-4,93]	0,02
Lipoprotéines (CH <sub>2</sub> )	1,1663	NOESY1D	0,31 [0,12-0,85]	0,02
Pic non défini	0,0337	NOESY1D	2,57 [1,14-5,78]	0,02
Pic non défini	0,2863	NOESY1D	2,22 [1,1-4,46]	0,03
Pic non défini	0,3663	NOESY1D	2,1 [1,08-4,08]	0,03
Pic non défini	0,4063	NOESY1D	2,19 [1,08-4,43]	0,03
Créatine	3,9180	NOESY1D	2,14 [1,08-4,25]	0,03
Pic non défini	0,0063	NOESY1D	2,19 [1,06-4,5]	0,03
Pic non défini	0,0863	NOESY1D	2,2 [1,06-4,57]	0,03
Glucose	5,2061	NOESY1D	0,44 [0,21-0,94]	0,04
Pic non défini	0,0463	NOESY1D	2,1 [1,04-4,2]	0,04
Lipides insaturés (CH <sub>2</sub> -C=C)	1,9263	CPMG	0,44 [0,2-0,95]	0,04
Lactate, thréonine	1,3103	NOESY1D	0,45 [0,21-0,96]	0,04
Pic non défini	2,8463	CPMG	2,29 [1,01-5,16]	0,046
Lipoprotéines (CH <sub>2</sub> )	1,1663	CPMG	0,48 [0,23-0,99]	0,046
Pic non défini	0,0863	CPMG	2,35 [1,01-5,45]	0,047

OR, odds ratio ; IC, intervalle de confiance

<sup>1</sup> Les modèles de régression logistique conditionnelle ont été ajustés sur les caractéristiques suivantes : âge, IMC, statut tabagique, nombre d'enregistrements alimentaires sur 24 heures et moyenne de l'apport énergétique quotidien pendant les 2 années suivant le prélèvement sanguin. Des tests de tendance linéaire ont été effectués sur les variables continues.

<sup>2</sup> 158 individus ont été analysés.

<sup>3</sup> OR<1 : associé à une faible adéquation aux recommandations nutritionnelles françaises

OR>1 : associé à une forte adéquation aux recommandations nutritionnelles françaises

L'ensemble des corrélations ( $p < 0.05$ ,  $N = 326$  associations) entre les composantes alimentaires du mPNNS-GS et les signaux RMN sont présentées en Annexe 8. Les 15 plus fortes sont présentées dans le Tableau 23. Des associations positives ont été observées entre la leucine et la consommation de produits de la mer, entre l'arginine et la lysine (superposition de signaux) et la consommation de produits céréaliers complets ; entre la tyrosine et la consommation de pain, céréales, pommes de terre et légumineuses ; entre la phénylalanine et la consommation de fruits et légumes et entre le glucose et la consommation de matière grasse totale ajoutée. Des associations négatives ont été observées entre l'isobutyrate, la valine et les lipoprotéines et la consommation de pain, céréales, pommes de terre et légumineuses. Toutes ces associations atteignaient le seuil  $FDR < 0.05$ , sauf celles du glucose et de la tyrosine ( $FDR < 0.15$ ).



Tableau 23 Associations entre les signaux RMN et les composantes alimentaires du mPNNs-GS<sup>1</sup> chez les femmes – cohorte SU.VI.MAX, 1994-2007

Assignations	Déplacement chimique (ppm)	Séquence <sup>2</sup>	R	P	Composantes mPNNs-GS
Leucine	0,9103	CPMG	0,30	0,0001	Produits de la mer (fréquence de consommation/ semaine)
Leucine	0,9100	NOESY1D	0,30	0,0002	
<i>Arginine, Lysine</i>	1,6463	CPMG	0,30	0,0002	Produits céréaliers complets (proportion de féculents)
<i>Non identifié</i>	6,3310	NOESY1D	0,30	0,0002	Pain, céréales, pommes de terre et légumineuses (portions/jour)
<i>Isobutyrate</i>	1,1263	NOESY1D	-0,27	0,0006	
<i>Valine</i>	1,0213	NOESY1D	-0,27	0,0007	
<i>Lipoprotéines (CH<sub>3</sub>)</i>	0,8063	NOESY1D	-0,27	0,0008	
<i>Tyrosine</i>	6,8600	CPMG	0,27	0,0009	
<i>Non identifié</i>	8,9103	NOESY1D	0,29	0,0002	Fruits et légumes (portions/jour)
<i>Non identifié</i>	6,7260	CPMG	0,28	0,0005	
<i>Phénylalanine</i>	7,3659	NOESY1D	0,27	0,0007	
<i>Non identifié</i>	8,0459	NOESY1D	0,27	0,0008	
<i>Non identifié</i>	2,8863	CPMG	0,27	0,0008	
<i>Glucose</i>	5,2061	CPMG	0,29	0,0003	
<i>Glucose</i>	5,2061	NOESY1D	0,28	0,0006	

R, coefficient de corrélation. La validation 2D des signaux en italique est en cours sur la plateforme RMN.

<sup>1</sup> Les corrélations de Spearman ont été ajustées sur les caractéristiques suivantes : âge, IMC, statut tabagique, statut ménopausique, saison de prélèvement sanguin, nombre d'enregistrements alimentaires sur 24 heures et moyenne de l'apport énergétique quotidien pendant les 2 années suivant le prélèvement sanguin.

<sup>2</sup> 159 individus ont été analysés.

L'ensemble des corrélations de Spearman ( $p < 0.05$ ,  $N = 103$  associations) entre les 3 profils alimentaires obtenus par ACP et les signaux RMN est présenté en Annexe 9. Aucune association n'atteignait le seuil  $FDR < 0.05$ . Les 15 plus fortes associations sont présentées dans le Tableau 24. La leucine était associée négativement à un profil alimentaire « occidental », les lipoprotéines et les lipides insaturés étaient associés négativement à un profil alimentaire « sain » et le citrate était associé positivement à un régime à base de produits sucrés.

Tableau 24 Associations entre les signaux RMN et les profils alimentaires obtenus par ACP<sup>1</sup> chez les femmes – cohorte SU.VI.MAX, 1994-2007

Assignations	Déplacement chimique (ppm)	Séquence <sup>2</sup>	R	P	Profils alimentaires obtenus par ACP
<i>Non identifié</i>	7,6459	CPMG	-0,23	0,005	Composante 1 Profil alimentaire "occidental"
Leucine	0,9103	CPMG	-0,23	0,005	
Leucine	0,9100	NOESY1D	-0,22	0,006	
<i>Non identifié</i>	2,8863	CPMG	0,28	0,0004	Composante 2 Profil alimentaire "sain"
Lipoprotéines (CH <sub>2</sub> )	1,1663	NOESY1D	-0,26	0,001	
<i>Non identifié</i>	6,2150	CPMG	-0,23	0,004	
Lipoprotéines (CH <sub>2</sub> )	1,1663	CPMG	-0,23	0,005	
<i>Lipides insaturés (CH<sub>2</sub>-C=C)</i>	1,9263	CPMG	-0,22	0,006	
<i>Non identifié</i>	6,1280	CPMG	0,26	0,001	Composante 3 Régime à base de produits sucrés
<i>Citrate</i>	2,6460	NOESY1D	0,26	0,001	
<i>Non identifié</i>	7,6859	CPMG	0,25	0,002	
<i>Non identifié</i>	2,5693	CPMG	0,25	0,002	
<i>Non identifié</i>	6,2150	CPMG	0,25	0,002	
<i>Non identifié</i>	6,9660	CPMG	0,24	0,003	
<i>Non identifié</i>	6,2460	CPMG	0,23	0,005	

R, coefficient de corrélation. La validation 2D des signaux en italique est en cours sur la plateforme RMN.

<sup>1</sup> Les corrélations de Spearman ont été ajustées sur les caractéristiques suivantes : âge, IMC, statut tabagique, statut ménopausique, saison de prélèvement sanguin, nombre d'enregistrements alimentaires sur 24 heures et moyenne de l'apport énergétique quotidien pendant les 2 années suivant le prélèvement sanguin.

<sup>2</sup> 159 individus ont été analysés.

## II. Signature de l'alimentation globale chez les hommes – analyses en H-RMN

### A. Méthodes

La sélection de la population d'étude pour ces analyses a été détaillée dans la partie Méthodologie générale (III.B.2.). Il s'agit d'une étude transversale dans la population SU.VI.MAX incluant 70 hommes du 1<sup>er</sup> décile du mPNNS-GS et 70 hommes du 10<sup>ème</sup> décile, appariés. Les échantillons de plasma ont été analysés en H-RMN avec une séquence NOESY1D et une séquence CPMG. Le détail de ces analyses est présenté dans la partie Méthodologie générale (IV.B.1.). Les données de 2 paires de participants ont été exclues en raison de problème pendant la préparation des échantillons ou l'acquisition des spectres pour la séquence NOESY1D. La procédure "intelligent bucketing" (253) a été appliquée sur les spectres de 0 à 10 ppm afin de reconstituer les pics. Après prétraitements des données 219 signaux RMN et 218 ont été obtenus pour la séquence NOESY1D et CPMG respectivement.

Les analyses statistiques réalisées étaient les mêmes que pour les femmes (Partie Objectif 2 I.A.1.). Les régressions logistiques conditionnelles ont été ajustées sur l'âge (en continu), l'IMC (en continu), le statut tabagique (non fumeurs, arrêt de la cigarette depuis plus de 16 ans, arrêt de la cigarette depuis moins de 16 ans et plus de 6, arrêt de la cigarette depuis moins de 6 ans, fumeurs), la moyenne de l'apport énergétique et le nombre d'enregistrements de 24h validés pendant les 2 ans suivant le prélèvement sanguin. Les analyses de corrélations ont été ajustées sur l'âge (en continu), l'IMC (en continu), le statut tabagique (non fumeurs, arrêt de la cigarette depuis plus de 16 ans, arrêt de la cigarette depuis moins de 16 ans et plus de 6, arrêt de la cigarette depuis moins de 6 ans, fumeurs), la saison de prélèvement sanguin

et la moyenne de l'apport énergétique et le nombre d'enregistrements de 24h validés pendant les 2 ans suivant le prélèvement sanguin.

## B. Résultats

Le Tableau 25 résume la comparaison des caractéristiques des deux groupes de participants inclus dans l'étude. Les hommes du 10<sup>e</sup> décile du mPNNS-GS avaient de meilleurs scores d'adéquation aux recommandations pour les composantes fruits et légumes ; produits céréaliers complets ; lait et produits laitiers ; viande, volaille, produits de la mer et œufs ; produits de la mer ; matière grasse totale ajoutée ; matière grasse végétale ajoutée ; boissons sucrées ; alcool et sel. Les composantes pain, céréales, pommes de terre et légumineuses et produits sucrés ne permettaient pas de discriminer les deux groupes. Les hommes du 1<sup>er</sup> décile avaient une moins bonne adéquation aux recommandations concernant l'activité physique et avaient une dépense énergétique plus faible et un apport énergétique plus élevé et donc un ratio apport/dépense énergétique plus élevé.

Tableau 25 Caractéristiques<sup>1</sup> à l'inclusion des hommes du 1<sup>e</sup> et du 10<sup>ème</sup> déciles du mPNNS-GS<sup>2</sup> – cohorte SU.VI.MAX, 1994-2007

	Hommes du 1 <sup>er</sup> décile (N=70)	Hommes du 10 <sup>ème</sup> décile (N=70)	p <sup>3</sup>
mPNNS-GS	4,0±0,8	9,8±0,7	Non applicable
Age	52,0±4,6	52,4±4,4	0,06
IMC (kg/m <sup>2</sup> )	24,8±2,7	24,8±2,4	0,9
IMC (catégories)			Non applicable
<18.5 kg/m <sup>2</sup> (sous-poids)	0 (0)	0 (0)	
≥18.5 to <25 kg/m <sup>2</sup> (poids normal)	39 (55,7)	39 (55,7)	
≥25 kg/m <sup>2</sup> (surpoids)	31 (44,3)	31 (44,3)	
Statut tabagique			Non applicable
Anciens ou non fumeurs	66 (94,3)	66 (94,3)	
Fumeurs	4 (5,7)	4 (5,7)	
Mois de prélèvement sanguin			Non applicable
Mars-Avril	21 (30,0)	21 (30,0)	
Octobre-Novembre	13 (18,6)	13 (18,6)	
Décembre-Janvier-Février	36 (51,4)	36 (51,4)	
Nombre d'enregistrements de 24h	11,9±1,2	12,3±1,4	0,1
Moyenne de l'apport énergétique quotidien (kcal)	2945,6±512,1	2132,1±477,1	<0,0001
<b>Fruits et légumes</b>	0,6±0,6	1,5±0,6	<0,0001
[0-3.5[ = 0	20 (28,6)	1 (1,4)	0,004
[3.5-5[ = 0.5	24 (34,3)	6 (8,6)	
[5-7.5[ = 1	19 (27,1)	23 (32,9)	
≥7.5 = 2	7 (10,0)	40 (57,1)	
<b>Pain, céréales, pommes de terre et légumineuses</b>	0,8±0,3	0,8±0,2	0,1
[0-1[ = 0	0 (0,0)	0 (0,0)	0,1
[1-3[ or ≥6 = 0.5	32 (45,7)	24 (34,3)	
[3-6[ = 1	38 (54,3)	46 (65,7)	
<b>Produits céréaliers complets</b>	0,1±0,2	0,4±0,5	0,0002
[0-1/3[ = 0	60 (85,7)	35 (50,0)	0,001
[1/3-2/3[ = 0.5	7 (10,0)	11 (15,7)	
≥2/3 = 1	3 (4,3)	24 (34,3)	
<b>Lait et produits laitiers</b>	0,3±0,4	0,7±0,4	<0,0001
[0-1[ ou >3.5 (>4.5 si ≥55 ans) = 0	41 (58,6)	15 (21,4)	0,0001
[1-2.5[ = 0.5	18 (25,7)	14 (20,0)	
[2.5-3.5[ ([2.5-4.5] si ≥55 ans) = 1	11 (15,7)	41 (58,6)	
<b>Viande, volaille, produits de la mer et œufs</b>	0,3±0,5	0,9±0,3	<0,0001
0 ou >2 = 0	48 (68,6)	8 (11,4)	<0,0001
]0-1[ = 0.5	1 (1,4)	1 (1,4)	
[1-2[ = 1	21 (30,0)	61 (87,1)	

<b>Produits de la mer</b>	0,2±0,4	0,8±0,4	<0,0001
<2 portions/semaine = 0	59 (84,3)	15 (21,4)	<0,0001
≥2 portions/semaine = 1	11 (15,7)	55 (78,6)	
<b>Matière grasse ajoutée totale</b>	0,6±0,5	0,9±0,3	0,002
Lipides des matières grasses animales et végétales >16% AE/j = 0	27 (38,6)	8 (11,4)	0,002
Lipides des matières grasses animales et végétales ≤16% AE/j = 1	43 (61,4)	62 (88,6)	
<b>Matière grasse ajoutée végétale</b>	0,3±0,4	0,9±0,2	0,0001
Pas de consommation de matière grasse ajoutée végétale ou ratio matière grasse ajoutée végétale / totale ≤0.5 = 0	52 (74,3)	4 (5,7)	0,0001
Pas de consommation de matière grasse ajoutée totale ou ratio matière grasse ajoutée végétale / totale >0.5 = 1	18 (25,7)	66 (94,3)	
<b>Produits sucrés</b>	0,8±0,4	1,0±0,0	0,9
Sucre ajouté issu des produits sucrés ≥15% AE/jour = -0.5	1 (1,4)	0 (0,0)	0,9
Sucre ajouté issu des produits sucrés : 10 à 15% AE/jour = 0	16 (22,9)	0 (0,0)	
Sucre ajouté issu des produits sucrés <10% AE/jour = 1	53 (75,7)	70 (100,0)	
<b>Boissons sucrées</b>	0,8±0,1	0,9±0,1	0,0002
<1l eau et >250 ml soda/jour = 0	0 (0,0)	0 (0,0)	0,002
≥1l eau et >250 ml soda/ jour = 0.5	1 (1,4)	0 (0,0)	
<1l eau et ≤250 ml soda/ jour = 0.75	45 (64,3)	21 (30,0)	
≥1l eau et ≤250 ml soda/ jour = 1	24 (34,3)	49 (70,0)	
<b>Alcool</b>	0,2±0,4	0,7±0,3	<0,0001
Ethanol >20 g/jour pour les femmes et >30g/jour pour les hommes = 0	53 (75,7)	8 (11,4)	<0,0001
Ethanol ≤20 g/jour pour les femmes et ≤30g/jour pour les hommes = 0.8	16 (22,9)	57 (81,4)	
Non-consommateurs ou consommateurs irréguliers (<1 verre de vin/semaine) = 1	1 (1,4)	5 (7,1)	
<b>Sel</b>	-0,4±0,4	0,2±0,6	<0,0001
>12 g/j = -0.5	59 (84,3)	21 (30,0)	0,0008
]10-12] g/j = 0	6 (8,6)	16 (22,9)	
]8-10] g/j = 0.5	3 (4,3)	15 (21,4)	
]6-8] g/j = 1	1 (1,4)	15 (21,4)	
≤6 g/j = 1.5	1 (1,4)	3 (4,3)	
<b>Activité physique</b>	0,4±0,6	0,9±0,6	0,0003
[0-30[ min/j = 0	46 (65,7)	22 (31,4)	0,001
[30-60[ min/j = 1	16 (22,9)	21 (30,0)	
≥60 min/j = 1.5	8 (11,4)	27 (38,6)	
<b>Métabolisme basal</b>	1724,3±107,5	1715,6±124,9	0,5
<b>Dépense énergétique (kcal)</b>	2869,3±341,5	3138,6±440,8	0,0008
<b>Ratio apport/dépense énergétique</b>	1,0±0,2	0,7±0,2	<0,0001

AE, apport énergétique

<sup>1</sup> Les valeurs présentées sont les moyennes ± écart type ou l'effectif (%)

<sup>2</sup> Les composantes du mPNNS-GS sont identifiées en gras

<sup>3</sup> P de tendance linéaire pour la comparaison entre les hommes du 1<sup>e</sup> décile et du 10<sup>e</sup> décile du mPNNS-GS grâce aux modèles de régression logistique conditionnelle. Non applicable pour les variables d'appariement sauf si variables plus précises (ex. âge et IMC).

Les résultats des analyses multivariées réalisées au sein de la plateforme de métabolomique n'ont pas permis de distinguer le statut des participants grâce aux signaux RMN analysés (ni par ACP, ni par PLS-DA).

Les associations (p<0.05) entre les signaux RMN et le statut mPNNS-GS des participants (1<sup>er</sup> vs 10<sup>ème</sup> décile) issues des régressions logistiques conditionnelles sont présentées dans le Tableau 26. Les taux plasmatiques de leucine et d'albumine étaient plus élevés chez les hommes ayant une meilleure

adéquation aux recommandations nutritionnelles. Toutefois, aucune association n’atteignait le seuil  $FDR < 0.05$ .

Tableau 26 Associations entre les signaux RMN et le niveau d’adéquation aux recommandations nutritionnelles mesuré par le mPNNS-GS<sup>1</sup> chez les hommes – cohorte SU.VI.MAX, 1994-2007

Assignations	Déplacement chimique (ppm)	Séquence <sup>2</sup>	OR [IC95%] <sup>3</sup>	P
Leucine	0,9107	CPMG	3,80 [1,35-10,66]	0,01
Pic non défini	9,0231	CPMG	6,74 [1,27-35,78]	0,03
Non identifié	1,4371	CPMG	0,31 [0,10-0,94]	0,04
Albumine	2,873	CPMG	3,94 [1,03-15,12]	0,046

OR, odds ratio ; IC, intervalle de confiance

<sup>1</sup> Les modèles de régression logistique conditionnelle ont été ajustés sur les caractéristiques suivantes : âge, IMC, statut tabagique, nombre d’enregistrements alimentaires sur 24 heures et moyenne de l’apport énergétique quotidien pendant les 2 années suivant le prélèvement sanguin. Des tests de tendance linéaire ont été effectués sur les variables continues.

<sup>2</sup> 140 individus ont été analysés pour la séquence CPMG et 138 pour la séquence NOESY1D.

<sup>3</sup> OR<1 : associé à une faible adéquation aux recommandations nutritionnelles françaises

OR>1 : associé à une forte adéquation aux recommandations nutritionnelles françaises

L’ensemble des corrélations ( $p < 0.05$ ,  $N = 254$  associations) entre les composantes alimentaires du mPNNS-GS et les signaux RMN est présenté en Annexe 10 et les 15 plus fortes sont présentées dans le Tableau 27. La consommation de pain, céréales, pommes de terre et légumineuses était associée positivement à la choline et à un signal regroupant la créatine et le glycéryl (superposition de signaux) et négativement au lactate, aux lipides, au pyruvate et à un signal regroupant la choline et le glucose (superposition de signaux). La consommation d’alcool était associée positivement au propylène glycol, à la choline et glucose (superposition de signaux) et aux lipides. La consommation de sel était associée négativement au lactate et des associations positives ont été observées entre le glucose et la consommation de lait et produits laitiers et entre la tyrosine et les matières grasses ajoutées totales. Toutes ces associations atteignaient le seuil  $FDR < 0.05$ , sauf celles entre lait et glucose et entre pyruvate et féculents ( $FDR < 0.15$ ).

Tableau 27 Associations entre les signaux RMN et les composantes alimentaires du mPNNS-GS<sup>1</sup> chez les hommes – cohorte SU.VI.MAX, 1994-2007

Assignations	Déplacement chimique (ppm)	Séquence <sup>2</sup>	R	P	Composantes mPNNS-GS
<i>Choline</i>	4,055	CPMG	0,37	0,00001	
<i>Créatine, glycéryl</i>	4,031	NOESY1D	0,33	0,0001	
<i>Lactate</i>	4,075	NOESY1D	-0,30	0,0004	Pain, céréales, pommes de terre et légumineuses (nombre de portions/jour)
<i>Lipides (CH2-CH)</i>	1,927	CPMG	-0,29	0,0008	
<i>Pyruvate</i>	2,3511	NOESY1D	-0,29	0,0009	
<i>Choline, glucose</i>	3,167	CPMG	-0,28	0,0009	
<i>Non identifié</i>	1,145	CPMG	-0,35	0,00003	
<i>Propylène glycol</i>	1,1271	CPMG	0,33	0,00008	
<i>Choline, glucose</i>	3,167	NOESY1D	0,32	0,0002	Alcool (g/jour)
<i>Lipides (CH2CH2COOC)</i>	1,9271	NOESY1D	0,30	0,0004	
<i>Choline, glucose</i>	3,167	CPMG	0,29	0,0006	
<i>Lactate</i>	4,075	NOESY1D	-0,32	0,0002	Sel (g/jour)
<i>Non identifié</i>	6,1058	CPMG	-0,30	0,0005	Produits céréaliers complets (proportion sur produits céréaliers totaux)
<i>Glucose</i>	3,366	NOESY1D	0,30	0,0006	<i>Lait et produits laitiers (nombre de portions/jour)</i>
<i>Tyrosine</i>	7,159	CPMG	-0,28	0,0009	Matière grasse ajoutée totale (% de l'apport énergétique sans l'alcool/jour)

R, coefficient de corrélation. La validation 2D des signaux en italique est en cours sur la plateforme RMN.

<sup>1</sup> Les corrélations de Spearman ont été ajustées sur les caractéristiques suivantes : âge, IMC, statut tabagique, saison de prélèvement sanguin, nombre d'enregistrements alimentaires sur 24 heures et moyenne de l'apport énergétique quotidien pendant les 2 années suivant le prélèvement sanguin.

<sup>2</sup> 140 individus ont été analysés pour la séquence CPMG et 138 pour la séquence NOESY1D

L'ensemble des corrélations de Spearman ( $p < 0.05$ ,  $N = 66$  associations) entre les 3 profils alimentaires obtenus par ACP et les signaux RMN est présenté en Annexe 11. Aucune association n'atteignait le seuil  $FDR < 0.05$ . Les 15 plus fortes associations sont présentées dans le Tableau 28. Très peu de signaux ont pu être identifiés, notamment en NOESY1D. La créatine était associée positivement au profil alimentaire « occidental ». La choline et le glucose (superposition de signaux) et les lipoprotéines étaient associés négativement au profil alimentaire « sain ».

Tableau 28 Associations entre les signaux RMN et les profils alimentaires obtenus par ACP<sup>1</sup> chez les hommes – cohorte SU.VI.MAX, 1994-2007

Assignations	Déplacement chimique (ppm)	Séquence <sup>2</sup>	R	P	Profils alimentaires obtenus par ACP
<i>Créatine</i>	3,028	CPMG	0,29	0,0008	Composante 1 Profil alimentaire "occidental"
<i>Non identifié</i>	7,5097	CPMG	0,21	0,02	
<i>Non identifié</i>	0,1271	NOESY1D	0,28	0,001	Composante 2 Profil alimentaire "sain"
<i>Non identifié</i>	0,2471	NOESY1D	0,28	0,001	
<i>Non identifié</i>	6,2068	CPMG	-0,24	0,005	
<i>Non identifié</i>	0,0871	NOESY1D	0,24	0,005	
<i>Choline, glucose</i>	3,167	NOESY1D	-0,23	0,009	
<i>Non identifié</i>	0,2071	NOESY1D	0,22	0,01	
<i>Non identifié</i>	6,2478	CPMG	-0,22	0,01	
Lipoprotéines (CH2)	1,1671	CPMG	-0,21	0,01	
<i>Non identifié</i>	6,3398	CPMG	-0,21	0,02	
<i>Non identifié</i>	7,2707	NOESY1D	-0,27	0,002	
<i>Non identifié</i>	8,9511	NOESY1D	0,25	0,004	
<i>Non identifié</i>	6,1668	CPMG	0,23	0,006	
<i>Non identifié</i>	1,6311	CPMG	-0,21	0,01	

R, coefficient de corrélation. La validation 2D des signaux en italique est en cours sur la plateforme RMN.

<sup>1</sup> Les corrélations de Spearman ont été ajustées sur les caractéristiques suivantes : âge, IMC, statut tabagique, saison de prélèvement sanguin, nombre d'enregistrements alimentaires sur 24 heures et moyenne de l'apport énergétique quotidien pendant les 2 années suivant le prélèvement sanguin.

<sup>2</sup> 140 individus ont été analysés pour la séquence CPMG et 138 pour la séquence NOESY1D

### III. Discussion

Concernant l'alimentation habituelle globale, nous avons observé que les femmes ayant une meilleure adéquation aux recommandations nutritionnelles avaient des taux plasmatiques plus élevés en acide pipécolique, en acide 3-indolepropionique, en leucine et en créatine. Les hommes ayant une meilleure adéquation aux recommandations avaient également des taux plasmatiques plus élevés en leucine mais aussi en albumine. Par ailleurs, les femmes ayant une faible adéquation aux recommandations avaient des taux plasmatiques plus élevés en lysophosphatidylcholine (17 :1), en acylcarnitines C11 :1 et C11 :9, en 2-déoxyglucose, en glucose, en lipoprotéines, en lipides insaturés, en lactate et en thréonine. Un profil alimentaire « occidental » était associé positivement à la pipérine et à la créatine (chez les femmes et chez les hommes respectivement). De plus, un profil alimentaire « sain » était associé positivement à l'acide pipécolique et au dihydro-4-mercapto-3(2H) furanone plasmatiques chez les femmes. Par ailleurs, des corrélations négatives ont été observées entre le taux plasmatique de leucine et d'acide pipécolique et un profil alimentaire « occidental » chez les femmes. Le taux plasmatique de citrate était associé positivement au régime à base de produits sucrés chez les femmes. Enfin, des corrélations négatives ont été observées entre un profil alimentaire « sain » et les taux plasmatiques d'acides gras insaturés et d'acylcarnitine C9 :1 chez les femmes, de choline et de glucose chez les hommes et de lipoprotéines chez les femmes et chez les hommes.

De manière plus précise, des corrélations positives ont été observées entre la consommation d'alcool et les taux plasmatiques d'ethyl-β-D-glucopyranoside chez les femmes et de propylène glycol, de lipides, de choline et de glucose chez les hommes. Chez les femmes, la consommation de fruits et légumes était

corrélée positivement aux taux plasmatiques d'acide pipécolique et de phénylalanine (en H-RMN) et négativement aux taux de phénylalanine (en LC-MS) et d'acylcarnitine C9 :1. La consommation de lait et produits laitiers était associée à une diminution du dipeptide phénylalanine-phénylalanine plasmatique et une augmentation du glucose plasmatique chez les femmes et chez les hommes respectivement. Une consommation plus élevée de pain, céréales, pommes de terre et légumineuses était associée à une augmentation de la tyrosine plasmatique chez les femmes et de la choline, la créatine et le glycéryl plasmatiques chez les hommes. Elle était également associée à une diminution des taux d'isobutyrate, de valine et de lipoprotéines chez les femmes et de lactate, de lipides, de pyruvate, de choline et de glucose chez les hommes. Nous avons également observé des corrélations positives entre la consommation de matière grasse totale ajoutée et le glucose et la tyrosine plasmatiques chez les femmes et chez les hommes respectivement. La consommation de matière grasse végétale ajoutée était associée négativement à la lysoPC(17 :1). Enfin, des corrélations positives ont été observées entre la consommation de produits de la mer et de produits céréaliers complets et les taux plasmatiques de leucine et d'arginine et lysine respectivement ; alors que des corrélations négatives ont été observées entre la consommation de produits sucrés et de sel et les taux plasmatiques d'acide cholique et de lactate respectivement.

Parallèlement, dans notre objectif secondaire en LC-MS, nous avons pu mettre en évidence des associations positives entre plusieurs métabolites plasmatiques et la consommation d'aliments spécifiques, notamment pour le café (trigonelline et paraxanthine par exemple), les agrumes (proline bêtaïne et hydroxyproline bêtaïne), le lait (glucokinase) ou encore les fibres, les pommes, les poires et le yaourt (acide pipécolique) ou la charcuterie (N6,N6,N6-Triméthyl-L-lysine et pipérine).

De manière générale, les associations entre les métabolites plasmatiques et les groupes d'aliments semblent être plus fortes qu'avec l'alimentation globale. Cette observation peut s'expliquer par le fait qu'au sein d'un régime alimentaire, les aliments peuvent agir en synergie sur le métabolisme mais également de manière antagoniste. On peut également observer une certaine hétérogénéité de consommation dans le fait de manger « bien » ou « mal », notamment lorsque ce qualificatif est basé sur une combinaison de score. Par exemple, les femmes du 10<sup>ème</sup> décile du mPNNG-GS ont un score moyen de 10.1±0.7 mais elles peuvent avoir atteint ce score de plusieurs manières. Notamment la composante « boissons sucrées » ne permet pas de discriminer les deux groupes. Une analyse complémentaire de classification ascendante hiérarchique, présentée en Annexe 6 a notamment montré que 3 femmes ont été classées dans le 10<sup>ème</sup> décile malgré une consommation élevée de boissons sucrées ; leur bon score dans les autres composantes ayant permis une compensation au global. De plus, pour certaines composantes, les habitudes de consommation extrêmes étaient bien marquées entre les groupes, mais un grand nombre de sujets avaient le même niveau de consommation moyenne dans les deux groupes. Par exemple, dans le groupe des femmes du 1<sup>er</sup> décile, 48,8 % des participantes consommaient moins de 5 portions quotidiennes de fruits et légumes contre 6,3 % dans le groupe des femmes du 10<sup>ème</sup> décile et seulement 10 % contre 51,3 % consommaient plus de 7,5 portions quotidiennes. Toutefois, une forte proportion de participantes des deux groupes (41,3-42,5 %) avait une consommation intermédiaire de 5 à 7,5 portions par jour. Par ailleurs, pour certaines composantes du mPNNN-GS, le système de notation n'est pas linéaire (comme par exemple, pour le sel), ce qui peut compliquer l'interprétation biologique.

L'association entre la leucine et une meilleure adéquation aux recommandations a été observée à la fois chez les hommes et chez les femmes en H-RMN. De la même manière, les lipoprotéines étaient associées négativement au profil alimentaire « sain » chez les hommes et chez femmes. Les autres associations n'ont pas été répliquées entre les hommes et les femmes ni chez les femmes entre les analyses LC-MS et H-



RMN. L'impact de la méthode analytique est donc bien visible ici puisque chez les femmes le reste de la méthodologie était identique. Par ailleurs, en LC-MS nous avons observé une association négative entre la consommation de fruits et légumes et la phénylalanine alors que cette association était positive en H-RMN. L'origine de cette incohérence pourrait venir des différences dans le protocole analytique, dans les prétraitements réalisés, d'un signal impur ou encore d'une erreur dans l'annotation du signal (le signal de la phénylalanine est en cours de validation sur la plateforme RMN). La comparaison des résultats obtenus à la littérature semble donc limitée, d'autant plus que les études sont peu nombreuses sur cette problématique. Une standardisation des méthodes, aussi bien analytiques que statistiques, semblerait nécessaire. On remarque aussi qu'entre les deux méthodes d'estimation de l'alimentation globale utilisée (score mPNNS-GS et profils alimentaires obtenus par ACP) les résultats différaient au sein des mêmes populations d'étude et des mêmes méthodes analytiques. Bien que cela donne des résultats complémentaires, l'extrapolation à des populations et des protocoles analytiques différents ainsi que la comparaison à l'échelle internationale est donc à considérer avec prudence. Il est tout de même opportun de comparer nos résultats aux quelques études ayant un objectif similaire (présentées dans le Tableau 4 en Introduction générale (IV.A.2.)) et de les interpréter à la lumière des connaissances actuelles.

En utilisant soit le mPNNS-GS, soit les profils alimentaires obtenus par ACP, l'acide pipécolique (AP) était associé à une alimentation globale plus favorable à la santé. L'origine de l'AP plasmatique est incertaine. Les haricots secs et les légumes crucifères représentent la source alimentaire la plus riche signalée jusqu'à présent (43 et 12-22 mg/kg, respectivement (296)). Il a été identifié en utilisant la métabolomique comme le meilleur biomarqueur plasmatique candidat de l'ingestion de haricots secs dans une étude d'intervention de 4 semaines avec 250 g/jour de haricots secs et l'ingestion de haricots secs autodéclarée (0 à 300 g/jour) dans une population de 212 sujets (297). Cependant, il est peu probable que la consommation de haricots secs explique entièrement les différences observées pour l'AP plasmatique dans notre étude. En effet, l'apport déclaré de légumineuses n'était pas corrélé avec sa concentration plasmatique (coefficient de corrélation  $< |0,05|$ ) et ne différait pas entre les groupes 1<sup>er</sup> et 10<sup>ème</sup> déciles du mPNNS-GS (données non présentées), cependant les haricots secs ne sont pas le principal contributeur du groupe des légumineuses. L'AP peut également provenir de la dégradation de la lysine alimentaire par le microbiome intestinal (296). Dans quatre études utilisant une approche métabolomique exploratoire, les concentrations plasmatiques d'AP ont augmenté de façon significative chez les sujets ayant consommé un régime « sain » enrichi en céréales complètes pendant 12 semaines (159), un régime à faible indice glycémique pendant 4 semaines (163), ayant une activité physique quotidienne supérieure (298) ou ayant une masse musculaire supérieure (299). Dans notre étude, l'AP était plus spécifiquement positivement associée à la consommation de fruits et légumes, de fibres, de pommes et poires et de yaourt. Il était également corrélé à un apport en nutriments de fibres totales et solubles, potassium, glucides simples, vitamines C, E, B9, bêta-carotène et autres micronutriments (voir Annexe 6).

L'acide indolépropionique (AIP) augmentait également avec une meilleure adéquation aux recommandations alimentaires. La production de l'AIP est médiée par le microbiote intestinal et dérive du tryptophane alimentaire (300). Dans une étude finlandaise sur la prévention du diabète, un lien entre le taux d'AIP sérique élevé et un moindre risque de développer un diabète de type 2 a été observé chez des sujets présentant une tolérance au glucose réduite, avec une réplique dans deux cohortes indépendantes (301,302). Jennis et al. ont également rapporté des concentrations sanguines plus faibles en AIP chez les sujets obèses atteints de diabète de type 2 comparativement à des sujets minces ou après une chirurgie bariatrique (303). Un rôle préventif direct entre l'AIP et le diabète n'a pas encore été établi à notre connaissance. Cependant, il a été lié à une augmentation de la sécrétion d'insuline et à une

inflammation de bas grade plus faible (301). Dans des modèles animaux, la supplémentation alimentaire d'AIP réduisait de façon significative la glycémie à jeun (304) et a permis de réduire la perméabilité intestinale en augmentation suite à une alimentation riche en graisses (303). En ce qui concerne les associations alimentaires, l'AIP sérique a été corrélé à la consommation de fibres alimentaires et de céréales complètes dans les populations nordiques (301,302), de bananes, de pommes et de poires dans une population de femmes britanniques (305), ainsi qu'à un régime « Baltic Sea Diet », « sain » et aux fibres alimentaires chez les hommes fumeurs finlandais de l'étude ATBC (171). Il a également été inversement corrélé à la consommation d'œufs et de viande rouge dans une population américaine (306) et augmentait suite à un régime à faible indice glycémique de 4 semaines (163). Dans notre étude, l'AIP plasmatique était corrélé avec le bêta-carotène plasmatique (voir Annexe 6), ce qui pourrait refléter une association positive avec la consommation de fruits et légumes (307). Nous pouvons émettre l'hypothèse que l'augmentation des taux plasmatiques d'AIP et d'AP dans l'échantillon de femmes de la population SU.VI.MAX présentant une meilleure adéquation aux recommandations nutritionnelles reflèterait une modification favorable du microbiote intestinal par l'alimentation, possiblement liée à la consommation de fruits et légumes, de fibres alimentaires, de yaourt ou autres composants alimentaires végétaux à activité prébiotique comme les polyphénols (308). Après validation dans des études indépendantes à grande échelle, l'AIP et l'AP pourraient être proposés comme des biomarqueurs candidats d'une alimentation saine et/ou d'un mode de vie sain.

La lysoPC(17:1) est une lyso-phosphatidylcholine ayant un nombre impair d'atomes de carbone, non synthétisée par le corps humain. Les lysoPC, aussi appelées lysolécithines, résultent de l'hydrolyse partielle des phosphatidylcholines, qui élimine l'un des groupes d'acides gras. Elles sont présentes dans l'alimentation, notamment dans les féculents (elles représentent environ 70% des lipides de l'amidon d'avoine (309)) et les graines de soja. Les lysoPC sont aussi utilisées comme émulsifiant en alimentation animale, notamment des volailles, afin d'augmenter la digestibilité des nutriments et en particulier des lipides (310). Les phospholipides contenant de la choline ont été associés à des apports élevés en beurre et en matières grasses dans les produits laitiers (174,305), ce qui semble cohérent avec les associations observées dans notre étude entre la lysoPC(17:1) et un faible score mPNNS-GS et sa corrélation positive avec les apports en acides gras saturés (voir Annexe 6) et négative avec la consommation de matière grasse végétale ajoutée. Par ailleurs, des sujets obèses ayant consommé des régimes riches en fructose pendant 4 semaines ont vu leur taux de lysoPC plasmatique augmenté (311). Dans notre étude, la choline plasmatique était associée positivement à la consommation d'alcool et à la consommation de pain, céréales, pommes de terre et légumineuses. La superposition de signaux de la choline et du glucose était également associée négativement à cette dernière et à un profil alimentaire « sain ». La choline est un nutriment essentiel apporté par divers aliments. Dans les populations européennes, les principales sources alimentaires sont la viande, le lait, les céréales, les œufs et le poisson (312). La choline peut également être formée par biosynthèse de novo *via* la méthylation de la phosphatidylethanolamine, catalysée par la phosphatidylethanolamine N-méthyltransférase. Le taux circulant de choline a été associé à des maladies liées au mode de vie, y compris les maladies cardiovasculaires (313,314).

Dans notre étude, la consommation d'alcool était associée positivement à des taux plasmatiques plus élevés en ethyl- $\beta$ -d-glucopyranoside et en propylène glycol chez les femmes et les hommes respectivement. Le premier appartient à la classe des composés organiques o-glycosylés dans lesquels un sucre est lié à un autre groupe par l'intermédiaire d'une liaison O-glycosidique (315) (l'autre groupe pouvant être un alcool, un phénol, une substance à fonction aminée ou à fonction thiol, un stéroïde, etc). L'ethyl- $\beta$ -d-glucopyranoside est un constituant des écorces d'agrumes, du cortex racinaire frais du

manioc et d'autres sous-espèces végétales. Il se trouve dans de nombreux aliments, comme les légumes-racines, les agrumes, les boissons alcoolisées et les fruits (315). Le propylène glycol, aussi appelé 1,2-dihydroxypropane ou méthyl glycol, est généralement issu de la pétrochimie. C'est un émoulliant et émulsifiant (E 1520) que l'on trouve dans les cosmétiques, les médicaments et les aliments (316). D'après l'EFSA, l'exposition maximum chez l'homme serait de 9.5mg/kg par jour et l'apport journalier acceptable a été fixé à 25mg/kg par jour (316). Le propylène glycol est également utilisé dans les cigarettes électroniques, cependant celles-ci n'étant pas d'usage au moment du prélèvement sanguin des participants n'expliquent probablement pas les variations observées.

Nous avons observé un taux plasmatique de pipérine associé positivement au profil alimentaire « occidental » et à la consommation de charcuterie. La pipérine, un alcaloïde actif exogène sans origine endogène signalé jusqu'à présent, est présente dans le poivre noir et le poivrier long (317) et est utilisée comme additif alimentaire dans l'alimentation animale, en particulier pour la volaille (318,319). L'exposition humaine à la pipérine par cette dernière source a été estimée à 0,93 µg/poids corporel métabolique par jour (319). Par ailleurs, la charcuterie, contenant du poivre, fait partie du profil alimentaire « occidental ».

Nous avons également observé une association entre deux acylcarnitines à moyenne chaîne et un faible score mPNNS-GS. Les acylcarnitines permettent le transport des acides gras impliqués dans la production d'énergie (320). La carnitine corporelle totale provient principalement de la carnitine alimentaire (que l'on trouve en majorité dans les aliments d'origine animale) et de sa synthèse endogène à partir de la lysine et de la méthionine (75%) (320). Les acylcarnitines sont des substrats et des produits de la bêta-oxydation mitochondriale (174) et les acylcarnitines C9:1 et C11:1 sont exclusivement dérivées du métabolisme des acides gras. Une étude a montré des associations positives entre des acylcarnitines à courte chaîne et un profil alimentaire « occidental » et entre des acylcarnitines à moyenne et longue chaîne et un profil alimentaire « sain » (167). Une autre étude a mis en évidence des associations entre 6 acylcarnitines et la consommation de viande (170). La concentration sanguine de C9:1 a été associée négativement à la consommation de fruits et légumes et positivement à l'ingestion de beurre ainsi qu'à un gène impliqué dans l'oxydation des lipides (174,305,321).

L'interprétation des variations observées d'acides aminés et dérivés, de protéines, de glucose, de lactate, de citrate, de lipides et de lipoprotéines est délicate puisqu'ils sont impliqués dans de nombreuses voies métaboliques, éventuellement modulées par divers facteurs et combinant des origines exogènes et endogènes. Dans notre étude, la leucine plasmatique a été associée positivement à une meilleure adéquation aux recommandations nutritionnelles et négativement à un profil alimentaire « occidental ». Un taux plasmatique plus élevé de leucine était également associé à la consommation de produits de la mer ; une composante de l'alimentation apportant des points positifs au score mPNNS-GS et peu présente dans le profil alimentaire « occidental ». Par ailleurs, une autre étude a montré une augmentation du taux plasmatique de leucine chez les consommateurs de poissons et les végétariens (170). Cependant, la leucine plasmatique a aussi été associée à un profil alimentaire « occidental » (167) et à un régime riche en viande, saucisse, sauces, œuf, glace et faible en noix, céréales, champignon et légumineuses (176). La leucine est un acide aminé essentiel et, avec les autres acides aminés branchés (l'isoleucine et la valine), constituent environ 40% des acides aminés essentiels libres du plasma sanguin (322). La leucine semble jouer un rôle important dans la réponse immunitaire adaptative *via* l'activation des lymphocytes T. Elle est aussi connue comme activateur de la voie mTORC1 (323). La leucine est généralement plus abondante dans les

protéines animales que végétales et la L-leucine est un additif alimentaire, utilisé comme exhausteur de goût (E641).

Nous avons observé une association positive entre la créatine plasmatique et une meilleure adéquation aux recommandations chez les femmes et à un profil alimentaire « occidental » chez les hommes. Dans la littérature, la créatine circulante a plutôt été associée à des régimes « sains ». En effet, une étude d'observation a trouvé une association positive avec le régime méditerranéen dans l'urine (172) et trois études d'intervention ont observé une augmentation de son taux suite à un régime en accord avec les recommandations nutritionnelles (160) et à un régime méditerranéen (164) dans les urines ainsi qu'à faible index glycémique dans le plasma (163). Dans notre étude, elle a également été associée positivement à la consommation de pain, céréales, pommes de terre et légumineuses, mais son signal était superposé avec celui du glycéryl, rendant impossible l'attribution de l'association observée.

La superposition des signaux de l'arginine et la lysine était associée positivement aux produits céréaliers complets mais dans la littérature, les associations entre ces métabolites et les consommations alimentaires sont très mitigées. En effet, la lysine a été associée positivement à un régime enrichi en produits complets (159) mais aussi à un régime riche en viande, saucisse, sauces, œuf, glace et faible en noix, céréales, champignon et légumineuses dans une étude (176), à un régime riche en viande dans une autre étude (175), mais aussi à un régime en accord avec les recommandations (160), à un régime riche en poisson et à un régime végétarien (170) et négativement au régime méditerranéen (172). D'autre part, le taux d'arginine a également été associé à un régime riche en viande en observationnel (175), à un régime enrichi en produits complets (159) et à faible index glycémique (163).

Le taux circulant de phénylalanine, qui était associé à la consommation de fruits et légumes dans notre étude, a également été associé à un régime omnivore (165), à un profil alimentaire « occidental » (167) et à un régime riche en viande, saucisse, sauces, œuf, glace et faible en noix, céréales, champignon et légumineuses (176).

Nous avons également observé des associations positives entre la tyrosine et la consommation de pain, céréales, pommes de terre et légumineuses et l'ajout de matière grasse. Dans la littérature, la tyrosine circulante a entre autres été associée à la consommation de lait écrémé (175) et à un régime riche en viande, saucisse, sauces, œuf, glace et faible en noix, céréales, champignon et légumineuses (176).

Dans notre étude, le glucose plasmatique a été associé avec une moindre adéquation aux recommandations ainsi qu'à la consommation d'alcool, de lait, et de matières grasses totales. Chez 1717 hommes, il a été montré que les personnes en surpoids et consommant plus de 30g d'éthanol par jour, ou que les hommes obèses qui consommaient même modérément de l'éthanol avaient plus de risque d'avoir une altération du glucose à jeun ou de développer un diabète de type 2 (324). Par ailleurs, une étude d'intervention de 3 mois a montré qu'un régime riche en graisse et en énergie augmentait le taux de glucose plasmatique à jeun (325). Une autre étude d'intervention a montré qu'une moindre adéquation aux recommandations nutritionnelles augmentait le taux urinaire de glucose (160) et une étude d'observation a montré une association positive entre ce dernier et une moindre adéquation au régime méditerranéen (172).

Nous avons par ailleurs observé une association positive entre le citrate plasmatique et la consommation de produits sucrés. Le citrate circulant a été associé positivement au régime méditerranéen dans une étude

d'intervention (164), au régime lactovégétarien (152), à un régime « sain » (168) et aux produits animaux (175) dans trois études d'observation ; des résultats encore très mitigés.

La H-RMN n'a pas permis d'identifier spécifiquement les composés lipidiques associés à l'alimentation. Nous avons notamment observé une association positive entre la consommation d'alcool et les lipides plasmatiques ainsi qu'une association négative entre les lipides insaturés plasmatiques et une meilleure adéquation aux recommandations ou un profil alimentaire « sain ». Deux études d'intervention ont montré qu'un régime « sain » augmentait le taux plasmatique d'acides gras n-3 et diminuait le taux d'acides gras n-6 et n-7 (158), et augmentait le taux d'acides gras polyinsaturés (159). Des investigations supplémentaires sur les dosages à l'inclusion des participants de SU.VI.MAX ont montré que les signaux des lipoprotéines mis en évidence chez les femmes en NOESY1D (à 0.8063 et à 1.163 ppm) étaient corrélés positivement aux taux de triglycérides sanguins ( $r=0.60$  et  $r=0.48$  respectivement), de cholestérol ( $r=0.48$  et  $r=0.28$  respectivement), d'apolipoprotéine A1 ( $r=0.42$  et  $r=0.43$  respectivement) et aux taux d'apolipoprotéine B pour le signal à 0.8063 ( $r=0.29$ ) mais pas de la lipoprotéine(a) ( $r=0$  et  $r=-0.05$  respectivement). En CPMG, le signal des lipoprotéines mis en évidence chez les femmes (à 1.663 ppm) était corrélé positivement aux taux sanguins de triglycérides ( $r=0.46$ ) et d'apolipoprotéine A1 ( $r=0.29$ ) mais pas de la lipoprotéine(a) ( $r=-0.03$ ). Chez les hommes, le signal des lipoprotéines à 1.1671 ppm (CPMG) était corrélé aux taux sanguins de cholestérol, de triglycérides, d'apolipoprotéine A1 et d'apolipoprotéine B ( $r=0.34$ ,  $0.48$ ,  $0.24$  et  $0.29$  respectivement) mais pas de la lipoprotéine(a) ( $r=-0.09$ ).

Notre objectif secondaire, s'intéressant à des groupes d'aliments spécifiques a permis de répliquer des associations déjà connues comme par exemple, la trigonelline, la paraxanthine, le cyclo(isoleucyl-prolyl), l'atractyligenin glucuronide pour la consommation de café (326) et la proline bêtaïne et l'hydroxyproline pour la consommation d'agrumes (135) ; renforçant leur potentiel de biomarqueurs de consommation. D'autres associations *a priori* nouvelles comme la glucokinase et la consommation de lait ; la N6,N6,N6-Triméthyl-L-lysine et la pipérine et la charcuterie ; et l'acide pipécolique et les consommations de pommes et de poires, de yaourt et de fibres sont à valider dans d'autres études indépendantes. Ces résultats pourraient permettre l'amélioration de l'évaluation de l'exposition nutritionnelle dans les études épidémiologiques grâce à des marqueurs biologiques. Cependant, certains métabolites ne sont pas spécifiques d'un seul aliment comme l'acide pipécolique et mériteraient d'être associés à d'autres biomarqueurs pour une meilleure spécificité.

Trop peu d'études voire aucune étude ont observé des variations des autres métabolites que nous avons associés à l'alimentation dans le cadre de cette thèse, ne permettant pas de comparer nos résultats. En outre, dans toutes les études métabolomiques non ciblées, y compris la nôtre, une proportion élevée de métabolites demeure non identifiée. De plus, le processus d'identification révèle souvent des structures chimiques pour lesquelles on ne dispose que de peu d'informations sur les origines possibles et leurs implications pour la nutrition et la santé. Un effort majeur d'enrichissement des bases de données est essentiel pour tirer pleinement profit de la métabolomique exploratoire. Dans tous les cas, toutes ces associations observées nécessitent une validation dans des études indépendantes de large ampleur ainsi que des études d'intervention afin d'élucider de potentielles associations causales.

Quelques autres variations de métabolites ont été observées de manière assez récurrente dans des études s'intéressant au reflet de l'alimentation globale sur les profils métabolomique que nous avons pu manquer faute de puissance ou d'identification comme par exemple, le triméthylamine-N-oxyde (TMAO) (150,156,157,160,163–166,172), l'hippurate (152,159–164,168,172,327) ou encore la

phénylacetylglutamine (160,164,168,172,328). Cependant, malgré cette récurrence, la cohérence entre les études des associations de ces métabolites avec l'alimentation encore limitée, et plus largement, des signatures métabolomiques robustes reflétant l'alimentation globale ne semblent pas encore établies dans la littérature. En plus du faible nombre d'études ayant investigué cette problématique, la disparité des résultats s'explique probablement par les nombreux paramètres variables cités plus haut, à savoir les méthodes d'estimation de l'alimentation globale, les populations et design d'études, des analyses et protocoles métabolomiques, des biofluides analysés, des noms pour même métabolite, des analyses statistiques, des facteurs de confusion pris en considération, etc.

Par ailleurs, nous remarquons que parmi tous les métabolites associés à l'alimentation globale que nous avons observés, très peu étaient des métabolites purement exogènes. Cela contraste avec plusieurs études antérieures qui ont identifié des métabolites spécifiques d'aliments (notamment du poisson, des céréales complètes ou de baies), associés à l'alimentation globale (159,160). Toutefois, ces études étaient des études d'interventions dans un environnement contrôlé et surveillé. Dans de telles études, bien que les habitudes alimentaires puissent être complexes, elles sont plutôt homogènes au sein des groupes expérimentaux. Dans notre étude, les sujets sont en conditions réelles, et notre design d'étude a permis de maximiser la variabilité intergroupe mais pas de minimiser la variabilité intragroupe, comme discuté plus haut. Dans de telles études d'observation, l'utilisation de combinaisons de biomarqueurs spécifiques de l'apport alimentaire afin de discriminer les sujets selon leurs propres habitudes alimentaires semble donc difficile. Par conséquent, nos résultats suggèrent que les signatures métabolomiques des régimes alimentaires globaux en observationnel seraient plutôt composés des métabolites endogènes et issus du métabolisme microbien reflétant l'impact à long terme du régime alimentaire sur les voies métaboliques et sur le microbiome intestinal. Ce type d'études, ainsi que des études d'intervention à long terme représentant au maximum les conditions réelles, permettraient de mieux capter l'impact du régime alimentaire global par ses effets bénéfiques ou néfastes à long terme sur l'état nutritionnel et métabolique des individus apportant donc des éléments sur les relations nutrition/santé.

### Objectif 3 - Identifier des signatures métabolomiques plasmatiques liées à la fois à l'alimentation et au risque à long terme de cancers du sein (objectif 3.I.) et de la prostate (objectif 3.II.)

---

Ce troisième objectif visait à explorer les données de métabolomique à partir de prélèvements plasmatiques d'individus *a priori* en bonne santé afin de mettre en évidence de potentiels biomarqueurs du risque de développer un cancer du sein ou de la prostate dans la décennie qui suit, mais étant également liés à l'alimentation. En effet, l'alimentation faisant partie des facteurs de risque modifiables de ces cancers, identifier une signature du risque sur laquelle il est possible d'agir par des interventions nutritionnelles aurait un impact très important en termes de santé publique et de prévention personnalisée. L'intérêt ici serait de sélectionner un ensemble restreint de biomarqueurs liés à l'alimentation permettant de discriminer les sujets à plus haut risque de développer ces cancers. Ces potentiels biomarqueurs peuvent toujours appartenir aux trois grandes classes : biomarqueurs d'exposition (notamment nutritionnelle), d'effet (impact de la nutrition sur le métabolisme) ou de susceptibilité. Afin de pouvoir répondre à cette problématique, nous avons cherché à prendre en compte diverses composantes de l'alimentation, qu'elles soient spécifiques ou globales avec diverses manières de l'estimer. Par ailleurs, puisque le but principal de cet objectif était l'aspect prévention, nous avons choisi d'utiliser une méthode de sélection de variables, afin non seulement d'identifier des métabolites qui prédisent conjointement le risque, mais également de sélectionner un nombre limité de potentiels biomarqueurs. Au-delà de l'intérêt potentiel en prévention, ces découvertes peuvent également apporter des éléments pour la compréhension de l'impact de l'alimentation sur la santé et notamment l'étiologie des cancers du sein et de la prostate. Cependant, cette sélection restreinte de métabolites n'a pas pour but de mettre en évidence un ensemble de voies métaboliques permettant d'approfondir la compréhension mécanistique des liens nutrition/cancer. Par ailleurs, ces méthodes de sélection de variables permettent de construire des modèles prédictifs sans que les prédicteurs soient forcément associés de manière causale.

#### I. Signature du risque de cancer du sein liée à l'alimentation

##### A. Analyses en LC-MS

Cette étude a fait l'objet d'un article soumis pour publication, présenté en Annexe 12.

*Lécuyer L, Dalle C, Lefevre-Arbogast S, Micheau P, Lyan B, Rossary A, Demidem A, Petera M, Lagree M, Centeno D, Galan P, Herberg S, Samieri C, Assi N, Ferrari P, Viallon V, Deschasaux M, Partula V, Srour B, Latino-Martel P, Kesse-Guyot E, Druetne-Pecollo N, Vasson MP, Durand S, Pujos-Guillot E, Manach C, Touvier M. **Diet-related metabolomic signature of long-term breast cancer risk using penalized regression: an exploratory study in the SU.VI.MAX cohort***

#### 1. Méthodes

La sélection de la population d'étude pour ces analyses a été détaillée dans la partie Méthodologie générale (III.A.1.). Il s'agit de la même étude cas-témoins nichée dans la population SU.VI.MAX que celle de l'objectif 1 (215 cas de cancer du sein incidents et 1 témoin apparié par cas).

Les échantillons de plasma des individus prélevés à l'inclusion ont été analysés en UPLC-MS en mode positif et négatif. Le protocole d'analyse était le même que celui de l'objectif 1 (I.B.1.). Après prétraitements, 528 et 690 ions ont été utilisés dans les analyses statistiques.

La description de la population d'étude a été réalisée dans l'objectif 1. Dans un premier temps, les associations entre les ions et l'alimentation ont été estimées grâce à des analyses de corrélation partielle de Spearman. Quatre méthodes d'estimation de l'exposition nutritionnelle ont été utilisées. Deux scores complémentaires reflétant l'adéquation aux recommandations nutritionnelles ont été calculés (méthodes *a priori*) : le score mPNNS-GS incluant 12 composantes de groupes d'aliments ainsi que le DQI-I incluant des composantes de variété, d'adéquation, de modération et d'équilibre global. Le calcul de ces scores a été détaillé dans la partie Méthodologie générale (III.B. et V.). Par ailleurs, l'alimentation globale a également été estimée grâce à une ACP avec rotation varimax sur la population générale de SU.VI.MAX ayant des données alimentaires disponibles pendant les 2 ans suivant l'inclusion (plus de 3 enregistrements de 24h, N=8112) permettant de mettre en évidence des profils alimentaires caractérisant notre population d'étude. Les deux premières composantes ont été sélectionnées. Enfin, nous avons utilisé les 74 groupes d'aliments détaillés dans la partie Méthodologie générale (V.), spécifiques et généraux. Les modèles de corrélation ont été ajustés sur l'âge (continu), l'IMC (continu), le statut ménopausique, le statut tabagique (fumeuses, anciennes fumeuses, non-fumeuses), la saison de prélèvement sanguin (3 saisons) et le nombre d'enregistrements de 24h validés et la moyenne de l'apport énergétique (sans la composante alcool pour les corrélations avec la consommation d'alcool) pendant les 2 ans suivant le prélèvement sanguin. La correction de test multiple de Benjamini Hochberg a été appliquée. 68 participantes ont été exclues pour les analyses de corrélation avec les groupes alimentaires et les composantes de l'ACP en raison de données alimentaires manquantes. 4 participantes supplémentaires ont été exclues pour les corrélations avec le DQI-I et 72 avec le mPNNS-GS en raison d'autres données manquantes comme par exemple, l'activité physique.

Les individus n'ayant pas de données métabolomiques disponibles issues à la fois du mode positif et négatif ont été exclus pour les analyses statistiques suivantes puisque tous les ions ayant été extraits ont été analysés conjointement. Les données de 200 cas et 200 témoins ont donc été analysées. Parmi les ions associés à l'alimentation, les meilleurs prédicteurs du risque de cancer du sein ont été sélectionnés grâce à une régression logistique pénalisée de type Elastic Net (détails présentés dans la partie Méthodologie générale VI.). Les modèles incluaient les potentiels facteurs de confusion suivant : âge, IMC, saison de prélèvement sanguin, statut ménopausique, statut tabagique, groupe d'intervention SU.VI.MAX, taille, activité physique, niveau d'éducation, consommation d'alcool, antécédents familiaux de cancer du sein, nombre d'enfants et utilisation de la thérapie de substitution hormonale de la ménopause. Les associations entre les ions sélectionnés et l'exposition alimentaire ont été vérifiées par pairs-plot. Afin de déterminer la stabilité des ions sélectionnés, une méthode de ré-échantillonnage par bootstrap a été réalisée en répétant la procédure de sélection 1000 fois et en enregistrant le pourcentage de fois où chaque ion était sélectionné (détails présentés dans la partie Méthodologie générale VI.). Le sens des associations entre les ions sélectionnés et le risque de cancer du sein a été établi grâce à des modèles de régression logistique sur l'ensemble de ces ions et les facteurs de confusion précédents. Bien que ces modèles génèrent des p-values, elles ne peuvent pas être utilisées pour l'inférence étant donné l'étape de sélection précédente par Elastic Net. Des analyses complémentaires ont été réalisées avec des modèles de régression logistique sur chaque ion ajustés sur les facteurs de confusion afin d'estimer leurs associations individuelles. Un résumé des analyses est disponible en Annexe 12.



Des estimations de la performance prédictives ont été calculées par les deux méthodes détaillées dans la partie Méthodologie générale (VI.) ; grâce à l'AUC par validation croisée simple ou double validation croisée. Les modèles de régression logistique utilisés pour calculer ces AUC étaient ajustés sur les facteurs de confusion cités plus haut.

## 2. Résultats

Les deux premières composantes de l'ACP permettaient d'expliquer 17% de la variabilité totale des données alimentaires de la population générale de SU.VI.MAX. La première composante était caractérisée par une consommation plus élevée de boissons alcoolisées, de matière grasse animale, de pain, de fromage, de pâtes, de riz et de céréales, de pommes de terre et de tubercules, de charcuterie, de viande rouge et d'abats, de produits salés et sucrés, de sauces, de matière grasse végétale et de volaille. Elle a donc été considérée comme la représentation d'un profil alimentaire « occidental ». La deuxième composante était caractérisée par une consommation plus élevée de fruits, de fruits secs, de légumes, de céréales pour petit-déjeuner, de matière grasse végétale, de céréales complètes et de yaourt – tendant globalement à représenter plutôt un profil alimentaire « sain ».

Parmi les 1218 ions analysés, 595 étaient associés à l'alimentation ( $p < 0.05$ , coefficient de corrélation  $r > |0.15|$ ) et ont été ensuite analysés en régression logistique pénalisée (Elastic Net). L'ensemble de ces associations est présenté en Annexe 12 (N=1085 associations). 14 ions ont été sélectionnés par Elastic Net. Parmi ceux-là, 2 ont été identifiés (niveau 1) : la pipérine (M286T989) et l'AcétylTributylCitrato (ATBC) (M425T1158), 2 ont été annotés de manière présumée : l'acide 2-amino-cyano-butanoïque (M153T116) et le pregnene-triol sulfate (M413T967) ; les autres n'ont pas pu être identifiés (M192T181, M265T186, M335T864, M364T125, M97T134, M166T144, M201T1091, M415T1344, M475T122, et M587T121). La pipérine était notamment positivement associée à la consommation totale d'alcool ( $r=0.19$ ) et avec le profil alimentaire « occidental » ( $r=0.22$ ), l'ATBC était positivement associé à la consommation de café ( $r=0.17$ ), l'acide 2-amino-cyano-butanoïque était associé négativement à la consommation de gâteaux et biscuits ( $r=-0.16$ ) et le pregnene-triol sulfate était associé positivement à la consommation totale d'alcool ( $r=0.17$ ). Les composés non identifiés étaient associés à diverses expositions alimentaires comme les pâtes et les céréales, les produits sucrés, la charcuterie, les tomates, les agrumes et le fromage à pâte pressée cuite (voir détails en Annexe 12). Le Tableau 29 présente les associations entre les 14 ions sélectionnés et le risque de cancer du sein incluant les ions ensemble dans le modèle de régression logistique ou chaque ion séparément.

Des taux plasmatiques plus faibles en pipérine et 6 composés non identifiés (M335T864, M364T125, M166T144, M415T1344, M475T122 and M587T121) et des taux plus élevés en acide 2-amino-butanoïque, ATBC, pregnene-triol sulfate et 4 composés non identifiés (M192T181, M265T186, M97T134 and M201T1091) ont été observés chez les femmes ayant développé un cancer du sein pendant le suivi.

Tableau 29 Associations entre les 14 ions sélectionnés par Elastic Net et le risque de cancer du sein<sup>1</sup> – cohorte SU.VI.MAX, 1994-2007

Ions (annotation <sup>2</sup> )	Masse/temps de rétention	Mode de détection	Tous les ions sélectionnés ensemble		Ions sélectionnés un par un	
			OR [IC95%]	P	OR [IC95%]	P
M153T116 (niveau 2: <b>acide 2-amino-cyano-butanoïque</b> )	153,0162/1,93	ESI Positif	1,23 [0,93-1,62]	0,1	1,08 [1,03-1,14]	0,002
M192T181 (niveau 4: formule moléculaire ; C <sub>9</sub> H <sub>6</sub> O <sub>4</sub> N)	192,0305/3,01	ESI Positif	1,37 [1,05-1,79]	0,02	1,08 [1,03-1,13]	0,003
M265T186 (niveau 4 : non identifié)	265,3932/3,1	ESI Positif	1,28 [0,93-1,76]	0,1	1,09 [1,04-1,14]	0,0008
M286T989 (niveau 1: <b>pipérine</b> )	286,143/16,48	ESI Positif	0,76 [0,58-0,99]	0,04	0,94 [0,89-0,99]	0,01
M335T864 (niveau 4 : non identifié)	335,2404/14,4	ESI Positif	0,77 [0,61-0,99]	0,04	0,94 [0,89-0,98]	0,009
M364T125 (niveau 4 : NaCOOH adduit)	363,9292/2,09	ESI Positif	0,74 [0,58-0,94]	0,01	0,93 [0,89-0,98]	0,007
M425T1158 (niveau 1 : <b>Acetyl Tributyl Citrate ATBC</b> )	425,2172/19,31	ESI Positif	1,39 [1,08-1,79]	0,01	1,07 [1,02-1,12]	0,006
M97T134 (niveau 4 : NaCOOH adduit)	96,9218/2,24	ESI Positif	1,23 [0,94-1,60]	0,1	1,10 [1,04-1,15]	0,0003
M166T144 (niveau 4 : non identifié)	166,0387/2,4	ESI Négatif	0,80 [0,62-1,03]	0,09	0,94 [0,89-0,98]	0,01
M201T1091 (niveau 4 : non identifié)	201,149/18,18	ESI Négatif	1,21 [0,94-1,57]	0,1	1,10 [1,04-1,15]	0,0003
M413T967 (niveau 2 : <b>Pregnene-triol Sulfate</b> )	413,2002/16,11	ESI Négatif	1,40 [1,08-1,84]	0,01	1,08 [1,03-1,14]	0,004
M415T1344 (niveau 4 : non identifié)	415,2078/22,4	ESI Négatif	0,83 [0,64-1,07]	0,2	0,92 [0,87-0,96]	0,0006
M475T122 (niveau 4 : non identifié)	474,7265/2,04	ESI Négatif	0,89 [0,70-1,14]	0,4	0,93 [0,88-0,97]	0,003
M587T121 (niveau 4 : non identifié)	586,6522/2,02	ESI Négatif	0,69 [0,54-0,89]	0,004	0,93 [0,89-0,98]	0,005

<sup>1</sup> Le modèle de régression logistique principale prenant en compte les 14 ions sélectionnés en même temps. Des modèles de régression logistique complémentaires considéreraient séparément les ions sélectionnés (un modèle pour chacun des 14 ions). Tous ces modèles ont été ajustés sur l'âge (continu), l'IMC (continu), la saison de prélèvement sanguin (3 saisons), le statut ménopausique, le statut tabagique (fumeuses, anciennes fumeuses et non-fumeuses), taille (continu), activité physique (faible, moyenne, intense), niveau de scolarité (primaire, secondaire, supérieure), consommation d'alcool (continu), utilisation de l'hormonothérapie substitutive pour la ménopause, nombre d'enfants et antécédents familiaux de cancer du sein à la prise de sang et groupe d'intervention SU.VI.MAX (placebo versus supplément). Des tests de tendance linéaire ont été effectués à l'aide des variables en continu. Les OR sont été présentés pour une augmentation de 1 écart-type de la variable continue.

<sup>2</sup> Niveaux de confiance de l'identification : niveau 1 = composé identifié et validé par standard, niveau 2 = composé identifié de manière présumée, niveau 3 = classe du composé annotée de manière présumée et niveau 4 = composé inconnu.

Concernant la stabilité des 14 ions sélectionnés, leur fréquence de sélection sur les 1000 jeux de données ré-échantillonnés était supérieure à 50% sauf pour l'acide 2-amino-cyano-butanoïque (27%). En outre, la fréquence de sélection de la pipérine, de l'ATBC, du pregnene-triol sulfate et de 8 composés non identifiés étaient supérieure à 70% (voir Figure 24 présentant les fréquences de sélection des ions >40%). Des recherches complémentaires ont été réalisées pour l'annotation des ions sélectionnés à plus de 70% ne faisant pas partie de la sélection initiale (14 ions) c'est-à-dire pour les ions M114T115, M498T1040, M230T171, M423T120, M454T1064, M196T953, M485T1437, M491T120 et M581T123, mais aucun n'a pu être identifié. Les associations entre ces ions et les expositions nutritionnelles ainsi que les associations avec le risque de cancer du sein sont présentées en Annexe 12.

La performance prédictive du modèle incluant les 14 ions sélectionnés et les facteurs de confusion, estimée par validation croisée simple, atteignait une AUC moyenne de  $0.76\pm 0.03$ . Par ailleurs, la performance prédictive de la procédure par double validation croisée atteignait une AUC de  $0.60\pm 0.02$ .

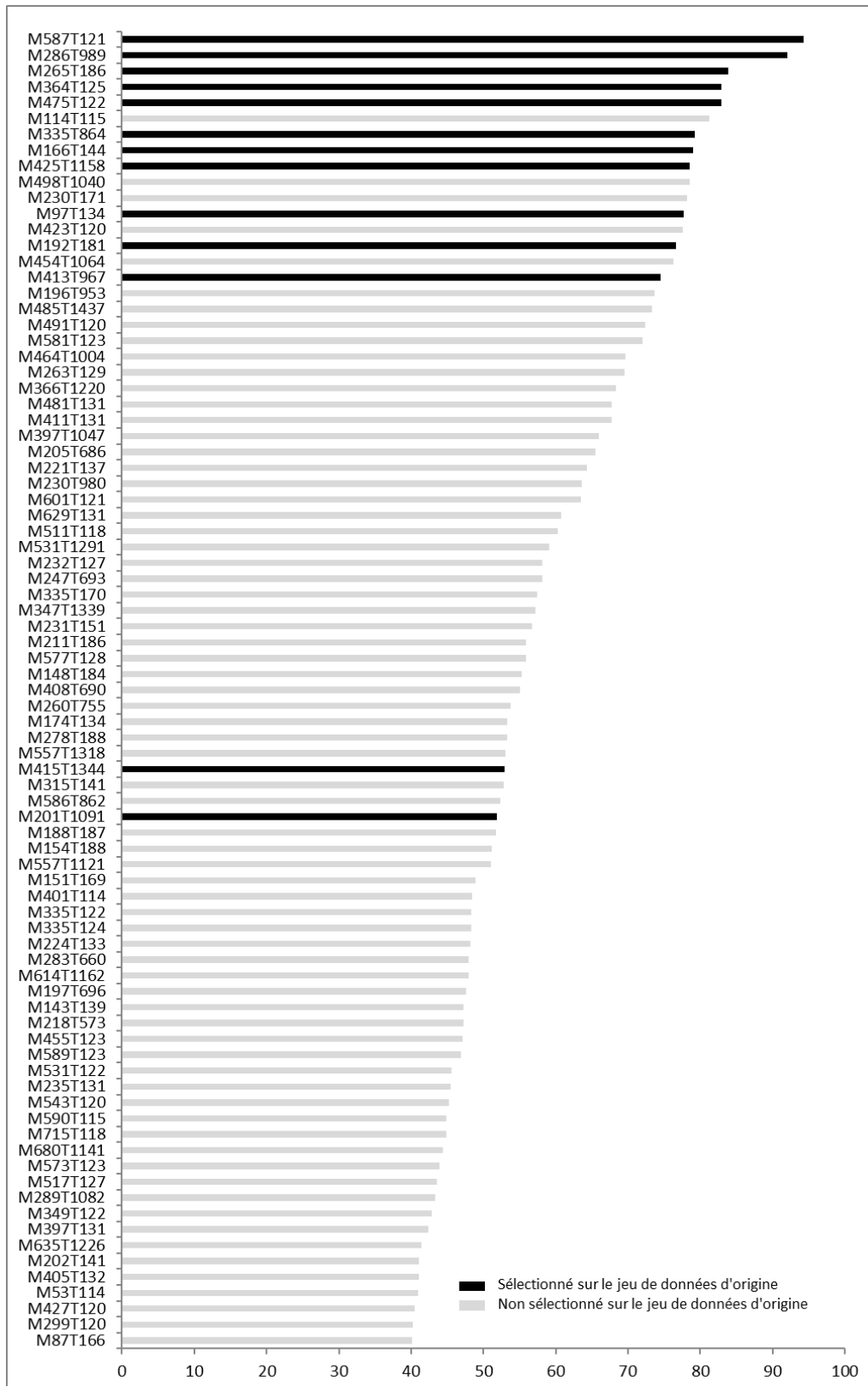


Figure 24 Fréquence de sélection  $\geq 40\%$  des ions pour le risque de cancer du sein (% sur 1000 jeux de données ré-échantillonnés)

## B. Analyses en H-RMN

### 1. Méthodes

La sélection de la population d'étude pour ces analyses a été détaillée dans la partie Méthodologie générale (III.A.1.). Il s'agit de la même étude cas-témoins nichée dans la population SU.VI.MAX que celle de l'objectif 1 (228 cas de cancer du sein incidents et 2 témoins appariés par cas). Après prétraitements des données 114 signaux RMN et 96 ont été obtenus pour la séquence NOESY1D et CPMG respectivement.

Les mêmes analyses statistiques que pour la LC-MS ont été réalisées. 84 et 105 participantes ont été exclues pour les analyses de corrélation avec les groupes d'aliments, 91 et 111 avec le DQI-I et 206 et 222 avec le mPNNs-GS pour les séquences NOESY1D et CPMG respectivement, en raison de données manquantes (alimentaires ou autre comme l'activité physique). Les individus n'ayant pas de données métabolomiques disponibles issues des deux séquences NOESY1D et CPMG à la fois ont été exclus pour les analyses de régression Elastic Net, puisque tous les signaux RMN ont été analysés conjointement. Par ailleurs, pour ces analyses Elastic Net, un seul témoin par cas a été sélectionné aléatoirement parmi les 2 témoins appariés. Les données de 216 cas et 216 témoins ont donc été analysées.

### 2. Résultats

Parmi les 210 signaux RMN analysés, 24 étaient associés à l'alimentation ( $p < 0.05$ ,  $r > |0.15|$ ) et ont été ensuite analysés en régression logistique pénalisée (Elastic Net). L'ensemble de ces associations est présenté en Annexe 13 (N=36 associations). Un seul signal a été sélectionné par Elastic Net mais n'a pas été identifié (région entre 2.3515 et 2.3865 en CPMG). Ce signal était associé négativement au profil alimentaire « sain » ( $r = -0.17$ ) et était augmenté dans le plasma des femmes ayant développé un cancer du sein au cours du suivi (OR=1.21, IC95%=[0.99-1.50],  $p = 0.06$ ). Ce signal avait une fréquence de sélection de 59% sur les 1000 jeux de données ré-échantillonnés. Tous les signaux avaient une fréquence sélection inférieure à 70%.

## II. Signature du risque de cancer de la prostate liée à l'alimentation – analyses en H-RMN

### A. Méthodes

La sélection de la population d'étude pour ces analyses a été détaillée dans la partie Méthodologie générale (III.A.2.). Il s'agit de la même étude cas-témoins nichée dans la population SU.VI.MAX que celle de l'objectif 1 (171 cas de cancer de la prostate incidents et 1 témoin apparié par cas). Après prétraitements des données, 159 signaux RMN et 178 ont été obtenus pour les séquences NOESY1D et CPMG respectivement.

Les mêmes analyses statistiques que pour le cancer du sein ont été réalisées. Les modèles d'analyse de corrélation ont été ajustés sur l'âge (en continu), l'IMC (en continu), le statut tabagique (non fumeurs, arrêt de la cigarette depuis plus de 16 ans, arrêt de la cigarette depuis moins de 16 ans et plus de 6, arrêt de la cigarette depuis moins de 6 ans, fumeurs), la saison de prélèvement sanguin et le nombre d'enregistrements de 24h validés et la moyenne de l'apport énergétique (sans la composante alcool pour les corrélations avec la consommation d'alcool) pendant les 2 ans suivant le prélèvement sanguin. 49 et 47 participants ont été exclus pour les analyses de corrélation avec les groupes d'aliments, 54 et 52 avec le DQI-I et 94 et 92 avec le mPNNs-GS pour les séquences NOESY1D et CPMG respectivement, en raison de données manquantes (alimentaires ou autre comme l'activité physique).

Les individus n'ayant pas de données métabolomiques disponibles issues des deux séquences NOESY1D et CPMG à la fois ont été exclus pour les analyses de régression Elastic Net, puisque tous les signaux RMN ont été analysés conjointement. Les données de 157 cas et 157 témoins ont donc été analysées. Tous les modèles de régression logistique ont été ajustés sur l'âge (en continu), l'IMC (en continu), le statut tabagique (non-fumeurs, arrêt de la cigarette depuis plus de 16 ans, arrêt de la cigarette depuis moins de 16 ans et plus de 6, arrêt de la cigarette depuis moins de 6 ans, fumeurs), groupe d'intervention SU.VI.MAX, taille, consommation d'alcool, activité physique (irrégulière, <1h/jour de marche ou équivalent ou ≥1h/jour de marche ou équivalent), niveau d'éducation (primaire, secondaire, supérieur), antécédents familiaux de cancer de la prostate.

## B. Résultats

Parmi les 337 signaux RMN analysés, 268 étaient associés à l'alimentation ( $p < 0.05$ ,  $r > |0.15|$ ) et ont ensuite été analysés en régression logistique pénalisée (Elastic Net). L'ensemble de ces associations est présenté en Annexe 14 (N=851). Trois signaux ont été sélectionnés par Elastic Net en CPMG dont un seul a pu être identifié : l'albumine. Elle était associée positivement avec la consommation de châtaigne, tapioca et banane plantain (0.15) et négativement avec la consommation d'agrumes et de jus de fruits ( $r = -0.17$  et  $-0.15$  respectivement). Cependant, les pairs-plots ont montré que la première corrélation était associée à une très faible consommation de ce groupe d'aliments dans notre étude (une des limites du test des rangs de Spearman) et n'a donc pas été validée. Les deux autres signaux (régions 7,2107-7,2697 et 7.9837-8.0177) étaient associés négativement avec le score mPNNS-GS, les légumineuses et les produits céréaliers complets pour le premier et positivement avec les condiments et aides culinaires pour le deuxième. Le Tableau 30 présente les associations entre les 3 signaux RMN sélectionnés et le risque de cancer de la prostate incluant les signaux ensemble dans le modèle de régression logistique ou chaque signal séparément. Un taux plasmatique plus élevé en albumine était observé dans le plasma des hommes ayant développé un cancer de la prostate au cours du suivi.

Tableau 30 Associations entre les 3 signaux RMN sélectionnés par Elastic Net et le risque de cancer de la prostate<sup>1</sup> – cohorte SU.VI.MAX, 1994-2007

Assignations	Déplacement chimique (ppm)	Séquence	Tous les signaux sélectionnés ensemble		Signaux sélectionnés un par un	
			OR [IC95%]	P	OR [IC95%]	P
<i>Albumine</i>	2,873	CPMG	1,28 [0,96-1,72]	0,09	1,51 [1,16-1,96]	0,002
<i>Non identifié</i>	7,2107	CPMG	1,42 [1,03-1,97]	0,03	1,71 [1,24-2,35]	0,0009
<i>Non identifié</i>	7,9837	CPMG	0,73 [0,55-0,97]	0,03	0,64 [0,49-0,83]	0,0009

La validation 2D des signaux en italique est en cours sur la plateforme RMN.

<sup>1</sup> Le modèle de régression logistique principale prenant en compte les 3 signaux RMN sélectionnés en même temps. Des modèles de régression logistique complémentaires considéraient séparément les signaux sélectionnés (un modèle pour chacun des 3 signaux). Tous ces modèles ont été ajustés sur l'âge (en continu), l'IMC (en continu), le statut tabagique (non fumeurs, arrêt de la cigarette depuis plus de 16 ans, arrêt de la cigarette depuis moins de 16 ans et plus de 6, arrêt de la cigarette depuis moins de 6 ans, fumeurs), groupe d'intervention SU.VI.MAX, taille, consommation d'alcool, activité physique (irrégulière, <1h/jour de marche ou équivalent ou ≥1h/jour de marche ou équivalent), niveau d'éducation (primaire, secondaire, supérieur), antécédents familiaux de cancer de la prostate. Des tests de tendance linéaire ont été effectués à l'aide des variables en continu. Les OR sont été présentés pour une augmentation de 1 écart-type de la variable continue.

Concernant la stabilité des 3 signaux sélectionnés, leur fréquence de sélection sur les 1000 jeux de données ré-échantillonnés était supérieure à 60% (voir Figure 25 présentant les fréquences de sélection des signaux

>40%). Tous les autres signaux (ne faisant pas partie de la sélection initiale) avaient une fréquence de sélection inférieure à 70%.

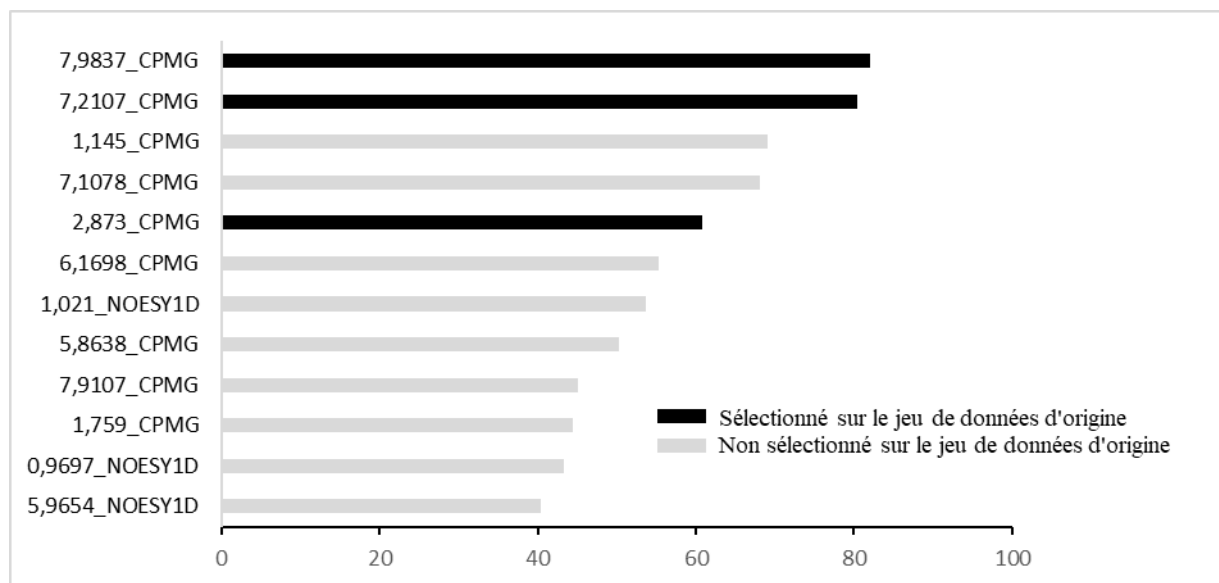


Figure 25 Fréquence de sélection  $\geq 40\%$  des signaux RMN pour le risque de cancer de la prostate (% sur 1000 jeux de données ré-échantillonnés)

### III. Discussion

Les analyses réalisées dans cet objectif ont permis de sélectionner des métabolites associés à l'alimentation et qui maximisaient la discrimination du risque de cancers du sein et de la prostate. Notamment, les femmes ayant développé un cancer du sein au cours du suivi avaient un taux plasmatique plus faible en pipérine et des taux plus élevés en acide 2-amino-butanoïque, ATBC et pregnene-triol sulfate. Ces composés étaient associés à la consommation d'alcool, de café, de biscuits et à un profil alimentaire « occidental ». De plus, les hommes ayant développé un cancer de la prostate au cours du suivi présentaient un taux plasmatique plus élevé en albumine, celui-ci étant corrélé négativement à la consommation d'agrumes et de jus de fruits.

Peu d'études ont examiné les associations entre les signatures métabolomiques liées à la nutrition et le risque de cancer. Quelques-unes ont observé des associations entre des métabolites et, d'une part, des expositions nutritionnelles (notamment la consommation de café, d'alcool, de fibres, de vitamine E et d'aliments frits, l'IMC, l'activité physique), et d'autre part, le risque de carcinome hépatocellulaire (192,194,196), de cancer colorectal (207,211) et de cancer du sein (204,205). La plupart de ces études ont notamment mis en évidence des métabolites liés à la consommation d'alcool et associés au risque de cancer du sein (205) et de carcinome hépatocellulaire (192,194,196). Dans notre étude sur le cancer du sein, parmi les ions sélectionnés ; 3 étaient positivement associés à la consommation d'alcool et notamment la pipérine et le pregnene-triol sulfate. De plus, l'ion M230T171, fréquemment sélectionné lors du ré-échantillonnage était également associé la consommation d'alcool. En revanche, nous n'avons pas observé d'association avec l'alcool parmi les métabolites prédicteurs du risque de cancer de la prostate. Le niveau de preuve est encore limité concernant la consommation d'alcool et le risque de cancer de la prostate contrairement à beaucoup d'autres cancers et notamment le cancer du sein (329). En comparant avec la littérature, certaines associations observées dans cette étude n'avaient jamais été identifiées auparavant (comme les associations positives entre l'ATBC, la consommation de café et le

risque de cancer du sein), d'autres ont été répliquées (détaillées plus loin) tandis que plusieurs autres associations observées dans l'étude de Playdon et al (205) (seule étude sur le risque de cancer du sein et de l'alimentation ; pas d'étude pour le cancer de la prostate et l'alimentation) n'ont pas été reproduites. Cependant, dans cette dernière, les analyses métabolomiques et statistiques étaient différentes (voir détails en Introduction générale IV.B.) ainsi que leur méthode d'évaluation de l'exposition nutritionnelle (fréquentiel alimentaire). Par ailleurs, il est fort possible que la différence de population (issues des Etats-Unis) et donc de mode de vie ait pu impacter les résultats obtenus. Nous avons également pu manquer de puissance dans la détection d'associations (effectif 3 fois plus important dans l'étude de Playdon et al.).

Dans la présente étude, la pipérine, un alcaloïde actif purement exogène (dans l'état des connaissances actuelles), a été mise en évidence comme prédicteur potentiel du risque de cancer du sein (association inverse). Plusieurs études animales ou cellulaires ont suggéré de nombreuses propriétés prometteuses à la pipérine telles que des activités antioxydantes (330,331), anti-inflammatoires (317,330,331), d'immunomodulateur et de promoteur de biodisponibilité et d'absorption de nombreuses molécules actives (317,332), anti-asthmatique, anti-convulsant, anti-mutagène, anti-mycobactérienne et anti-cancer (317) (propriétés chimiopréventives dont l'inhibition de l'angiogenèse et l'amplification de l'apoptose) – notamment sur des modèles de tumeur du sein (333–336). Comme spécifié dans l'objectif 2, la pipérine est présente dans le poivre (317) et est utilisée comme additif alimentaire dans l'alimentation de la volaille (318,319). Dans notre étude, la pipérine était notamment positivement associée à la consommation d'alcool et à un profil alimentaire « occidental ». Ces associations s'expliquent probablement par le fait que plusieurs aliments (par exemple, la viande transformée, les sauces, le fromage industriel, la volaille) contenant de la pipérine comme additif alimentaire pour l'alimentation animale ou *via* le poivre font partie d'un profil alimentaire « occidental » ou sont consommés en association avec des boissons alcoolisées. L'association avec la consommation d'alcool pourrait également s'expliquer par une plus grande solubilité des alcaloïdes dans l'éthanol (337). Conformément à nos résultats, Playdon et al. ont observé une association positive entre la pipérine circulante et la consommation d'alcool et une association négative entre la pipérine circulante et le risque de cancer du sein (205). Une association directe a également été trouvée entre la pipérine sanguine et la consommation de vin dans une cohorte de jumelles (305). Cependant, l'origine de la pipérine circulante n'est pas limitée aux aliments issus d'un profil alimentaire « occidental » et peut aussi être le résultat de l'ajout de poivre dans les aliments, notamment pour l'assaisonnement d'aliments plus sains pour la santé. Malheureusement, le niveau de détail des enregistrements de 24h de la cohorte SU.VI.MAX ne nous a pas permis d'estimer directement avec assez de précision la consommation de poivre ajouté.

L'ATBC (AcetylTributylCitrato) est un plastifiant alternatif aux phtalates (285), couramment utilisé dans les résines polyvinyliques et autorisé comme arôme alimentaire et dans les matériaux de contact alimentaire (338). La migration de l'ATBC des matériaux d'emballage alimentaires vers les aliments a été observée pour le fromage, les gâteaux emballés, la soupe réchauffée au micro-ondes et durant la cuisson au micro-ondes de cookies (338,339). De plus, son taux de lixiviation (techniques d'extraction de produits solubles par un solvant) à partir d'équipements médicaux a été observé comme dix fois plus rapide que le perturbateur endocrinien di-2-ethylhexyl phtalate (287,338,340,341). Dans notre étude, nous avons trouvé une association positive entre l'ATBC plasmatique et la consommation de café. Cette nouvelle association doit être confirmée dans des études d'observation indépendantes chez l'humain et des études d'intervention animale *in vivo* ou *in vitro*. Toutefois, elle pourrait refléter la migration du contaminant ATBC des gobelets en plastique dans le café. En outre, le lait ajouté dans le café pourrait faciliter cette migration puisqu'une

étude a suggéré que l'ATBC a tendance à migrer dans les solutions protéiques, comme les solutions aqueuses de lait écrémé (342). Son augmentation dans le plasma des femmes ayant développé un cancer du sein au cours du suivi pourrait également provenir d'autres expositions que nous n'avons pas pu détecter dans cette étude et cette observation doit également être validée dans d'autres études. Une activité biologique potentielle de l'ATBC sur la croissance tissulaire a récemment été observée (343) ainsi qu'une perturbation potentielle de la fonction ovarienne chez les souris femelles en raison d'une exposition à l'ATBC à faible dose imitant celle de l'exposition humaine (287).

À l'heure actuelle, la consommation d'alcool est le seul facteur de risque alimentaire établi du cancer du sein pour lequel il existe un niveau de preuve « convainquant » (344,345). Certains mécanismes sous-jacents sont déjà décrits, mais d'autres restent à élucider (344). Dans notre étude, le pregnene-triol sulfate, une hormone stéroïdienne sulfatée appartenant à la famille de la progestérone, était positivement associé à la consommation d'alcool et au risque de cancer du sein. En accord avec nos résultats, une étude métabolomique récente a trouvé des corrélations positives entre la consommation d'alcool et plusieurs stéroïdes sériques, notamment le pregnene-diol sulfate, qui étaient associés à une augmentation du risque de cancer du sein (205). Par ailleurs, l'augmentation des niveaux des stéroïdes sexuels semble fortement associée au risque de cancer du sein postménopausique (91). La consommation d'alcool pourrait augmenter les niveaux circulants d'hormones stéroïdiennes et affecter la susceptibilité à la transformation cellulaire ou favoriser la croissance tumorale (347). Une étude interventionnelle chez des femmes ménopausées a notamment montré que la consommation d'alcool augmentait le taux sérique de sulfate de déhydroépiandrostérone (DHEA) (348), un précurseur de l'androstènedione, de la testostérone, de l'estrone et de l'estradiol. En outre, le pregnene-triol sulfate, trouvé dans notre étude, pourrait provenir de la 17-hydroxy-pregnenolone, dont la perte de la chaîne latérale peut produire du DHEA. D'autres facteurs pourraient influencer le niveau de stéroïdes sexuels. Ces derniers sont susceptibles de jouer un rôle dans l'obésité et le cancer car après la ménopause, le tissu adipeux est la principale source d'œstrogènes endogènes (344). Cependant, dans notre étude, l'IMC a été pris en compte comme facteur de confusion potentiel. L'allaitement pourrait également modifier les taux d'hormones, en particulier les androgènes (344). De plus, il a été démontré que l'ATBC activait fortement les récepteurs stéroïdiens et xénobiotiques (SXR) chez l'humain et le rat et pourrait modifier le métabolisme des hormones stéroïdiennes endogènes (285).

Dans notre étude, nous avons constaté une augmentation du taux plasmatique d'acide 2-amino-4-cyano butanoïque dans le plasma de femmes qui ont développé un cancer du sein lors du suivi. Ce composé est aussi appelé acide alpha-amino-gamma-cyanobutanoïque, c'est un acide alpha-aminé non protéinogène (acide 2-aminobutanoïque) substitué en position 4 par un groupe cyano et un nitrile aliphatique. Il provient de l'acide butyrique (appelé aussi acide butanoïque) (349). Le butyrate est un acide gras à courte chaîne synthétisé par la fermentation de fibres non digestibles par des bactéries du côlon (350). Le butyrate a récemment fait l'objet d'une attention particulière pour ses effets bénéfiques sur l'homéostasie intestinale et le métabolisme énergétique. Avec ses propriétés anti-inflammatoires, le butyrate améliore la fonction de barrière intestinale et l'immunité des muqueuses (351). Les acides gras à courte chaîne pourraient notamment avoir un rôle protecteur contre l'état inflammatoire associé aux modes de vie occidentaux (352). De nombreuses études soutiennent un effet chimiopréventif du butyrate, cependant certains résultats restent controversés, notamment entre les études *in vivo* et *in vitro* ; ce que l'on appelle le « paradoxe du butyrate » (353). Dans notre étude, nous avons trouvé une association inverse entre l'acide 2-amino-4-cyano butanoïque plasmatique et la consommation de gâteaux et de biscuits. À notre



connaissance, aucune association directe entre l'acide 2-amino-cyano butanoïque plasmatique et l'exposition alimentaire n'a été établie à ce jour. La production et les effets du butyrate sembleraient être liés à l'alimentation, notamment au type de fibres et de graisse consommés respectivement (353). Par ailleurs, l'acide butyrique est présent dans divers types d'aliments, dont la vanille, l'avoine, l'arachide et plusieurs fruits (354). N'ayant pas de données disponibles sur le taux de butyrate plasmatique dans notre étude, nous ne pouvons pas conclure sur un possible lien entre les variations observées et une perturbation du métabolisme du butyrate *via* le microbiote. Nos résultats doivent être reproduits dans d'autres études d'observation indépendantes et des études d'intervention sur modèles animaux permettraient de mieux comprendre l'origine de ces variations et les liens avec l'alimentation. Par ailleurs, la modélisation *in silico* des relations dynamiques complexes entre l'alimentation, la composition du microbiote et production de substrat semblerait permettre de prédire la production d'acides gras à courte chaîne issue des interactions entre le microbiote intestinal et le régime alimentaire (352). Son association avec le risque de cancer du sein et l'élucidation d'une relation causale potentielle pourrait être étudiées dans des études mécanistiques cellulaires.

Au-delà de sa fonction de régulateur de la pression oncotique sanguine, l'albumine a également un rôle de transporteur se liant avec diverses molécules endogènes (telles que les acides gras à longue chaîne et les stéroïdes), des ions (zinc, calcium) ou encore divers composés d'origine exogène (médicaments, xénobiotiques etc.) impactant son activité et sa demi-vie (355). En outre, certains composés alimentaires peuvent se fixer à l'albumine sanguine notamment les polyphénols (356). La naringinine et l'hespérétine sont les principaux polyphénols (flavanones) des agrumes et des jus de fruits (357). Ces composés sont absorbés et métabolisés en naringinine et hespérétine glucuronides lesquels peuvent se fixer à l'albumine circulante permettant leur distribution aux tissus (357). Par ailleurs, plusieurs études ont montré des effets anti-inflammatoires et protecteurs des flavones dans le développement des cancers et les maladies cardiovasculaires (358). Dans notre étude, nous avons observé que les hommes ayant développé un cancer de la prostate au cours du suivi avaient un taux plasmatique d'albumine augmenté par rapport aux témoins. Nous avons également observé une corrélation négative entre l'albumine plasmatique et la consommation d'agrumes et de jus de fruits. Cette association nécessite une validation dans des études indépendantes et notamment dans des études d'intervention, cependant la fixation des dérivés polyphénoliques issus de ces aliments à l'albumine circulante pourrait expliquer une modification de l'activité / du métabolisme de cette protéine ou encore de sa détection en spectroscopie RMN.

De manière générale, les signaux issus des analyses en H-RMN semblent être de moins bons prédicteurs du risque de cancers du sein et de la prostate que ceux issus des analyses en LC-MS. Cela peut s'expliquer par le fait que l'on détecte moins de composés exogènes en RMN (impliqués dans de moins nombreuses voies métaboliques et donc plus spécifiques). En outre, parmi les métabolites sélectionnés au global et identifiés, 3 sur 5 sont d'origine purement exogène ou issus du métabolisme microbien (pipérine, ATBC, acide 2-amino-cyano-butanoïque) et les 2 autres (pregnene-triol sulfate et albumine) sont impliqués dans des voies relativement spécifiques comparées à des composés tels que les acides aminés. On remarque par ailleurs que certaines composantes de l'alimentation sont associées à de plus nombreux métabolites. Par exemple, en spectrométrie de masse, on a observé de nombreuses associations avec le café et le vin rouge. Cela peut s'expliquer par la grande sensibilité de cette méthode analytique et le fait que ces composantes contiennent des molécules bioactives purement exogènes comme les polyphénols, qui sont très spécifiques.

Les associations observées entre les métabolites et l'alimentation d'une part et les métabolites et le risque de cancer d'autre part, ne sont pas forcément directes entre l'alimentation et le risque de cancer, comme montré pour la pipérine et l'ATBC par exemple. Ces effets peuvent être le proxi d'autres facteurs de confusion résiduels. Pour le vérifier, il conviendrait de réaliser des analyses de médiations ainsi que des études d'intervention, notamment chez l'homme lorsque les hypothèses sont protectrices et *in vivo/in vitro* lorsqu'elles sont délétères.

En raison de l'approche non ciblée, les autres métabolites sélectionnés par la régression Elastic Net n'ont malheureusement pas pu être identifiés. Cependant, les résultats concernant ces métabolites pourraient être utiles à l'avenir, avec le développement de ce type d'approches et les efforts de compilation des annotations dans les bases de données métabolomiques. Ces métabolites non identifiés discriminants du risque de cancers du sein et de la prostate dans notre étude étaient liés à diverses expositions alimentaires comme les profils alimentaires « sain » ou « occidental » ou bien leurs composantes comme les légumineuses, les produits céréaliers complets, les poissons gras ou encore les produits sucrés. Encore trop peu de preuves sont disponibles à ce jour pour établir un lien avéré entre ces facteurs et le risque de cancers du sein et de la prostate (18,21), cependant si nos résultats sont répliqués dans d'autres études à grande échelle et qu'un effet médiateur est établi, ils pourront apporter un appui supplémentaire d'un point de vue épidémiologique et moléculaire.

Enfin, nous constatons une grande différence entre les deux méthodes d'estimation de la performance prédictive pour le risque de cancer du sein avec les données de spectrométrie de masse. L'estimation de l'AUC pour la validation croisée simple (0.76) surestime probablement la valeur qui serait obtenue sur un jeu de données indépendant. Par ailleurs, les sélections de métabolites par double validation croisée étant différentes pour chaque sous-échantillon, la valeur d'AUC de 0.60 ne reflète probablement pas celle de notre signature d'intérêt ; elle souligne toutefois une robustesse limitée de la méthode. Cependant, lors du processus de double validation croisée, l'effectif est réduit par rapport au processus de sélection sur l'échantillon initial et cela pourrait donc s'expliquer par un manque de puissance. En outre, le processus de ré-échantillonnage a montré une bonne stabilité des métabolites sélectionnés sur l'échantillon initial.

### I. Bilan des résultats

Cette thèse avait pour objectif d'utiliser la métabolomique de manière exploratoire afin de proposer de potentielles signatures plasmatiques du risque de cancers du sein et de la prostate, de l'alimentation ou des deux. La première partie avait un objectif étiologique, explorant toutes les associations possibles entre chaque métabolite et le risque de cancers. La deuxième partie visait à mettre en évidence les associations entre alimentation et profils métabolomiques pouvant apporter soit des éléments de compréhension des liens nutrition/métabolisme et santé soit des mesures biologiques pour l'amélioration de l'évaluation de l'exposition nutritionnelle. Enfin, la troisième partie avait pour objectif de sélectionner quelques métabolites les plus discriminants du risque de cancers du sein et de la prostate qui soient liés à l'exposition nutritionnelle, dans une optique future de contribuer à l'amélioration de la prévention de ces pathologies *via* l'alimentation.

Les principaux résultats et éléments de discussion de ces trois parties sont présentés dans le Tableau 31 ci-dessous. Ces travaux ont permis de révéler des associations entre des métabolites d'origines endogène et exogène en H-RMN et en LC-MS, et exclusivement exogènes ou issus du métabolisme du microbiote en LC-MS, qui se sont révélés être associés au risque de cancers du sein et de la prostate, à l'alimentation ou aux deux. Ces résultats, relativement innovants dans le domaine, n'ont que peu d'éléments de comparaison pour pouvoir être validés à ce jour de manière robuste. Par ailleurs, de nombreuses sources de variabilité existent entre les quelques études similaires publiées dans la littérature et rendent les comparaisons limitées. Cependant, plusieurs associations ont tout de même été retrouvées dans d'autres études soit de manière très similaire, soit plus éloignées (par exemple entre différentes localisations de cancer ou différentes composantes de l'alimentation faisant partie d'un même régime). Il est dans ce cas, intéressant de pouvoir disposer de signatures globales (comme le permet la métabolomique) plutôt qu'un seul métabolite afin d'augmenter la spécificité pour le paramètre d'intérêt, d'où l'avantage de la métabolomique.

Tableau 31 Résumé des résultats et éléments de discussion

	Objectifs	Méthode analytique	Principaux résultats concernant les métabolites identifiés	Éléments de discussion
1	Identifier une signature métabolomique plasmatique associée au risque à long terme de <b>cancer du sein</b>	LC-MS	<ul style="list-style-type: none"> <li>- Association négative avec l'O-succinyl-homoserine</li> <li>- Associations positives avec la valine, la glutamine, l'acide 5-aminovalérique, la phénylalanine, le tryptophane, la <math>\gamma</math>-glutamyl-threonine (GGT), l'acetyl tributyl citrate (ATBC) et le pregnene-triol sulfate</li> </ul>	<ul style="list-style-type: none"> <li>- Quelques résultats reproduits dans la littérature mais peu d'études similaires</li> <li>- Certains métabolites communs entre les 2 cancers</li> <li>- Reflet d'une perturbation métabolique jusqu'à 13 ans avant diagnostic du cancer</li> <li>- Rôle potentiel de facteurs exogènes et du microbiote</li> </ul>
		H-RMN	<ul style="list-style-type: none"> <li>-Associations négatives avec les lipoprotéines, les lipides, les glycoprotéines, l'acétone, les composés dérivés du glycérol, les lipides insaturés</li> <li>- Associations positives avec la valine, la lysine, l'arginine, la glutamine, la créatine, la créatinine et le glucose</li> </ul>	
	Identifier une signature métabolomique plasmatique associée au risque à long terme de <b>cancer de la prostate</b>	H-RMN	<ul style="list-style-type: none"> <li>- Associations négatives avec l'urée, le citrate et les lipoprotéines</li> <li>- Associations positives avec la valine, la glutamine, la tyrosine, la créatine, l'albumine lysyl, la phénylalanine, l'histidine, l'albumine, la 3-méthylhistidine</li> </ul>	
2	Identifier une signature métabolomique plasmatique associée à l' <b>alimentation chez les femmes</b>	LC-MS	<ul style="list-style-type: none"> <li>-Associations positives entre une alimentation saine et l'acide pipécolique, l'acide 3-indolepropionique et le dihydro4mercapto-3(2H) furanone</li> <li>- Associations négatives entre une alimentation saine et la lysophosphatidylcholine (17 :1), les acylcarnitines C11 :1 et C11 :9, le 2-déoxyglucose</li> <li>- Association positive entre le profil alimentaire « occidental » et la pipérine</li> <li>- Associations positives entre l'alcool et l'ethyl-<math>\beta</math>-D-glucopyranoside; l'acide pipécolique et les fruits et légumes, les fibres, les pommes et les poires et le yaourt; la trigonelline, la paraxanthine, le cyclo(isoleucyl-propyl), l'atractyligénine glucuronide et le café; la proline bétaine, l'hydroxyproline bétaine et les agrumes ; la glucokinase et le lait; la N6,N6,N6-Trimethyl-L-lysine et la pipérine et la charcuterie</li> <li>- Associations négatives entre la phenylalanine-phenylalanine et les produits laitiers; l'acide cholique et les produits sucrés ; la lysoPC(17:1) et les matières grasses végétales ajoutées</li> </ul>	<ul style="list-style-type: none"> <li>- Associations plus fortes entre les métabolites et les composantes spécifiques de l'alimentation qu'avec l'alimentation globale</li> <li>- Hétérogénéité au sein des régimes alimentaires</li> <li>- Comparaison difficile entre les études car nombreuses sources de variabilités (méthodes analytiques et statistiques, estimation de l'alimentation globale, population d'étude)</li> <li>- Plusieurs origines possibles des variations</li> </ul>

		H-RMN	<ul style="list-style-type: none"> <li>-Associations positives entre une alimentation saine et la leucine et la créatine</li> <li>- Associations négatives entre une alimentation saine et le glucose, les lipoprotéines, les lipides insaturés, le lactate et la thréonine</li> <li>- Association positive entre le citrate et un régime à base de produits sucrés</li> <li>- Associations positives entre la leucine et les produits de la mer; l'arginine et la lysine et les produits céréaliers complets ; la tyrosine et le pain, les céréales, les pommes de terre et les légumineuses ; la phénylalanine et les fruits et légumes; le glucose et les matières grasses totales ajoutées</li> <li>- Associations négatives entre l'isobutyrate, la valine et les lipoprotéines et le pain, les céréales, les pommes de terre et les légumineuses</li> </ul>	<p>observées selon les métabolites (endogène, exogène, microbiotique)</p> <ul style="list-style-type: none"> <li>- Peu de métabolites marqueurs de la consommation alimentaire (purement exogène) associés à l'alimentation globale / reflet d'un effet long terme</li> <li>- Réplication de plusieurs associations concernant des marqueurs de consommation alimentaire</li> </ul>
	Identifier une signature métabolomique plasmatique associée à l' <b>alimentation chez les hommes</b>	H-RMN	<ul style="list-style-type: none"> <li>- Associations positives entre une alimentation saine et la leucine et l'albumine</li> <li>- Associations négatives entre une alimentation saine et la choline, le glucose et les lipoprotéines</li> <li>- Association positive entre la créatine et le profil alimentaire « occidental »</li> <li>- Associations positives entre la choline, la créatine et le glycéryl et le pain, les céréales, les pommes de terre et les légumineuses; l'alcool et le propylène glycol, les lipides, la choline et le glucose; le glucose et les produits laitiers; la tyrosine et les matières grasses ajoutées totales</li> <li>- Associations négatives entre le lactate, les lipides, le pyruvate, la choline et le glucose et le pain, les céréales, les pommes de terre et les légumineuses; le sel et le lactate</li> </ul>	
3	Identifier une signature métabolomique plasmatique liée à l' <b>alimentation</b> et au risque à long terme de <b>cancer du sein</b>	LC-MS	<ul style="list-style-type: none"> <li>- Associations positives entre le risque de cancer du sein et l'ATBC, le pregnene-triol sulfate, l'acide 2-amino-cyano-butanoïque acide</li> <li>- Association négative entre le risque de cancer du sein et la pipérine</li> <li>- Associations positives entre l'ATBC et le café; le pregnene-triol sulfate et l'alcool; la pipérine et l'alcool et le profil alimentaire « occidental »</li> <li>- Association négative entre l'acide 2-amino-cyano-butanoïque et les biscuits et gâteaux</li> </ul>	<ul style="list-style-type: none"> <li>- Signaux issus de la LC-MS meilleurs prédicteurs du risque que ceux de la H-RMN --&gt; plus spécifiques</li> <li>- Peu d'études similaires mais réplication de certaines associations</li> <li>- Certaines associations avec l'alimentation pourraient provenir de modification des propriétés physico-chimiques des métabolites, d'autres pourraient être le reflet d'autres expositions (comme le café et les gobelets en plastique)</li> </ul>
		H-RMN	Pas de métabolite identifié	
	Identifier une signature métabolomique plasmatique liée à l' <b>alimentation</b> et au risque à long terme de <b>cancer de la prostate</b>	H-RMN	<ul style="list-style-type: none"> <li>- Association positive entre l'albumine et le risque de cancer de la prostate</li> <li>- Association négative entre l'albumine et les agrumes et jus de fruits</li> </ul>	

## II. Considérations méthodologiques

### A. Limites inhérentes aux analyses métabolomiques

La métabolomique peut être utilisée dans le cadre de plusieurs approches. Elle peut s'utiliser sous forme de kit, où une centaine de métabolites sont analysés et en général quantifiés. Cette approche, relativement peu coûteuse (par rapport aux autres) est très adaptée pour des études cherchant à répondre à une hypothèse particulière. Même si elle peut couvrir plusieurs familles de métabolites, cette approche limite le potentiel de découverte de nouveaux biomarqueurs. Une autre approche semi-ciblée, très coûteuse pour le moment, permet de détecter un grand nombre de métabolites dont certains sont identifiés grâce à des standards commerciaux. Une majorité des études utilisant ce type de technologie se concentre en général sur les métabolites identifiés. Dans le cadre de cette thèse exploratoire, l'approche métabolomique non ciblée a été appliquée afin de maximiser le potentiel de nouvelles découvertes. Cependant, une limite majeure inhérente à ce type d'approche est le nombre important de métabolites qui restent non identifiés. Elle peut parfois également révéler des structures chimiques très spécifiques pour lesquelles il est difficile de trouver des informations en termes d'origine, de recherches originales, etc. Les résultats concernant ces métabolites non identifiés demeurent d'intérêt pour les études futures avec l'enrichissement des bases de données qui pourront permettre la comparaison des résultats une fois identifiés. Une autre limite de cette approche non ciblée est la production et l'analyse d'une grande quantité de signaux correspondant à du bruit, qui ne sont pas forcément décelables en amont des analyses. Cela peut avoir un impact statistique, notamment sur les questions de correction des tests multiples qui dépendent du nombre de signaux analysés pouvant donc pénaliser la significativité des associations correspondantes aux métabolites « réels ». Cependant, les prétraitements réalisés par les plateformes permettent de limiter un maximum cette production de bruits.

Pour des raisons financières, la technique GC-MS n'a pas été appliquée. Il est donc possible que la couverture des métabolites détectés en LC-MS ait limité l'observation de certaines associations. Même si la LC-MS est très adaptée pour les échantillons biologiques en détectant de nombreuses classes de métabolites telles que les vitamines et les acides aminés, la GC-MS est très adaptée pour détecter des molécules peu polaires et très volatiles, telles que les alcools, les cétones et les esters. La RMN est moins sensible que la MS et n'a pas permis d'identifier assez précisément plusieurs métabolites, notamment lipidiques. Par ailleurs, de nombreux signaux se superposent dans les spectres RMN. Le cas échéant il est donc difficile d'attribuer les associations observées à un seul métabolite. De plus, les signaux de la H-RMN représentant la fréquence de résonance des noyaux d'hydrogènes, l'effet d'une même molécule peut être dilué dans plusieurs signaux ce qui pourrait entraîner une diminution de la significativité de l'association.

### B. Limites inhérentes à la population et au design d'études

#### 1. Causalité

Dans le cadre de cette thèse, les données de la cohorte SU.VI.MAX ont été utilisées selon un design observationnel. Dans ce type d'études il n'est pas possible d'affirmer de manière certaine que les associations observées sont effectivement causales. En outre, les origines des variations observées ne peuvent pas être élucidées dans ce type d'étude d'observation, notamment pour de potentiels biomarqueurs d'effet. Des hypothèses concernant les origines des biomarqueurs d'exposition candidats peuvent être suggérées notamment lorsque ce sont des composés exclusivement exogènes et suffisamment spécifiques d'une exposition particulière. Par ailleurs, il n'est pas non plus possible dans ce

type d'étude d'établir une causalité certaine pour d'éventuels biomarqueurs de susceptibilité, notamment avec le risque de cancer. En outre, certaines variations observées peuvent sembler contradictoires, cependant les effets observés dans ces cas là n'ont pas forcément de lien direct (implication simultanée dans plusieurs voies métaboliques par exemple) ou peuvent être dus à de la confusion résiduelle (reflet d'une autre exposition par exemple). Par ailleurs, un design prospectif pour le risque de cancer permet d'établir une certaine temporalité dans les associations.

Sans connaissance *a priori* des métabolites analysés (approche non ciblée), il est difficile d'évaluer tous les facteurs confondants potentiels. D'autre part, certaines données n'étaient pas disponibles et auraient permis de prendre en compte d'autres facteurs tels que le taux de créatinine qui renseigne sur d'éventuelles insuffisances rénales pouvant entraîner des variations d'autres métabolites. Toutefois, de nombreux facteurs de confusion potentiels ont été pris en compte dans les analyses (appariement et ajustement) pour limiter les biais de confusion. En plus de la comparaison systématique des résultats à la littérature, la plausibilité biologique est un élément très important à prendre en compte. Cependant, lorsqu'il s'agit de métabolites ayant des origines multiples et étant impliqués dans de nombreuses voies métaboliques, comme c'est le cas de nombreux métabolites observés dans notre étude, les hypothèses d'interprétations biologiques et mécanistiques sont à prendre avec précaution. Dans ce type de problématique, plusieurs approches pourraient permettre de valider ces hypothèses, elles seront détaillées plus loin (perspectives de recherche).

## 2. Fiabilité des profils métabolomiques

Les analyses métabolomiques réalisées dans le cadre de cette thèse se basent sur un seul prélèvement sanguin (à l'inclusion). Ainsi, la variabilité intra-individuelle et la stabilité dans le temps des profils métabolomiques n'ont pas pu être investiguées. Néanmoins, plusieurs études ont montré une bonne stabilité et reproductibilité des mesures métabolomiques pour la plupart des classes de métabolites et en particulier des acides aminés, hexoses, phosphatidylcholines et sphingomyélines et de manière plus limitée pour les acylcarnitines (359–362). Une étude a notamment montré que la reproductibilité serait meilleure dans le plasma que le sérum bien que ce dernier fournirait des résultats plus sensibles dans la détection de biomarqueurs au vu de ses concentrations en métabolites plus importantes (362). Ces résultats semblent soutenir le fait qu'un seul prélèvement sanguin serait suffisant pour l'évaluation du risque dans études épidémiologiques chez des sujets en bonne santé (359). Cependant, l'établissement de profils métabolomiques issus de plusieurs prélèvements sanguins au cours du suivi aurait permis une détection plus précise des perturbations métaboliques aux différents stades de la cancérogenèse.

D'autre part, bien que la stabilité de certaines classes de métabolites, comme les acides aminés, ne semble pas particulièrement être affectée par le statut de jeûne (reflétant des processus hautement régulés) (363), la prise alimentaire peut influencer la composition du métabolome, à court terme, et à long terme (51). En outre, une plus grande variabilité intra-individuelle est attendue pour les biomarqueurs ayant une courte demi-vie comme par exemple, pour les polyphénols et les alkylresorcinol (51,364–367). Ces biomarqueurs seraient donc utiles dans les études d'observation uniquement dans les populations qui consomment régulièrement et fréquemment ces sources alimentaires (51). L'utilisation de prélèvements sanguins à jeun aura donc plutôt tendance à refléter l'impact à long terme de l'alimentation.

Enfin, nous ne pouvons pas exclure que les échantillons aient pu être dégradés (oxydation des lipides par exemple) au cours du temps (une vingtaine d'années depuis le prélèvement). En revanche, comme les

prélèvements ont été aliquotés juste après le prélèvement, les échantillons analysés qui étaient conservés dans l'azote liquide n'ont jamais été décongelés avant cette étude.

### *3. Aspects statistiques*

En raison des coûts élevés de la métabolomique, la taille des populations d'études était relativement limitée. Bien que certainement suffisante pour les analyses principales, nous avons pu manquer de puissance pour les analyses stratifiées et de sensibilité. Ces effectifs restreints peuvent expliquer certains intervalles de confiance larges. Par ailleurs, pour des questions de puissance nous n'avons pas pu exclure les participants ayant d'autres pathologies chroniques prévalentes ou qui ont développé d'autres types de cancer durant le suivi. Comme expliqué plus haut, nous ne pouvons pas non plus exclure la possibilité de facteurs de confusion résiduels.

Par ailleurs, ne disposant pas de jeux de données indépendants, une estimation de la performance prédictive n'a pas pu être établie de manière non biaisée, malgré les processus de validation croisée, de répétition et de ré-échantillonnage. Le fait de ne pas avoir disposé d'échantillon indépendant pour la validation est notamment problématique pour cette estimation de la performance prédictive qui risque d'être trop spécifique de la population étudiée, d'où la nécessité de valider ces observations dans des cohortes indépendantes et notamment dans différents pays.

Enfin, nous avons appliqué de manière systématique la correction de Benjamini Hochberg lorsque des tests multiples ont été réalisés. Cependant, tous les résultats qui nous ont semblé pertinents d'un point de vue biologique et épidémiologique ont été commentés dans ces travaux exploratoires (même ceux ayant un  $FDR > 0.05$ ) afin de limiter le risque de ne pas détecter des associations qui existent vraiment (erreur de seconde espèce, risque de faux négatifs). Ce type de correction est d'ailleurs peu adapté à l'approche métabolomique non ciblée en raison du très grand nombre de signaux produits et de sa proportion importante en signaux relatifs au bruit. Tous les résultats de cette thèse devront être répliqués et validés dans des échantillons indépendants afin notamment de confirmer qu'ils n'étaient pas dus au hasard par la multiplicité des tests.

### *4. Potentiels biais épidémiologiques*

Dans les études épidémiologiques, l'estimation des consommations alimentaires est soumise à plusieurs biais, notamment les biais de mémoire et de désirabilité sociale (fait de se présenter d'une manière plus favorable aux autres, de manière consciente ou non) (368,369). Cependant, les biais de mémoire sont beaucoup plus importants dans les questionnaires de fréquences (effort de mémoire sur une période longue) que dans les enregistrements de 24h (38). Par ailleurs, l'utilisation de photographies validées permet de limiter ce biais. Le biais de désirabilité sociale peut amener, par exemple, à sous-estimer les consommations d'aliments considérés comme « malsains ». Une étude a d'ailleurs montré que ce biais semblait être réduit pour les études menées sur internet (370). Une procédure standardisée (240,371) est appliquée de manière systématique sur les données de la cohorte SU.VI.MAX afin d'exclure les participants ayant des apports énergétiques globaux non plausibles (apports énergétiques globaux inférieurs à 100 kcal par jour ou supérieurs à 6000 kcal par jour). De plus, les hommes rapportant moins de 800 kcal par jour et les femmes déclarant moins de 500 kcal par jour dans au moins un tiers de leurs enregistrements de 24 heures étaient considérés comme sous-déclarants en énergie et ont donc été exclus. Dans tous les cas, les apports alimentaires habituels ne peuvent être estimés de manière fiable que par l'administration répétée d'enregistrements de 24h sur des périodes suffisamment longues (372,373). Ainsi, la méthode utilisée pour ces travaux et pour les autres études de SU.VI.MAX consistait à calculer les apports moyens de tous



les enregistrements de 24h disponibles au cours des deux premières années de suivi, permettant notamment de couvrir les différents jours de la semaine et les différentes saisons de l'année.

Un seul enregistrement de 24h ne peut suffire à représenter les apports usuels étant donné la variabilité intra-individuelle des apports. Il a été montré qu'un nombre minimum de 3 enregistrements de 24h serait nécessaire pour estimer les apports énergétiques et en macronutriments usuels (373). Pour une estimation très précise des apports nutritionnels usuels, une étude méthodologique a montré qu'au moins 8 enregistrements de 24h seraient nécessaires (374). Dans notre étude, tous les cas incidents de cancer du sein ou de la prostate ont été inclus dans les études cas-témoins nichées, même s'ils avaient moins de 3 enregistrements de 24h disponibles. Tous les autres participants avaient au moins 3 enregistrements de 24h disponibles. Les données alimentaires n'ont été calculées que chez les participants ayant ce critère, les autres ont été exclus pour les analyses nutritionnelles de l'objectif 3. Pour l'objectif 2, seuls les participants ayant au moins 10 enregistrements de 24h disponibles ont été inclus.

Cependant, le recrutement étant basé sur le volontariat, les participants de l'étude et notamment ceux ayant rempli un grand nombre d'enregistrements alimentaires de 24h pourraient présenter un intérêt plus particulier pour la nutrition et la santé et donc un comportement plus sain et un meilleur état de santé. Les échantillons des études transversales pour l'objectif 2 ont probablement des comportements proportionnellement plus sains en moyenne que ceux de la population générale française, même au sein du groupe du 1<sup>er</sup> décile du mPNNS-GS. Par ailleurs, le processus d'appariement a tendance à rendre semblables les deux groupes comparés sur les paramètres qui ne sont pas celui étudié, ce qui n'est probablement pas le cas de la population générale. Par exemple, les personnes qui développent des cancers ou qui ont une alimentation défavorable à la santé auront souvent des caractéristiques plus délétères par ailleurs (plus âgés, plus grands consommateurs de tabac, etc). L'extrapolation des résultats doit donc être conduite avec prudence. Cependant, les limites de représentativité ne conduisent pas nécessairement à un biais concernant les associations observées entre des expositions et les paramètres d'intérêt, mais peuvent conduire à une variabilité réduite des variables d'exposition (375,376). Par exemple, on peut supposer que si dans les études transversales, les groupes du 1<sup>er</sup> décile du mPNNS-GS comprenaient plus d'individus dont la qualité alimentaire globale est très faible, les associations observées entre les indicateurs *a priori* et *a posteriori* de l'alimentation globale auraient été encore plus fortes (dans l'hypothèse que la qualité de l'alimentation globale a un rôle dans les variations métaboliques observées).

### III. Perspectives de recherche

Un schéma de l'ensemble des perspectives de recherche est disponible à la fin de ce chapitre.

#### A. Au sein du laboratoire

##### 1. Analyses supplémentaires dans le cadre du projet (données disponibles)

Une méthode innovante « Meeting in the Middle » (196) développée au Centre International de Recherche contre le Cancer pourrait être appliquée aux données issues de ces études (voir Figure 26). Elle consiste dans un premier temps à réaliser une PLS entre des variables de mode de vie, dont la nutrition, et les données de métabolomiques afin de construire des composantes maximisant la variabilité des deux jeux de données. Pour chaque composante, on obtient des « scores » consistant en des combinaisons linéaires des variables de mode de vie d'une part et des métabolites d'autre part. Ces composantes sont ensuite reliées au risque de cancer grâce à des régressions logistiques conditionnelles. Enfin, une analyse de médiation est réalisée enfin d'estimer si les profils métabolomiques sont sur le chemin causal de l'effet

potentiel entre le mode de vie et le risque d’avoir un cancer. Il est également possible de réaliser une Sparse PLS (377) à la place de la PLS permettant de faire simultanément de la réduction de dimension (combinaison de variables) mais également de la sélection de variables au sein de ces combinaisons. Cela permet notamment d’obtenir des résultats plus facilement interprétables. Il pourrait également être intéressant de faire des analyses de médiation par métabolite et exposition nutritionnelle spécifique selon les hypothèses soulevées (par exemple entre l’alcool, le pregnene-triol sulfate et le risque de cancer du sein).

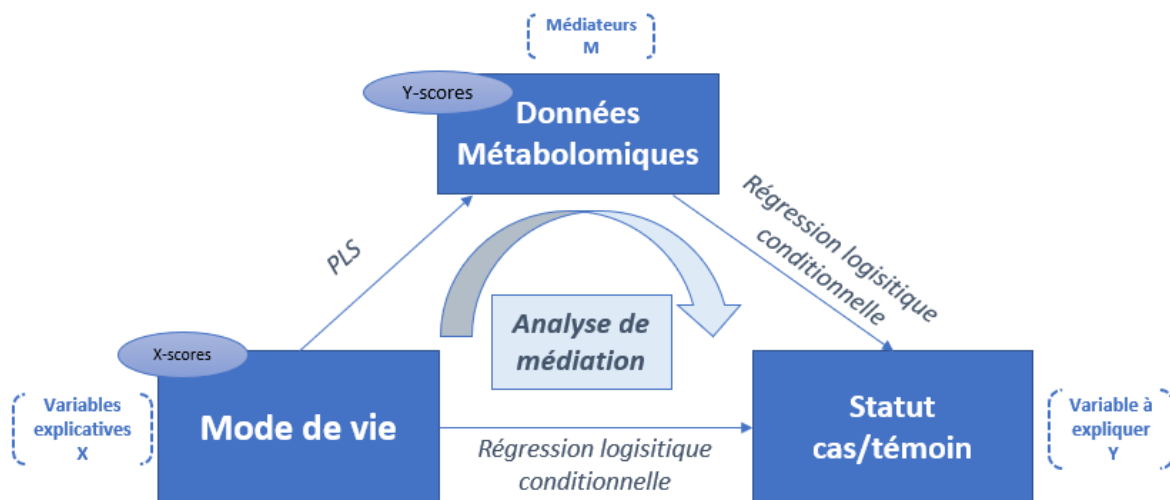


Figure 26 Approche « Meeting in the Middle”  
Source : Assi, Mutagenesis, 2015 (196)

Par ailleurs, des outils de bioinformatique sont disponibles afin de mettre en évidence des voies métaboliques à partir de données métabolomiques tels que Metabolite Set Enrichment Analysis (378) et KEGG Pathway Database (379) ou des réseaux de corrélation avec le Gaussian graphical model (380). Il pourrait être intéressant de faire des analyses par voies métaboliques et de mettre en évidence des réseaux afin d’appuyer les hypothèses mécanistiques avancées et éventuellement en proposer de nouvelles.

Avec les données déjà disponibles dans la cohorte SU.VI.MAX, il pourrait être pertinent d’étudier d’autres paramètres de l’alimentation associés aux profils métabolomiques ; par exemple, d’autres régimes, comme le régime méditerranéen (très étudié dans la littérature), ou encore des régimes d’exclusions (végétariens, végétan etc.).

Un partenariat a été mis en place avec l’équipe “Key Laboratory of Separation Science for Analytical Chemistry” au Dalian Institute of Chemical Physics (Chinese Academy of Sciences) en Chine afin d’analyser les échantillons plasmatiques des participants de l’objectif 1 (sur le cancer de la prostate) en LC-MS et GC-MS semi-ciblée. Les analyses de laboratoires sont achevées et les analyses statistiques des données sont actuellement en cours. Ces travaux apporteront des informations complémentaires et probablement plus spécifiques sur cet objectif.

La Fondation de France a également financé une répliation des objectifs 1, 2 et 3 chez les hommes dans la cohorte Nutrinet-Santé en H-RMN. Les analyses statistiques seront réalisées à horizon 2020. L’étude

NutriNet-Santé, conduite par l'EREN, est la première cohorte basée sur internet dans le monde et à si grande échelle (n = 164 000 en 2019). C'est une cohorte d'observation permettant d'étudier de manière générale les relations entre la nutrition et la santé. Elle est notamment caractérisée par une estimation très fine de l'exposition nutritionnelle et du comportement alimentaire. Une biobanque contenant des échantillons de sérum, plasma, buffy-coat et urine a été constituée. Ces analyses permettront donc de comparer les résultats obtenus sur une cohorte indépendante.

## 2. *Elargissement du projet*

Une recherche de financement est en cours afin de pouvoir répliquer les objectifs 1, 2 et 3 chez les femmes également dans la cohorte Nutrinet-Santé. Toutes les analyses de cette thèse pourraient être reproduites sur d'autres localisations de cancer, dans les cohortes SU.VI.MAX et Nutrinet-Santé, permettant de mettre en évidence des variations métaboliques communes à tous les cancers ou spécifiques d'un organe. Ces analyses pourraient apporter des éléments sur l'impact des facteurs extérieurs et du métabolisme sur la cancérogenèse en général, ou sur des processus plus spécifiques à chaque localisation de cancer. Ce projet étant pionnier sur la thématique au sein du laboratoire, il serait également possible d'étendre cette approche à d'autres maladies chroniques telles que les maladies cardiovasculaires, le diabète, le syndrome métabolique, l'hypertension artérielle, l'obésité, la santé mentale etc.

Par ailleurs, dans la cohorte SU.VI.MAX, plusieurs prélèvements sanguins ont été réalisés au cours du suivi pour certains participants. Des analyses de métabolomique pourraient être réalisées sur ces multiples prélèvements pour ce groupe de participants afin d'évaluer la stabilité des profils et de prendre en compte la variabilité intra-individuelle.

D'autre part, les informations collectées dans le cadre de la cohorte Nutrinet-Santé forment une base de données très riche permettant d'étudier différentes composantes spécifiques de l'alimentation. Notamment, un projet est actuellement en cours sur l'étude des expositions aux additifs alimentaires (projet « Additives »). Dans le cadre de ce projet, il est notamment prévu d'investiguer, grâce à une étude cas-cohorte, les associations entre signatures métabolomiques et exposition aux additifs et leurs liens avec le développement de maladies chroniques. Par ailleurs, grâce à un questionnaire de fréquence administré à un sous-échantillon de la cohorte (projet « BioNutrinet »), il est également possible d'étudier les relations entre consommation de produits issus l'agriculture biologique ou exposition aux pesticides et les profils métabolomiques (projet en cours). Enfin, le laboratoire possède également des données sur la consommation de compléments alimentaires (substance, produits, fréquence, nombres d'unités/jour) grâce notamment à une table de composition des compléments alimentaires (projet « Quantica »). Il pourrait être pertinent d'étudier l'impact d'une telle consommation sur les profils métabolomiques et les comparer avec les composantes nutritionnelles s'y rapportant (exemple : comparer l'impact de la consommation de vitamines issues des compléments *versus* de l'alimentation sur les profils métabolomiques). Pour étudier de tels aspects plus spécifiques de l'alimentation, des approches plus ciblées peuvent être envisagées notamment par métabolomique sous forme de Kit ou encore par la lipidomique ou protéomique.

Grâce à la biobanque de Nutrinet-Santé, il serait envisageable d'élargir la thématique de recherche à la génétique permettant d'étudier plus spécifiquement les liens entre facteurs extérieurs ou processus endogènes et composante génétique et l'impact sur le risque de cancer. En effet, des échantillons de buffy-coat (fraction de sang non coagulée après centrifugation par densité de gradient) sont disponibles permettant d'étudier divers polymorphismes génétiques. Par ailleurs, dans le cadre de cette thèse nous

avons pu mettre en évidence des métabolites issus de l'activité du microbiote intestinal associés au risque de cancer du sein mais également à la qualité de l'alimentation. Ceci évoque la pertinence d'étudier plus particulièrement cette thématique afin de mieux caractériser les interactions complexes entre nutrition, composition du microbiote (abondance, diversité) et santé de l'hôte. Notamment, dans une autre étude à laquelle j'ai participé nous avons pu mettre en évidence certains métabolites plasmatiques (issus d'analyses en H-RMN) associés à plusieurs caractéristiques du microbiote intestinal établies par séquençage du gène 16S de l'ARNr dans des échantillons de selles d'une cohorte française de participants en bonne santé (étude *Milieu Intérieur*). La biobanque de Nutrinet-Santé dispose pour le moment d'échantillons de selles de 150 participants.

## B. Etudes complémentaires

Une seule étude, quelle qu'en soit la méthodologie, ne suffit pas à tirer des conclusions fermes sur les résultats obtenus. Des études mobilisant des méthodologies différentes (études d'observation, d'intervention et mécanistiques) sont nécessaires pour qualifier le niveau de preuve des associations observées (approche portfolio) (16). Dans un premier temps, les résultats obtenus dans le cadre de cette thèse soulèvent des hypothèses qu'il est nécessaire de valider dans d'autres études d'observation menées dans des cohortes incluant un nombre suffisant de participants pour bénéficier d'une puissance statistique adéquate dans les analyses étiologiques. Ces jeux de données indépendants peuvent notamment permettre d'évaluer la performance prédictive de signatures métabolomiques de manière non biaisée. En outre, bien qu'une validation dans la cohorte Nutrinet-Santé soit une première étape dans le processus de validation, il est nécessaire de répliquer ses résultats dans des pays différents afin de pouvoir les extrapoler à diverses populations ayant des modes de vie et comportements alimentaires divers et variables.

Des analyses poolées permettent de regrouper les données obtenues dans différentes populations. Notamment, des groupes collaboratifs existent et prennent en charge ce type d'analyse. Par exemple, le groupe collaboratif « Endogenous Hormones, Nutritional Biomarkers and Prostate Cancer Collaborative Group » dont l'EREN fait partie, effectue des analyses poolées des relations entre les hormones sexuelles endogènes, les facteurs de croissance analogues à l'insuline, les biomarqueurs nutritionnels et le risque de cancer de la prostate.

Les résultats issus de ces études et analyses poolées permettront de mettre en avant les métabolites discriminants les plus robustes et qui nécessiteront des études complémentaires plus spécifiques.

Pour les biomarqueurs de risque candidats, des études expérimentales, *in vivo* ou *in vitro* permettront notamment d'investiguer les hypothèses mécanistiques et biologiques soulevées (16). Elles pourront fournir davantage d'éléments pour la détermination de l'origine des variations observées impliquant une augmentation ou diminution de risque. En plus des études d'observation, elles permettront de répondre aux questions : ces potentiels biomarqueurs sont-ils sur le chemin causal de la pathologie ? Permettent-ils d'améliorer la prédiction du risque ou sont-ils simplement associés au risque ?

Pour les biomarqueurs associés à l'alimentation, des études d'interventions sont nécessaires afin d'évaluer, entre autres, le potentiel modifiable de ces biomarqueurs. Afin de mettre en évidence des biomarqueurs de l'exposition nutritionnelle, plusieurs types d'intervention sont possibles. Elles peuvent être réalisées à court ou long terme, sur des modèles animaux ou humains. Cependant, si l'hypothèse d'exposition est susceptible d'avoir des effets néfastes, les modèles humains ne pourront pas être utilisés. Lors de ces interventions, il faudra notamment évaluer des effets de dose/réponse (effets seuils ?), si les variations dépendent des facteurs de l'hôte, s'ils sont spécifiques d'une exposition ou s'il est préférable qu'ils soient utilisés en combinaison avec d'autres biomarqueurs, le niveau basal habituel (idéalement nul

si pas d'exposition) ou encore sa durée de vie et sa cinétique. L'origine du biomarqueur est une question essentielle. En effet, les métabolites transformés en nutriments simples et absorbés sont en général endogènes (sauf exception comme les acides aminés et acides gras essentiels), leur origine n'est donc pas assez spécifique. Si ces métabolites sont traités comme corps étrangers (xénobiotiques), ils pourraient constituer de bons candidats (non synthétisable par l'organisme). Si ces métabolites sont issus du métabolisme microbien, ils pourraient être considérés comme d'éventuels biomarqueurs d'exposition, cependant des précautions sont nécessaires car les interactions entre microbiote, nutrition et santé sont complexes (51).

Pour l'identification de potentiels biomarqueurs d'effet (impact de la nutrition sur le métabolisme et la santé), des études d'intervention à plus long terme, se rapprochant au maximum des conditions réelles sont nécessaires ; là encore, en respectant les conditions d'éthique. Pour ce type de biomarqueurs, il est préférable d'utiliser des échantillons biologiques répétés sur un même individu afin de prendre en compte la variabilité intra-individuelle et contrôler les facteurs qui contribuent aux variations physiologiques normales. De cette façon, les fluctuations métaboliques basales ne sont pas confondues avec les biomarqueurs qui représentent un changement métabolique dû à une intervention nutritionnelle (51,125).

En plus d'autres études d'observation du même type que celles de cette thèse et incluant des analyses de médiation, toutes ces informations collectées par ces études complémentaires pourront servir à l'identification de biomarqueurs à la fois liés à l'alimentation et au risque de cancer ou autre pathologie. Dans une optique de prévention, seuls les métabolites modifiables après une intervention seront d'intérêt afin de diminuer le risque. Ces informations pourront là aussi permettre de répondre aux questions ; ces potentiels biomarqueurs sont-ils sur le chemin causal de la pathologie et l'exposition ? Permettent-ils d'améliorer la prédiction du risque ou sont-ils simplement associés au risque ?

Des méta-analyses et des travaux d'expertise collective permettront d'agréger les différents résultats disponibles et d'établir des niveaux de preuve associés et, le cas échéant, d'établir des recommandations (16). Afin d'être validés comme biomarqueurs, les différents candidats devront répondre à un certain nombre de critères, à savoir la plausibilité biologique, la robustesse, la fiabilité, la stabilité, la sensibilité, la spécificité et la reproductibilité (63).

Ce domaine de recherche nécessite l'apport de spécialistes de diverses disciplines, dont la chimométrie, les biostatistiques, la bioinformatique, la nutrition, l'épidémiologie et la biologie. Avec l'évolution rapide des techniques de profilage métabolomique, il est nécessaire, non seulement de développer des collaborations multidisciplinaires au sein des projets, mais également de pouvoir échanger de nouvelles informations et de nouveaux outils, grâce notamment à des réseaux. Ces réseaux peuvent notamment participer à la création de bases de données permettant de compiler toutes les informations nécessaires (données spectrales, bibliothèques de standards chimiques, algorithmes pour l'analyse des données, dépôts de données brutes et de métadonnées) et initier des protocoles de normalisation pour définir les bonnes pratiques (51). La Metabolomics Standards Initiative (MSI) (381) lancée par la Metabolomics Society et le Consortium of METabolomics Studies (COMETS) (382) lancée par l'institut national du cancer aux USA sont des exemples d'une telle initiative. En effet, afin que la comparaison des résultats issus de diverses populations soit la plus pertinente possible, la standardisation des protocoles analytiques et des analyses statistiques serait nécessaire.

De nombreux programmes européens et internationaux comme le Joint Programme Initiative (JPI) permettent de réunir de nombreuses disciplines et d'inclure dans un même projet des études

complémentaires (méthodologique, observation, intervention par exemple). Par exemple, le projet FoodBALL - Food Biomarkers Alliance - est un projet multicentrique financé par le JPI qui vise à développer des stratégies de découverte et de validation de biomarqueurs alimentaires (383). Le projet OXYGENATE, également financé par le JPI, vise à découvrir et valider des signatures d'oxylipines (superclasse de médiateurs lipidiques produits à partir de l'oxygénation d'acides gras polyinsaturés comme les prostanoïdes) qui soient à la fois influencées par l'alimentation et liées au syndrome cardiométabolique (384). L'EREN est leader d'un volet dans ce projet européen et j'ai donc eu l'occasion de contribuer à cette collaboration au cours de ma thèse (voir encadré sur ci-dessous).

#### **Le projet Oxygenate**

Identifier et valider des signatures d'oxylipines (i) permettant de discriminer des individus à différents stades du syndrome métabolique et des individus ayant des habitudes alimentaires de qualité différente et (ii) modifiables suite à une intervention nutritionnelle affectant les paramètres cardiométaboliques.

- *Coordinateur : Dr Cécile Gladine INRA/UMR 1019 Unité de Nutrition Humaine, Nutrition & Fonction vasculaire, France ; Expertise : nutritionniste*

#### **Objectifs**

WP1. Optimiser le profilage des oxylipines par lipidomique ciblée en (i) élargissant le profil actuel des oxylipines, (ii) déterminant les conditions optimales de stockage et de préparation, (iii) effectuant le transfert de méthodes et la validation inter-laboratoires et (iv) effectuant des tests d'automatisation pour les étapes de limitation de débit.

- *Leader : Dr Nils Schebb Univ. of Hannover, Institut de chimie alimentaire, Allemagne; Expertise : chimiste analytique*

WP2. Identifier, dans une première cohorte (PURE), les oxylipines candidates reflétant la transition cardiométabolique et ses relations avec l'alimentation en mesurant le profil complet des oxylipines chez les individus (i) ayant des régimes alimentaires différents et (ii) avec ou sans syndrome métabolique (étude transversale) ainsi que (iii) avant et après développement de syndrome métabolique (analyse longitudinale).

- *Co-leader : Dr Cécile Gladine*
- *Co-leader : Pr Andrzej Szuba Univ. Wroclaw, Département médecine interne, Pologne; Expertise : Professeur de médecine*

WP3. Valider la fiabilité et la reproductibilité des oxylipines candidates identifiées (i) en répliquant les analyses transversales dans une cohorte indépendante (NutriNet-Santé) et (ii) en mesurant les variations des oxylipines candidates en réponse à une intervention alimentaire en utilisant les biobanques existantes de 2 études d'intervention réalisées chez des individus ayant un statut cardiométabolique perturbé : SHOPUS (Department for Nutrition, Exercise and Sports, université de Copenhague, Danemark) et iMAPS (USDA Davis).

- *Leader : Dr Mathilde Touvier, EREN, France ; Expertise : épidémiologiste*

#### **Contribution personnelle**

- Sélection des populations d'études transversales (études appariées) dans la cohorte PURE et Nutrinet-Santé
- Participation aux réunions de projets
- Présentation du protocole, des sélections des populations et des analyses préliminaires à la réunion annuelle de 2018 à Bruxelles
- Elaboration de protocole d'analyses statistiques sur les données issues des cohortes en collaboration avec l'équipe coordinatrice (UNH, UMR 1019)

### C. Impacts et applications futures

L'identification et la validation de biomarqueurs dans cette thématique de recherche auront divers impacts. Dans un premier temps, la mise en évidence de biomarqueurs de risque pourra déboucher sur la recherche de nouvelles cibles thérapeutiques comme par exemple, des dispositifs permettant de réguler le taux plasmatique de divers métabolites (dans le cadre de biomarqueurs causalement associés au risque) et sur la surveillance à l'échelle de la santé publique de certains paramètres biologiques.

Cette approche exploratoire peut permettre d'émettre des hypothèses sur de potentielles nouvelles molécules cancérigènes qui, suite à des études complémentaires, pourraient amener à la mise en place de nouvelles mesures de santé publique afin de limiter l'exposition à ces substances.

Les biomarqueurs d'exposition nutritionnelle pourront permettre l'amélioration de l'évaluation nutritionnelle dans le cadre d'études épidémiologiques débouchant sur des résultats et une comparaison plus robuste entre les études. Certains de ces biomarqueurs, mesurés en combinaison pourraient également évaluer l'exposition à des composantes plus globales de l'alimentation (régimes alimentaires).

La mise en évidence de biomarqueurs reflétant l'impact de l'alimentation sur le métabolisme pourrait contribuer à une meilleure compréhension des relations nutrition santé et éventuellement déboucher sur l'élaboration de nouvelles recommandations nutritionnelles ou dispositifs de santé publique afin de diminuer l'incidence de ces pathologies par l'alimentation.

Enfin, dans un contexte de prévention personnalisée, il serait possible d'envisager des kits de dosage de paramètres biologiques renseignant sur le risque de développer un cancer dans les années à venir et sur lesquels il est possible d'agir grâce à une intervention, notamment nutritionnelle. Plusieurs études ont montré que les modèles prédictifs actuels concernant le risque de cancer sont peu nombreux, manquent de validation externe et de précision et sont peu performants (39,40). De plus, ils ne contiennent que peu ou pas de facteurs modifiables. A terme, il serait donc pertinent d'évaluer le gain de performance de tels modèles en y ajoutant une signature métabolique sur laquelle il est possible d'agir pour modifier le risque. Cependant, la prévention personnalisée possède plusieurs limites. Comme suggéré précédemment, cette approche n'a d'intérêt que s'il existe des moyens efficaces pour lutter contre le risque ; ce genre d'approche pouvant devenir anxiogène. De plus, de nombreuses personnes ne sont pas forcément prêtes à changer leur comportement. Un suivi poussé est donc nécessaire. Le développement de kits de dosage est très coûteux et complexe à mettre en place (âge des premiers tests, fréquence, durée) et sera coûteux également pour les bénéficiaires. Dans un contexte international, il pourrait entraîner une discrimination notamment dans les pays en développement où les interventions ultérieures resteront dans tous les cas limitées (45).

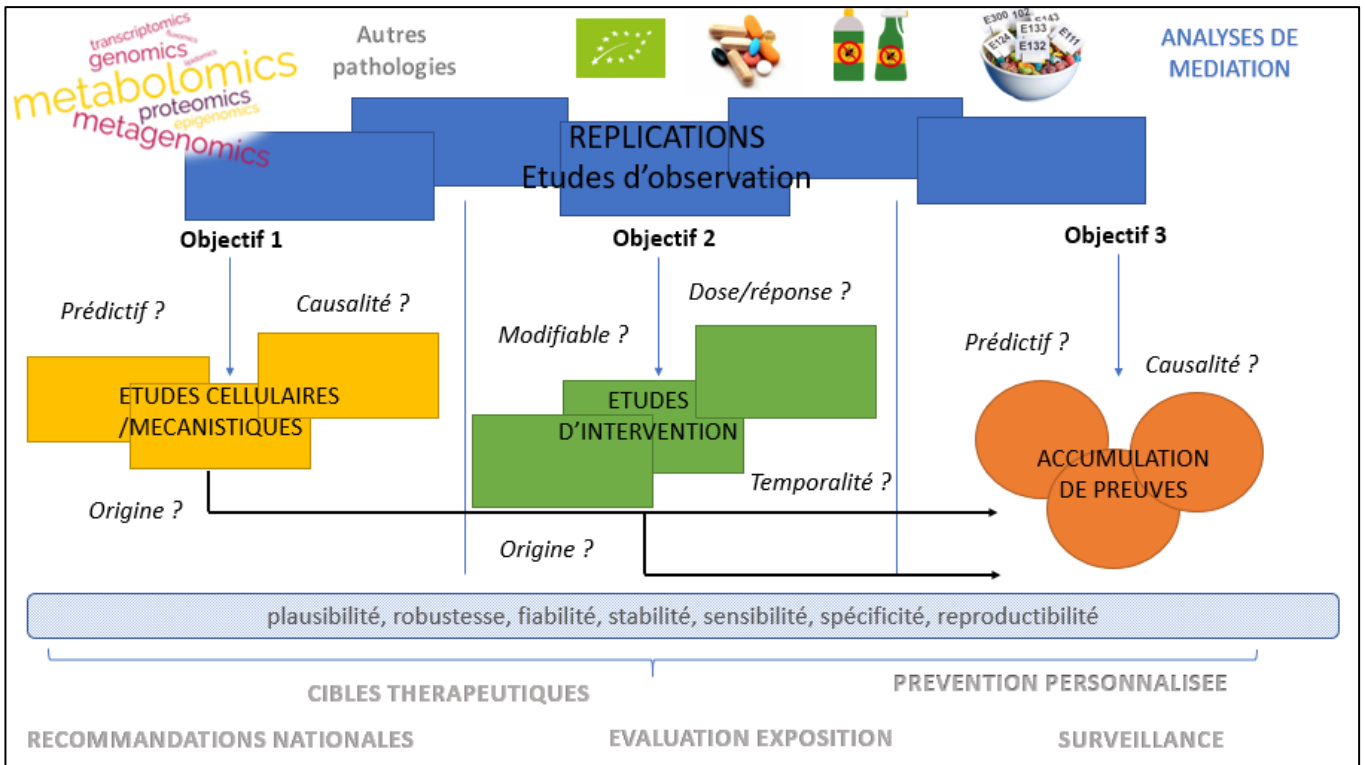


Figure 27 Synthèse des perspectives de recherche sur la thématique



## Conclusion

---

Cette thèse s'inscrit dans un contexte où l'incidence des cancers du sein et de la prostate est un problème majeur de santé publique et où de nouveaux outils de technologie à haut débit sont maintenant disponibles pour permettre de mieux comprendre l'impact de l'environnement extérieur (au sens large, incluant la nutrition) et du métabolisme sur leur étiologie et d'améliorer l'identification de personnes à plus haut risque. Nous avons notamment utilisé le profilage métabolomique de manière exploratoire afin de mettre en évidence des signatures plasmatiques du risque de ces cancers et de l'alimentation. Nous avons identifié une combinaison de métabolites endogènes et de métabolites issus du métabolisme microbien qui semblerait refléter l'impact de l'alimentation globale par ses effets bénéfiques ou néfastes à long terme sur l'état nutritionnel et métabolique des individus. Nous avons également pu mettre en évidence de potentiels biomarqueurs candidats de l'exposition alimentaire.

Par ailleurs, nous avons identifié des métabolites associés au risque de cancers du sein et de la prostate, principalement synthétisables par l'organisme mais également reflétant l'exposition environnementale ou microbienne et suggérant une perturbation métabolique jusqu'à 13 ans avant le diagnostic du cancer. Quelques métabolites plasmatiques sembleraient permettre de discriminer les individus à plus haut risque de développer un cancer du sein ou de la prostate et pour lesquels les facteurs alimentaires pourraient jouer un rôle dans leur variation.

Bien que l'origine des variations plasmatiques de la plupart des métabolites observés ne peut être déterminée dans le cadre de cette thèse, nos résultats montrent que la métabolomique permet d'intégrer des informations concernant l'influence des facteurs extrinsèques et intrinsèques, comme l'exposition au mode de vie et le métabolisme de base, et aborde ainsi la question complexe des interactions environnement-santé.

Encore relativement peu d'études ont été réalisées sur cette thématique, nos résultats sont donc à répliquer dans d'autres cohortes. Cependant, une standardisation des procédures semble nécessaire pour une comparaison optimale des résultats. Ce type de signatures permettra de mettre en évidence de nouvelles hypothèses qui feront la cible d'études expérimentales afin de valider les observations et d'établir par la suite des biomarqueurs robustes du risque de cancers du sein et de la prostate, de l'impact de l'alimentation globale sur le métabolisme, de l'exposition nutritionnelle et de l'influence de l'alimentation sur le risque de ces deux cancers.

Ces travaux ouvrent de nombreuses perspectives de recherche, notamment l'étude de composantes plus spécifiques de l'alimentation comme les additifs alimentaires, les pesticides ou les compléments alimentaires, l'extrapolation de la thématique à d'autres méthodes de omiques ou encore la transposition à d'autres localisations de cancer afin de mettre en évidence des signatures communes à plusieurs localisations de cancers ou au contraire spécifiques d'un organe.

Les résultats de cette thèse s'inscrivant dans le cadre de la prévention primaire des cancers pourraient permettre de contribuer à terme à l'amélioration de la compréhension des relations entre exposition (notamment à l'alimentation), métabolisme et cancérogenèse, ainsi qu'une optimisation de l'évaluation de l'exposition nutritionnelle voire la mise en place de nouvelles recommandations en matière de santé publique en vue de la diminution de l'incidence de ces pathologies.

## Références

---

1. Bray F, Jemal A, Torre LA, Forman D, Vineis P. Long-term Realism and Cost-effectiveness: Primary Prevention in Combatting Cancer and Associated Inequalities Worldwide. *J Natl Cancer Inst.* 2015 Dec;107(12):djv273.
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018 Sep 12;
3. National Cancer Institute. Cancer Statistics [Internet]. 2018 [cited 2019 Apr 2]. Available from: <https://www.cancer.gov/about-cancer/understanding/statistics>
4. Jemal A, editor. The cancer atlas. Second edition. Atlanta, Georgia: American Cancer Society/Health Promotion; 2015.
5. Wu S, Zhu W, Thompson P, Hannun YA. Evaluating intrinsic and non-intrinsic cancer risk factors. *Nat Commun.* 2018 28;9(1):3490.
6. WHO | Cancer prevention [Internet]. WHO. [cited 2019 Feb 18]. Available from: <http://www.who.int/cancer/prevention/en/>
7. Vineis P, Wild CP. Global cancer patterns: causes and prevention. *Lancet Lond Engl.* 2014 Feb 8;383(9916):549–57.
8. World Health Organization. Cancer Control: Knowledge into Action: WHO Guide for Effective Programmes: Module 2: Prevention [Internet]. Geneva: WHO Press; 2007 [cited 2019 Apr 2]. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK195368/>
9. International Agency for Research on Cancer. IARC Medium-Term Strategy for 2016–2020, Including Implementation Plans. IARC; 2017.
10. World Cancer Research Fund International/American Institute for Cancer Research. Diet, Nutrition, Physical Activity and Cancer: a Global Perspective [Internet]. Continuous Update Project Expert Report 2018. Our Cancer Prevention Recommendations. [cited 2019 Apr 2]. Available from: <https://www.wcrf.org/dietandcancer>
11. IARC. Les cancers attribuables au mode de vie et à l'environnement en France métropolitaine. [Internet]. Lyon: International Agency for Research on Cancer; 2018. Available from: [http://gco.iarc.fr/includes/PAF/PAF\\_FR\\_report.pdf](http://gco.iarc.fr/includes/PAF/PAF_FR_report.pdf)
12. Réseau NACRe - Réseau National Alimentation Cancer Recherche - Cancers attribuables aux facteurs nutritionnels [Internet]. [cited 2019 Apr 8]. Available from: <https://www6.inra.fr/nacre/Prevention-primaire/Vous-informer-sur/Cancers-attribuables-aux-facteurs-nutritionnels>
13. Shield KD, Freisling H, Boutron-Ruault M-C, Touvier M, Marant Micallef C, Jenab M, et al. New cancer cases attributable to diet among adults aged 30-84 years in France in 2015. *Br J Nutr.* 2018 Nov;120(10):1171–80.
14. Christopher Paul Wild. Precision in the Fight Against the Global Cancer Problem [Internet]. 2015. Available from: <https://healthmanagement.org/c/healthmanagement/issuearticle/precision-in-the-fight-against-the-global-cancer-problem>

15. World Cancer Research Fund/American Institute for Cancer Research. Continuous Update Project Expert Report 2018. The cancer process. [Internet]. 2018. Available from: [dietandcancerreport.org](http://dietandcancerreport.org)
16. ANSES, rapport d'expertise collective. Nutrition et cancer : légitimité de recommandations nutritionnelles dans le cadre de la prévention des cancers. 2011;78.
17. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011 Mar 4;144(5):646–74.
18. World Cancer Research Fund/American Institute for Cancer Research. Continuous Update Project Expert Report 2018. Diet, nutrition, physical activity and prostate cancer. [Internet]. 2014. Available from: <https://www.wcrf.org/sites/default/files/Prostate-cancer-report.pdf>
19. Rowlands M-A, Gunnell D, Harris R, Vatten LJ, Holly JMP, Martin RM. Circulating insulin-like growth factor peptides and prostate cancer risk: a systematic review and meta-analysis. *Int J Cancer*. 2009 May 15;124(10):2416–29.
20. Roddam AW, Allen NE, Appleby P, Key TJ, Ferrucci L, Carter HB, et al. Insulin-like growth factors, their binding proteins, and prostate cancer risk: analysis of individual patient data from 12 prospective studies. *Ann Intern Med*. 2008 Oct 7;149(7):461–71, W83-88.
21. World Cancer Research Fund/American Institute for Cancer Research. Continuous Update Project Expert Report 2018. Diet, nutrition, physical activity and breast cancer. [Internet]. 2017. Available from: <https://www.wcrf.org/sites/default/files/Breast-cancer-report.pdf>
22. Modan B, Chetrit A, Alfandary E, Katz L. Increased risk of breast cancer after low-dose irradiation. *Lancet Lond Engl*. 1989 Mar 25;1(8639):629–31.
23. Ronckers CM, Erdmann CA, Land CE. Radiation and breast cancer: a review of current evidence. *Breast Cancer Res BCR*. 2005;7(1):21–32.
24. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Combined estrogen-progestogen contraceptives and combined estrogen-progestogen menopausal therapy. *IARC Monogr Eval Carcinog Risks Hum*. 2007;91:1–528.
25. Reeves GK, Beral V, Green J, Gathani T, Bull D, Million Women Study Collaborators. Hormonal therapy for menopause and breast-cancer risk by histological type: a cohort study and meta-analysis. *Lancet Oncol*. 2006 Nov;7(11):910–8.
26. Benkhadra K, Mohammed K, Al Nofal A, Carranza Leon BG, Alahdab F, Faubion S, et al. Menopausal Hormone Therapy and Mortality: A Systematic Review and Meta-Analysis. *J Clin Endocrinol Metab*. 2015 Nov;100(11):4021–8.
27. Cavicchia PP, Steck SE, Hurley TG, Hussey JR, Ma Y, Ockene IS, et al. A New Dietary Inflammatory Index Predicts Interval Changes in Serum High-Sensitivity C-Reactive Protein. *J Nutr*. 2009 Dec 1;139(12):2365–72.
28. Zahedi H, Djalalinia S, Sadeghi O, Asayesh H, Noroozi M, Gorabi AM, et al. Dietary Inflammatory Potential Score and Risk of Breast Cancer: Systematic Review and Meta-analysis. *Clin Breast Cancer*. 2018;18(4):e561–70.
29. Fowler ME, Akinyemiju TF. Meta-analysis of the association between dietary inflammatory index (DII) and cancer outcomes. *Int J Cancer*. 2017 01;141(11):2215–27.

30. Kim JH, Kim J. Index-Based Dietary Patterns and the Risk of Prostate Cancer. *Clin Nutr Res*. 2017 Oct;6(4):229–46.
31. Bastide NM, Pierre FHF, Corpet DE. Heme Iron from Meat and Risk of Colorectal Cancer: A Meta-analysis and a Review of the Mechanisms Involved. *Cancer Prev Res (Phila Pa)*. 2011 Feb 1;4:177–84.
32. Druesne-Pecollo N, Latino-Martel P, Hercberg S. Prévention nutritionnelle des cancers axée sur les comportements alimentaires. *Réalités en nutrition*. 2007;(4):31–6.
33. Tao W, Lagergren J. Clinical management of obese patients with cancer. *Nat Rev Clin Oncol*. 2013 Sep;10(9):519–33.
34. Tucker KL. Assessment of usual dietary intake in population studies of gene-diet interaction. *NutrMetab Cardiovasc*. 2007 Feb;17:74–81.
35. Bingham SA, Luben R, Welch A, Wareham N, Khaw KT, Day N. Are imprecise methods obscuring a relation between fat and breast cancer? *Lancet*. 2003 Jul 19;362:212–4.
36. Bingham S, Luben R, Welch A, Low YL, Khaw KT, Wareham N, et al. Associations between dietary methods and biomarkers, and between fruits and vegetables and risk of ischaemic heart disease, in the EPIC Norfolk Cohort Study. *IntJEpidemiol*. 2008 Oct;37:978–87.
37. Freedman LS, Potischman N, Kipnis V, Midthune D, Schatzkin A, Thompson FE, et al. A comparison of two dietary instruments for evaluating the fat-breast cancer relationship. *IntJEpidemiol*. 2006 Aug;35:1011–21.
38. Kristal AR, Peters U, Potter JD. Is it time to abandon the food frequency questionnaire? *Cancer EpidemiolBiomarkers Prev*. 2005 Dec;14:2826–8.
39. Meads C, Ahmed I, Riley RD. A systematic review of breast cancer incidence risk prediction models with meta-analysis of their performance. *Breast Cancer Res Treat*. 2012 Apr;132(2):365–77.
40. Alblas M, Velt KB, Pashayan N, Widschwendter M, Steyerberg EW, Vergouwe Y. Prediction models for endometrial cancer for the general population or symptomatic women: A systematic review. *Crit Rev Oncol Hematol*. 2018 Jun;126:92–9.
41. Hill AB. THE ENVIRONMENT AND DISEASE: ASSOCIATION OR CAUSATION? *Proc R Soc Med*. 1965 May;58:295–300.
42. Section of Mechanisms of Carcinogenesis – IARC [Internet]. [cited 2019 Jul 5]. Available from: <https://www.iarc.fr/research-groups-mca-mmb-rationale/>
43. Spitz MR, Bondy ML. The evolving discipline of molecular epidemiology of cancer. *Carcinogenesis*. 2010 Jan;31(1):127–34.
44. Wild CP. Complementing the Genome with an ‘Exposome’: The Outstanding Challenge of Environmental Exposure Measurement in Molecular Epidemiology. *Cancer Epidemiol Biomarkers Prev*. 2005 Aug 1;14(8):1847–50.
45. Wild CP. The exposome: from concept to utility. *Int J Epidemiol*. 2012 Feb;41(1):24–32.

46. Schulte PA, Perera FP. *Molecular Epidemiology: Principles and Practices*. San Diego, CA, USA: Academic Press, Inc.; 1993.
47. Athersuch TJ, Keun HC. Metabolic profiling in human exposome studies. *Mutagenesis*. 2015 Nov;30(6):755–62.
48. Athersuch T. Metabolome analyses in exposome studies: Profiling methods for a vast chemical space. *Arch Biochem Biophys*. 2016 Jan 1;589:177–86.
49. Rutishauser IHE. Dietary intake measurements. *Public Health Nutr*. 2005 Oct;8(7A):1100–7.
50. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, et al. HMDB 3.0--The Human Metabolome Database in 2013. *Nucleic Acids Res*. 2013 Jan;41:D801–7.
51. Scalbert A, Brennan L, Manach C, Andres-Lacueva C, Dragsted LO, Draper J, et al. The food metabolome: a window over dietary exposure. *AmJClinNutr*. 2014 Jun;99(6):1286–308.
52. Fardet A, Llorach R, Orsoni A, Martin J-F, Pujos-Guillot E, Lapiere C, et al. Metabolomics provide new insight on the metabolism of dietary phytochemicals in rats. *J Nutr*. 2008 Jul;138(7):1282–7.
53. Nicholson JK, Holmes E, Kinross J, Burcelin R, Gibson G, Jia W, et al. Host-gut microbiota metabolic interactions. *Science*. 2012 Jun 8;336:1262–7.
54. Gao Q, Praticò G, Scalbert A, Vergères G, Kolehmainen M, Manach C, et al. A scheme for a flexible classification of dietary and health biomarkers. *Genes Nutr*. 2017;12:34.
55. Biesalski H-K, Dragsted LO, Elmadfa I, Grossklaus R, Müller M, Schrenk D, et al. Bioactive compounds: definition and assessment of activity. *Nutr Burbank Los Angel Cty Calif*. 2009 Dec;25(11–12):1202–5.
56. Aggett PJ, Antoine J-M, Asp N-G, Bellisle F, Contor L, Cummings JH, et al. PASSCLAIM: consensus on criteria. *Eur J Nutr*. 2005 Jun;44 Suppl 1:i5-30.
57. Vineis P, Perera F. Molecular epidemiology and biomarkers in etiologic cancer research: the new in light of the old. *Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol*. 2007 Oct;16(10):1954–65.
58. Boffetta P. Biomarkers in cancer epidemiology: an integrative approach. *Carcinogenesis*. 2010 Jan;31(1):121–6.
59. Jenab M, Slimani N, Bictash M, Ferrari P, Bingham SA. Biomarkers in nutritional epidemiology: applications, needs and new horizons. *HumGenet*. 2009 Jun;125:507–25.
60. Kaaks R, Ferrari P, Ciampi A, Plummer M, Riboli E. Uses and limitations of statistical accounting for random error correlations, in the validation of dietary questionnaire assessments. *Public Health Nutr*. 2002 Dec;5(6A):969–76.
61. Sugar EA, Wang CY, Prentice RL. Logistic regression with exposure biomarkers and flexible measurement error. *Biometrics*. 2007 Mar;63:143–51.
62. Day N, McKeown N, Wong M, Welch A, Bingham S. Epidemiological assessment of diet: a comparison of a 7-day diary with a food frequency questionnaire using urinary markers of nitrogen, potassium and sodium. *Int J Epidemiol*. 2001 Apr;30(2):309–17.

63. Dragsted LO, Gao Q, Scalbert A, Vergères G, Kolehmainen M, Manach C, et al. Validation of biomarkers of food intake-critical assessment of candidate biomarkers. *Genes Nutr.* 2018;13:14.
64. Freedman LS, Tasevska N, Kipnis V, Schatzkin A, Mares J, Tinker L, et al. Gains in statistical power from using a dietary biomarker in combination with self-reported intake to strengthen the analysis of a diet-disease association: an example from CAREDS. *Am J Epidemiol.* 2010 Oct 1;172(7):836–42.
65. Moco S, Martin F-PJ, Rezzi S. Metabolomics view on gut microbiome modulation by polyphenol-rich foods. *J Proteome Res.* 2012 Oct 5;11(10):4781–90.
66. Perez-Jimenez J, Hubert J, Hooper L, Cassidy A, Manach C, Williamson G, et al. Urinary metabolites as biomarkers of polyphenol intake in humans: a systematic review. *AmJClinNutr.* 2010 Oct;92:801–9.
67. Brouwer-Brolsma EM, Brennan L, Drevon CA, van Kranen H, Manach C, Dragsted LO, et al. Combining traditional dietary assessment methods with novel metabolomics techniques: present efforts by the Food Biomarker Alliance. *Proc Nutr Soc.* 2017;76(4):619–27.
68. Neveu V, Moussy A, Rouaix H, Wedekind R, Pon A, Knox C, et al. Exposome-Explorer: a manually-curated database on biomarkers of exposure to dietary and environmental factors. *Nucleic Acids Res.* 2017 Jan 4;45(D1):D979–84.
69. Potischman N. Biologic and Methodologic Issues for Nutritional Biomarkers. *J Nutr.* 2003 Mar 1;133(3):875S–880S.
70. Brennan L, Hu FB. Metabolomics-Based Dietary Biomarkers in Nutritional Epidemiology-Current Status and Future Opportunities. *Mol Nutr Food Res.* 2019 Jan;63(1):1701064.
71. de Koning L, Chiuve SE, Fung TT, Willett WC, Rimm EB, Hu FB. Diet-quality scores and the risk of type 2 diabetes in men. *Diabetes Care.* 2011 May;34(5):1150–6.
72. InterAct Consortium. Adherence to predefined dietary patterns and incident type 2 diabetes in European populations: EPIC-InterAct Study. *Diabetologia.* 2014 Feb;57(2):321–33.
73. Nanri A, Mizoue T, Shimazu T, Ishihara J, Takachi R, Noda M, et al. Dietary patterns and all-cause, cancer, and cardiovascular disease mortality in Japanese men and women: The Japan public health center-based prospective study. *PLoS One.* 2017;12(4):e0174848.
74. World Health Organization (WHO). *Guideline: Dietary Patterns.* WHO: Geneva, Switzerland; 2017.
75. Cespedes EM, Hu FB. Dietary patterns: from nutritional epidemiologic analysis to national guidelines. *Am J Clin Nutr.* 2015 May;101(5):899–900.
76. McMichael AJ, Hall AJ. The use of biological markers as predictive early-outcome measures in epidemiological research. *IARC Sci Publ.* 1997;(142):281–9.
77. Valeri L, Vanderweele TJ. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. *Psychol Methods.* 2013 Jun;18(2):137–50.
78. MacKinnon DP. *Introduction to statistical mediation analysis.* 2008.

79. Gupta V, Walia GK, Sachdeva MP. 'Mendelian randomization': an approach for exploring causal relations in epidemiology. *Public Health*. 2017 Apr;145:113–9.
80. Bonassi S, Au WW. Biomarkers in molecular epidemiology studies for health risk prediction. *Mutat Res Mutat Res*. 2002 Mar;511(1):73–86.
81. Corella D, Ordovás JM. Biomarkers: background, classification and guidelines for applications in nutritional epidemiology. *Nutr Hosp*. 2015 Feb 26;31 Suppl 3:177–88.
82. Miller AB, International Agency for Research on Cancer, editors. *Biomarkers in cancer chemoprevention*. Lyon: International Agency for Research on Cancer; 2001. 294 p. (IARC scientific publications).
83. Spitz MR, Etzel CJ, Dong Q, Amos CI, Wei Q, Wu X, et al. An expanded risk prediction model for lung cancer. *Cancer Prev Res Phila Pa*. 2008 Sep;1(4):250–4.
84. Li G, Hu J, Hu G. Biomarker Studies in Early Detection and Prognosis of Breast Cancer. *Adv Exp Med Biol*. 2017;1026:27–39.
85. Walsh MF, Nathanson KL, Couch FJ, Offit K. Genomic Biomarkers for Breast Cancer Risk. *Adv Exp Med Biol*. 2016;882:1–32.
86. Verma M. Biomarkers for risk assessment in molecular epidemiology of cancer. *Technol Cancer Res Treat*. 2004 Oct;3(5):505–14.
87. Terry MB, McDonald JA, Wu HC, Eng S, Santella RM. Epigenetic Biomarkers of Breast Cancer Risk: Across the Breast Cancer Prevention Continuum. *Adv Exp Med Biol*. 2016;882:33–68.
88. Kleibl Z, Kristensen VN. Women at high risk of breast cancer: Molecular characteristics, clinical presentation and management. *Breast Edinb Scotl*. 2016 Aug;28:136–44.
89. Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, et al. Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci U S A*. 2008 Jul 29;105(30):10513–8.
90. Lee JD, Cai Q, Shu XO, Nechuta SJ. The Role of Biomarkers of Oxidative Stress in Breast Cancer Risk and Prognosis: A Systematic Review of the Epidemiologic Literature. *J Womens Health* 2002. 2017;26(5):467–82.
91. Key T, Appleby P, Barnes I, Reeves G, Endogenous Hormones and Breast Cancer Collaborative Group. Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. *J Natl Cancer Inst*. 2002 Apr 17;94(8):606–16.
92. Chlebowski RT, Kuller LH, Prentice RL, Stefanick ML, Manson JE, Gass M, et al. Breast cancer after use of estrogen plus progestin in postmenopausal women. *N Engl J Med*. 2009 Feb 5;360(6):573–87.
93. Uray IP, Brown PH. Prevention of breast cancer: current state of the science and future opportunities. *Expert Opin Investig Drugs*. 2006 Dec;15(12):1583–600.
94. Rose DP, Vona-Davis L. The cellular and molecular mechanisms by which insulin influences breast cancer risk and progression. *Endocr Relat Cancer*. 2012 Dec;19(6):R225–241.

95. Patterson RE, Rock CL, Kerr J, Natarajan L, Marshall SJ, Pakiz B, et al. Metabolism and breast cancer risk: frontiers in research and practice. *J Acad Nutr Diet*. 2013 Feb;113(2):288–96.
96. Pan H, Deng L-L, Cui J-Q, Shi L, Yang Y-C, Luo J-H, et al. Association between serum leptin levels and breast cancer risk: An updated systematic review and meta-analysis. *Medicine (Baltimore)*. 2018 Jul;97(27):e11345.
97. Makridakis NM, Reichardt JKV. Molecular epidemiology of androgen-metabolic loci in prostate cancer: predisposition and progression. *J Urol*. 2004 Feb;171(2 Pt 2):S25-28; discussion S28-29.
98. Fleshner NE, Lawrentschuk N. Risk of developing prostate cancer in the future: overview of prognostic biomarkers. *Urology*. 2009 May;73(5 Suppl):S21-27.
99. Costa-Pinheiro P, Patel HRH, Henrique R, Jerónimo C. Biomarkers and personalized risk stratification for patients with clinically localized prostate cancer. *Expert Rev Anticancer Ther*. 2014 Nov;14(11):1349–58.
100. Osses DF, Roobol MJ, Schoots IG. Prediction Medicine: Biomarkers, Risk Calculators and Magnetic Resonance Imaging as Risk Stratification Tools in Prostate Cancer Diagnosis. *Int J Mol Sci*. 2019 Apr 2;20(7).
101. Muppala S, Konduru SKP, Merchant N, Ramsoondar J, Rampersad CK, Rajitha B, et al. Adiponectin: Its role in obesity-associated colon and prostate cancers. *Crit Rev Oncol Hematol*. 2017 Aug;116:125–33.
102. Hu M-B, Xu H, Hu J-M, Zhu W-H, Yang T, Jiang H-W, et al. Genetic polymorphisms in leptin, adiponectin and their receptors affect risk and aggressiveness of prostate cancer: evidence from a meta-analysis and pooled-review. *Oncotarget*. 2016 Dec 6;7(49):81049–61.
103. Gann PH, Chatterton RT, Lee C. Peptide growth factors as biomarkers of prostate cancer risk. *Epidemiol Rev*. 2001;23(1):67–71.
104. Prentice RL, Huang Y. Nutritional Epidemiology Methods and Related Statistical Challenges and Opportunities. *Stat Theory Relat Fields*. 2018;2(1):2–10.
105. Dunn WB, Goodacre R, Neyses L, Mamas M. Integration of metabolomics in heart disease and diabetes research: current achievements and future outlook. *Bioanalysis*. 2011 Oct;3:2205–22.
106. Barton RH. A decade of advances in metabolomics. *Expert Metab Toxicol*. 2011 Feb;7:129–36.
107. Claudino WM, Quattrone A, Biganzoli L, Pestrin M, Bertini I, Di LA. Metabolomics: available results, current research projects in breast cancer, and future applications. *J Clin Oncol*. 2007 Jul 1;25:2840–6.
108. Fiehn O. Metabolomics--the link between genotypes and phenotypes. *Plant Mol Biol*. 2002 Jan;48(1–2):155–71.
109. Siroux V, Agier L, Slama R. The exposome concept: a challenge and a potential driver for environmental health research. *Eur Respir Rev Off J Eur Respir Soc*. 2016 Jun;25(140):124–9.
110. Al-Delaimy WK, Ferrari P, Slimani N, Pala V, Johansson I, Nilsson S, et al. Plasma carotenoids as biomarkers of intake of fruits and vegetables: individual-level correlations in the European



Prospective Investigation into Cancer and Nutrition (EPIC). *Eur J Clin Nutr.* 2005 Dec;59(12):1387–96.

111. Chajès V, Biessy C, Byrnes G, Deharveng G, Saadatian-Elahi M, Jenab M, et al. Ecological-level associations between highly processed food intakes and plasma phospholipid elaidic acid concentrations: results from a cross-sectional study within the European prospective investigation into cancer and nutrition (EPIC). *Nutr Cancer.* 2011 Nov;63(8):1235–50.
112. Rappaport SM. Biomarkers intersect with the exposome. *Biomark Biochem Indic Expo Response Susceptibility Chem.* 2012 Sep;17(6):483–9.
113. Wild CP, Scalbert A, Herceg Z. Measuring the exposome: a powerful basis for evaluating environmental exposures and cancer risk. *Environ Mol Mutagen.* 2013 Aug;54(7):480–99.
114. Mennen LI, Sapinho D, Ito H, Bertrais S, Galan P, Hercberg S, et al. Urinary flavonoids and phenolic acids as biomarkers of intake for polyphenol-rich foods. *Br J Nutr.* 2006 Jul;96(1):191–8.
115. Wishart DS, Feunang YD, Marcu A, Guo AC, Liang K, Vázquez-Fresno R, et al. HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res.* 2018 Jan 4;46(D1):D608–17.
116. Wishart D. Systems Biology Resources Arising from the Human Metabolome Project. In: Suhre K, editor. *Genetics Meets Metabolomics: from Experiment to Systems Biology* [Internet]. New York, NY: Springer New York; 2012 [cited 2019 Jun 2]. p. 157–75. Available from: [https://doi.org/10.1007/978-1-4614-1689-0\\_11](https://doi.org/10.1007/978-1-4614-1689-0_11)
117. Rothwell JA, Urpi-Sarda M, Boto-Ordoñez M, Knox C, Llorach R, Eisner R, et al. Phenol-Explorer 2.0: a major update of the Phenol-Explorer database integrating data on polyphenol metabolism and pharmacokinetics in humans and experimental animals. *Database J Biol Databases Curation.* 2012;2012:bas031.
118. Institut National de la Recherche Agronomique. PhytoHub [Internet]. [cited 2019 Jun 2]. Available from: <http://phytohub.eu/>
119. Guo AC, Jewison T, Wilson M, Liu Y, Knox C, Djoumbou Y, et al. ECMDDB: the E. coli Metabolome Database. *Nucleic Acids Res.* 2013 Jan;41(Database issue):D625-630.
120. Jewison T, Knox C, Neveu V, Djoumbou Y, Guo AC, Lee J, et al. YMDB: the Yeast Metabolome Database. *Nucleic Acids Res.* 2012 Jan;40(Database issue):D815-820.
121. Nicholson JK, Holmes E, Elliott P. The Metabolome-Wide Association Study: A New Look at Human Disease Risk Factors. *J Proteome Res.* 2008 Sep 5;7(9):3637–8.
122. Shurubor YI, Matson WR, Willett WC, Hankinson SE, Kristal BS. Biological variability dominates and influences analytical variance in HPLC-ECD studies of the human plasma metabolome. *BMC Clin Pathol.* 2007 Nov 12;7:9.
123. van Dorsten FA, Grün CH, van Velzen EJJ, Jacobs DM, Draijer R, van Duynhoven JPM. The metabolic fate of red wine and grape juice polyphenols in humans assessed by metabolomics. *Mol Nutr Food Res.* 2010 Jul;54(7):897–908.
124. Jacobs DM, Fuhrmann JC, van Dorsten FA, Rein D, Peters S, van Velzen EJJ, et al. Impact of short-term intake of red wine and grape polyphenol extract on the human metabolome. *J Agric Food Chem.* 2012 Mar 28;60(12):3078–85.

125. Ulaszewska MM, Weinert CH, Trimigno A, Portmann R, Andres Lacueva C, Badertscher R, et al. Nutrismetabolomics: An Integrative Action for Metabolomic Analyses in Human Nutritional Studies. *Mol Nutr Food Res*. 2019 Jan;63(1):1800384.
126. Guasch-Ferré M, Bhupathiraju SN, Hu FB. Use of Metabolomics in Improving Assessment of Dietary Intake. *Clin Chem*. 2018 Jan;64(1):82–98.
127. Sansone S-A, Rocca-Serra P, Field D, Maguire E, Taylor C, Hofmann O, et al. Toward interoperable bioscience data. *Nat Genet*. 2012 Jan 27;44(2):121–6.
128. González-Peña D, Brennan L. Recent Advances in the Application of Metabolomics for Nutrition and Health. *Annu Rev Food Sci Technol*. 2019 Mar 25;10:479–519.
129. Gibbons H, Brennan L. Metabolomics as a tool in the identification of dietary biomarkers. *Proc Nutr Soc*. 2017;76(1):42–53.
130. Rasmussen LG, Winning H, Savorani F, Ritz C, Engelsen SB, Astrup A, et al. Assessment of dietary exposure related to dietary GI and fibre intake in a nutritional metabolomic study of human urine. *Genes Nutr*. 2012 Apr;7(2):281–93.
131. Johansson-Persson A, Barri T, Ulmius M, Onning G, Dragsted LO. LC-QTOF/MS metabolomic profiles in human plasma after a 5-week high dietary fiber intake. *Anal Bioanal Chem*. 2013 May;405(14):4799–809.
132. Menni C, Zhai G, Macgregor A, Prehn C, Römisch-Margl W, Suhre K, et al. Targeted metabolomics profiles are strongly correlated with nutritional patterns in women. *Metabolomics Off J Metabolomic Soc*. 2013 Apr;9(2):506–14.
133. Stanstrup J, Schou SS, Holmer-Jensen J, Hermansen K, Dragsted LO. Whey protein delays gastric emptying and suppresses plasma fatty acids and their metabolites compared to casein, gluten, and fish protein. *J Proteome Res*. 2014 May 2;13(5):2396–408.
134. Andersen M-BS, Kristensen M, Manach C, Pujos-Guillot E, Poulsen SK, Larsen TM, et al. Discovery and validation of urinary exposure markers for different plant foods by untargeted metabolomics. *Anal Bioanal Chem*. 2014 Mar;406(7):1829–44.
135. Pujos-Guillot E, Hubert J, Martin JF, Lyan B, Quintana M, Claude S, et al. Mass Spectrometry-based Metabolomics for the Discovery of Biomarkers of Fruit and Vegetable Intake: Citrus Fruit as a Case Study. *JProteomeRes*. 2013 Mar 5;
136. Lloyd AJ, Beckmann M, Favé G, Mathers JC, Draper J. Proline betaine and its biotransformation products in fasting urine samples are potential biomarkers of habitual citrus fruit consumption. *Br J Nutr*. 2011 Sep;106(6):812–24.
137. Heinzmann SS, Brown IJ, Chan Q, Bictash M, Dumas ME, Kochhar S, et al. Metabolic profiling strategy for discovery of nutritional biomarkers: proline betaine as a marker of citrus consumption. *AmJClinNutr*. 2010 Aug;92:436–43.
138. Andersen M-BS, Reinbach HC, Rinnan Å, Barri T, Mithril C, Dragsted LO. Discovery of exposure markers in urine for Brassica-containing meals served with different protein sources by UPLC-qTOF-MS untargeted metabolomics. *Metabolomics*. 2013 Oct;9(5):984–97.

139. Edmands WMB, Beckonert OP, Stella C, Campbell A, Lake BG, Lindon JC, et al. Identification of human urinary biomarkers of cruciferous vegetable consumption by metabonomic profiling. *J Proteome Res.* 2011 Oct 7;10(10):4513–21.
140. Llorach R, Urpi-Sarda M, Jauregui O, Monagas M, Andres-Lacueva C. An LC-MS-based metabolomics approach for exploring urinary metabolome modifications after cocoa consumption. *J Proteome Res.* 2009 Nov;8(11):5060–8.
141. Llorach R, Urpi-Sarda M, Tulipani S, Garcia-Aloy M, Monagas M, Andres-Lacueva C. Metabolomic fingerprint in patients at high risk of cardiovascular disease by cocoa intervention. *Mol Nutr Food Res.* 2013 Jun;57(6):962–73.
142. Tulipani S, Llorach R, Jáuregui O, López-Uriarte P, Garcia-Aloy M, Bullo M, et al. Metabolomics unveils urinary changes in subjects with metabolic syndrome following 12-week nut consumption. *J Proteome Res.* 2011 Nov 4;10(11):5047–58.
143. Lang R, Wahl A, Stark T, Hofmann T. Urinary N-methylpyridinium and trigonelline as candidate dietary biomarkers of coffee consumption. *Mol Nutr Food Res.* 2011 Nov;55(11):1613–23.
144. Lloyd AJ, Beckmann M, Haldar S, Seal C, Brandt K, Draper J. Data-driven strategy for the discovery of potential urinary biomarkers of habitual dietary exposure. *AmJClinNutr.* 2013 Feb;97:377–89.
145. Vázquez-Fresno R, Llorach R, Perera A, Mandal R, Feliz M, Tinahones FJ, et al. Clinical phenotype clustering in cardiovascular risk patients for the identification of responsive metabolotypes after red wine polyphenol intake. *J Nutr Biochem.* 2016 Feb;28:114–20.
146. van Velzen EJJ, Westerhuis JA, van Duynhoven JPM, van Dorsten FA, Grün CH, Jacobs DM, et al. Phenotyping tea consumers by nutrikinetic analysis of polyphenolic end-metabolites. *J Proteome Res.* 2009 Jul;8(7):3317–30.
147. Law WS, Huang PY, Ong ES, Ong CN, Li SFY, Pasikanti KK, et al. Metabonomics investigation of human urine after ingestion of green tea with gas chromatography/mass spectrometry, liquid chromatography/mass spectrometry and (1)H NMR spectroscopy. *Rapid Commun Mass Spectrom RCM.* 2008 Aug;22(16):2436–46.
148. Van Dorsten FA, Daykin CA, Mulder TPJ, Van Duynhoven JPM. Metabonomics approach to determine metabolic differences between green tea and black tea consumption. *J Agric Food Chem.* 2006 Sep 6;54(18):6929–38.
149. Lloyd AJ, Favé G, Beckmann M, Lin W, Tailliant K, Xie L, et al. Use of mass spectrometry fingerprinting to identify urinary metabolites after consumption of specific foods. *Am J Clin Nutr.* 2011 Oct;94(4):981–91.
150. Stella C, Beckwith-Hall B, Cloarec O, Holmes E, Lindon JC, Powell J, et al. Susceptibility of human metabolic phenotypes to dietary modulation. *J Proteome Res.* 2006 Oct;5(10):2780–8.
151. Bertram HC, Hoppe C, Petersen BO, Duus JØ, Mølgaard C, Michaelsen KF. An NMR-based metabonomic investigation on effects of milk and meat protein diets given to 8-year-old boys. *Br J Nutr.* 2007 Apr;97(4):758–63.
152. Xu J, Yang S, Cai S, Dong J, Li X, Chen Z. Identification of biochemical changes in lactovegetarian urine using 1H NMR spectroscopy and pattern recognition. *Anal Bioanal Chem.* 2010 Feb;396(4):1451–63.

153. Svensson BG, Akesson B, Nilsson A, Paulsson K. Urinary excretion of methylamines in men with varying intake of fish from the Baltic Sea. *J Toxicol Environ Health*. 1994 Apr;41(4):411–20.
154. Koeth RA, Wang Z, Levison BS, Buffa JA, Org E, Sheehy BT, et al. Intestinal microbiota metabolism of l-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nat Med*. 2013 Apr 7;19(5):576–85.
155. Brennan L. Metabolomics: a tool to aid dietary assessment in nutrition. *Curr Opin Food Sci*. 2017 Aug;16:96–9.
156. Andersen M-BS, Rinnan Å, Manach C, Poulsen SK, Pujos-Guillot E, Larsen TM, et al. Untargeted metabolomics as a screening tool for estimating compliance to a dietary pattern. *J Proteome Res*. 2014 Mar 7;13(3):1405–18.
157. Acar E, Gürdeniz G, Khakimov B, Savorani F, Korndal SK, Larsen TM, et al. Biomarkers of Individual Foods, and Separation of Diets Using Untargeted LC-MS-based Plasma Metabolomics in a Randomized Controlled Trial. *Mol Nutr Food Res*. 2019 Jan;63(1):1800215.
158. Lankinen M, Kolehmainen M, Jääskeläinen T, Paananen J, Joukamo L, Kangas AJ, et al. Effects of Whole Grain, Fish and Bilberries on Serum Metabolic Profile and Lipid Transfer Protein Activities: A Randomized Trial (Sysdimet). Berthold HK, editor. *PLoS ONE*. 2014 Feb 28;9(2):e90352.
159. Hanhineva K, Lankinen MA, Pedret A, Schwab U, Kolehmainen M, Paananen J, et al. Nontargeted Metabolite Profiling Discriminates Diet-Specific Biomarkers for Consumption of Whole Grains, Fatty Fish, and Bilberries in a Randomized Controlled Trial. *J Nutr*. 2015 Jan 1;145(1):7–17.
160. Garcia-Perez I, Posma JM, Gibson R, Chambers ES, Hansen TH, Vestergaard H, et al. Objective assessment of dietary patterns by use of metabolic phenotyping: a randomised, controlled, crossover trial. *Lancet Diabetes Endocrinol*. 2017;5(3):184–95.
161. Bondia-Pons I, Martinez JA, de la Iglesia R, Lopez-Legarrea P, Poutanen K, Hanhineva K, et al. Effects of short- and long-term Mediterranean-based dietary treatment on plasma LC-QTOF/MS metabolic profiling of subjects with metabolic syndrome features: The Metabolic Syndrome Reduction in Navarra (RESMENA) randomized controlled trial. *Mol Nutr Food Res*. 2015 Apr;59(4):711–28.
162. May DH, Navarro SL, Ruczinski I, Hogan J, Ogata Y, Schwarz Y, et al. Metabolomic profiling of urine: response to a randomised, controlled feeding study of select fruits and vegetables, and application to an observational study. *BrJ Nutr*. 2013 Nov;110:1760–70.
163. Esko T, Hirschhorn JN, Feldman HA, Hsu Y-HH, Deik AA, Clish CB, et al. Metabolomic profiles as reliable biomarkers of dietary composition. *Am J Clin Nutr*. 2017 Mar;105(3):547–54.
164. Vázquez-Fresno R, Llorach R, Urpi-Sarda M, Lupianez-Barbero A, Estruch R, Corella D, et al. Metabolomic pattern analysis after mediterranean diet intervention in a nondiabetic population: a 1- and 3-year follow-up in the PREDIMED study. *J Proteome Res*. 2015 Jan 2;14(1):531–40.
165. Xu J, Yang S, Cai S, Dong J, Li X, Chen Z. Identification of biochemical changes in lactovegetarian urine using <sup>1</sup>H NMR spectroscopy and pattern recognition. *Anal Bioanal Chem*. 2010 Feb;396(4):1451–63.
166. O’Sullivan A, Gibney MJ, Brennan L. Dietary intake patterns are reflected in metabolomic profiles: potential role in dietary assessment studies. *Am J Clin Nutr*. 2011 Feb 1;93(2):314–21.

167. Bouchard-Mercier A, Rudkowska I, Lemieux S, Couture P, Vohl M-C. The metabolic signature associated with the Western dietary pattern: a cross-sectional study. *Nutr J.* 2013 Dec 11;12:158.
168. Gibbons H, Carr E, McNulty BA, Nugent AP, Walton J, Flynn A, et al. Metabolomic-based identification of clusters that reflect dietary patterns. *Mol Nutr Food Res.* 2017;61(10).
169. Bhupathiraju SN, Guasch-Ferré M, Gadgil MD, Newgard CB, Bain JR, Muehlbauer MJ, et al. Dietary Patterns among Asian Indians Living in the United States Have Distinct Metabolomic Profiles That Are Associated with Cardiometabolic Risk. *J Nutr* [Internet]. 2018 Jun 10 [cited 2019 Jan 22]; Available from: <https://academic.oup.com/jn/advance-article/doi/10.1093/jn/nxy074/5033850>
170. Schmidt JA, Rinaldi S, Ferrari P, Carayol M, Achaintre D, Scalbert A, et al. Metabolic profiles of male meat eaters, fish eaters, vegetarians, and vegans from the EPIC-Oxford cohort. *Am J Clin Nutr.* 2015 Dec;102(6):1518–26.
171. Playdon MC, Moore SC, Derkach A, Reedy J, Subar AF, Sampson JN, et al. Identifying biomarkers of dietary patterns by using metabolomics. *Am J Clin Nutr.* 2017;105(2):450–65.
172. Almanza-Aguilera E, Urpi-Sarda M, Llorach R, Vázquez-Fresno R, Garcia-Aloy M, Carmona F, et al. Microbial metabolites are associated with a high adherence to a Mediterranean dietary pattern using a 1H-NMR-based untargeted metabolomics approach. *J Nutr Biochem.* 2017 Oct;48:36–43.
173. Altmaier E, Kastenmüller G, Römisch-Margl W, Thorand B, Weinberger KM, Illig T, et al. Questionnaire-based self-reported nutrition habits associate with serum metabolism as revealed by quantitative targeted metabolomics. *Eur J Epidemiol.* 2011 Feb;26(2):145–56.
174. Floegel A, von Ruesten A, Drogan D, Schulze MB, Prehn C, Adamski J, et al. Variation of serum metabolites related to habitual diet: a targeted metabolomic approach in EPIC-Potsdam. *Eur J Clin Nutr.* 2013 Oct;67(10):1100–8.
175. Peré-Trepat E, Ross AB, Martin F-P, Rezzi S, Kochhar S, Hasselbalch AL, et al. Chemometric strategies to assess metabolomic imprinting of food habits in epidemiological studies. *Chemom Intell Lab Syst.* 2010 Nov;104(1):95–100.
176. Merz B, Frommherz L, Rist MJ, Kulling SE, Bub A, Watzl B. Dietary Pattern and Plasma BCAA-Variations in Healthy Men and Women-Results from the KarMeN Study. *Nutrients.* 2018 May 15;10(5).
177. Wei R, Ross AB, Su M, Wang J, Guiraud S-P, Draper CF, et al. Metabotypes Related to Meat and Vegetable Intake Reflect Microbial, Lipid and Amino Acid Metabolism in Healthy People. *Mol Nutr Food Res.* 2018 Nov;62(21):1800583.
178. Tzoulaki I, Ebbels TMD, Valdes A, Elliott P, Ioannidis JPA. Design and analysis of metabolomics studies in epidemiologic research: a primer on -omic technologies. *Am J Epidemiol.* 2014 Jul 15;180(2):129–39.
179. Franceschi S, Wild CP. Meeting the global demands of epidemiologic transition - the indispensable role of cancer prevention. *Mol Oncol.* 2013 Feb;7(1):1–13.
180. Liesenfeld DB, Habermann N, Owen RW, Scalbert A, Ulrich CM. Review of mass spectrometry-based metabolomics in cancer research. *Cancer Epidemiol Biomarkers Prev.* 2013 Dec;22:2182–201.

181. Howell A. Can metabolomics in addition to genomics add to prognostic and predictive information in breast cancer? *BMC Med*. 2010;8:73-.
182. Budczies J, Brockmoller SF, Muller BM, Barupal DK, Richter-Ehrenstein C, Kleine-Tebbe A, et al. Comparative metabolomics of estrogen receptor positive and estrogen receptor negative breast cancer: alterations in glutamine and beta-alanine metabolism. *J Proteomics*. 2013 Oct 11;94C:279–88.
183. Wood SL, Westbrook JA, Brown JE. Omic-profiling in breast cancer metastasis to bone: implications for mechanisms, biomarkers and treatment. *Cancer Treat Rev*. 2014 Feb;40:139–52.
184. Jobard E, Pontoizeau C, Blaise BJ, Bachelot T, Elena-Herrmann B, Tredan O. A serum nuclear magnetic resonance-based metabolomic signature of advanced metastatic human breast cancer. *Cancer Lett*. 2014 Feb 1;343:33–41.
185. Brockmoller SF, Bucher E, Muller BM, Budczies J, Hilvo M, Griffin JL, et al. Integration of metabolomics and expression of glycerol-3-phosphate acyltransferase (GPAM) in breast cancer-link to patient survival, hormone receptor status, and metabolic profiling. *J Proteome Res*. 2012 Feb 3;11:850–60.
186. Maxeiner A, Adkins CB, Zhang Y, Taupitz M, Halpern EF, McDougal WS, et al. Retrospective analysis of prostate cancer recurrence potential with tissue metabolomic profiles. *Prostate*. 2010 May 15;70:710–7.
187. Sreekumar A, Poisson LM, Rajendiran TM, Khan AP, Cao Q, Yu J, et al. Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature*. 2009 Feb 12;457:910–4.
188. Denkert C, Bucher E, Hilvo M, Salek R, Oresic M, Griffin J, et al. Metabolomics of human breast cancer: new approaches for tumor typing and biomarker discovery. *Genome Med*. 2012;4:37-.
189. Beaumont JJ, Steenland K, Minton A, Meyer S. A computer program for incidence density sampling of controls in case-control studies nested within occupational cohort studies. *Am J Epidemiol*. 1989 Jan;129(1):212–9.
190. Shu X, Zheng W, Yu D, Li H-L, Lan Q, Yang G, et al. Prospective metabolomics study identifies potential novel blood metabolites associated with pancreatic cancer risk. *Int J Cancer*. 2018 01;143(9):2161–7.
191. Nakagawa T, Kobayashi T, Nishiumi S, Hidaka A, Yamaji T, Sawada N, et al. Metabolome analysis for pancreatic cancer risk in nested case-control study: Japan Public Health Center-based prospective Study. *Cancer Sci*. 2018 May;109(5):1672–81.
192. Assi N, Thomas DC, Leitzman M, Stepien M, Chajes V, Philip T, et al. Are metabolic signatures mediating the relationship between lifestyle factors and hepatocellular carcinoma risk? Results from a nested case-control study in EPIC. *Cancer Epidemiol Biomarkers Prev*. 2018 Mar 21;27(5):531–40.
193. Zhang W, Zhou L, Yin P, Wang J, Lu X, Wang X, et al. A weighted relative difference accumulation algorithm for dynamic metabolomics data: long-term elevated bile acids are risk factors for hepatocellular carcinoma. *Sci Rep*. 2015 Mar 11;5:8984.

194. Assi N, Gunter MJ, Thomas DC, Leitzmann M, Stepien M, Chajès V, et al. Metabolic signature of healthy lifestyle and its relation with risk of hepatocellular carcinoma in a large European cohort. *Am J Clin Nutr.* 2018 Jul 1;108(1):117–26.
195. Fages A, Duarte-Salles T, Stepien M, Ferrari P, Fedirko V, Pontoizeau C, et al. Metabolomic profiles of hepatocellular carcinoma in a European prospective cohort. *BMC Med.* 2015 Sep;13:242.
196. Assi N. A statistical framework to model the meeting-in-the-middle principle using metabolomic data: application to hepatocellular carcinoma in the EPIC study. *Mutagenesis.* 2015 Nov;30(6):743–53.
197. Stepien M, Duarte-Salles T, Fedirko V, Floegel A, Barupal DK, Rinaldi S, et al. Alteration of amino acid and biogenic amine metabolism in hepatobiliary cancers: Findings from a prospective cohort study. *Int J Cancer.* 2016 Jan 15;138:348–60.
198. Mayers JR, Wu C, Clish CB, Kraft P, Torrence ME, Fiske BP, et al. Elevation of circulating branched-chain amino acids is an early event in human pancreatic adenocarcinoma development. *Nat Med.* 2014 Oct;20(10):1193–8.
199. Yuan JM, Gao YT, Murphy SE, Carmella SG, Wang R, Zhong Y, et al. Urinary levels of cigarette smoke constituent metabolites are prospectively associated with lung cancer development in smokers. *Cancer Res.* 2011 Nov 1;71:6749–57.
200. Wikoff WR, Hanash S, DeFelice B, Miyamoto S, Barnett M, Zhao Y, et al. Diacetylspermine Is a Novel Prediagnostic Serum Biomarker for Non-Small-Cell Lung Cancer and Has Additive Performance With Pro-Surfactant Protein B. *J ClinOncol.* 2015 Nov 20;33:3880–6.
201. Huang J, Weinstein SJ, Kitahara CM, Karoly ED, Sampson JN, Albanes D. A prospective study of serum metabolites and glioma risk. *Oncotarget.* 2017 Sep 19;8(41):70366–77.
202. Bjorkblom B, Wibom C, Jonsson P, Moren L, Andersson U, Johannesen TB, et al. Metabolomic screening of pre-diagnostic serum samples identifies association between alpha- and gamma-tocopherols and glioblastoma risk. *Oncotarget.* 2016 Jun 14;7:37043–53.
203. Bro R, Kamstrup-Nielsen MH, Engelsen SB, Savorani F, Rasmussen MA, Hansen L, et al. Forecasting individual breast cancer risk using plasma metabolomics and biocontours. *Metabolomics.* 2015 Oct;11(5):1376–80.
204. Moore SC, Playdon MC, Sampson JN, Hoover RN, Trabert B, Matthews CE, et al. A Metabolomics Analysis of Body Mass Index and Postmenopausal Breast Cancer Risk. *JNCI J Natl Cancer Inst.* 2018 Jan 9;110(6):dix244.
205. Playdon MC, Ziegler RG, Sampson JN, Stolzenberg-Solomon R, Thompson HJ, Irwin ML, et al. Nutritional metabolomics and breast cancer risk in a prospective study. *AmJClinNutr.* 2017 Aug;106:637–49.
206. Kuhn T, Floegel A, Sookthai D, Johnson T, Rolle-Kampczyk U, Otto W, et al. Higher plasma levels of lysophosphatidylcholine 18:0 are related to a lower risk of common cancers in a prospective metabolomics study. *BMC Med.* 2016 Jan 28;14:13-.
207. Chadeau-Hyam M, Athersuch TJ, Keun HC, De Iorio M, Ebbels TMD, Jenab M, et al. Meeting-in-the-middle using metabolic profiling - a strategy for the identification of intermediate biomarkers in cohort studies. *Biomark Biochem Indic Expo Response Susceptibility Chem.* 2011 Feb;16(1):83–8.

208. Shu X, Xiang Y-B, Rothman N, Yu D, Li H-L, Yang G, et al. Prospective study of blood metabolites associated with colorectal cancer risk. *Int J Cancer*. 2018 01;143(3):527–34.
209. Cross AJ, Moore SC, Boca S, Huang WY, Xiong X, Stolzenberg-Solomon R, et al. A prospective study of serum metabolites and colorectal cancer risk. *Cancer*. 2014 Oct 1;120:3049–57.
210. Cross AJ, Boca S, Freedman ND, Caporaso NE, Huang W-Y, Sinha R, et al. Metabolites of tobacco smoking and colorectal cancer risk. *Carcinogenesis*. 2014 Jul;35(7):1516–22.
211. Guertin KA, Loftfield E, Boca SM, Sampson JN, Moore SC, Xiao Q, et al. Serum biomarkers of habitual coffee consumption may provide insight into the mechanism underlying the association between coffee consumption and colorectal cancer. *AmJ ClinNutr*. 2015 May;101:1000–11.
212. Troche JR, Mayne ST, Freedman ND, Shebl FM, Guertin KA, Cross AJ, et al. Alcohol Consumption-Related Metabolites in Relation to Colorectal Cancer and Adenoma: Two Case-Control Studies Using Serum Biomarkers. *PloS One*. 2016;11(3):e0150962.
213. Mondul AM, Moore SC, Weinstein SJ, Mannisto S, Sampson JN, Albanes D. 1-stearoylglycerol is associated with risk of prostate cancer: results from serum metabolomic profiling. *Metabolomics*. 2014 Oct 1;10:1036–41.
214. Mondul AM, Moore SC, Weinstein SJ, Karoly ED, Sampson JN, Albanes D. Metabolomic analysis of prostate cancer risk in a prospective cohort: The alpha-tocolpherol, beta-carotene cancer prevention (ATBC) study. *IntJ Cancer*. 2015 Nov 1;137:2124–32.
215. Huang J, Mondul AM, Weinstein SJ, Koutros S, Derkach A, Karoly E, et al. Serum metabolomic profiling of prostate cancer risk in the prostate, lung, colorectal, and ovarian cancer screening trial. *BrJ Cancer*. 2016 Oct 25;115:1087–95.
216. Schmidt JA, Fensom GK, Rinaldi S, Scalbert A, Appleby PN, Achaintre D, et al. Pre-diagnostic metabolite concentrations and prostate cancer risk in 1077 cases and 1077 matched controls in the European Prospective Investigation into Cancer and Nutrition. *BMC Med*. 2017 05;15(1):122.
217. McGee EE, Kiblawi R, Playdon MC, Eliassen AH. Nutritional Metabolomics in Cancer Epidemiology: Current Trends, Challenges, and Future Directions. *Curr Nutr Rep*. 2019 May 25;
218. Bonferroni CE. Teoria statistica delle classi e calcolo delle probabilità. *Pubblicazioni R Ist Super Sci Econ E Commer Firenze*. 1936;8.
219. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. *Stat Med*. 1990 Jul;9(7):811–8.
220. Benjamini Y, Heller R, Yekutieli D. Selective inference in complex research. *Philos Trans R Soc Math Phys Eng Sci*. 2009 Nov 13;367(1906):4255–71.
221. Chadeau-Hyam M, Campanella G, Jombart T, Bottolo L, Portengen L, Vineis P, et al. Deciphering the complex: methodological overview of statistical models to derive OMICS-based biomarkers. *Environ Mol Mutagen*. 2013 Aug;54(7):542–57.
222. Worley B, Halouska S, Powers R. Utilities for quantifying separation in PCA/PLS-DA scores plots. *Anal Biochem*. 2013 Feb 15;433(2):102–4.



223. Pearson K. LIII. *On lines and planes of closest fit to systems of points in space*. Lond Edinb Dublin Philos Mag J Sci. 1901 Nov;2(11):559–72.
224. Hotelling H. Analysis of a complex of statistical variables into principal components. J Educ Psychol. 1933;24(7):498–520.
225. Hanhineva K, Barri T, Kolehmainen M, Pekkinen J, Pihlajamäki J, Vesterbacka A, et al. Comparative nontargeted profiling of metabolic changes in tissues and biofluids in high-fat diet-fed Ossabaw pig. J Proteome Res. 2013 Sep 6;12(9):3980–92.
226. Sugimoto M, Kawakami M, Robert M, Soga T, Tomita M. Bioinformatics Tools for Mass Spectroscopy-Based Metabolomic Data Processing and Analysis. Curr Bioinforma. 2012 Mar;7(1):96–108.
227. Ståhle L, Wold S. Partial least squares analysis with cross-validation for the two-class problem: A Monte Carlo study. J Chemom. 1987 Jul;1(3):185–96.
228. Barker M, Rayens W. Partial least squares for discrimination. J Chemom. 2003 Mar 24;17(3):166–73.
229. Tibshirani R. Regression Shrinkage and Selection via the Lasso. JRSSB. 1996;58:267–88.
230. Hoerl A, Kennard R. Ridge regression: biased estimation for nonorthogonal problems. Technometrics. 1970;55–67.
231. Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Ser B Stat Methodol. 2005 Apr;67(2):301–20.
232. Hercberg S, Galan P, Preziosi P, Bertrais S, Mennen L, Malvy D, et al. The SU.VI.MAX study: a randomized, placebo-controlled trial of the health effects of antioxidant vitamins and minerals. ArchInternMed. 2004 Nov 22;164:2335–42.
233. Hercberg S, Preziosi P, Briancon S, Galan P, Triol I, Malvy D, et al. A primary prevention trial using nutritional doses of antioxidant vitamins and minerals in cardiovascular diseases and cancers in a general population: the SU.VI.MAX study--design, methods, and participant characteristics. SUpplementation en Vitamines et Mineraux AntioXydants. Control ClinTrials. 1998 Aug;19:336–51.
234. Kriska AM, Knowler WC, LaPorte RE, Drash AL, Wing RR, Blair SN, et al. Development of questionnaire to examine relationship of physical activity and diabetes in Pima Indians. Diabetes Care. 1990 Apr;13:401–11.
235. Schulz LO, Harper IT, Smith CJ, Kriska AM, Ravussin E. Energy intake and physical activity in Pima Indians: comparison with energy expenditure measured by doubly-labeled water. Obes Res. 1994 Nov;2(6):541–8.
236. Vuillemin A, Oppert JM, Guillemin F, Essermeant L, Fontvieille AM, Galan P, et al. Self-administered questionnaire compared with interview to assess past-year physical activity. MedSciSports Exerc. 2000 Jun;32:1119–24.
237. Le Moullec N, Deheeger M, Preziosi P, Montero P, Valeix P, Rolland-Cachera MF, et al. Validation du manuel photos utilisé pour l'enquête alimentaire de l'étude SU.VI.MAX (Validation of the food portion size booklet used in the SU.VI.MAX study). CahNutrDiet. 1996;31:158–64.

238. Hercberg S. Table de Composition SU.VI.MAX des Aliments. Paris: Les éditions INSERM/Economica; 2005.
239. Kesse-Guyot E, Touvier M, Henegar A, Czernichow S, Galan P, Hercberg S, et al. Higher adherence to French dietary guidelines and chronic diseases in the prospective SU.VI.MAX cohort. *Eur J Clin Nutr*. 2011 Aug;65:887–94.
240. Lassale C, Fezeu L, Andreeva VA, Hercberg S, Kengne A-P, Czernichow S, et al. Association between dietary scores and 13-year weight change and obesity risk in a French prospective cohort. *Int J Obes* 2005. 2012 Nov;36(11):1455–62.
241. Lasfargues G, Vol S, Le Clesiau H, Bedouet M, Hagel L, Constans T, et al. [Validity of a short self-administered dietary questionnaire compared with a dietetic interview]. *Presse Med*. 1990 May 19;19:953–7.
242. WHO. ICD-10, International Classification of Diseases and Related Health Problems. 10th revision. Geneva: World Health Organization; 1993.
243. Estaquio C, Kesse-Guyot E, Deschamps V, Bertrais S, Dauchet L, Galan P, et al. Adherence to the French Programme National Nutrition Sante Guideline Score is associated with better nutrient intake and nutritional status. *JAmDiet Assoc*. 2009 Jun;109:1031–41.
244. Lassale C, Galan P, Julia C, Fezeu L, Hercberg S, Kesse-Guyot E. Association between adherence to nutritional guidelines, the metabolic syndrome and adiposity markers in a French adult general population. *PLoS One*. 2013;8(10):e76349.
245. Lavalette C, Adjibade M, Srouf B, Sellem L, Fiolet T, Hercberg S, et al. Cancer-Specific and General Nutritional Scores and Cancer Risk: Results from the Prospective NutriNet-Santé Cohort. *Cancer Res*. 2018 Aug 1;78(15):4427–35.
246. Lelong H, Blacher J, Menai M, Galan P, Fezeu L, Hercberg S, et al. Association Between Blood Pressure and Adherence to French Dietary Guidelines. *Am J Hypertens*. 2016 Aug;29(8):948–58.
247. Lassale C, Galan P, Castetbon K, Péneau S, Méjean C, Hercberg S, et al. Differential association between adherence to nutritional recommendations and body weight status across educational levels: a cross-sectional study. *Prev Med*. 2013 Nov;57(5):488–93.
248. Lindon JC, Nicholson JK. Spectroscopic and Statistical Techniques for Information Recovery in Metabonomics and Metabolomics. *Annu Rev Anal Chem*. 2008 Jul;1(1):45–69.
249. Giraudeau P, Silvestre V, Akoka S. Optimizing water suppression for quantitative NMR-based metabolomics: a tutorial review. *Metabolomics*. 2015;2.
250. McKay RT. How the 1D-NOESY suppresses solvent signal in metabonomics NMR spectroscopy: An examination of the pulse sequence components and evolution. *Concepts Magn Reson Part A*. 2011 Sep;38A(5):197–220.
251. Carr HY, Purcell EM. Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments. *Phys Rev*. 1954;94:630–8.
252. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *JBiomolNMR*. 1995 Nov;6:277–93.

253. Sousa SA, Magalhaes A, Ferreira MM. Optimized bucketing for NMR spectra: Three case studies. *Chemom Intell Lab Syst.* 2013 Mar 15;122:93–102.
254. Beckonert O, Keun HC, Ebbels TMD, Bundy J, Holmes E, Lindon JC, et al. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat Protoc.* 2007;2(11):2692–703.
255. Giacomoni F, Le Corguille G, Monsoor M, Landi M, Pericard P, Petera M, et al. Workflow4Metabolomics: a collaborative research infrastructure for computational metabolomics. *Bioinformatics.* 2015 May 1;31(9):1493–5.
256. Van der Kloet FM, Bobeldijk I, Verheij ER, Jellema RH. Analytical error reduction using single point calibration for accurate and precise metabolomic phenotyping. *JProteomeRes.* 2009 Nov;8:5132–41.
257. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, et al. Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics.* 2007 Sep 19;3(3):211–21.
258. Kim S, Haines PS, Siega-Riz AM, Popkin BM. The Diet Quality Index-International (DQI-I) provides an effective tool for cross-national comparison of diet quality as illustrated by China and the United States. *J Nutr.* 2003 Nov;133(11):3476–84.
259. Martin A. Apports nutritionnels conseillés pour la population française. Tec & Doc Lavoisier/AFSSA: Paris edn. 2001.
260. Cattell RB. The scree test for the number of factors. *Multivar BehavRes.* 1966 Mar;1:245–76.
261. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *J R Stat Soc B.* 1995;57(1):289–300.
262. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw.* 2010;33(1):1–22.
263. Hastie T, Tibshirani R, Friedman JH. The elements of statistical learning: data mining, inference, and prediction. 2nd ed. New York, NY: Springer; 2009. 745 p. (Springer series in statistics).
264. Bunea F, She Y, Ombao H, Gongvatana A, Devlin K, Cohen R. Penalized least squares regression methods and applications to neuroimaging. *NeuroImage.* 2011 Apr;55(4):1519–27.
265. Pereira F, Martin JF, Joly C, Sébédio JL, Pujos-Guillot E. Development and validation of a UPLC/MS method for a nutritional metabolomic study of human plasma. *Metabolomics.* 2010 Feb 19;6(2):207–18.
266. Dumas M-E, Maibaum EC, Teague C, Ueshima H, Zhou B, Lindon JC, et al. Assessment of analytical reproducibility of <sup>1</sup>H NMR spectroscopy based metabonomics for large-scale epidemiological research: the INTERMAP Study. *Anal Chem.* 2006 Apr 1;78(7):2199–208.
267. Llanos AA, Makambi KH, Tucker CA, Wallington SF, Shields PG, Adams-Campbell LL. Cholesterol, lipoproteins, and breast cancer risk in African American women. *Ethn Dis.* 2012;22(3):281–7.

268. His M, Zelek L, Deschasaux M, Pouchieu C, Kesse-Guyot E, Hercberg S, et al. Prospective associations between serum biomarkers of lipid metabolism and overall, breast and prostate cancer risk. *Eur J Epidemiol.* 2014 Feb;29(2):119–32.
269. Kucharska-Newton AM, Rosamond WD, Mink PJ, Alberg AJ, Shahar E, Folsom AR. HDL-cholesterol and incidence of breast cancer in the ARIC cohort study. *Ann Epidemiol.* 2008 Sep;18(9):671–7.
270. Boyle P, Koehler A, Pizot C, Boniol M, Robertson C, Mullie P, et al. Blood glucose concentrations and breast cancer risk in women without diabetes: a meta-analysis. *Eur J Nutr.* 2013 Aug;52:1533–40.
271. Kühn T, Sookthai D, Graf ME, Schübel R, Freisling H, Johnson T, et al. Albumin, bilirubin, uric acid and cancer risk: results from a prospective population-based study. *Br J Cancer.* 2017 Nov 7;117(10):1572–9.
272. Katzke VA, Sookthai D, Johnson T, Kühn T, Kaaks R. Blood lipids and lipoproteins in relation to incidence and mortality risks for CVD and cancer in the prospective EPIC-Heidelberg cohort. *BMC Med.* 2017 Dec 19;15(1):218.
273. Radišauskas R, Kuzmickienė I, Milinavičienė E, Everatt R. Hypertension, serum lipids and cancer risk: A review of epidemiological evidence. *Med Kaunas Lith.* 2016;52(2):89–98.
274. Mondul AM, Weinstein SJ, Virtamo J, Albanes D. Serum total and HDL cholesterol and risk of prostate cancer. *Cancer Causes Control CCC.* 2011 Nov;22(11):1545–52.
275. Van Hemelrijck M, Walldius G, Jungner I, Hammar N, Garmo H, Binda E, et al. Low levels of apolipoprotein A-I and HDL are associated with risk of prostate cancer in the Swedish AMORIS study. *Cancer Causes Control CCC.* 2011 Jul;22(7):1011–9.
276. YuPeng L, YuXue Z, PengFei L, Cheng C, YaShuang Z, DaPeng L, et al. Cholesterol Levels in Blood and the Risk of Prostate Cancer: A Meta-analysis of 14 Prospective Studies. *Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol.* 2015 Jul;24(7):1086–93.
277. Magnusson M, Lewis GD, Ericson U, Orho-Melander M, Hedblad B, Engstrom G, et al. A diabetes-predictive amino acid score and future cardiovascular disease. *EurHeart J.* 2013 Jul;34:1982–9.
278. Newgard CB, An J, Bain JR, Muehlbauer MJ, Stevens RD, Lien LF, et al. A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. *Cell Metab.* 2009 Apr;9:311–26.
279. Wurtz P, Soininen P, Kangas AJ, Ronnema T, Lehtimäki T, Kahonen M, et al. Branched-Chain and Aromatic Amino Acids Are Predictors of Insulin Resistance in Young Adults. *Diabetes Care.* 2013 Mar 1;36(3):648–55.
280. O’Connell TM. The complex role of branched chain amino acids in diabetes and cancer. *Metabolites.* 2013 Oct 14;3:931–45.
281. Costello LC, Franklin RB. The intermediary metabolism of the prostate: a key to understanding the pathogenesis and progression of prostate malignancy. *Oncology.* 2000 Nov;59(4):269–82.
282. Ren J-G, Seth P, Ye H, Guo K, Hanai J-I, Husain Z, et al. Citrate Suppresses Tumor Growth in Multiple Models through Inhibition of Glycolysis, the Tricarboxylic Acid Cycle and the IGF-1R Pathway. *Sci Rep.* 2017 03;7(1):4537.

283. Salazar JH. Overview of Urea and Creatinine. *Lab Med*. 2014 Feb;45(1):e19–20.
284. Leverve X, Carpentier F, Barnoud D, Fontaine E, Guignier M. Intérêt et limites de la 3-méthylhistidine pour la mesure du catabolisme musculaire. *Nutr Clin Métabolisme*. 1992 Jan;6(4):219–25.
285. Takeshita A, Igarashi-Migitaka J, Nishiyama K, Takahashi H, Takeuchi Y, Koibuchi N. Acetyl tributyl citrate, the most widely used phthalate substitute plasticizer, induces cytochrome p450 3a through steroid and xenobiotic receptor. *Toxicol Sci Off J Soc Toxicol*. 2011 Oct;123(2):460–70.
286. Bui TT, Giovanoulis G, Cousins AP, Magnér J, Cousins IT, de Wit CA. Human exposure, hazard and risk of alternative plasticizers to phthalate esters. *Sci Total Environ*. 2016 Jan 15;541:451–67.
287. Rasmussen LM, Sen N, Liu X, Craig ZR. Effects of oral exposure to the phthalate substitute acetyl tributyl citrate on female reproduction in mice. *J Appl Toxicol JAT*. 2017 Jun;37(6):668–75.
288. Kambia N, Farce A, Belarbi K, Gressier B, Luyckx M, Chavatte P, et al. Docking study: PPARs interaction with the selected alternative plasticizers to di(2-ethylhexyl) phthalate. *J Enzyme Inhib Med Chem*. 2016;31(3):448–55.
289. Guenin S, Schwartz L, Morvan D, Steyaert JM, Poinet A, Madelmont JC, et al. PP2A activity is controlled by methylation and regulates oncoprotein expression in melanoma cells: a mechanism which participates in growth inhibition induced by chloroethylnitrosourea treatment. *Int J Oncol*. 2008 Jan;32:49–57.
290. Lange CA, Yee D. Progesterone and breast cancer. *Womens Health Lond Engl*. 2008 Mar;4(2):151–62.
291. Zhenyukh O, Civantos E, Ruiz-Ortega M, Sánchez MS, Vázquez C, Peiró C, et al. High concentration of branched-chain amino acids promotes oxidative stress, inflammation and migration of human peripheral blood mononuclear cells via mTORC1 activation. *Free Radic Biol Med*. 2017;104:165–77.
292. Lin H-M, Barnett MPG, Roy NC, Joyce NI, Zhu S, Armstrong K, et al. Metabolomic analysis identifies inflammatory and noninflammatory metabolic effects of genetic modification in a mouse model of Crohn's disease. *J Proteome Res*. 2010 Apr 5;9(4):1965–75.
293. Messerschmidt A, Worbs M, Steegborn C, Wahl MC, Huber R, Laber B, et al. Determinants of enzymatic specificity in the Cys-Met-metabolism PLP-dependent enzymes family: crystal structure of cystathionine gamma-lyase from yeast and intrafamilial structure comparison. *Biol Chem*. 2003 Mar;384(3):373–86.
294. Coe DM, Viola RE. Assessing the roles of essential functional groups in the mechanism of homoserine succinyltransferase. *Arch Biochem Biophys*. 2007 May;461(2):211–8.
295. Wei X, Jiang S, Zhao X, Li H, Lin W, Li B, et al. Community-Metabolome Correlations of Gut Microbiota from Child-Turcotte-Pugh of A and B Patients. *Front Microbiol*. 2016;7:1856.
296. Fujita T, Hada T, Higashino K. Origin of D- and L-pipecolic acid in human physiological fluids: a study of the catabolic mechanism to pipecolic acid using the lysine loading test. *Clin Chim Acta Int J Clin Chem*. 1999 Sep;287(1–2):145–56.

297. Perera T, Young MR, Zhang Z, Murphy G, Colburn NH, Lanza E, et al. Identification and monitoring of metabolite markers of dry bean consumption in parallel human and mouse studies. *Mol Nutr Food Res*. 2015 Apr;59(4):795–806.
298. Fukai K, Harada S, Iida M, Kurihara A, Takeuchi A, Kuwabara K, et al. Metabolic Profiling of Total Physical Activity and Sedentary Behavior in Community-Dwelling Men. *PLoS One*. 2016;11(10):e0164877.
299. Zhao Q, Shen H, Su K-J, Tian Q, Zhao L-J, Qiu C, et al. A joint analysis of metabolomic profiles associated with muscle mass and strength in Caucasian women. *Aging*. 2018 Oct 14;10(10):2624–35.
300. Wikoff WR, Anfora AT, Liu J, Schultz PG, Lesley SA, Peters EC, et al. Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. *Proc Natl Acad Sci*. 2009 Mar 10;106(10):3698–703.
301. de Mello VD, Paananen J, Lindström J, Lankinen MA, Shi L, Kuusisto J, et al. Indolepropionic acid and novel lipid metabolites are associated with a lower risk of type 2 diabetes in the Finnish Diabetes Prevention Study. *Sci Rep [Internet]*. 2017 Dec [cited 2019 Feb 8];7(1). Available from: <http://www.nature.com/articles/srep46337>
302. Tuomainen M, Lindström J, Lehtonen M, Auriola S, Pihlajamäki J, Peltonen M, et al. Associations of serum indolepropionic acid, a gut microbiota metabolite, with type 2 diabetes and low-grade inflammation in high-risk individuals. *Nutr Diabetes [Internet]*. 2018 Dec [cited 2019 Feb 8];8(1). Available from: <http://www.nature.com/articles/s41387-018-0046-9>
303. Jennis M, Cavanaugh CR, Leo GC, Mabus JR, Lenhard J, Hornby PJ. Microbiota-derived tryptophan indoles increase after gastric bypass surgery and reduce intestinal permeability in vitro and in vivo. *Neurogastroenterol Motil*. 2018 Feb;30(2):e13178.
304. Abildgaard A, Elfving B, Hokland M, Wegener G, Lund S. The microbial metabolite indole-3-propionic acid improves glucose metabolism in rats, but does not affect behaviour. *Arch Physiol Biochem*. 2018 Oct;124(4):306–12.
305. Pallister T, Jennings A, Mohny RP, Yarand D, Mangino M, Cassidy A, et al. Characterizing Blood Metabolomics Profiles Associated with Self-Reported Food Intakes in Female Twins. *PLoS One*. 2016;11(6):e0158568.
306. Guertin KA, Moore SC, Sampson JN, Huang WY, Xiao Q, Stolzenberg-Solomon RZ, et al. Metabolomics in nutritional epidemiology: identifying metabolites associated with diet and quantifying their potential to uncover diet-disease relations in populations. *AmJClinNutr*. 2014 Apr 16;100:208–17.
307. Eggersdorfer M, Wyss A. Carotenoids in human nutrition and health. *Arch Biochem Biophys*. 2018 Aug 15;652:18–26.
308. Tomás-Barberán FA, Selma MV, Espín JC. Interactions of gut microbiota with dietary polyphenols and consequences to human health: *Curr Opin Clin Nutr Metab Care*. 2016 Nov;19(6):471–6.
309. BeMiller JN, Whistler RL. *Starch: chemistry and technology [Internet]*. London: Academic; 2009 [cited 2019 Jun 23]. Available from: <http://site.ebrary.com/id/10305274>

310. Siyal FA, Babazadeh D, Wang C, Arain MA, Saeed M, Ayasan T, et al. Emulsifiers in the poultry industry. *Worlds Poult Sci J.* 2017 Sep;73(3):611–20.
311. Gonzalez-Granda A, Damms-Machado A, Basrai M, Bischoff S. Changes in Plasma Acylcarnitine and Lysophosphatidylcholine Levels Following a High-Fructose Diet: A Targeted Metabolomics Study in Healthy Women. *Nutrients.* 2018 Sep 6;10(9):1254.
312. Vennemann FBC, Ioannidou S, Valsta LM, Dumas C, Ocké MC, Mensink GBM, et al. Dietary intake and food sources of choline in European populations. *Br J Nutr.* 2015 Dec 28;114(12):2046–55.
313. Ueland PM. Choline and betaine in health and disease. *J Inherit Metab Dis.* 2011 Feb;34(1):3–15.
314. Schartum-Hansen H, Pedersen ER, Svingen GFT, Ueland PM, Seifert R, Ebbing M, et al. Plasma choline, smoking, and long-term prognosis in patients with stable angina pectoris. *Eur J Prev Cardiol.* 2015 May;22(5):606–14.
315. Human Metabolome Database: Showing metabocard for Ethyl beta-D-glucopyranoside (HMDB0029968) [Internet]. [cited 2019 Jun 23]. Available from: <http://www.hmdb.ca/metabolites/HMDB0029968>
316. McGowan MA, Scheman A, Jacob SE. Propylene Glycol in Contact Dermatitis: A Systematic Review. *Dermat Contact Atopic Occup Drug.* 2018 Feb;29(1):6–12.
317. Zakerali T, Shahbazi S. Rational Druggability Investigation Toward Selection of Lead Molecules: Impact of the Commonly Used Spices on Inflammatory Diseases. *ASSAY Drug Dev Technol.* 2018 Oct;16(7):397–407.
318. European Union. Register of Feed Additives, pursuant to Regulation (EC) No 1831/2003, Annex I : List of additives. 2019.
319. EFSA, Panel on Additives and Products or Substances used in Animal Feed. Safety and efficacy of pyridine and pyrrole derivatives belonging to chemical group 28 when used as flavourings for all animal species. *EFSA Journal.* 2016;14(2).
320. Steiber A, Kerner J, Hoppel CL. Carnitine: a nutritional, biosynthetic, and functional perspective. *Mol Aspects Med.* 2004 Dec;25(5–6):455–73.
321. Sun H, Liu D. Hydrodynamic delivery of interleukin 15 gene promotes resistance to high fat diet-induced obesity, fatty liver and improves glucose homeostasis. *Gene Ther.* 2015 Apr;22(4):341–7.
322. Hutson SM, Sweatt AJ, Lanoue KF. Branched-chain [corrected] amino acid metabolism: implications for establishing safe intakes. *J Nutr.* 2005;135(6 Suppl):1557S-64S.
323. Ananieva EA, Powell JD, Hutson SM. Leucine Metabolism in T Cell Activation: mTOR Signaling and Beyond. *Adv Nutr Int Rev J.* 2016 Jul;7(4):798S-805S.
324. Roh W-G, Shin H-C, Choi J-H, Lee YJ, Kim K. Alcohol consumption and higher incidence of impaired fasting glucose or type 2 diabetes in obese Korean men. *Alcohol Fayettev N.* 2009 Dec;43(8):643–8.
325. Larsen MO, Rolin B, Wilken M, Carr RD, Svendsen O. High-fat high-energy feeding impairs fasting glucose and increases fasting insulin levels in the Göttingen minipig: results from a pilot study. *Ann N Y Acad Sci.* 2002 Jun;967:414–23.

326. Rothwell JA, Madrid-Gambin F, Garcia-Aloy M, Andres-Lacueva C, Logue C, Gallagher AM, et al. Biomarkers of intake for coffee, tea, and sweetened beverages. *Genes Nutr.* 2018;13:15.
327. Andersen MB, Rinnan A, Manach C, Poulsen SK, Pujos-Guillot E, Larsen TM, et al. Untargeted metabolomics as a screening tool for estimating compliance to a dietary pattern. *JProteomeRes.* 2014 Mar 7;13:1405–18.
328. O’Sullivan A, Gibney MJ, Brennan L. Dietary intake patterns are reflected in metabolomiprofiles : potential role in dietary assessment studies. *AmJClinNutr.* 2011 Feb;93:314–21.
329. Alcoholic drinks [Internet]. World Cancer Research Fund. 2018 [cited 2019 Jun 25]. Available from: <https://www.wcrf.org/dietandcancer/exposures/alcoholic-drinks>
330. Diwan V, Poudyal H, Brown L. Piperine Attenuates Cardiovascular, Liver and Metabolic Changes in High Carbohydrate, High Fat-Fed Rats. *Cell Biochem Biophys.* 2013 Nov;67(2):297–304.
331. Rather RA, Bhagat M. Cancer Chemoprevention and Piperine: Molecular Mechanisms and Therapeutic Opportunities. *Front Cell Dev Biol* [Internet]. 2018 Feb 15 [cited 2019 Feb 28];6. Available from: <http://journal.frontiersin.org/article/10.3389/fcell.2018.00010/full>
332. Ajazuddin null, Alexander A, Qureshi A, Kumari L, Vaishnav P, Sharma M, et al. Role of herbal bioactives as a potential bioavailability enhancer for Active Pharmaceutical Ingredients. *Fitoterapia.* 2014 Sep;97:1–14.
333. Greenshields AL, Doucette CD, Sutton KM, Madera L, Annan H, Yaffe PB, et al. Piperine inhibits the growth and motility of triple-negative breast cancer cells. *Cancer Lett.* 2015 Feb;357(1):129–40.
334. Do MT, Kim HG, Choi JH, Khanal T, Park BH, Tran TP, et al. Antitumor efficacy of piperine in the treatment of human HER2-overexpressing breast cancer cells. *Food Chem.* 2013 Dec;141(3):2591–9.
335. Doucette CD, Hilchie AL, Liwski R, Hoskin DW. Piperine, a dietary phytochemical, inhibits angiogenesis. *J Nutr Biochem.* 2013 Jan;24(1):231–9.
336. Zheng J, Zhou Y, Li Y, Xu D-P, Li S, Li H-B. Spices for Prevention and Treatment of Cancers. *Nutrients.* 2016 Aug 12;8(8):495.
337. Alkaloids [Internet]. [cited 2019 May 24]. Available from: [https://www.chemicalbook.com/ProductCatalog\\_EN/2322.htm](https://www.chemicalbook.com/ProductCatalog_EN/2322.htm)
338. United States Consumer Product Safety Commission. Review of Exposure and Toxicity Data for Phthalate Substitutes. 2010.
339. Sheftel VO. Indirect food additives and polymers: migration and toxicology [Internet]. 2000 [cited 2019 May 24]. Available from: [https://nls.ldls.org.uk/welcome.html?ark:/81055/vdc\\_100056086042.0x000001](https://nls.ldls.org.uk/welcome.html?ark:/81055/vdc_100056086042.0x000001)
340. Welle F, Wolz G, Franz R. Migration of plasticizers from PVC tubes into enteral feeding solutions. *Pharma International.* 2005;(3):17–21.
341. Testai E, Ms Scientific Committee SCENIHR. Electronic address: SANTE-C2-SCENIHR@ec.europa.eu, Hartemann P, Rastogi SC, Bernauer U, Piersma A, et al. The safety of medical devices containing



DEHP plasticized PVC or other plasticizers on neonates and other groups possibly at risk (2015 update). *Regul Toxicol Pharmacol RTP*. 2016 Apr;76:209–10.

342. Nara K, Nishiyama K, Natsugari H, Takeshita A, Takahashi H. Leaching of the Plasticizer, Acetyl Tributyl Citrate: (ATBC) from Plastic Kitchen Wrap. *J Health Sci*. 2009;55(2):281–4.
343. Rasmussen LM, Sen N, Vera JC, Liu X, Craig ZR. Effects of in vitro exposure to dibutyl phthalate, mono-butyl phthalate, and acetyl tributyl citrate on ovarian antral follicle growth and viability. *Biol Reprod*. 2017 May 1;96(5):1105–17.
344. World Cancer Research Fund/American Institute for Cancer Research. Continuous Update Project Expert Report 2018. Diet, nutrition, physical activity and breast cancer [Internet]. Available from: [dietandcancerreport.org](http://dietandcancerreport.org)
345. Latino-Martel P, Cottet V, Druesne-Pecollo N, Pierre FHF, Touillaud M, Touvier M, et al. Alcoholic beverages, obesity, physical activity and other nutritional factors, and cancer risk: A review of the evidence. *Crit Rev Oncol Hematol*. 2016 Mar;99:308–23.
346. Jung S, Wang M, Anderson K, Baglietto L, Bergkvist L, Bernstein L, et al. Alcohol consumption and breast cancer risk by estrogen receptor status: in a pooled analysis of 20 studies. *Int J Epidemiol*. 2016;45(3):916–28.
347. Singletary KW, Gapstur SM. Alcohol and breast cancer: review of epidemiologic and experimental evidence and potential mechanisms. *JAMA*. 2001 Nov 7;286(17):2143–51.
348. Dorgan JF, Baer DJ, Albert PS, Judd JT, Brown ED, Corle DK, et al. Serum hormones and the alcohol-breast cancer association in postmenopausal women. *J Natl Cancer Inst*. 2001 May 2;93(9):710–5.
349. PubChem. 2-Amino-4-cyanobutanoic acid [Internet]. [cited 2019 Jun 26]. Available from: <https://pubchem.ncbi.nlm.nih.gov/compound/440770>
350. Bourassa MW, Alim I, Bultman SJ, Ratan RR. Butyrate, neuroepigenetics and the gut microbiome: Can a high fiber diet improve brain health? *Neurosci Lett*. 2016 20;625:56–63.
351. Liu H, Wang J, He T, Becker S, Zhang G, Li D, et al. Butyrate: A Double-Edged Sword for Health? *Adv Nutr*. 2018 Jan 1;9(1):21–9.
352. Morrison DJ, Preston T. Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism. *Gut Microbes*. 2016 03;7(3):189–200.
353. Lupton JR. Microbial Degradation Products Influence Colon Cancer Risk: the Butyrate Controversy. *J Nutr*. 2004 Feb 1;134(2):479–82.
354. Api AM, Belmonte F, Belsito D, Botelho D, Bruze M, Burton GA, et al. RIFM fragrance ingredient safety assessment, butyric acid, CAS Registry Number 107-92-6. *Food Chem Toxicol* [Internet]. 2019 Jan [cited 2019 Mar 21]; Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0278691519300420>
355. Merlot AM, Kalinowski DS, Richardson DR. Unraveling the mysteries of serum albumin-more than just a serum protein. *Front Physiol*. 2014;5:299.

356. Dufour C, Dangles O. Flavonoid-serum albumin complexation: determination of binding constants and binding sites by fluorescence spectroscopy. *Biochim Biophys Acta*. 2005 Jan 18;1721(1–3):164–73.
357. Khan MK, Rakotomanomana N, Dufour C, Dangles O. Binding of citrus flavanones and their glucuronides and chalcones to human serum albumin. *Food Funct*. 2011;2(10):617.
358. Benavente-García O, Castillo J. Update on uses and properties of citrus flavonoids: new findings in anticancer, cardiovascular, and anti-inflammatory activity. *J Agric Food Chem*. 2008 Aug 13;56(15):6185–205.
359. Carayol M, Licaj I, Achaintre D, Sacerdote C, Vineis P, Key TJ, et al. Reliability of Serum Metabolites over a Two-Year Period: A Targeted Metabolomic Approach in Fasting and Non-Fasting Samples from EPIC. *PLoSOne*. 2015;10:e0135437.
360. Floegel A, Drogan D, Wang-Sattler R, Prehn C, Illig T, Adamski J, et al. Reliability of serum metabolite concentrations over a 4-month period using a targeted metabolomic approach. *PLoSOne*. 2011;6:e21103.
361. Breier M, Wahl S, Prehn C, Fugmann M, Ferrari U, Weise M, et al. Targeted metabolomics identifies reliable and stable metabolites in human serum and plasma samples. *PloS One*. 2014;9(2):e89728.
362. Yu Z, Kastenmüller G, He Y, Belcredi P, Möller G, Prehn C, et al. Differences between human plasma and serum metabolite profiles. *PloS One*. 2011;6(7):e21230.
363. McBride KL, Belmont JW, O'Brien WE, Amin TJ, Carter S, Lee BH. Heritability of plasma amino acid levels in different nutritional states. *Mol Genet Metab*. 2007 Feb;90(2):217–20.
364. Andersson A, Marklund M, Diana M, Landberg R. Plasma alkylresorcinol concentrations correlate with whole grain wheat and rye intake and show moderate reproducibility over a 2- to 3-month period in free-living Swedish adults. *J Nutr*. 2011 Sep;141(9):1712–8.
365. Zamora-Ros R, Touillaud M, Rothwell JA, Romieu I, Scalbert A. Measuring exposure to the polyphenol metabolome in observational epidemiologic studies: current tools and applications and their limits. *Am J Clin Nutr*. 2014 Jul;100(1):11–26.
366. Manach C, Williamson G, Morand C, Scalbert A, Rémésy C. Bioavailability and bioefficacy of polyphenols in humans. I. Review of 97 bioavailability studies. *Am J Clin Nutr*. 2005;81(1 Suppl):230S–242S.
367. Soderholm PP, Koskela AH, Lundin JE, Tikkanen MJ, Adlercreutz HC. Plasma pharmacokinetics of alkylresorcinol metabolites: new candidate biomarkers for whole-grain rye and wheat intake. *AmJClinNutr*. 2009 Nov;90:1167–71.
368. Gibson R. Principles of nutritional assessment. 2nd Ed. Oxford, United Kingdom: Oxford University Press; 2005.
369. Hebert JR, Hurley TG, Peterson KE, Resnicow K, Thompson FE, Yaroch AL, et al. Social desirability trait influences on self-reported dietary measures among diverse participants in a multicenter multiple risk factor trial. *J Nutr*. 2008 Jan;138(1):226S–234S.
370. Joinson A. Social desirability, anonymity, and Internet-based questionnaires. *BehavResMethods InstrumComput*. 1999 Aug;31:433–8.

371. Kesse-Guyot E, Fezeu L, Jeandel C, Ferry M, Andreeva V, Amieva H, et al. French adults' cognitive performance after daily supplementation with antioxidant vitamins and minerals at nutritional doses: a post hoc analysis of the Supplementation in Vitamins and Mineral Antioxidants (SU.VI.MAX) trial. *Am J Clin Nutr*. 2011 Sep;94(3):892–9.
372. Andreeva V, Kesse-Guyot E. Nutrition and Cognition in the Context of Ageing: Role of Dietary Patterns. In: Best T, Dye L, editors. *Nutrition for Brain Health and Cognitive Performance* [Internet]. CRC Press; 2015 [cited 2019 Jun 28]. p. 11–40. Available from: <http://www.crcnetbase.com/doi/10.1201/b18563-4>
373. Willett WC. *Nutritional Epidemiology*. Vol. 2nd. New York, NY: Oxford University Press; 1998.
374. Mennen LI, Bertrais S, Galan P, Arnault N, Potier de C, Hercberg S. The use of computerised 24 h dietary recalls in the French SU.VI.MAX Study: number of recalls required. *EurJClinNutr*. 2002 Jul;56:659–65.
375. Nohr EA, Frydenberg M, Henriksen TB, Olsen J. Does low participation in cohort studies induce bias? *Epidemiol Camb Mass*. 2006 Jul;17(4):413–8.
376. Rothman KJ, Greenland S, Lash TL. *Modern epidemiology*. 3rd ed., thoroughly rev. and updated. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2008. 758 p.
377. Chun H, Keleş S. Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *J R Stat Soc Ser B Stat Methodol*. 2010 Jan;72(1):3–25.
378. Xia J, Wishart DS. MSEA: a web-based tool to identify biologically meaningful patterns in quantitative metabolomic data. *Nucleic Acids Res*. 2010 Jul 1;38(Web Server):W71–7.
379. KEGG PATHWAY Database [Internet]. [cited 2019 Jun 28]. Available from: <https://www.genome.jp/kegg/pathway.html>
380. Krumsiek J, Suhre K, Illig T, Adamski J, Theis FJ. Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data. *BMC Syst Biol*. 2011 Jan 31;5:21.
381. Fiehn O, Robertson D, Griffin J, van der Werf M, Nikolau B, Morrison N, et al. The metabolomics standards initiative (MSI). *Metabolomics*. 2007 Sep;3(3):175–8.
382. COnsortium of METabolomics Studies (COMETS) | EGRP/DCCPS/NCI/NIH [Internet]. [cited 2019 Jun 29]. Available from: <https://epi.grants.cancer.gov/comets/>
383. The FOOTBALL Portal [Internet]. [cited 2019 Jun 29]. Available from: <http://foodmetabolome.org/>
384. Oxygenate - The Project [Internet]. [cited 2019 Jun 29]. Available from: <https://www6.inra.fr/jpi-hdhl-biomarkers-oxygenate/The-Project>

## Annexes

---

Annexe 1 Texte intégral “NMR metabolomic signatures reveal predictive plasma metabolites associated with long-term risk of developing breast cancer” .....	187
Annexe 2 Texte intégral : “Plasma metabolomic signatures associated with long-term breast cancer risk in the SU.VI.MAX prospective cohort” .....	216
Annexe 3 Associations entre les signaux RMN et le risque de cancer de la prostate issues des régressions logistiques conditionnelles.....	244
Annexe 4 Corrélations significatives entre les signaux RMN 2 à 2 identifiés (cas-témoins niché, cancer de la prostate) .....	252
Annexe 5 Signaux RMN associés avec le risque de cancer de la prostate (intensités catégorisées en tertiles) .....	254
Annexe 6 Texte intégral: “Untargeted plasma metabolomic profiles associated with overall diet in women from the SU.VI.MAX cohort.” .....	256
Annexe 7 Ions associés à la consommation de groupes d’aliments spécifiques chez les femmes (analyses en LC-MS) .....	342
Annexe 8 Associations entre les signaux RMN et les composantes alimentaires du mPNNS-GS chez les femmes.....	344
Annexe 9 Associations entre les signaux RMN et les profils alimentaires obtenus par ACP chez les femmes .....	350
Annexe 10 Associations entre les signaux RMN et les composantes alimentaires du mPNNS-GS chez les hommes.....	352
Annexe 11 Associations entre les signaux RMN et les profils alimentaires obtenus par ACP chez les hommes.....	356
Annexe 12 Texte intégral: “Diet-related metabolomic signature of long-term breast cancer risk using penalized regression: an exploratory study in the SU.VI.MAX cohort” .....	358
Annexe 13 Signaux RMN associés à l’alimentation (score DQI-I, mPNNS-GS, profil alimentaire « occidental », profil alimentaire « sain », groupes d’aliments) chez les femmes .....	413
Annexe 14 Signaux RMN associés à l’alimentation (score DQI-I, mPNNS-GS, profil alimentaire « occidental », profil alimentaire « sain », groupes d’aliments) chez les hommes .....	414



## Cancer

## NMR metabolomic signatures reveal predictive plasma metabolites associated with long-term risk of developing breast cancer

Lucie Lécuyer,<sup>1\*</sup> Agnès Victor Bala,<sup>2</sup> Mélanie Deschasaux,<sup>1</sup>  
 Nadia Bouchemal,<sup>2</sup> Mohamed Nawfal Triba,<sup>2</sup> Marie-Paule Vasson,<sup>3,4</sup>  
 Adrien Rossary,<sup>3</sup> Aicha Demidem,<sup>3</sup> Pilar Galan,<sup>1</sup> Serge Hercberg,<sup>1,5</sup>  
 Valentin Partula,<sup>1</sup> Laurence Le Moyec,<sup>6</sup> Bernard Srour,<sup>1</sup>  
 Thibault Fiolet,<sup>1</sup> Paule Latino-Martel,<sup>1</sup> Emmanuelle Kesse-Guyot,<sup>1</sup>  
 Philippe Savarin<sup>2</sup> and Mathilde Touvier<sup>1</sup>

<sup>1</sup>Sorbonne Paris Cité Epidemiology and Statistics Research Center (CRESS), French National Institute of Health and Medical Research (Inserm) U1153, French National Institute for Agricultural Research (Inra) U1125, French National Conservatory of Arts and Crafts (Cnam), Paris 13 University, Nutritional Epidemiology Research Team (EREN), 93017 Bobigny Cedex, France, <sup>2</sup>Chemistry Structures Properties of Biomaterials and Therapeutic Agents (CSPBAT), The National Center for Scientific Research (CNRS) 7244, Paris 13 University, Spectroscopy Biomolecules and Biological Environment (SBMB), 93017 Bobigny Cedex, France, <sup>3</sup>Clermont Auvergne University, INRA, Human Nutrition Unit (UNH), CRNH Auvergne, 63009 Clermont-Ferrand Cedex, France, <sup>4</sup>Anticancer Center Jean-Perrin, CHU Clermont-Ferrand, 63011 Clermont-Ferrand Cedex, France, <sup>5</sup>Public Health Department, Avicenne Hospital, 93000 Bobigny, France and <sup>6</sup>UBIAE, INSERM, Evry University, Paris-Saclay University, 91025 Evry, France

\*Corresponding author. EREN, Inserm U1153, SMBH Paris 13, 74 rue Marcel Cachin, F-93017 Bobigny Cedex, France.  
 E-mail: l.lecuyer@eren.smbh.univ-paris13.fr

Editorial decision 3 December 2017; Accepted 4 January 2018

### Abstract

**Background:** Combination of metabolomics and epidemiological approaches opens new perspectives for ground-breaking discoveries. The aim of the present study was to investigate for the first time whether plasma untargeted metabolomic profiles, established from a simple blood draw from healthy women, could contribute to predict the risk of developing breast cancer within the following decade and to better understand the aetiology of this complex disease.

**Methods:** A prospective nested case–control study was set up in the Supplémentation en Vitamines et Minéraux Antioxydants (SU.VI.MAX) cohort, including 206 breast cancer cases diagnosed during a 13-year follow-up and 396 matched controls. Untargeted nuclear magnetic resonance (NMR) metabolomic profiles were established from baseline plasma samples. Multivariable conditional logistic regression models were computed for each individual NMR variable and for combinations of variables derived by principal component analysis.

**Results:** Several metabolomic variables from 1D NMR spectroscopy were associated with breast cancer risk. Women characterized by higher fasting plasma levels of valine, lysine, arginine, glutamine, creatine, creatinine and glucose, and lower plasma levels of lipoproteins, lipids, glycoproteins, acetone, glycerol-derived compounds and unsaturated lipids had a higher risk of developing breast cancer. *P*-values ranged from 0.00007 [odds ratio (OR)<sub>T3vsT1</sub>=0.37 (0.23–0.61) for glycerol-derived compounds] to 0.04 [OR<sub>T3vsT1</sub>=1.61 (1.02–2.55) for glutamine].

**Conclusion:** This study highlighted associations between baseline NMR plasma metabolomic signatures and long-term breast cancer risk. These results provide interesting insights to better understand complex mechanisms involved in breast carcinogenesis and evoke plasma metabolic disorders favourable for carcinogenesis initiation. This study may contribute to develop screening strategies for the identification of at-risk women for breast cancer well before symptoms appear.

**Key words:** metabolomics, breast neoplasm, nuclear magnetic resonance, plasma, prospective study

#### Key Messages

- The present study investigated for the first time whether plasma fully untargeted metabolomic profiles (established from a simple blood draw from healthy women) could contribute both to predict the risk of developing breast cancer within the following decade and to better understand the aetiology of this complex disease.
- Baseline nuclear magnetic resonance (NMR) plasma metabolomic signatures were associated with long-term breast cancer risk : women characterized by higher fasting plasma levels of valine, lysine, arginine, glutamine, creatine, creatinine and glucose, and lower plasma levels of lipoproteins, lipids, glycoproteins, acetone, glycerol-derived compounds and unsaturated lipids had a higher risk of developing breast cancer during the 13-year follow-up.
- This study provides interesting insights into the understanding of complex early mechanism pathways involved in breast cancer carcinogenesis. If they are replicated and validated in other independent prospective studies, this could lead to the identification of strong biomarkers of breast cancer risk and contribute to the identification of at-risk women for breast cancer well before symptoms appear and improvement of prevention strategies.

#### Introduction

Breast cancer is a major cause of death in women of economically developed nations.<sup>1</sup> Current predictive models based on well-established risk factors of breast cancer (genetic, environmental and lifestyle) appear largely incomplete.<sup>2</sup> In order to improve breast cancer prevention, key challenges lie in a better comprehension of mechanisms involved in its aetiology, and in the identification of new biomarkers or metabolomic profiles associated with subsequent long-term breast cancer risk.

Metabolomics, one of the newest 'omics' sciences, is an emerging and powerful discipline that provides a comprehensive and detailed picture of the phenotype of living beings through a snapshot of endogenous and exogenous metabolites in cells, tissues or biofluids.<sup>3,4</sup> In contrast with the traditional approach, in which molecules are studied individually, metabolomic platforms allow the study of more complex interactions in biological systems. Metabolomic profiling is a promising tool to discover new

disease-risk biomarkers and to provide novel insights into disease aetiology, biological mechanisms and pathways.<sup>5–8</sup>

The European FP7 project METAcancer previously showed that metabolomics can provide information on the stage and grade of breast tumours.<sup>9</sup> Furthermore, cross-sectional studies highlighted differences between metabolomic profiles of breast cancer patients and controls.<sup>10,11</sup> These results could be useful to improve early diagnosis of breast cancer, neoplasms classification, identification of new prognostic markers or the discovery of new targets for therapeutic interventions. However, since the presence of a breast tumour itself may alter metabolomic profiles,<sup>12</sup> this type of cross-sectional study limits the ability to discover predictive biomarkers and to draw causal inference. Breast cancer prevention relies on the ability to identify at-risk individuals, long before the development of overt cancers. A recent prospective study<sup>13</sup> obtained very promising results regarding the ability of metabolomics to improve short-term breast cancer risk prediction and

therefore early diagnosis. Another one focused on the associations between nutrition-related metabolites and breast cancer risk.<sup>14</sup> However, to our knowledge, no prospective study in the field of primary prevention has previously investigated the associations between baseline fully untargeted metabolomic profiles and subsequent long-term risk of breast cancer, in order to discover new potential predictive metabolomic profiles and to better understand the aetiology of this complex disease.

Thus, the aim of the present study was to investigate whether metabolomic profiles, established from a simple blood draw from healthy women, could contribute to predict the risk of developing breast cancer in the 10–15 subsequent years. To this end, an untargeted nuclear magnetic resonance (NMR) metabolomic analysis, a highly reliable and reproducible technology,<sup>15</sup> was performed in a case–control study nested within the Supplémentation en Vitamines et Minéraux Antioxydants (SU.VI.MAX) prospective cohort.

## Methods

### Population study

The SU.VI.MAX study (clinicaltrials.gov; NCT00272428) was initially designed as a double-blind placebo-controlled trial with the purpose of assessing the influence of a daily supplementation with nutritional doses of antioxidants on the incidence of cardiovascular diseases and cancers. The study design and methods have been previously detailed.<sup>16,17</sup> Briefly, a total of 13 017 participants were enrolled in 1994–95 for an 8-year intervention trial and were followed up for health events until September 2007. The written informed consent for each participant was obtained. Participants were advised against taking any self-prescribed supplements during the trial. The study was conducted according to the Declaration of Helsinki guidelines and was approved by the Ethics Committee for Studies with Human Subjects of Paris-Cochin Hospital (CCPPRB 706/2364) and the ‘Commission Nationale de l’Informatique et des Libertés’ (CNIL 334641/907094).

### Baseline data collection

At enrolment, all participants underwent a clinical examination by the study nurses and physicians, with anthropometric measurements and a blood draw, occurring after a 12-hour fasting period. Information on socio-demographics, smoking habits, physical activity, medication use and health status were collected through self-administered questionnaires. Age at menopause was collected from the participants by follow-up questionnaires.

A 35-mL venous blood sample was collected in sodium heparin Vacutainer tubes (Becton Dickinson, Rungis, France)

from all fasting participants. After centrifugation at +4°C, plasma aliquots were immediately prepared and stored frozen at –20°C during less than 2 days and then stored in liquid nitrogen.

### Case ascertainment

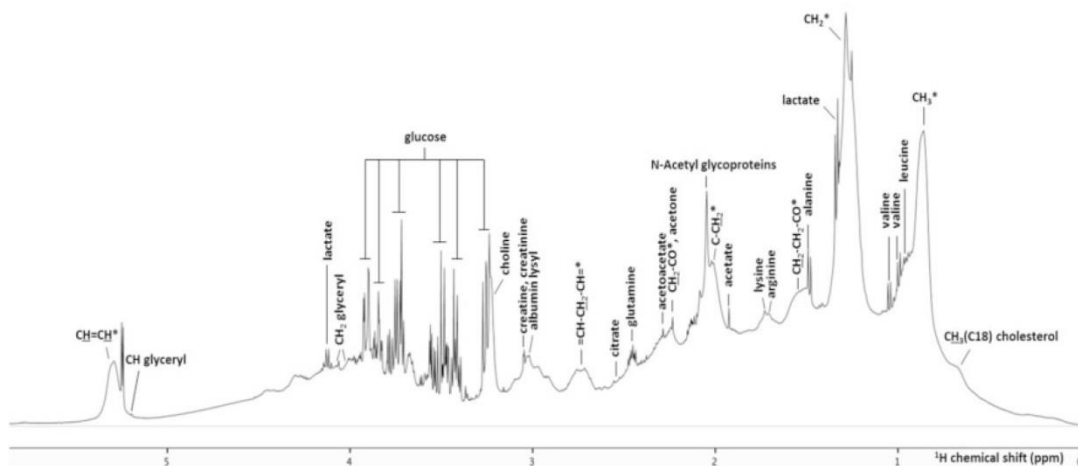
Health events were self-reported by the participants in regular follow-up questionnaires. Then, all relevant medical information and pathological reports were gathered through participants, physicians and/or hospitals and subsequently reviewed by an independent physician expert committee. Validated cases were classified according to the International Chronic Diseases Classification, 10th Revision (ICD-10).<sup>18</sup>

### Nested case–control study

All participants with a first incident invasive breast cancer, diagnosed between 1 year after their inclusion in the SU.VI.MAX cohort in 1994–95 and September 2007, were included in this nested case–control study ( $n=206$ ). Incident breast cancers diagnosed during the first year of follow-up were excluded to avoid reverse causality bias and guarantee the prospective design. For each case, two controls were randomly selected and matched for baseline age (35–39 years/40–44 years/45–49 years/50–54 years/55–59 years/>60 years), menopausal status (pre/post-menopause status at baseline and at diagnosis), body mass index (BMI) (underweight, normal weight and overweight/obese), intervention group of the initial SU.VI.MAX trial (placebo/supplemented), smoking status (current smokers and non-smokers) and season of blood draw (a priori-defined periods: October–November/December–January–February/March–April–May). The method for control selection was density sampling, i.e. every time a case was diagnosed, two controls were selected from other members of the cohort who, at that time, did not have diagnosed breast cancer.

### NMR metabolomic analysis

Plasma samples were then analysed by proton nuclear magnetic resonance (<sup>1</sup>H NMR). NMR spectra from two complementary experiments—One-dimensional <sup>1</sup>H Nuclear Overhauser Effect Spectroscopy NOESY1D presat (NOESY1dgppr sequence)<sup>19</sup> and Carr-Purcell-Meiboom-Gill (CPMG presat)<sup>20</sup>—were acquired for each plasma sample at 500 MHz on a Bruker AVANCE III spectrometer using automatic sample changer. Details on sample preparation, NMR data acquisition and spectra processing are available in Supplementary File 1, available as Supplementary Data at *IJE* online, and quality controls are presented in Supplementary File 2, available as Supplementary Data at



**Figure 1.** Proton NMR spectrum of plasma sample with metabolites' assignment. The figure represents a  $^1\text{H}$  NMR spectrum (one-dimensional) from 500 MHz  $^1\text{H}$  NOESY1D experiments of plasma sample with chemical shift (ppm) on the horizontal axis. \*, moieties of lipids.

*IJE* online. After processing and calibration, each 1D NMR spectrum was sliced into buckets of 0.001 ppm, containing NMR signals, using NMRPipe software.<sup>21</sup> An 'intelligent bucketing' procedure<sup>22</sup> was then applied over a chemical shift range of 0–9.0 ppm to limit the number of buckets and facilitate their assignment in gathering buckets corresponding to the same peak. The intelligent buckets were scaled to the total summed integrals for each spectrum and then used as the final input for the multivariable models (= 'NMR variables'). NMR variables thus corresponded to continuous variables representing the integrals of peak entities from NMR spectra and were proportional to metabolite concentrations in plasma. Overall, after the 'intelligent bucketing' procedure, 114 (from NOESY1D experiments) and 96 (from CPMG experiments) NMR variables were used in further statistical analyses. The most probable assignment of NMR signals was achieved by gathering information from additional 2D NMR experiments ( $^1\text{H}$ – $^1\text{H}$  TOCSY and J-resolved), acquired on a set of representative samples. CHENOMX software, the Human Metabolome Database<sup>23</sup> and ad hoc literature were used to confirm the assignments. Figure 1 shows a 500-MHz  $^1\text{H}$  NOESY1D spectrum of plasma sample with metabolites' assignment. Overall, 33 subjects—30 controls (4.9%) and 3 cases (0.5%)—were excluded due to spectra acquisition problems and/or to low volume of plasma, resulting in 206 cases and 396 controls spectra for NOESY1D experiments and 203 cases and 378 controls spectra for the CPMG experiments.

### Statistical analysis

Participants' baseline characteristics were compared between cases and controls using  $\chi^2$  test for categorical variables and Student test for continuous variables.

Prospective associations between each NMR variable from both NOESY1D and CPMG experiments and breast

cancer risk were characterized using multivariable conditional logistic regression models. Odds ratios (ORs) for tertiles of NMR variables and their 95% confidence intervals (95% CIs) were computed. The lowest tertile was used as reference. Multivariable models were adjusted for the following baseline characteristics: age, BMI, menopausal status, intervention group of the SU.VI.MAX trial, smoking status, season of blood draw, height, alcohol intake, physical activity, education level, family history of breast cancer, number of children and use of hormone replacement therapy for menopause; all these covariates constituted the reference model. NMR variables associated with breast cancer risk corresponding to different peaks of the same metabolite were combined into a single variable by summing up the buckets intensities. A Spearman correlation matrix of these metabolites was computed for NOESY1D and CPMG experiments (see Supplementary File 3, available as Supplementary Data at *IJE* online).

In order to handle the False Discovery Rate (FDR) due to multiple testing, a Benjamini-Hochberg (BH) correction was tested (FDR < 0.05 for 114 tests for NOESY1D experiments and for 96 tests for CPMG experiments, calculated separately).<sup>24</sup> Sensitivity analyses were performed by excluding cases diagnosed before 2 years of follow-up. Exploratory stratified analyses were also performed according to menopausal status at diagnosis and according to the median follow-up duration in order to investigate whether some metabolites were more specifically associated with short- or long-term breast cancer risk.

A principal component analysis (PCA) was then conducted on the NMR variables associated with breast cancer, identified by logistic regressions from CPMG or NOESY1D experiments. The first two components were retained according to Cattell's scree plots (plots of the total variance explained by each pattern)<sup>25,26</sup> and introduced in



multivariable conditional logistic regression models (with same adjustments as described above), in order to investigate the associations between these metabolomic profiles and breast cancer risk.

Last, AUC (Area Under the Curve) and NRI (Net Reclassification Improvement) indexes<sup>27</sup> were calculated to assess the improvement in the predictive performance of breast cancer risk logistic models when single metabolites or metabolomic profiles (single NMR variables or components from PCA) were added, compared with the reference model based on known risk factors only, as defined above.

All analyses were performed using SAS software (v9.3, Cary, NC). All tests were two-sided.

## Results

Table 1 summarizes the baseline characteristics of breast cancer cases and controls. Compared with controls, cases were taller and were more likely to have a family history of breast cancer.

NMR variables associated with breast cancer risk in conditional logistic regression models are presented in Table 2, sorted by chemical shift (over a chemical shift range of 0–5.5 ppm). The NMR spectrum area ranging from 5.5 to 9.0 ppm was withdrawn in further analysis because it showed only very weak signals close to noise level. The results for the complete list of NMR variables are available in Supplementary File 4, available as Supplementary Data at *IJE* online. For the NOESY1D experiments, lipoproteins, lipids, glycoproteins, acetone, glycerol-derived compounds (such as glycerides and phospholipids) and unsaturated lipids were associated with a decreased risk of developing breast cancer. For the CPMG experiments, lipoproteins, lipids and acetone were also associated with decreased risk, whereas valine, lysine, arginine, glutamine, creatine, creatinine and glucose were associated with elevated risk of developing breast cancer. Lipids ( $p$ -corrected=0.02), acetone ( $p$ -corrected=0.02) and glycerol-derived compounds ( $p$ -corrected=0.04 for CH<sub>2</sub> moiety and  $p$ -corrected=0.008 for CH moiety) (NOESY1D experiments) were still associated after BH correction (not tabulated). Results of sensitivity and stratification analyses are presented in Supplementary File 5, available as Supplementary Data at *IJE* online. Most of the results were unchanged when excluding cases diagnosed during the first 2 years of follow-up.

All NMR variables associated with breast cancer risk in logistic regression were then analysed by PCA. The first two PCA components ('metabolomic profiles', explaining >78% of total variability) were integrated into multivariable conditional logistic regressions (Table 3). The first component was characterized by elevated plasma levels of lipoproteins, lipids, acetone, unsaturated lipids, glycerol-derived compounds and glycoproteins (loading values  $\geq 0.15$ ) and lower

plasma levels of glutamine, lysine, creatine and creatinine (loading value  $\leq -0.15$ ). Only this first metabolomic profile was associated with a decreased risk of breast cancer [OR<sub>T3vsT1</sub>=0.45 (0.28–0.72),  $p=0.0008$ ]. Loading values are displayed in Supplementary File 6, available as Supplementary Data at *IJE* online.

Regarding NRI indexes, the predictive performance of breast cancer risk models were most of the time improved after inclusion of single metabolites or metabolomic patterns into the reference models, whereas the improvement in AUCs was more limited (Table 4).

## Discussion

To our knowledge, this study was the first to prospectively investigate the relationship between baseline fully untargeted metabolomic profiles and subsequent long-term breast cancer risk. Women characterized by higher fasting plasma levels of valine, lysine, arginine, glutamine, creatine, creatinine and glucose, and lower plasma levels of lipoproteins, lipids, glycoproteins, acetone, glycerol-derived compounds and unsaturated lipids had a higher risk of developing breast cancer during the 13-year follow-up.

Improvements in AUCs were limited, even for metabolites that were strongly associated with breast cancer risk in logistic models. This illustrates the fact that the main interest of these results lies in a better understanding of breast cancer aetiology rather than in an improvement in discrimination models. Although no previous study conducted similar analysis for long-term risk of breast cancer, our results could be interpreted in light of several prospective metabolomic studies: one based on NMR metabolomics and short-term risk of breast cancer,<sup>13</sup> one based on targeted mass spectrometry (MS) metabolomics focusing on 120 specific circulating and already known metabolites (acylcarnitines, amino acids, biogenic amines, phosphatidylcholines, sphingolipids and hexoses) and breast, prostate and colorectal cancer risk<sup>28</sup> and others based on NMR or untargeted MS and risk of other cancer locations (hepatocellular,<sup>29–31</sup> prostate,<sup>32–34</sup> glioblastoma,<sup>35</sup> colorectal<sup>28,36,37</sup> and lung<sup>38</sup>).

Regarding lipid compounds and consistently with our results, several prospective metabolomic studies observed inverse associations between plasma or serum total lipids,<sup>29,33,34,36</sup> unsaturated lipids,<sup>29,30,33,34,36</sup> glycerides,<sup>33,34</sup> phospholipids<sup>34</sup> and glycoproteins,<sup>30</sup> and hepatocellular,<sup>29,30</sup> prostate<sup>33,34</sup> and colorectal<sup>36</sup> cancer risk. Bro *et al.* also observed an inverse association between glyceryl and cholesterol contents and short-term breast cancer risk.<sup>13</sup> Kühn *et al.* observed inverse association with lysophosphatidylcholine (18:0) and direct association with phosphatidylcholine (30:0)—both are glycerol-derived compounds—and breast,

**Table 1.** Baseline characteristics of breast cancer cases and controls,<sup>a</sup> SU.VI.MAX cohort, France (1994–2007)

	Breast cancer cases (N=206)	Controls (N=396)	P <sup>b</sup>
Age at baseline (years)	49.3±6.0	49.1±6.1	0.8
BMI (kg/m <sup>2</sup> )	23.1±3.8	23.4±3.8	0.4
			0.6
<18.5 (underweight)	9 (4.4)	11 (2.8)	
≥18.5 to <25 (normal weight)	147 (71.4)	292 (73.7)	
≥25 (overweight)	50 (24.3)	93 (23.5)	
Height (cm)	162.8±6.3	160.8±5.9	0.0002
<b>Intervention group</b>			1
Placebo	101 (49.0)	194 (49.0)	
Active group: Antioxidants	105 (51.0)	202 (51.0)	
<b>Smoking status</b>			0.8
Never and former	166 (80.6)	322 (81.3)	
Current	40 (19.4)	74 (18.7)	
<b>Physical activity</b>			0.6
Irregular	63 (30.6)	105 (26.5)	
< 1 hour/day walking equivalent	67 (32.5)	135 (34.1)	
≥ 1 hour/day walking equivalent	76 (36.9)	156 (39.4)	
<b>Educational level</b>			0.2
Primary	39 (18.9)	88 (22.2)	
Secondary	81 (39.3)	171 (43.2)	
Superior	86 (41.7)	137 (34.6)	
<b>Number of biological children</b>	1.9±1.2	2.0±1.1	0.4
<b>Hormonal treatment for menopause (yes)</b>	78 (37.9)	152 (38.4)	0.9
<b>Menopausal status at baseline</b>			0.8
Pre-menopausal	122 (59.2)	238 (60.1)	
Post-menopausal	84 (40.8)	158 (39.9)	
<b>Menopausal status at diagnosis</b>			1
Pre-menopausal	68 (33.0)	131 (33.1)	
Post-menopausal	138 (67.0)	265 (66.9)	
<b>Family history of breast cancer<sup>c</sup> (yes)</b>	34 (16.5)	41 (10.4)	0.03
<b>Alcohol intake (g/day)</b>	10.7±11.1	11.0±12.7	0.7
<b>Month of blood draw</b>			1
March–April–May	78 (37.9)	151 (38.1)	
October–November	33 (16.0)	63 (15.9)	
December–January–February	95 (46.1)	182 (46.0)	

SU.VI.MAX, Supplémentation en Vitamines et Minéraux Antioxydants; BMI, body mass index; SD, standard deviation.

<sup>a</sup>Values are means ± SDs or n (%).

<sup>b</sup>P for the comparison between breast cancer cases and controls using  $\chi^2$  tests for categorical variables or Student tests (from analysis of variance models) for continuous variables.

<sup>c</sup>Among first-degree female relatives.

prostate and colorectal cancer risk in a targeted MS metabolomic analysis.<sup>28</sup>

In line with our findings, two NMR metabolomic studies<sup>29,30</sup> found inverse association between acetone serum level and hepatocellular cancer risk.

Results of metabolomic studies are more contrasted regarding amino acids. For instance, glutamine,<sup>30,31</sup> lysine,<sup>31,33</sup> valine<sup>30</sup> and arginine<sup>32</sup> were inversely associated with hepatocellular<sup>30,31</sup> and prostate<sup>32,33</sup> cancer risk in some prospective metabolomic studies. But arginine was also associated with increased aggressive prostate cancer risk in another study based on untargeted MS metabolomics.<sup>34</sup> Valine is one of the branched-chain amino acids

(BCAAs) previously suggested as novel markers of cardiovascular disease development,<sup>39</sup> insulin resistance and the future development of diabetes.<sup>40</sup>

Similarly, findings were mitigated for creatine and creatinine in two prospective NMR studies. Bro *et al.* observed an inverse association between a NMR signal corresponding to a spectral overlap of creatine and creatinine and short-term breast cancer risk, but they also observed a positive association with a signal related to creatine alone.<sup>13</sup> Assi *et al.* reported an inverse association between creatinine and hepatocellular carcinoma risk.<sup>29</sup> However, these comparisons of results should be interpreted with caution, since the aetiology may vary according to cancer locations.

**Table 2.** NMR metabolites associated with breast cancer risk from multivariable conditional logistic regression,<sup>a</sup> SU.VI.MAX cohort, France (1994–2007)

Assignments	Chemical shift (ppm)	T1 (ref.)		T2		T3		P-trend
		Cases/controls	OR	Cases/controls	OR (95% CI)	Cases/controls	OR (95% CI)	
Metabolites highlighted with NOESY1D experiments								
Lipoproteins (CH <sub>3</sub> )	0.86	84/114	1.00	60/141	0.52 (0.33–0.82)	62/141	0.56 (0.36–0.88)	0.01
Lipoproteins (CH <sub>2</sub> )	1.27	84/117	1.00	67/133	0.69 (0.44–1.07)	55/146	0.52 (0.33–0.82)	0.004
Lipids (CH <sub>2</sub> CH <sub>2</sub> CO)	1.56	84/117	1.00	72/128	0.7 (0.45–1.09)	50/151	0.43 (0.27–0.68)	0.0004
Undefined region	1.83–1.85	78/120	1.00	67/133	0.72 (0.46–1.12)	61/143	0.55 (0.35–0.89)	0.01
Lipids (CCH <sub>2</sub> ) and Glycoproteins (N-acetyl) <sup>b</sup>	2.001 and 2.035	79/121	1.00	72/124	0.86 (0.55–1.33)	55/151	0.55 (0.35–0.87)	0.01
Lipids (CH <sub>2</sub> CO) and Acetone <sup>b</sup>	2.22	86/113	1.00	69/135	0.62 (0.4–0.95)	51/148	0.43 (0.27–0.69)	0.0004
Glycerol-derived compounds (CH <sub>2</sub> )	4.04 and 4.25	90/113	1.00	60/141	0.54 (0.35–0.83)	56/142	0.49 (0.32–0.77)	0.001
Glycerol-derived compounds (CH)	5.18	88/111	1.00	68/132	0.6 (0.38–0.94)	50/153	0.37 (0.23–0.61)	0.00007
Unsaturated lipids (CH=CH)	5.28	80/122	1.00	68/129	0.79 (0.51–1.22)	58/145	0.6 (0.38–0.94)	0.03
Metabolites highlighted with GPMG experiments								
Assignments								
Chemical shift (ppm)								
		T1 (ref.)		T2		T3		P-trend
		Cases/controls	OR	Cases/controls	OR (95% CI)	Cases/controls	OR (95% CI)	
Valine	0.98 and 1.03	53/134	1.00	69/129	1.35 (0.86–2.11)	81/115	1.83 (1.15–2.92)	0.01
Lipoproteins (CH <sub>2</sub> )	1.26	78/114	1.00	70/128	0.83 (0.53–1.29)	55/136	0.58 (0.36–0.92)	0.02
Lysine and Arginine <sup>b</sup>	1.70	57/139	1.00	63/121	1.19 (0.77–1.85)	83/118	1.62 (1.05–2.48)	0.03
Lipids (CH <sub>2</sub> CO) and Acetone <sup>b</sup>	2.22	72/123	1.00	80/110	1.16 (0.76–1.78)	51/145	0.58 (0.36–0.93)	0.03
Unassigned	2.33	54/140	1.00	72/118	1.58 (1–2.51)	77/120	1.67 (1.08–2.6)	0.02
Glutamine	2.45	59/131	1.00	64/129	1.17 (0.74–1.84)	80/118	1.61 (1.02–2.55)	0.04
Undefined region	2.55–2.57	52/140	1.00	77/115	1.74 (1.11–2.74)	74/123	1.63 (1.04–2.55)	0.04
Undefined region	2.57–2.61	56/131	1.00	65/133	1.11 (0.71–1.76)	82/114	1.79 (1.15–2.8)	0.009
Lysine, Creatine and Creatinine <sup>b</sup>	3.01 and 3.03	52/141	1.00	70/120	1.64 (1.02–2.63)	81/117	1.84 (1.19–2.85)	0.007
Unassigned	3.34	56/140	1.00	72/115	1.72 (1.1–2.69)	75/123	1.59 (1.01–2.5)	0.04
Glucose	3.36–3.86	57/137	1.00	66/124	1.17 (0.73–1.86)	80/117	1.61 (1.04–2.5)	0.03
Unassigned	5.18	81/112	1.00	64/133	0.6 (0.38–0.94)	58/133	0.61 (0.39–0.97)	0.04
Undefined region	5.38–5.43	81/115	1.00	61/129	0.68 (0.43–1.07)	61/134	0.63 (0.41–0.98)	0.046

SU.VI.MAX, Supplémentation en Vitamines et Minéraux Antioxydants; T, tertile; ref., reference; OR, odds ratio; CI, confidence interval; NRI, Net Reclassification Improvement.

<sup>a</sup>Multivariable models were additionally adjusted for age (continuous), BMI (continuous), menopausal status (pre/post-menopause), intervention group of the SU.VI.MAX trial (placebo/supplemented), smoking status (current smokers and non-smokers), season of blood draw (defined above), height (continuous), alcohol intake (continuous), physical activity (irregular/<1 hour/day walking equivalent/≥1 hour/day walking equivalent), education level (primary/secondary/superior), family history of breast cancer (yes/no), number of children (continuous) and use of hormone replacement therapy for menopause (yes/no) (reference model). Tests for linear trend were performed using the ordinal score on tertiles of buckets intensity.

<sup>b</sup>Multiple assignments for the same NMR variable due to spectral overlap.

**Table 3.** Associations between the metabolomic patterns derived by PCA and breast cancer risk from multivariable conditional logistic regression,<sup>a</sup> SU.VI.MAX cohort, France (1994–2007)

Components	T1 (ref.)		T2		T3		P-trend
	Cases/controls	OR	Cases/controls	OR (95% CI)	Cases/controls	OR (95% CI)	
Component 1 <sup>b</sup>	82/111	1.00	74/120	0.77 (0.48–1.22)	47/146	0.45 (0.28–0.72)	0.0008
Component 2 <sup>c</sup>	67/126	1.00	65/129	0.93 (0.61–1.42)	71/122	1.07 (0.69–1.66)	0.8

SU.VI.MAX, Supplémentation en Vitamines et Minéraux Antioxydants; T, tertile; ref., reference; OR, odds ratio; CI, confidence interval; PCA, principal component analysis.

<sup>a</sup>Multivariable models were additionally adjusted for age (continuous), BMI (continuous), menopausal status (pre-/post-menopause), intervention group of the SU.VI.MAX trial (placebo/supplemented), smoking status (current smokers and non-smokers), season of blood draw (defined above), height (continuous), alcohol intake (continuous), physical activity (irregular/<1 hour/day walking equivalent/≥1 hour/day walking equivalent), education level (primary/secondary/superior), family history of breast cancer (yes/no), number of children (continuous) and use of hormone replacement therapy for menopause (yes/no) (reference model). Tests for linear trend were performed using the ordinal score on tertiles of components from PCA.

<sup>b</sup>Component 1 corresponds to the first metabolomic pattern and was characterized by elevated plasma levels of lipoproteins, lipids, acetone, unsaturated lipids, glycerol-derived compounds and glycoproteins (loading values ≥0.15) and lower plasma levels of glutamine, lysine, creatine and creatinine (loading value ≤−0.15).

<sup>c</sup>Component 2 corresponds to the second metabolomic pattern and was characterized by elevated plasma levels of valine, lipoproteins, lysine, arginine, lipids, acetone, glutamine, creatine, creatinine and glucose (loading values ≥0.15).

In line with our findings concerning glucose level, a meta-analysis based on prospective (non-metabolomic) studies showed a positive, although weak, association between fasting serum glucose levels and the risk of breast cancer,<sup>41</sup> echoing a suggested role of chronic hyperinsulinemia and insulin resistance in breast cancer aetiology.<sup>42</sup>

Our study provides interesting insights into potential early mechanistic pathways involved in breast carcinogenesis.

First, these results suggest a disruption of energetic homeostasis and amino acids metabolism in women who will develop a breast cancer. The decrease in plasmatic levels of glycerol-derived compounds, fatty acids and lipoproteins may reflect a decrease in inter-organ lipid exchange and an increase in beta-oxidation.<sup>43–45</sup> The latter is supported by the increase in plasmatic level of lysine, a carnitine precursor involved in fatty acids transportation into mitochondrion for beta-oxidation. The increase in lysine content—a ketogenic amino acid—also suggests an increase in ketogenesis. Lipids and ketone bodies are alternative energy providers replacing glucose metabolism, which may partly explain the increase in plasma glucose level. Furthermore, this increase in glucose could also result from gluconeogenesis as revealed by the increase in glutamine and valine, two glucogenic amino acids. Altogether, these metabolic changes may reflect a setting-up of insulin resistance.

Second, the increase in valine, glutamine, creatine and arginine plasma levels in women who will develop a breast cancer could participate to the establishment of favourable conditions for cell proliferation. Indeed, valine and glutamine are activators of the mTOR pathway (one of the main cell proliferation pathways), in particular mTORC1 (which is breast-cancer-specific).<sup>40</sup> Creatine, considered as a methyl donor, may play a role in the oncoprotein PP2A methylation that activates cell proliferation and neoplastic transformation pathways.<sup>46</sup> The cell proliferation is also favoured by the

production of citrate from the glutamine pathway to support fatty acids production contributing to membrane synthesis.<sup>40</sup>

Third, the increase in plasma glutamine level in subjects who will be affected by breast cancer may indicate a decrease in antioxidant defences.<sup>47,48</sup> It has also been reported that elevated glutamine ties in with changes in cancer cell metabolism.<sup>47</sup>

Other interpretations of these metabolic variations may also be considered given that these metabolites participate in many metabolic pathways.

Strengths of this study pertained to the NMR metabolomic approach (from two complementary experiments) combined with a prospective design and long follow-up. Besides, most of the results were unchanged after the exclusion of cases diagnosed during the first 2 years of follow-up, showing the robustness of the results and the fact that the observed associations were not likely to entirely reflect the presence of an already existing tumour.

Nevertheless, limitations should be acknowledged. First, although the number of cases was appropriate for the main analysis, it may have limited our ability to detect some associations, especially in stratified analyses. Second, chance findings due to multiple testing as well as the risk of overfitting of the models cannot be entirely excluded. Thus, replication of this study in independent prospective cohorts (as planned in the NutriNet-Santé cohort<sup>49</sup>) in different countries and settings will definitely strengthen the results and constitutes an important research perspective of the present study. Third, the precision of buckets annotation was limited for some metabolites, especially for lipids. For instance, CH<sub>2</sub>CH<sub>2</sub>CO corresponds to a chemical function in the aliphatic chain of all fatty acids esterifying glycerol. Because the protons of these groups resonate at approximately the same frequency for every fatty acid, it is impossible to differentiate among the different esterified

**Table 4.** Predictive performances of breast cancer risk models including single metabolites or metabolomic patterns assessed by AUCs (Area Under the ROC Curves) and NRI (Net Reclassification Improvement) indexes

Metabolites highlighted with NOESY1D experiments						
Assignments	Chemical shift (ppm)	AUC <sup>a</sup>	95% CI for difference in AUC	P-value for AUC difference <sup>b</sup>	NRI <sup>c</sup> (95% CI)	NRI p-value <sup>d</sup>
Lipoproteins (CH <sub>3</sub> )	0.86	0.68	(-0.01-0.04)	0.3	0.19 (0.03-0.36)	0.02
Lipoproteins (CH <sub>2</sub> )	1.27	0.69	(0-0.05)	0.1	0.22 (0.05-0.39)	0.01
Lipids (CH <sub>2</sub> CH <sub>2</sub> CO)	1.56	0.69	(0-0.06)	0.048	0.29 (0.12-0.46)	0.0007
Undefined region	1.83-1.85	0.68	(-0.01-0.03)	0.3	0.26 (0.1-0.43)	0.002
Lipids (CCH <sub>2</sub> ) and Glycoproteins (N-acetyl)	2.001 and 2.035	0.68	(-0.01-0.04)	0.3	0.27 (0.1-0.44)	0.002
Lipids (CH <sub>2</sub> CO) and Acetone	2.22	0.69	(0-0.06)	0.07	0.38 (0.21-0.54)	0.00001
Glycerol-derived compounds (CH <sub>2</sub> )	4.04 and 4.25	0.69	(0-0.05)	0.08	0.41 (0.25-0.57)	0.000002
Glycerol-derived compounds (CH)	5.18	0.7	(0-0.06)	0.04	0.39 (0.23-0.56)	0.000004
Unsaturated lipids (CH=CH)	5.28	0.67	(-0.01-0.03)	0.4	0.17 (0-0.34)	0.05
Metabolites highlighted with CPMG experiments						
Assignments	Chemical shift (ppm)	AUC <sup>a</sup>	95% CI for difference in AUC	P-value for AUC difference <sup>b</sup>	NRI <sup>c</sup> (95% CI)	NRI p-value <sup>d</sup>
Valine	0.98 and 1.03	0.68	(0-0.04)	0.08	0.27 (0.1-0.44)	0.002
Lipoproteins (CH <sub>2</sub> )	1.26	0.68	(-0.01-0.03)	0.2	0.21 (0.04-0.38)	0.01
Lysine and Arginine	1.70	0.68	(-0.01-0.03)	0.2	0.22 (0.05-0.39)	0.01
Lipids (CH <sub>2</sub> CO) and Acetone	2.22	0.67	(-0.01-0.03)	0.4	0.21 (0.04-0.38)	0.01
Unassigned	2.33	0.68	(0-0.04)	0.1	0.26 (0.09-0.43)	0.003
Glutamine	2.45	0.68	(-0.01-0.03)	0.2	0.14 (-0.03-0.31)	0.1
Undefined region	2.55-2.57	0.68	(0-0.04)	0.1	0.18 (0.01-0.35)	0.04
Undefined region	2.57-2.61	0.68	(0-0.04)	0.1	0.33 (0.16-0.5)	0.0001
Lysine, Creatine and Creatinine	3.01 and 3.03	0.68	(0-0.05)	0.09	0.21 (0.04-0.38)	0.02
Unassigned	3.34	0.67	(-0.01-0.03)	0.3	0.21 (0.04-0.38)	0.02
Glucose	3.36-3.86	0.67	(-0.01-0.03)	0.2	0.26 (0.09-0.43)	0.003
Unassigned	5.18	0.67	(-0.01-0.03)	0.5	0.29 (0.12-0.46)	0.001
Undefined region	5.38-5.43	0.67	(-0.01-0.03)	0.2	0.22 (0.05-0.39)	0.01
Metabolomic patterns derived by PCA						
Components		AUC <sup>a</sup>	95% CI for difference in AUC	P-value for AUC difference <sup>b</sup>	NRI <sup>c</sup> (95% CI)	NRI p-value <sup>d</sup>
Component 1		0.69	(0-0.06)	0.09	0.31 (0.14-0.48)	0.0004
Component 2		0.66	(0-0)	0.5	0.03 (-0.14-0.2)	0.7

AUC, area under the curve; CI, confidence interval; NRI, Net Reclassification Improvement.

<sup>a</sup>AUCs after inclusion of single metabolites (single NMR variables) or metabolomic patterns (components from PCA) into the reference model based on known risk factors. AUCs before inclusion of metabolomic variables was 0.66.

<sup>b</sup>P-value for AUC difference between before and after inclusion of single metabolites or metabolomic patterns into the reference model.

<sup>c</sup>Net Reclassification Improvement (NRI) assesses the improvement in the predictive performance of breast cancer risk logistic models when single metabolites or metabolomic patterns were added, compared with the reference model based on known risk factors.

<sup>d</sup>P-value for the Net Reclassification Improvement between before and after inclusion of single metabolites or metabolomic patterns into the reference model.

fatty acids present in the sample.<sup>50</sup> Mechanistic interpretation regarding metabolites for which the exact nature could not be determined should be considered with caution. A similar future study using MS analyses (more sensitive) is ongoing in the SU.VI.MAX cohort and will allow

us to strengthen and precise these results. Next, this metabolomic analysis is based on a single blood draw. Thus, the intra-individual variability of metabolomic profile over time was not controlled in this study. Nevertheless, several studies showed a good stability and reproducibility of

metabolomic measurements for most of metabolites.<sup>51,52</sup> Finally, a quantification analysis of plasma levels of key identified metabolites will be necessary to propose potential predictive biomarkers for cancer prevention.

In conclusion, this prospective study suggested that baseline NMR plasma metabolomic signatures may reveal predictive metabolites associated with long-term breast cancer risk. These results constitute interesting preliminary findings that should be replicated in other independent prospective studies, to allow the identification of strong biomarkers of breast cancer risk. Nevertheless, these results provide interesting insights to better understand complex mechanisms involved in breast carcinogenesis. Discriminant metabolites evoke plasma metabolic disorders favourable for carcinogenesis initiation preceding tumour onset, such as disruption in energetic metabolism, oxidative stress and pre-proliferation conditions. This study may contribute to develop screening strategies for the identification of at-risk women for breast cancer well before symptoms appear. In the shorter term, correlation studies between metabolomic profiles and detailed dietary and lifestyle patterns are underway in the same dataset. This will allow us to identify which of the breast cancer risk discriminant metabolites are also associated with modifiable risk factors, the latter being particularly relevant to improve prevention strategies.

### Supplementary Data

Supplementary data are available at *IJE* online.

### Funding

This work was supported by the French National Cancer Institute (grant number INCa\_8085 for the project, and PhD grant number INCa\_11323 for L.L.); the Federative Institute for Biomedical Research IFRB Paris 13; and the Cancéropôle Ile-de-France/Région Ile de France (PhD grant for M.D.). The funders had no role in the design, analysis or writing of this article.

### Acknowledgements

The authors thank Nathalie Arnault, Younes Esseddik, Paul Flanzy, Yasmina Chelghoum and Van Than Duong (computer scientists) and Véronique Gourlet, Fabien Szabo and Laurent Bourhis (statisticians) for their technical contribution to the SU.VI.MAX study. We also thank all the volunteers of the SU.VI.MAX cohort. This work was conducted in the framework of the French network for Nutrition And Cancer Research (NACRe network [www.inra.fr/nacre](http://www.inra.fr/nacre)) and received the Partnership Label of the NACRe network. The authors' responsibilities were as follow—L.L. and M.T.: designed the research; S.H., P.G., M.T.: conducted the research; N.B., A.V.B.: acquired NMR spectra; L.L., B.S., A.V.B.: performed statistical analysis; M.T., M.D.: supervised statistical analysis; L.L., M.T.: wrote the paper; L.L., A.V.B., M.D., N.B., M.N.T., M.P.V., A.R., A.D., P.G., S.H., V.P., L.L.M., B.S., T.F., P.L.M., E.K.G., P.S., M.T.: contributed to the data

interpretation and revised each draft for important intellectual content; and M.T. had primary responsibility for final content. All authors read and approved the final manuscript.

**Conflict of interest:** The authors have no conflicts of interest to declare.

### References

1. Ferlay J, Soerjomataram I, Ervik M *et al.* *GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11*. Lyon, France: International Agency for Research on Cancer, 2013.
2. Howell A, Anderson AS, Clarke RB *et al.* Risk determination and prevention of breast cancer. *Breast Cancer Res* 2014;**16**:446.
3. Dunn WB, Goodacre R, Neyses L, Mamas M. Integration of metabolomics in heart disease and diabetes research: current achievements and future outlook. *Bioanalysis* 2011;**3**:2205–22.
4. Suhre K, Shin SY, Petersen AK *et al.* Human metabolic individuality in biomedical and pharmaceutical research. *Nature* 2011;**477**:54–60.
5. Bictash M, Ebbels TM, Chan Q *et al.* Opening up the 'Black Box': metabolic phenotyping and metabolome-wide association studies in epidemiology. *J Clin Epidemiol* 2010;**63**:970–79.
6. Norvig P, Relman DA, Goldstein DB *et al.* 2020 visions. *Nature* 2010;**463**:26–32.
7. Holmes E, Loo RL, Stampler J *et al.* Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature* 2008;**453**:396–400.
8. Claudino WM, Quattrone A, Biganzoli L, Pestrin M, Bertini I, Di LA. Metabolomics: available results, current research projects in breast cancer, and future applications. *J Clin Oncol* 2007;**25**:2840–46.
9. Denkert C, Bucher E, Hilvo M *et al.* Metabolomics of human breast cancer: new approaches for tumor typing and biomarker discovery. *Genome Med* 2012;**4**:37.
10. Budczies J, Denkert C, Muller BM *et al.* Remodeling of central metabolism in invasive breast cancer compared to normal breast tissue—a GC-TOFMS based metabolomics study. *BMC Genomics* 2012;**13**:334.
11. Hennes C, Bullinger D, Fux R *et al.* Prediction of breast cancer by profiling of urinary RNA metabolites using Support Vector Machine-based feature selection. *BMC Cancer* 2009;**9**:104.
12. Brauer HA, Makowski L, Hoadley KA *et al.* Impact of tumor microenvironment and epithelial phenotypes on metabolism in breast cancer. *Clin Cancer Res* 2013;**19**:571–85.
13. Bro R, Kamstrup-Nielsen MH, Engelsen SB *et al.* Forecasting individual breast cancer risk using plasma metabolomics and biocountours. *Metabolomics* 15 A.D. Mar 10.
14. Playdon MC, Ziegler RG, Sampson JN *et al.* Nutritional metabolomics and breast cancer risk in a prospective study. *Am J Clin Nutr* 2017;**106**:637–49.
15. Kruk J, Doskocz M, Jodłowska E *et al.* NMR techniques in metabolomic studies: a quick overview on examples of utilization. *Appl Magn Reson* 2017;**48**:1–21.
16. Hercberg S, Galan P, Preziosi P *et al.* The SU.VI.MAX study: a randomized, placebo-controlled trial of the health effects of antioxidant vitamins and minerals. *Arch Intern Med* 2004;**164**:2335–42.

17. Hercberg S, Preziosi P, Briancon S *et al.* A primary prevention trial using nutritional doses of antioxidant vitamins and minerals in cardiovascular diseases and cancers in a general population: the SU.VI.MAX study—design, methods, and participant characteristics. *Supplementation en Vitamines et Minéraux AntioXydants. Control Clin Trials* 1998;19:336–51.
18. WHO. *ICD-10, International Classification of Diseases and Related Health Problems*, 10th revision. Geneva: World Health Organization, 1993.
19. Giraudeau P, Silvestre V, Akoka S. Optimizing water suppression for quantitative NMR-based metabolomics: a tutorial review. *Metabolomics* 2015;11:1041–55.
20. Carr H, Purcell E. Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Phys Rev* 1954;94:630–38.
21. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 1995;6:277–93.
22. Sousa SA, Magalhaes A, Ferreira MM. Optimized bucketing for NMR spectra: three case studies. *Chemometr Intell Lab Syst* 2013;122:93–102.
23. Wishart DS, Jewison T, Guo AC *et al.* HMDB 3.0—The Human Metabolome Database in 2013. *Nucleic Acids Res* 2013; 41(Database issue):D801–07.
24. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc B* 1995;57:289–300.
25. Cattell RB. The scree test for the number of factors. *Multivariate Behav Res* 1966;1:245–76.
26. Kim J-O, Mueller C. *Factor Analysis: Statistical Methods and Practical Issues*. Newbury Park: SAGE Publications, Inc., 1978.
27. Pencina MJ, D'Agostino RB, Sr, D'Agostino RB, Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008;27:157–72.
28. Kuhn T, Floegel A, Sookthai D *et al.* Higher plasma levels of lysophosphatidylcholine 18:0 are related to a lower risk of common cancers in a prospective metabolomics study. *BMC Med* 2016;14:13.
29. Assi N, Fages A, Vineis P *et al.* A statistical framework to model the meeting-in-the-middle principle using metabolomic data: application to hepatocellular carcinoma in the EPIC study. *Mutagenesis* 2015;30:743–53.
30. Fages A, Duarte-Salles T, Stepien M *et al.* Metabolomic profiles of hepatocellular carcinoma in a European prospective cohort. *BMC Med* 2015;13:242.
31. Stepien M, Duarte-Salles T, Fedirko V *et al.* Alteration of amino acid and biogenic amine metabolism in hepatobiliary cancers: findings from a prospective cohort study. *Int J Cancer* 2016;138:348–60.
32. Huang J, Mondul AM, Weinstein SJ *et al.* Serum metabolomic profiling of prostate cancer risk in the prostate, lung, colorectal, and ovarian cancer screening trial. *Br J Cancer* 2016;115:1087–95.
33. Mondul AM, Moore SC, Weinstein SJ, Mannisto S, Sampson JN, Albanes D. 1-stearoylglycerol is associated with risk of prostate cancer: results from serum metabolomic profiling. *Metabolomics* 2014;10:1036–41.
34. Mondul AM, Moore SC, Weinstein SJ, Karoly ED, Sampson JN, Albanes D. Metabolomic analysis of prostate cancer risk in a prospective cohort: the alpha-tocopherol, beta-carotene cancer prevention (ATBC) study. *Int J Cancer* 2015;137:2124–32.
35. Bjorkblom B, Wibom C, Jonsson P *et al.* Metabolomic screening of pre-diagnostic serum samples identifies association between alpha- and gamma-tocopherols and glioblastoma risk. *Oncotarget* 2016;7:37043–53.
36. Cross AJ, Moore SC, Boca S *et al.* A prospective study of serum metabolites and colorectal cancer risk. *Cancer* 2014;120:3049–57.
37. Guertin KA, Loftfield E, Boca SM *et al.* Serum biomarkers of habitual coffee consumption may provide insight into the mechanism underlying the association between coffee consumption and colorectal cancer. *Am J Clin Nutr* 2015;101:1000–11.
38. Wikoff WR, Hanash S, DeFelice B *et al.* Diacetylspermine is a novel prediagnostic serum biomarker for non-small-cell lung cancer and has additive performance with pro-surfactant protein B. *J Clin Oncol* 2015;33:3880–86.
39. Magnusson M, Lewis GD, Ericson U *et al.* A diabetes-predictive amino acid score and future cardiovascular disease. *Eur Heart J* 2013;34:1982–89.
40. O'Connell TM. The complex role of branched chain amino acids in diabetes and cancer. *Metabolites* 2013;3:931–45.
41. Boyle P, Koechlin A, Pizot C *et al.* Blood glucose concentrations and breast cancer risk in women without diabetes: a meta-analysis. *Eur J Nutr* 2013;52:1533–40.
42. Mullie P, Koechlin A, Boniol M, Autier P, Boyle P. Relation between breast cancer and high glycemic index or glycemic load: a meta-analysis of prospective cohort studies. *Crit Rev Food Sci Nutr* 2016;56:152–59.
43. Rohrig F, Schulze A. The multifaceted roles of fatty acid synthesis in cancer. *Nat Rev Cancer* 2016;16:732–49.
44. His M, Zelek L, Deschasaux M *et al.* Prospective associations between serum biomarkers of lipid metabolism and overall, breast and prostate cancer risk. *Eur J Epidemiol* 2014;29:119–32.
45. Gunther UL. Metabolomics biomarkers for breast cancer. *Pathobiology* 2015;82:153–65.
46. Guenin S, Schwartz L, Morvan D *et al.* PP2A activity is controlled by methylation and regulates oncoprotein expression in melanoma cells: a mechanism which participates in growth inhibition induced by chloroethylnitrosourea treatment. *Int J Oncol* 2008;32:49–57.
47. Jerby L, Wolf L, Denkert C *et al.* Metabolic associations of reduced proliferation and oxidative stress in advanced breast cancer. *Cancer Res* 2012;72:5712–20.
48. Levy PL, Duponchel S, Eischeid H *et al.* Hepatitis C virus infection triggers a tumor-like glutamine metabolism. *Hepatology* 2017;65: 789–803.
49. Hercberg S, Castetbon K, Czernichow S *et al.* The Nutrinet-Sante Study: a web-based prospective study on the relationship between nutrition and health and determinants of dietary patterns and nutritional status. *BMC Public Health* 2010;10:242.
50. Oostendorp M, Engelke UF, Willemsen MA, Wevers RA. Diagnosing inborn errors of lipid metabolism with proton nuclear magnetic resonance spectroscopy. *Clin Chem* 2006;52:1395–405.
51. Carayol M, Licaj I, Achaintre D *et al.* Reliability of serum metabolites over a two-year period: a targeted metabolomic approach in fasting and non-fasting samples from EPIC. *PLoS One* 2015; 10:e0135437.
52. Floegel A, Drogan D, Wang-Sattler R *et al.* Reliability of serum metabolite concentrations over a 4-month period using a targeted metabolomic approach. *PLoS One* 2011;6:e21103.

*Additional File 1 – L. Lecuyer et al. NMR metabolomic signatures reveal predictive plasma metabolites associated with long-term risk of developing breast cancer*

### Supplementary methods for NMR metabolomic analysis

#### Sample preparation and spectral acquisition

Due to the prospective design, when the samples were stored in liquid nitrogen, we did not know who were the future cases and controls, thus all samples were treated in the same way. All samples were extracted the same day, by the same technician. The technician who extracted the sample and the one who performed NMR spectra acquisition did not have access to the information regarding case/control status of the subjects. No thawing was carried out previously, since the samples were aliquoted in several plastic straws of 500 microliters per subject, immediately after blood draw and before storage in liquid nitrogen. Randomization was carried out (cases and controls mixed together at random) for the positioning in batches. A volume of 0.35 mL of defrosted plasma was placed into a 5- mm diameter-specific NMR tube with deuterated solution of PBS (0.25 mL) with NaN<sub>3</sub>, permitted the use of deuterium frequency-field lock. Because of the number of samples, a strict protocol was followed. Automation of recording of NMR spectra was performed. The proton spectra were acquired at 500 MHz on a Bruker AVANCE III spectrometer at 300 K. The spectral window was 10 kHz. The water signal was suppressed by a presaturation sequence using low-power irradiation on the water-signal frequency during the relaxation delay. The NOESY1D experiments (a combination of both macromolecules and other metabolite signals) and the CPMG experiments (suppressing the broad signal of macromolecules according to their short relaxation time) were used with 128 transients. Each signal on the spectrum corresponds to a different proton chemical environment. One metabolite can therefore produce several signals with several chemical shifts according to the number of different chemical group containing protons in the molecule.

#### Spectra processing

A Fourier transformation was applied on NMR data with an exponential window function to produce a 0.3 Hz broadening line. Spectra were phased and a baseline correction was applied with 10 points (from 0 to 9 ppm). All spectra were aligned on glucose doublet at 5.23 ppm. Spectra were normalized to the sum of intensities. These processing were carried out using NMRPipe software (1).

#### REFERENCES

- (1) Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 1995;6(3):277-93.

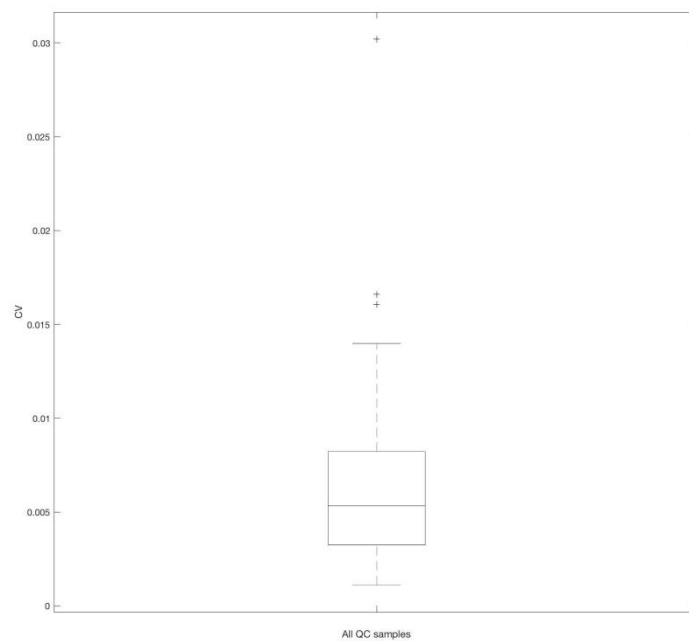


*Additional File 2 – L. Lecuyer et al. NMR metabolomic signatures reveal predictive plasma metabolites associated with long-term risk of developing breast cancer*

### Quality controls

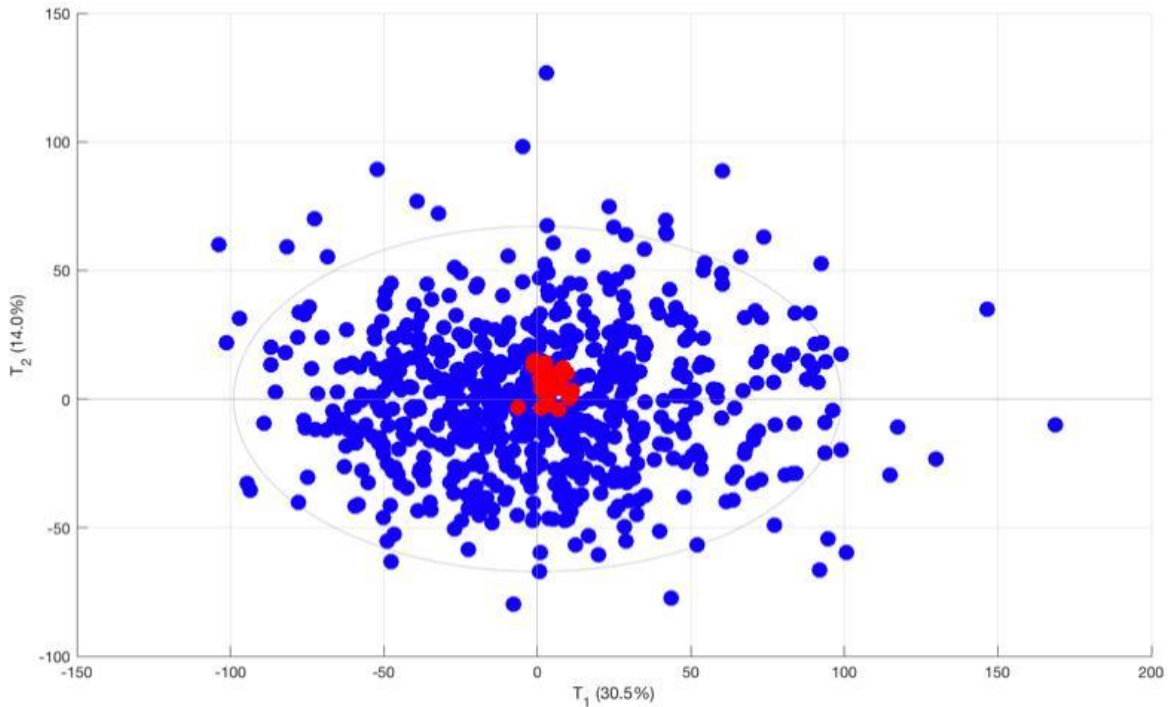
Quality control (QC) samples (from 39 plasma samples of subjects from the SU.VI.MAX cohort, pooled together) were recorded every 20 samples to ensure reproducibility testing of the NMR data acquisition.

**Figure 1s** Boxplot of the distribution of the coefficients of variation across the buckets of quality control samples



The **Figure 1s** represents the CV distribution's boxplot across the buckets after the "intelligent bucketing" procedure of QC samples (obtained from 39 plasma samples of subjects from the SU.VI.MAX cohort, pooled together) which were recorded every 20 samples.

**Figure 2s** Principal Component Analysis score plot of spectra from all samples included in the NMR analyses (samples from the nested case-control study and from the quality control)



The **Figure 2s** represents the score plot of a Principal Component Analysis of all samples included in the NMR analyses. The spectra corresponding to samples from the nested case-control study are represented by blue dots and the spectra corresponding to samples from the QC are represented by red dots. All the spectra corresponding to the QC samples were located in the same area.

*Additional File 3 – L. Lecuyer et al. NMR metabolomic signatures reveal predictive plasma metabolites associated with long-term risk of developing breast cancer*

### Supplementary Spearman correlation matrices of the metabolites for NOESY1D and CPMG experiments

**Supplementary Table 1** Spearman correlation matrix of the metabolites highlighted with NOESY1D experiments (correlation coefficient [95%CI])

<u>Metabolites [ppm]</u>	Unsaturated lipids (CH=CH) [5.28]	Glycerol-derived compounds (CH) [5.18]	Lipids (CH <sub>2</sub> CO) & Acetone [2.22]	Lipids (CCH <sub>2</sub> ) & Glycoproteins (N-acetyl) [2.001] & [2.035]	Undefined region [1.83-1.85]	Lipoproteins (CH <sub>3</sub> ) [0.86]	Lipoproteins (CH <sub>2</sub> ) [1.27]	Lipids (CH <sub>2</sub> CH <sub>2</sub> CO) [1.56]
Glycerol-derived compounds (CH) [5.18]	0.77 [0.74-0.8]							
Lipids (CH <sub>2</sub> CO) & Acetone [2.22]	0.8 [0.77-0.83]	0.9 [0.89-0.92]						
Lipids (CCH <sub>2</sub> ) & Glycoproteins (N-acetyl) [2.001] & [2.035]	0.92 [0.9-0.93]	0.86 [0.84-0.88]	0.89 [0.87-0.9]					
Undefined region [1.83-1.85]	0.54 [0.48-0.6]	0.25 [0.17-0.32]	0.4 [0.33-0.47]	0.52 [0.46-0.57]				
Lipoproteins (CH <sub>3</sub> ) [0.86]	0.93 [0.92-0.94]	0.73 [0.69-0.77]	0.76 [0.72-0.79]	0.91 [0.89-0.92]	0.66 [0.61-0.7]			
Lipoproteins (CH <sub>2</sub> ) [1.27]	0.9 [0.89-0.92]	0.85 [0.82-0.87]	0.92 [0.91-0.93]	0.94 [0.93-0.95]	0.5 [0.44-0.56]	0.92 [0.9-0.93]		
Lipids (CH <sub>2</sub> CH <sub>2</sub> CO) [1.56]	0.81 [0.78-0.84]	0.91 [0.89-0.92]	0.97 [0.96-0.97]	0.89 [0.88-0.91]	0.5 [0.43-0.55]	0.81 [0.78-0.84]	0.92 [0.91-0.93]	
Glycerol-derived compounds (CH <sub>2</sub> ) [4.04 & 4.25]	0.67 [0.63-0.71]	0.73 [0.69-0.77]	0.9 [0.88-0.91]	0.79 [0.76-0.82]	0.42 [0.35-0.48]	0.64 [0.58-0.68]	0.82 [0.79-0.85]	0.84 [0.82-0.86]

**Supplementary Table 2** Spearman correlation matrix of the metabolites highlighted with CPMG experiments (correlation coefficient [95% CI])

<b>Metabolites [ppm]</b>	<b>Valine [0.98] &amp; [1.03]</b>	<b>Lipoproteins (CH<sub>2</sub>) [1.26]</b>	<b>Lysine &amp; Arginine [1.7]</b>	<b>Lipids (CH<sub>2</sub>CO) &amp; Acetone [2.22]</b>	<b>Unassigned [2.33]</b>	<b>Glutamine [2.45]</b>	<b>Undefined region [2.55-2.57]</b>	<b>Undefined region [2.57-2.61]</b>	<b>Lysine, Creatine &amp; Creatinine [3.01] &amp; [3.03]</b>	<b>Unassigned [3.34]</b>	<b>Glucose [3.36-3.86]</b>	<b>Unassigned [5.18]</b>
<b>Lipoproteins (CH<sub>2</sub>) [1.26]</b>	0.29 [0.22-0.36]											
<b>Lysine &amp; Arginine [1.7]</b>	0.74 [0.7-0.78]	0.2 [0.12-0.28]										
<b>Lipids (CH<sub>2</sub>CO) &amp; Acetone [2.22]</b>	0.33 [0.25-0.4]	0.89 [0.87-0.9]	0.34 [0.26-0.41]									
<b>Unassigned [2.33]</b>	0.7 [0.65-0.74]	0.19 [0.11-0.27]	0.85 [0.83-0.87]	0.29 [0.21-0.36]								
<b>Glutamine [2.45]</b>	0.45 [0.38-0.51]	-0.04 [-0.12-0.04]	0.71 [0.67-0.75]	0.07 [-0.01-0.15]	0.56 [0.5-0.62]							
<b>Undefined region [2.55-2.57]</b>	0.61 [0.55-0.66]	0.16 [0.08-0.24]	0.82 [0.79-0.84]	0.25 [0.17-0.32]	0.82 [0.8-0.85]	0.71 [0.67-0.75]						
<b>Undefined region [2.57-2.61]</b>	0.45 [0.39-0.52]	0.14 [0.06-0.22]	0.64 [0.58-0.68]	0.23 [0.15-0.31]	0.54 [0.48-0.59]	0.73 [0.69-0.76]	0.76 [0.72-0.79]					
<b>Lysine, Creatine &amp; Creatinine [3.01] &amp; [3.03]</b>	0.58 [0.52-0.63]	-0.09 [-0.17--0.01]	0.88 [0.86-0.9]	0.06 [-0.02-0.15]	0.8 [0.77-0.83]	0.79 [0.75-0.81]	0.83 [0.8-0.85]	0.65 [0.6-0.7]				
<b>Unassigned [3.34]</b>	0.61 [0.56-0.66]	0.2 [0.12-0.28]	0.72 [0.68-0.75]	0.22 [0.14-0.3]	0.67 [0.62-0.71]	0.63 [0.58-0.68]	0.69 [0.64-0.73]	0.57 [0.51-0.62]	0.64 [0.59-0.69]			
<b>Glucose [3.36-3.86]</b>	0.61 [0.56-0.66]	0.08 [0-0.16]	0.81 [0.78-0.84]	0.17 [0.09-0.25]	0.72 [0.68-0.76]	0.61 [0.55-0.66]	0.72 [0.68-0.76]	0.55 [0.49-0.6]	0.73 [0.69-0.77]	0.68 [0.63-0.72]		
<b>Unassigned [5.18]</b>	-0.22 [-0.29--0.14]	0.52 [0.46-0.58]	-0.45 [-0.51--0.38]	0.46 [0.39-0.52]	-0.41 [-0.47--0.34]	-0.47 [-0.53--0.4]	-0.42 [-0.48--0.35]	-0.23 [-0.31--0.15]	-0.6 [-0.65--0.54]	-0.34 [-0.41--0.27]	-0.48 [-0.54--0.41]	
<b>Undefined region [5.38-5.43]</b>	-0.52 [-0.57--0.45]	0.12 [0.03-0.19]	-0.84 [-0.86--0.81]	-0.05 [-0.13-0.03]	-0.75 [-0.78--0.71]	-0.7 [-0.74--0.66]	-0.77 [-0.8--0.73]	-0.55 [-0.6--0.49]	-0.9 [-0.91--0.88]	-0.62 [-0.67--0.57]	-0.73 [-0.77--0.69]	0.71 [0.67-0.75]

*Additional File 4 – L. Lecuyer et al. NMR metabolomic signatures reveal predictive plasma metabolites associated with long-term risk of developing breast cancer*

**Supplementary results of conditional logistic regressions for the complete list of the buckets**

**Supplementary Table 3** Associations between NMR metabolites and breast cancer risk from multivariable conditional logistic regression<sup>1</sup> - SU.VI.MAX cohort, France (1994-2007)

Metabolites highlighted with NOESY1D experiments							
Chemical shift (ppm) (metabolites) <sup>2</sup>	T1 (ref)		T2		T3		P-trend
	cases/ controls	OR	cases/ controls	OR [95%CI]	cases/ controls	OR [95%CI]	
0.0003	64/127	1.00	74/130	1.07 [0.69-1.66]	68/139	0.91 [0.57-1.44]	0.6890
0.0471	65/128	1.00	76/131	1.06 [0.68-1.66]	65/137	0.9 [0.57-1.42]	0.6357
0.0871	68/126	1.00	75/130	1.01 [0.64-1.58]	63/140	0.76 [0.48-1.2]	0.2418
0.1271	62/128	1.00	78/126	1.2 [0.76-1.87]	66/142	0.91 [0.57-1.45]	0.6388
0.1671	64/128	1.00	70/134	1.04 [0.66-1.62]	72/134	1.07 [0.68-1.68]	0.7699
0.2071	55/135	1.00	79/125	1.58 [1-2.48]	72/136	1.34 [0.84-2.14]	0.2491
0.2471	55/133	1.00	81/125	1.54 [0.98-2.41]	70/138	1.24 [0.78-1.97]	0.4342
0.2871	56/133	1.00	76/126	1.51 [0.97-2.35]	74/137	1.26 [0.78-2.03]	0.3261
0.3271	57/134	1.00	81/117	1.75 [1.11-2.75]	68/145	1.16 [0.71-1.9]	0.5761
0.3671	65/124	1.00	75/127	1.19 [0.77-1.86]	66/145	0.88 [0.55-1.4]	0.5945
0.4071	61/125	1.00	73/137	1.06 [0.69-1.64]	72/134	1.17 [0.74-1.86]	0.5053
0.4471	54/135	1.00	84/122	1.6 [1.04-2.48]	68/139	1.2 [0.75-1.93]	0.3469
0.4871	57/135	1.00	80/124	1.42 [0.92-2.19]	69/137	1.18 [0.74-1.88]	0.4287
0.5271	61/130	1.00	82/124	1.36 [0.88-2.1]	63/142	0.91 [0.56-1.48]	0.7761
0.5671	68/125	1.00	75/129	1.1 [0.71-1.7]	63/142	0.76 [0.47-1.23]	0.3001
0.6071	70/129	1.00	81/116	1.3 [0.84-2.01]	55/151	0.65 [0.4-1.05]	0.0742
0.6471	70/127	1.00	76/129	1.04 [0.68-1.6]	60/140	0.76 [0.48-1.2]	0.2346
0.6871	73/124	1.00	66/131	0.81 [0.52-1.24]	67/141	0.82 [0.53-1.27]	0.3587
0.7271	68/124	1.00	74/126	1.04 [0.68-1.6]	64/146	0.8 [0.51-1.27]	0.3614
0.7671	68/124	1.00	73/126	1.21 [0.75-1.93]	65/146	0.86 [0.53-1.38]	0.4623
0.86 (Lipoproteins (CH <sub>3</sub> ))	84/114	1.00	60/141	0.52 [0.33-0.82]	62/141	0.56 [0.36-0.88]	0.0119
0.9111	72/126	1.00	70/127	0.96 [0.61-1.52]	64/143	0.75 [0.48-1.19]	0.2261
0.9501	73/129	1.00	64/132	0.91 [0.59-1.41]	69/135	0.92 [0.59-1.43]	0.6995
0.9841	76/124	1.00	65/132	0.82 [0.53-1.25]	65/140	0.73 [0.47-1.14]	0.1653
1.0211	73/125	1.00	66/132	0.92 [0.6-1.42]	67/139	0.83 [0.53-1.31]	0.4251
1.0511	70/128	1.00	70/129	1 [0.65-1.54]	66/139	0.84 [0.53-1.32]	0.4505
1.0821	69/128	1.00	70/129	1.03 [0.66-1.6]	67/139	0.87 [0.56-1.37]	0.5567
1.0881	73/125	1.00	67/131	0.87 [0.56-1.33]	66/140	0.76 [0.48-1.19]	0.2319
1.1271	74/124	1.00	67/131	0.83 [0.54-1.29]	65/141	0.74 [0.47-1.17]	0.2032
1.1671	71/128	1.00	70/128	1.07 [0.69-1.67]	65/140	0.86 [0.54-1.37]	0.5262
1.27 (Lipoproteins (CH <sub>2</sub> ))	84/117	1.00	67/133	0.69 [0.44-1.07]	55/146	0.52 [0.33-0.82]	0.0043
1.4281	76/121	1.00	67/131	0.76 [0.49-1.18]	63/144	0.66 [0.42-1.05]	0.0780
1.4471	78/125	1.00	63/129	0.78 [0.5-1.2]	65/142	0.72 [0.47-1.12]	0.1392
1.56 (Lipids (CH <sub>2</sub> CH <sub>2</sub> CO))	84/117	1.00	72/128	0.7 [0.45-1.09]	50/151	0.43 [0.27-0.68]	0.0004
1.6311	72/126	1.00	71/124	1.04 [0.67-1.6]	63/146	0.76 [0.48-1.21]	0.2636
1.6361	72/125	1.00	71/123	1.06 [0.69-1.63]	63/148	0.72 [0.45-1.15]	0.1894
1.6471	69/124	1.00	73/131	0.98 [0.63-1.54]	64/141	0.8 [0.5-1.29]	0.3580

## ANNEXE 1

1.7511	73/122	1.00	65/136	0.74 [0.46-1.17]	68/138	0.73 [0.45-1.2]	0.2258
1.7911	76/122	1.00	67/134	0.79 [0.51-1.23]	63/140	0.62 [0.38-1]	0.0522
1.83-1.85 (Undefined region)	78/120	1.00	67/133	0.72 [0.46-1.12]	61/143	0.55 [0.35-0.89]	0.0137
1.8521	71/124	1.00	72/133	0.87 [0.57-1.33]	63/139	0.72 [0.45-1.15]	0.1752
1.8871	73/121	1.00	73/130	0.86 [0.56-1.32]	60/145	0.63 [0.4-1]	0.0533
2.001 & 2.035 (Lipids (CCH <sub>2</sub> ) & Glycoproteins (N-acetyl) <sup>3</sup> )	79/121	1.00	72/124	0.86 [0.55-1.33]	55/151	0.55 [0.35-0.87]	0.0095
2.0631	75/123	1.00	70/135	0.82 [0.52-1.28]	61/138	0.69 [0.43-1.11]	0.1267
2.1101	73/124	1.00	71/129	0.92 [0.58-1.45]	62/143	0.67 [0.41-1.09]	0.1044
2.1511	75/123	1.00	71/129	0.91 [0.59-1.41]	60/144	0.67 [0.43-1.05]	0.0787
2.22 (Lipids (CH <sub>2</sub> CO) & Acetone <sup>3</sup> )	86/113	1.00	69/135	0.62 [0.4-0.95]	51/148	0.43 [0.27-0.69]	0.0004
2.2641	72/128	1.00	76/123	1.17 [0.74-1.85]	58/145	0.65 [0.41-1.05]	0.0765
2.3111	73/129	1.00	73/123	1.08 [0.69-1.68]	60/144	0.69 [0.44-1.08]	0.1153
2.3511	74/129	1.00	62/137	0.84 [0.55-1.3]	70/130	0.94 [0.6-1.46]	0.7784
2.3881	65/131	1.00	73/131	1.12 [0.71-1.76]	68/134	0.98 [0.6-1.59]	0.9184
2.4141	72/125	1.00	63/138	0.8 [0.51-1.27]	71/133	0.94 [0.59-1.51]	0.8436
2.4711	64/136	1.00	67/131	1.06 [0.66-1.69]	75/129	1.21 [0.76-1.92]	0.4215
2.5111	62/137	1.00	74/129	1.44 [0.9-2.32]	70/130	1.3 [0.81-2.09]	0.3247
2.5511	64/133	1.00	76/122	1.34 [0.84-2.13]	66/141	0.99 [0.61-1.59]	0.9083
2.5821	61/135	1.00	68/129	1.22 [0.77-1.95]	77/132	1.41 [0.89-2.25]	0.1450
2.6271	62/134	1.00	68/130	1.17 [0.75-1.84]	76/132	1.34 [0.86-2.08]	0.2031
2.6471	77/124	1.00	69/131	0.83 [0.54-1.27]	60/141	0.75 [0.48-1.16]	0.1867
2.6871	72/130	1.00	69/126	0.99 [0.64-1.54]	65/140	0.84 [0.54-1.32]	0.4478
2.7911	70/127	1.00	70/130	0.88 [0.56-1.37]	66/139	0.84 [0.53-1.31]	0.4354
2.8311	71/129	1.00	70/128	1 [0.65-1.52]	65/139	0.86 [0.55-1.34]	0.5070
2.8361	53/138	1.00	88/122	1.96 [1.23-3.13]	65/136	1.35 [0.84-2.17]	0.2668
2.8471	60/137	1.00	70/126	1.3 [0.82-2.05]	76/133	1.33 [0.84-2.1]	0.2203
2.8871	56/141	1.00	79/123	1.54 [0.98-2.42]	71/132	1.36 [0.84-2.18]	0.1926
2.9271	64/137	1.00	68/129	1.06 [0.68-1.65]	74/130	1.16 [0.75-1.79]	0.5126
2.9791	65/132	1.00	66/138	0.95 [0.62-1.43]	75/126	1.21 [0.77-1.91]	0.4484
3.069	72/127	1.00	64/138	0.77 [0.49-1.2]	70/131	0.87 [0.55-1.38]	0.5515
3.111	63/134	1.00	72/130	1.14 [0.73-1.77]	71/132	1.13 [0.72-1.77]	0.5880
3.13	67/126	1.00	66/138	0.89 [0.57-1.4]	73/132	1.07 [0.68-1.69]	0.7610
3.167	72/128	1.00	69/134	0.91 [0.59-1.4]	65/134	0.85 [0.54-1.33]	0.4739
3.269	60/137	1.00	76/126	1.41 [0.91-2.17]	70/133	1.24 [0.79-1.95]	0.3091
3.311	61/136	1.00	82/125	1.44 [0.94-2.22]	63/135	1.05 [0.67-1.65]	0.7949
3.332	59/138	1.00	72/130	1.41 [0.92-2.16]	75/128	1.42 [0.91-2.22]	0.1146
3.368	64/141	1.00	67/126	1.14 [0.73-1.77]	75/129	1.27 [0.81-2]	0.2924
3.437	62/142	1.00	67/129	1.19 [0.77-1.85]	77/125	1.46 [0.92-2.3]	0.1086
3.497	59/144	1.00	72/125	1.3 [0.83-2.04]	75/127	1.43 [0.91-2.25]	0.1114
3.531	64/141	1.00	76/120	1.45 [0.93-2.26]	66/135	1.09 [0.69-1.7]	0.6989
3.588	66/138	1.00	69/131	1.3 [0.83-2.03]	71/127	1.27 [0.8-2.01]	0.3212
3.616	66/141	1.00	72/125	1.23 [0.8-1.91]	68/130	1.06 [0.67-1.67]	0.7961
3.714	56/146	1.00	80/122	1.61 [1.03-2.52]	70/128	1.36 [0.84-2.2]	0.1862
3.791	65/137	1.00	73/129	1.19 [0.77-1.83]	68/130	1.11 [0.7-1.74]	0.6507
3.793	67/135	1.00	72/129	1.09 [0.71-1.68]	67/132	1.03 [0.66-1.61]	0.8897
3.807	61/142	1.00	76/122	1.42 [0.9-2.25]	69/132	1.17 [0.74-1.85]	0.5068
3.859	61/140	1.00	78/126	1.41 [0.89-2.22]	67/130	1.15 [0.73-1.8]	0.5483
3.951	71/127	1.00	66/137	0.85 [0.55-1.33]	69/132	0.93 [0.59-1.48]	0.7673

## ANNEXE 1

3.953	65/140	1.00	73/128	1.25 [0.8-1.96]	68/128	1.14 [0.73-1.78]	0.5707
4.031	72/131	1.00	80/123	1.23 [0.78-1.92]	54/142	0.65 [0.4-1.05]	0.0843
4.04 & 4.25 (Glycerol-derived compounds (CH <sub>2</sub> ))	90/113	1.00	60/141	0.54 [0.35-0.83]	56/142	0.49 [0.32-0.77]	0.0014
4.094	67/132	1.00	68/133	1.07 [0.69-1.66]	71/131	1.11 [0.71-1.71]	0.6534
4.149	64/137	1.00	74/129	1.19 [0.76-1.86]	68/130	1.07 [0.66-1.74]	0.7872
4.191	65/136	1.00	74/129	1.22 [0.8-1.88]	67/131	1.01 [0.64-1.6]	0.9381
4.311	69/134	1.00	77/126	1.08 [0.71-1.66]	60/136	0.81 [0.52-1.27]	0.3703
4.351	72/130	1.00	73/128	1.01 [0.66-1.54]	61/138	0.79 [0.51-1.22]	0.2898
4.381	66/141	1.00	72/126	1.14 [0.74-1.73]	68/129	1.15 [0.75-1.77]	0.5164
4.426	72/135	1.00	69/132	0.95 [0.62-1.47]	65/129	0.95 [0.62-1.46]	0.8213
4.4679	70/132	1.00	72/130	0.93 [0.59-1.44]	64/134	0.89 [0.55-1.44]	0.6431
4.5109	75/126	1.00	66/137	0.82 [0.53-1.28]	65/133	0.89 [0.56-1.4]	0.6143
4.5509	72/128	1.00	67/140	0.91 [0.59-1.41]	67/128	0.97 [0.6-1.54]	0.8898
4.5909	68/134	1.00	66/136	0.93 [0.59-1.46]	72/126	1.2 [0.75-1.91]	0.4610
4.7979	65/133	1.00	71/133	1.34 [0.86-2.1]	70/130	1.31 [0.83-2.08]	0.2478
4.8709	66/132	1.00	74/132	1.18 [0.75-1.85]	66/132	1.11 [0.69-1.76]	0.6843
4.8879	70/127	1.00	67/136	0.96 [0.62-1.5]	69/133	1.07 [0.67-1.71]	0.7784
4.9909	70/130	1.00	71/130	1.12 [0.72-1.75]	65/136	0.98 [0.61-1.58]	0.9374
5.0309	70/124	1.00	67/138	0.9 [0.58-1.39]	69/134	0.97 [0.61-1.56]	0.8981
5.0439	67/129	1.00	77/126	1.31 [0.83-2.07]	62/141	0.91 [0.56-1.49]	0.7543
5.0789	71/125	1.00	71/132	0.95 [0.61-1.49]	64/139	0.9 [0.56-1.45]	0.6775
5.0869	70/131	1.00	72/124	1 [0.64-1.57]	64/141	0.93 [0.58-1.49]	0.7614
5.1269	74/122	1.00	77/125	1.01 [0.64-1.58]	55/149	0.62 [0.39-1]	0.0505
5.18 (Glycerol-derived compounds (CH))	88/111	1.00	68/132	0.6 [0.38-0.94]	50/153	0.37 [0.23-0.61]	0.0001
5.2069	78/125	1.00	61/131	0.76 [0.47-1.23]	67/140	0.72 [0.44-1.17]	0.1780
5.28 (Unsaturated lipids (CH=CH))	80/122	1.00	68/129	0.79 [0.51-1.22]	58/145	0.6 [0.38-0.94]	0.0268
5.3899	68/133	1.00	73/120	1.3 [0.81-2.07]	65/143	0.93 [0.58-1.47]	0.7242
5.4309	68/132	1.00	76/121	1.44 [0.89-2.31]	62/143	0.92 [0.57-1.48]	0.7037
5.4709	70/129	1.00	74/124	1.24 [0.78-2]	62/143	0.84 [0.52-1.36]	0.5054

## Metabolites highlighted with CPMG experiments

Chemical shift (metabolites)	T1 (ref)		T2		T3		P-trend
	cases/ controls	OR	cases/ controls	OR [95%CI]	cases/ controls	OR [95%CI]	
0.0305	75/120	1.00	68/125	0.85 [0.54-1.33 ]	60/133	0.74 [0.47-1.14 ]	0.1734
0.0735	74/121	1.00	69/124	0.92 [0.59-1.42 ]	60/133	0.75 [0.48-1.16 ]	0.1921
0.1310	74/119	1.00	66/127	0.83 [0.52-1.31 ]	63/132	0.78 [0.5-1.21 ]	0.2711
0.2050	72/123	1.00	66/123	0.91 [0.58-1.43 ]	65/132	0.84 [0.54-1.3 ]	0.4294
0.2415	71/125	1.00	65/123	0.9 [0.58-1.4 ]	67/130	0.92 [0.6-1.42 ]	0.7156
0.2830	69/124	1.00	72/119	1.03 [0.67-1.59 ]	62/135	0.82 [0.53-1.28 ]	0.3730
0.3235	71/123	1.00	71/118	0.95 [0.62-1.46 ]	61/137	0.78 [0.5-1.21 ]	0.2651
0.3480	73/121	1.00	66/126	0.88 [0.56-1.36 ]	64/131	0.79 [0.51-1.24 ]	0.3074
0.3875	74/121	1.00	65/128	0.82 [0.53-1.26 ]	64/129	0.82 [0.53-1.27 ]	0.3607
0.4275	73/122	1.00	69/121	0.9 [0.59-1.38 ]	61/135	0.75 [0.48-1.17 ]	0.2042
0.4675	74/120	1.00	66/122	0.82 [0.54-1.26 ]	63/136	0.75 [0.48-1.15 ]	0.1828
0.5075	72/123	1.00	69/121	0.94 [0.61-1.44 ]	62/134	0.79 [0.51-1.24 ]	0.3148
0.5475	72/121	1.00	66/125	0.84 [0.54-1.29 ]	65/132	0.82 [0.52-1.28 ]	0.3674
0.5875	73/121	1.00	64/126	0.82 [0.53-1.27 ]	66/131	0.83 [0.54-1.29 ]	0.4102
0.6275	72/121	1.00	63/127	0.78 [0.51-1.21 ]	68/130	0.91 [0.58-1.43 ]	0.6494
0.6675	73/119	1.00	67/130	0.83 [0.54-1.27 ]	63/129	0.8 [0.52-1.25 ]	0.3206

## ANNEXE 1

0.7075	78/113	1.00	63/129	0.72 [0.47-1.12 ]	62/136	0.69 [0.45-1.06 ]	0.0933
0.7475	73/113	1.00	61/135	0.67 [0.43-1.06 ]	69/130	0.86 [0.54-1.34 ]	0.5123
0.7875	70/113	1.00	71/127	0.96 [0.62-1.49 ]	62/138	0.75 [0.47-1.19 ]	0.2145
0.8595	75/123	1.00	62/125	0.85 [0.55-1.31 ]	66/130	0.84 [0.54-1.31 ]	0.4409
0.9310	61/127	1.00	70/128	1.08 [0.69-1.7 ]	72/123	1.18 [0.74-1.87 ]	0.4852
0.98 & 1.03 (Valine)	53/134	1.00	69/129	1.35 [0.86-2.11 ]	81/115	1.83 [1.15-2.92 ]	0.0113
1.0730	69/130	1.00	65/121	1.06 [0.68-1.66 ]	69/127	1.06 [0.67-1.68 ]	0.7953
1.0845	74/121	1.00	57/133	0.68 [0.43-1.08 ]	72/124	0.97 [0.61-1.53 ]	0.8907
1.1080	74/120	1.00	63/130	0.81 [0.52-1.26 ]	66/128	0.85 [0.54-1.35 ]	0.4692
1.1475	78/112	1.00	65/132	0.75 [0.49-1.14 ]	60/134	0.66 [0.42-1.04 ]	0.0687
1.1875	74/115	1.00	67/129	0.75 [0.48-1.15 ]	62/134	0.73 [0.46-1.14 ]	0.1625
1.26 (Lipoproteins (CH <sub>2</sub> ))	78/114	1.00	70/128	0.83 [0.53-1.29 ]	55/136	0.58 [0.36-0.92 ]	0.0209
1.3715	68/126	1.00	80/116	1.31 [0.84-2.04 ]	55/136	0.74 [0.46-1.2 ]	0.2137
1.4115	61/131	1.00	77/116	1.48 [0.95-2.29 ]	65/131	1.09 [0.69-1.7 ]	0.7153
1.4340	63/129	1.00	71/122	1.21 [0.79-1.86 ]	69/127	1.16 [0.74-1.8 ]	0.5114
1.4420	62/133	1.00	71/119	1.25 [0.81-1.93 ]	70/126	1.25 [0.81-1.92 ]	0.3164
1.4850	62/133	1.00	68/122	1.22 [0.78-1.89 ]	73/123	1.31 [0.84-2.06 ]	0.2337
1.5570	70/122	1.00	79/117	1.14 [0.74-1.76 ]	54/139	0.68 [0.42-1.09 ]	0.1029
1.6115	67/124	1.00	76/117	1.27 [0.82-1.97 ]	60/137	0.85 [0.54-1.35 ]	0.5250
1.6340	61/134	1.00	67/122	1.12 [0.72-1.75 ]	75/122	1.36 [0.88-2.1 ]	0.1691
1.6420	58/134	1.00	72/119	1.33 [0.86-2.07 ]	73/125	1.38 [0.89-2.14 ]	0.1465
1.7 (Lysine & Arginine <sup>3</sup> )	57/139	1.00	63/121	1.19 [0.77-1.85 ]	83/118	1.62 [1.05-2.48 ]	0.0287
1.8070	60/133	1.00	65/122	1.08 [0.7-1.68 ]	78/123	1.36 [0.88-2.11 ]	0.1666
1.8300	62/131	1.00	67/120	1.11 [0.72-1.71 ]	74/127	1.22 [0.79-1.88 ]	0.3706
1.8435	60/131	1.00	70/119	1.24 [0.8-1.9 ]	73/128	1.25 [0.8-1.94 ]	0.3208
1.8685	64/128	1.00	62/126	0.98 [0.61-1.57 ]	77/124	1.26 [0.81-1.94 ]	0.2936
1.9075	61/130	1.00	68/126	1.11 [0.7-1.75 ]	74/122	1.25 [0.8-1.95 ]	0.3241
1.9475	75/121	1.00	64/126	0.84 [0.54-1.31 ]	64/131	0.8 [0.51-1.24 ]	0.3122
2.0150	70/125	1.00	76/114	1.31 [0.84-2.04 ]	57/139	0.74 [0.47-1.17 ]	0.2067
2.0790	64/128	1.00	66/127	1.01 [0.65-1.56 ]	73/123	1.19 [0.76-1.87 ]	0.4553
2.1235	60/131	1.00	70/122	1.31 [0.83-2.05 ]	73/125	1.21 [0.77-1.92 ]	0.3951
2.1600	66/128	1.00	69/120	1.11 [0.71-1.71 ]	68/130	0.98 [0.64-1.51 ]	0.9404
2.22 (Lipids (CH <sub>2</sub> CO) & Acetone <sup>3</sup> )	72/123	1.00	80/110	1.16 [0.76-1.78 ]	51/145	0.58 [0.36-0.93 ]	0.0264
2.2880	54/137	1.00	77/115	1.67 [1.07-2.62 ]	72/126	1.45 [0.92-2.28 ]	0.1222
2.33 (Unassigned)	54/140	1.00	72/118	1.58 [1-2.51 ]	77/120	1.67 [1.08-2.6 ]	0.0233
2.3690	60/133	1.00	69/125	1.22 [0.77-1.94 ]	74/120	1.37 [0.88-2.13 ]	0.1699
2.45 (Glutamine)	59/131	1.00	64/129	1.17 [0.74-1.84 ]	80/118	1.61 [1.02-2.55 ]	0.0423
2.4860	60/132	1.00	66/125	1.2 [0.77-1.88 ]	77/121	1.37 [0.88-2.13 ]	0.1635
2.5260	61/131	1.00	69/127	1.15 [0.73-1.81 ]	73/120	1.34 [0.87-2.05 ]	0.1837
2.55-2.57 (Undefined region)	52/140	1.00	77/115	1.74 [1.11-2.74 ]	74/123	1.63 [1.04-2.55 ]	0.0355
2.57-2.61 (Undefined region)	56/131	1.00	65/133	1.11 [0.71-1.76 ]	82/114	1.79 [1.15-2.8 ]	0.0094
2.6295	59/133	1.00	72/121	1.44 [0.92-2.27 ]	72/124	1.38 [0.88-2.16 ]	0.1836
2.6675	61/133	1.00	62/129	1.04 [0.65-1.66 ]	80/116	1.52 [0.99-2.33 ]	0.0483
2.7395	64/129	1.00	75/117	1.27 [0.84-1.94 ]	64/132	0.93 [0.59-1.48 ]	0.8406
2.8115	61/131	1.00	80/114	1.36 [0.88-2.1 ]	62/133	0.94 [0.59-1.48 ]	0.7972
2.8395	61/133	1.00	73/119	1.42 [0.92-2.21 ]	69/126	1.17 [0.76-1.8 ]	0.4804
2.8675	55/136	1.00	71/119	1.56 [0.98-2.48 ]	77/123	1.52 [0.99-2.35 ]	0.0645
2.9075	58/133	1.00	68/128	1.17 [0.75-1.82 ]	77/117	1.56 [1-2.44 ]	0.0532



## ANNEXE 1

3.01 & 3.03 (Lysine, Creatine & Creatinine <sup>3</sup> )	52/141	1.00	70/120	1.64 [1.02-2.63 ]	81/117	1.84 [1.19-2.85 ]	0.0066
3.0915	63/130	1.00	65/128	1.05 [0.67-1.64 ]	75/120	1.34 [0.86-2.09 ]	0.1934
3.1210	63/131	1.00	64/124	1.03 [0.65-1.63 ]	76/123	1.28 [0.83-1.98 ]	0.2595
3.1490	70/124	1.00	61/125	0.84 [0.53-1.33 ]	72/129	1.03 [0.66-1.59 ]	0.8666
3.2195	68/125	1.00	65/133	0.9 [0.57-1.42 ]	70/120	1.08 [0.68-1.7 ]	0.7271
3.2915	58/139	1.00	69/118	1.43 [0.91-2.27 ]	76/121	1.43 [0.92-2.22 ]	0.1029
3.3125	55/140	1.00	77/112	1.79 [1.14-2.83 ]	71/126	1.4 [0.89-2.21 ]	0.1245
3.3230	59/137	1.00	74/115	1.54 [0.97-2.45 ]	70/126	1.27 [0.81-1.99 ]	0.3019
3.34 (Unassigned)	56/140	1.00	72/115	1.72 [1.1-2.69 ]	75/123	1.59 [1.01-2.5 ]	0.0433
3.36-3.86 (Glucose)	57/137	1.00	66/124	1.17 [0.73-1.86 ]	80/117	1.61 [1.04-2.5 ]	0.0312
3.5425	60/136	1.00	66/123	1.11 [0.7-1.76 ]	77/119	1.44 [0.92-2.26 ]	0.1080
3.7935	60/133	1.00	66/127	1.09 [0.69-1.73 ]	77/118	1.44 [0.92-2.26 ]	0.1101
3.9650	58/137	1.00	65/123	1.21 [0.78-1.88 ]	80/118	1.53 [0.98-2.39 ]	0.0611
4.0050	57/137	1.00	70/119	1.4 [0.9-2.16 ]	76/122	1.41 [0.9-2.21 ]	0.1187
4.0330	58/134	1.00	70/121	1.38 [0.87-2.17 ]	75/123	1.37 [0.87-2.15 ]	0.1571
4.0575	63/130	1.00	68/125	1.16 [0.73-1.83 ]	72/123	1.18 [0.75-1.85 ]	0.4663
4.1160	64/128	1.00	59/134	0.92 [0.58-1.46 ]	80/116	1.39 [0.89-2.16 ]	0.1561
4.1670	57/137	1.00	73/118	1.51 [0.96-2.36 ]	73/123	1.34 [0.86-2.1 ]	0.1756
4.1905	60/134	1.00	68/122	1.3 [0.84-2.02 ]	75/122	1.32 [0.85-2.07 ]	0.1945
4.2030	60/134	1.00	70/120	1.37 [0.88-2.13 ]	73/124	1.23 [0.78-1.94 ]	0.3134
4.2595	59/132	1.00	69/124	1.34 [0.85-2.11 ]	75/122	1.4 [0.88-2.22 ]	0.1435
5.0185	75/123	1.00	69/119	0.99 [0.62-1.59 ]	59/136	0.72 [0.46-1.14 ]	0.1419
5.0610	77/122	1.00	67/121	0.93 [0.58-1.49 ]	59/135	0.72 [0.46-1.13 ]	0.1413
5.0850	74/123	1.00	70/120	1.03 [0.64-1.67 ]	59/135	0.76 [0.48-1.2 ]	0.2027
5.1075	77/120	1.00	65/125	0.84 [0.52-1.36 ]	61/133	0.74 [0.47-1.17 ]	0.1983
5.1475	77/119	1.00	66/125	0.83 [0.52-1.33 ]	60/134	0.7 [0.44-1.11 ]	0.1322
5.18 (Unassigned)	81/112	1.00	64/133	0.6 [0.38-0.94 ]	58/133	0.61 [0.39-0.97 ]	0.0359
5.2275	65/125	1.00	67/130	1.01 [0.64-1.59 ]	71/123	1.18 [0.75-1.85 ]	0.4709
5.2995	79/113	1.00	64/134	0.67 [0.42-1.06 ]	60/131	0.65 [0.41-1.02 ]	0.0564
5.38-5.43 (Undefined region)	81/115	1.00	61/129	0.68 [0.43-1.07 ]	61/134	0.63 [0.41-0.98 ]	0.0457
5.4515	77/121	1.00	69/120	0.91 [0.57-1.43 ]	57/137	0.66 [0.42-1.03 ]	0.0622
5.4915	77/122	1.00	67/118	0.85 [0.54-1.35 ]	59/138	0.68 [0.44-1.04 ]	0.0758

SU.VI.MAX, Supplémentation en Vitamines et Minéraux Antioxydants ; T, tertile ; ref, reference ; OR, odds ratio ; CI, confidence interval

<sup>1</sup> Multivariable models were additionally adjusted for age (continuous), BMI (continuous), menopausal status (pre/post-menopause), intervention group of the SU.VI.MAX trial (placebo/supplemented), smoking status (current smokers and non-smokers), season of blood draw (defined above), height (continuous), alcohol intake (continuous), physical activity (irregular/< 1h/d walking equivalent/≥ 1h/d walking equivalent), education level (primary/secondary/superior), family history of breast cancer (yes/no), number of children (continuous), and use of hormone replacement therapy for menopause (yes/no) (reference model). Tests for linear trend were performed using the ordinal score on tertiles of buckets intensity.

<sup>2</sup> As usually done in this type of study, the variables which were not associated with the risk of breast cancer were not annotated but were assigned with the lower value of its chemical shift range.

<sup>3</sup> Multiple assignments for the same NMR variable due to spectral overlap.

*Additional File 5 – L. Lecuyer et al. NMR metabolomic signatures reveal predictive plasma metabolites associated with long-term risk of developing breast cancer*

#### **Supplementary results of sensitivity and stratification analyses**

After the exclusion of cases diagnosed during the two first years of follow-up, the 9 NMR variables from NOESY1D experiments and 11 (instead of 13) variables from CPMG experiments remained associated with breast cancer risk (**Supplementary Table 4**).

In women with a follow-up duration < median (i.e. 6.5 years), lipoproteins, lipids, acetone, glutamine, glycerol-derived compounds and glycoproteins were associated with breast cancer risk (**Supplementary Table 5**). Valine, lysine, arginine, lipids, creatine, creatinine and glycerol-derived compounds were associated with breast cancer risk in women with a follow-up duration  $\geq$  median (**Supplementary Table 6**).

In women who were premenopausal at cancer diagnosis, glycerol-derived compounds, lipoproteins, lipids, acetone and glutamine were associated with breast cancer risk (**Supplementary Table 7**); while in postmenopausal women at cancer diagnosis glycerol-derived compounds, lipids, acetone, valine, lysine, creatine and creatinine were associated with breast cancer (**Supplementary Table 8**).

**Supplementary Table 4** NMR metabolites associated with breast cancer risk from multivariable conditional logistic regression<sup>1</sup> after the exclusion of cases diagnosed during the second year of follow-up - SU.VI.MAX cohort, France (1994-2007)

<b>Metabolites highlighted with NOESY1D experiments</b>								
Assignments	Chemical shift (ppm)	T1 (ref)		T2		T3		P-trend
		cases/controls	OR	cases/controls	OR [95% CI]	cases/controls	OR [95% CI]	
Lipoproteins (CH <sub>3</sub> )	0.86	77/109	1.00	58/128	0.58 [0.36-0.92]	56/130	0.57 [0.35-0.91]	0.02
Lipoproteins (CH <sub>2</sub> )	1.27	76/112	1.00	66/122	0.79 [0.5-1.25]	49/133	0.54 [0.34-0.87]	0.01
Lipids (CH <sub>2</sub> CH <sub>2</sub> CO)	1.56	79/111	1.00	67/119	0.71 [0.45-1.12]	45/137	0.43 [0.26-0.69]	0.0005
Undefined region	1.83-1.85	73/114	1.00	63/121	0.77 [0.49-1.21]	55/132	0.55 [0.33-0.89]	0.02
Lipids (CCH <sub>2</sub> ) & Glycoproteins (N-acetyl) <sup>2</sup>	2.001 & 2.035	72/113	1.00	69/116	0.91 [0.57-1.43]	50/138	0.56 [0.35-0.9]	0.02
Lipids (CH <sub>2</sub> CO) & Acetone <sup>2</sup>	2.22	79/108	1.00	66/125	0.66 [0.42-1.03]	46/134	0.44 [0.27-0.72]	0.001
Glycerol-derived compounds (CH <sub>2</sub> )	4.04 & 4.25	85/105	1.00	57/130	0.54 [0.35-0.84]	49/132	0.46 [0.28-0.73]	0.0007
Glycerol-derived compounds (CH)	5.18	81/105	1.00	65/123	0.63 [0.4-1]	45/139	0.38 [0.23-0.63]	0.0002
Unsaturated lipids (CH=CH)	5.28	73/116	1.00	67/116	0.89 [0.57-1.41]	51/135	0.58 [0.36-0.94]	0.03

<b>Metabolites highlighted with CPMG experiments</b>								
Assignments	Chemical shift (ppm)	T1 (ref)		T2		T3		P-trend
		cases/controls	OR	cases/controls	OR [95% CI]	cases/controls	OR [95% CI]	
Valine	0.98 & 1.03	48/124	1.00	65/114	1.48 [0.92-2.38]	75/111	1.8 [1.11-2.93]	0.02
Lipoproteins (CH <sub>2</sub> )	1.26	71/108	1.00	68/117	0.91 [0.57-1.44]	49/124	0.59 [0.36-0.96]	0.04
Lysine & Arginine <sup>2</sup>	1.70	53/128	1.00	58/110	1.24 [0.78-1.96]	77/111	1.62 [1.04-2.53]	0.03
Lipids (CH <sub>2</sub> CO) & Acetone <sup>2</sup>	2.22	65/117	1.00	78/99	1.35 [0.87-2.11]	45/133	0.59 [0.36-0.97]	0.04
Unassigned	2.33	51/130	1.00	66/107	1.64 [1.01-2.65]	71/112	1.63 [1.03-2.56]	0.04
Undefined region	2.55-2.57	48/130	1.00	70/104	1.81 [1.13-2.91]	70/115	1.69 [1.06-2.67]	0.03
Undefined region	2.57-2.61	52/120	1.00	61/122	1.15 [0.72-1.85]	75/107	1.73 [1.09-2.76]	0.02
Lysine, Creatine & Creatinine <sup>2</sup>	3.01 & 3.03	50/130	1.00	62/107	1.6 [0.98-2.61]	76/112	1.76 [1.13-2.76]	0.01
Glucose	3.36-3.86	54/127	1.00	59/111	1.13 [0.7-1.83]	75/111	1.59 [1-2.5]	0.04
Unassigned	5.18	76/105	1.00	59/120	0.6 [0.38-0.96]	53/124	0.59 [0.36-0.95]	0.03
Undefined region	5.38-5.43	77/110	1.00	57/115	0.72 [0.45-1.14]	54/124	0.6 [0.38-0.95]	0.03

SU.VI.MAX, Supplémentation en Vitamines et Minéraux Antioxydants ; T, tertile ; ref, reference ; OR, odds ratio ; CI, confidence interval

<sup>1</sup> Multivariable models were additionally adjusted for age (continuous), BMI (continuous), menopausal status (pre/post-menopause), intervention group of the SU.VI.MAX trial (placebo/supplemented), smoking status (current smokers and non-smokers), season of blood draw (defined above), height (continuous), alcohol intake (continuous), physical activity (irregular/< 1h/d walking equivalent/≥ 1h/d walking equivalent), education level (primary/secondary/superior), family history of breast cancer (yes/no), number of children (continuous), and use of hormone replacement therapy for menopause (yes/no) (reference model). Tests for linear trend were performed using the ordinal score on tertiles of buckets intensity.

<sup>2</sup> Multiple assignments for the same NMR variable due to spectral overlap

**Supplementary Table 5** NMR Metabolites Associated with Breast Cancer Risk in the Women with a Follow-up Duration < Median (6.5 years) from Multivariable Conditional Logistic Regression<sup>1</sup> - SU.VI.MAX Cohort, France (1994-2007)

<b>Metabolites highlighted with NOESY1D experiments</b>								
Assignments	Chemical shift (ppm)	T1 (ref)		T2		T3		P-trend
		cases/controls	OR	cases/controls	OR [95% CI]	cases/controls	OR [95% CI]	
Lipoproteins (CH <sub>3</sub> )	0.86	46/57	1.00	27/65	0.41 [0.2-0.84]	28/62	0.48 [0.24-0.97]	0.03
Lipoproteins (CH <sub>2</sub> )	1.27	44/54	1.00	32/64	0.63 [0.32-1.26]	25/66	0.39 [0.19-0.79]	0.009
Lipids (CH <sub>2</sub> -CH <sub>2</sub> CO)	1.56	41/52	1.00	38/64	0.64 [0.32-1.26]	22/68	0.31 [0.15-0.66]	0.002
Lipids (CCH <sub>2</sub> ) & Glycoproteins (N-acetyl) <sup>2</sup>	2.001 & 2.035	40/57	1.00	36/59	0.8 [0.4-1.63]	25/68	0.47 [0.23-0.94]	0.03
Lipids (CH <sub>2</sub> CO) & Acetone <sup>2</sup>	2.22	44/50	1.00	35/68	0.56 [0.28-1.09]	22/66	0.32 [0.15-0.67]	0.003
Glycerol-derived compounds (CH <sub>2</sub> )	4.04 & 4.25	42/52	1.00	33/65	0.72 [0.38-1.38]	26/67	0.41 [0.2-0.83]	0.01
Glycerol-derived compounds (CH)	5.18	44/48	1.00	33/70	0.35 [0.16-0.77]	24/66	0.25 [0.11-0.6]	0.002

<b>Metabolites highlighted with CPMG experiments</b>								
Assignments	Chemical shift (ppm)	T1 (ref)		T2		T3		P-trend
		cases/controls	OR	cases/controls	OR [95% CI]	cases/controls	OR [95% CI]	
Lipoproteins (CH <sub>2</sub> )	1.26	44/58	1.00	29/60	0.72 [0.37-1.42]	29/68	0.39 [0.19-0.8]	0.01
Lipids (CH <sub>2</sub> CO) & Acetone <sup>2</sup>	2.22	38/57	1.00	38/51	0.98 [0.49-1.95]	26/78	0.4 [0.19-0.86]	0.02
Glutamine	2.45	27/63	1.00	36/68	1.61 [0.79-3.28]	39/55	2.27 [1.13-4.57]	0.02
Unassigned	3.34	26/67	1.00	39/57	2.3 [1.15-4.59]	37/62	2.02 [1.01-4.07]	0.048

SU.VI.MAX, Supplémentation en Vitamines et Minéraux Antioxydants ; T, tertile ; ref, reference ; OR, odds ratio ; CI, confidence interval

<sup>1</sup> Multivariable models were additionally adjusted for age (continuous), BMI (continuous), menopausal status (pre/post-menopause), intervention group of the SU.VI.MAX trial (placebo/supplemented), smoking status (current smokers and non-smokers), season of blood draw (defined above), height (continuous), alcohol intake (continuous), physical activity (irregular/< 1h/d walking equivalent/≥ 1h/d walking equivalent), education level (primary/secondary/superior), family history of breast cancer (yes/no), number of children (continuous), and use of hormone replacement therapy for menopause (yes/no) (reference model). Tests for linear trend were performed using the ordinal score on tertiles of buckets intensity.

<sup>2</sup> Multiple assignments for the same NMR variable due to spectral overlap

**Supplementary Table 6** NMR metabolites associated with breast cancer risk in the women with a follow-up duration  $\geq$  median from multivariable conditional logistic regression<sup>1</sup> - SU.VI.MAX cohort, France (1994-2007)

<b>Metabolites highlighted with NOESY1D experiments</b>								
Assignments	Chemical shift (ppm)	T1 (ref)		T2		T3		P-trend
		cases/controls	OR	cases/controls	OR [95% CI]	cases/controls	OR [95% CI]	
Lipids (CH <sub>2</sub> CH <sub>2</sub> CO)	1.56	42/63	1.00	34/60	0.71 [0.37-1.35]	26/70	0.48 [0.25-0.94]	0.03
Glycerol-derived compounds (CH <sub>2</sub> )	4.04 & 4.25	47/59	1.00	26/70	0.41 [0.21-0.8]	29/64	0.51 [0.26-1.01]	0.04
Glycerol-derived compounds (CH)	5.18	43/61	1.00	34/59	0.79 [0.42-1.46]	25/73	0.39 [0.19-0.78]	0.01

<b>Metabolites highlighted with CPMG experiments</b>								
Assignments	Chemical shift (ppm)	T1 (ref)		T2		T3		P-trend
		cases/controls	OR	cases/controls	OR [95% CI]	cases/controls	OR [95% CI]	
Valine	0.98 & 1.03	28/70	1.00	30/67	1.15 [0.6-2.23]	43/54	2.22 [1.12-4.43]	0.02
Lysine & Arginine <sup>2</sup>	1.70	32/79	1.00	27/60	1.2 [0.65-2.2]	42/52	1.89 [1.03-3.45]	0.04
Undefined region	2.55-2.57	25/77	1.00	36/59	1.98 [1.04-3.75]	40/55	2.21 [1.16-4.21]	0.02
Undefined region	2.57-2.61	30/69	1.00	24/68	0.78 [0.39-1.59]	47/54	1.97 [1.07-3.64]	0.02
Lysine, Creatine & Creatinine <sup>2</sup>	3.01 & 3.03	28/76	1.00	30/59	1.21 [0.63-2.32]	43/56	2.03 [1.1-3.76]	0.03

SU.VI.MAX, Supplémentation en Vitamines et Minéraux Antioxydants ; T, tertile ; ref, reference ; OR, odds ratio ; CI, confidence interval

<sup>1</sup> Multivariable models were additionally adjusted for age (continuous), BMI (continuous), menopausal status (pre/post-menopause), intervention group of the SU.VI.MAX trial (placebo/supplemented), smoking status (current smokers and non-smokers), season of blood draw (defined above), height (continuous), alcohol intake (continuous), physical activity (irregular/< 1h/d walking equivalent/ $\geq$  1h/d walking equivalent), education level (primary/secondary/superior), family history of breast cancer (yes/no), number of children (continuous), and use of hormone replacement therapy for menopause (yes/no) (reference model). Tests for linear trend were performed using the ordinal score on tertiles of buckets intensity.

<sup>2</sup> Multiple assignments for the same NMR variable due to spectral overlap

**Supplementary Table 7** NMR metabolites associated with breast cancer risk in premenopausal women from multivariable conditional logistic regression<sup>1</sup> - SU.VI.MAX cohort, France (1994-2007)

<b>Metabolites highlighted with NOESY1D experiments</b>								
Assignments	Chemical shift (ppm)	T1 (ref)		T2		T3		P-trend
		cases/controls	OR	cases/controls	OR [95% CI]	cases/controls	OR [95% CI]	
Lipoproteins (CH <sub>2</sub> )	1.27	35/40	1.00	16/42	0.47 [0.21-1.07]	15/38	0.4 [0.17-0.94]	0.02
Lipids (CH <sub>2</sub> -CH <sub>2</sub> CO)	1.56	35/46	1.00	18/35	0.62 [0.26-1.45]	13/39	0.33 [0.14-0.82]	0.02
Lipids (CH <sub>2</sub> CO) & Acetone <sup>2</sup>	2.22	33/40	1.00	19/38	0.5 [0.22-1.13]	14/42	0.31 [0.13-0.76]	0.009
Glycerol-derived compounds (CH <sub>2</sub> )	4.04 & 4.25	31/35	1.00	20/43	0.56 [0.26-1.21]	15/42	0.31 [0.12-0.77]	0.01
Glycerol-derived compounds (CH)	5.18	37/45	1.00	18/37	0.49 [0.21-1.16]	11/38	0.27 [0.1-0.71]	0.006

<b>Metabolites highlighted with CPMG experiments</b>								
Assignments	Chemical shift (ppm)	T1 (ref)		T2		T3		P-trend
		cases/controls	OR	cases/controls	OR [95% CI]	cases/controls	OR [95% CI]	
Lipoproteins (CH <sub>2</sub> )	1.26	32/42	1.00	18/36	0.7 [0.32-1.5]	16/42	0.38 [0.15-0.93]	0.03
Lipids (CH <sub>2</sub> CO) & Acetone <sup>2</sup>	2.22	27/43	1.00	27/32	1.37 [0.63-2.97]	12/45	0.3 [0.11-0.79]	0.04
Glutamine	2.45	21/47	1.00	19/44	1.27 [0.52-3.09]	26/29	2.6 [1.07-6.33]	0.04
Unassigned	3.34	18/50	1.00	24/37	2.76 [1.14-6.67]	24/33	2.62 [1.05-6.57]	0.048

SU.VI.MAX, Supplémentation en Vitamines et Minéraux Antioxydants ; T, tertile ; ref, reference ; OR, odds ratio ; CI, confidence interval

<sup>1</sup> Multivariable models were additionally adjusted for age (continuous), BMI (continuous), menopausal status (pre/post-menopause), intervention group of the SU.VI.MAX trial (placebo/supplemented), smoking status (current smokers and non-smokers), season of blood draw (defined above), height (continuous), alcohol intake (continuous), physical activity (irregular/< 1h/d walking equivalent/≥ 1h/d walking equivalent), education level (primary/secondary/superior), family history of breast cancer (yes/no), number of children (continuous), and use of hormone replacement therapy for menopause (yes/no) (reference model). Tests for linear trend were performed using the ordinal score on tertiles of buckets intensity.

<sup>2</sup> Multiple assignments for the same NMR variable due to spectral overlap

**Supplementary Table 8** NMR metabolites associated with breast cancer risk in postmenopausal women from multivariable conditional logistic regression<sup>1</sup> - SU.VI.MAX cohort, France (1994-2007)

<b>Metabolites highlighted with NOESY1D experiments</b>								
Assignments	Chemical shift (ppm)	T1 (ref)		T2		T3		P-trend
		cases/controls	OR	cases/controls	OR [95% CI]	cases/controls	OR [95% CI]	
Lipids (CH <sub>2</sub> CH <sub>2</sub> CO)	1.56	48/69	1.00	54/89	0.77 [0.45-1.33]	35/99	0.49 [0.27-0.88]	0.02
Lipids (CH <sub>2</sub> CO) & Acetone <sup>2</sup>	2.22	52/71	1.00	50/93	0.67 [0.39-1.14]	35/93	0.5 [0.28-0.9]	0.02
Glycerol-derived compounds (CH)	5.18	50/64	1.00	49/92	0.66 [0.37-1.15]	38/101	0.44 [0.24-0.82]	0.009

<b>Metabolites highlighted with CPMG experiments</b>								
Assignments	Chemical shift (ppm)	T1 (ref)		T2		T3		P-trend
		cases/controls	OR	cases/controls	OR [95% CI]	cases/controls	OR [95% CI]	
Valine	0.98 & 1.03	35/88	1.00	41/85	1.24 [0.69-2.25]	61/84	1.77 [1.01-3.1]	0.04
Undefined region	2.57-2.61	36/86	1.00	42/90	1.11 [0.62-1.99]	59/81	1.82 [1.04-3.18]	0.03
Lysine, Creatine & Creatinine <sup>2</sup>	3.01 & 3.03	33/94	1.00	49/82	1.61 [0.91-2.86]	55/81	1.73 [1-2.98]	0.049

SU.VI.MAX, Supplémentation en Vitamines et Minéraux Antioxydants ; T, tertile ; ref, reference ; OR, odds ratio ; CI, confidence interval

<sup>1</sup> Multivariable models were additionally adjusted for age (continuous), BMI (continuous), menopausal status (pre/post-menopause), intervention group of the SU.VI.MAX trial (placebo/supplemented), smoking status (current smokers and non-smokers), season of blood draw (defined above), height (continuous), alcohol intake (continuous), physical activity (irregular/< 1h/d walking equivalent/ $\geq$  1h/d walking equivalent), education level (primary/secondary/superior), family history of breast cancer (yes/no), number of children (continuous), and use of hormone replacement therapy for menopause (yes/no) (reference model). Tests for linear trend were performed using the ordinal score on tertiles of buckets intensity.

<sup>2</sup> Multiple assignments for the same NMR variable due to spectral overlap

*Additional File 6 – L. Lecuyer et al. NMR metabolomic signatures reveal predictive plasma metabolites associated with long-term risk of developing breast cancer*

### Supplementary results of Principal Component Analysis

**Supplementary Table 9** NMR metabolites loading values for the first component from principal component analysis

Chemical shift (ppm)	Assignments	NMR Experiments	PCA Component 1
0.98 & 1.03	Valine	CPMG	-0.038867
1.26	Lipoproteins (CH <sub>2</sub> )	CPMG	0.248058
1.70	Lysine & Arginine <sup>1</sup>	CPMG	-0.114249
2.22	Lipids (CH <sub>2</sub> CO) & Acetone <sup>1</sup>	CPMG	0.21357
2.33	Unassigned	CPMG	-0.104279
2.45	Glutamine	CPMG	-0.151966
2.55-2.57	Undefined region	CPMG	-0.119194
2.57-2.61	Undefined region	CPMG	-0.077247
3.01 & 3.03	Lysine, Creatine & Creatinine <sup>1</sup>	CPMG	-0.199923
3.34	Unassigned	CPMG	-0.08283
3.36-3.86	Glucose	CPMG	-0.130886
5.18	Unassigned	CPMG	0.263793
5.38-5.43	Undefined region	CPMG	0.20871
5.28	Unsaturated lipids (CH=CH)	NOESY1D	0.270088
5.18	Glycerol-derived compounds (CH)	NOESY1D	0.273545
2.22	Lipids (CH <sub>2</sub> CO) & Acetone <sup>1</sup>	NOESY1D	0.29047
2.001 & 2.035	Lipids (CCH <sub>2</sub> ) & Glycoproteins (N-acetyl) <sup>1</sup>	NOESY1D	0.281568
1.83-1.85	Undefined region	NOESY1D	0.159127
0.86	Lipoproteins (CH <sub>3</sub> )	NOESY1D	0.26042
1.27	Lipoproteins (CH <sub>2</sub> )	NOESY1D	0.29239
1.56	Lipids (CH <sub>2</sub> CH <sub>2</sub> CO)	NOESY1D	0.292266
4.04 & 4.25	Glycerol-derived compounds (CH <sub>2</sub> )	NOESY1D	0.267513

<sup>1</sup> Multiple assignments for the same NMR variable due to spectral overlap



**Supplementary Table 10** NMR metabolites loading values for the second component from principal component analysis

Chemical shift (ppm)	Assignments	NMR Experiments	PCA Component 2
0.98 & 1.03	Valine	CPMG	0.297927
1.26	Lipoproteins (CH <sub>2</sub> )	CPMG	0.204464
1.70	Lysine & Arginine <sup>1</sup>	CPMG	0.333531
2.22	Lipids (CH <sub>2</sub> CO) & Acetone <sup>1</sup>	CPMG	0.212858
2.33	Unassigned	CPMG	0.304811
2.45	Glutamine	CPMG	0.250894
2.55-2.57	Undefined region	CPMG	0.309708
2.57-2.61	Undefined region	CPMG	0.278493
3.01 & 3.03	Lysine, Creatine & Creatinine <sup>1</sup>	CPMG	0.26854
3.34	Unassigned	CPMG	0.284837
3.36-3.86	Glucose	CPMG	0.277569
5.18	Unassigned	CPMG	-0.022763
5.38-5.43	Undefined region	CPMG	-0.232528
5.28	Unsaturated lipids (CH=CH)	CPMG	0.098363
5.18	Glycerol-derived compounds (CH)	NOESY1D	0.125688
2.22	Lipids (CH <sub>2</sub> CO) & Acetone <sup>1</sup>	NOESY1D	0.114236
2.001 & 2.035	Lipids (CCH <sub>2</sub> ) & Glycoproteins (N-acetyl) <sup>1</sup>	NOESY1D	0.120958
1.83-1.85	Undefined region	NOESY1D	0.023759
0.86	Lipoproteins (CH <sub>3</sub> )	NOESY1D	0.096261
1.27	Lipoproteins (CH <sub>2</sub> )	NOESY1D	0.110016
1.56	Lipids (CH <sub>2</sub> CH <sub>2</sub> CO)	NOESY1D	0.108695
4.04 & 4.25	Glycerol-derived compounds (CH <sub>2</sub> )	NOESY1D	0.108481

<sup>1</sup> Multiple assignments for the same NMR variable due to spectral overlap

## Annexe 2 Texte intégral : “Plasma metabolomic signatures associated with long-term breast cancer risk in the SU.VI.MAX prospective cohort”

Research Article

Cancer  
Epidemiology,  
Biomarkers  
& Prevention

## Plasma Metabolomic Signatures Associated with Long-term Breast Cancer Risk in the SU.VI.MAX Prospective Cohort



Lucie Lécuyer<sup>1</sup>, Céline Dalle<sup>2</sup>, Bernard Lyan<sup>2</sup>, Aicha Demidem<sup>3</sup>, Adrien Rossary<sup>3</sup>, Marie-Paule Vasson<sup>3,4</sup>, Mélanie Petera<sup>2</sup>, Marie Lagree<sup>5</sup>, Thomas Ferreira<sup>3</sup>, Delphine Centeno<sup>2</sup>, Pilar Galan<sup>1</sup>, Serge Hercberg<sup>1,6</sup>, Mélanie Deschasaux<sup>1</sup>, Valentin Partula<sup>1</sup>, Bernard Srour<sup>1</sup>, Paule Latino-Martel<sup>1</sup>, Emmanuelle Kesse-Guyot<sup>1</sup>, Nathalie Druesne-Pecollo<sup>1</sup>, Stéphanie Durand<sup>2</sup>, Estelle Pujos-Guillot<sup>2</sup>, and Mathilde Touvier<sup>1</sup>

## Abstract

**Background:** Breast cancer is a major cause of death in occidental women. The role of metabolism in breast cancer etiology remains unclear. Metabolomics may help to elucidate novel biological pathways and identify new biomarkers to predict breast cancer long before symptoms appear. The aim of this study was to investigate whether untargeted metabolomic signatures from blood draws of healthy women could contribute to better understand and predict the long-term risk of developing breast cancer.

**Methods:** A nested case-control study was conducted within the SU.VI.MAX prospective cohort (13 years of follow-up) to analyze baseline plasma samples of 211 incident breast cancer cases and 211 matched controls by LC/MS. Multivariable conditional logistic regression models were computed.

**Results:** A total of 3,565 ions were detected and 1,221 were retained for statistical analysis. A total of 73 ions were associated with breast cancer risk ( $P < 0.01$ ;  $FDR \leq 0.2$ ). Notably,

we observed that a lower plasma level of O-succinyl-homoserine [0.70 (0.55–0.89)] and higher plasma levels of valine/norvaline [1.45 (1.15–1.83)], glutamine/isoglutamine [1.33 (1.07–1.66)], 5-aminovaleic acid [1.46 (1.14–1.87)], phenylalanine [1.43 (1.14–1.78)], tryptophan [1.40 (1.10–1.79)],  $\gamma$ -glutamyl-threonine [1.39 (1.09–1.77)], ATBC [1.41 (1.10–1.79)], and pregnene-triol sulfate [1.38 (1.08–1.77)] were associated with an increased risk of developing breast cancer during follow-up.

**Conclusion:** Several prediagnostic plasmatic metabolites were associated with long-term breast cancer risk and suggested a role of microbiota metabolism and environmental exposure.

**Impact:** After confirmation in other independent cohort studies, these results could help to identify healthy women at higher risk of developing breast cancer in the subsequent decade and to propose a better understanding of the complex mechanisms involved in its etiology.

<sup>1</sup>Sorbonne Paris Cité Epidemiology and Statistics Research Center (CRESS), French National Institute of Health and Medical Research (Inserm) U1153, French National Institute for Agricultural Research (Inra) U1125, French National Conservatory of Arts and Crafts (Cnam), Paris 13 University, Nutritional Epidemiology Research Team (EREN), Bobigny, France. <sup>2</sup>Clermont Auvergne University, INRA, UNH, Plateforme d'Exploration du Métabolisme, MetaboHUB Clermont, Clermont-Ferrand, France. <sup>3</sup>Clermont Auvergne University, INRA, UMR 1019, Human Nutrition Unit (UNH), CRNH Auvergne, Cellular Micro-Environment, Immunomodulation and Nutrition (ECREIN), Clermont-Ferrand, France. <sup>4</sup>Anti-cancer Center Jean-Perrin, CHU Clermont-Ferrand, France. <sup>5</sup>Clermont Auvergne University, Institut de Chimie de Clermont-Ferrand, Plateforme d'Exploration du Métabolisme, MetaboHUB-Clermont, BP 80026, Aubière, France. <sup>6</sup>Public Health Department, Avicenne Hospital, Bobigny, France.

**Note:** Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (<http://cebp.aacrjournals.org/>).

Clinical Trials.gov registration ID: SU.VI.MAX, NCT00272428. Registered on January 3, 2006.

**Corresponding Author:** Lucie Lécuyer, Inserm - U1153 team 3/INRA U1125/Paris 13 University, 74 rue Marcel Cachin, Bobigny Cedex 93000, France. Phone: 331-4838-7638; Fax: 331-4838-8931; E-mail: [llecuyer@eren.smbh.univ-paris13.fr](mailto:llecuyer@eren.smbh.univ-paris13.fr)

Cancer Epidemiol Biomarkers Prev 2019;XX:XX-XX

doi: 10.1158/1055-9965.EPI-19-0154

©2019 American Association for Cancer Research.

## Introduction

Breast cancer is the most common female cancer in the world ([www.dietandcancerreport.org](http://www.dietandcancerreport.org)) and ranks as the fifth cause of death from cancer overall (626,679 deaths worldwide in 2018; ref. 1). Despite some well-known risk factors such as adult height, reproductive life, and hormonal factors, and, for postmenopausal women, alcoholic drinks, body fatness, and adult weight gain ([www.wcrf.org/breast-cancer-2017](http://www.wcrf.org/breast-cancer-2017)), the role of metabolism in breast cancer etiology remains unclear and hypothesis-driven research has not yet allowed to entirely elucidate it.

Recent advances in modern explorative technologies, such as metabolomics, increase the capacity to identify additional risk factors and better understand mechanisms involved (2). Indeed, mass spectrometry (MS) and NMR allows detecting a wide range of low molecular weight biochemicals in biofluids, cells, or tissues (3) and thus can provide a signature of both exogenous exposures and endogenous metabolism. Therefore, metabolomics allows identifying biomarkers relevant to cancer etiology and elucidating novel biological pathways (4–6). Several cross-sectional epidemiologic studies highlighted differences between metabolomic profiles and patients with breast cancer and

controls (7, 8). An NMR-based prospective study (5 years of follow-up) recently showed the benefit of metabolomic signatures of breast cancer risk (9). However metabolomic signatures may be altered by the presence of a breast tumor itself (10), arguing for similar investigations with a long follow-up. Improving the ability to discriminate profiles at higher risk of developing breast cancer in the subsequent decades through the identification of prediagnostic biomarkers, to establish preventive approaches, remains an important focus in breast cancer research.

Two recent semiuntargeted MS studies investigated the associations between diet- (11) or body mass index (BMI)-related metabolites (12) and long-term breast cancer risk, and found promising results supporting the role of diet or BMI in postmenopausal breast cancer etiology. We previously set-up a prospective nested case-control study within the SU.VI.MAX cohort to investigate whether baseline metabolomic profiles from plasma of apparently healthy women could contribute to identify women who would develop breast cancer during the subsequent years, and better understand the etiology of this complex disease. Using an NMR metabolomics analysis, higher fasting plasma levels of valine, lysine, glutamine, and glucose and lower plasma levels of lipoproteins and glycerol-derived compounds were notably found to be associated with a higher risk of developing breast cancer (13). To increase the number of metabolites detected both in terms of sensitivity and metabolome coverage, this study applied for the first time a fully untargeted and nonoriented MS analysis, with the objective to highlight new endogenous or exogenous metabolites associated with long-term risk of breast cancer.

## Materials and Methods

### Population study

This analysis was based on a case-control study nested within the SU.VI.MAX prospective cohort (clinicaltrials.gov; NCT00272428). The latter was initially designed to investigate the influence of a daily supplementation with nutritional doses of antioxidants on the incidence of cardiovascular diseases and cancers. The study design and methods have been previously detailed (14, 15). Briefly, this study is a population-based, double-blind, placebo-controlled, randomized trial enrolling a total of 13,017 participants recruited in 1994–1995. The intervention study lasted 8 years, and observational follow-up of health events was maintained until September 2007. Written informed consent was obtained from all participants. The trial was approved by the Ethics Committee for Studies with Human Subjects of Paris-Cochin Hospital (CCPPRB 706/2364) and the 'Commission Nationale de l'Informatique et des Libertés' (CNIL 334641/907094) and was conducted according to the Declaration of Helsinki guidelines.

### Baseline data collection

At inclusion, self-administered questionnaires about sociodemographic characteristics, smoking status, medication use, health status, and family history of cancer were filled-in by participants. During a baseline clinical examination, all participants underwent anthropometric measurements (height and weight) and a blood draw (occurring after a 12-hour fasting period) by study nurses and physicians. These 35 mL venous blood samples were collected in sodium heparin Vacutainer Tubes (Becton Dickinson,

son). After centrifugation at +4°C, plasma aliquots were immediately prepared and stored frozen at –20°C during less than 2 days and then stored in liquid nitrogen.

### Case ascertainment

Health events that occurred during follow-up (from 1994 to 2007) were self-reported by participants. Medical data were gathered through participants, physicians, and hospitals and reviewed by an independent physician expert committee. Pathologic reports were used to validate cases and extract cancer characteristics. Cancers were classified by using the International Classification of Diseases, 10th Revision, Clinical Modification (16).

### Nested case-control study

This analysis initially included 215 participants with a first incident invasive breast cancer matched 1:1 with controls for baseline age (35–39 years/40–44 years/45–49 years/50–54 years/55–59 years/>60 years), menopausal status (pre-/postmenopause status at baseline and at diagnosis), BMI (<18.5 kg/m<sup>2</sup>, underweight; ≥18.5–<25 kg/m<sup>2</sup>, normal weight; and ≥25 kg/m<sup>2</sup>, overweight/obese), arm assignment to the initial SU.VI.MAX clinical trial (placebo/supplemented), smoking status (current smokers and non/former smokers at baseline), and season of blood draw (*a priori*-defined periods: October–November/December–January–February/March–April–May). Controls selection was based on the classical density sampling method (17), that is, every time a case was diagnosed, one control was selected from other participants from the cohort with no breast cancer at that time.

### Metabolomic analyses

Metabolomic profiling was conducted using ultrahigh performance liquid chromatography/mass spectroscopy with the Metabolic Profiler Platform (Bruker Daltonique). Details on chemicals and reagents, biological samples preparation, metabolic profiling, raw data extraction, and metabolite identification are available in Supplementary Materials and Methods and Supplementary Fig. S1. Samples were randomized within the analytic sequence based on a Williams Latin Square strategy defined according to the main factors of the study. The stability of the analytic system was monitored using pooled samples as quality control and analyzed in the same sequence than participant samples (injected one time at the beginning of each sequence and then after each set of 12 samples). This metabolomic discovery approach allows semiquantitative measurements, representing relative ion intensities in arbitrary unit. However, even if it is not absolute quantitation, there are linear relations between ion intensity for each metabolite and its concentration level in the samples.

### Statistical analysis

Participants' baseline characteristics were compared between cases and controls using  $\chi^2$  tests for categorical variables and Student *t* test for continuous variables. Less than 5% of values were missing for all covariates and missing values were replaced by the mode.

Prospective associations between each ion's signal intensity and breast cancer risk were characterized using multivariable conditional logistic regression models. OR for an increment of 1 SD for each metabolite and their 95% confidence intervals (95%

CI) were computed, as well as tests for linear trend. Multivariable models were adjusted for the following baseline characteristics: age (continuous), BMI (continuous), smoking status (current, former, and nonsmokers), height (continuous), alcohol intake (continuous), physical activity (low, moderate, and intense), education level (primary, secondary, and superior), family history of breast cancer, number of children, and use of hormone replacement therapy for menopause; all these covariates constituting the reference model. Models were adjusted for matching factors only when the variable was more precise than the one used for the matching, to avoid residual confounding. The models were not additionally adjusted for menopausal status (pre-/postmenopause status at baseline and at diagnosis), season of blood draw (October–November/December–January–February/March–April–May), and arm assignment to the initial SU.VI. MAX clinical trial (yes/no) because cases and controls were already matched for these variables with the same coding. To limit the false negative rate (type II error = risk of not detecting real associations) in this exploratory untargeted metabolomic study, a FDR level of 0.2 was retained from the Benjamini Hochberg procedure (ref. 18; 1,221 tests: 530 ions in ESI-positive mode and 691 ions in ESI-negative mode). Linearity assumption was checked by plotting the ion's intensity as a function of logit(predicted values of breast cancer risk). ORs for categorized intensities in tertiles were also computed, as well as a Spearman correlation matrix of the identified associated metabolites.

Sensitivity analysis was performed on the same models by excluding cases diagnosed during the first year of follow-up, as well as the matched controls to prevent reverse causality bias. Exploratory stratified and heterogeneity (interaction) analyses were also performed according to menopausal status at diagnosis and according to the median follow-up duration: 5.8 years (to investigate differentially short- or long-term associations with breast cancer risk). Interaction was tested as the produce of each ion (continuous/SD) by these two factors (two categories each).

A multivariate analysis using a principal component analysis (PCA) with rotated factors by orthogonal transformation was performed on ions associated with breast cancer from logistic regressions. The four first factors were retained according to the plot of the total explained variance (Cattel scree plots; ref. 19). The multivariable conditional logistic regression models were run on these factors (with the same adjustments as described above) to investigate the associations between metabolomic patterns (combinations of metabolites variations) and breast cancer risk.

Analyses were performed using SAS software (v9.3) and plots were generated using R software (v3.5.2). All of the statistical tests were two-sided.

#### Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

#### Informed consent

Informed consent was obtained from all individual participants included in the study.

## Results

The baseline characteristics of breast cancer cases and controls in the study population are summarized in Table 1. The cases were more likely to be taller with more frequent family history of breast cancer than controls. A total of 25% of cases were diagnosed before 3.5 years, 50% before 5.8 years, and 75% before 9.4 years. The maximum length of follow-up was 12.8 years.

A total of 3,565 ions were detected and 1,221 were retained for statistical analysis. From logistic regression analyses, 53 ions were positively associated with breast cancer risk and 20 ions were inversely associated ( $P \leq 0.01$ ;  $FDR \leq 0.2$ ). Results of all associated ions are shown in Supplementary Table S1. Among these ions, some have been identified using available databases (in-house database, Metlin; <https://metlin.scripps.edu>), Human Metabolome Database ([www.hmdb.ca](http://www.hmdb.ca)), Kyoto Encyclopedia of Genes and Genomes database), and analytic data (as MS2); the associations between these ions ( $N = 9$ ) and breast cancer risk are displayed in Table 2. A lower plasma level of O-succinyl-homoserine [0.70 (0.55–0.89)] and higher plasma levels of valine/norvaline [1.45 (1.15–1.83)], glutamine/isoglutamine [1.33 (1.07–1.66)], 5-aminovaleric acid [1.46 (1.14–1.87)], phenylalanine [1.43 (1.14–1.78)], tryptophan [1.40 (1.10–1.79)],  $\gamma$ -glutamyl-threonine [GGT; 1.39 (1.09–1.77)], acetyl tributyl citrate ATBC [1.41 (1.10–1.79)], and the putative pregnenetriol sulfate [1.38 (1.08–1.77)] were associated with an increased risk of developing breast cancer during follow-up. Linearity assessments, as well as analyses performed with ions categorized in tertiles were provided in Supplementary Fig. S2A and S2B and Supplementary Table S2. The Spearman correlation matrix of the identified associated metabolites is provided in Supplementary Table S3. Positive correlations ( $P < 0.05$ ) were found for pregnenetriol sulfate with 5-aminovaleric acid, tryptophan, and GGT; for phenylalanine with valine, tryptophan, 5-aminovaleric acid, and glutamine; for valine with 5-aminovaleric acid, tryptophan, and glutamine; for 5-aminovaleric acid with GGT; for tryptophan with glutamine; and for GGT with glutamine. Inverse correlations were found for O-succinyl-homoserine with 5-aminovaleric acid and ATBC; and for 5-aminovaleric acid with ATBC. Results of sensitivity, stratification, and heterogeneity analyses are shown in Supplementary Tables S4, S5, and S6. All identified ions remained associated with breast cancer when excluding cases diagnosed before 1 year of follow-up. No interaction was found between breast cancer risk and duration of follow-up. Higher plasma level of glutamine was associated with an increased risk of breast cancer only in premenopausal women ( $P_{\text{interaction}} = 0.003$ ;  $FDR = 0.2$ ).

All ions associated with breast cancer risk in logistic regressions were then analyzed by PCA. The four first PCA factors (metabolomic patterns) explained more than 36% of total variability. Loading values are displayed in Supplementary Table S7. The first factor was notably characterized by higher plasma levels of valine and 5-aminovaleric acid, the second factor was in particular characterized by higher plasma levels of phenylalanine, valine, 5-aminovaleric acid, and tryptophan, the third by lower plasma level of pregnenetriol sulfate, and the fourth by higher plasma levels of phenylalanine, valine 5-aminovaleric acid, GGT, and glutamine (loading values  $\geq |0.20|$ ). Moreover, several unidentified ions also characterized these four factors. These data allowed distinguishing three metabolomic patterns (factors 1, 2, and 4)

**Table 1.** Baseline characteristics of breast cancer cases and controls<sup>a</sup>, SU.VI. MAX cohort, France (1994–2007)

	Breast cancer cases (N = 211)	Controls (N = 211)	P <sup>b</sup>
Age at baseline (y)	49.0 ± 6.0	48.9 ± 6.0	0.9
BMI (kg/m <sup>2</sup> )	23.2 ± 4.0	23.6 ± 4.2	0.3
<18.5 kg/m <sup>2</sup> (underweight)	9 (4.3)	6 (2.8)	0.7
≥18.5–<25 kg/m <sup>2</sup> (normal weight)	148 (70.1)	151 (71.6)	
≥25 kg/m <sup>2</sup> (overweight)	54 (25.6)	54 (25.6)	
Height (cm)	163.0 ± 6.2	160.9 ± 5.9	0.0004
Intervention group			1
Placebo	105 (49.8)	105 (49.8)	
Antioxidants	106 (50.2)	106 (50.2)	
Smoking status			1
Never and former	164 (77.7)	164 (77.7)	
Current smoker	47 (22.3)	47 (22.3)	
Physical activity			0.8
Irregular	67 (31.8)	60 (28.4)	
<1 h/d walking equivalent	69 (32.7)	72 (34.1)	
≥1 h/d walking equivalent	75 (35.5)	79 (37.4)	
Educational level			0.3
Primary	37 (17.5)	44 (20.9)	
Secondary	82 (38.9)	91 (43.1)	
Superior	92 (43.6)	76 (36.0)	
Number of biological children	1.8 ± 1.2	2 ± 1.2	0.1
Hormonal treatment for menopause (yes)	76 (36.0)	75 (35.5)	0.9
Menopausal status at baseline			1
Premenopausal	129 (61.1)	129 (61.1)	
Postmenopausal	82 (38.9)	82 (38.9)	
Menopausal status at diagnosis			1
Premenopausal	78 (37.0)	78 (37.0)	
Postmenopausal	133 (63.0)	133 (63.0)	
Family history of breast cancer <sup>c</sup> (yes)	35 (16.6)	21 (10.0)	0.04
Alcohol intake (g/d)	11.2 ± 12.0	11.5 ± 13.0	0.8
Month of blood draw			1
March–April–May	77 (36.5)	77 (36.5)	
October–November	31 (14.7)	32 (15.2)	
December–January–February	103 (48.8)	102 (48.3)	

<sup>a</sup>Values are means ± SD or n (%).

<sup>b</sup>P for the comparison between breast cancer cases and controls using  $\chi^2$  tests for categorical variables or Student *t* test (from analysis of variance models) for continuous variables.

<sup>c</sup>Among first-degree female relatives.

associated with an increased risk of breast cancer in multivariable conditional logistic regression models and one (factor 3) with a decreased risk (FDR < 0.05; Table 3).

## Discussion

To our knowledge, this prospective study was the first fully untargeted and nonoriented LC/MS-based analysis aiming to investigate the associations between prediagnosis metabolic profiles and long-term breast cancer risk. Untargeted metabolomics, by the detection of a large number of metabolites, allowed discriminating early phenotypes and revealing a plasma metabolic signature announcing the evolution toward breast cancer in the subsequent 13 years. This could be relevant for detecting specific plasmatic profiles of women at higher risk for this disease and improve personalized prevention. A profile (signature) of circulating levels of multiple metabolites provides insights to draw an overall picture of the different pathways involved in the

etiology of this cancer (20). Even if the origin of metabolites variations (exogenous or endogenous) cannot be determined in this study, our results also illustrate that metabolomics integrates information regarding the influence of both extrinsic and intrinsic factors, such as lifestyle exposures and basal metabolism, and consequently address the complex issue of environment and health interaction.

In this study, we found that a lower plasma level of O-succinyl-homoserine and higher plasma levels of valine/norvaline, glutamine/isoglutamine, 5-aminovaleric acid, phenylalanine, tryptophan,  $\gamma$ -glutamyl-threonine, ATBC, and pregnene-triol sulfate were associated with an increased risk of developing breast cancer during the following decade. Three metabolomic patterns characterized by several concurrently metabolites variations were associated with an increased risk of breast cancer, as well as another metabolomic pattern associated with a decreased risk. These univariate analysis (analysis of one by one metabolite variation) and multivariate analysis (analysis of combinations of simultaneous metabolites variations) allowed getting complementary overviews of metabolic signatures discriminating phenotypes at higher risk for this disease well before breast cancer clinical onset.

We previously highlighted a higher plasma level of glutamine and valine in the future breast cancer cases with NMR analyses on the same population study (13), which was consistent with our current results. An increase in plasmatic levels of valine and three  $\gamma$ -glutamyl dipeptides was also recently found to be positively associated with BMI and estrogen receptor–positive breast cancer risk in postmenopausal women through a semiuntargeted MS study (12). The pregnene-triol sulfate is a steroid hormone and belongs to progestative family compounds. In line with our results, two semiuntargeted MS studies found a positive association between steroid hormones and breast cancer risk in postmenopausal women (11, 12). Four metabolomic studies reported that an increased level of phenylalanine was positively associated with hepatocellular carcinoma (HCC) risk (21–24) and three of these studies found that a higher level of glutamine was associated with a decreased risk of HCC (21–23). Concerning valine, findings were more contrasted; one of these studies showed a positive association with HCC risk (21), whereas another observed an inverse association (22). An increase in plasmatic valine level was also reported in two metabolomic studies to be positively associated with pancreatic adenoma risk (25) and inversely associated with prostate cancer risk (26). In three metabolomic studies an inverse association was found between phenylalanine (26), tryptophan (27),  $\gamma$ -glutamyl-histidine (28), and  $\gamma$ -glutamyl-phenylalanine (26, 27) and prostate cancer risk, whereas a positive association was reported between  $\gamma$ -glutamyl-glutamine and nonaggressive prostate cancer risk (27). The comparisons of our results with other prospective metabolomic studies should be interpreted with caution, taking into account methodologic and technological variations (as the wide range of metabolomics platforms and types of analyses), because the etiology may vary according to cancer locations.

The findings highlighted in this study provide evidence for circulating metabolites as putative risk biomarkers for breast cancer. Given that metabolites concerned by the observed variations are not specific of one metabolic pathway and as their variations could arise from several types of exposure and endogenous reactions, precaution is needed in the mechanistic interpretation of our epidemiologic findings. Even so, we propose

**Table 2.** Identified ions associated with breast cancer risk from multivariable conditional logistic regression<sup>a</sup>, SU.VI.MAX cohort, France (1994–2007)

Name (level of confidence for identification <sup>b</sup> )	Annotation	Measured mass m/z	Retention time (minute)	Mode	Cases/controls	Mean ± SD	OR (95% CI)	P <sub>trend</sub>
L-phenylalanine (1)	[M+H] <sup>+</sup>	166.08618	10.11	ESI+	211/211	227,718 ± 42,535	1.43 (1.14–1.78)	0.002 <sup>c</sup>
L-valine/norvaline (1)	[M+H-CH <sub>2</sub> O <sub>2</sub> ] <sup>+</sup>	72.08089	3.08	ESI+	211/211	115,033 ± 24,435	1.45 (1.15–1.83)	0.002 <sup>c</sup>
5-aminovaleric acid (1)	[M+Na] <sup>+</sup>	140.06816	2.29	ESI+	211/211	10,219 ± 5,139	1.46 (1.14–1.87)	0.003 <sup>d</sup>
O-succinyl-L-homoserine (1)	[M+Na-2H] <sup>-</sup>	240.04755	2.86	ESI-	203/203	919 ± 400	0.70 (0.55–0.89)	0.004 <sup>d</sup>
ATBC (acetyl tributyl citrate) (1)	[M+Na] <sup>+</sup>	425.21694	19.31	ESI+	211/211	2,198 ± 964	1.41 (1.10–1.79)	0.006 <sup>d</sup>
L-tryptophan (1)	[M+Na] <sup>+</sup>	227.07887	11.09	ESI+	211/211	799 ± 180	1.40 (1.10–1.79)	0.007 <sup>d</sup>
L-γ-glutamyl-L-threonine (1)	[M+H-H <sub>2</sub> O] <sup>+</sup>	231.1451	2.52	ESI+	211/211	1,054 ± 601	1.39 (1.09–1.77)	0.008 <sup>d</sup>
L-glutamine/L-isoglutamine (1)	[M+H-C <sub>4</sub> H <sub>8</sub> O] <sup>+</sup>	90.0551	2.32	ESI+	211/211	13,187 ± 2,982	1.33 (1.07–1.66)	0.009 <sup>d</sup>
Pregnene-triol sulfate (2)	C <sub>21</sub> H <sub>33</sub> O <sub>6</sub> S	413.19771	16.11	ESI-	203/203	2,441 ± 1,254	1.38 (1.08–1.77)	0.01 <sup>d</sup>

<sup>a</sup>Multivariable models were adjusted for age (continuous), BMI (continuous), smoking status (current smokers, former smokers, and nonsmokers), season of blood draw (*a priori*-defined periods: October–November/December–January–February/March–April–May), height (continuous), alcohol intake (continuous), physical activity (irregular/<1 h/d walking equivalent/≥1 h/d walking equivalent), education level (primary/secondary/superior), family history of breast cancer (yes/no), number of children (continuous), and use of hormone replacement therapy for menopause (yes/no; reference model). Tests for linear trend were performed using the continuous variables. ORs were presented for an increment of 1 SD for each metabolite.

<sup>b</sup>Metabolites were classified according to Sumner and colleagues (46) concerning the levels of confidence in the identification process: identified (level 1: confirmed by standard), putatively annotated (level 2: based upon physicochemical properties and/or spectral similarity with public/commercial spectral libraries).

<sup>c</sup>FDR after Benjamini Hochberg correction for multiple testing = 0.1.

<sup>d</sup>FDR after Benjamini Hochberg correction for multiple testing = 0.2.

interpretation hypotheses for some presumed pathways, summarized in Fig. 1. Among the possible interpretations, our results could suggest a perturbation of energy and amino acids metabolism, a set-up of conditions favoring cell proliferation, inflammation, and oxidative stress in women who will develop breast cancer involving environmental exposure and microbiota metabolism. Glutamine and valine (one of the branched chain amino acids; BCAA) positively correlated in our study, are two glycolytic amino acids; their plasma increase could suggest an amplified phenomenon of gluconeogenesis. BCAAs were previously suggested as markers of cardiovascular disease risk (29), insulin resistance (30), and diabetes (31). Moreover there is convincing evidence on the effect of obesity on an increased risk of post-

menopausal breast cancer, and pathways related to insulin resistance is one of the number of plausible mechanisms that may explain this relationship (32). Moreover, valine and glutamine are regulators of many cell signaling pathways including activators of the mTOR pathway; one of the main cell proliferation routes (31). ATBC is an alternative plasticizer to phthalates (33) widely used in products such as food wrap, vinyl toys, and pharmaceutical excipients. Although this contaminant does not seem to accumulate in tissues (34), recent data revealed potential biological activity on tissue growth (35). The combination of plasmatic increase in both ATBC and in a progestative sulfate steroid may interact either with PPAR signaling involved in cell proliferation and inflammation, and/or with steroid signaling (36). Indeed, ATBC has been shown to influence steroid levels (35) and is known to interact with several signaling pathways as a ligand for steroid xenobiotic receptor or sexual hormone-binding globulin involved in steroid signaling (33). Phenylalanine and tryptophan are aromatic amino acids (AAA, correlated in our study;  $R = 0.27$ ) and high level of AAAs in blood have been considered as an inflammatory marker in patients with liver cirrhosis (37). Disruptions in AAAs and BCAAs levels are found in several pathologies with chronic low-grade inflammation such as obesity and diabetes (38, 39); in our study valine (BCAA) is highly positively correlated with AAAs (phenylalanine;  $R = 0.47$  and tryptophan;  $R = 0.24$ ). Data on the kynurenine/tryptophan ratio, unfortunately unavailable in this study, would be useful to support this hypothesis, as it reflects IDO1 gene expression, which increases in situations of chronic low-grade inflammation. 5-aminovaleric acid is a catabolism product from lysine; its increase could witness either the presence of cellular necrosis or microbiota metabolism. In mice, higher urinary levels of 5-aminovaleric acid were associated with intestinal inflammation (40). Concerning identified metabolites, the second PCA-derived factor highlighted in our study could represent an inflammatory pattern, as it was notably characterized by higher plasma levels of phenylalanine, valine, 5-aminovaleric acid, and tryptophan. Our findings could also reflect an influence of gut microbiota, because O-succinyl-homoserine is a metabolite exclusively synthesized by bacteria. Another metabolite synthesized by gut microbiota, 2-amino-4-

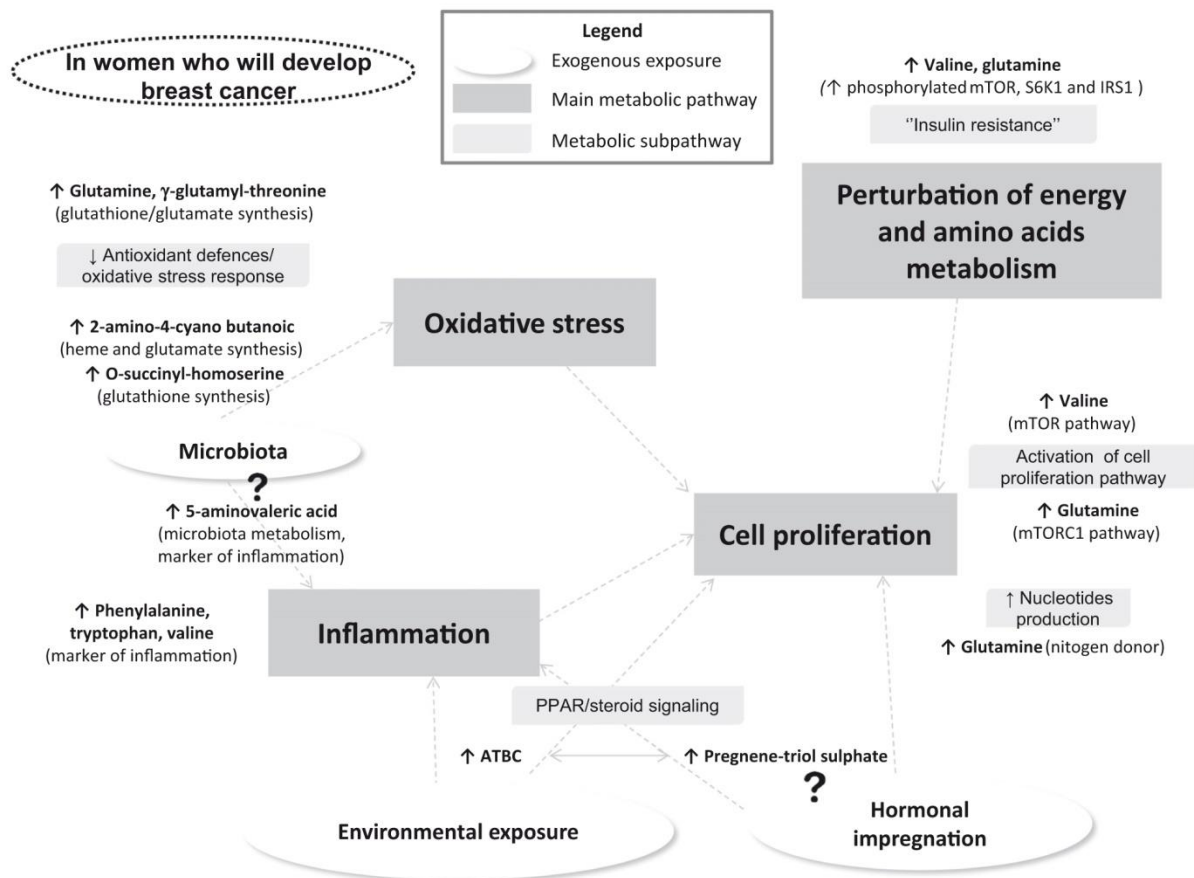
**Table 3.** Associations between the metabolomic patterns derived by PCA and breast cancer risk from multivariable conditional logistic regression<sup>a</sup>, SU.VI.MAX cohort, France (1994–2007)

PCA Factors <sup>b</sup>	Cases/controls	OR (95% CI)	P <sub>trend</sub>	Corrected P <sub>trend</sub> <sup>c</sup>
Factor 1	200/200	1.38 (1.11–1.73)	0.004	0.004
Factor 2	200/200	1.47 (1.16–1.85)	0.001	0.002
Factor 3	200/200	0.49 (0.37–0.65)	<0.0001	<0.0001
Factor 4	200/200	1.49 (1.17–1.89)	0.001	0.002

<sup>a</sup>Multivariable models were adjusted for age (continuous), BMI (continuous), smoking status (current smokers, former smokers, and nonsmokers), season of blood draw (*a priori*-defined periods: October–November/December–January–February/March–April–May), height (continuous), alcohol intake (continuous), physical activity (irregular/<1 h/d walking equivalent/≥1 h/d walking equivalent), education level (primary/secondary/superior), family history of breast cancer (yes/no), number of children (continuous), and use of hormone replacement therapy for menopause (yes/no; reference model). Tests for linear trend were performed using the continuous variables. ORs are computed for a 1 point increment of the continuous variable.

<sup>b</sup>The four first PCA factors (metabolomic patterns) explained more than 36% of total variability. The first factor was notably characterized by higher plasma levels of valine and 5-aminovaleric acid, the second factor was in particular characterized by higher plasma levels of phenylalanine, valine, 5-aminovaleric acid, and tryptophan, the third by lower plasma level of pregnene-triol sulfate, and the fourth by higher plasma levels of phenylalanine, valine 5-aminovaleric acid, GGT, and glutamine (loading values ≥|0.20|). Moreover, several unidentified ions also characterized these four factors.

<sup>c</sup>P<sub>trend</sub> after Benjamini Hochberg correction for multiple testing.



**Figure 1.**

Summary of proposed mechanistic interpretations for some putative pathways regarding observed metabolites variations in plasma of women who will develop breast cancer in subsequent years. Among the possible interpretations, our results could suggest a perturbation of energy and amino acids metabolism, a set-up of conditions favoring cell proliferation, inflammation, and oxidative stress in women who will develop breast cancer involving environmental exposure and microbiota metabolism.

cyano butanoic acid, was associated with increased breast cancer risk in this study before FDR correction only [OR = 1.28 (1.03–1.58);  $P = 0.02$ ; FDR = 0.3; data not tabulated]. The variation of these two microbiota-related metabolites could be interpreted as an oxidative stress response given their potential role in glutamate and glutathione synthesis. Indeed, humans' metabolism can produce cysteine (one component of glutathione) using O-succinyl-homoserine instead of homocysteine (41). 2-amino-4-cyano butanoic acid is the substrate of nitrilase responsible for the *de novo* synthesis of glutamate in bacteria and fungus (42). Moreover it is also implied in heme synthesis, recognized as an oxidative biocompound (43). Likewise, the increase in plasmatic glutamine and in  $\gamma$ -glutamyl-threonine (positively correlated in our study) can suggest both a decrease in antioxidant defenses and an oxidative stress response ( $\gamma$ -glutamyl-dipeptides may indicate increased glutathione synthesis). The data available in this study do not allow to draw firm conclusions about these mechanistic interpretations, which need further investigations for their validation, such as complementary analyses of other metabolites variation involved in the same pathway (but not detected with the current analytic technic), fluxomic analyses to determine the rates

of metabolic reactions, or lipidomic and proteomic analyses to cover a wide range of biggest molecules.

The major strengths of this study pertained to the fully untargeted MS metabolomics approach and the prospective design. However, several limitations should be acknowledged. First, the relatively small sample size was probably sufficient for main analyses, but it was certainly limited for secondary stratified analyses. For instance, a larger sample size could have helped to investigate more precisely the relation between the metabolites and the duration before cancer development (e.g., short, medium, and long term). Second, we cannot rule out the possibility of residual or unmeasured confounding despite the large spectrum of covariates taken into account, and we cannot exclude the possibility of overfitting of the models. Third, because the size of the sample did not allow keeping apart a validation set, we did not compute predictive performance such as AUC. However the main interest of our study lied in a better understanding of breast cancer etiology rather than in an improvement in discrimination models. Then, only LC/MS and not gas chromatography–mass spectroscopy was applied. This may have limited our ability to detect some categories of metabolites which could have

supported certain mechanistic interpretations. Moreover, as the metabolic profiles were established a decade before the diagnosis, some small variations may not have been detected. Some of the metabolites highlighted in our study are at the cross-roads of several metabolic pathways, thus these mechanistic interpretations resulting from literature should be considered with caution and need further investigations to be validated. Other interpretations could also be considered. Furthermore, this metabolomic analysis was based on a single blood draw so we could not investigate the stability of the profiles. Nevertheless, several studies showed a good stability and reproducibility of metabolomic measurements for most of metabolites (44, 45). Finally, as it is usually the case in a fully untargeted metabolomic approach, several associated ions could not be identified to date. However, these findings will be useful in the future in a context of strong development of metabolites databases and to compare our results with other future LC/MS studies.

In conclusion, this prospective study highlighted associations between baseline untargeted metabolomic profiles and subsequent long-term breast cancer risk. These findings generate mechanistic hypotheses regarding the etiology of this complex disease, which were consistent with those reported in our NMR analysis, and include a possible implication of gut microbiota activity and environmental contamination/exposure. If replicated in other independent prospective studies and after quantification of key metabolites to set discrimination thresholds, these results might contribute to develop screening strategies for the identification, before tumor set up, of women at higher risk of developing a breast cancer in the decade following the blood draw.

## Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

## Disclaimer

The funders had no role in the design, analysis, or writing of this article.

## References

- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68:394–424.
- Lindon JC, Nicholson JK. The emergent role of metabolic phenotyping in dynamic patient stratification. *Expert Opin Drug Metab Toxicol* 2014;10:915–9.
- Suhre K, Shin SY, Petersen AK, Mohny RP, Meredith D, Wagle B, et al. Human metabolic individuality in biomedical and pharmaceutical research. *Nature* 2011;477:54–60.
- Norvig P, Relman DA, Goldstein DB, Kammen DM, Weinberger DR, Aiello LC, et al. 2020 visions. *Nature* 2010;463:26–32.
- Holmes E, Loo RL, Stamler J, Bictash M, Yap IK, Chan Q, et al. Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature* 2008;453:396–400.
- Claudino WM, Quattrone A, Biganzoli L, Pestrin M, Bertini I, Di LA. Metabolomics: available results, current research projects in breast cancer, and future applications. *J Clin Oncol* 2007;25:2840–6.
- Budczies J, Denkert C, Muller BM, Brockmoller SF, Klauschen F, Gyorffy B, et al. Remodeling of central metabolism in invasive breast cancer compared to normal breast tissue - a GC-TOFMS based metabolomics study. *BMC Genomics* 2012;13:334.
- Henneges C, Bullinger D, Fux R, Friese N, Seeger H, Neubauer H, et al. Prediction of breast cancer by profiling of urinary RNA metabolites using Support Vector Machine-based feature selection. *BMC Cancer* 2009;9:104.
- Bro R, Kamstrup-Nielsen MH, Engelsen SB, Savorani F, Rasmussen MA, Hansen L, et al. Forecasting individual breast cancer risk using plasma metabolomics and biocontours. *Metabolomics* 2015;11:1376–80.
- Brauer HA, Makowski L, Hoadley KA, Casbas-Hernandez P, Lang LJ, Roman-Perez E, et al. Impact of tumor microenvironment and epithelial phenotypes on metabolism in breast cancer. *Clin Cancer Res* 2013;19:571–85.
- Playdon MC, Ziegler RG, Sampson JN, Stolzenberg-Solomon R, Thompson HJ, Irwin ML, et al. Nutritional metabolomics and breast cancer risk in a prospective study. *Am J Clin Nutr* 2017;106:637–49.
- Moore SC, Playdon MC, Sampson JN, Hoover RN, Trabert B, Matthews CE, et al. A metabolomics analysis of body mass index and postmenopausal breast cancer risk. *J Natl Cancer Inst* 2018;110:djx244.
- Lécuyer L, Victor Bala A, Deschasaux M, Bouchemal N, Nawfal Triba M, Vasson M-P, et al. NMR metabolomic signatures reveal predictive plasma metabolites associated with long-term risk of developing breast cancer. *Int J Epidemiol* 2018;47:484–94.
- Hercberg S, Galan P, Preziosi P, Bertrais S, Mennen L, Malvy D, et al. The SU.VI.MAX study: a randomized, placebo-controlled trial of the health effects

## Authors' Contributions

**Conception and design:** L. Lécuyer, M. Touvier

**Development of methodology:** L. Lécuyer, E. Pujos-Guillot, M. Touvier

**Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.):** B. Lyan, M. Petera, D. Centeno, P. Galan, S. Hercberg, E. Kesse-Guyot, M. Touvier

**Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis):** L. Lécuyer, C. Dalle, B. Lyan, A. Demidem, A. Rossary, M.-P. Vasson, M. Petera, D. Centeno, M. Deschasaux, V. Partula, B. Srour, P. Latino-Martel, N. Druesne-Pecollo, S. Durand, E. Pujos-Guillot, M. Touvier

**Writing, review, and/or revision of the manuscript:** L. Lécuyer, C. Dalle, B. Lyan, A. Demidem, A. Rossary, M.-P. Vasson, M. Petera, T. Ferreira, D. Centeno, P. Galan, S. Hercberg, M. Deschasaux, V. Partula, P. Latino-Martel, E. Kesse-Guyot, N. Druesne-Pecollo, S. Durand, E. Pujos-Guillot, M. Touvier

**Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases):** L. Lécuyer, E. Pujos-Guillot, M. Touvier

**Study supervision:** L. Lécuyer, A. Demidem, P. Galan, S. Hercberg, M. Touvier

## Acknowledgments

The authors thank Younes Esseddik, Frédéric Coffinieres, Thi Hong Van Duong, Paul Flanzy, Régis Gatibelza, Jagatjit Mohinder, and Maithyly Sivapalan (computer scientists), Rachida Mehroug and Frédérique Ferrat (logistic assistants), Nathalie Arnault, Véronique Gourlet, PhD, Fabien Szabo, PhD, Julien Allegre, and Laurent Bourhis (data-manager/statisticians), and Cédric Agaesse (dietitian) for their technical contribution to the SU.VI.MAX study. We also thank Nathalie Druesne-Pecollo, PhD (operational coordination), as well as all participants of the SU.VI.MAX study. The authors thank Dan Chaltiel for his help in R programming. This work was conducted in the framework of the French network for Nutrition And Cancer Research (NACRe network), [www.inra.fr/nacre](http://www.inra.fr/nacre) and received the NACRe Partnership Label. Metabolomic analysis was performed within the metaboHUB French infrastructure (ANR-INBS-0010). This work was supported by the French National Cancer Institute (grant no. INCa\_8085 for the project, and PhD grant no. INCa\_11323 for L. Lécuyer), the Federative Institute for Biomedical Research IFRB Paris 13, and the Cancéropôle Ile-de-France/Région Ile de France.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received February 5, 2019; revised April 12, 2019; accepted May 28, 2019; published first xx xx, xxxx.



- of antioxidant vitamins and minerals. *Arch Intern Med* 2004;164:2335–42.
15. Hercberg S, Preziosi P, Briançon S, Galan P, Triol I, Malvy D, et al. A primary prevention trial using nutritional doses of antioxidant vitamins and minerals in cardiovascular diseases and cancers in a general population: the SU.VI.MAX study—design, methods, and participant characteristics. *Supplementation en Vitamines et Minéraux AntioXydants. Control Clin Trials* 1998;19:336–51.
  16. World Health Organisation. ICD-10, International classification of diseases and related health problems. 10th Revision. Geneva, Switzerland: World Health Organization; 1993.
  17. Vandembroucke JP, Pearce N. Case-control studies: basic concepts. *Int J Epidemiol* 2012;41:1480–9.
  18. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B* 1995;57:289–300.
  19. Cattell RB. The scree test for the number of factors. *Multivar Behav Res* 1966;1:245–76.
  20. Ramautar R, Berger R, van der Greef J, Hankemeier T. Human metabolomics: strategies to understand biology. *Curr Opin Chem Biol* 2013;17:841–6.
  21. Assi N, Thomas DC, Leitzman M, Stepien M, Chajes V, Philip T, et al. Are metabolic signatures mediating the relationship between lifestyle factors and hepatocellular carcinoma risk? Results from a nested case-control study in EPIC. *Cancer Epidemiol Biomarkers Prev* 2018;27:531–40.
  22. Pages A, Duarte-Salles T, Stepien M, Ferrari P, Fedirko V, Pontoizeau C, et al. Metabolomic profiles of hepatocellular carcinoma in a European prospective cohort. *BMC Med* 2015;13:242.
  23. Stepien M, Duarte-Salles T, Fedirko V, Floegel A, Barupal DK, Rinaldi S, et al. Alteration of amino acid and biogenic amine metabolism in hepatobiliary cancers: findings from a prospective cohort study. *Int J Cancer* 2016;138:348–60.
  24. Assi N. A statistical framework to model the meeting-in-the-middle principle using metabolomic data: application to hepatocellular carcinoma in the EPIC study. *Mutagenesis* 2015;30:743–53.
  25. Mayers JR, Wu C, Clish CB, Kraft P, Torrence ME, Fiske BP, et al. Elevation of circulating branched-chain amino acids is an early event in human pancreatic adenocarcinoma development. *Nat Med* 2014;20:1193–8.
  26. Mondul AM, Moore SC, Weinstein SJ, Mannisto S, Sampson JN, Albanes D. 1-stearoylglycerol is associated with risk of prostate cancer: results from serum metabolomic profiling. *Metabolomics* 2014;10:1036–41.
  27. Huang J, Mondul AM, Weinstein SJ, Koutros S, Derkach A, Karoly E, et al. Serum metabolomic profiling of prostate cancer risk in the prostate, lung, colorectal, and ovarian cancer screening trial. *Br J Cancer* 2016;115:1087–95.
  28. Mondul AM, Moore SC, Weinstein SJ, Karoly ED, Sampson JN, Albanes D. Metabolomic analysis of prostate cancer risk in a prospective cohort: the alpha-tocopherol, beta-carotene cancer prevention (ATBC) study. *Int J Cancer* 2015;137:2124–32.
  29. Magnusson M, Lewis GD, Ericson U, Orho-Melander M, Hedblad B, Engstrom G, et al. A diabetes-predictive amino acid score and future cardiovascular disease. *Eur Heart J* 2013;34:1982–9.
  30. Newgard CB, An J, Bain JR, Muehlbauer MJ, Stevens RD, Lien LF, et al. A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. *Cell Metab* 2009;9:311–26.
  31. O'Connell TM. The complex role of branched chain amino acids in diabetes and cancer. *Metabolites* 2013;3:931–45.
  32. Pischon T, Nimpf K. Obesity and risk of cancer: an introductory overview. *Recent Results Cancer Res* 2016;208:1–15.
  33. Takeshita A, Igarashi-Migitaka J, Nishiyama K, Takahashi H, Takeuchi Y, Koibuchi N. Acetyl tributyl citrate, the most widely used phthalate substitute plasticizer, induces cytochrome p450 3a through steroid and xenobiotic receptor. *Toxicol Sci* 2011;123:460–70.
  34. Bui TT, Giovanoulis G, Cousins AP, Magnér J, Cousins IT, de Wit CA. Human exposure, hazard and risk of alternative plasticizers to phthalate esters. *Sci Total Environ* 2016;541:451–67.
  35. Rasmussen LM, Sen N, Vera JC, Liu X, Craig ZR. Effects of *in vitro* exposure to dibutyl phthalate, mono-butyl phthalate, and acetyl tributyl citrate on ovarian antral follicle growth and viability. *Biol Reprod* 2017;96:1105–17.
  36. Kambia N, Farce A, Belarbi K, Gressier B, Luyckx M, Chavatte P, et al. Docking study: PPARs interaction with the selected alternative plasticizers to di(2-ethylhexyl) phthalate. *J Enzyme Inhib Med Chem* 2016;31:448–55.
  37. Amathieu R, Nahon P, Triba M, Bouchemal N, Trinchet JC, Beaugrand M, et al. Metabolomic approach by 1H NMR spectroscopy of serum for the assessment of chronic liver failure in patients with cirrhosis. *J Proteome Res* 2011;10:3239–45.
  38. Zhenyukh O, Civantos E, Ruiz-Ortega M, Sánchez MS, Vázquez C, Peiró C, et al. High concentration of branched-chain amino acids promotes oxidative stress, inflammation and migration of human peripheral blood mononuclear cells via mTORC1 activation. *Free Radic Biol Med* 2017;104:165–77.
  39. Wang Q, Holmes MV, Davey Smith G, Ala-Korpela M. Genetic support for a causal role of insulin resistance on circulating branched-chain amino acids and inflammation. *Diabetes Care* 2017;40:1779–86.
  40. Lin H-M, Barnett MPG, Roy NC, Joyce NI, Zhu S, Armstrong K, et al. Metabolomic analysis identifies inflammatory and noninflammatory metabolic effects of genetic modification in a mouse model of Crohn's disease. *J Proteome Res* 2010;9:1965–75.
  41. Messerschmidt A, Worbs M, Steegborn C, Wahl MC, Huber R, Laber B, et al. Determinants of enzymatic specificity in the Cys-Met-metabolism PLP-dependent enzymes family: crystal structure of cystathionine gamma-lyase from yeast and intrafamilial structure comparison. *Biol Chem* 2003;384:373–86.
  42. Strobel GA. 4-Amino-4-cyanobutyric acid as an intermediate in glutamate biosynthesis. *J Biol Chem* 1967;244:3265–9.
  43. Jeney V, Balla J, Yachie A, Varga Z, Vercellotti GM, Eaton JW, et al. Pro-oxidant and cytotoxic effects of circulating heme. *Blood* 2002;100:879–87.
  44. Carayol M, Licaj I, Achaintre D, Sacerdote C, Vineis P, Key TJ, et al. Reliability of serum metabolites over a two-year period: a targeted metabolomic approach in fasting and non-fasting samples from EPIC. *PLoS One* 2015;10:e0135437.
  45. Floegel A, Drogan D, Wang-Sattler R, Prehn C, Illig T, Adamski J, et al. Reliability of serum metabolite concentrations over a 4-month period using a targeted metabolomic approach. *PLoS One* 2011;6:e21103.
  46. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, et al. Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* 2007;3:211–21.

**Supplementary Methods and Figure S1 – Plasma metabolomic signatures associated with long-term breast cancer risk in the SU.VI.MAX prospective cohort.****Supplementary methods for MS metabolomic analysis*****Chemicals and reagents***

Methanol (Lichrosolv Hypergrade for LC-MS) was purchased from Merck. Acetonitrile (ULC-MS), water (ULC-MS) and formic acid (99%) were purchased from Biosolve. All standard compounds were purchased from Sigma-Aldrich (Saint Quentin Fallavier, France), except for L- $\gamma$ -Glutamyl-L-Threonine, which was obtained from Genecust (Ellange, Luxembourg). All were analytical standard grade ( $\geq 98\%$ ), except 5-aminovaleric acid ( $\geq 97\%$ ) and L- $\gamma$ -Glutamyl-L-Threonine (95%).

***Biological samples preparation***

Untargeted semi-quantitative metabolomics was performed on plasma samples following a slightly modified procedure described in Pereira et al. (1). In brief, plasma samples (200  $\mu$ L) were deproteinized using cold methanol; after centrifugation and evaporation under nitrogen, the dry residues were redissolved in 50/50 (v/v) acetonitrile/water containing 0.1% formic acid. Pooled quality control samples were prepared by mixing 20  $\mu$ L from each of the plasma samples and prepared similarly.

***UPLC-MS analyses for metabolic profiling***

Metabolic profiles were determined using the Metabolic profiler® platform (Bruker Daltonique, Wissembourg, France). Briefly, six  $\mu$ L of each sample were injected into a fast LC system (Agilent 1200 series) coupled to a microTOF mass spectrometer equipped with electrospray ionization source (Bruker). Chromatographic separation was performed using a 150 x 2.1mm Acquity HSS T3 C18 column (Waters, Saint Quentin en Yvelines, France), with a particle size of 1.8 $\mu$ m. The flow rate was set at 0.2 mL/min with mobile phase A (100% water) and mobile phase B (100% acetonitrile) both containing 0.1% formic acid. The elution gradient was carried out as follows: 0-1min, 100% A; 1-13min linear gradient to 100% B; 13-20min 100% B; 20-26 min 100% A for re equilibration. The column temperature was set at 30°C. Samples were kept at 4°C during analyses. Data were acquired in positive and negative ion modes with a scan range from 50 to 1000 mass-to-charge ratio (m/z).

Samples were randomized within the analytical sequence based on a Williams Latin Square strategy defined according to the main factors of the study. The stability of the analytical system was monitored using pooled samples as quality control, injected one time at the beginning of each sequence and then after each set of 12 samples.

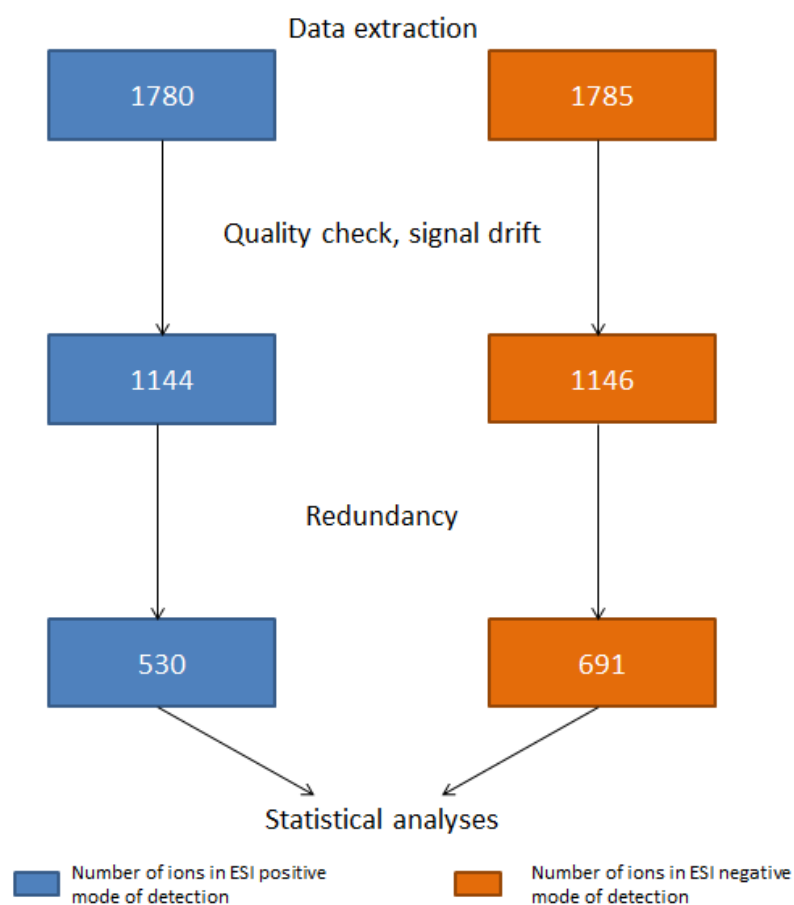
***Data extraction***

After a fast overview of all chromatograms, 4 pairs of individuals in positive mode and 12 pairs in negative mode were excluded due to problems during sample preparation, too much ions with null intensity (>50%) or pollution on chromatograms, leaving 211 cases and 211 controls samples for positive mode and 203 cases and 203 controls samples for negative mode analyses. Raw data files were processed and converted to NetCDF format using the Bruker Data Analysis software. Data were processed under the Galaxy web-based platform Workflow4metabolomics (2), using first XCMS, followed by quality checks and signal drift correction according to the algorithm described by Van der Kloet et al (3), to yield a data matrix containing retention times, masses and peak intensities that have been corrected for batch effects. These steps include noise filtering, automatic peak detection and chromatographic alignment allowing the appropriate comparison of multiple samples by further processing methods. In order to take into account redundancy inherent to the untargeted approach (parent ion, fragments, isotopes and so on) and to keep only relevant ions (removing ions with a too low intensity, non-robust and so on) only features present in  $\geq 90\%$  of either case or control population, or with an intensity features  $\geq 1,500$  a.u. for the 90th higher percentile were kept.

Moreover, analytical correlations were filtered using the ‘Metabolites Correlation Analysis’ tool (Galaxy Version 1.0.1) with a correlation coefficient threshold of 0.9. After data processing and filtering steps, 530 and 691 ions (for positive and negative modes respectively) were considered for statistical analyses. Details about number of ions in matrices for each step of processing are shown in **Supplementary Figure 1**.

### ***Metabolite identification***

Identification experiments were performed on a ThermoScientific LTQ Orbitrap Velos hybrid mass spectrometer (Thermo Fisher Scientific, San José, CA, USA) using high resolution, full MS ( $m/z$  70–1,000) at 100,000 resolving power, with acquisitions carried out for a limited selection of samples of interest. Metabolites contributing to the discrimination of the different phenotypes were first identified using an in-house database containing the reference spectra of more than 1,000 authentic standard compounds analysed in the same analytical conditions. Then, the remaining unknown compounds were identified on the basis of their exact masses which were compared to those registered in Metlin (<https://metlin.scripps.edu>), in the Human Metabolome Database (HMDB; [www.hmdb.ca](http://www.hmdb.ca)) (4), or in Kyoto Encyclopedia of Genes and Genomes (KEGG) database (<http://www.genome.jp/kegg/>). Database queries were performed with a mass error of 0.005 Da and a retention time difference of 0.1 min (for the in-house database). Database results were confirmed using appropriate standards when available, isotopic patterns, and mass fragmentation analyses using collision induced dissociation MS<sup>2</sup> data acquired on the same samples. For unidentified ions, the number of plausible elemental compositions are restricted to a small number (or uniquely identified) with the support of additional chemical information-i.e. the molecular formula of the parent and knowledge of possible metabolic pathways. Metabolites were classified accordingly to Sumner et al (5) concerning the levels of confidence in the identification process: identified (confirmed by standard=level 1), putatively annotated (based upon physicochemical properties and/or spectral similarity with public/commercial spectral libraries=level 2).

**Supplementary Figure S1** Number of ions in matrices for each step of processing

## REFERENCES

- Pereira F, Martin JF, Joly C, Sébédio JL, Pujos-Guillot E. Development and validation of a UPLC/MS method for a nutritional metabolomic study of human plasma. *Metabolomics*. 2010;6:207–18.
- Giacomoni F, Le Corguille G, Monsoor M, Landi M, Pericard P, Petera M, et al. Workflow4Metabolomics: a collaborative research infrastructure for computational metabolomics. *Bioinformatics*. 2015;31:1493–5.
- Van der Kloet FM, Bobeldijk I, Verheij ER, Jellema RH. Analytical error reduction using single point calibration for accurate and precise metabolomic phenotyping. *JProteomeRes*. 2009;8:5132–41.
- Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, et al. HMDB 3.0--The Human Metabolome Database in 2013. *Nucleic Acids Res*. 2013;41:D801–7.
- Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, et al. Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics*. 2007;3:211–21.

**Supplementary Table S1 – Plasma metabolomic signatures associated with long-term breast cancer risk in the SU.VI.MAX prospective cohort.**

**Supplementary Table S1** Ions associated with breast cancer risk from multivariable conditional logistic regression<sup>a</sup> (FDR≤0.2)

Name (mass_retention time for unknown)	Mode	Cases/Controls	Mean	SD	OR [95% CI]	P-trend	Corrected P-trend <sup>b</sup>
M98.919T2.24	ESI +	211/211	2627	339	1.63 [1.28-2.09]	0.0001	0.06
M96.9218T2.24	ESI +	211/211	6326	751	1.63 [1.28-2.08]	0.0001	0.06
M415.2078T22.4	ESI -	203/203	3432	639	0.61 [0.47-0.80]	0.0003	0.1
M80.9474T2.25	ESI +	211/211	50550	4469	1.49 [1.20-1.87]	0.0004	0.1
M201.149T18.18	ESI -	203/203	1449	918	1.51 [1.19-1.91]	0.0006	0.1
M265.3932T3.1	ESI +	211/211	829	1038	1.87 [1.31-2.69]	0.0006	0.1
M548.8294T2.17	ESI +	211/211	1937	183	0.66 [0.53-0.84]	0.0007	0.1
M424.8961T2.05	ESI +	211/211	1727	305	1.47 [1.17-1.85]	0.001	0.1
M153.0162T1.93	ESI +	211/211	2824	438	1.46 [1.16-1.84]	0.001	0.1
M160.9894T1.9	ESI +	211/211	2683	383	1.45 [1.16-1.81]	0.001	0.1
M114.9883T1.9	ESI +	211/211	10204	1454	1.43 [1.15-1.77]	0.001	0.1
M483.1958T22.45	ESI -	203/203	1400	444	0.66 [0.52-0.86]	0.002	0.1
M474.7265T2.04	ESI -	203/203	3160	354	0.68 [0.54-0.86]	0.002	0.1
L-Phenylalanine M166.0869T10.11	ESI +	211/211	227718	42535	1.43 [1.14-1.78]	0.002	0.1
M556.3244T16.85	ESI -	203/203	3113	847	1.46 [1.15-1.86]	0.002	0.1
M320.068T2.31	ESI +	211/211	2292	1195	1.47 [1.16-1.88]	0.002	0.1
M332.8363T2.21	ESI +	211/211	3017	380	1.44 [1.14-1.81]	0.002	0.1
L-valine/norvaline M72.0803T3.08	ESI +	211/211	115033	24435	1.45 [1.15-1.83]	0.002	0.1
5-aminovaleric acid M140.0686T2.29	ESI +	211/211	10219	5139	1.46 [1.14-1.86]	0.003	0.2
M485.2107T23.95	ESI -	203/203	1268	330	0.70 [0.55-0.88]	0.003	0.2
M629.6992T2.23	ESI +	211/211	926	1028	1.71 [1.20-2.43]	0.003	0.2
M152.0367T1.91	ESI +	211/211	3377	525	1.40 [1.12-1.74]	0.003	0.2
M476.8958T2.06	ESI +	211/211	1697	245	1.39 [1.12-1.74]	0.003	0.2
M74.528T1.89	ESI +	211/211	17127	2227	1.39 [1.12-1.73]	0.003	0.2
O-succinyl-L-homoserine M240.0464T2.86	ESI -	203/203	919	400	0.70 [0.55-0.89]	0.004	0.2
M131.0476T10.12	ESI +	211/211	1353	277	1.43 [1.12-1.81]	0.004	0.2
M488.8728T2.15	ESI +	211/211	6003	409	0.72 [0.58-0.90]	0.004	0.2
M634.8771T2.08	ESI +	211/211	6580	528	0.72 [0.58-0.90]	0.004	0.2
M430.9143T2.09	ESI +	211/211	24103	1302	0.73 [0.58-0.90]	0.004	0.2
M222.8676T2.21	ESI +	211/211	1376	189	1.42 [1.11-1.80]	0.005	0.2
M518.7449T2.21	ESI +	211/211	1579	258	1.38 [1.10-1.72]	0.005	0.2
M111.0109T1.91	ESI +	211/211	4464	681	1.38 [1.10-1.72]	0.005	0.2
M190.0186T2.77	ESI -	203/203	1074	887	1.41 [1.11-1.80]	0.005	0.2
ATBC (acetyl tributyl citrate) M425.2172T19.31	ESI +	211/211	2198	964	1.40 [1.10-1.79]	0.006	0.2

## ANNEXE 2

M544.8888T2.06	ESI +	211/211	1398	203	1.36 [1.09-1.68]	0.006	0.2
M586.6522T2.02	ESI -	203/203	1427	199	0.72 [0.57-0.91]	0.006	0.2
M121.0844T10.12	ESI +	211/211	3317	634	1.36 [1.09-1.69]	0.006	0.2
M208.8913T2.22	ESI +	211/211	5336	576	1.34 [1.09-1.66]	0.006	0.2
M68.0197T1.9	ESI +	211/211	7421	1019	1.35 [1.09-1.68]	0.006	0.2
M359.0438T3.08	ESI +	211/211	1782	575	1.43 [1.11-1.86]	0.007	0.2
L-tryptophan M227.0784T11.09	ESI +	211/211	799	180	1.40 [1.10-1.79]	0.007	0.2
M294.9383T2.1	ESI +	211/211	12723	681	0.73 [0.59-0.92]	0.007	0.2
M90.9764T2.13	ESI +	211/211	126879	3773	0.73 [0.58-0.92]	0.007	0.2
M285.2066T19.04	ESI -	203/203	948	1114	1.41 [1.10-1.82]	0.007	0.2
M391.2072T23.51	ESI -	203/203	3549	950	0.72 [0.56-0.91]	0.007	0.2
M192.0305T3.01	ESI +	211/211	1705	1224	1.37 [1.09-1.73]	0.007	0.2
M169.0362T3.1	ESI +	211/211	41324	9207	1.41 [1.10-1.83]	0.008	0.2
M88.0579T1.94	ESI +	211/211	975	754	1.37 [1.08-1.72]	0.008	0.2
L-γ-glutamyl-L-threonine M231.1436T2.52	ESI +	211/211	1054	601	1.39 [1.09-1.78]	0.008	0.2
M609.3187T19.97	ESI -	203/203	2815	419	0.75 [0.60-0.93]	0.009	0.2
M410.8577T2.18	ESI +	211/211	5805	350	0.73 [0.58-0.93]	0.009	0.2
M458.7861T2.2	ESI +	211/211	2052	254	1.33 [1.07-1.65]	0.009	0.2
L-glutamine / L-isoglutamine M90.0547T2.32	ESI +	211/211	13187	2982	1.33 [1.07-1.66]	0.009	0.2
M363.9292T2.09	ESI +	211/211	1431	140	0.75 [0.61-0.93]	0.009	0.2
M82.9449T2.24	ESI +	211/211	18486	4576	1.38 [1.08-1.77]	0.009	0.2
M276.888T2.21	ESI +	211/211	3901	345	1.32 [1.07-1.63]	0.009	0.2
M53.0185T1.89	ESI +	211/211	5522	837	1.34 [1.07-1.68]	0.01	0.2
pregnene-triol sulfate M413.2002T16.11	ESI -	203/203	2441	1254	1.38 [1.08-1.77]	0.01	0.2
M286.143T16.48	ESI +	211/211	1931	2491	0.71 [0.55-0.92]	0.01	0.2
M166.0387T2.4	ESI -	203/203	1697	703	0.73 [0.57-0.93]	0.01	0.2
M201.8018T2.34	ESI -	203/203	2321	598	1.38 [1.08-1.77]	0.01	0.2
M588.3285T19.13	ESI -	203/203	38789	12941	1.37 [1.07-1.75]	0.01	0.2
M204.9569T2.03	ESI +	211/211	2579	364	1.31 [1.06-1.61]	0.01	0.2
M112.0076T1.91	ESI +	211/211	3917	627	1.32 [1.06-1.64]	0.01	0.2
M590.3434T19.69	ESI -	203/203	11805	5362	1.38 [1.07-1.77]	0.01	0.2
M540.3295T19.89	ESI -	203/203	168053	32521	1.38 [1.07-1.78]	0.01	0.2
M73.5317T1.89	ESI +	211/211	108235	14722	1.31 [1.06-1.62]	0.01	0.2
M412.8538T2.18	ESI +	211/211	3853	277	0.74 [0.58-0.94]	0.01	0.2
M412.8055T2.02	ESI -	203/203	3495	361	1.34 [1.06-1.68]	0.01	0.2
M185.1292T10.52	ESI +	211/211	1379	433	1.32 [1.06-1.65]	0.01	0.2
M558.9914T2.38	ESI +	211/211	1347	243	1.33 [1.06-1.67]	0.01	0.2
M794.4134T2.12	ESI -	203/203	1887	210	0.74 [0.58-0.94]	0.01	0.2
M112.0389T2.78	ESI -	203/203	822	635	1.32 [1.06-1.65]	0.01	0.2

<sup>a</sup> Multivariable models were adjusted for age (continuous), BMI (continuous), smoking status (current smokers, former smokers and non-smokers), season of blood draw (a priori-defined periods: October-November/December-January-February/March-April-May), height (continuous), alcohol intake (continuous), physical activity (irregular/< 1h/d walking equivalent/≥ 1h/d walking equivalent), education level (primary/secondary/superior), family history of breast cancer (yes/no), number of children (continuous), and use of hormone replacement therapy for menopause (yes/no)

## ANNEXE 2

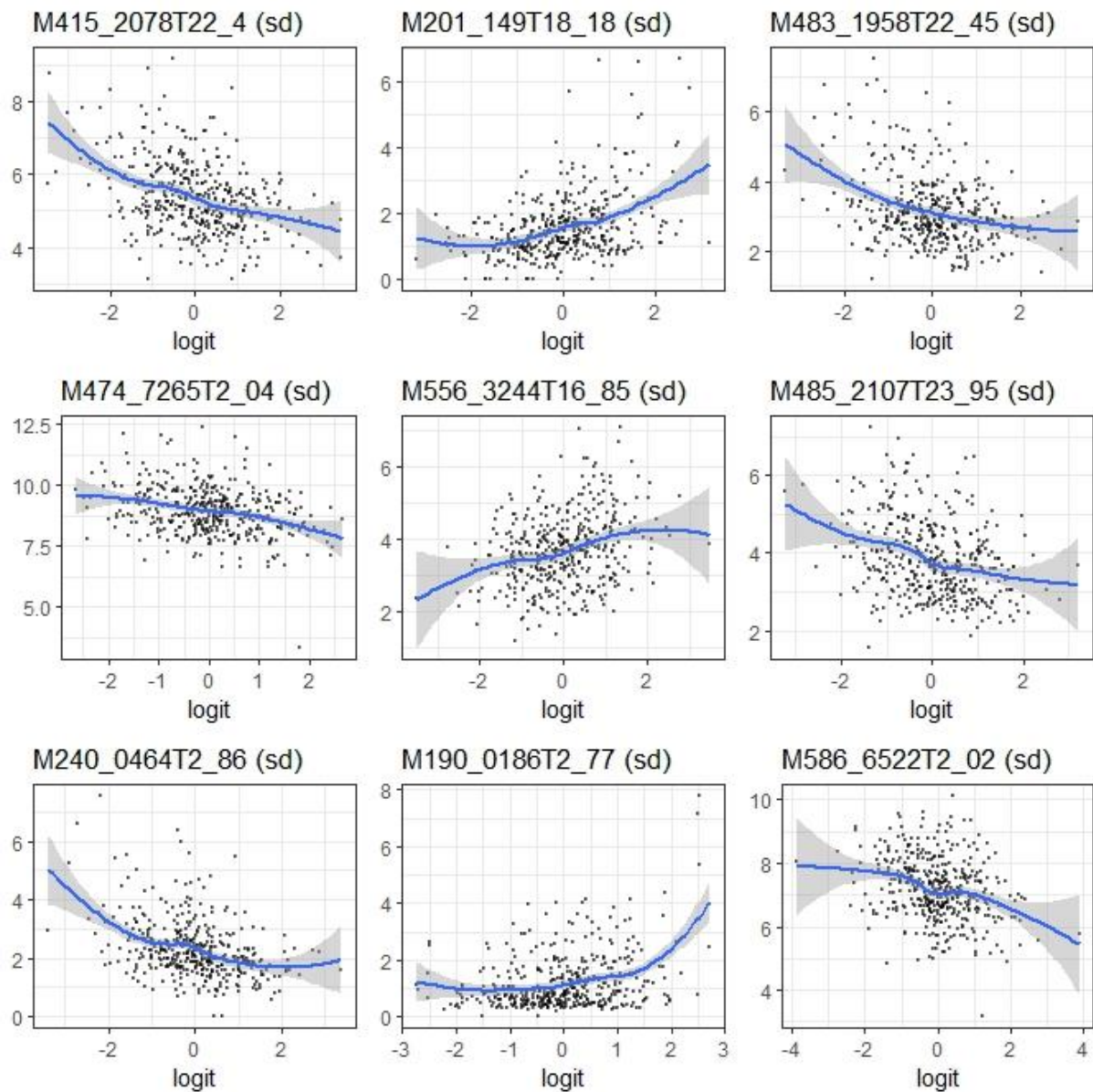
(reference model). Tests for linear trend were performed using continuous variables. ORs were presented for an increment of 1 SD for each metabolite.

<sup>b</sup> P-trend after Benjamini Hochberg correction for multiple testing  
SD, Standard Deviation

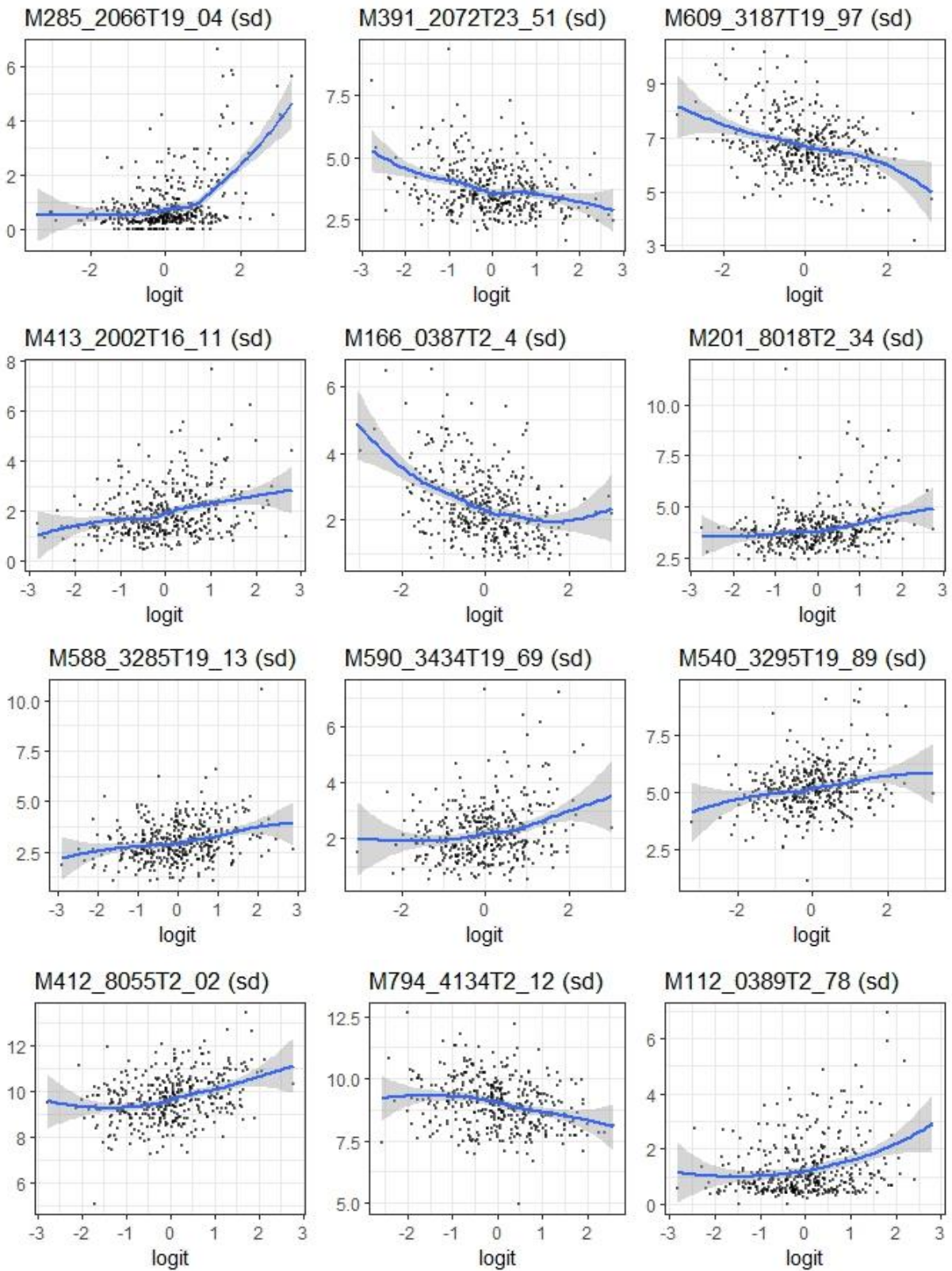
**Supplementary Figures S2.a and S2.b and Table S2 – Plasma metabolomic signatures associated with long-term breast cancer risk in the SU.VI.MAX prospective cohort.**

**Linearity assumption between ions' intensity<sup>a</sup> and the logit of predicted values<sup>b</sup> of breast cancer risk**

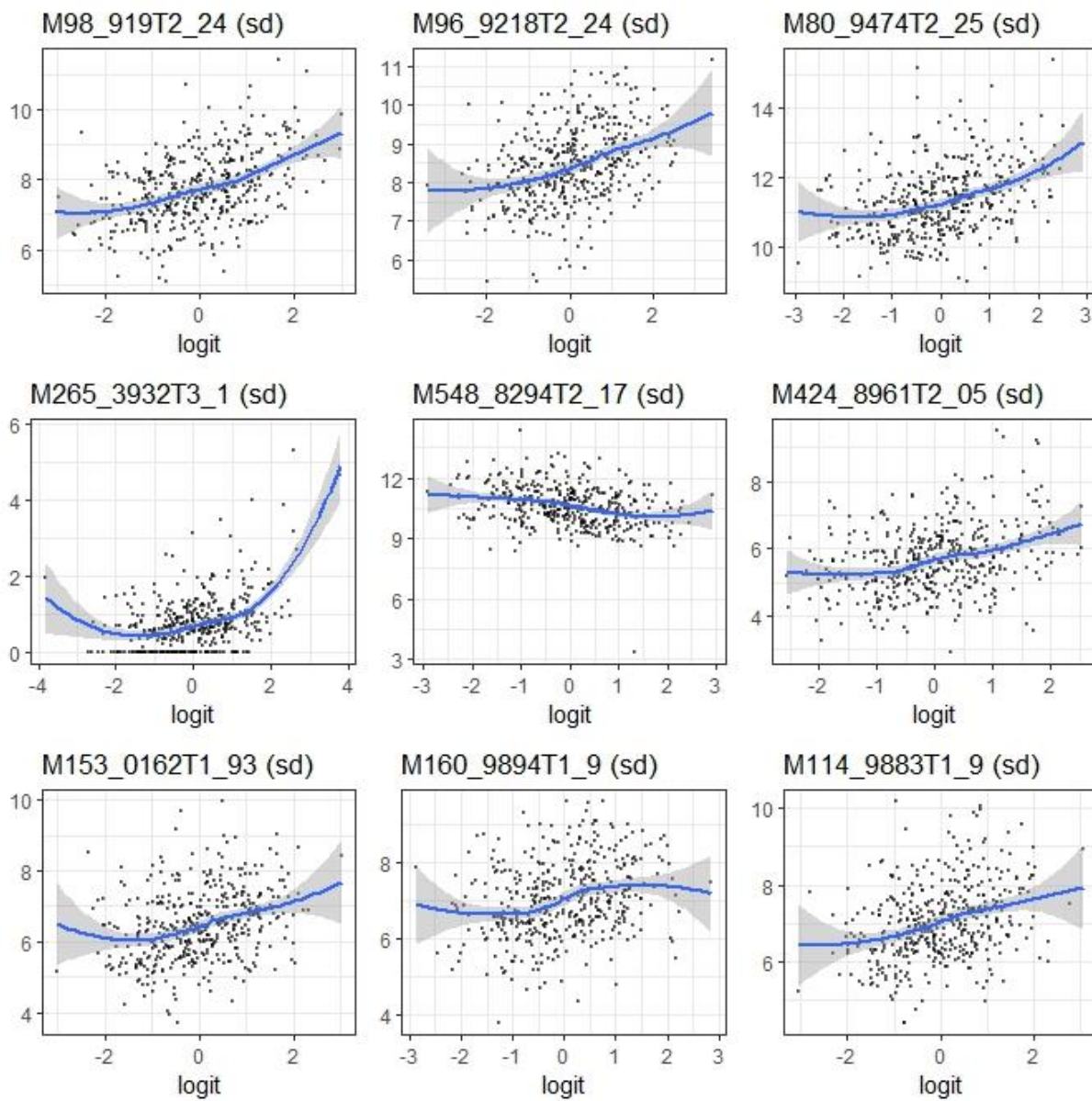
**Supplementary figure S2.a** Plots<sup>c</sup> of ESI negative mode ions' intensity as a function of the logit(predicted values of breast cancer risk)

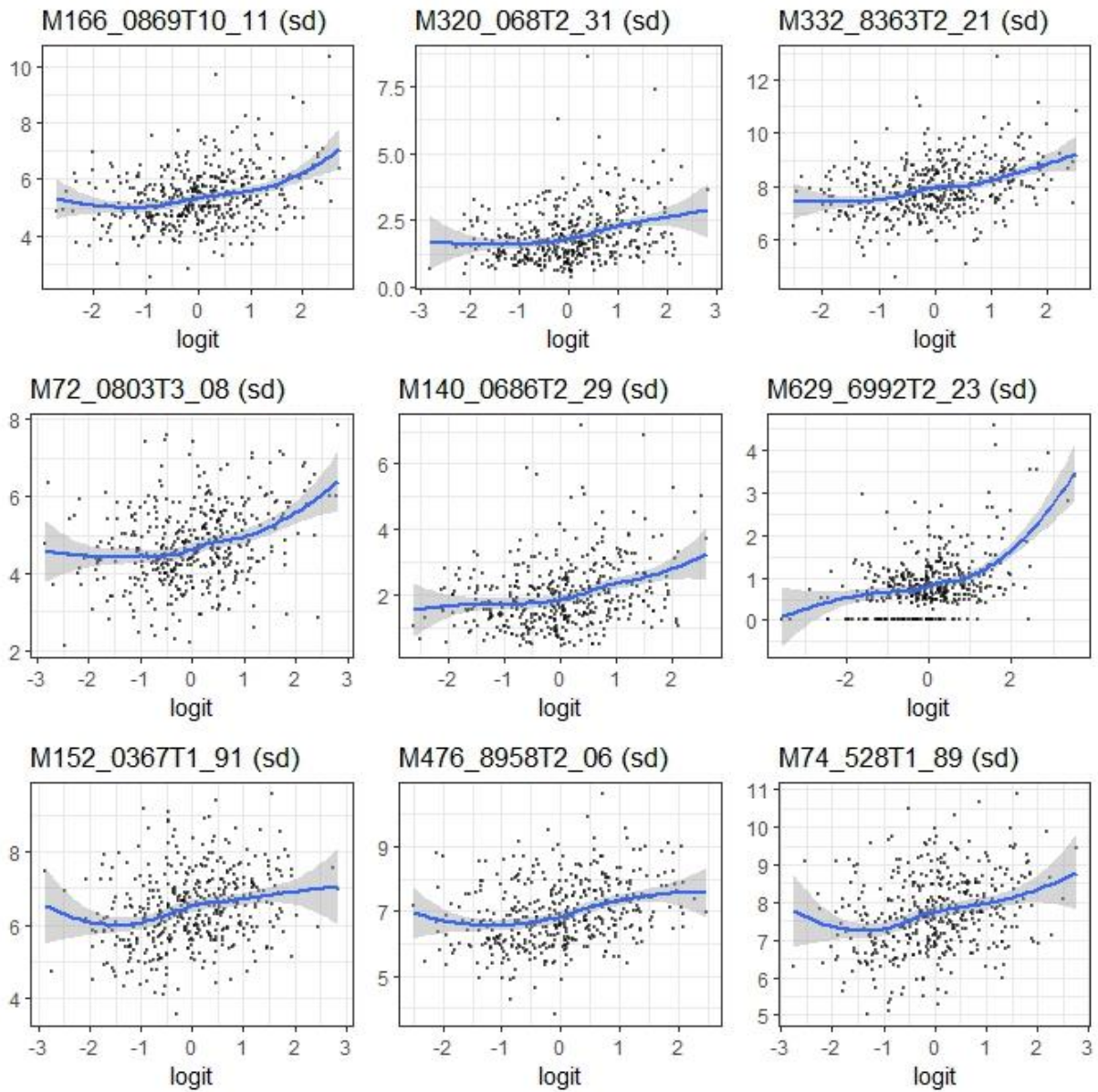


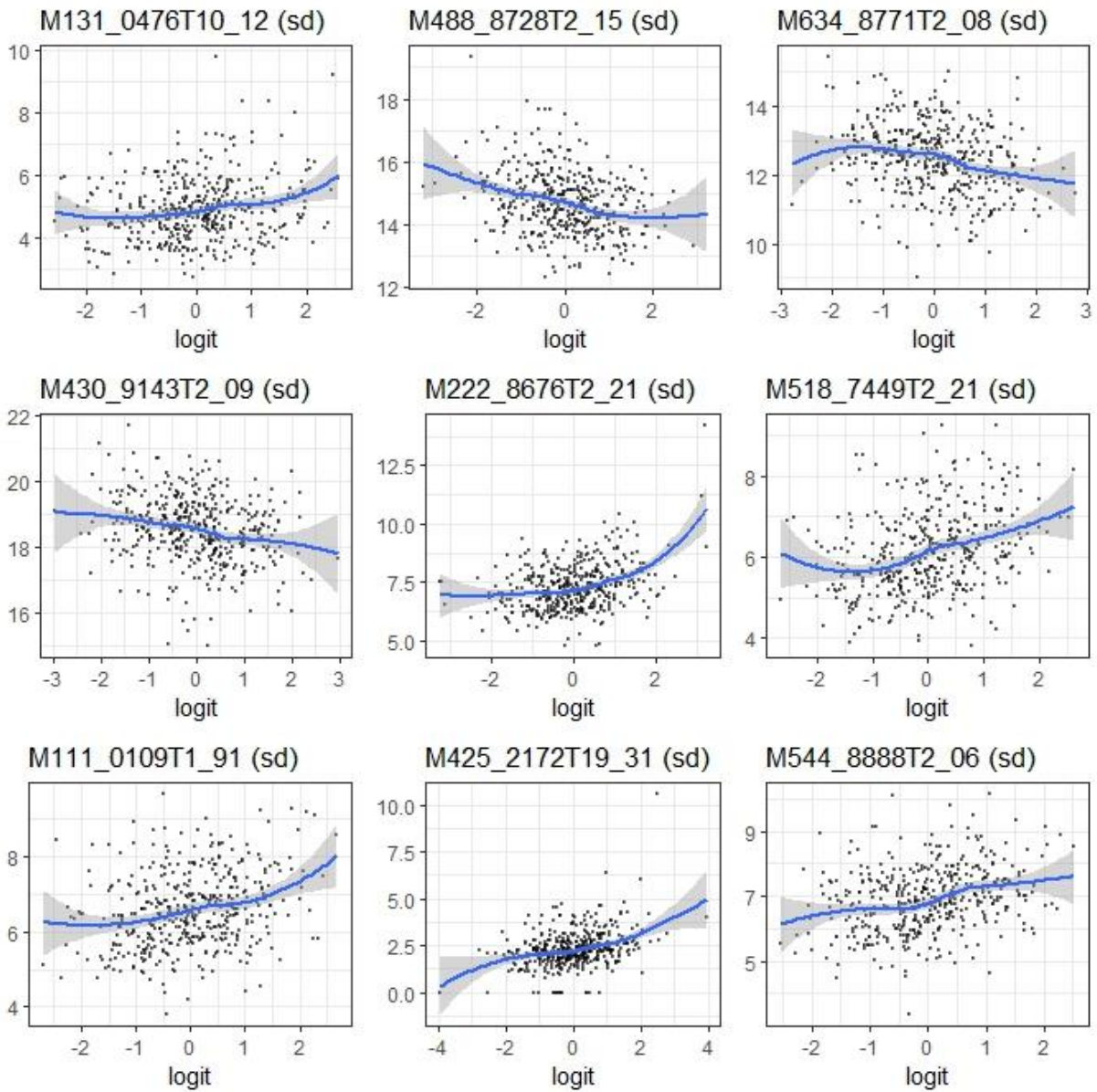


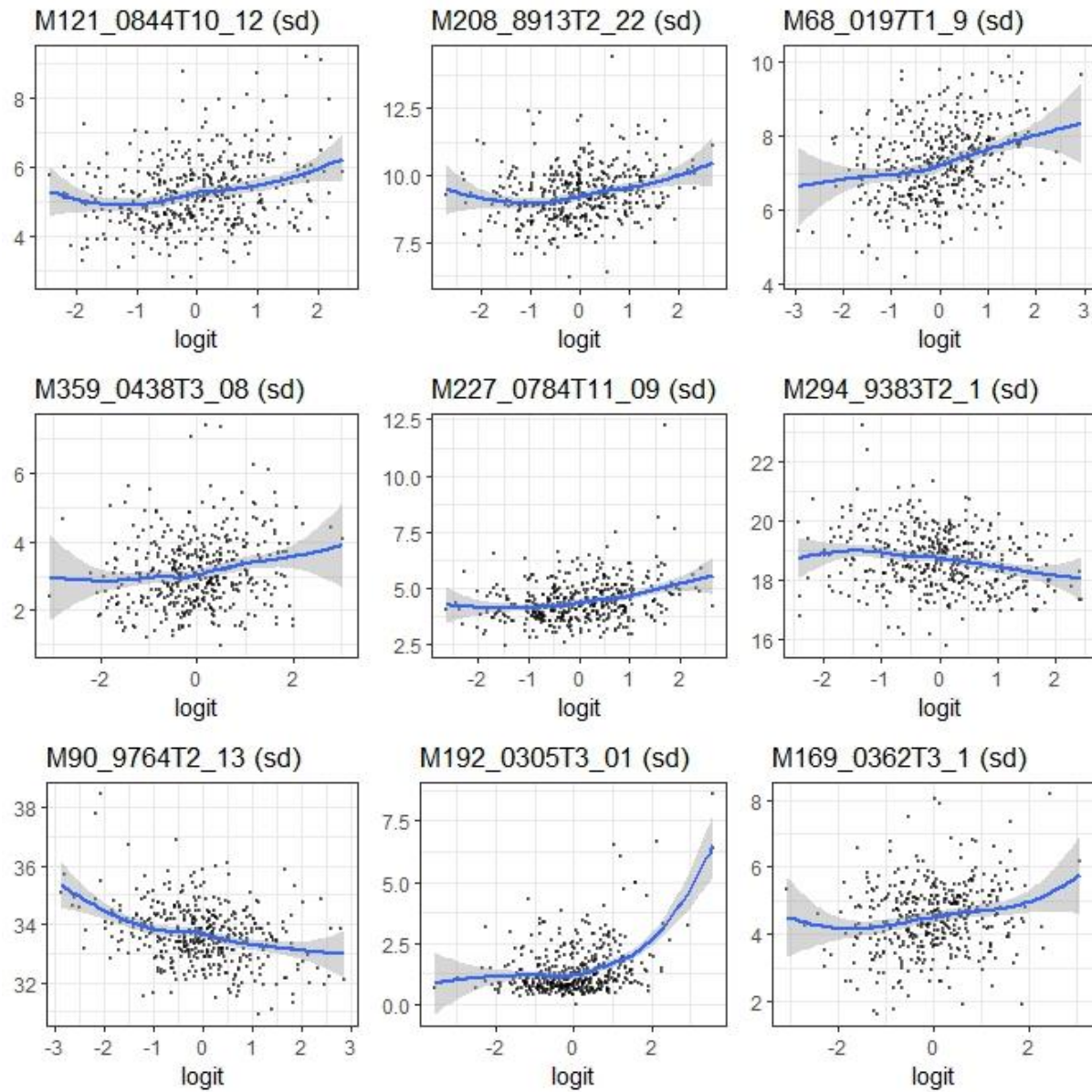


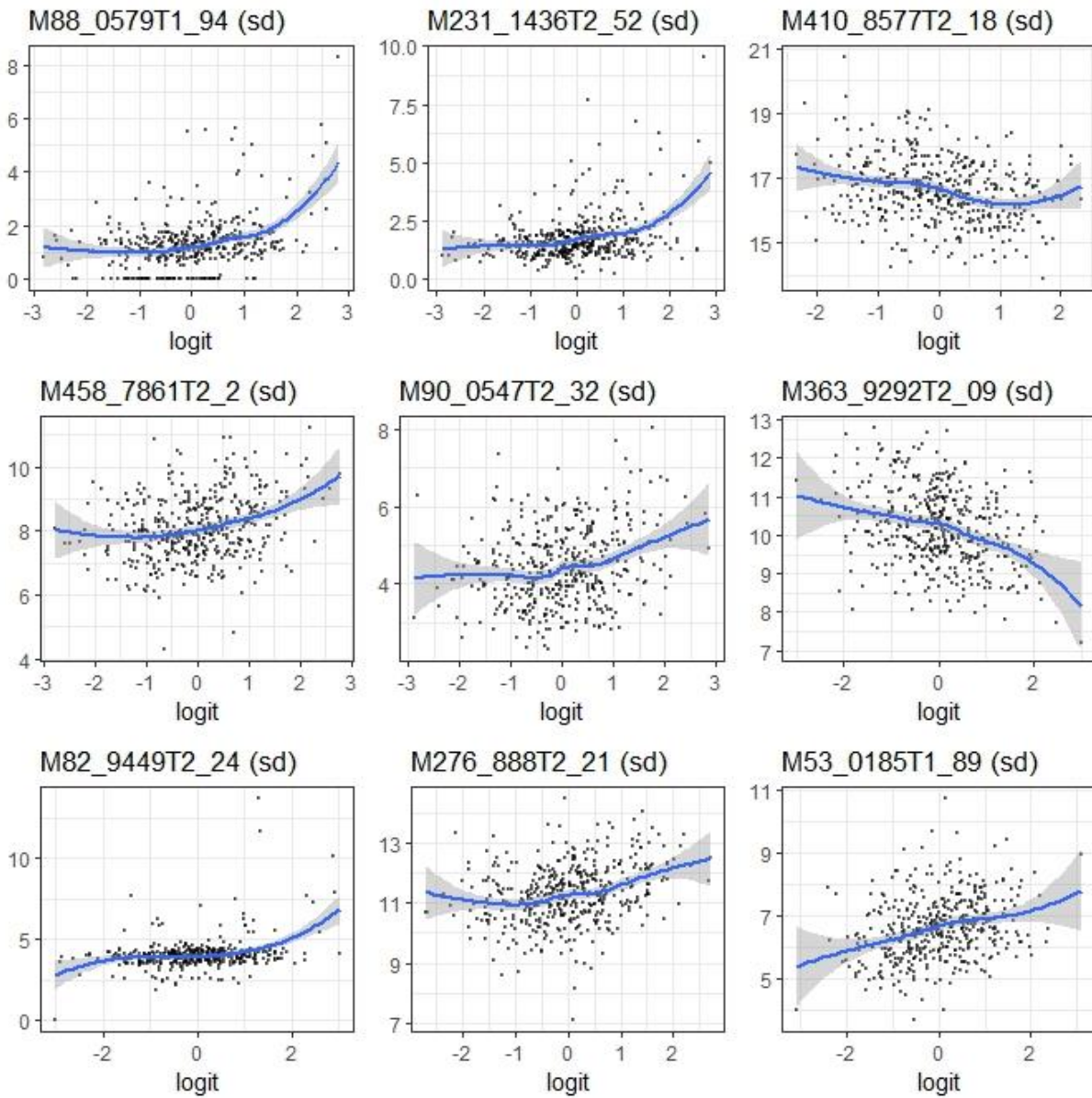
**Supplementary figure S2.b** Plots<sup>c</sup> of ESI positive mode ions' intensity as a function of the logit(predicted values of breast cancer risk)

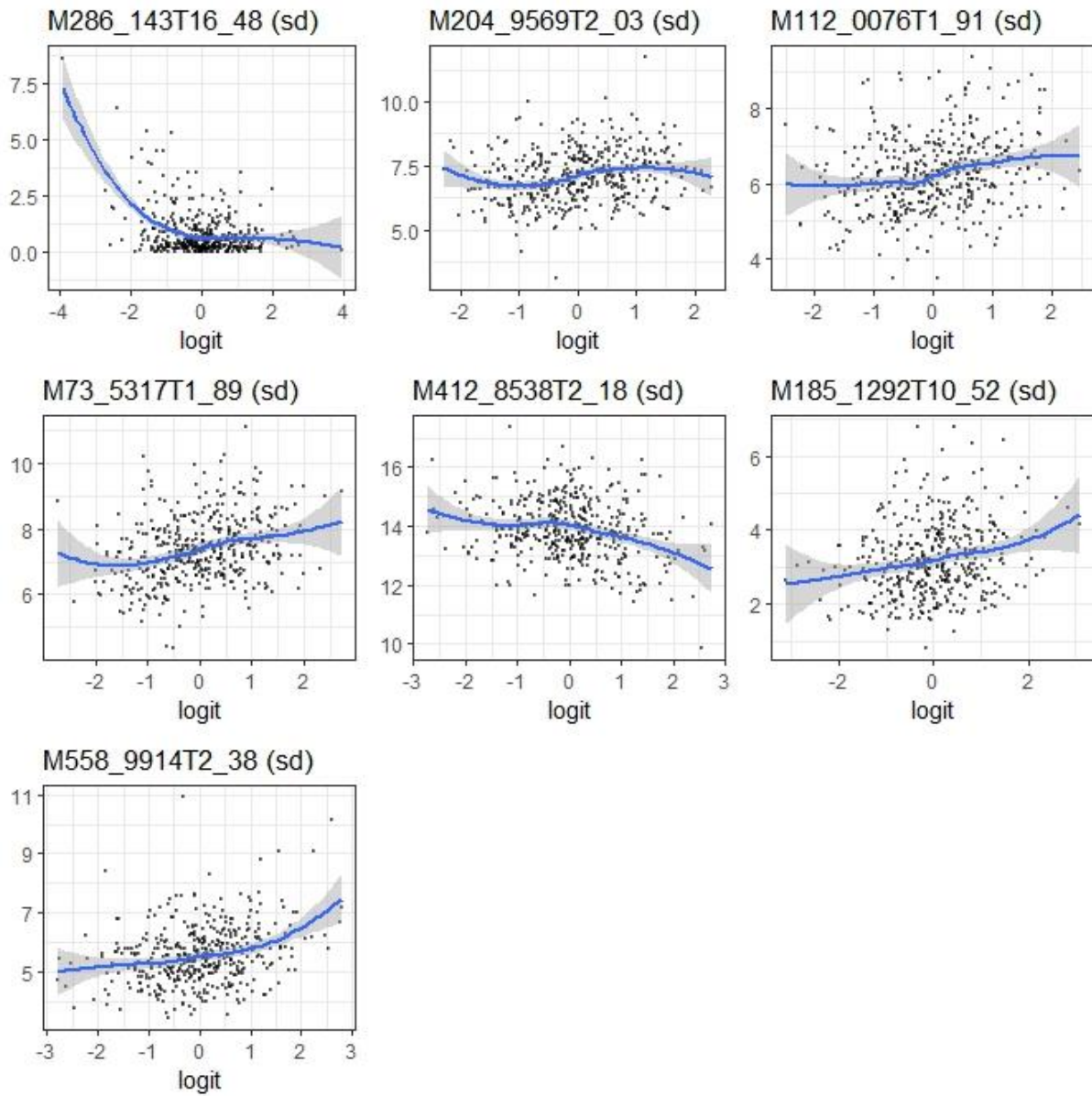












**Supplementary Table S2** Associations between categorized ions' intensity<sup>d</sup> and breast cancer risk from multivariable conditional logistic regression<sup>b</sup> (FDR≤0.2)

Name (mass retention time for unknown)	Mode	T1 (ref)		T2		T3		P-nontrend	Corrected P-nontrend <sup>e</sup>
		cases/controls	OR	cases/controls	OR [95% CI]	cases/controls	OR [95% CI]		
M265.3932T3.1	ESI +	49/91	1.00	76/65	2.33 [1.35-4.04]	86/55	3.97 [2.15-7.36]	0.00004	0.003
M586.6522T2.02	ESI -	77/58	1.00	79/57	0.84 [0.47-1.5]	47/88	0.31 [0.17-0.55]	0.0001	0.004
M96.9218T2.24	ESI +	51/89	1.00	78/63	2.63 [1.48-4.67]	82/59	3.06 [1.7-5.5]	0.0003	0.007
M98.919T2.24	ESI +	56/84	1.00	68/73	1.37 [0.81-2.3]	87/54	2.71 [1.56-4.72]	0.001	0.02
M412.8055T2.02	ESI -	64/71	1.00	54/82	0.66 [0.38-1.15]	85/50	1.89 [1.07-3.33]	0.002	0.02
M88.0579T1.94	ESI +	55/85	1.00	81/60	2.55 [1.47-4.44]	75/66	1.96 [1.16-3.31]	0.003	0.03
M548.8294T2.17	ESI +	81/59	1.00	71/70	0.53 [0.32-0.89]	59/82	0.37 [0.2-0.65]	0.003	0.03
M80.9474T2.25	ESI +	54/86	1.00	76/65	1.93 [1.15-3.25]	81/60	2.4 [1.41-4.07]	0.003	0.03
M430.9143T2.09	ESI +	77/63	1.00	74/67	0.86 [0.52-1.43]	60/81	0.4 [0.23-0.7]	0.004	0.03
M53.0185T1.89	ESI +	58/82	1.00	70/71	1.33 [0.82-2.18]	83/58	2.4 [1.42-4.05]	0.004	0.03
M201.149T18.18	ESI -	54/81	1.00	68/68	1.58 [0.9-2.75]	81/54	2.51 [1.44-4.38]	0.005	0.03
M476.8958T2.06	ESI +	55/85	1.00	72/69	1.41 [0.85-2.34]	84/57	2.36 [1.4-3.97]	0.005	0.03
M424.8961T2.05	ESI +	56/84	1.00	74/67	1.73 [1.05-2.85]	81/60	2.34 [1.38-3.98]	0.005	0.03
M609.3187T19.97	ESI -	77/58	1.00	69/67	0.8 [0.47-1.37]	57/78	0.4 [0.22-0.72]	0.006	0.03
M74.528T1.89	ESI +	65/75	1.00	63/78	0.88 [0.54-1.44]	83/58	2.05 [1.21-3.46]	0.007	0.03
M412.8538T2.18	ESI +	77/63	1.00	74/67	0.82 [0.5-1.37]	60/81	0.41 [0.23-0.73]	0.008	0.04
M285.2066T19.04	ESI -	59/76	1.00	63/73	0.99 [0.58-1.69]	81/54	2.1 [1.22-3.61]	0.009	0.04
M415.2078T22.4	ESI -	81/54	1.00	63/73	0.49 [0.29-0.85]	59/76	0.45 [0.26-0.79]	0.01	0.04
M153.0162T1.93	ESI +	61/79	1.00	67/74	1.16 [0.69-1.93]	83/58	2.19 [1.28-3.75]	0.01	0.04
M629.6992T2.23	ESI +	59/81	1.00	69/72	1.29 [0.77-2.15]	83/58	2.23 [1.3-3.82]	0.01	0.04
M160.9894T1.9	ESI +	59/81	1.00	75/66	1.64 [0.99-2.7]	77/64	2.23 [1.29-3.84]	0.01	0.04
M556.3244T16.85	ESI -	59/76	1.00	68/68	1.73 [1.02-2.93]	76/59	2.29 [1.3-4.03]	0.01	0.04
M208.8913T2.22	ESI +	69/71	1.00	60/81	0.81 [0.48-1.36]	82/59	1.87 [1.08-3.23]	0.01	0.04
M363.9292T2.09	ESI +	81/59	1.00	70/71	0.7 [0.42-1.19]	60/81	0.47 [0.28-0.78]	0.01	0.04
M794.4134T2.12	ESI -	73/62	1.00	74/62	1 [0.59-1.71]	56/79	0.47 [0.27-0.84]	0.02	0.05
L-tryptophan M227.0784T11.09	ESI +	62/78	1.00	64/77	1.06 [0.65-1.74]	85/56	2.04 [1.21-3.43]	0.02	0.05
M590.3434T19.69	ESI -	62/73	1.00	64/72	1.17 [0.7-1.96]	77/58	2.23 [1.26-3.94]	0.02	0.05
L-valine/norvaline M72.0803T3.08	ESI +	66/74	1.00	63/78	0.98 [0.57-1.7]	82/59	2.01 [1.14-3.53]	0.02	0.05
M474.7265T2.04	ESI -	76/59	1.00	69/67	0.69 [0.4-1.17]	58/77	0.46 [0.26-0.79]	0.02	0.05
M131.0476T10.12	ESI +	67/73	1.00	66/75	1.06 [0.62-1.82]	78/63	2.13 [1.17-3.91]	0.02	0.05
M222.8676T2.21	ESI +	58/82	1.00	75/66	1.8 [1.07-3.02]	78/63	2.05 [1.18-3.55]	0.02	0.05
M458.7861T2.2	ESI +	60/80	1.00	70/71	1.59 [0.92-2.72]	81/60	2.11 [1.23-3.62]	0.02	0.05
M391.2072T23.51	ESI -	73/62	1.00	73/63	1.04 [0.62-1.74]	57/78	0.52 [0.3-0.91]	0.03	0.06
L-Phenylalanine M166.0869T10.11	ESI +	62/78	1.00	70/71	1.43 [0.85-2.4]	79/62	2.07 [1.21-3.52]	0.03	0.06
M90.9764T2.13	ESI +	73/67	1.00	79/62	1.07 [0.65-1.78]	59/82	0.57 [0.34-0.95]	0.03	0.06
M320.068T2.31	ESI +	59/81	1.00	74/67	1.73 [1-2.99]	78/63	1.93 [1.13-3.3]	0.04	0.07
L-glutamine / L-isoglutamine M90.0547T2.32	ESI +	65/75	1.00	66/75	1.13 [0.66-1.92]	80/61	1.9 [1.14-3.19]	0.04	0.07
M488.8728T2.15	ESI +	72/68	1.00	78/63	1.11 [0.66-1.86]	61/80	0.59 [0.35-0.99]	0.04	0.1
M68.0197T1.9	ESI +	60/80	1.00	75/66	1.67 [0.98-2.82]	76/65	1.84 [1.09-3.09]	0.046	0.1
M634.8771T2.08	ESI +	80/60	1.00	69/72	0.74 [0.44-1.23]	62/79	0.52 [0.31-0.88]	0.049	0.1



## ANNEXE 2

M410.8577T2.18	ESI +	76/64	1.00	73/68	0.72 [0.41-1.26]	62/79	0.49 [0.28-0.87]	0.1	0.1
M190.0186T2.77	ESI -	62/73	1.00	65/71	1.11 [0.65-1.91]	76/59	1.81 [1.07-3.05]	0.1	0.1
5-aminovaleric acid M140.0686T2.29	ESI +	64/76	1.00	67/74	1.18 [0.68-2.05]	80/61	1.89 [1.11-3.2]	0.1	0.1
M152.0367T1.91	ESI +	61/79	1.00	73/68	1.57 [0.93-2.66]	77/64	1.85 [1.1-3.13]	0.1	0.1
M332.8363T2.21	ESI +	57/83	1.00	76/65	1.6 [0.97-2.65]	78/63	1.83 [1.08-3.1]	0.1	0.1
M166.0387T2.4	ESI -	75/60	1.00	73/63	0.92 [0.53-1.58]	55/80	0.54 [0.31-0.94]	0.1	0.1
M588.3285T19.13	ESI -	62/73	1.00	64/72	1.16 [0.69-1.93]	77/58	1.86 [1.08-3.21]	0.1	0.1
ATBC (acetyl tributyl citrate) M425.2172T19.31	ESI +	61/79	1.00	68/73	1.29 [0.75-2.19]	82/59	1.83 [1.08-3.12]	0.1	0.1
M558.9914T2.38	ESI +	64/76	1.00	64/77	1.04 [0.61-1.76]	83/58	1.75 [1.03-3]	0.1	0.1
M169.0362T3.1	ESI +	67/73	1.00	67/74	0.93 [0.53-1.63]	77/64	1.66 [0.93-2.96]	0.1	0.1
M82.9449T2.24	ESI +	60/80	1.00	74/67	1.64 [0.98-2.74]	77/64	1.69 [0.98-2.91]	0.1	0.1
M544.8888T2.06	ESI +	58/82	1.00	79/62	1.63 [0.97-2.71]	74/67	1.69 [1-2.87]	0.1	0.1
M112.0389T2.78	ESI -	63/72	1.00	64/72	1.01 [0.58-1.76]	76/59	1.65 [0.97-2.8]	0.1	0.1
M483.1958T22.45	ESI -	75/60	1.00	67/69	0.73 [0.43-1.24]	61/74	0.54 [0.3-0.95]	0.1	0.1
M114.9883T1.9	ESI +	62/78	1.00	72/69	1.46 [0.84-2.56]	77/64	1.72 [1.04-2.86]	0.1	0.1
M192.0305T3.01	ESI +	63/77	1.00	68/73	1.08 [0.65-1.79]	80/61	1.66 [1.01-2.73]	0.1	0.1
M185.1292T10.52	ESI +	65/75	1.00	68/73	1.2 [0.71-2.01]	78/63	1.73 [1.01-2.95]	0.1	0.1
M112.0076T1.91	ESI +	64/76	1.00	68/73	1.19 [0.71-1.99]	79/62	1.73 [1.02-2.94]	0.1	0.1
M121.0844T10.12	ESI +	68/72	1.00	66/75	1.03 [0.59-1.78]	77/64	1.62 [0.97-2.71]	0.1	0.1
M485.2107T23.95	ESI -	75/60	1.00	67/69	0.72 [0.42-1.24]	61/74	0.58 [0.34-0.98]	0.1	0.2
M294.9383T2.1	ESI +	76/64	1.00	70/71	0.73 [0.43-1.23]	65/76	0.57 [0.33-0.99]	0.1	0.2
M276.888T2.21	ESI +	64/76	1.00	68/73	1.03 [0.59-1.78]	79/62	1.55 [0.94-2.56]	0.2	0.2
pregnene-triol sulfate M413.2002T16.11	ESI -	60/75	1.00	67/69	1.24 [0.71-2.15]	76/59	1.68 [0.96-2.94]	0.2	0.2
M540.3295T19.89	ESI -	67/68	1.00	62/74	0.95 [0.57-1.59]	74/61	1.52 [0.85-2.71]	0.2	0.2
M73.5317T1.89	ESI +	63/77	1.00	70/71	1.13 [0.67-1.9]	78/63	1.57 [0.94-2.63]	0.2	0.2
M204.9569T2.03	ESI +	64/76	1.00	72/69	1.35 [0.81-2.25]	75/66	1.58 [0.95-2.62]	0.2	0.2
M111.0109T1.91	ESI +	65/75	1.00	71/70	1.24 [0.73-2.11]	75/66	1.58 [0.94-2.64]	0.2	0.2
L- $\gamma$ -glutamyl-L- threonine M231.1436T2.52	ESI +	68/72	1.00	65/76	1.03 [0.61-1.74]	78/63	1.49 [0.9-2.49]	0.2	0.2
O-succinyl-L- homoserine M240.0464T2.86	ESI -	75/60	1.00	70/66	0.83 [0.49-1.4]	58/77	0.64 [0.38-1.07]	0.2	0.2
M286.143T16.48	ESI +	68/72	1.00	79/62	1.24 [0.72-2.13]	64/77	0.78 [0.45-1.37]	0.2	0.3
M201.8018T2.34	ESI -	62/73	1.00	67/69	1.27 [0.73-2.22]	74/61	1.52 [0.9-2.57]	0.2	0.3
M359.0438T3.08	ESI +	68/72	1.00	67/74	0.96 [0.56-1.64]	76/65	1.43 [0.81-2.54]	0.2	0.3
M518.7449T2.21	ESI +	61/79	1.00	74/67	1.17 [0.69-1.97]	76/65	1.43 [0.84-2.43]	0.4	0.4

<sup>a</sup> All these additional information were provided for ions associated with breast cancer risk using continuous variables (FDR $\leq$ 2)

<sup>b</sup> Predicted values of breast cancer risk were obtained with conditional logistic regression models including ion intensity (continuous), age (continuous), BMI (continuous), smoking status (current smokers, former smokers and non-smokers), season of blood draw (a priori-defined periods: October-November/December-January-February/March-April-May), height (continuous), alcohol intake (continuous), physical activity (irregular/< 1h/d walking equivalent/ $\geq$  1h/d walking equivalent), education level (primary/secondary/superior), family history of breast cancer (yes/no), number of children (continuous), and use of hormone replacement therapy for menopause (yes/no) (reference model).

<sup>c</sup> logit $\geq$ 3SD were excluded from the plot

<sup>d</sup> Ions' intensities were categorized in tertiles. The lowest tertile was used as reference. Tests for non-linear trend were performed using the ordinal score.

<sup>e</sup> P-nontrend after Benjamini Hochberg correction for multiple testing.

**Supplementary Table S3 – Plasma metabolomic signatures associated with long-term breast cancer risk in the SU.VI.MAX prospective cohort.****Supplementary Table S3** Spearman correlation matrix of the identified associated metabolites

Metabolites	O-succinyl-L-homoserine		pregnene-triol sulfate		L-phenylalanine		L-valine/norvaline		5-aminovaleric acid		ATBC (acetyl tributyl citrate)		L-tryptophan		L-γ-glutamyl-L-threonine		
	R [95%CI]	P	R [95%CI]	P	R [95%CI]	P	R [95%CI]	P	R [95%CI]	P	R [95%CI]	P	R [95%CI]	P	R [95%CI]	P	
pregnene-triol sulfate	-0.03 [-0.12-0.07]	0.6															
L-phenylalanine	0.09 [-0.01-0.18]	0.1	0.09 [-0.01-0.18]	0.08													
L-valine/norvaline	-0.05 [-0.14-0.05]	0.4	0.04 [-0.06-0.14]	0.4	0.47 [0.39-0.54]	<0.0001											
5-aminovaleric acid	-0.18 [-0.28--0.09]	0.0002	0.11 [0.01-0.20]	0.03	0.19 [0.09-0.28]	0.0001	0.28 [0.19-0.37]	<0.0001									
ATBC (acetyl tributyl citrate)	-0.14 [-0.24--0.05]	0.004	0.06 [-0.04-0.16]	0.2	-0.06 [-0.15-0.04]	0.2	-0.09 [-0.19-0.00]	0.06	-0.10 [-0.19-0.00]	0.04							
L-tryptophan	0.00 [-0.10-0.10]	1.0	0.14 [0.04-0.23]	0.006	0.27 [0.18-0.36]	<0.0001	0.24 [0.15-0.33]	<0.0001	0.08 [-0.01-0.18]	0.08	0.05 [-0.05-0.14]	0.3					
L-γ-glutamyl-L-threonine	-0.06 [-0.16-0.04]	0.2	0.12 [0.02-0.21]	0.02	-0.01 [-0.11-0.08]	0.8	-0.02 [-0.12-0.07]	0.7	0.30 [0.21-0.39]	<0.0001	0.06 [-0.03-0.16]	0.2	-0.06 [-0.16-0.03]	0.2			
L-glutamine / L-isoglutamine	0.06 [-0.04-0.16]	0.2	0.09 [-0.01-0.19]	0.07	0.21 [0.12-0.30]	<0.0001	0.16 [0.06-0.25]	0.001	0.09 [0.00-0.18]	0.06	0.03 [-0.06-0.13]	0.5	0.20 [0.11-0.29]	<0.0001	0.14 [0.05-0.24]	0.003	

R, correlation coefficient

**Supplementary Tables S4, S5 and S6 – Plasma metabolomic signatures associated with long-term breast cancer risk in the SU.VI.MAX prospective cohort.**

**Supplementary results of sensitivity, stratification and heterogeneity analyses**

Results of sensitivity analysis after exclusion of cases diagnosed during the first year of follow-up are shown in **Supplementary Table S4**. All identified ions remained associated with breast cancer (FDR <0.05). No interaction was found between breast cancer risk and duration of follow-up (**Supplementary Table S5**). **Supplementary Table S6** showed that higher plasma level of glutamine was associated with an increased risk of breast cancer risk only in premenopausal women (P-interaction=0.003, FDR=0.2). The directions of all these associations were the same than the ones for the main analysis.

**Supplementary Table S4** Association between identified ions and breast cancer risk from multivariable conditional logistic regression<sup>a</sup> after the exclusion of cases diagnosed during the first year of follow-up - SU.VI.MAX cohort, France (1994-2007)

Name (Level of confidence for identification <sup>b</sup> )	Cases/controls	OR [95% CI]	P-trend	Corrected P-trend <sup>c</sup>
L-valine/norvaline (1)	197/197	1.45 [1.13-1.85]	0.003	0.009
L-glutamine / L-isoglutamine (1)	197/197	1.29 [1.04-1.60]	0.02	0.02
5-aminovaleic acid (1)	197/197	1.40 [1.09-1.80]	0.009	0.01
L-phenylalanine (1)	197/197	1.43 [1.15-1.79]	0.002	0.007
L-tryptophan (1)	197/197	1.36 [1.06-1.74]	0.02	0.02
L-γ-glutamyl-L-threonine (1)	197/197	1.41 [1.09-1.83]	0.009	0.01
O-succinyl-L-homoserine (1)	189/189	0.72 [0.56-0.91]	0.007	0.01
ATBC (acetyl tributyl citrate) (1)	197/197	1.41 [1.10-1.80]	0.007	0.01
pregnene-triol sulfate (2)	189/189	1.38 [1.07-1.77]	0.01	0.02

<sup>a</sup> Multivariable models were adjusted for age (continuous), BMI (continuous), smoking status (current smokers, former smokers and non-smokers), season of blood draw (a priori-defined periods: October-November/December-January-February/March-April-May), height (continuous), alcohol intake (continuous), physical activity (irregular/< 1h/d walking equivalent/≥ 1h/d walking equivalent), education level (primary/secondary/superior), family history of breast cancer (yes/no), number of children (continuous), and use of hormone replacement therapy for menopause (yes/no) (reference model). Tests for linear trend were performed using the continuous variables. ORs were presented for an increment of 1 SD for each metabolite.

<sup>b</sup> Metabolites were classified accordingly to Sumner et al (1) concerning the levels of confidence in the identification process: identified (level 1: confirmed by standard), putatively annotated (level 2: based upon physicochemical properties and/or spectral similarity with public/commercial spectral libraries).

<sup>c</sup> P-trend after Benjamini Hochberg correction for multiple testing.

**Supplementary Table S5** Association between identified ions and breast cancer risk in the women with a follow-up duration < or ≥ median (5.8 years) from multivariable conditional logistic regression<sup>a</sup> - SU.VI.MAX Cohort, France (1994-2007)

Name (Level of confidence for identification <sup>b</sup> )	Follow-up duration	Cases/controls	OR [95% CI]	P-trend	Corrected P-trend <sup>c</sup>	P-interaction	Corrected P-interaction <sup>c</sup>
L-valine/norvaline (1)	< median	105/105	1.36 [0.97-1.91]	0.08	0.1	0.5	0.9
	≥ median	106/106	1.68 [1.15-2.45]	0.008	0.05	0.5	0.9
L-glutamine / L-isoglutamine (1)	< median	105/105	1.60 [1.15-2.24]	0.006	0.04	0.2	0.8
	≥ median	106/106	1.16 [0.83-1.63]	0.4	0.4	0.2	0.8
5-aminovaleric acid (1)	< median	105/105	1.89 [1.24-2.89]	0.003	0.04	0.2	0.8
	≥ median	106/106	1.30 [0.92-1.83]	0.1	0.2	0.2	0.8
L-phenylalanine (1)	< median	105/105	1.40 [1.00-1.96]	0.05	0.09	0.6	0.9
	≥ median	106/106	1.54 [1.10-2.16]	0.01	0.06	0.6	0.9
L-tryptophan (1)	< median	105/105	1.53 [1.06-2.22]	0.02	0.06	0.6	0.9
	≥ median	106/106	1.35 [0.96-1.90]	0.09	0.2	0.6	0.9
L-γ-glutamyl-L-threonine (1)	< median	105/105	1.43 [1.03-2.00]	0.03	0.07	0.5	0.9
	≥ median	106/106	1.54 [1.00-2.38]	0.05	0.1	0.5	0.9
O-succinyl-L-homoserine (1)	< median	102/102	0.61 [0.41-0.91]	0.02	0.06	0.7	0.9
	≥ median	101/101	0.66 [0.48-0.93]	0.02	0.06	0.7	0.9
ATBC (acetyl tributyl citrate) (1)	< median	105/105	1.36 [0.97-1.92]	0.07	0.1	1.0	1.0
	≥ median	106/106	1.46 [1.00-2.13]	0.05	0.1	1.0	1.0
pregnene-triol sulfate (2)	< median	102/102	1.62 [1.05-2.51]	0.03	0.07	0.5	0.9
	≥ median	101/101	1.24 [0.89-1.74]	0.2	0.3	0.5	0.9

<sup>a</sup> Multivariable models were adjusted for age (continuous), BMI (continuous), smoking status (current smokers, former smokers and non-smokers), season of blood draw (a priori-defined periods: October-November/December-January-February/March-April-May), height (continuous), alcohol intake (continuous), physical activity (irregular/< 1h/d walking equivalent/≥ 1h/d walking equivalent), education level (primary/secondary/superior), family history of breast cancer (yes/no), number of children (continuous), and use of hormone replacement therapy for menopause (yes/no) (reference model). Tests for linear trend were performed using the continuous variables. ORs were presented for an increment of 1 SD for each metabolite.

<sup>b</sup> Metabolites were classified accordingly to Sumner et al (1) concerning the levels of confidence in the identification process: identified (level 1: confirmed by standard), putatively annotated (level 2: based upon physicochemical properties and/or spectral similarity with public/commercial spectral libraries).

<sup>c</sup> P after Benjamini Hochberg correction for multiple testing.

**Supplementary Table S6** Association between identified ions and breast cancer risk in premenopausal or postmenopausal women (at the time of cancer diagnosis) from multivariable conditional logistic regression<sup>a</sup> - SU.VI.MAX cohort, France (1994-2007)

Name (Level of confidence for identification <sup>b</sup> )	Menopausal status at cancer diagnosis	Cases/controls	OR [95% CI]	P-trend	Corrected P-trend <sup>c</sup>	P-interaction	Corrected P-interaction <sup>c</sup>
L-valine/norvaline (1)	premenopausal	78/78	1.54 [1.00-2.38]	0.05	0.1	0.9	1.0
	postmenopausal	133/133	1.50 [1.10-2.05]	0.01	0.03	0.9	1.0
L-glutamine / L-isoglutamine (1)	premenopausal	78/78	2.49 [1.49-4.17]	0.0005	0.04	0.003	0.2
	postmenopausal	133/133	1.03 [0.78-1.35]	0.9	0.9	0.003	0.2
5-aminovaleric acid (1)	premenopausal	78/78	1.57 [1.01-2.45]	0.05	0.1	0.8	1.0
	postmenopausal	133/133	1.39 [1.03-1.87]	0.03	0.06	0.8	1.0
L-phenylalanine (1)	premenopausal	78/78	1.66 [1.10-2.49]	0.01	0.1	0.4	0.7
	postmenopausal	133/133	1.30 [0.98-1.71]	0.07	0.1	0.4	0.7
L-tryptophan (1)	premenopausal	78/78	1.67 [1.01-2.76]	0.04	0.1	0.3	0.7
	postmenopausal	133/133	1.32 [0.99-1.77]	0.06	0.1	0.3	0.7
L-γ-glutamyl-L-threonine (1)	premenopausal	78/78	1.45 [0.96-2.20]	0.08	0.2	0.9	1.0
	postmenopausal	133/133	1.25 [0.90-1.74]	0.2	0.2	0.9	1.0
O-succinyl-L-homoserine (1)	premenopausal	78/78	0.81 [0.52-1.27]	0.4	0.4	0.1	0.7
	postmenopausal	125/125	0.62 [0.44-0.86]	0.004	0.02	0.1	0.7
ATBC (acetyl tributyl citrate) (1)	premenopausal	78/78	1.28 [0.83-1.98]	0.3	0.3	0.5	0.9
	postmenopausal	133/133	1.48 [1.09-2.02]	0.01	0.03	0.5	0.9
	postmenopausal	133/133	1.27 [0.96-1.67]	0.03	0.1	0.3	0.7
pregnene-triol sulfate (2)	premenopausal	78/78	1.67 [1.06-2.60]	0.2	0.2	0.3	0.7
	postmenopausal	125/125	1.25 [0.92-1.69]	0.05	0.1	0.9	1.0

<sup>a</sup> Multivariable models were adjusted for age (continuous), BMI (continuous), smoking status (current smokers, former smokers and non-smokers), season of blood draw (a priori-defined periods: October-November/December-January-February/March-April-May), height (continuous), alcohol intake (continuous), physical activity (irregular/< 1h/d walking equivalent/≥ 1h/d walking equivalent), education level (primary/secondary/superior), family history of breast cancer (yes/no), number of children (continuous), and use of hormone replacement therapy for menopause (yes/no) (reference model). Tests for linear trend were performed using the continuous variables. ORs were presented for an increment of 1 SD for each metabolite.

<sup>b</sup> Metabolites were classified accordingly to Sumner et al (1) concerning the levels of confidence in the identification process: identified (level 1: confirmed by standard), putatively annotated (level 2: based upon physicochemical properties and/or spectral similarity with public/commercial spectral libraries).

<sup>c</sup> P after Benjamini Hochberg correction for multiple testing.

## REFERENCES

1. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, et al. Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics*. 2007;3:211–21.

## Annexe 3 Associations entre les signaux RMN et le risque de cancer de la prostate issues des régressions logistiques conditionnelles

Métabolites mis en évidence avec la séquence NOESY1D (N=162 cas/162 témoins)			
Déplacements chimiques (ppm) (assignations)	OR <sup>a</sup> [IC95%]	P	P-corrigée
0,4867	1,27 [0,92-1,74]	0,2	0,6
0,5267	1,25 [0,91-1,73]	0,2	0,6
0,5667	1,12 [0,84-1,5]	0,4	0,8
0,6067	1,1 [0,84-1,45]	0,5	0,8
0,6467	1,04 [0,8-1,35]	0,8	0,9
0,6867	0,93 [0,74-1,18]	0,6	0,8
0,7267	0,98 [0,77-1,25]	0,9	1,0
0,7667	1,02 [0,8-1,3]	0,9	1,0
0,8067	0,88 [0,69-1,13]	0,3	0,7
0,9107	0,87 [0,69-1,11]	0,3	0,7
0,9697, 1,021 (valine)	1,37 [1,07-1,76]	0,01	0,3
0,9997	0,89 [0,7-1,12]	0,3	0,7
1,051	1,1 [0,86-1,4]	0,5	0,8
1,0667	1,06 [0,81-1,38]	0,7	0,9
1,0887	0,97 [0,75-1,25]	0,8	1,0
1,1267	0,92 [0,71-1,17]	0,5	0,8
1,1667	0,88 [0,68-1,14]	0,3	0,7
1,2067	0,88 [0,69-1,12]	0,3	0,7
1,3107	0,88 [0,7-1,12]	0,3	0,7
1,3507	0,88 [0,69-1,13]	0,3	0,7
1,3907	0,92 [0,72-1,17]	0,5	0,8
1,4287	0,93 [0,73-1,19]	0,6	0,8
1,4467	0,95 [0,74-1,21]	0,7	0,9
1,5217	0,9 [0,71-1,14]	0,4	0,7
1,5907	0,94 [0,74-1,19]	0,6	0,8
1,6307	1,07 [0,84-1,37]	0,6	0,8
1,6357	1,13 [0,88-1,46]	0,3	0,7
1,6467 (pic non-défini)	1,37 [1,05-1,8]	0,02	0,3
1,677	0,95 [0,74-1,23]	0,7	0,9
1,707	1,07 [0,85-1,36]	0,6	0,8
1,7507	1,24 [0,94-1,63]	0,1	0,6
1,7907	1,15 [0,89-1,48]	0,3	0,7
1,8307	1,11 [0,87-1,42]	0,4	0,7
1,8507	1,25 [0,97-1,62]	0,09	0,6
1,8867	1,14 [0,89-1,45]	0,3	0,7
1,927	1,1 [0,85-1,44]	0,5	0,8
2,018	1,03 [0,81-1,31]	0,8	0,9
2,057	1,18 [0,93-1,51]	0,2	0,6
2,1507	0,92 [0,72-1,16]	0,5	0,8
2,1737	0,89 [0,7-1,13]	0,4	0,7
2,2637	1,01 [0,8-1,28]	0,9	1,0

## ANNEXE 3

2,3107	1,01 [0,78-1,31]	0,9	1,0
2,3507	0,92 [0,72-1,17]	0,5	0,8
2,374 (pic non-défini)	0,73 [0,57-0,95]	0,02	0,3
2,4137 (glutamine)	1,3 [1-1,7]	0,05	0,4
2,4707	1,1 [0,83-1,44]	0,5	0,8
2,5017	1,05 [0,79-1,39]	0,7	0,9
2,5507	1,2 [0,92-1,58]	0,2	0,6
2,5797 (pic non-défini)	1,42 [1,07-1,88]	0,02	0,3
2,6267	0,91 [0,71-1,17]	0,5	0,8
2,647	1,08 [0,85-1,38]	0,5	0,8
2,708	0,85 [0,67-1,08]	0,2	0,6
2,8467	1,24 [0,95-1,61]	0,1	0,6
2,8867 (albumin lysyl)	1,48 [1,12-1,95]	0,0061	0,3
2,9267 (non identifié)	1,39 [1,06-1,82]	0,02	0,3
2,9777 (albumin lysyl, créatine <sup>b</sup> )	1,37 [1,04-1,8]	0,02	0,3
3,0706	1,27 [0,98-1,63]	0,07	0,6
3,1106	1,24 [0,97-1,58]	0,09	0,6
3,1296	1,26 [0,98-1,63]	0,07	0,6
3,1666	1,02 [0,79-1,32]	0,9	1,0
3,2696	0,95 [0,73-1,23]	0,7	0,9
3,3106	1,23 [0,96-1,58]	0,1	0,6
3,3136	1,14 [0,89-1,46]	0,3	0,7
3,3656	1,12 [0,88-1,43]	0,4	0,7
3,3666	1,01 [0,8-1,27]	0,9	1,0
3,4376	1 [0,8-1,26]	1,0	1,0
3,4966	1 [0,79-1,27]	1,0	1,0
3,5316	1,03 [0,77-1,37]	0,8	1,0
3,5876	1,09 [0,84-1,42]	0,5	0,8
3,6106	1,01 [0,78-1,31]	0,9	1,0
3,7066	0,95 [0,73-1,23]	0,7	0,9
3,7906	1,12 [0,86-1,45]	0,4	0,7
3,7926	1,04 [0,79-1,36]	0,8	0,9
3,8066	1,07 [0,84-1,37]	0,6	0,8
3,8606	0,98 [0,75-1,28]	0,9	1,0
3,918	1,01 [0,79-1,3]	0,9	1,0
3,9506	0,92 [0,72-1,17]	0,5	0,8
3,9516	1,21 [0,93-1,57]	0,1	0,6
4,031	0,84 [0,65-1,07]	0,2	0,6
4,075	1,07 [0,83-1,39]	0,6	0,8
4,1506	1,2 [0,91-1,57]	0,2	0,6
4,1906	1,19 [0,9-1,56]	0,2	0,6
4,2136	0,94 [0,74-1,19]	0,6	0,8
4,3106	1,24 [0,96-1,59]	0,1	0,6
4,3506	1,16 [0,91-1,48]	0,2	0,6
4,3815	1,12 [0,87-1,43]	0,4	0,7
5,0465	0,85 [0,65-1,11]	0,2	0,6

## ANNEXE 3

5,0865	0,83 [0,63-1,1]	0,2	0,6
5,1265	0,81 [0,62-1,07]	0,1	0,6
5,1665	0,86 [0,67-1,11]	0,2	0,6
5,2065	0,9 [0,7-1,17]	0,4	0,8
5,2465	0,86 [0,68-1,1]	0,2	0,6
5,3505	0,85 [0,67-1,09]	0,2	0,6
5,3905	0,83 [0,63-1,08]	0,2	0,6
5,4305	0,81 [0,61-1,07]	0,1	0,6
5,4705	0,8 [0,6-1,07]	0,1	0,6
5,5105	0,81 [0,61-1,08]	0,2	0,6
5,5505	0,85 [0,64-1,12]	0,2	0,7
5,5905	0,87 [0,67-1,15]	0,3	0,7
5,6305	0,9 [0,7-1,17]	0,4	0,8
5,6385	0,87 [0,66-1,14]	0,3	0,7
5,6465	0,81 [0,62-1,08]	0,2	0,6
5,6865, 5,7265, 5,8304 (urée)	0,7 [0,54-0,92]	0,009	0,3
5,8354 (pic non-défini)	0,71 [0,52-0,96]	0,02	0,3
5,8614 (pic non-défini)	0,67 [0,49-0,91]	0,01	0,3
5,8874	1,01 [0,8-1,26]	1,0	1,0
5,9654	1,11 [0,86-1,42]	0,4	0,8
5,9664	1,01 [0,8-1,27]	0,9	1,0
6,0704 (pic non-défini)	0,63 [0,45-0,89]	0,008	0,3
6,1104	0,93 [0,72-1,2]	0,6	0,8
6,1184	0,81 [0,62-1,07]	0,1	0,6
6,1264	0,94 [0,73-1,21]	0,6	0,8
6,1664	1,06 [0,85-1,32]	0,6	0,8
6,2084	1,07 [0,83-1,37]	0,6	0,8
6,2464	1,05 [0,82-1,34]	0,7	0,9
6,2864	0,94 [0,73-1,22]	0,6	0,8
6,3274	0,98 [0,75-1,28]	0,9	1,0
6,3664	0,99 [0,74-1,3]	0,9	1,0
6,4064	1,02 [0,78-1,34]	0,9	1,0
6,4464	1,05 [0,8-1,36]	0,7	0,9
6,4864	1,17 [0,89-1,54]	0,3	0,7
6,5264 (pic non-défini)	1,32 [1,03-1,69]	0,03	0,3
6,5664	1,18 [0,92-1,52]	0,2	0,6
6,6064	0,88 [0,67-1,15]	0,3	0,7
6,6464	0,81 [0,62-1,07]	0,1	0,6
6,6864	0,9 [0,68-1,19]	0,5	0,8
6,7264	0,96 [0,73-1,27]	0,8	0,9
6,7664	1,25 [0,96-1,63]	0,09	0,6
6,8634	1,19 [0,91-1,57]	0,2	0,6
6,9054	0,97 [0,75-1,26]	0,8	1,0
6,9264	1,15 [0,9-1,48]	0,3	0,7
6,9664 (pic non-défini)	1,36 [1,05-1,76]	0,02	0,3
7,0074	1,16 [0,88-1,53]	0,3	0,7



## ANNEXE 3

7,1104	0,98 [0,75-1,27]	0,9	1,0
7,1503, 7,1833 (tyrosine)	1,33 [1,03-1,71]	0,03	0,3
7,2093 (pic non-défini)	1,33 [1,02-1,73]	0,03	0,3
7,2703	1,09 [0,86-1,4]	0,5	0,8
7,2763	1,23 [0,95-1,59]	0,1	0,6
7,3503	1,2 [0,92-1,58]	0,2	0,6
7,3673	1,18 [0,9-1,54]	0,2	0,6
7,4303	1 [0,77-1,29]	1,0	1,0
7,4703	0,84 [0,65-1,09]	0,2	0,6
7,5103	0,82 [0,64-1,06]	0,1	0,6
7,5433	1,08 [0,84-1,39]	0,6	0,8
7,5663	1,01 [0,78-1,32]	0,9	1,0
7,6063	0,95 [0,73-1,25]	0,7	0,9
7,6463	1,12 [0,86-1,45]	0,4	0,7
7,6863 (histidine)	1,34 [1,03-1,73]	0,03	0,3
7,7723	1,1 [0,84-1,43]	0,5	0,8
7,8303	0,92 [0,71-1,19]	0,5	0,8
7,8703	0,85 [0,66-1,11]	0,2	0,6
7,9083	0,88 [0,68-1,15]	0,3	0,7
7,9173	0,96 [0,73-1,25]	0,8	0,9
7,9263	0,87 [0,67-1,14]	0,3	0,7
7,9663	0,82 [0,63-1,07]	0,2	0,6
<b>Métabolites mis en évidence avec la séquence CPMG (N=195 cas/195 témoins)</b>			
Déplacement chimique (ppm) (assignations)	OR [IC95%]	P	P-corrigée
0,0003	1,01 [0,79-1,29]	0,9	1,0
0,0581	1,13 [0,88-1,43]	0,3	0,7
0,0871	1,05 [0,82-1,35]	0,7	0,9
0,1721	1,13 [0,88-1,46]	0,3	0,7
0,2321	1,05 [0,82-1,34]	0,7	0,9
0,2471	1,06 [0,83-1,36]	0,6	0,9
0,3181	1,06 [0,83-1,36]	0,6	0,9
0,3271	1,07 [0,83-1,37]	0,6	0,9
0,3681	1,04 [0,81-1,34]	0,7	1,0
0,4091	1,07 [0,84-1,38]	0,6	0,9
0,4471	1,07 [0,83-1,37]	0,6	0,9
0,4871	1,14 [0,88-1,47]	0,3	0,7
0,5271	1,16 [0,9-1,5]	0,2	0,7
0,5671	1,08 [0,83-1,4]	0,6	0,9
0,6071	1,06 [0,83-1,37]	0,6	0,9
0,6471	0,97 [0,77-1,23]	0,8	1,0
0,7131	1,04 [0,8-1,33]	0,8	1,0
0,7271	1,03 [0,8-1,33]	0,8	1,0
0,7671	1,03 [0,8-1,33]	0,8	1,0
0,8071	0,88 [0,7-1,12]	0,3	0,7
0,9107	1,02 [0,8-1,3]	0,9	1,0
0,9697	1,26 [0,99-1,6]	0,06	0,4

## ANNEXE 3

0,9997	0,89 [0,71-1,13]	0,3	0,7
1,021 (valine)	1,29 [1,02-1,64]	0,04	0,3
1,051	0,98 [0,77-1,25]	0,9	1,0
1,0531	1,13 [0,89-1,43]	0,3	0,7
1,0811	1,07 [0,85-1,35]	0,6	0,9
1,0881	1,01 [0,8-1,27]	0,9	1,0
1,1271	0,76 [0,58-1,01]	0,06	0,4
1,145	1,16 [0,9-1,5]	0,3	0,7
1,1671	0,94 [0,74-1,19]	0,6	0,9
1,2071	0,89 [0,71-1,13]	0,3	0,7
1,3111	0,89 [0,7-1,12]	0,3	0,7
1,3511	0,88 [0,7-1,12]	0,3	0,7
1,3911	0,92 [0,72-1,17]	0,5	0,8
1,4311	0,9 [0,71-1,14]	0,4	0,7
1,4371	0,88 [0,7-1,12]	0,3	0,7
1,4471	0,86 [0,68-1,08]	0,2	0,7
1,5211	0,9 [0,71-1,14]	0,4	0,7
1,5271	0,9 [0,71-1,13]	0,4	0,7
1,6311	1,02 [0,81-1,29]	0,9	1,0
1,6361	1,1 [0,87-1,4]	0,4	0,8
1,6471	1,12 [0,87-1,46]	0,4	0,7
1,707	1,01 [0,8-1,27]	0,9	1,0
1,759	0,92 [0,73-1,17]	0,5	0,8
1,7911	1,18 [0,93-1,49]	0,2	0,7
1,8211	1,14 [0,9-1,44]	0,3	0,7
1,849	1,22 [0,96-1,55]	0,1	0,5
1,899	1,02 [0,78-1,33]	0,9	1,0
1,927	1,16 [0,89-1,51]	0,3	0,7
2,017	1,01 [0,79-1,28]	0,9	1,0
2,0631	1,05 [0,8-1,38]	0,7	1,0
2,0941	1,11 [0,85-1,43]	0,5	0,8
2,1511 (pic non-défini)	0,77 [0,6-0,98]	0,04	0,3
2,1661	0,84 [0,65-1,07]	0,2	0,7
2,1671	0,9 [0,71-1,14]	0,4	0,7
2,2641	1,13 [0,88-1,45]	0,3	0,7
2,277	0,93 [0,73-1,19]	0,6	0,9
2,3111	0,87 [0,67-1,12]	0,3	0,7
2,315	1 [0,78-1,27]	1,0	1,0
2,354	0,97 [0,76-1,24]	0,8	1,0
2,372	0,84 [0,66-1,07]	0,2	0,7
2,411	1,31 [0,98-1,73]	0,06	0,4
2,4711	0,79 [0,62-1,02]	0,07	0,4
2,5011	0,79 [0,6-1,03]	0,08	0,4
2,5511	0,99 [0,77-1,27]	0,9	1,0
2,5711	1,2 [0,93-1,55]	0,2	0,7
2,6121	0,89 [0,69-1,14]	0,3	0,7

## ANNEXE 3

2,6471	1,08 [0,85-1,37]	0,5	0,8
2,705	1,03 [0,79-1,35]	0,8	1,0
2,715	0,84 [0,67-1,07]	0,2	0,7
2,8481	0,98 [0,76-1,27]	0,9	1,0
2,873 (albumine)	1,51 [1,13-2,02]	0,005	0,2
2,9271	1,24 [0,96-1,61]	0,1	0,5
2,9671	1,24 [0,97-1,58]	0,08	0,4
2,98	1,31 [0,98-1,74]	0,07	0,4
3,028	1 [0,78-1,28]	1,0	1,0
3,044	1,06 [0,83-1,35]	0,6	0,9
3,071	1,07 [0,84-1,36]	0,6	0,9
3,111	0,99 [0,78-1,27]	1,0	1,0
3,129	1,03 [0,8-1,33]	0,8	1,0
3,167	1,05 [0,8-1,36]	0,7	1,0
3,269	0,84 [0,65-1,09]	0,2	0,7
3,311	0,86 [0,67-1,12]	0,3	0,7
3,312	0,99 [0,78-1,26]	0,9	1,0
3,367	1,01 [0,8-1,27]	0,9	1,0
3,437	1 [0,79-1,26]	1,0	1,0
3,496	0,97 [0,76-1,25]	0,8	1,0
3,587	1,01 [0,8-1,28]	0,9	1,0
3,616	0,97 [0,76-1,25]	0,8	1,0
3,713	1 [0,79-1,26]	1,0	1,0
3,791	0,98 [0,76-1,25]	0,9	1,0
3,795	1,06 [0,84-1,33]	0,6	0,9
3,86	1,02 [0,79-1,31]	0,9	1,0
3,919	1,02 [0,8-1,3]	0,9	1,0
3,932	1,09 [0,85-1,39]	0,5	0,8
3,951	0,95 [0,75-1,21]	0,7	0,9
3,978	0,92 [0,71-1,2]	0,5	0,9
4,021	0,9 [0,72-1,13]	0,4	0,7
4,034	0,87 [0,68-1,12]	0,3	0,7
4,055	0,92 [0,72-1,17]	0,5	0,8
4,081	0,89 [0,69-1,14]	0,4	0,7
4,151	0,86 [0,66-1,12]	0,3	0,7
4,184	0,83 [0,64-1,08]	0,2	0,7
4,198	0,88 [0,68-1,15]	0,4	0,7
4,208	0,86 [0,68-1,09]	0,2	0,7
4,311	0,86 [0,66-1,12]	0,3	0,7
4,351	0,84 [0,64-1,09]	0,2	0,7
4,39	0,83 [0,64-1,09]	0,2	0,7
5,6869, 5,7269 (urée)	0,75 [0,59-0,97]	0,03	0,3
5,0069 (pic non-défini)	1,32 [1,01-1,72]	0,04	0,3
5,0809 (pic non-défini)	1,34 [1,02-1,75]	0,04	0,3
5,0869	1,29 [0,99-1,69]	0,06	0,4
5,1269	1,13 [0,88-1,45]	0,4	0,7

## ANNEXE 3

5,1669	0,92 [0,73-1,16]	0,5	0,8
5,2069	0,95 [0,75-1,22]	0,7	0,9
5,2469	0,86 [0,68-1,09]	0,2	0,7
5,3509	0,88 [0,69-1,11]	0,3	0,7
5,3889	0,9 [0,7-1,15]	0,4	0,8
5,4309	0,91 [0,71-1,17]	0,5	0,8
5,4709	1,02 [0,81-1,3]	0,8	1,0
5,5099	0,94 [0,74-1,19]	0,6	0,9
5,5279	1,11 [0,88-1,39]	0,4	0,7
5,6299	0,88 [0,7-1,11]	0,3	0,7
5,6339	1,09 [0,87-1,36]	0,5	0,8
5,6469	0,89 [0,69-1,14]	0,3	0,7
5,8308	0,85 [0,67-1,07]	0,2	0,7
5,8638 (pic non-défini)	0,72 [0,55-0,94]	0,01	0,3
5,8888	1,1 [0,86-1,4]	0,4	0,8
5,9688	1,12 [0,89-1,41]	0,4	0,7
6,0698	1,02 [0,81-1,29]	0,9	1,0
6,0998	1,19 [0,94-1,51]	0,2	0,7
6,1328	1,06 [0,83-1,35]	0,6	0,9
6,1698	1,27 [1-1,63]	0,05	0,4
6,2068	1,13 [0,89-1,45]	0,3	0,7
6,2478	1,12 [0,88-1,43]	0,3	0,7
6,3398	1,08 [0,84-1,39]	0,5	0,9
6,3668	1,13 [0,87-1,47]	0,4	0,7
6,4068	0,99 [0,78-1,26]	0,9	1,0
6,4468	1,26 [0,98-1,61]	0,07	0,4
6,4918	1,31 [1-1,72]	0,05	0,4
6,5268 (pic non-défini)	1,34 [1,02-1,77]	0,03	0,3
6,5668	1,1 [0,84-1,43]	0,5	0,8
6,6128	1,04 [0,81-1,35]	0,8	1,0
6,6478	1,08 [0,83-1,41]	0,6	0,9
6,6868	1,19 [0,91-1,56]	0,2	0,7
6,7278	1,08 [0,82-1,42]	0,6	0,9
6,7668 (pic non-défini)	1,35 [1,03-1,77]	0,03	0,3
6,857 (tyrosine)	1,4 [1,06-1,85]	0,02	0,3
6,912	0,84 [0,65-1,08]	0,2	0,7
6,9428 (pic non-défini)	1,35 [1,06-1,72]	0,02	0,3
6,968 (pic non-défini)	1,45 [1,1-1,92]	0,009	0,2
7,023	1,01 [0,8-1,28]	0,9	1,0
7,054	1,32 [0,97-1,81]	0,08	0,4
7,0708	0,93 [0,73-1,19]	0,6	0,9
7,1078	1,13 [0,89-1,44]	0,3	0,7
7,1198	0,9 [0,7-1,16]	0,4	0,8
7,128	0,87 [0,67-1,12]	0,3	0,7
7,159	1,11 [0,86-1,42]	0,4	0,8
7,2107 (pic non-défini)	1,76 [1,25-2,48]	0,001	0,1

## ANNEXE 3

7,2697 (pic non-défini)	1,58 [1,18-2,12]	0,002	0,1
7,2787 (phenylalanine)	1,39 [1,08-1,79]	0,01	0,2
7,3437	1,03 [0,81-1,3]	0,8	1,0
7,3747	0,99 [0,77-1,27]	0,9	1,0
7,4307	1,02 [0,8-1,3]	0,8	1,0
7,4707	0,82 [0,63-1,07]	0,1	0,7
7,5077	0,87 [0,67-1,12]	0,3	0,7
7,5387 (3-methylhistidine)	1,37 [1,05-1,8]	0,02	0,3
7,6287	0,97 [0,77-1,22]	0,8	1,0
7,6467	1,11 [0,87-1,41]	0,4	0,7
7,6867 (histidine)	1,46 [1,12-1,88]	0,004	0,2
7,7717	1 [0,79-1,28]	1,0	1,0
7,8307 (pic non-défini)	0,73 [0,56-0,95]	0,02	0,3
7,8707 (pic non-défini)	0,7 [0,54-0,92]	0,01	0,2
7,9107 (pic non-défini)	0,68 [0,51-0,9]	0,007	0,2
7,9207 (pic non-défini)	0,76 [0,59-0,98]	0,03	0,3
7,9837 (pic non-défini)	0,66 [0,5-0,86]	0,002	0,1

OR, Odds Ratio ; IC, intervalle de confiance

<sup>a</sup> Modèles de régression logistique conditionnelle sur les signaux RMN en continu divisé par l'écart type ajustés sur l'âge, l'IMC, le statut tabagique, la taille, la consommation d'alcool intake, l'activité physique, le niveau d'éducation, les antécédents familiaux de cancer de la prostate. Des tests de tendance linéaire ont été réalisés.

<sup>b</sup> Assignations multiples pour le même signal RMN dû à la superposition spectrale

## Annexe 4 Corrélations significatives entre les signaux RMN 2 à 2 identifiés (cas-témoins niché, cancer de la prostate)

Métabolite	Métabolite	R[IC95%]	P
Glutamine_NOESY	Valine_CPMG	0,29 [ 0,19-0,39 ]	<0,0001
Glutamine_NOESY	Albumine_CPMG	0,51 [ 0,43-0,59 ]	<0,0001
Glutamine_NOESY	Tyrosine_CPMG	0,18 [ 0,07-0,28 ]	0,001
Glutamine_NOESY	Phénylalanine_CPMG	0,14 [ 0,03-0,24 ]	0,01
Glutamine_NOESY	3-methylhistidine_CPMG	0,39 [ 0,29-0,47 ]	<0,0001
Glutamine_NOESY	Histidine_CPMG	0,58 [ 0,5-0,64 ]	<0,0001
Glutamine_NOESY	Urée_CPMG	-0,37 [ -0,46--0,27 ]	<0,0001
Albumine_CPMG	Valine_CPMG	0,3 [ 0,2-0,39 ]	<0,0001
Albumine Lysyl_NOESY	Valine_CPMG	0,21 [ 0,1-0,31 ]	0,0001
Albumine Lysyl_NOESY	Glutamine_NOESY	0,62 [ 0,55-0,68 ]	<0,0001
Albumine Lysyl_NOESY	Albumine_CPMG	0,52 [ 0,44-0,6 ]	<0,0001
Albumine Lysyl_NOESY	Phénylalanine_CPMG	0,17 [ 0,06-0,27 ]	0,002
Albumine Lysyl_NOESY	3-methylhistidine_CPMG	0,54 [ 0,46-0,61 ]	<0,0001
Albumine Lysyl_NOESY	Histidine_CPMG	0,44 [ 0,35-0,52 ]	<0,0001
Albumine Lysyl_NOESY	Urée_CPMG	-0,29 [ -0,39--0,19 ]	<0,0001
Albumine Lysyl, créatine_NOESY	Valine_CPMG	0,27 [ 0,17-0,37 ]	<0,0001
Albumine Lysyl, créatine_NOESY	Glutamine_NOESY	0,43 [ 0,34-0,51 ]	<0,0001
Albumine Lysyl, créatine_NOESY	Albumine_CPMG	0,28 [ 0,18-0,38 ]	<0,0001
Albumine Lysyl, créatine_NOESY	Albumine Lysyl_NOESY	0,81 [ 0,77-0,85 ]	<0,0001
Albumine Lysyl, créatine_NOESY	Tyrosine_CPMG	0,18 [ 0,07-0,28 ]	0,001
Albumine Lysyl, créatine_NOESY	Phénylalanine_CPMG	0,34 [ 0,24-0,44 ]	<0,0001
Albumine Lysyl, créatine_NOESY	3-methylhistidine_CPMG	0,42 [ 0,33-0,51 ]	<0,0001
Albumine Lysyl, créatine_NOESY	Histidine_CPMG	0,36 [ 0,27-0,45 ]	<0,0001
Albumine Lysyl, créatine_NOESY	Urée_CPMG	-0,11 [ -0,22-0 ]	0,047
Tyrosine_CPMG	Valine_CPMG	0,42 [ 0,33-0,51 ]	<0,0001
Phénylalanine_CPMG	Valine_CPMG	0,23 [ 0,12-0,33 ]	<0,0001
Phénylalanine_CPMG	Albumine_CPMG	0,27 [ 0,16-0,37 ]	<0,0001
Phénylalanine_CPMG	Tyrosine_CPMG	0,18 [ 0,07-0,28 ]	0,001
3-methylhistidine_CPMG	Valine_CPMG	0,21 [ 0,1-0,31 ]	0,0002
3-methylhistidine_CPMG	Albumine_CPMG	0,49 [ 0,41-0,57 ]	<0,0001
3-methylhistidine_CPMG	Phénylalanine_CPMG	0,43 [ 0,33-0,51 ]	<0,0001
Histidine_CPMG	Valine_CPMG	0,38 [ 0,28-0,47 ]	<0,0001
Histidine_CPMG	Albumine_CPMG	0,38 [ 0,28-0,47 ]	<0,0001
Histidine_CPMG	Tyrosine_CPMG	0,31 [ 0,21-0,4 ]	<0,0001
Histidine_CPMG	Phénylalanine_CPMG	0,4 [ 0,31-0,49 ]	<0,0001
Histidine_CPMG	3-methylhistidine_CPMG	0,6 [ 0,53-0,67 ]	<0,0001
Tyrosine_NOESY	Valine_CPMG	0,24 [ 0,14-0,34 ]	<0,0001
Tyrosine_NOESY	Glutamine_NOESY	0,54 [ 0,46-0,61 ]	<0,0001
Tyrosine_NOESY	Albumine_CPMG	0,45 [ 0,36-0,53 ]	<0,0001
Tyrosine_NOESY	Albumine Lysyl_NOESY	0,74 [ 0,69-0,79 ]	<0,0001
Tyrosine_NOESY	Albumine Lysyl, créatine_NOESY	0,77 [ 0,72-0,81 ]	<0,0001
Tyrosine_NOESY	Tyrosine_CPMG	0,29 [ 0,19-0,39 ]	<0,0001
Tyrosine_NOESY	Phénylalanine_CPMG	0,25 [ 0,15-0,35 ]	<0,0001

## ANNEXE 4

Tyrosine_NOESY	3-methylhistidine_CPMG	0,37 [ 0,27-0,46 ]	<0,0001
Tyrosine_NOESY	Histidine_CPMG	0,3 [ 0,19-0,39 ]	<0,0001
Urée_CPMG	Phénylalanine_CPMG	0,12 [ 0,02-0,23 ]	0,02
Urée_CPMG	3-methylhistidine_CPMG	-0,16 [ -0,26--0,05 ]	0,003
Urée_CPMG	Histidine_CPMG	-0,48 [ -0,56--0,4 ]	<0,0001
Urée_NOESY	Glutamine_NOESY	-0,55 [ -0,62--0,47 ]	<0,0001
Urée_NOESY	Albumine_CPMG	-0,23 [ -0,33--0,12 ]	<0,0001
Urée_NOESY	Albumine Lysyl_NOESY	-0,58 [ -0,64--0,5 ]	<0,0001
Urée_NOESY	Albumine Lysyl, créatine_NOESY	-0,38 [ -0,47--0,28 ]	<0,0001
Urée_NOESY	3-methylhistidine_CPMG	-0,32 [ -0,41--0,22 ]	<0,0001
Urée_NOESY	Histidine_CPMG	-0,54 [ -0,61--0,46 ]	<0,0001
Urée_NOESY	Tyrosine_NOESY	-0,36 [ -0,45--0,26 ]	<0,0001
Urée_NOESY	Urée_CPMG	0,9 [ 0,88-0,92 ]	<0,0001
Valine_NOESY	Valine_CPMG	0,85 [ 0,82-0,88 ]	<0,0001
Valine_NOESY	Glutamine_NOESY	0,38 [ 0,28-0,47 ]	<0,0001
Valine_NOESY	Albumine_CPMG	0,27 [ 0,17-0,37 ]	<0,0001
Valine_NOESY	Albumine Lysyl_NOESY	0,21 [ 0,11-0,31 ]	0,0001
Valine_NOESY	Albumine Lysyl, créatine_NOESY	0,19 [ 0,09-0,3 ]	0,0004
Valine_NOESY	Tyrosine_CPMG	0,36 [ 0,26-0,45 ]	<0,0001
Valine_NOESY	Phénylalanine_CPMG	0,13 [ 0,02-0,23 ]	0,02
Valine_NOESY	3-methylhistidine_CPMG	0,14 [ 0,03-0,24 ]	0,01
Valine_NOESY	Histidine_CPMG	0,28 [ 0,18-0,38 ]	<0,0001
Valine_NOESY	Tyrosine_NOESY	0,24 [ 0,13-0,34 ]	<0,0001

R, coefficient de corrélation; IC, intervalle de confiance

## Annexe 5 Signaux RMN associés avec le risque de cancer de la prostate (intensités catégorisées en tertiles)

Métabolites mis en évidence par la séquence NOESY1D (N=162 cas/162 témoins)									
Assignations	Déplacement chimique (ppm)	T1 (ref)		T2		T3		P	P-corrigée
		cas/témoins	OR <sup>a</sup>	cas/témoins	OR [IC95%]	cas/témoins	OR [IC95%]		
valine	0,9697, 1,021	46/62	1,00	56/52	1,54 [0,86-2,74]	60/48	1,91 [1,05-3,48]	0,03	0,3
Undefined region	1,6467	45/63	1,00	61/47	2,3 [1,22-4,34]	56/52	1,99 [1,04-3,82]	0,04	0,3
glutamine	2,057	49/59	1,00	48/60	0,71 [0,38-1,31]	65/43	2,18 [1,17-4,04]	0,02	0,3
Undefined region	2,374	63/45	1,00	53/55	0,63 [0,36-1,11]	46/62	0,46 [0,25-0,88]	0,02	0,3
glutamine	2,4137	44/64	1,00	59/49	1,96 [1,04-3,71]	59/49	2,09 [1,12-3,9]	0,02	0,3
Unassigned	2,8467	48/60	1,00	53/55	1,38 [0,78-2,45]	61/47	2,02 [1,05-3,88]	0,03	0,3
albumin lysil	2,8867	43/65	1,00	58/50	2,59 [1,37-4,89]	61/47	2,82 [1,45-5,49]	0,002	0,2
Unassigned	2,9267	46/62	1,00	57/51	2,16 [1,14-4,11]	59/49	2,39 [1,26-4,52]	0,006	0,2
Unassigned	3,0706	47/61	1,00	57/51	2,04 [1,09-3,82]	58/50	1,98 [1,05-3,74]	0,03	0,3
Unassigned	3,1296	46/62	1,00	57/51	1,88 [0,99-3,59]	59/49	2,09 [1,12-3,89]	0,02	0,3
Undefined region	4,3815	47/61	1,00	53/55	1,72 [0,91-3,24]	62/46	2,33 [1,23-4,43]	0,01	0,2
urea	5,8354	58/50	1,00	56/52	0,79 [0,43-1,43]	48/60	0,49 [0,26-0,92]	0,03	0,3
Undefined region	6,0704	58/50	1,00	59/49	1 [0,54-1,85]	45/63	0,42 [0,22-0,82]	0,01	0,2
Undefined region	6,5264	51/57	1,00	50/58	1,02 [0,55-1,9]	61/47	1,89 [1,02-3,52]	0,05	0,4
Undefined region	6,7664	43/65	1,00	66/42	2,41 [1,32-4,41]	53/55	1,86 [0,97-3,57]	0,05	0,4
Undefined region	6,9664	44/64	1,00	61/47	2,03 [1,07-3,86]	57/51	2,41 [1,25-4,62]	0,009	0,2
tyrosine	7,1503, 7,1833	45/63	1,00	60/48	2,14 [1,17-3,93]	57/51	2,37 [1,25-4,49]	0,008	0,2
phenylalanine	7,2763	47/61	1,00	56/52	1,64 [0,87-3,08]	59/49	2,57 [1,3-5,08]	0,007	0,2
histidine	7,6863	43/65	1,00	61/47	2,03 [1,12-3,68]	58/50	2,62 [1,34-5,12]	0,004	0,2

Métabolites mis en évidence par la séquence CPMG (N=159 cas/159 témoins)									
Assignations	Déplacement chimique (ppm)	T1 (ref)		T2		T3		P	P-corrigée
		cas/témoins	OR	cas/témoins	OR [IC95%]	cas/témoins	OR [IC95%]		
valine	1,021	44/62	1,00	56/50	1,69 [0,95-2,99]	59/47	1,73 [0,99-3,02]	0,047	0,4
CH2 lipoproteins	1,3511	60/46	1,00	53/53	0,63 [0,34-1,16]	46/60	0,53 [0,29-0,97]	0,04	0,4
Undefined region	1,8211	49/57	1,00	52/54	1,51 [0,78-2,91]	58/48	1,88 [1,02-3,47]	0,04	0,4
Undefined region	2,1511	58/48	1,00	55/51	0,83 [0,45-1,53]	46/60	0,52 [0,28-0,96]	0,04	0,4
Undefined region	2,3111	57/49	1,00	59/47	1,26 [0,67-2,37]	43/63	0,53 [0,28-1]	0,04	0,4
citrate	2,5011	62/44	1,00	53/53	0,68 [0,38-1,22]	44/62	0,41 [0,21-0,8]	0,008	0,4
albumine	2,873	48/58	1,00	49/57	1,2 [0,63-2,29]	62/44	2,4 [1,18-4,86]	0,01	0,4
Undefined region	6,4468	47/59	1,00	52/54	1,46 [0,8-2,66]	60/46	1,81 [1,01-3,26]	0,049	0,4
Undefined region	6,5268	45/61	1,00	55/51	1,57 [0,89-2,77]	59/47	2,07 [1,06-4,03]	0,03	0,4
Undefined region	6,9428	45/61	1,00	51/55	1,43 [0,76-2,71]	63/43	2,22 [1,19-4,16]	0,01	0,4
Undefined region	6,968	47/59	1,00	52/54	1,19 [0,67-2,14]	60/46	1,91 [1,01-3,6]	0,047	0,4
Undefined region	7,2107	44/62	1,00	49/57	1,26 [0,7-2,24]	66/40	3,01 [1,57-5,79]	0,001	0,2
Undefined region	7,2697	44/62	1,00	53/53	1,4 [0,78-2,53]	62/44	2,65 [1,37-5,14]	0,004	0,4
phenylalanine	7,2787	47/59	1,00	49/57	0,99 [0,55-1,76]	63/43	1,95 [1,08-3,52]	0,03	0,4
histidine	7,6867	47/59	1,00	48/58	0,84 [0,44-1,59]	64/42	2,08 [1,11-3,9]	0,02	0,4
Undefined region	7,8307	58/48	1,00	57/49	0,98 [0,53-1,84]	44/62	0,53 [0,28-0,99]	0,04	0,4
Undefined region	7,9107	63/43	1,00	51/55	0,59 [0,33-1,06]	45/61	0,47 [0,25-0,88]	0,02	0,4
Undefined region	7,9207	61/45	1,00	55/51	0,82 [0,46-1,47]	43/63	0,49 [0,27-0,91]	0,03	0,4



## ANNEXE 5

---

Undefined region	7,9837	65/41	1,00	50/56	0,65 [0,37-1,12]	44/62	0,47 [0,26-0,84]	0,01	0,4
------------------	--------	-------	------	-------	------------------	-------	------------------	------	-----

---

OR, Odds Ratio ; CI, Confidence Interval; T, Tertile

OR, Odds Ratio ; IC, intervalle de confiance ; T, tertile

<sup>a</sup> Modèles de régression logistique conditionnelle sur les signaux RMN en continu divisé par l'écart type ajustés sur l'âge, l'IMC, le statut tabagique, la taille, la consommation d'alcool intake, l'activité physique, le niveau d'éducation, les antécédents familiaux de cancer de la prostate. Des tests de tendance linéaire ont été réalisés.

Annexe 6 Texte intégral: “Untargeted plasma metabolomic profiles associated with overall diet in women from the SU.VI.MAX cohort.”

### **Untargeted plasma metabolomic profiles associated with overall diet in women from the SU.VI.MAX cohort**

Lucie Lécuyer\*<sup>1</sup>, Céline Dalle\*<sup>2</sup>, Pierre Micheau<sup>2</sup>, Mélanie Pétéra<sup>3</sup>, Delphine Centeno<sup>3</sup>, Bernard Lyan<sup>3</sup>, Marie Lagree<sup>4</sup>, Pilar Galan<sup>1</sup>, Serge Hercberg<sup>1,5</sup>, Adrien Rossary<sup>6</sup>, Aicha Demidem<sup>6</sup>, Marie-Paule Vasson<sup>6,7</sup>, Valentin Partula<sup>1</sup>, Mélanie Deschasaux<sup>1</sup>, Bernard Srour<sup>1</sup>, Paule Latino-Martel<sup>1</sup>, Nathalie Druesne-Pecollo<sup>1</sup>, Emmanuelle Kesse-Guyot<sup>1</sup>, Stéphanie Durand<sup>3</sup>, Estelle Pujos-Guillot<sup>3</sup>, Claudine Manach<sup>#2</sup>, Mathilde Touvier<sup>#1</sup>

\* and # : equal contributions

#### **Authors' Affiliations:**

<sup>1</sup> Sorbonne Paris Cité Epidemiology and Statistics Research Center (CRESS), French National Institute of Health and Medical Research (Inserm) U1153, French National Institute for Agricultural Research (Inra) U1125, French National Conservatory of Arts and Crafts (Cnam), Paris 13 University, Nutritional Epidemiology Research Team (EREN), Bobigny, France. (LL, PG, SH, VP, MD, BS, PLM, NDP, EKG, MT)

<sup>2</sup> Université Clermont Auvergne, INRA, UMR 1019, Human Nutrition Unit, F-63000 Clermont-Ferrand, France. (CD, PM, CM)

<sup>3</sup> Clermont Auvergne University, INRA, UNH, Plateforme d'Exploration du Métabolisme, MetaboHUB Clermont, F-63000 Clermont-Ferrand, France. (MP, DC, BL, SD, EPG)

<sup>4</sup> Clermont Auvergne University, Institut de Chimie de Clermont-Ferrand, Plateforme d'Exploration du Métabolisme, MetaboHUB-Clermont, BP 80026, F-63171 Aubière, France. (ML)

<sup>5</sup> Public Health Department, Avicenne Hospital, Bobigny, France. (SH)

<sup>6</sup> Clermont Auvergne University, INRA, UMR 1019, Human Nutrition Unit (UNH), CRNH Auvergne, Cellular Micro-Environment, Immunomodulation and Nutrition (ECREIN), F-63000 Clermont-Ferrand, France. (AR, AD, MPV)

<sup>7</sup> Anticancer Center Jean-Perrin, CHU Clermont-Ferrand, France. (MPV)

#### **Corresponding author contact information:**

Lucie Lécuyer

EREN, Inserm U1153, SMBH Paris 13, 74 rue Marcel Cachin, F-93017 Bobigny Cedex, France

Telephone number: +33 1 48 38 89 54

Fax number: +33 1 48 38 89 31

e-mail: l.lecuyer@eren.smbh.univ-paris13.fr

**Running title:** Metabolomics and overall diet

**Number of figures:** 2 / **Number of table:** 4

#### **Funding**

This work was supported by the French National Cancer Institute [grant number INCa\_8085 for the project, and PhD grant number INCa\_11323 for L. Lecuyer]; the Federative Institute for Biomedical Research IFRB Paris 13; and the Cancéropôle Ile-de-France/Région Ile de France [PhD grant for M. Deschasaux]. The funders had no role in the design, analysis, or writing of this article.

**Abbreviations:**

SU.VI.MAX, SUPplémentation en Vitamines et Minéraux AntioXydants; mPNNS-GS; modified Programme National Nutrition Santé- Guideline Score; UPLC, Ultrahigh Performance Liquid Chromatography; Lysophosphatidylcholine, LysoPC; ; FFQ, Food Frequency Questionnaire; WHO, World Health Organization; NMR, Nuclear Magnetic Resonance; LC-MS, Liquid Chromatography Mass Spectrometry; PCA, Principal Component Analysis; CVD, CardioVascular disease; BMI, Body Mass Index; HCA, Hierarchical Clustering Analysis; OR, Odds Ratios; SD, Standard Deviation; CI, Confidence Interval; FDR, False Discovery Rate; BH, Benjamini-Hochberg; PA, Pipecolic acid; IPA, 3-indolepropionic acid; 2DG, 2-deoxy-D-glucose; T2D, Type 2 Diabetes; HEI, Healthy Eating Index; aMED, alternate Mediterranean Diet Score; HDI, Healthy Diet Indicator; BSD, Baltic Sea Diet

**Trial registration:** SU.VI.MAX, clinicaltrials.gov NCT00272428. Registered 3 January 2006, <https://clinicaltrials.gov/show/NCT00272428>

## ANNEXE 6

**Abstract**

**Background:** Dietary intakes are reflected in plasma by the presence of hundreds of exogenous metabolites and variations in endogenous metabolites. The exploration of diet-related plasma metabolic profiles could help to better understand the impact of overall diet on health.

**Objective:** Our aim was to identify metabolomic signatures reflecting overall diet in women from the French general population.

**Design:** This cross-sectional study included 160 women in the SU.VI.MAX cohort with detailed dietary data ( $\geq 10$  24h-dietary records) selected according to their level of adherence to the French dietary recommendations, represented by the validated score mPNNS-GS; 80 women from the 10<sup>th</sup> decile of the score were matched with 80 women from the 1<sup>st</sup> decile. Plasma metabolomic profiles were acquired using untargeted UPLC-QToF mass spectrometry analysis. The associations between metabolomic profiles and the mPNNS-GS, its components and Principal Component Analyses-derived dietary patterns were investigated using multivariable conditional logistic regression models and partial correlations.

**Results:** Adherence to the dietary recommendations was positively associated with 3-indolepropionic acid and piperolic acid (also positively associated with fruit and vegetable intake and a healthy diet) – 2 metabolites linked to microbiota; and inversely associated with lysophosphatidylcholine (LysoPC(17:1)), acylcarnitine C9:1 (also inversely associated with a healthy diet), acylcarnitine C11:1 and 2-deoxy-D-glucose. Increased plasma levels of piperine and Dihydro4mercapto-3(2H) furanone were observed in women who consumed a Western diet and a healthy diet respectively. Ethyl- $\beta$ -D-glucopyranoside was positively associated with alcohol intake. Plasma levels of LysoPC(17:1), cholic acid, phenylalanine-phenylalanine and phenylalanine and carnitine C9:1 decreased with the consumption of vegetable added fat, sweetened food, milk and dairy products and fruit and vegetable intakes respectively.

**Conclusion:** This study highlighted several metabolites from both host and microbial metabolism reflecting the long-term impact of the overall diet.

**Keywords:** Metabolomics, overall diet, dietary patterns, mass spectrometry, 24h-dietary records

## INTRODUCTION

Accumulating evidence links diet to chronic disease risk. Today, nutrition-related behaviours are recognized as key modifiable determinants of health. However, important challenges remain to better elucidate the relationships between nutrition and health and underlying mechanisms. Randomized controlled trials and observational epidemiological studies based on repeated 24h-dietary records as well as food frequency questionnaires (FFQ) allowed major discoveries and will continue to increase knowledge in this field. Nevertheless, to complete self-reported data from questionnaire, biomarkers directly reflecting dietary intakes or nutritional status have been proven useful. Nutritional metabolomics allows to obtain metabolic signatures reflecting dietary exposures, including non-nutrients and contaminants, and their potential interactions with both host and microbial metabolisms (1). These metabolic signatures would provide objective elements for a better understanding of nutrition and health interactions. So far, many studies highlighting associations between diet and metabolomic profiles focused on specific food groups, often with the objective of discovering biomarkers of food intake (2–4). Over the last decades, dietary patterns have gained increasing attention since the cumulative effects of multiple dietary components could act synergistically to impact health (5,6). Due to the holistic nature of metabolomics, it has naturally been applied to find associations with the overall diet (7). A limited number of controlled trials, from single meal to medium term dietary interventions, analysed the changes in metabolomic profiles after consumption of diets such as Mediterranean diet, New Nordic Diet, Average Danish Diet, diet enriched in whole-grain products, diet based on WHO healthy eating guidelines or vegetarian diet (8–16). The comparison between studies is not straightforward due to the variety of non-standardized NMR or LC-MS methods used, the different types of samples analysed, and the uniqueness of the dietary shifts tested. Beyond randomised controlled trials, which focused on very specific dietary interventions in experimental conditions, observational epidemiological studies also provide complementary information regarding complex real-life dietary intakes (17). Several observational studies investigated the relationships between overall diet assessed by FFQs or multiple 24h-dietary records, and blood or urine metabolomic profiles (17–31). A recent review compared the results from 8 observational studies (32) and reported that although many putative biomarkers of individual foods and dietary patterns have been identified, those biomarkers may not be sensitive or specific enough. Indeed, while some results were consistent across studies (e.g. urinary anserine associated with a “healthy” dietary pattern (30) and a high adherence to Mediterranean diet (22)), results for most metabolites were mitigated or not replicated. The inconstancy of some associations in the literature may partly be explained by the diversity of the studied populations and the different conceptual approaches used to define and characterized the dietary patterns. Some studies classified overall diet through a posteriori approaches (i.e. dietary patterns observed in the population study) and others were based on a priori approaches using dietary scores or predefined diets.

The aim of our study was to determine the associations between untargeted plasma metabolomic profiles assessed by a UPLC-QToF sensitive method and overall diet in women from the French general population – assessed using a detailed dietary assessment method. This study compared two complementary approaches for the overall diet modelling: a priori through the adherence to the 2001 French national dietary guidelines score and a posteriori through Principal Component Analyses (PCA)-derived dietary patterns.

## SUBJECTS AND METHODS

### Population

The study population was derived from the French prospective “SUpplémentation en Vitamines et Minéraux AntioXydants” (SU.VI.MAX) cohort. Initially the SU.VI.MAX study was a randomised, double-blind, placebo-controlled, primary prevention trial (Trial Registration number in the clinicaltrials.gov database: NCT00272428) enrolling a total of 13,017 participants (33,34). In brief, a national recruitment campaign

## ANNEXE 6

started in 1994, lasted for 8 years and was designed to evaluate the effect of a supplementation in antioxidant vitamins and minerals at nutritional doses on the incidence of CVD and cancer. Additional observational follow-up went on until 2007. The study was conducted according to the guidelines laid down in the Declaration of Helsinki and approved by the Ethics Committee for Studies with Human Subjects of Paris-Cochin Hospital (CCPPRB 706/2364) and the 'Commission Nationale de l'Informatique et des Libertés' (CNIL 334641/907094). All subjects gave written informed consent to participate in the study.

**Baseline data collection**

Data on sociodemographic characteristics, smoking status, medication use, health status were collected at study inclusion, using self-administered questionnaires. Anthropometric measurements as well as a blood draw (occurring after a 12h-fasting period) were performed by study nurses and physicians during a baseline clinical examination. More precisely, 35-ml venous blood samples were collected in sodium heparin Vacutainer tubes (Becton Dickinson, Rungis, France). Plasma aliquots were prepared immediately after centrifugation at +4°C, and subsequently stored frozen at -20°C for less than two days before being transferred into liquid nitrogen.

**Dietary data assessment**

During the trial phase, all participants were invited to complete a 24-h dietary record every 2 months, covering all days of the week and all seasons of the year, using computerised questionnaires. To indicate the portion size consumed, participants were assisted by a validated instruction manual with photographs of >250 generic food items corresponding to 990 specific foods (35). A French published food composition table was used to calculate nutrient intakes (36). Dietary habits were estimated by averaging intakes from all dietary records collected during the first 2 years of participation in the SU.VI.MAX study. Thus, intra-individual variability of intake was accounted for. The foods were categorized into 29 *a priori*-defined food groups (see **Supplemental file 1**).

**Selection of the study sample**

The selection of the study population was made according to the mPNNS-GS (modified Programme National Nutrition Santé-Guideline Score) reflecting adherence to the French national dietary recommendations (details on score computation based on 12 components are available in **Supplemental file 2**). This score has been previously validated (37) and associated with lower risk of chronic diseases (38–41). The present cross-sectional study initially included women having at least 10 validated 24h-dietary records during the 2 years following blood draw and randomly selected among all eligible women participants of the cohort (see the flow-chart **Figure 1**): 80 women within the 1<sup>st</sup> decile (indicating a lower quality of diet) and 80 women within the 10<sup>th</sup> decile (indicating a better quality of diet) of the score. The two groups were matched for age (35-39y/40-44y/45-49y/50-54y/55-59y/>60y), menopausal status (pre/post-menopause), BMI (<18.5kg/m<sup>2</sup> : underweight, ≥18.5 kg/m<sup>2</sup> to <25 kg/m<sup>2</sup> : normal weight and ≥25 kg/m<sup>2</sup> : overweight/obese), smoking status (current smokers and non/former smokers) and season (a priori-defined periods: October-November/December-January-February/March-April) at time of blood draw.

**Metabolomic analyses**

Two complementary plasma extraction protocols were used for untargeted metabolomics: one developed for endogenous metabolites following a slightly modified procedure described in Pereira et al. (42), and the other developed for exogenous metabolites with precipitation of proteins using acidified methanol (**Supplemental file 3**). Plasma samples were analyzed using ultrahigh performance liquid chromatography (UPLC, 26 min gradient) coupled to a high-resolution mass spectrometer QToF Impact HDII (Bruker) equipped with an electrospray ionization source. Metabolomic profiles were acquired in positive and negative modes with a scan range from 50 to 1000 mass-to-charge ratio (m/z). The stability of the analytical system was monitored using pooled samples as quality control, analyzed in the same sequence than participant samples. The detailed analytical conditions are given in **Supplemental file 3**.

*Processing of UPLC-QToF data*

## ANNEXE 6

Data were processed under the Galaxy web-based platform Workflow4Metabolomics (43), using first XCMS module for peak detection, grouping and alignment, to yield a data matrix containing variables (retention times, masses) and peak intensities that have been corrected for batch effects. The whole procedure, detailed in **Supplemental file 3**, was applied independently to the four analyses, i.e. endogenous or exogenous protocol with positive and negative mode. Four samples (exogenous protocol, positive ionization mode) were excluded due to abnormal profiles revealed by processing controls.

*Metabolites identification*

Ions of interest were identified using the multi-step procedure described in **Supplemental file 3**. Briefly, fragmentation spectra were obtained with the QToF and an ultra-high resolution LTQ Orbitrap Velos hybrid mass spectrometer (Thermo Fisher Scientific, San José, CA, USA) and compared to in-house spectral libraries, online databases, data from the literature, and in silico predicted fragmentation pathways generated with MassFrontier (Thermo-Scientific). Levels of confidence for every identification were given accordingly to Sumner et al. (44): level 1; formally identified compound (confirmed with analysis of authentic standard), level 2; putatively identified compound (based upon spectral similarity with public/commercial spectral libraries or reference compound in the literature and/or physicochemical properties), level 3; putatively characterized compound classes and level 4; unknown compound.

**Statistical analyses**

Participants' baseline characteristics and scores of the mPNNS-GS components were compared between the two groups (i.e. 1<sup>st</sup> and 10<sup>th</sup> deciles of mPNNS-GS) using conditional logistic regression.

*A priori PNNS-GS*

In order to characterize the two mPNNS-GS compared groups in terms of dietary intakes, a Hierarchical Clustering Analysis (HCA) of the mPNNS-GS components for the 160 participants was built using Euclidean distance and Ward aggregation method as clustering algorithm.

We estimated the association between the signal intensity of each ion and mPNNS-GS status (1<sup>st</sup> decile *versus* 10<sup>th</sup> decile) using multivariable conditional logistic regression models. Odds ratios (OR) for an increment of 1 SD for each metabolite and their 95% confidence intervals (95% CI) were computed as well as tests for linear trend. Multivariable models were adjusted for the following baseline characteristics: age, BMI, smoking status, number of 24h-dietary records and mean of daily energy intake during the 2 years following the blood draw. In order to avoid residual confounding, models were adjusted for matching factors only when the variable was more precise than the one used for the matching (e.g. 6 classes of age for matching factor: 35-39y/40-44y/45-49y/50-54y/55-59y/>60y and age in continuous for adjustment). Models were not additionally adjusted for menopausal status and season of blood draw because the compared groups were already matched for those factors.

We estimated correlations between the ion intensities and the intake of the food groups (continuous variables) corresponding to the 12 dietary components of mPNNS-GS using partial Spearman correlation analyses, adjusted for the following variables measured at time of blood draw: age, BMI, menopausal status and smoking status, season, number of 24h-dietary records and mean of daily energy intake during the 2 years following blood draw.

*A posteriori dietary patterns*

Dietary patterns were obtained using a principal component analysis (PCA) with varimax rotated factors by orthogonal transformation on the 29 food groups on the general SU.VI.MAX population having at least 10 validated 24h-dietary records during the 2 years following baseline (N=4645). The three first factors (="patterns") were retained according to the plot of the total variance explained (Cattel's scree plots) (45). For interpreting the data, we considered food groups with a factor loading  $\geq 0.25$ . Then, correlations between the ions and the three dietary patterns obtained from the PCA were estimated using partial Spearman correlation, adjusted for the following variables measured at time of blood draw: age, BMI, menopausal status and smoking status, season, number of 24h-dietary records and mean of daily energy intake during the 2 years following blood draw.

## ANNEXE 6

In order to handle the False Discovery Rate (FDR) due to multiple testing, a Benjamini-Hochberg (BH) (46) correction was tested.

Partial Spearman correlations were also estimated between identified ions and nutrient intakes and clinical and biological parameters measured at baseline (details in **Supplemental file 4**).

The analyses were performed using SAS software (v9.3, Cary, NC; conditional logistic regressions and correlations) and the web-based pipeline MetaboAnalyst 3.0 (HCA) (47). All statistical tests were 2-sided.

## RESULTS

The baseline characteristics of the participants are summarized in **Table 1**. Women from the 1<sup>st</sup> mPNNS-GS decile group (low adherence to the French dietary recommendations) tended to have a higher daily energy intake than women from the 10<sup>th</sup> decile (high adherence).

After careful pre-processing of the metabolomics data, 1344 and 767 ions were obtained for samples extracted with the exogenous and endogenous protocols respectively (for both modes of ionisation) and used for all statistical analyses.

### *A priori PNNS-GS*

The HCA heatmap (**Figure 2**) of dietary intakes relative to the 12 components of the mPNNS-GS for the 160 included subjects illustrates the relatively high heterogeneity of consumptions existing between and within each of the two mPNNS-GS groups. Participants from the 10<sup>th</sup> decile group had a mean mPNNS-GS of  $10.1 \pm 0.7$  but they could have reached this score with different combinations of components. Some extreme cases exist, such as the three participants from the 10<sup>th</sup> decile group, misclassified in the HCA, who had a markedly higher consumption of sweetened beverages and total added fat, but were classified in the 10<sup>th</sup> decile group due to good scores in the other components. In addition, for some components extreme consumption were well marked between groups, but a large number of subjects have the same level of medium consumption in both groups. For example, in the 1<sup>st</sup> decile group 48.8% of the participants consumed less than 5 portions/day of fruits and vegetables vs 6.3% in the 10<sup>th</sup> decile group, and only 10% vs 51.3% consumed more than 7.5 portions/day. However, a large proportion of participants in both groups (41.3-42.5%) consumed between 5 and 7.5 portions/d. **Table 1** provides the baseline scores in the two groups for the 12 dietary components of the mPNNS-GS. In the study population, all the components except for “bread, cereals, potatoes and legumes” and “sweetened beverages” components (i.e. “fruits and vegetables”, “whole grain food”, “milk and dairy products”, “meat, poultry, seafood and eggs”, “seafood”, “total added fat”, “vegetable added fat”, “sweetened food”, “alcohol”, “salt”) discriminated the 10<sup>th</sup> and 1<sup>st</sup> decile groups.

For clarity, the 15 strongest associations between ions and mPNNS-GS status from conditional logistic regression are presented in **Table 2** and the extended list of associations at a  $p < 0.05$  level of statistical significance (N=102) are presented in **Supplemental file 5**. None achieved a FDR  $< 0.05$  level. Pipecolic acid (PA, HMDB0000070) and 3-indolepropionic acid (IPA, HMDB0002302) had increased plasma levels in women with higher adherence to the dietary recommendations while lysophosphatidylcholine (LysoPC(17:1), LMGP01050002), 2,6-dimethylheptanoyl carnitine (HMDB0006320) / nonanoyl-l-carnitine (HMDB0013288) (acylcarnitine C9:1), 4,8-dimethylnonanoyl carnitine (HMDB0006202) / undecanoylcarnitine (HMDB0013321) (acylcarnitine C11:1) and 2-deoxy-D-glucose (2DG, HMDB0062477) had increased plasma levels in women with lower adherence to the dietary recommendations. We further investigated the correlations between metabolomic profiles and the intakes of the 12 food groups of the dietary components of mPNNS-GS (correlations at a  $p < 0.05$  level of statistical significance are shown in **Supplemental file 6**, N=1522). The strongest correlations are presented in **Table 3**. Positive associations between alcohol intake and ethyl- $\beta$ -D-glucopyranoside (HMDB0029968) and between seafood intake and the ions M157.0118T23.44 and M810.5654T24.77 as well as negative associations between seafood intake and the ions M136.8774T0.94, M318.1622T15.23 and



## ANNEXE 6

M241.1031T0.89, achieved a FDR level  $<0.05$  (see **Supplemental file 6**). Positive correlations were found for PA and fruit and vegetable intake. Negative correlations were found for phenylalanine (HMDB0000159), acylcarnitine C9:1 and fruit and vegetable intake, phenylalanyl-phenylalanine (HMDB0013302) and milk and dairy products intakes, cholic acid (HMDB00619) and sweetened food intake, and for LysoPC(17:1) and vegetable added fat intake.

#### *A posteriori dietary patterns*

The first three PCA factors (dietary patterns) derived from the general SU.VI.MAX population explained  $>23\%$  of total variability in dietary data. Loading values are displayed in **Supplemental file 1**. The first factor was characterized by higher intakes of alcoholic drinks, animal fat, bread, cheese, pasta, rice and cereals, potatoes and tubers, processed meat, legumes, red meat and offal, salted and sweetened products, sauces, vegetable fat and white meat – and therefore considered as a “Western diet” pattern. The second factor was characterized by higher intakes of fruit, vegetable, dried fruit, fish and seafood, breakfast cereals, vegetable fat, whole grain and yoghurt – and therefore considered as a “Healthy diet” pattern. The third factor was characterized by higher intake of cakes and biscuits, dried fruits, fruit juice, animal fat, milk, nuts and sweetened products. Therefore, this pattern was considered as a “Sweetened diet” pattern. Associations between ions and dietary patterns from Spearman correlation analyses at a  $p<0.05$  level of statistical significance are shown in **Supplemental file 7** (N=336) and the strongest correlations are presented in **Table 4**. Positive association between pipecolic acid and a Healthy diet as well as negative association between PA and a Western diet achieved a FDR level  $<0.05$  (see **Supplemental file 7**). Positive correlations were also observed for piperine (HMDB0029377) and a Western diet and for dihydro4mercapto-3(2H) furanone (HMDB39786) and a Healthy diet; and negative correlation was found for acylcarnitine C9:1 and a Healthy diet.

## DISCUSSION

The present study revealed new associations between plasma metabolomic profiles and overall diet characterized by complementary methods in a prospective cohort with detailed information on food intake over 13 years.

Using either a score reflecting adherence to the French national dietary recommendations or PCA-derived dietary patterns, pipecolic acid (PA) was associated with a healthy overall diet. The origin of plasma PA is uncertain. Dry bean and cruciferous vegetable represent the richest dietary source reported so far for PA (43 and 12-22 mg/kg, respectively (48)). It was identified using metabolomics as the best candidate plasma biomarker of dry bean intake in a 4-week intervention study with 250g/day of dry bean and reflected self-reported dry bean intake over a range of 0-300 g/day in a population of 212 subjects (49). However, dry bean consumption is unlikely to explain the observed differences in PA in our study as reported intake of legumes was not correlated with plasma level of PA ( $r<|0.05|$ ) and did not differ between the 1st and 10th PNNS-GS groups (data not shown). PA may also originate from degradation of dietary lysine by the gut microbiome (48). In four studies applying an exploratory metabolomic approach (13,16,50,51), plasma concentration of PA was significantly increased in subjects who consumed a whole grain-enriched “Healthy diet” for 12 weeks (13), a low-glycemic index diet for 4 weeks (16), who reported higher daily physical activity (50) or with higher muscle mass (51). In our study, PA was positively associated with the “Fruit and vegetable” mPNNS-GS component as well as with intakes of total and soluble fibres, potassium, simple carbohydrates, vitamins C, E, B9, beta-carotene and other micronutrients (see **Supplemental file 4**). Indolepropionic acid (IPA) was the other identified metabolite that increased with higher adherence to dietary recommendations. IPA is a gut microbial metabolite of dietary tryptophan (52). Higher serum IPA was associated with lower risk of developing Type 2 Diabetes (T2D) in subjects with impaired glucose tolerance from the Finnish Diabetes Prevention Study (53), with replication in two independent cohorts

## ANNEXE 6

(53,54). Jennis et al. also reported lower blood concentrations of IPA in obese T2D subjects compared to lean subjects or to post-bariatric surgery levels (55). Whether IPA has a direct preventive role towards diabetes is unclear. It was linked to better insulin secretion and lower low-grade inflammation (53). In animal models, diet supplementation with IPA significantly lowered fasting blood glucose (56) and reduced the increased intestinal permeability induced by a high fat diet (55). Regarding dietary associations, serum IPA was correlated to intake of dietary fibres and whole grains in Nordic populations (53,54), of banana, apple and pear in a UK female population (57), as well as with healthy Baltic Sea Diet and dietary fibres in male Finnish smokers of the ATBC study (23). It was inversely correlated with eggs and red meat intake in a US population (4). In our study, plasma IPA was correlated with plasma beta-carotene at baseline (see **Supplemental file 4**), which probably reflected a positive association with fruit and vegetable intake (58). We hypothesize that the increased plasma levels of IPA and PA in SU.VI.MAX women with a higher adherence to the dietary recommendations reflected a favourable modification of the gut microbiome by the diet, possibly linked to fruit and vegetable intake, dietary fibres or other plant food components with prebiotic activity such as polyphenols (59). IPA and PA deserve further validation as candidate biomarkers of healthy diet and/or healthy lifestyle.

Some lipids metabolites were associated to a lower adherence to dietary recommendations. LysoPC(17:1) is a lyso-phosphatidylcholine with an odd number of carbons atoms, not synthesized by the human body. Choline-containing phospholipids were previously linked to high butter and high-fat dairy intakes (20,57), which seems consistent with the LysoPC(17:1) association with a low mPNNS-GS score and the correlation with intakes of saturated fatty acids in our study (see **Supplemental File 4**). Two acylcarnitines were also associated with a low mPNNS-GS score. Acylcarnitines are a transport form of fatty acids involved in energy production (60). Their plasma concentration is partly determined by dietary intakes. C9:1 and C11:1 acylcarnitines are medium-chain species that are exclusively derived from the fatty acid metabolism. Blood level of C9:1 was recently linked with butter intake and associated with a gene involved in lipid oxidation (20,57), again consistent with our findings. The interpretation of the negative association of plasma cholic acid, a major primary bile acid, with the intake of sweetened food is less straightforward and complementary data would be needed to propose a clear explanation. Variations of endogenous metabolites are sometimes difficult to interpret when they involved in several metabolic pathways, possibly modulated by diverse factors, and when they combine exogenous and endogenous origins.

In all untargeted metabolomics studies, including ours, a high proportion of metabolites remain unidentified. Furthermore, the identification process often reveal chemical structures for which there is only scarce information available on possible origins and implication for nutrition and health. This is the case for 2-deoxy-D-glucose, a non-metabolizable glucose analogue, and dihydro4mercapto-3(2H) furanone, for which we could not find sufficient relevant information to elaborate on their presence among the metabolites correlated with the overall diet. A major effort for further enriching databases is critical to take full advantage of exploratory metabolomics.

A large range of candidate biomarkers of food intake have recently been identified with metabolomics and are currently under validation for robustness across populations (2,61). Although many of them were detectable in our analytical conditions we did not find food intake biomarkers among the compounds discriminating the overall diet in our study. This contrasts with several former studies that reported food-specific metabolites, e.g. of from fish, whole-grain products or berries associated with dietary patterns (8,13) or multivariate models based on food-derived metabolites able to classify subjects into dietary patterns (8,10,11). However, these studies were dietary interventions in which subjects were assigned to the consumption of specific dietary patterns in a controlled environment (food was generally provided) with monitoring of compliance. In such design, although the dietary patterns can be complex, they are rather homogenous within experimental groups. In our study where subjects are in free-living conditions, intra-

## ANNEXE 6

group variability is not minimized when the inter-group variability is maximized. Our dietary data clearly showed the high heterogeneity and variety of food choices that exist within the 1st and 10th mPNNS-GS decile groups, and the absence of clear boundaries between dietary patterns. The use of combinations of food intake biomarkers to stratify subjects into predefined dietary patterns in observational studies may thus be challenging. Accordingly, our findings suggest that metabolomic signatures of dietary patterns in observational studies may rather comprise endogenous and gut microbial metabolites reflecting the long-term impact of the diet on metabolic pathways and on the gut microbiome. To our knowledge, only two other studies linked untargeted LC-MS metabolomic profiles to dietary recommendation or dietary patterns in a cross-sectional design. In the ATBC cohort (n=1336), Playdon et al. found 23, 46, 23 and 33 metabolites correlated ( $-0.30 < r < 0.20$ ) with the four diet quality indexes HEI2010, aMED, HDI and BSD respectively, with little overlap between correlated metabolites (23). Indolepropionic acid was found associated with the Baltic Sea Diet index (23). In a case-control study (n=421 pairs) nested in a Swedish cohort, Shi et al. identified 31 plasma metabolites associated ( $-0.19 < r < 0.21$ ) with two a priori defined healthy Nordic dietary indexes (BSDS and HNFI) (62). Only 6 were associated with both indexes, among which DHA was also previously found associated with two healthy dietary pattern indexes in the study by Playdon et al (23). More metabolomic studies are needed in populations with diverse dietary backgrounds to analyze the consistency of metabolites found correlated with dietary patterns. Comparison between studies is challenged by the various scoring systems used for dietary patterns, by the variety of analytical method applied, and by the type of biospecimen collected, as urine and blood differently reflect dietary habits. In addition, as many synonyms exist for each compound, it is essential that studies describe metabolites in an unambiguous way, for example using an identifier in a database.

Strengths of this study mainly pertained to the untargeted metabolomic approach and the detailed data of overall dietary intakes (at least ten 24h-dietary records over a 2-year period) representing real-life conditions. However, some major limitations should be acknowledged. First, the number of participants, and thus the statistical power, may have limited our ability to detect some associations and may explain the relatively large confidence intervals in logistic regression models. Second, this metabolomic analysis was based on a single blood draw; several blood draws would have allowed investigating the stability of the profiles. Nevertheless, several studies showed a good stability and reproducibility of metabolomic measurements for most of endogenous metabolites (63,64). Third, we cannot exclude the possibilities of residual or unmeasured confounding or overfitting of our models. Finally, due to the untargeted approach, several metabolites of interest could not be identified.

Defining a global and universal metabolomic signature that could reflect a “healthy” or “western” dietary pattern for use in large cohort studies, with the objective to complement dietary questionnaire for stratification of subjects is challenging. Our study and the available literature showed that some combinations of endogenous and microbial metabolites might reflect the overall diet though its long-term beneficial or detrimental effects on the nutritional and metabolic status of individuals. Additional observational studies and long term controlled intervention studies highly representative of real-life conditions are still needed to consolidate and validate the metabolomics findings.

### Acknowledgments

The authors thank Younes Esseddik, Frédéric Coffinieres, Thi Hong Van Duong, Paul Flanzy, Régis Gatibelza, Jagatjit Mohinder and Maithily Sivapalan (computer scientists), Rachida Mehroug and Frédérique Ferrat (logistic assistants), Nathalie Arnault, Véronique Gourlet, PhD, Fabien Szabo, PhD, Julien Allegre, and Laurent Bourhis (data-manager/statisticians) and Cédric Agaesse (dietitian) for their technical contribution to the SU.VI.MAX study. We also thank all participants of the SU.VI.MAX study. This work was conducted in the framework of the French network for Nutrition And Cancer Research (NACRe

## ANNEXE 6

network), [www.inra.fr/nacre](http://www.inra.fr/nacre) and received the NACRe Partnership Label. . Metabolomic analyses were performed within the metaboHUB French infrastructure (ANR-INBS-0010).

**Conflicts of interest:** The authors have no conflict of interest to disclose.

**Authors' contributions**

The author's responsibilities were as follow CM and MT: designed the research and supervised data interpretation; LL and CD conducted the research; SH, PG, MT, EKG: conducted the SU.VI.MAX cohort and provided essential data and samples; DC, ML, SD, BL and EPG: performed metabolomic analyses; LL, CD and PM: performed statistical analysis; MT, CM, MP: supervised statistical analysis; LL, MT, CD, CM and EPG: wrote the paper; all authors provided critical intellectual input for data interpretation, read and approved the final manuscript; CM and MT had primary responsibility for final content.

VERSION SOUMISE

## REFERENCES

1. Jones DP, Park Y, Ziegler TR. Nutritional metabolomics: progress in addressing complexity in diet and health. *Annu Rev Nutr.* 2012;32:183–202.
2. González-Peña D, Brennan L. Recent Advances in the Application of Metabolomics for Nutrition and Health. *Annu Rev Food Sci Technol.* 2019;10:479–519.
3. Playdon MC, Sampson JN, Cross AJ, Sinha R, Guertin KA, Moy KA, Rothman N, Irwin ML, Mayne ST, Stolzenberg-Solomon R, et al. Comparing metabolite profiles of habitual diet in serum and urine. *Am J Clin Nutr.* 2016;104:776–89.
4. Guertin KA, Moore SC, Sampson JN, Huang WY, Xiao Q, Stolzenberg-Solomon RZ, Sinha R, Cross AJ. Metabolomics in nutritional epidemiology: identifying metabolites associated with diet and quantifying their potential to uncover diet-disease relations in populations. *AmJClinNutr.* 2014;100:208–17.
5. Cespedes EM, Hu FB. Dietary patterns: from nutritional epidemiologic analysis to national guidelines. *Am J Clin Nutr.* 2015;101:899–900.
6. World Health Organization (WHO). *Guideline: Dietary Patterns.* WHO: Geneva, Switzerland; 2017.
7. Brennan L. Metabolomics: a tool to aid dietary assessment in nutrition. *Curr Opin Food Sci.* 2017;16:96–9.
8. Garcia-Perez I, Posma JM, Gibson R, Chambers ES, Hansen TH, Vestergaard H, Hansen T, Beckmann M, Pedersen O, Elliott P, et al. Objective assessment of dietary patterns by use of metabolic phenotyping: a randomised, controlled, crossover trial. *Lancet Diabetes Endocrinol.* 2017;5:184–95.
9. Bondia-Pons I, Martinez JA, de la Iglesia R, Lopez-Legarrea P, Poutanen K, Hanhineva K, Zulet M de los Á. Effects of short- and long-term Mediterranean-based dietary treatment on plasma LC-QTOF/MS metabolic profiling of subjects with metabolic syndrome features: The Metabolic Syndrome Reduction in Navarra (RESMENA) randomized controlled trial. *Mol Nutr Food Res.* 2015;59:711–28.
10. Andersen MB, Rinnan A, Manach C, Poulsen SK, Pujos-Guillot E, Larsen TM, Astrup A, Dragsted LO. Untargeted metabolomics as a screening tool for estimating compliance to a dietary pattern. *JProteomeRes.* 2014;13:1405–18.
11. Acar E, Gürdeniz G, Khakimov B, Savorani F, Korndal SK, Larsen TM, Engelsen SB, Astrup A, Dragsted LO. Biomarkers of Individual Foods, and Separation of Diets Using Untargeted LC-MS-based Plasma Metabolomics in a Randomized Controlled Trial. *Mol Nutr Food Res.* 2019;63:1800215.
12. Vázquez-Fresno R, Llorach R, Urpi-Sarda M, Lupianez-Barbero A, Estruch R, Corella D, Fitó M, Arós F, Ruiz-Canela M, Salas-Salvadó J, et al. Metabolomic pattern analysis after mediterranean diet intervention in a nondiabetic population: a 1- and 3-year follow-up in the PREDIMED study. *J Proteome Res.* 2015;14:531–40.
13. Hanhineva K, Lankinen MA, Pedret A, Schwab U, Kolehmainen M, Paananen J, de Mello V, Sola R, Lehtonen M, Poutanen K, et al. Nontargeted Metabolite Profiling Discriminates Diet-Specific Biomarkers for Consumption of Whole Grains, Fatty Fish, and Bilberries in a Randomized Controlled Trial. *J Nutr.* 2015;145:7–17.
14. Lankinen M, Kolehmainen M, Jääskeläinen T, Paananen J, Joukamo L, Kangas AJ, Soininen P, Poutanen K, Mykkänen H, Gylling H, et al. Effects of Whole Grain, Fish and Bilberries on Serum

## ANNEXE 6

- Metabolic Profile and Lipid Transfer Protein Activities: A Randomized Trial (Sysdimet). Berthold HK, editor. PLoS ONE. 2014;9:e90352.
15. Stella C, Beckwith-Hall B, Cloarec O, Holmes E, Lindon JC, Powell J, van der Ouderaa F, Bingham S, Cross AJ, Nicholson JK. Susceptibility of human metabolic phenotypes to dietary modulation. *J Proteome Res.* 2006;5:2780–8.
  16. Esko T, Hirschhorn JN, Feldman HA, Hsu Y-HH, Deik AA, Clish CB, Ebbeling CB, Ludwig DS. Metabolomic profiles as reliable biomarkers of dietary composition. *Am J Clin Nutr.* 2017;105:547–54.
  17. Peré-Trepat E, Ross AB, Martin F-P, Rezzi S, Kochhar S, Hasselbalch AL, Kyvik KO, Sørensen TIA. Chemometric strategies to assess metabonomic imprinting of food habits in epidemiological studies. *Chemom Intell Lab Syst.* 2010;104:95–100.
  18. Altmaier E, Kastenmüller G, Römisch-Margl W, Thorand B, Weinberger KM, Illig T, Adamski J, Döring A, Suhre K. Questionnaire-based self-reported nutrition habits associate with serum metabolism as revealed by quantitative targeted metabolomics. *Eur J Epidemiol.* 2011;26:145–56.
  19. Merz B, Frommherz L, Rist MJ, Kulling SE, Bub A, Watzl B. Dietary Pattern and Plasma BCAA-Variations in Healthy Men and Women-Results from the KarMeN Study. *Nutrients.* 2018;10.
  20. Floegel A, von Ruesten A, Drogan D, Schulze MB, Prehn C, Adamski J, Pischon T, Boeing H. Variation of serum metabolites related to habitual diet: a targeted metabolomic approach in EPIC-Potsdam. *Eur J Clin Nutr.* 2013;67:1100–8.
  21. Bouchard-Mercier A, Rudkowska I, Lemieux S, Couture P, Vohl M-C. The metabolic signature associated with the Western dietary pattern: a cross-sectional study. *Nutr J.* 2013;12:158.
  22. Almanza-Aguilera E, Urpi-Sarda M, Llorach R, Vázquez-Fresno R, Garcia-Aloy M, Carmona F, Sanchez A, Madrid-Gambin F, Estruch R, Corella D, et al. Microbial metabolites are associated with a high adherence to a Mediterranean dietary pattern using a 1H-NMR-based untargeted metabolomics approach. *J Nutr Biochem.* 2017;48:36–43.
  23. Playdon MC, Moore SC, Derkach A, Reedy J, Subar AF, Sampson JN, Albanes D, Gu F, Kontto J, Lassale C, et al. Identifying biomarkers of dietary patterns by using metabolomics. *Am J Clin Nutr.* 2017;105:450–65.
  24. O’Sullivan A, Gibney MJ, Brennan L. Dietary intake patterns are reflected in metabolomic profiles: potential role in dietary assessment studies. *AmJClinNutr.* 2011;93:314–21.
  25. Xu J, Yang S, Cai S, Dong J, Li X, Chen Z. Identification of biochemical changes in lactovegetarian urine using 1H NMR spectroscopy and pattern recognition. *Anal Bioanal Chem.* 2010;396:1451–63.
  26. Schmidt JA, Rinaldi S, Ferrari P, Carayol M, Achaintre D, Scalbert A, Cross AJ, Gunter MJ, Fensom GK, Appleby PN, et al. Metabolic profiles of male meat eaters, fish eaters, vegetarians, and vegans from the EPIC-Oxford cohort. *Am J Clin Nutr.* 2015;102:1518–26.
  27. Menni C, Zhai G, Macgregor A, Prehn C, Römisch-Margl W, Suhre K, Adamski J, Cassidy A, Illig T, Spector TD, et al. Targeted metabolomics profiles are strongly correlated with nutritional patterns in women. *Metabolomics Off J Metabolomic Soc.* 2013;9:506–14.
  28. Bhupathiraju SN, Guasch-Ferré M, Gadgil MD, Newgard CB, Bain JR, Muehlbauer MJ, Ilkayeva OR, Scholtens DM, Hu FB, Kanaya AM, et al. Dietary Patterns among Asian Indians Living in the United States Have Distinct Metabolomic Profiles That Are Associated with Cardiometabolic Risk. *J Nutr*

## ANNEXE 6

- [Internet]. 2018 [cited 2019 Jan 22]; Available from: <https://academic.oup.com/jn/advance-article/doi/10.1093/jn/nxy074/5033850>
29. Wei R, Ross AB, Su M, Wang J, Guiraud S-P, Draper CF, Beaumont M, Jia W, Martin F-P. Metabotypes Related to Meat and Vegetable Intake Reflect Microbial, Lipid and Amino Acid Metabolism in Healthy People. *Mol Nutr Food Res*. 2018;62:1800583.
  30. Gibbons H, Carr E, McNulty BA, Nugent AP, Walton J, Flynn A, Gibney MJ, Brennan L. Metabolomic-based identification of clusters that reflect dietary patterns. *Mol Nutr Food Res*. 2017;61.
  31. Sedlmeier A, Kluttig A, Giegling I, Prehn C, Adamski J, Kastenmüller G, Lacruz ME. The human metabolic profile reflects macro- and micronutrient intake distinctly according to fasting time. *Sci Rep* [Internet]. 2018 [cited 2019 Jan 22];8. Available from: <http://www.nature.com/articles/s41598-018-30764-4>
  32. Guasch-Ferré M, Bhupathiraju SN, Hu FB. Use of Metabolomics in Improving Assessment of Dietary Intake. *Clin Chem*. 2018;64:82–98.
  33. Hercberg S, Galan P, Preziosi P, Bertrais S, Mennen L, Malvy D, Roussel AM, Favier A, Briancon S. The SU.VI.MAX study: a randomized, placebo-controlled trial of the health effects of antioxidant vitamins and minerals. *ArchInternMed*. 2004;164:2335–42.
  34. Hercberg S, Preziosi P, Briancon S, Galan P, Triol I, Malvy D, Roussel AM, Favier A. A primary prevention trial using nutritional doses of antioxidant vitamins and minerals in cardiovascular diseases and cancers in a general population: the SU.VI.MAX study--design, methods, and participant characteristics. *SUPPLEMENTATION EN VITAMINES ET MINÉRAUX ANTIOXYDANTS. CONTROL CLIN TRIALS*. 1998;19:336–51.
  35. Le Moullec N, Deheeger M, Preziosi P, Montero P, Valeix P, Rolland-Cachera MF, Potier de Courcy G, Christides JP, Galan P, Hercberg S. Validation du manuel photos utilisé pour l'enquête alimentaire de l'étude SU.VI.MAX (Validation of the food portion size booklet used in the SU.VI.MAX study). *CahNutrDiet*. 1996;31:158–64.
  36. Hercberg S. Table de Composition SU.VI.MAX des Aliments. Paris: Les éditions INSERM/Economica; 2005.
  37. Estaquio C, Kesse-Guyot E, Deschamps V, Bertrais S, Dauchet L, Galan P, Hercberg S, Castetbon K. Adherence to the French Programme National Nutrition Sante Guideline Score is associated with better nutrient intake and nutritional status. *JAmDiet Assoc*. 2009;109:1031–41.
  38. Kesse-Guyot E, Touvier M, Henegar A, Czernichow S, Galan P, Hercberg S, Castetbon K. Higher adherence to French dietary guidelines and chronic diseases in the prospective SU.VI.MAX cohort. *EurJ ClinNutr*. 2011;65:887–94.
  39. Lassale C, Galan P, Julia C, Fezeu L, Hercberg S, Kesse-Guyot E. Association between adherence to nutritional guidelines, the metabolic syndrome and adiposity markers in a French adult general population. *PLoS One*. 2013;8:e76349.
  40. Lassale C, Fezeu L, Andreeva VA, Hercberg S, Kengne A-P, Czernichow S, Kesse-Guyot E. Association between dietary scores and 13-year weight change and obesity risk in a French prospective cohort. *Int J Obes* 2005. 2012;36:1455–62.
  41. Lavalette C, Adjibade M, Srour B, Sellem L, Fiolet T, Hercberg S, Latino-Martel P, Fassier P, Deschasaux M, Kesse-Guyot E, et al. Cancer-Specific and General Nutritional Scores and Cancer Risk: Results from the Prospective NutriNet-Santé Cohort. *Cancer Res*. 2018;78:4427–35.

## ANNEXE 6

42. Pereira F, Martin JF, Joly C, Sébédio JL, Pujos-Guillot E. Development and validation of a UPLC/MS method for a nutritional metabolomic study of human plasma. *Metabolomics*. 2010;6:207–18.
43. Giacomoni F, Le Corguille G, Monsoor M, Landi M, Pericard P, Petera M, Duperier C, Tremblay-Franco M, Martin J-F, Jacob D, et al. Workflow4Metabolomics: a collaborative research infrastructure for computational metabolomics. *Bioinformatics*. 2015;31:1493–5.
44. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, Fan TW-M, Fiehn O, Goodacre R, Griffin JL, et al. Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics*. 2007;3:211–21.
45. Cattell RB. The scree test for the number of factors. *Multivar Behav Res*. 1966;1:245–76.
46. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *J R Stat Soc B*. 1995;57:289–300.
47. Xia J, Wishart DS. Using MetaboAnalyst 3.0 for Comprehensive Metabolomics Data Analysis. *Curr Protoc Bioinforma*. 2016;55:14.10.1-14.10.91.
48. Fujita T, Hada T, Higashino K. Origin of D- and L-pipecolic acid in human physiological fluids: a study of the catabolic mechanism to pipecolic acid using the lysine loading test. *Clin Chim Acta Int J Clin Chem*. 1999;287:145–56.
49. Perera T, Young MR, Zhang Z, Murphy G, Colburn NH, Lanza E, Hartman TJ, Cross AJ, Bobe G. Identification and monitoring of metabolite markers of dry bean consumption in parallel human and mouse studies. *Mol Nutr Food Res*. 2015;59:795–806.
50. Fukai K, Harada S, Iida M, Kurihara A, Takeuchi A, Kuwabara K, Sugiyama D, Okamura T, Akiyama M, Nishiwaki Y, et al. Metabolic Profiling of Total Physical Activity and Sedentary Behavior in Community-Dwelling Men. *PLoS One*. 2016;11:e0164877.
51. Zhao Q, Shen H, Su K-J, Tian Q, Zhao L-J, Qiu C, Garrett TJ, Liu J, Kakhniashvili D, Deng H-W. A joint analysis of metabolomic profiles associated with muscle mass and strength in Caucasian women. *Aging*. 2018;10:2624–35.
52. Wikoff WR, Anfora AT, Liu J, Schultz PG, Lesley SA, Peters EC, Siuzdak G. Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. *Proc Natl Acad Sci*. 2009;106:3698–703.
53. de Mello VD, Paananen J, Lindström J, Lankinen MA, Shi L, Kuusisto J, Pihlajamäki J, Auriola S, Lehtonen M, Rolandsson O, et al. Indolepropionic acid and novel lipid metabolites are associated with a lower risk of type 2 diabetes in the Finnish Diabetes Prevention Study. *Sci Rep* [Internet]. 2017 [cited 2019 Feb 8];7. Available from: <http://www.nature.com/articles/srep46337>
54. Tuomainen M, Lindström J, Lehtonen M, Auriola S, Pihlajamäki J, Peltonen M, Tuomilehto J, Uusitupa M, de Mello VD, Hanhineva K. Associations of serum indolepropionic acid, a gut microbiota metabolite, with type 2 diabetes and low-grade inflammation in high-risk individuals. *Nutr Diabetes* [Internet]. 2018 [cited 2019 Feb 8];8. Available from: <http://www.nature.com/articles/s41387-018-0046-9>
55. Jennis M, Cavanaugh CR, Leo GC, Mabus JR, Lenhard J, Hornby PJ. Microbiota-derived tryptophan indoles increase after gastric bypass surgery and reduce intestinal permeability in vitro and in vivo. *Neurogastroenterol Motil*. 2018;30:e13178.



## ANNEXE 6

56. Abildgaard A, Elfving B, Hokland M, Wegener G, Lund S. The microbial metabolite indole-3-propionic acid improves glucose metabolism in rats, but does not affect behaviour. *Arch Physiol Biochem.* 2018;124:306–12.
57. Pallister T, Jennings A, Mohny RP, Yarand D, Mangino M, Cassidy A, MacGregor A, Spector TD, Menni C. Characterizing Blood Metabolomics Profiles Associated with Self-Reported Food Intakes in Female Twins. *PloS One.* 2016;11:e0158568.
58. Eggersdorfer M, Wyss A. Carotenoids in human nutrition and health. *Arch Biochem Biophys.* 2018;652:18–26.
59. Tomás-Barberán FA, Selma MV, Espín JC. Interactions of gut microbiota with dietary polyphenols and consequences to human health: *Curr Opin Clin Nutr Metab Care.* 2016;19:471–6.
60. Steiber A, Kerner J, Hoppel CL. Carnitine: a nutritional, biosynthetic, and functional perspective. *Mol Aspects Med.* 2004;25:455–73.
61. Scalbert A, Brennan L, Manach C, Andres-Lacueva C, Dragsted LO, Draper J, Rappaport SM, van der Hoof JJ, Wishart DS. The food metabolome: a window over dietary exposure. *AmJClinNutr.* 2014;99:1286–308.
62. Shi L, Brunius C, Johansson I, Bergdahl IA, Lindahl B, Hanhineva K, Landberg R. Plasma metabolites associated with healthy Nordic dietary indexes and risk of type 2 diabetes—a nested case-control study in a Swedish population. *Am J Clin Nutr.* 2018;108:564–75.
63. Carayol M, Licaj I, Achaintre D, Sacerdote C, Vineis P, Key TJ, Onland Moret NC, Scalbert A, Rinaldi S, Ferrari P. Reliability of Serum Metabolites over a Two-Year Period: A Targeted Metabolomic Approach in Fasting and Non-Fasting Samples from EPIC. *PLoSOne.* 2015;10:e0135437.
64. Floegel A, Drogan D, Wang-Sattler R, Prehn C, Illig T, Adamski J, Joost HG, Boeing H, Pischon T. Reliability of serum metabolite concentrations over a 4-month period using a targeted metabolomic approach. *PLoSOne.* 2011;6:e21103.

## ANNEXE 6

**Table 1** Baseline characteristics<sup>1</sup> and scores of mPNNS-GS components<sup>1,2</sup> of participants according to their degree of adherence to the French dietary recommendations (mPNNS-GS, 1st decile=low compliance / 10th decile=high compliance) - SU.VI.MAX cohort, France (1994-2007)

	1st mPNNS-GS decile (N=80)	10th mPNNS-GS decile (N=80)	P <sup>3</sup>
mPNNS-GS	4.4±0.8	10.1±0.7	Not applicable
Age at baseline (y)	48.3±6.7	48±6.4	0.2
BMI (kg/m <sup>2</sup> )	22.2±3.1	22.4±2.7	0.5
BMI (categorized)			Not applicable
<18.5 kg/m <sup>2</sup> (underweight)	1 (1.3)	1 (1.3)	
≥18.5 to <25 kg/m <sup>2</sup> (normal weight)	67 (83.8)	67 (83.8)	
≥25 kg/m <sup>2</sup> (overweight)	12 (15.0)	12 (15.0)	
Smoking status			Not applicable
Never	58 (72.5)	53 (66.3)	
Former	21 (26.3)	26 (32.5)	
Current	1 (1.3)	1 (1.3)	
Month of blood draw			Not applicable
March-April	31 (38.8)	31 (38.8)	
October-November	12 (15.0)	12 (15.0)	
December-January-February	37 (46.3)	37 (46.3)	
Menopausal status at baseline			Not applicable
Premenopausal	52 (65.0)	52 (65.0)	
Postmenopausal	28 (35.0)	28 (35.0)	
Number of 24h-dietary records	12.1±1.1	12.1±1.2	0.8
Mean of daily energy intake (kcal)	2369.5±506.1	1771.2±272.7	<.0001
<b>Fruits and vegetables (portions/day)</b>	0.7 ±0.6	1.5 ±0.5	<.0001
[0-3.5[ = 0	20 (25.0)	0 (0.0)	0.0005
[3.5-5[ = 0.5	19 (23.8)	5 (6.3)	
[5-7.5[ = 1	33 (41.3)	34 (42.5)	
≥7.5 = 2	8 (10.0)	41 (51.3)	
<b>Bread, cereals, potatoes and legumes (portions/day)</b>	0.8 ±0.3	0.7 ±0.3	0.5
[0-1[ = 0	1 (1.3)	2 (2.5)	0.7
[1-3[ or ≥6 = 0.5	35 (43.8)	38 (47.5)	
[3-6[ = 1	44 (55.0)	40 (50.0)	
<b>Whole grain food (proportion of starches)</b>	0.1 ±0.3	0.4 ±0.4	<.0001
[0-1/3[ = 0	73 (91.3)	37 (46.3)	0.0003
[1/3-2/3[ = 0.5	2 (2.5)	20 (25.0)	
≥2/3 = 1	5 (6.3)	23 (28.8)	
<b>Milk and dairy products (portions/day)</b>	0.3 ±0.4	0.7 ±0.4	<.0001
[0-1[ or >3.5 (>4.5 if ≥55y) = 0	43 (53.8)	18 (22.5)	<.0001
[1-2.5[ = 0.5	19 (23.8)	11 (13.8)	
[2.5-3.5[ ([2.5-4.5] if ≥55y) = 1	18 (22.5)	51 (63.8)	
<b>Meat, poultry, seafood and eggs (portions/day)</b>	0.7 ±0.4	0.9 ±0.2	0.0004
0 or >2 = 0	20 (25.0)	2 (2.5)	0.007
]0-1[ = 0.5	12 (15.0)	8 (10.0)	
[1-2[ = 1	48 (60.0)	70 (87.5)	
<b>Seafood</b>	0.1 ±0.3	0.7 ±0.4	<.0001
<2 portions/week = 0	71 (88.8)	22 (27.5)	<.0001
≥2 portions/week = 1	9 (11.3)	58 (72.5)	
<b>Total added fat</b>	0.5 ±0.5	0.9 ±0.4	0.0002
Lipids from animal and vegetable added fat >16% EI/d = 0	38 (47.5)	12 (15.0)	0.0002
Lipids from animal and vegetable added fat ≤16% EI/d = 1	42 (52.5)	68 (85.0)	
<b>Vegetable added fat</b>	0.4 ±0.5	0.9 ±0.2	0.0002
No consumption of vegetable oil or ratio vegetable oil/total added fats ≤0.5 = 0	48 (60.0)	5 (6.3)	0.0002
No consumption of added fats or ratio vegetable oil/total added fats >0.5 = 1	32 (40.0)	75 (93.8)	
<b>Sweetened food</b>	0.5 ±0.5	1.0 ±0.2	0.0005
Added sugar from sweetened foods ≥15% EI/d = -0.5	3 (3.8)	0 (0.0)	0.003

## ANNEXE 6

Added sugar from sweetened foods 10 to 15% EI/d = 0	34 (42.5)	2 (2.5)	
Added sugar from sweetened foods <10% EI/d = 1	43 (53.8)	78 (97.5)	
<b>Sweetened beverages</b>	0.9 ±0.1	0.9 ±0.1	0.06
<11 water and >250 ml soda/d = 0	0 (0.0)	0 (0.0)	0.06
≥11 water and >250 ml soda/d = 0.5	0 (0.0)	0 (0.0)	
<11 water and ≤250 ml soda/d = 0.75	33 (41.3)	21 (26.3)	
≥11 water and ≤250 ml soda/d = 1	47 (58.8)	59 (73.8)	
<b>Alcohol</b>	0.5 ±0.4	0.8 ±0.2	<.0001
Ethanol >20 g/d for women and >30g/d for men = 0	34 (42.5)	5 (6.3)	0.0005
Ethanol ≤20 g/d for women and ≤30g/d for men = 0.8	41 (51.3)	66 (82.5)	
Abstainers and irregular consumers (<1 glass of wine/week) = 1	5 (6.3)	9 (11.3)	
<b>Salt</b>	-0.1 ±0.5	0.6 ±0.5	<.0001
>12 g/d = -0.5	43 (53.8)	3 (3.8)	<.0001
]10-12] g/d = 0	16 (20.0)	13 (16.3)	
]8-10] g/d = 0.5	14 (17.5)	28 (35.0)	
]6-8] g/d = 1	5 (6.3)	30 (37.5)	
≤6 g/d = 1.5	2 (2.5)	6 (7.5)	
<b>Physical activity</b>	0.3 ±0.5	0.6 ±0.6	0.002
[0-30[ min/d = 0	62 (77.5)	41 (51.3)	0.007
[30-60[ min/d = 1	10 (12.5)	23 (28.8)	
≥60 min/d = 1.5	8 (10.0)	16 (20.0)	
Basal Metabolic Rate	1314.2 ±80.3	1326.2 ±65.0	0.2
Energy expenditure	2097.9 ±172.8	2167.9 ±163.1	0.007
Ratio energy intake/expenditure	1.1 ±0.2	0.8 ±0.1	<.0001

## SU.VI.MAX, Supplémentation en Vitamines et Minéraux Antioxydants;

<sup>1</sup> Values are means ± SDs or N (%).

<sup>2</sup> mPNNS-GS components are highlighted in bold.

<sup>3</sup> P for the comparison between women of 1<sup>st</sup> and 10<sup>th</sup> mPNNS-GS deciles using conditional logistic regression. Not applicable for matching factors except for more precise variables (e.g. age and BMI).

## ANNEXE 6

**Table 2** Associations between ions and the level of adherence to the French dietary recommendations, as measured by the *a priori* mPNNS-GS, from multivariable conditional logistic regression<sup>1</sup> - SU.VI.MAX cohort, France (1994-2007)

Ions Mass (M). retention time (T)	Identification <sup>2</sup>	Protocol/ ionisation mode <sup>4</sup>	OR <sup>5</sup> [95% CI]	P- trend
M131.0895T1.16	<sup>3</sup> Pipecolic acid, [M+H] <sup>+</sup> - <sup>13</sup> C HMDB0000070	Endogenous ESI+	13.12 [2.58-66.60]	0.002
M130.0861T1.16	<sup>3</sup> Pipecolic acid, [M+H] <sup>+</sup> HMDB0000070	Endogenous ESI+	11.39 [2.39-54.32]	0.002
M190.0859T10.97	<sup>3</sup> 3-indolepropionic acid, [M+H] <sup>+</sup> HMDB0002302	Exogenous ESI+	4.39 [1.69-11.42]	0.002
M84.0808T1.17	<sup>3</sup> Pipecolic acid, [M+H-(HCOOH)] <sup>+</sup> HMDB0000070	Endogenous ESI+	10.43 [2.23-48.73]	0.003
M508.3395T15.27	(level 2: LysoPC (17:1), [M+H] <sup>+</sup> ) LMGP01050002	Exogenous ESI+	0.23 [0.09-0.63]	0.004
M330.2639T11.6	(level 3 : 4,8 dimethylnonanoyl carnitine, [M+H] <sup>+</sup> HMDB0006202/ undecanoyl-carnitine, [M+H] <sup>+</sup> HMDB0013321; acylcarnitine C11:1)	Exogenous ESI+	0.37 [0.18-0.75]	0.006
M510.4363T17.5	(level 4: unknown)	Endogenous ESI+	0.33 [0.15-0.74]	0.006
M303.2357T10.49	(level 3: 2,6 dimethylheptanoyl carnitine, [M+H] <sup>+</sup> - <sup>13</sup> C HMDB0006320/ nonanoyl-l-carnitine, [M+H] <sup>+</sup> - <sup>13</sup> C HMDB0013288; acylcarnitine C9:1)	Exogenous ESI+	0.28 [0.11-0.70]	0.007
M76.0393T1.17	(level 4 : unknown)	Endogenous ESI+	3.18 [1.37-7.35]	0.007
M189.0784T10.98	(level 4: unknown)	Exogenous ESI+	3.49 [1.40-8.67]	0.007
M737.8604T0.77	NaCOOH adduct	Exogenous ESI+	3.58 [1.40-9.17]	0.008
M252.0874T18.1	(level 4 : unknown)	Exogenous ESI+	3.43 [1.38-8.53]	0.008
M187.0577T0.97	<sup>3</sup> 2-deoxy-D-glucose, [M+Na] <sup>+</sup> HMDB0062477	Endogenous ESI+	0.36 [0.17-0.77]	0.009
M302.2325T10.49	(level 3: 2,6 dimethylheptanoyl carnitine, [M+H] <sup>+</sup> HMDB0006320/ nonanoyl-l-carnitine, [M+H] <sup>+</sup> HMDB0013288; acylcarnitine C9:1)	Exogenous ESI+	0.35 [0.16-0.78]	0.01
M325.1129T0.87	(level 3 : carbohydrates class C <sub>12</sub> H <sub>21</sub> O <sub>10</sub> )	Endogenous ESI+	0.29 [0.11-0.75]	0.01

SU.VI.MAX, Supplémentation en Vitamines et Minéraux Antioxydants; OR, odds ratio; CI, confidence interval

Associations presented in this table are the 15 strongest, see **Supplemental File 5** for all associations with a p<0.05.

<sup>1</sup> Subjects from the 1<sup>st</sup> and 10<sup>th</sup> deciles of mPNNS-GS were matched for age (35-39y/40-44y/45-49y/50-54y/55-59y/>60y), menopausal status (pre/post-menopause), BMI (<18.5kg/m<sup>2</sup> : underweight, ≥18.5 kg/m<sup>2</sup> to <25 kg/m<sup>2</sup> : normal weight and ≥25 kg/m<sup>2</sup> : overweight/obese), smoking status (current smokers and non/former smokers) and season (*a priori*-defined periods: October-November/December-January-February/March-April) at time of blood draw. Multivariable models were additionally adjusted for the following characteristics: age (continuous), BMI (continuous), smoking status (current/former smokers vs non-smokers), number of 24h-dietary records and mean of daily energy intake (kcal) during the 2 years following the blood draw. Tests for linear trend were performed using continuous variables.

<sup>2</sup> Levels of confidence for every identification were given accordingly to Sumner et al. (44): level 1; formally identified compound (confirmed with analysis of authentic standard), level 2; putatively identified compound (based upon spectral similarity with public/commercial spectral libraries or reference compound in the literature and/or physicochemical properties), level 3; putatively characterized compound classes and level 4; unknown compound.

<sup>3</sup>Metabolites confirmed by standard compounds (level 1)

## ANNEXE 6

<sup>4</sup>160 samples for endogenous protocol and exogenous protocol in negative ionization mode and 152 samples for exogenous protocol in positive ionization mode were analyzed (due to conditional analyses, the pairs corresponding to the samples excluded during pre-processing (N=4) were not analyzed).

<sup>5</sup> OR<1: associated with low adherence to the French dietary recommendations (lower quality of the overall diet)

OR>1: associated with high adherence to the French dietary recommendations (higher quality of the overall diet)

VERSION SOUMISE

## ANNEXE 6

**Table 3** Associations between ions and components of the *a priori* mPNNS-GS, from spearman correlation analyses<sup>1</sup> - SU.VI.MAX cohort, France (1994-2007)

Ions Mass (M), retention time (T)	Identification <sup>2</sup>	Protocol/ ionisation mode <sup>4</sup>	R	P-trend	mPNNS-GS Component
M231.0839T1.16	(level 2 : <b>Ethyl-β-D-glucopyranoside</b> , [M+Na] <sup>+</sup> ) HMDB0029968	Exogenous ESI+	0.37	0.000003	
M136.8774T0.94	(level 4: unknown)	Endogenous ESI-	-0.32	0.00006	Alcohol (g/day)
M138.8751T0.92	(level 4: unknown)	Exogenous ESI-	-0.30	0.0001	
M131.0895T1.16	<sup>3</sup> <b>Pipecolic acid</b> , [M+H] <sup>+</sup> <sup>13</sup> C HMDB0000070	Endogenous ESI +	0.31	0.00009	
M303.2357T10.49	(level 3: <b>2,6 dimethylheptanoyl carnitine</b> , [M+H] <sup>+</sup> <sup>13</sup> C, HMDB0006320 / <b>nonanoyl-L-carnitine</b> , [M+H] <sup>+</sup> <sup>13</sup> C HMDB0013288; <b>acylcarnitine C9:1</b> )	Exogenous ESI +	-0.31	0.0001	Fruit and vegetable (number of portions/day)
M149.0596T5.81	<sup>3</sup> <b>Phenylalanine</b> , [(M+H)-NH <sub>3</sub> ] <sup>+</sup> HMDB0000159	Exogenous ESI +	-0.30	0.0002	
M342.2637T11.93	(level 4: unknown)	Exogenous ESI +	-0.31	0.0001	Meat, poultry, seafood and eggs (number of portions/day)
M313.1546T8.83	<sup>3</sup> <b>Phenylalanyl-phenylalanine</b> , [M+H] <sup>+</sup> HMDB0013302	Exogenous ESI +	-0.31	0.0002	Milk and dairy products (number of portions/day)
M312.2261T15.13	(level 4: unknown)	Exogenous ESI -	0.30	0.0002	
M318.1622T15.23	(level 4: unknown)	Exogenous ESI +	-0.35	0.00001	
M241.1031T0.89	(level 4: unknown)	Exogenous ESI +	-0.32	0.00007	Seafood (intake frequency/week)
M157.0118T23.44	(level 4: unknown)	Exogenous ESI -	0.31	0.00009	
M810.5654T24.77	(level 4: unknown)	Exogenous ESI -	0.31	0.0001	
M373.2736T12.81	<sup>3</sup> <b>Cholic acid</b> , [(M+H)-2(H <sub>2</sub> O)] <sup>+</sup> HMDB00619	Exogenous ESI +	-0.31	0.00009	Sweetened food (% energy intake without alcohol/day)
M508.3395T15.27	(level 2: <b>LysoPC (17:1)</b> , [M+H] <sup>+</sup> ) LMGP01050002)	Exogenous ESI +	-0.30	0.0002	Vegetable added fat (% vegetable oil/total added fat)

SU.VI.MAX, Supplémentation en Vitamines et Minéraux Antioxydants; R; correlation coefficient

Associations presented in this table are the 15 strongest, see **Supplemental File 6** for all associations with a p<0.05.

<sup>1</sup> Partial spearman correlation adjusted for age (continuous), BMI (continuous), menopausal status (pre/post-menopause) and smoking status (current/former/never) at baseline, season at time of blood draw (a priori-defined periods: October/November, December/January/February, and March-April), number of 24h-dietary records and mean of daily energy intake (kcal) during the 2 years following the blood draw.

<sup>2</sup> Levels of confidence for every identification were given accordingly to Sumner et al. (44): level 1; formally identified compound (confirmed with analysis of authentic standard), level 2; putatively identified compound (based upon spectral similarity with public/commercial spectral libraries or reference compound in the literature and/or physicochemical properties), level 3; putatively characterized compound classes and level 4; unknown compound.

<sup>3</sup>Metabolites confirmed by standard compounds (level 1)

<sup>4</sup>160 samples for endogenous protocol and exogenous protocol in negative ionization mode and 156 samples for exogenous protocol in positive ionization mode were analyzed.

## ANNEXE 6

**Table 4** Associations between ions and *a posteriori* dietary patterns derived by principal component analysis from spearman correlation analyses<sup>1</sup> - SU.VI.MAX cohort, France (1994-2007)

Ions Mass (M). retention time (T)	Identification <sup>2</sup>	Protocol/ ionisation mode <sup>4</sup>	R	P-trend	PCA-derived dietary pattern
M131.0895T1.16	<sup>3</sup> Pipecolic acid, [M+H] <sup>+</sup> - <sup>13</sup> C HMDB0000070	Endogenous ESI+	-0.36	0.000006	
M130.0861T1.16	<sup>3</sup> Pipecolic acid, [M+H] <sup>+</sup> HMDB0000070	Endogenous ESI+	-0.35	0.000008	
M84.0808T1.17	<sup>3</sup> Pipecolic acid, [M+H-(HCOOH)] <sup>+</sup> HMDB0000070	Endogenous ESI+	-0.32	0.000005	Factor 1 “Western diet”
M246.9051T14.25	(level 4: C <sub>2</sub> H <sub>2</sub> O <sub>7</sub> N <sub>2</sub> Br)	Exogenous ESI-	-0.30	0.0002	
M286.1437T13.64	<sup>3</sup> Piperine, [M+H] <sup>+</sup> HMDB0029377	Exogenous ESI+	0.26	0.001	
M131.0895T1.16	<sup>3</sup> Pipecolic acid, [M+H] <sup>+</sup> - <sup>13</sup> C HMDB0000070	Endogenous ESI +	0.36	0.000004	
M130.0861T1.16	<sup>3</sup> Pipecolic acid, [M+H] <sup>+</sup> HMDB0000070	Endogenous ESI +	0.34	0.00002	
M84.0808T1.17	<sup>3</sup> Pipecolic acid, [M+H-(HCOOH)] <sup>+</sup> HMDB0000070	Endogenous ESI +	0.32	0.000005	
M303.2357T10.49	(level 3: <b>2,6 dimethylheptanoyl carnitine</b> , [M+H] <sup>+</sup> - <sup>13</sup> C HMDB0006320/ <b>nonanoyl-l-</b> <b>carnitine</b> , [M+H] <sup>+</sup> - <sup>13</sup> C HMDB0013288; <b>acylcarnitine C9:1</b> )	Exogenous ESI +	-0.30	0.0002	Factor 2 “Healthy diet”
M119.016T1.03	(level 2: <b>Dihydro4mercapto-3(2H)</b> <b>furanone</b> , [M+H] <sup>+</sup> ) HMDB0039786	Endogenous ESI +	0.28	0.0004	
M73.0284T1.02	(level 4 : unknown)	Endogenous ESI +	0.28	0.0006	
M918.3358T10.11	(level 4 : unknown)	Endogenous ESI +	-0.27	0.0009	
M224.0998T11.87	(level 4 : C <sub>12</sub> H <sub>15</sub> O <sub>4</sub> )	Endogenous ESI +	-0.30	0.0002	
M340.0238T11.87	(level 4 : unknown)	Endogenous ESI +	-0.26	0.001	Factor 3 “Sweetened products”
M498.9301T19.76	(level 4 : unknown)	Exogenous ESI -	-0.26	0.001	

SU.VI.MAX, Supplémentation en Vitamines et Minéraux Antioxydants; R; correlation coefficient

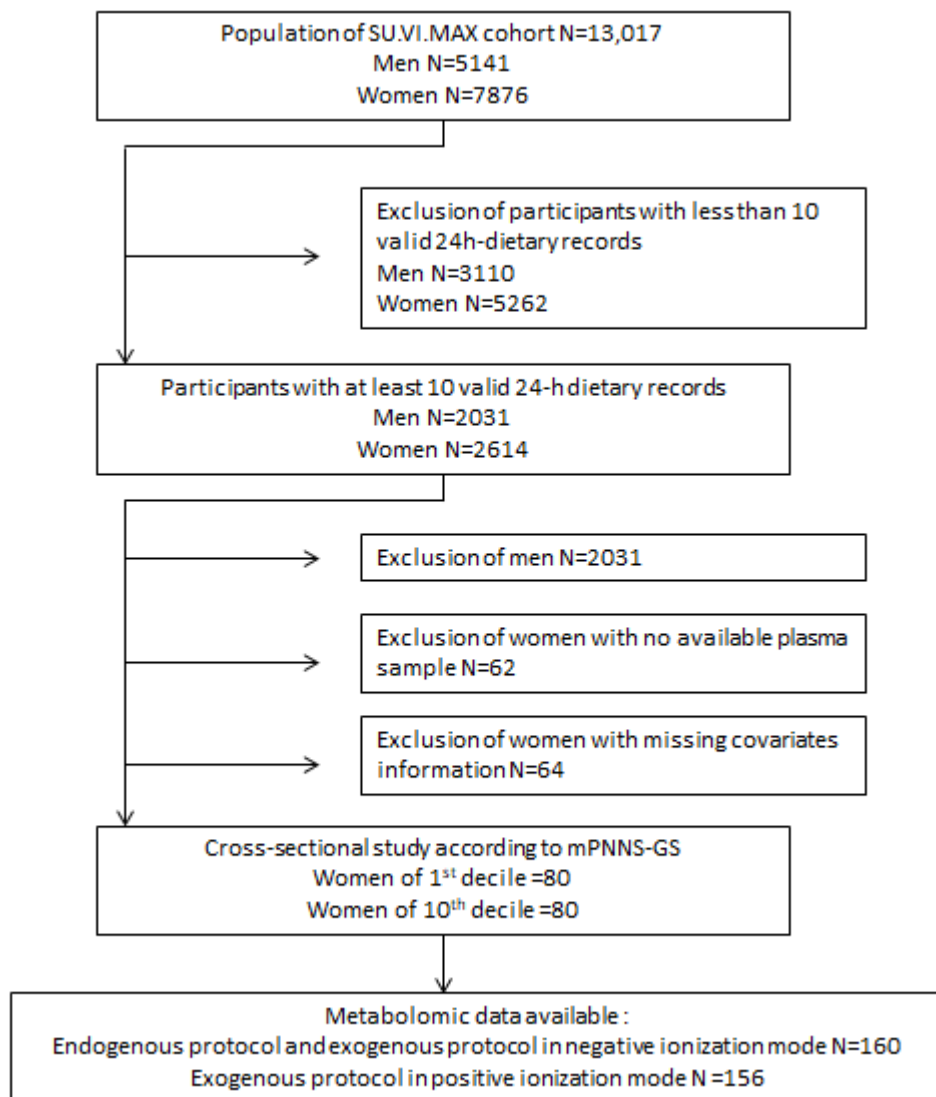
Associations presented in this table are the 15 strongest, see **Supplemental File 7** for all associations with a p<0.05.

<sup>1</sup> Partial spearman correlation adjusted for age (continuous), BMI (continuous), menopausal status (pre/post-menopause) and smoking status (current/former/never) at baseline, season at time of blood draw (a priori-defined periods: October/November, December/January/February, and March-April), number of 24h-dietary records and mean of daily energy intake (kcal) during the 2 years following the blood draw.

<sup>2</sup> Levels of confidence for every identification were given accordingly to Sumner et al. (44): level 1; formally identified compound (confirmed with analysis of authentic standard), level 2; putatively identified compound (based upon spectral similarity with public/commercial spectral libraries or reference compound in the literature and/or physicochemical properties), level 3; putatively characterized compound classes and level 4; unknown compound.

<sup>3</sup>Metabolites confirmed by standard compounds (level 1)

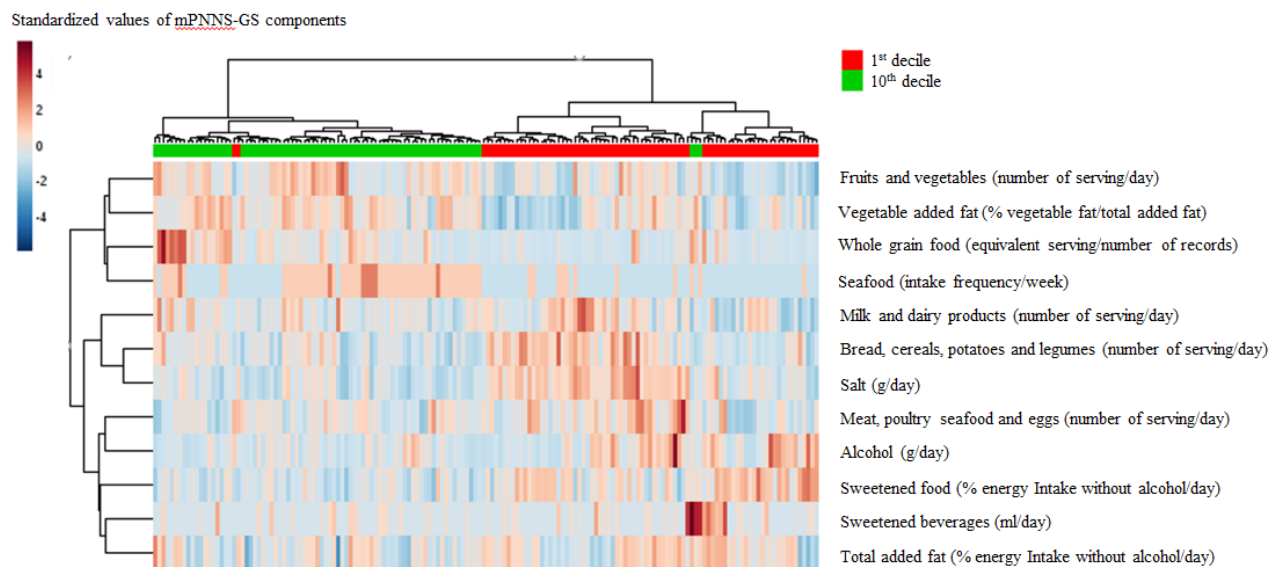
<sup>4</sup>160 samples for endogenous protocol and exogenous protocol in negative ionization mode and 156 samples for exogenous protocol in positive ionization mode were analyzed.

**Figure 1:** Participant flowchart, selection of the SU.VI.MAX study



## ANNEXE 6

**Figure 2:** Hierarchical Clustering Analysis (HCA) of the mPNNS-GS components using Euclidean distance and Ward aggregation method as clustering algorithm



**Supplemental file 1**

**Supplementary Table S1:** Loadings values of the *a posteriori* 3 dietary patterns derived by Principal Component Analysis<sup>1</sup> - SU.VI.MAX cohort, France (1994-2007)

Food group	Factor1 = “Western diet”	Factor2 = “Healthy diet”	Factor3 = “Sweetened diet”
Alcoholic drinks	0.57	-0.22	-0.26
Animal fat	0.44	-0.02	0.31
Bread	0.60	-0.02	0.14
Breakfast cereals	-0.19	0.27	0.17
Cakes, biscuits	0.08	-0.12	0.56
Cheese	0.48	0.18	0.04
Dried fruits	-0.02	0.25	0.35
Eggs	0.16	0.09	0.08
Fish, seafood	0.19	0.25	-0.05
Fruits	0.06	0.64	0.14
Fruit juice	-0.05	-0.03	0.31
Milk	0.02	0.02	0.29
Nuts	0.10	0.22	0.25
Other <sup>2</sup>	0.01	0.16	0.21
Pasta, rice, cereals	0.31	0.05	0.15
Potatoes, tubers	0.49	0.00	0.09
Processed meat	0.56	-0.11	0.04
Legumes	0.27	0.14	0.05
Red meat, offal	0.46	-0.13	-0.18
Salted products	0.31	-0.05	0.18
Sauces	0.33	-0.03	0.13
Sugar confectionery	0.27	0.03	0.60
Sweetened drinks	0.09	-0.08	0.21
Unsweetened drinks	-0.10	0.24	-0.04
Vegetables	0.19	0.71	-0.10
Vegetable fat	0.39	0.56	-0.20
White meat	0.44	0.06	-0.22
Whole grain	-0.07	0.43	-0.00
Yoghurt	-0.37	0.44	0.04

<sup>1</sup>Patterns of overall diet were obtained using a principal component analysis (PCA) with varimax rotated factors by orthogonal transformation on the 29 food groups on the general SU.VI.MAX population having at least 10 validated 24h-dietary records during the 2 years following baseline (N=4645). The three first factors (=“patterns”) were retained according to the plot of the total variance explained (Cattel’s scree plots) (1).

<sup>2</sup>Condiments, cooking aids, water, frogs’ legs, snail, seeds, Growing-up milk, meal replacement, hyperproteinated food, soya food.

## REFERENCES

1. Cattell RB. The scree test for the number of factors. *Multivar BehavRes.* 1966 Mar;1:245–76.

**Supplemental file 2:** Description and computation of mPNNS-GS

Adherence to the 2001 recommendations of the French ‘Programme National Nutrition Santé’, reflecting the quality of the overall diet was computed using the PNNS-GS. The score computation, including food grouping, serving sizes, scoring, cut-off values and penalties, has previously been described in detail (1). Briefly, the score includes 13 components and has a range of 0–15 points (see **Supplementary Table S2** below): fruit and vegetable; bread, cereals, potatoes and legumes; whole grain food; milk and dairy products; meat and poultry, seafood and eggs; seafood; vegetable added fats; total added fats; sweetened foods; sweetened beverages; salt; alcohol, and one component covers adherence to physical activity recommendations. A penalty for overconsumption was when energy intakes are  $\geq 5\%$  higher than estimates of energy expenditure based on basal metabolic rate and physical activity level. Higher scores indicate better adherence to the French national nutrition guidelines and quality of overall diet. Several publications (2–4) have used a modified score, the mPNNS-GS, which contains only the 12 dietary items and has a maximum value of 13.5, in order to specifically investigate dietary quality (and not a combination of dietary quality and physical activity). For the present study, we used this modified score.

**Supplementary Table S2:** Components and scoring for calculation of the PNNS-GS, according to PNNS recommendation scores

	Recommendation	Scoring criteria	Score
<b>1. Fruits and vegetables</b>	<b>5 servings/day</b>	<b>[0-3.5[</b>	<b>0</b>
		<b>[3.5-5[</b>	<b>0.5</b>
		<b>[5-7.5[</b>	<b>1</b>
		<b>≥7.5</b>	<b>2</b>
<b>2. Bread, cereals, potatoes and legumes</b>	<b>At each meal according to appetite</b>	<b>[0-1[</b>	<b>0</b>
		<b>[1-3[</b>	<b>0.5</b>
		<b>[3-6[</b>	<b>1</b>
		<b>≥6</b>	<b>0.5</b>
<b>3. Whole grain food</b>	<b>Preferentially choose whole grains and whole grain breads</b>	<b>[0-1/3[</b>	<b>0</b>
		<b>[1/3-2/3[</b>	<b>0.5</b>
		<b>≥2/3</b>	<b>1</b>
<b>4. Milk and dairy products</b>	<b>3 servings/day (individuals aged ≥55 years: 3 to 4 servings/day)</b>	<b>[0-1[</b>	<b>0</b>
		<b>[1-2.5[</b>	<b>0.5</b>
		<b>[2.5-3.5] (≥55-years-old: [2.5-4.5])</b>	<b>1</b>
		<b>&gt;3.5 (≥55-years-old : &gt;4.5)</b>	<b>0</b>
<b>5. Meat, poultry seafood and eggs</b>	<b>1 to 2 servings/day</b>	<b>0</b>	<b>0</b>
		<b>]0-1[</b>	<b>0.5</b>
		<b>[1-2]</b>	<b>1</b>
		<b>&gt;2</b>	<b>0</b>
<b>6. Seafood</b>	<b>At least 2 servings/week</b>	<b>&lt; 2/week</b>	<b>0</b>
		<b>≥ 2/week</b>	<b>1</b>
<b>7. Total added fat<sup>1</sup></b>	<b>Limit consumption</b>	<b>Lipids from animal and vegetable added fat &gt;16% EI/d</b>	<b>0</b>
		<b>Lipids from animal and vegetable added fat ≤16% EI/d</b>	<b>1</b>
<b>8. Vegetable added fat</b>	<b>Favour fats of vegetable origin</b>	<b>No consumption of vegetable oil or ratio vegetable oil/total added fats ≤0.5</b>	<b>0</b>
		<b>No consumption of added fats or ratio vegetable oil/total added fats &gt;0.5</b>	<b>1</b>

9. Sweetened food <sup>1</sup>	Limit consumption	Added sugar from sweetened foods ≥15% EI/d	-0,5
		Added sugar from sweetened foods 10 to 15% EI/d	0
		Added sugar from sweetened foods <10% EI/d	1
10. Sweetened beverages	Limit sweetened beverages to <1 glass/day	<1l water and >250 ml soda/d	0
		≥1l water and >250 ml soda/d	0.5
		<1l water and ≤250 ml soda/d	0.75
		≥1l water and ≤250 ml soda/d	1
11. Alcohol <sup>2</sup>	≤2 glasses of wine/day for women and ≤3 glasses of wine/day for men	Ethanol >20 g/d for women and >30g/d for men	0
		Ethanol ≤20 g/d for women and ≤30g/d for men	0.8
		Abstainers and irregular consumers (<1 glass of wine/week))	1
12. Salt	Limit consumption	>12g/d	-0.5
		]10-12] g/d	0
		]8-10] g/d	0.5
		]6-8] g/d	1
		≤6 g/d	1.5
13. Physical activity	At least the equivalent of 30 min of brisk walking/day	[0-30[ min/d	0
		[30-60[ min/d	1
		≥60 min/d	1.5

Abbreviations: EI, total energy intake without alcohol; PNNS-GS, Programme National Nutrition Santé Guideline Score.

<sup>1</sup>Established according to the French recommended daily allowance.

<sup>2</sup>Pregnant women are advised to abstain from all alcohol consumption during the entire pregnancy.

## REFERENCES

1. Estaquio C, Kesse-Guyot E, Deschamps V, Bertrais S, Dauchet L, Galan P, Hercberg S, Castetbon K. Adherence to the French Programme National Nutrition Sante Guideline Score is associated with better nutrient intake and nutritional status. *JAmDiet Assoc.* 2009;109:1031–41.
2. Lelong H, Blacher J, Menai M, Galan P, Fezeu L, Hercberg S, Kesse-Guyot E. Association Between Blood Pressure and Adherence to French Dietary Guidelines. *Am J Hypertens.* 2016;29:948–58.
3. Lassale C, Galan P, Castetbon K, Péneau S, Méjean C, Hercberg S, Kesse-Guyot E. Differential association between adherence to nutritional recommendations and body weight status across educational levels: a cross-sectional study. *Prev Med.* 2013;57:488–93.
4. Assmann KE, Lassale C, Galan P, Hercberg S, Kesse-Guyot E. Dietary quality and 6-year anthropometric changes in a sample of French middle-aged overweight and obese adults. *PLoSOne.* 2014;9:e87083-.

VERSION SOUMISE

### Supplemental file 3: methods for MS metabolomic analysis

#### *Chemicals and reagents*

Methanol (Lichrosolv Hypergrade for LC-MS) was purchased from Merck. Acetonitrile (ULC-MS), water (ULC-MS) and formic acid (99%) were purchased from Biosolve. All standard compounds were purchased from Sigma-Aldrich (Saint Quentin Fallavier, France). All were analytical standard grade ( $\geq 98\%$ ).

#### *Plasma sample preparation*

Two complementary protocols of plasma sample preparation were used to ensure a wider coverage of metabolites. The first protocol, developed for endogenous metabolites, follows a slightly modified version of the procedure described in Pereira et al. (1). In brief, plasma samples (100  $\mu\text{L}$ ) were deproteinized using cold methanol; after centrifugation and evaporation, the dry residues were redissolved in 500  $\mu\text{L}$  acetonitrile/water (50:50, v/v) containing 0.1% formic acid. In the second protocol, developed for exogenous metabolites, 100  $\mu\text{L}$  plasma were deproteinized using methanol containing 0.1% formic acid to break the hydrogen bonds between metabolites and plasma proteins, then centrifuged for 10 min at 13,000 rpm, 4°C. 250  $\mu\text{L}$  acetonitrile were added to the collected supernatant (250  $\mu\text{L}$ ). Extracted plasma were stored at -80°C until analysis. For each protocol, quality controls (QC) were prepared by pooling 10  $\mu\text{L}$  of all extracted plasma samples.

#### *Untargeted UPLC-MS metabolomics analyses*

Metabolomics profiles were acquired with an UHPLC-QToF Impact II (Bruker). Briefly, 5  $\mu\text{L}$  of sample were injected into an U3000 UHPLC system (Thermo Scientific, France) coupled to a high-resolution Bruker Impact HD II UHR-QToF equipped with electrospray ionization source (ESI). Chromatographic separation was performed using a WatersAcquity HSS T3 C18 column (150 x 2.1 mm), with a particle size of 1.8  $\mu\text{m}$ . The flow rate was set at 0.4 mL/min and 30°C. Mobile phases A and B were milliQ water and acetonitrile respectively, both containing 0.1% formic acid. The elution gradient was carried out as follows: 100% A (0-2 min); linear gradient to 100% B (2-15 min); 100% B (15-22 min); 100% to 0% B (22.1 min); 100% A (22.1-26 min for re-equilibration). Samples were kept at 4°C during analyses. Data were acquired in positive and negative ionization modes with a full scan range from 50 to 1000 mass-to-charge ratio (m/z). Samples were randomized within the analytical sequence based on a Williams Latin Square strategy defined according to the mPNNS-GS score (1<sup>st</sup> or 10<sup>th</sup> decile of the score). For each protocol, quality controls (QC) were prepared by pooling 10  $\mu\text{L}$  of all extracted plasma samples. The stability of the analytical system was monitored using these QC, injected at the beginning of each sequence and then every 12 samples.

#### *Data extraction*

Raw data were processed under the Galaxy web-based platform Workflow4metabolomics (2), using XCMS module (peak detection, grouping and chromatographic alignment), followed by quality checks and signal drift correction according to the algorithm described by Van der Kloet et al (3), to yield a data matrix containing retention times, masses and peak intensities that have been corrected for batch effects. For the exogenous protocol in positive mode, 4 samples were excluded due to abnormal low intensity of all ions. Details on processing steps and number of ions in each data matrix are presented in the **Supplementary Figure S1**. After signal drift correction, ions were discarded when intensities had more variability in quality controls than in samples and when ions were annotated as isotopes. Due to a wide extraction of ions, we also discarded ions when  $\geq 30\%$  of the quality controls per analytical batch had a null intensity. Moreover, since in this study we are interested in diet-derived metabolites, ions present in at least 40% of 1<sup>st</sup> or 10<sup>th</sup> decile samples were kept to deal with inter-individual variability in dietary habits. Highly correlated ions (correlation threshold at 0.8) from the same metabolite within the same retention time cluster were removed using the Metabolite Correlation Analysis (MCA) Galaxy module. These processing steps allowed to discard ions considered as noise, characterized by a too high variability, as well as to eliminate redundancy between signals (isotopes, adducts, fragments).

**Metabolite identification**

Ions of interest were first annotated as parent molecular ions, fragments or adducts by manual inspection of the QToF MS spectra (Bruker Compass DataAnalysis version 4.3). To determine the elemental formula of the parent ions of interest, a full scan MS ( $m/z$  70 – 1,000) at high resolving power ( $R = 100,000$  at  $m/z = 800$ ) was obtained for selected samples with a ThermoScientific LTQ Orbitrap Velos hybrid mass spectrometer (Thermo Fisher Scientific, San José, CA, USA). Hypotheses of identification were obtained by querying online databases with the elemental formula or the monoisotopic mass (mass window 0.005 Da): Metlin (<https://metlin.scripps.edu>), HMDB (<http://www.hmdb.ca>) (4), FooDB (<http://foodb.ca>), Dictionary of Food Compounds (<http://dfc.chemnetbase.com>), PhytoHub (<http://phytohub.eu>), ChEBI (<https://www.ebi.ac.uk/chebi/>), KNApSAcK (<http://kanaya.naist.jp/KNApSAcK>). Fragmentation spectra (MS/MS) were obtained on the LTQ Orbitrap (Higher-energy C-trap Dissociation (HCD) techniques) and/or QToF mass spectrometers (collision energies 12eV, 20eV, 30eV and 40eV;  $m/z$  width 7 Da) and compared to reference spectra in our in-house, online spectral libraries such as mzcloud (<https://www.mzcloud.org>), Metlin (<https://metlin.scripps.edu>), HMDB (<http://www.hmdb.ca>) data from the literature, and in silico predicted fragmentation pathways generated with MassFrontier (Thermo-Scientific). Identifications were confirmed with commercial standards when available. Levels of identification were used according to Sumner et al (5): level 1 refers to identified compounds (monoisotopic mass, MS/MS fragmentation pattern and retention time matching with those of authentic standards); level 2 refers to putatively annotated compounds (monoisotopic mass and MS/MS fragmentation pattern matching with spectra published in online libraries/publications); level 3 refers to putatively characterized compound classes (spectral similarity to a chemical class); level 4 refers to unknown compounds. Main fragments of the identified metabolites (level 1 & 2) are presented in the **Supplementary Table S3**.

**Supplementary Table S3** Main fragments of the identified metabolites (level 1 & 2)

Compound	Level	RT (min)	Measured Mass (Da)	Fragments/Adducts (Da)
Pipecolic acid	1	1.16	130.0861 [M+H] <sup>+</sup>	130.0861 [M+H] <sup>+</sup> 84.0808
3-indolepropionic acid	1	10.97	190.0859 [M+H] <sup>+</sup>	190.0859 [M+H] <sup>+</sup> 172.0772 130.0663
LysoPC (17:1)	2	15.27	508.3395 [M+H] <sup>+</sup>	508.3395 [M+H] <sup>+</sup> 320.8102 307.2632 295.2631 287.6331 279.2338 258.1158 251.2370 184.0774 104.1092
2-deoxy-D-glucose	1	0.97	187.0577 [M+Na] <sup>+</sup>	187.0577 [M+Na] <sup>+</sup> 351.1260
Ethyl-β-D-glucopyranoside	2	1.16	231.0839 [M+Na] <sup>+</sup>	231.0839 [M+Na] <sup>+</sup> 219.0260 204.0554 203.0526 198.0977 149.0231
Phenylalanine	1	5.81	166.0864 [M+H] <sup>+</sup>	166.0864 [M+H] <sup>+</sup> 149.0596 120.0809
Phenylalanyl-phenylalanine	1	8.83	313.1546 [M+H] <sup>+</sup>	313.1546 [M+H] <sup>+</sup> 120.0807

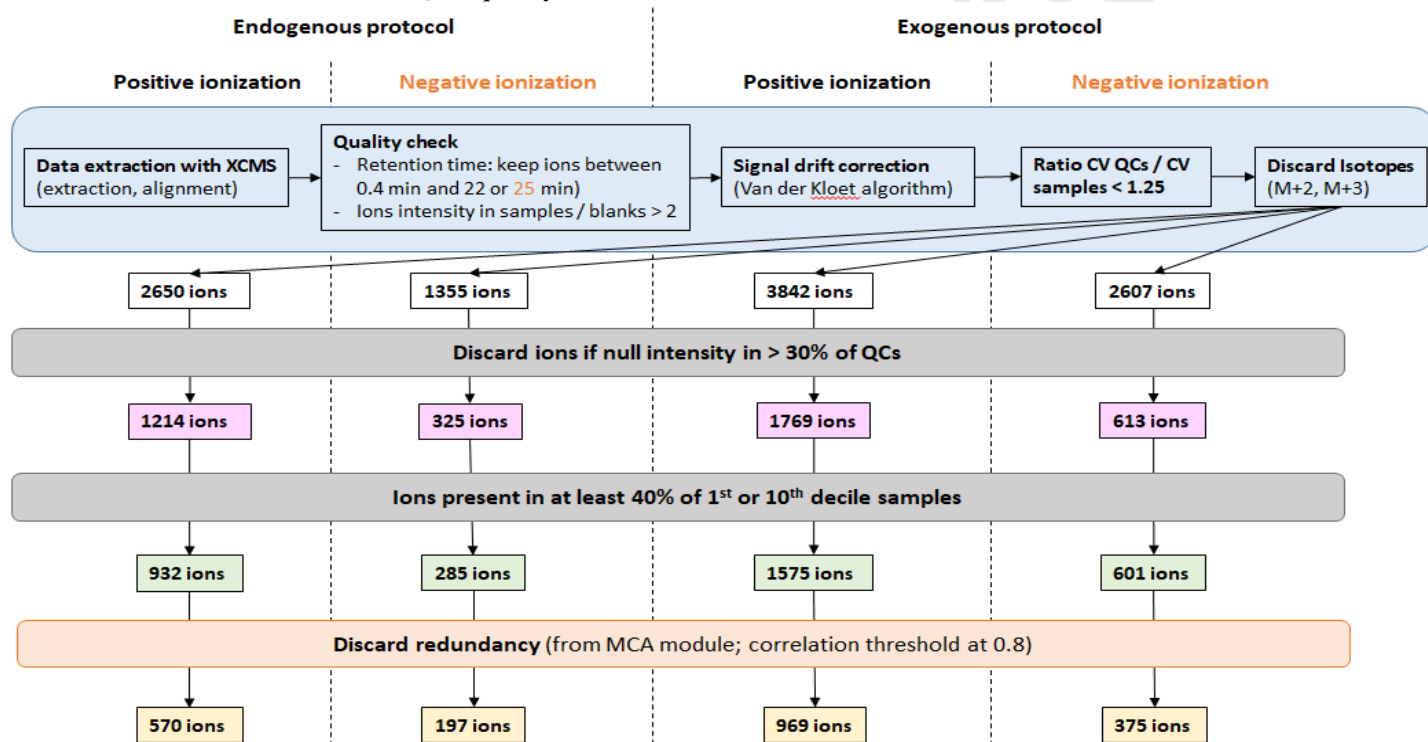


## ANNEXE 6

				426.3214 [M+H] <sup>+</sup>
				373.2736
<b>Cholic acid</b>	1	12.81	426.3214 [M+H] <sup>+</sup>	355.2635
				245.1532
				213.1634
				159.1164
<b>Piperine</b>	1	13.64	286.1437 [M+H] <sup>+</sup>	286.1437 [M+H] <sup>+</sup>
				201.0546
				135.0441
<b>Dihydro4mercapto-3(2H) furanone</b>	2	1.03	119.016 [M+H] <sup>+</sup>	119.016 [M+H] <sup>+</sup>
				77.0056
				74.9899
				73.0285

## Supplementary Figure S1 Data processing steps.

CV, coefficient of variation; QCs, quality controls.



## REFERENCES

1. Pereira F, Martin JF, Joly C, Sébédio JL, Pujos-Guillot E. Development and validation of a UPLC/MS method for a nutritional metabolomic study of human plasma. *Metabolomics*. 2010 Feb 19;6:207-218.
2. Giacomoni F, Le Corguille G, Monsoor M, Landi M, Pericard P, Petera M, et al. Workflow4Metabolomics: a collaborative research infrastructure for computational metabolomics. *Bioinformatics*. 2015 May 1;31(9):1493-5.
3. van der Kloet FM, Bobeldijk I, Verheij ER, Jellema RH. Analytical error reduction using single point calibration for accurate and precise metabolomic phenotyping. *JProteomeRes*. 2009 Nov;8:5132-41.
4. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, et al. HMDB 3.0--The Human Metabolome Database in 2013. *Nucleic Acids Res*. 2013 Jan;41:D801-7.
5. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, et al. Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics*. 2007 Sep 19;3(3):211-21.

## ANNEXE 6

**Supplemental file 4** Identified metabolites associated with baseline clinical and biological parameters<sup>1</sup> and nutrients intakes<sup>2</sup> from Spearman correlation analyses<sup>3</sup>  
- SU.VI.MAX cohort, France (1994-2007)

**Pipecolic acid**Clinical and biological parameters

	Body mass index (kg/m <sup>2</sup> )		Retinol (µmol/L)		Beta-carotene (µmol/L)	
	R	P	R	P	R	P
M130.0861T1.16			0.21	0.01		
M131.0895T1.16	-0.20	0.01	0.21	0.01		
M84.0808T1.17	-0.22	0.006	0.21	0.02	0.20	0.02

Nutrients intakes

	Carbohydrates (g/j)		Simple carbohydrates (g/j)		Total fibres (g/j)		Soluble fibres (g/j)		Beta-Carotene (µg/j)		Vit B5 - pantothenic acid (mg/j)		Vit B9 - folic acid (µg/j)		Vit C - ascorbic acid (mg/j)		Vit E - tocopherols (mg/j)	
	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P
M130.0861T1.16			0.29	0.0003	0.30	0.0001	0.28	0.0004	0.22	0.005	0.20	0.01	0.24	0.002	0.28	0.0004	0.25	0.002
M131.0895T1.16	0.20	0.01	0.28	0.0004	0.31	0.0001	0.28	0.0005	0.24	0.003	0.21	0.01	0.25	0.002	0.29	0.0002	0.26	0.001
M84.0808T1.17			0.25	0.002	0.28	0.0004	0.25	0.002					0.21	0.01	0.27	0.0008	0.23	0.003

	Calcium (mg/j)		Potassium (mg/j)		Magnesium (mg/j)		Sodium (mg/j)	
	R	P	R	P	R	P	R	P
M130.0861T1.16	0.22	0.008	0.26	0.001	0.23	0.004	-0.22	0.006
M131.0895T1.16	0.21	0.008	0.28	0.0005	0.22	0.006		
M84.0808T1.17			0.23	0.003	0.22	0.006		

**2-deoxy-D-glucose**

## ANNEXE 6

Nutrients intakes

	Total fibres (g/j)		Vit E - tocopherols (mg/j)		Potassium (mg/j)		Magnesium (mg/j)	
	R	P	R	P	R	P	R	P
M187.05773T0.9	-0.23	0.005	-0.23	0.005	-0.29	0.0004	-0.27	0.0007

**Dihydro4mercapto-3(2H) furanone**Nutrients intakes

	Alpha-linoleic acid 18:3 n-3 (g/j)		Total fibres (g/j)		Vit B5 - pantothenic acid (mg/j)		Vit B9 - folic acid (µg/j)		Vit C - ascorbic acid (mg/j)		Potassium (mg/j)	
	R	P	R	P	R	R	R	P	R	P	R	P
M119.016T1.03	0.21	0.008	-0.23	0.005	0.23	-0.29	0.28	0.0006	0.21	0.009	0.21	0.01

**2,6 Dimethylheptanoyl carnitine/ nonanoyl-l-carnitine**Clinical and biological parameters

	Cholesterol (mmol/L)		Apo-lipoprotein B (g/L)		Low density lipoprotein cholesterol (mmol/L)	
	R	P	R	P	R	P
M302.2325T10.49	0.22	0.006	0.25	0.002	0.26	0.002
M303.2357T10.49						

Nutrients intakes

	Saturated fatty acids(g/j)		Total fibres (g/j)		Soluble fibres (g/j)		beta-Carotene (µg/j)		Vit B1 - thiamine (mg/j)		Vit B5 - pantothenic acid (mg/j)		Vit B6 - pyridoxine (mg/j)	
	R	P	R	P	R	P	R	P	R	P	R	P	R	P
M302.2325T10.49	0.22	0.006	-0.21	0.01	-0.26	0.001	-0.24	0.003					-0.27	0.0008

## ANNEXE 6

M303.2357T10.49	0.21	0.01	-0.27	0.001	-0.27	0.0007	-0.24	0.004	-0.21	0.01	-0.20	0.01	-0.31	0.0001
-----------------	------	------	-------	-------	-------	--------	-------	-------	-------	------	-------	------	-------	--------

	Vit B9 - folic acid (µg/j)		Vit C - ascorbic acid (mg/j)		Vit E - tocopherols (mg/j)		Iron (mg/j)		Potassium (mg/j)		Magnesium (mg/j)	
	R	P	R	P	R	P	R	P	R	P	R	P
M302.2325T10.49	-0.20	0.01	-0.20	0.01	-0.25	0.003	-0.21	0.01				
M303.2357T10.49	-0.27	0.001	-0.22	0.007	-0.26	0.001	-0.23	0.004	-0.21	0.01	-0.22	0.008

### 4,8 dimethylnonanoyl carnitine / undecanoylcarnitine

#### Nutrients intakes

	Soluble fibres (g/j)		Vit B9 - folic acid (µg/j)	
	R	P	R	P
M330.2639T11.6	-0.22	0.008	-0.21	0.01

### 3-indolepropionic acid

#### Clinical and biological parameters

	beta-Carotene (µmol/L)	
	R	P
M190.0859T10.97	0.22	0.01

### Phosphatidylcholine (17:1)

#### Clinical and biological parameters

	Cholesterol (mmol/L)	
	R	P
M508.3395T15.27	0.20	0.01

## ANNEXE 6

Nutrients intakes

	Saturated fatty acids(g/j)		Cholesterol (mg/j)		Vit E - tocopherols (mg/j)		Energy intake without alcoholic component (Kcal)	
	R	P	R	P	R	P	R	P
M508.3395T15.27	0.21	0.01	0.21	0.01	-0.23	0.004	0.20	0.01

**Ethyl- $\beta$ -D-gluco pyranoside**Clinical and biological parameters

	Weight (kg)		Vitamin C ( $\mu$ g/mL),	
	R	P	R	P
M231.0839T1.16	0.25	0.002	0.21	0.02

Nutrients intakes

	Carbohydrates (g/j)		Vit C - ascorbic acid (mg/j)		Alcohol (g/j)	
	R	P	R	P	R	P
M231.0839T1.16	-0.22	0.008	0.20	0.01	0.38	<0.001

**Piperine**Nutrients intakes

	Carbohydrates (g/j)		Simple carbohydrates (g/j)		Total fibres (g/j)	
	R	P	R	P	R	P
M286.1437T13.64	-0.22	0.008	-0.23	0.005	-0.24	0.003

### Cholic acid

#### Nutrients intakes

	Cholesterol (mg/j)		Simple carbohydrates (g/j)		Simple added sugar (g/j)		Vit B5 - pantothenic acid (mg/j)		Vit B9 - folic acid (µg/j)	
	R	P	R	P	R	P	R	P	R	P
M373.2736T12.81	0.20	0.01	-0.23	0.005	-0.27	0.001	0.24	0.003	0.20	0.01

<sup>1</sup>Tested clinico-biological parameters at blood draw: Weight (kg), height (cm), body mass index (kg/m<sup>2</sup>) and blood rates of cholesterol (mmol/L), triglycerides (mmol/L), glucose (mmol/L), apo-lipoprotein A1 (g/L), apo-lipoprotein B (g/L), lipoprotein A (g/L), serum ferritin (µg/L), selenium (µmol/l), zinc (µmol/l), retinol (µmol/L), alpha-tocopherol (µmol/L), beta-carotene (µmol/L), vitamin C (µg/mL), low density lipoprotein cholesterol (mmol/L), high density lipoprotein cholesterol (mmol/L)

<sup>2</sup>Tested mean of daily nutrients intakes during the 2 years following the blood draw: energy intake without alcoholic component (Kcal), proteins (g/j), vegetal proteins (g/j), animal proteins (g/j), lipids (g/j), saturated fatty acids (g/j), monounsaturated fatty acids (g/j), polyunsaturated fatty acids (g/j), linoleic acid 18:2 n-6 (g/j), alpha-linoleic acid 18:3 n-3 (g/j), arachidonic acid 20:4 n-6 (g/j), EPA 20:5 n-3 (g/j), DPA 22:5 n-3 (g/j), DHA 22:6 n-3 (g/j), cholesterol (mg/j), carbohydrates (g/j), simple carbohydrates (g/j), simple added sugar (g/j), starch (g/j), total fibres (g/j), soluble fibres (g/j), alcohol (g/j), Vit A - retinol (µg/j), beta-carotene (µg/j), Vit B1 - thiamine (mg/j), Vit B2 - riboflavin (mg/j), Vit B3 - niacin (mg/j), Vit B5 - pantothenic acid (mg/j), Vit B6 - pyridoxine (mg/j), Vit B9 – folic acid (µg/j), Vit B12 – cobalamin (µg/j), Vit C – ascorbic acid (mg/j), Vit D - calciferol (µg/j), Vit E - tocopherols (mg/j), calcium (mg/j), iron (mg/j), iodine (µg/j), potassium (mg/j), magnesium (mg/j), sodium (mg/j), phosphorus (mg/j), water (g/j)

<sup>3</sup>Partial spearman correlation adjusted for age (continuous), BMI (continuous), menopausal status (pre/post-menopause) and smoking status (current/former/never) at baseline, season at time of blood draw (a priori-defined periods: October/November, December/January/February, and March-April), number of 24h-dietary records and mean of daily energy intake (kcal) during the 2 years following the blood draw. Ions were considered as associated when correlation coefficient >|0.2|. Correlations with height, weight and BMI were not adjusted for BMI, correlation with energy intake without alcoholic component was not adjusted for energy intake but alcohol intake, correlation with alcohol intake was not adjusted for energy intake but energy intake without alcoholic component.

## Supplemental file 5

**Supplementary Table S4** Associations ( $p < 0.05$ ) between ions from endogenous protocol (“ENDO”) and from exogenous protocol (“EXO”) and the level of adherence to the French dietary recommendations, as measured by the a priori mPNNS-GS from multivariable conditional logistic regression<sup>1</sup> - SU.VI.MAX cohort, France (1994-2007)

Ions Mass (M). retention time (T) (classified by mass)	Protocol <sup>2</sup>	Mode of detection	OR <sup>3</sup> [95% CI]	P-trend	P-corrected <sup>4</sup>
M67.522T0.67	EXO	ESI Positive	0.55 [0.32-0.96]	0.04	0.8
M72.0806T1.15	EXO	ESI Positive	1.88 [1.01-3.49]	0.04	0.8
M73.0284T1.02	ENDO	ESI Positive	1.97 [1.16-3.36]	0.01	0.7
M84.0807T0.98	EXO	ESI Positive	2.33 [1.09-4.98]	0.03	0.8
<b>M84.0808T1.17</b> ( <b>Pipecolic acid</b> )	ENDO	ESI Positive	10.43 [2.23-48.73]	0.003	0.4
M76.0391T1.15	EXO	ESI Positive	2.01 [1.12-3.62]	0.02	0.8
<b>M76.0393T1.17</b>	ENDO	ESI Positive	3.18 [1.37-7.35]	0.007	0.6
M103.9201T11.8	EXO	ESI Negative	1.91 [1.05-3.46]	0.03	0.8
M108.9875T0.78	EXO	ESI Positive	0.39 [0.16-0.92]	0.03	0.8
M111.9683T0.62	EXO	ESI Positive	0.56 [0.31-0.98]	0.04	0.8
M118.0348T0.68	EXO	ESI Positive	0.53 [0.3-0.94]	0.03	0.8
M118.0496T1.17	ENDO	ESI Positive	2.16 [1.14-4.11]	0.02	0.7
<b>M130.0861T1.16</b> ( <b>Pipecolic acid</b> )	ENDO	ESI Positive	11.39 [2.39-54.32]	0.002	0.4
<b>M131.0895T1.16</b> ( <b>Pipecolic acid</b> )	ENDO	ESI Positive	13.12 [2.58-66.6]	0.002	0.4
M136.8774T0.94	ENDO	ESI Negative	1.72 [1.03-2.87]	0.04	0.9
M155.8665T0.99	EXO	ESI Negative	2.4 [1.11-5.23]	0.03	0.8
M158.9639T0.77	EXO	ESI Positive	0.54 [0.3-0.98]	0.04	0.8
M166.8603T0.87	EXO	ESI Negative	0.42 [0.21-0.84]	0.01	0.8
M167.021T1.19	ENDO	ESI Negative	0.46 [0.25-0.88]	0.02	0.9
M173.9261T11.8	EXO	ESI Negative	0.55 [0.3-0.99]	0.045	0.8
M174.956T15.9	EXO	ESI Negative	2.69 [1.26-5.74]	0.01	0.8
M176.1029T0.87	ENDO	ESI Positive	0.52 [0.28-0.98]	0.04	0.8
M187.0577T0.95 (2-deoxy-D-glucose)	EXO	ESI Positive	0.55 [0.3-0.99]	0.04	0.8
<b>M187.0577T0.97</b> ( <b>2-deoxy-D-glucose</b> )	ENDO	ESI Positive	0.36 [0.17-0.77]	0.009	0.6
<b>M189.0784T10.9</b>	EXO	ESI Positive	3.49 [1.4-8.67]	0.007	0.8
<b>M190.0859T10.9</b> ( <b>3-indolepropionic acid</b> )	EXO	ESI Positive	4.39 [1.69-11.42]	0.002	0.8
M192.0435T0.82	EXO	ESI Positive	2.12 [1.08-4.15]	0.03	0.8
M192.1382T12.2	EXO	ESI Positive	2.15 [1.04-4.43]	0.04	0.8
M198.0972T1.1	EXO	ESI Positive	2.06 [1.06-3.98]	0.03	0.8
M203.139T1.16	ENDO	ESI Positive	0.5 [0.26-0.97]	0.04	0.8
M223.1077T7.2	EXO	ESI Positive	0.45 [0.23-0.88]	0.02	0.8
M223.1078T7.19	ENDO	ESI Positive	0.45 [0.22-0.92]	0.03	0.8
M224.0621T24.8	ENDO	ESI Negative	0.31 [0.11-0.86]	0.02	0.9
M227.1251T0.87	EXO	ESI Positive	0.52 [0.28-0.96]	0.04	0.8
M229.1182T0.97	EXO	ESI Positive	0.5 [0.26-1]	0.049	0.8



## ANNEXE 6

M244.9081T14.2	EXO	ESI Negative	1.95 [1.09-3.48]	0.02	0.8
M248.9989T0.84	ENDO	ESI Positive	2.32 [1.06-5.05]	0.03	0.8
M249.1696T9.21	EXO	ESI Positive	2.9 [1.27-6.63]	0.01	0.8
<b>M252.0874T18.1</b>	EXO	ESI Positive	3.43 [1.38-8.53]	0.008	0.8
M261.0112T1.05	EXO	ESI Positive	0.42 [0.19-0.93]	0.03	0.8
M261.1439T7.47	EXO	ESI Positive	0.51 [0.26-1]	0.049	0.8
M262.1649T1.17	EXO	ESI Positive	2.3 [1.13-4.67]	0.02	0.8
M267.1965T15.6	EXO	ESI Negative	0.55 [0.3-1]	0.049	0.8
M269.2218T20.1	EXO	ESI Positive	0.49 [0.25-0.96]	0.04	0.8
M283.0075T8.89	EXO	ESI Positive	0.33 [0.13-0.85]	0.02	0.8
M283.1077T20.5	EXO	ESI Positive	0.47 [0.23-0.94]	0.03	0.8
M285.9886T0.81	ENDO	ESI Positive	2.23 [1.09-4.56]	0.03	0.8
M287.2325T18.0	EXO	ESI Positive	0.43 [0.19-0.99]	0.046	0.8
M289.0031T10.9	EXO	ESI Positive	2.2 [1.13-4.29]	0.02	0.8
M292.6377T13.8	EXO	ESI Positive	0.53 [0.29-0.97]	0.04	0.8
M299.6254T14.0	ENDO	ESI Positive	2.68 [1.19-6]	0.02	0.7
M301.1184T20.5	EXO	ESI Positive	0.54 [0.29-0.99]	0.04	0.8
M301.2173T17.2	EXO	ESI Negative	2.78 [1-7.73]	0.050	0.8
<b>M302.2325T10.4</b>					
<b>(2,6 Dimethylheptanoyl carnitine/ nonanoyl-l-carnitine)</b>	EXO	ESI Positive	0.35 [0.16-0.78]	0.01	0.8
M303.0272T0.66	EXO	ESI Positive	0.46 [0.24-0.87]	0.02	0.8
<b>M303.2357T10.4</b>					
<b>(2,6 dimethylheptanoyl carnitine/ nonanoyl-l-carnitine)</b>	EXO	ESI Positive	0.28 [0.11-0.7]	0.01	0.8
M304.2304T12.3	EXO	ESI Positive	0.53 [0.29-0.98]	0.04	0.8
M310.9129T0.78	EXO	ESI Positive	0.49 [0.25-0.98]	0.04	0.8
M311.2227T15.1	EXO	ESI Negative	1.88 [1.04-3.42]	0.04	0.8
M312.2261T15.1	EXO	ESI Negative	1.93 [1.02-3.67]	0.04	0.8
M318.1622T15.2	EXO	ESI Positive	0.47 [0.24-0.93]	0.03	0.8
<b>M325.1129T0.87</b>	ENDO	ESI Positive	0.29 [0.11-0.75]	0.01	0.7
<b>M330.2639T11.6</b>					
<b>(4,8 dimethylnonanoyl carnitine / undecanoylcarnitine)</b>	EXO	ESI Positive	0.37 [0.18-0.75]	0.01	0.8
M352.1654T9.04	EXO	ESI Positive	0.45 [0.22-0.93]	0.03	0.8
M353.1686T9.04	EXO	ESI Positive	0.4 [0.17-0.92]	0.03	0.8
M357.2787T14.6	EXO	ESI Positive	0.51 [0.27-0.95]	0.03	0.8
M359.1513T17.2	EXO	ESI Positive	2.15 [1.07-4.33]	0.03	0.8
M367.121T0.97	ENDO	ESI Positive	0.46 [0.25-0.87]	0.02	0.7
M369.174T13.26	EXO	ESI Negative	2.46 [1.04-5.8]	0.04	0.8
M379.158T18.17	ENDO	ESI Negative	2.27 [1.13-4.58]	0.02	0.9
M383.116T0.87	ENDO	ESI Positive	0.46 [0.23-0.94]	0.03	0.8
M384.0503T11.8	EXO	ESI Positive	0.47 [0.25-0.87]	0.02	0.8
M391.2853T14.6	EXO	ESI Negative	0.53 [0.3-0.96]	0.04	0.8
M392.2793T12.2	EXO	ESI Positive	2.19 [1.01-4.74]	0.047	0.8
M429.301T15.49	EXO	ESI Negative	0.48 [0.23-1]	0.0496	0.8
M452.2771T14.1	EXO	ESI Positive	0.45 [0.23-0.89]	0.02	0.8
M464.1913T10.3	EXO	ESI Positive	0.3 [0.11-0.85]	0.02	0.8
M479.2838T24.5	EXO	ESI Negative	2.01 [1.04-3.91]	0.04	0.8

## ANNEXE 6

M500.277T13.97	ENDO	ESI Positive	1.99 [1.15-3.43]	0.01	0.7
<b>M508.3395T15.2 (LysoPC (17:1))</b>	EXO	ESI Positive	0.23 [0.09-0.63]	0.004	0.8
<b>M510.4363T17.5</b>	ENDO	ESI Positive	0.33 [0.15-0.74]	0.006	0.6
M512.2686T14.6	EXO	ESI Negative	0.58 [0.33-0.99]	0.05	0.8
M529.4072T20.1	EXO	ESI Positive	0.39 [0.17-0.91]	0.03	0.8
M550.3868T17.2	EXO	ESI Positive	2.67 [1.08-6.57]	0.03	0.8
M567.2254T13.8	EXO	ESI Positive	0.35 [0.14-0.85]	0.02	0.8
M568.3397T14.6	ENDO	ESI Positive	1.84 [1-3.4]	0.05	0.8
M570.3553T15.3	EXO	ESI Positive	0.27 [0.09-0.81]	0.02	0.8
M583.2549T20.5	EXO	ESI Positive	0.5 [0.28-0.89]	0.02	0.8
M584.26T20.56	EXO	ESI Positive	0.51 [0.29-0.89]	0.02	0.8
M584.3569T15.2	EXO	ESI Negative	0.4 [0.2-0.83]	0.01	0.8
M585.2705T11.2	EXO	ESI Positive	0.51 [0.26-1]	0.048	0.8
M585.2706T9.57	EXO	ESI Positive	0.53 [0.29-0.98]	0.04	0.8
M586.2737T11.2	EXO	ESI Positive	0.51 [0.26-1]	0.0495	0.8
M598.2651T15.2	ENDO	ESI Positive	1.98 [1.02-3.84]	0.04	0.8
M616.3618T15.7	EXO	ESI Negative	0.52 [0.29-0.94]	0.03	0.8
M681.2546T15.2	EXO	ESI Negative	0.48 [0.26-0.9]	0.02	0.8
<b>M737.8604T0.77</b>	EXO	ESI Positive	3.58 [1.4-9.17]	0.008	0.8
M758.867T0.78	EXO	ESI Negative	2.81 [1.06-7.46]	0.04	0.8
M793.5145T0.77	EXO	ESI Positive	2.13 [1.01-4.47]	0.046	0.8
M918.885T0.77	EXO	ESI Negative	0.5 [0.26-0.97]	0.04	0.8
M972.1613T11.2	ENDO	ESI Positive	0.44 [0.23-0.85]	0.01	0.7
M1004.8108T0.7	ENDO	ESI Negative	2.17 [1.11-4.25]	0.02	0.9

SU.VI.MAX, Supplémentation en Vitamines et Minéraux Antioxydants; M, monoisotopic mass ; T, retention time ; OR, odds ratio; CI, confidence interval ; ENDO, endogenous protocol ; EXO, exogenous protocol

Associations presented in **Table 2** in the manuscript are highlighted in bold.

<sup>1</sup> Subjects from the 1<sup>st</sup> and 10<sup>th</sup> deciles of mPNNS-GS score were matched for age (35-39y/40-44y/45-49y/50-54y/55-59y/>60y), menopausal status (pre/post-menopause), BMI (<18.5kg/m<sup>2</sup>: underweight, ≥18.5 kg/m<sup>2</sup> to <25 kg/m<sup>2</sup>: normal weight and ≥25 kg/m<sup>2</sup>: overweight/obese), smoking status (current smokers and non/former smokers) and season (a priori-defined periods: October-November/December-January-February/March-April) at time of blood draw. Multivariable models were additionally adjusted for the following characteristics: age (continuous), BMI (continuous), smoking status (current/former smokers vs non-smokers), number of 24h-dietary records and mean of daily energy intake during the 2 years following the blood draw. Tests for linear trend were performed using continuous variables.

<sup>2</sup>160 samples for endogenous protocol and exogenous protocol in negative ionization mode and 152 samples for exogenous protocol in positive ionization mode were analyzed (due to conditional analyses, the pairs corresponding to the samples excluded during pre-processing (N=4) were not analyzed).

<sup>3</sup> OR<1: associated with low adherence to the French dietary recommendations (lower quality of the overall diet) OR>1: associated with high adherence to the French dietary recommendations (higher quality of the overall diet)

<sup>4</sup> P-corrected: P-trend after Benjamini Hochberg correction for multiple testing

## Supplemental file 6

**Supplementary Table S5** Associations ( $p < 0.05$ ) between ions from endogenous protocol (“ENDO”) and from exogenous protocol (“EXO”) and components of the *a priori* mPNNS-GS, from spearman correlation analyses<sup>1</sup>, SU.VI.MAX cohort, France (1994-2007)

Ions Mass (M). retention time (T) (classified by mass)	Protocol <sup>2</sup>	Mode of detection	R	P-trend	P-corrected <sup>3</sup>	mPNNS-GS Component
M57.0697T9.73	EXO	ESI Positive	-0.18	0.03	0.9	Alcohol
M78.9187T0.9	EXO	ESI Negative	-0.27	0.0008	0.1	Alcohol
M98.9842T1.08	ENDO	ESI Positive	-0.19	0.02	0.97	Alcohol
M100.0757T1.17	ENDO	ESI Positive	-0.19	0.02	0.97	Alcohol
M100.0758T18.11	EXO	ESI Positive	0.16	0.048	0.9	Alcohol
M100.0759T8.78	EXO	ESI Positive	0.18	0.02	0.9	Alcohol
M102.0223T0.68	EXO	ESI Positive	0.18	0.03	0.9	Alcohol
M104.1072T15.22	EXO	ESI Positive	-0.17	0.04	0.9	Alcohol
M105.0038T0.68	EXO	ESI Positive	0.20	0.02	0.9	Alcohol
M105.0337T8.29	ENDO	ESI Positive	-0.21	0.009	0.97	Alcohol
M110.899T0.87	EXO	ESI Negative	0.20	0.01	0.7	Alcohol
M120.0806T2.6	ENDO	ESI Positive	-0.19	0.02	0.97	Alcohol
M121.9662T18.9	EXO	ESI Positive	0.16	0.047	0.9	Alcohol
M124.0242T0.67	ENDO	ESI Positive	-0.17	0.04	0.97	Alcohol
M124.9648T14.06	EXO	ESI Positive	-0.18	0.03	0.9	Alcohol
M129.0657T0.91	ENDO	ESI Positive	-0.17	0.04	0.97	Alcohol
M130.0861T1.16 (Pipelicolic acid)	ENDO	ESI Positive	-0.17	0.03	0.97	Alcohol
M130.0862T0.72	ENDO	ESI Positive	-0.22	0.005	0.97	Alcohol
M132.1018T2.06	EXO	ESI Positive	0.19	0.02	0.9	Alcohol
M133.0658T11.51	EXO	ESI Negative	0.18	0.03	0.96	Alcohol
<b>M136.8774T0.94</b>	ENDO	ESI Negative	-0.32	0.0001	0.01	Alcohol
<b>M138.8751T0.92</b>	EXO	ESI Negative	-0.30	0.0001	0.05	Alcohol
M138.8751T0.94	ENDO	ESI Negative	-0.21	0.009	0.6	Alcohol
M140.0681T0.76	ENDO	ESI Positive	-0.19	0.02	0.97	Alcohol
M148.0603T0.87	ENDO	ESI Positive	0.16	0.04	0.97	Alcohol
M158.9785T0.7	EXO	ESI Negative	0.23	0.004	0.4	Alcohol
M160.1332T8.17	EXO	ESI Positive	-0.20	0.02	0.9	Alcohol
M163.0601T0.86	EXO	ESI Positive	0.21	0.01	0.9	Alcohol
M166.0862T2.77	ENDO	ESI Positive	-0.24	0.003	0.97	Alcohol
M166.0862T5.82	EXO	ESI Positive	0.27	0.0008	0.4	Alcohol
M176.972T0.51	EXO	ESI Positive	0.16	0.046	0.9	Alcohol
M180.0655T8.29	EXO	ESI Positive	-0.16	0.049	0.9	Alcohol
M182.0811T2.06	EXO	ESI Positive	0.20	0.01	0.9	Alcohol
M185.0304T1.17	ENDO	ESI Positive	0.16	0.047	0.97	Alcohol
M209.1899T13.82	EXO	ESI Positive	-0.22	0.007	0.8	Alcohol
M210.8422T0.85	EXO	ESI Negative	0.17	0.04	0.97	Alcohol
M223.1339T13.31	ENDO	ESI Negative	-0.17	0.04	0.96	Alcohol

## ANNEXE 6

M229.1547T1.18	ENDO	ESI Positive	-0.16	0.046	0.97	Alcohol
<b>M231.0839T1.16</b>						
<b>(Ethyl-<math>\beta</math>-D-glucopyranoside)</b>	EXO	ESI Positive	0.37	0.000003	0.003	Alcohol
M248.9989T0.84	ENDO	ESI Positive	-0.18	0.02	0.97	Alcohol
M260.1856T8.88	EXO	ESI Positive	0.18	0.03	0.9	Alcohol
M261.1439T7.47	EXO	ESI Positive	0.24	0.003	0.8	Alcohol
M262.9874T10.03	EXO	ESI Positive	-0.23	0.005	0.8	Alcohol
M266.0169T0.67	EXO	ESI Positive	0.17	0.03	0.9	Alcohol
M268.104T1.17	ENDO	ESI Positive	0.19	0.02	0.97	Alcohol
M268.1543T12.79	ENDO	ESI Positive	-0.18	0.03	0.97	Alcohol
M268.2184T20.19	EXO	ESI Positive	0.21	0.01	0.9	Alcohol
M269.2218T20.19	EXO	ESI Positive	0.19	0.02	0.9	Alcohol
M269.2263T18.08	EXO	ESI Positive	0.16	0.048	0.9	Alcohol
M282.1911T1.18	ENDO	ESI Positive	-0.16	0.048	0.97	Alcohol
M286.9064T18.9	EXO	ESI Positive	0.20	0.01	0.9	Alcohol
M287.1471T13.51	ENDO	ESI Positive	0.22	0.007	0.97	Alcohol
M287.147T13.5	EXO	ESI Positive	0.17	0.04	0.9	Alcohol
M287.2325T18.08	EXO	ESI Positive	0.16	0.0498	0.9	Alcohol
M288.2168T10.25	EXO	ESI Positive	0.20	0.02	0.9	Alcohol
M292.198T0.79	EXO	ESI Positive	-0.22	0.006	0.8	Alcohol
M297.2435T17.36	EXO	ESI Negative	0.25	0.002	0.2	Alcohol
M301.2173T17.26	EXO	ESI Negative	0.18	0.03	0.96	Alcohol
M304.2304T12.33	EXO	ESI Positive	0.17	0.03	0.9	Alcohol
M304.8953T0.75	ENDO	ESI Positive	0.19	0.02	0.97	Alcohol
M311.1687T11.85	EXO	ESI Negative	0.20	0.01	0.7	Alcohol
M318.1622T15.23	EXO	ESI Positive	-0.20	0.01	0.9	Alcohol
M329.2485T18.05	EXO	ESI Negative	0.18	0.03	0.96	Alcohol
M331.2265T18.08	EXO	ESI Positive	0.17	0.03	0.9	Alcohol
M337.064T1.17	ENDO	ESI Positive	0.16	0.045	0.97	Alcohol
M357.9843T1.17	ENDO	ESI Positive	0.22	0.007	0.97	Alcohol
M368.0188T13.27	EXO	ESI Positive	-0.18	0.03	0.9	Alcohol
M394.8907T0.75	ENDO	ESI Negative	0.18	0.02	0.96	Alcohol
M395.2203T17.69	EXO	ESI Negative	0.16	0.046	0.97	Alcohol
M402.9023T0.72	EXO	ESI Positive	0.22	0.007	0.8	Alcohol
M426.3577T14.21	EXO	ESI Positive	0.17	0.04	0.9	Alcohol
M452.7371T8.77	ENDO	ESI Positive	0.17	0.03	0.97	Alcohol
M458.3474T12.43	EXO	ESI Positive	-0.18	0.03	0.9	Alcohol
M464.1913T10.3	EXO	ESI Positive	0.23	0.006	0.8	Alcohol
M468.3083T15.8	EXO	ESI Positive	-0.19	0.02	0.9	Alcohol
M477.9086T0.73	EXO	ESI Positive	0.17	0.04	0.9	Alcohol
M482.3241T14.5	ENDO	ESI Positive	-0.17	0.03	0.97	Alcohol
M526.3149T14.5	EXO	ESI Negative	-0.23	0.005	0.4	Alcohol
M529.4072T20.19	EXO	ESI Positive	0.18	0.02	0.9	Alcohol
M540.366T15.45	EXO	ESI Positive	-0.17	0.04	0.9	Alcohol
M556.9917T0.87	ENDO	ESI Positive	-0.17	0.04	0.97	Alcohol
M583.2548T12.41	ENDO	ESI Positive	-0.18	0.03	0.97	Alcohol
M596.7595T14.67	EXO	ESI Positive	-0.18	0.03	0.9	Alcohol

## ANNEXE 6

M599.2619T14.67	EXO	ESI Positive	-0.17	0.04	0.9	Alcohol
M619.2569T14.67	EXO	ESI Positive	-0.22	0.008	0.8	Alcohol
M734.8122T0.79	EXO	ESI Positive	-0.18	0.03	0.9	Alcohol
M788.5795T24.78	EXO	ESI Negative	-0.20	0.02	0.7	Alcohol
M944.8925T0.77	ENDO	ESI Negative	-0.17	0.04	0.96	Alcohol
M952.1505T0.78	ENDO	ESI Positive	-0.16	0.049	0.97	Alcohol
M954.1525T11.32	ENDO	ESI Positive	0.17	0.04	0.97	Alcohol
M960.9475T0.75	ENDO	ESI Negative	-0.24	0.003	0.3	Alcohol
M972.1634T10.64	ENDO	ESI Positive	0.19	0.02	0.97	Alcohol
M1004.8108T0.78	ENDO	ESI Negative	-0.18	0.03	0.96	Alcohol
M58.0649T0.91	EXO	ESI Positive	-0.23	0.004	0.8	Bread, cereals, potatoes and legumes
M71.9517T0.69	EXO	ESI Positive	-0.21	0.01	0.97	Bread, cereals, potatoes and legumes
M72.0806T1.15	EXO	ESI Positive	0.17	0.04	0.998	Bread, cereals, potatoes and legumes
M76.0391T1.15	EXO	ESI Positive	-0.24	0.004	0.8	Bread, cereals, potatoes and legumes
M76.0755T0.86	EXO	ESI Positive	-0.18	0.02	0.998	Bread, cereals, potatoes and legumes
M79.9572T9.41	ENDO	ESI Negative	0.21	0.009	0.4	Bread, cereals, potatoes and legumes
M84.0444T0.85	ENDO	ESI Positive	-0.17	0.04	0.9	Bread, cereals, potatoes and legumes
M91.0541T1.17	ENDO	ESI Positive	0.16	0.04	0.9	Bread, cereals, potatoes and legumes
M105.0337T8.29	ENDO	ESI Positive	0.17	0.04	0.9	Bread, cereals, potatoes and legumes
M108.9018T0.87	ENDO	ESI Negative	0.19	0.02	0.4	Bread, cereals, potatoes and legumes
M110.0714T0.8	EXO	ESI Positive	0.19	0.02	0.998	Bread, cereals, potatoes and legumes
M114.0659T0.88	EXO	ESI Positive	-0.16	0.049	0.998	Bread, cereals, potatoes and legumes
M118.0497T1.15	EXO	ESI Positive	-0.18	0.03	0.998	Bread, cereals, potatoes and legumes
M120.0031T0.82	ENDO	ESI Positive	-0.19	0.02	0.9	Bread, cereals, potatoes and legumes
M123.0442T1.17	ENDO	ESI Positive	0.19	0.02	0.9	Bread, cereals, potatoes and legumes
M126.0218T0.85	ENDO	ESI Positive	-0.21	0.008	0.9	Bread, cereals, potatoes and legumes
M132.0654T0.85	ENDO	ESI Positive	-0.16	0.047	0.9	Bread, cereals, potatoes and legumes
M136.0479T0.89	ENDO	ESI Positive	-0.17	0.04	0.9	Bread, cereals, potatoes and legumes
M143.9589T15.11	EXO	ESI Positive	-0.17	0.04	0.998	Bread, cereals, potatoes and legumes
M146.9659T21.59	ENDO	ESI Negative	-0.18	0.03	0.4	Bread, cereals, potatoes and legumes
M148.0037T0.83	ENDO	ESI Positive	-0.17	0.04	0.9	Bread, cereals, potatoes and legumes
M156.0767T0.79	ENDO	ESI Positive	0.17	0.03	0.9	Bread, cereals, potatoes and legumes
M156.0767T0.79	EXO	ESI Positive	0.17	0.04	0.998	Bread, cereals, potatoes and legumes
M156.8776T0.87	ENDO	ESI Negative	0.20	0.01	0.4	Bread, cereals, potatoes and legumes
M161.0454T0.92	ENDO	ESI Negative	0.18	0.02	0.4	Bread, cereals, potatoes and legumes

## ANNEXE 6

M167.021T1.16	EXO	ESI Negative	-0.18	0.03	0.9	Bread, cereals, potatoes and legumes
M169.0583T0.84	ENDO	ESI Positive	-0.20	0.01	0.9	Bread, cereals, potatoes and legumes
M170.0924T0.81	ENDO	ESI Positive	-0.17	0.04	0.9	Bread, cereals, potatoes and legumes
M175.1189T0.95	EXO	ESI Positive	-0.17	0.04	0.998	Bread, cereals, potatoes and legumes
M184.9158T14.94	EXO	ESI Negative	0.23	0.005	0.9	Bread, cereals, potatoes and legumes
M188.128T0.97	EXO	ESI Positive	-0.18	0.02	0.998	Bread, cereals, potatoes and legumes
M189.1597T0.79	EXO	ESI Positive	-0.22	0.007	0.9	Bread, cereals, potatoes and legumes
M192.0435T0.82	EXO	ESI Positive	-0.19	0.02	0.998	Bread, cereals, potatoes and legumes
M199.1703T16.25	EXO	ESI Negative	-0.16	0.045	0.9	Bread, cereals, potatoes and legumes
M202.9977T1	EXO	ESI Negative	-0.21	0.01	0.9	Bread, cereals, potatoes and legumes
M209.092T5.61	EXO	ESI Positive	-0.21	0.01	0.97	Bread, cereals, potatoes and legumes
M223.1339T13.26	EXO	ESI Negative	-0.21	0.01	0.9	Bread, cereals, potatoes and legumes
M227.1251T0.87	EXO	ESI Positive	-0.16	0.0497	0.998	Bread, cereals, potatoes and legumes
M229.1412T9.33	EXO	ESI Positive	0.24	0.003	0.8	Bread, cereals, potatoes and legumes
M230.0957T0.97	EXO	ESI Positive	-0.18	0.02	0.998	Bread, cereals, potatoes and legumes
M251.1277T13.29	ENDO	ESI Positive	-0.17	0.04	0.9	Bread, cereals, potatoes and legumes
M265.1183T8.28	ENDO	ESI Positive	0.17	0.04	0.9	Bread, cereals, potatoes and legumes
M276.1264T8.82	ENDO	ESI Positive	0.20	0.01	0.9	Bread, cereals, potatoes and legumes
M280.1542T8.31	EXO	ESI Positive	0.26	0.001	0.8	Bread, cereals, potatoes and legumes
M280.6547T15.59	ENDO	ESI Positive	-0.21	0.01	0.9	Bread, cereals, potatoes and legumes
M288.2533T13.95	ENDO	ESI Positive	0.17	0.03	0.9	Bread, cereals, potatoes and legumes
M299.6254T14.04	ENDO	ESI Positive	-0.16	0.046	0.9	Bread, cereals, potatoes and legumes
M310.1553T15.6	EXO	ESI Positive	-0.19	0.02	0.998	Bread, cereals, potatoes and legumes
M327.1268T17.68	ENDO	ESI Negative	-0.16	0.04	0.6	Bread, cereals, potatoes and legumes
M329.2485T18.05	EXO	ESI Negative	-0.17	0.04	0.9	Bread, cereals, potatoes and legumes
M339.6042T15.6	EXO	ESI Positive	-0.19	0.02	0.998	Bread, cereals, potatoes and legumes
M341.1538T16.26	EXO	ESI Positive	-0.16	0.0496	0.998	Bread, cereals, potatoes and legumes
M341.3162T15.4	ENDO	ESI Positive	-0.20	0.01	0.9	Bread, cereals, potatoes and legumes
M343.3318T16.15	ENDO	ESI Positive	-0.17	0.03	0.9	Bread, cereals, potatoes and legumes
M349.2412T24.51	EXO	ESI Negative	0.16	0.04	0.9	Bread, cereals, potatoes and legumes
M357.9843T1.17	ENDO	ESI Positive	-0.20	0.01	0.9	Bread, cereals, potatoes and legumes
M358.2951T12.6	ENDO	ESI Positive	-0.16	0.048	0.9	Bread, cereals, potatoes and legumes
M369.174T13.26	EXO	ESI Negative	-0.18	0.02	0.9	Bread, cereals, potatoes and legumes

## ANNEXE 6

M391.9529T0.81	ENDO	ESI Positive	-0.16	0.048	0.9	Bread, cereals, potatoes and legumes
M412.7516T15.63	EXO	ESI Positive	-0.23	0.005	0.8	Bread, cereals, potatoes and legumes
M436.3989T20.2	EXO	ESI Positive	-0.21	0.009	0.97	Bread, cereals, potatoes and legumes
M478.2939T15.43	EXO	ESI Negative	-0.17	0.04	0.9	Bread, cereals, potatoes and legumes
M500.2771T13.94	EXO	ESI Positive	-0.17	0.04	0.998	Bread, cereals, potatoes and legumes
M500.2782T14.57	EXO	ESI Negative	-0.16	0.04	0.9	Bread, cereals, potatoes and legumes
M526.3512T15.65	ENDO	ESI Negative	0.20	0.01	0.4	Bread, cereals, potatoes and legumes
M529.4072T20.19	EXO	ESI Positive	-0.23	0.004	0.8	Bread, cereals, potatoes and legumes
M536.8913T17.4	ENDO	ESI Positive	0.17	0.03	0.9	Bread, cereals, potatoes and legumes
M538.3151T14.23	ENDO	ESI Negative	0.18	0.03	0.4	Bread, cereals, potatoes and legumes
M542.3241T14.04	ENDO	ESI Positive	-0.16	0.046	0.9	Bread, cereals, potatoes and legumes
M562.2996T14.68	EXO	ESI Positive	-0.22	0.008	0.9	Bread, cereals, potatoes and legumes
M562.3151T13.99	EXO	ESI Negative	-0.18	0.03	0.9	Bread, cereals, potatoes and legumes
M566.3463T15.62	EXO	ESI Negative	-0.18	0.02	0.9	Bread, cereals, potatoes and legumes
M568.3621T17.55	EXO	ESI Negative	-0.18	0.02	0.9	Bread, cereals, potatoes and legumes
M575.8765T0.79	ENDO	ESI Positive	-0.19	0.02	0.9	Bread, cereals, potatoes and legumes
M586.315T14.02	EXO	ESI Negative	-0.16	0.046	0.9	Bread, cereals, potatoes and legumes
M588.3306T14.68	EXO	ESI Negative	-0.16	0.046	0.9	Bread, cereals, potatoes and legumes
M590.8865T0.73	EXO	ESI Positive	0.18	0.03	0.998	Bread, cereals, potatoes and legumes
M595.2571T15.21	EXO	ESI Positive	-0.18	0.03	0.998	Bread, cereals, potatoes and legumes
M598.837T0.79	EXO	ESI Positive	0.18	0.03	0.998	Bread, cereals, potatoes and legumes
M610.3727T15.62	EXO	ESI Negative	-0.20	0.02	0.9	Bread, cereals, potatoes and legumes
M621.2727T15.59	ENDO	ESI Positive	-0.20	0.01	0.9	Bread, cereals, potatoes and legumes
M634.4523T16.51	EXO	ESI Positive	0.17	0.04	0.998	Bread, cereals, potatoes and legumes
M640.1179T11.32	ENDO	ESI Positive	0.18	0.03	0.9	Bread, cereals, potatoes and legumes
M647.4933T11.28	ENDO	ESI Positive	0.17	0.04	0.9	Bread, cereals, potatoes and legumes
M655.4154T15.26	EXO	ESI Positive	-0.20	0.01	0.998	Bread, cereals, potatoes and legumes
M656.4174T15.26	EXO	ESI Positive	-0.17	0.04	0.998	Bread, cereals, potatoes and legumes
M663.9023T0.81	ENDO	ESI Positive	-0.19	0.02	0.9	Bread, cereals, potatoes and legumes
M668.4225T15.59	EXO	ESI Positive	-0.17	0.04	0.998	Bread, cereals, potatoes and legumes
M674.9819T0.75	ENDO	ESI Negative	-0.19	0.02	0.4	Bread, cereals, potatoes and legumes
M680.2484T15.26	EXO	ESI Negative	0.18	0.02	0.9	Bread, cereals, potatoes and legumes
M691.5033T15.62	EXO	ESI Negative	-0.17	0.03	0.9	Bread, cereals, potatoes and legumes

## ANNEXE 6

M729.5478T11.33	ENDO	ESI Positive	0.17	0.03	0.9	Bread, cereals, potatoes and legumes
M736.2589T15.62	EXO	ESI Negative	-0.17	0.03	0.9	Bread, cereals, potatoes and legumes
M835.3435T0.78	ENDO	ESI Negative	-0.21	0.008	0.4	Bread, cereals, potatoes and legumes
M861.3502T0.78	ENDO	ESI Negative	0.21	0.009	0.4	Bread, cereals, potatoes and legumes
M870.9506T0.76	EXO	ESI Positive	-0.18	0.03	0.998	Bread, cereals, potatoes and legumes
M949.8087T0.78	EXO	ESI Positive	-0.21	0.01	0.97	Bread, cereals, potatoes and legumes
M960.9475T0.75	ENDO	ESI Negative	0.19	0.02	0.4	Bread, cereals, potatoes and legumes
M1002.4839T0.78	ENDO	ESI Negative	0.21	0.008	0.4	Bread, cereals, potatoes and legumes
M71.0489T0.87	EXO	ESI Positive	0.18	0.03	0.4	Fruit and vegetable
M73.0284T1.02	ENDO	ESI Positive	0.24	0.003	0.2	Fruit and vegetable
M84.0807T0.98	EXO	ESI Positive	0.19	0.02	0.4	Fruit and vegetable
M84.0808T1.17 (Pipelic acid)	ENDO	ESI Positive	0.27	0.0006	0.1	Fruit and vegetable
M85.0283T0.87	EXO	ESI Positive	-0.19	0.02	0.4	Fruit and vegetable
M85.0285T0.88	ENDO	ESI Positive	-0.18	0.02	0.6	Fruit and vegetable
M86.0601T1.18	ENDO	ESI Positive	0.20	0.01	0.4	Fruit and vegetable
M93.0115T0.68	EXO	ESI Positive	-0.19	0.02	0.4	Fruit and vegetable
M99.044T0.86	EXO	ESI Positive	-0.20	0.01	0.4	Fruit and vegetable
M100.9334T19.23	EXO	ESI Negative	-0.19	0.02	0.6	Fruit and vegetable
M101.0607T11.87	EXO	ESI Negative	0.17	0.04	0.7	Fruit and vegetable
M102.0559T0.87	EXO	ESI Negative	-0.16	0.047	0.7	Fruit and vegetable
M104.1071T14.68	ENDO	ESI Positive	0.17	0.04	0.7	Fruit and vegetable
M105.0337T8.29	ENDO	ESI Positive	0.22	0.007	0.3	Fruit and vegetable
M118.0348T0.68	EXO	ESI Positive	-0.22	0.008	0.3	Fruit and vegetable
M119.016T0.97	EXO	ESI Positive	0.21	0.009	0.3	Fruit and vegetable
M119.016T1.03 (Dihydro4mercapto-3(2H) furanone)	ENDO	ESI Positive	0.27	0.0008	0.1	Fruit and vegetable
M121.9662T18.9	EXO	ESI Positive	-0.23	0.005	0.3	Fruit and vegetable
M130.0861T1.16 (Pipelic acid)	ENDO	ESI Positive	0.29	0.0003	0.07	Fruit and vegetable
M130.0862T0.98	EXO	ESI Positive	0.18	0.03	0.4	Fruit and vegetable
<b>M131.0895T1.16 (Pipelic acid)</b>	ENDO	ESI Positive	0.31	0.0001	0.05	Fruit and vegetable
M131.9616T18.9	EXO	ESI Positive	-0.22	0.007	0.3	Fruit and vegetable
M132.0654T0.85	ENDO	ESI Positive	-0.16	0.046	0.7	Fruit and vegetable
M132.0766T0.92	ENDO	ESI Positive	-0.23	0.004	0.2	Fruit and vegetable
M132.0767T0.91	EXO	ESI Positive	-0.19	0.02	0.4	Fruit and vegetable
M137.0457T1.17	EXO	ESI Positive	-0.29	0.0003	0.1	Fruit and vegetable
M138.8751T0.92	EXO	ESI Negative	0.16	0.049	0.7	Fruit and vegetable
M144.1018T0.96	EXO	ESI Positive	0.16	0.047	0.5	Fruit and vegetable
M144.1018T0.99	ENDO	ESI Positive	0.21	0.008	0.3	Fruit and vegetable
M145.0494T0.87	EXO	ESI Positive	-0.21	0.009	0.3	Fruit and vegetable
M146.0599T9.95	ENDO	ESI Positive	0.20	0.01	0.4	Fruit and vegetable
M146.1175T0.91	EXO	ESI Positive	-0.17	0.04	0.5	Fruit and vegetable
<b>M149.0596T5.81</b>	EXO	ESI Positive	-0.30	0.0002	0.1	Fruit and vegetable



## ANNEXE 6

M154.0585T0.9	EXO	ESI Positive	-0.19	0.02	0.4	Fruit and vegetable
M157.0118T23.44	EXO	ESI Negative	0.16	0.046	0.7	Fruit and vegetable
M157.087T17.71	EXO	ESI Negative	-0.18	0.03	0.6	Fruit and vegetable
M159.0277T1.16	EXO	ESI Positive	-0.18	0.03	0.4	Fruit and vegetable
M160.1331T0.97	EXO	ESI Positive	-0.18	0.03	0.4	Fruit and vegetable
M169.9858T0.83	ENDO	ESI Positive	-0.17	0.03	0.7	Fruit and vegetable
M171.0431T0.67	EXO	ESI Positive	-0.20	0.02	0.4	Fruit and vegetable
M175.1189T0.95	EXO	ESI Positive	-0.18	0.03	0.4	Fruit and vegetable
M176.8602T0.98	EXO	ESI Negative	-0.19	0.02	0.6	Fruit and vegetable
M182.0811T1.17	EXO	ESI Positive	-0.17	0.04	0.5	Fruit and vegetable
M182.0811T1.18	ENDO	ESI Positive	-0.16	0.046	0.7	Fruit and vegetable
M182.0811T2.06	EXO	ESI Positive	-0.27	0.0008	0.1	Fruit and vegetable
M182.0811T2.09	ENDO	ESI Positive	-0.18	0.02	0.6	Fruit and vegetable
M184.9855T18.9	EXO	ESI Positive	-0.19	0.02	0.4	Fruit and vegetable
M187.0577T0.95	EXO	ESI Positive	-0.22	0.008	0.3	Fruit and vegetable
M187.0577T0.97 (2-deoxy-D-glucose)	ENDO	ESI Positive	-0.24	0.002	0.2	Fruit and vegetable
M189.0784T10.98	EXO	ESI Positive	0.20	0.02	0.4	Fruit and vegetable
M190.0859T10.97 (3-indolepropionic acid)	EXO	ESI Positive	0.22	0.007	0.3	Fruit and vegetable
M199.851T20.43	EXO	ESI Negative	0.21	0.008	0.6	Fruit and vegetable
M202.858T20.4	EXO	ESI Negative	0.16	0.049	0.7	Fruit and vegetable
M203.1502T0.91	ENDO	ESI Positive	-0.16	0.047	0.7	Fruit and vegetable
M214.8362T0.85	ENDO	ESI Negative	-0.16	0.046	0.9	Fruit and vegetable
M219.0265T0.86	ENDO	ESI Positive	-0.16	0.048	0.7	Fruit and vegetable
M226.045T1.17	EXO	ESI Positive	-0.19	0.02	0.4	Fruit and vegetable
M227.1641T9.54	EXO	ESI Positive	0.19	0.02	0.4	Fruit and vegetable
M233.9467T11.85	EXO	ESI Negative	-0.18	0.03	0.6	Fruit and vegetable
M244.9081T14.25	EXO	ESI Negative	0.21	0.01	0.6	Fruit and vegetable
M245.0162T0.95	EXO	ESI Positive	-0.21	0.01	0.3	Fruit and vegetable
M246.1699T7.86	EXO	ESI Positive	-0.18	0.03	0.4	Fruit and vegetable
M246.9051T14.25	EXO	ESI Negative	0.25	0.002	0.6	Fruit and vegetable
M249.1557T0.9	EXO	ESI Positive	-0.17	0.04	0.5	Fruit and vegetable
M260.1856T8.88	EXO	ESI Positive	-0.26	0.002	0.2	Fruit and vegetable
M261.0112T1.05	EXO	ESI Positive	-0.21	0.01	0.3	Fruit and vegetable
M263.1389T8.18	EXO	ESI Positive	-0.16	0.048	0.6	Fruit and vegetable
M271.0137T8.24	EXO	ESI Positive	-0.19	0.02	0.4	Fruit and vegetable
M276.1991T11.45	EXO	ESI Positive	-0.18	0.03	0.4	Fruit and vegetable
M281.2486T18.96	EXO	ESI Negative	-0.19	0.02	0.6	Fruit and vegetable
M285.0055T8.89	EXO	ESI Positive	-0.17	0.04	0.5	Fruit and vegetable
M287.147T13.5	EXO	ESI Positive	-0.17	0.04	0.5	Fruit and vegetable
M287.9713T24.49	EXO	ESI Negative	0.19	0.02	0.6	Fruit and vegetable
M288.2168T10.25	EXO	ESI Positive	-0.25	0.002	0.2	Fruit and vegetable
M289.0031T10.98	EXO	ESI Positive	0.19	0.02	0.4	Fruit and vegetable
M299.6254T14.04	ENDO	ESI Positive	0.16	0.047	0.7	Fruit and vegetable
M302.2147T11.99	EXO	ESI Positive	-0.22	0.008	0.3	Fruit and vegetable
M302.2325T10.49	EXO	ESI Positive	-0.26	0.001	0.2	Fruit and vegetable

## ANNEXE 6

(2,6 dimethylheptanoyl carnitine/ nonanoyl-l-carnitine; acylcarnitine C9:1) <b>M303.2357T10.49</b>	EXO	ESI Positive	-0.31	0.0001	0.1	Fruit and vegetable
<b>(2,6 dimethylheptanoyl carnitine/ nonanoyl-l-carnitine; acylcarnitine C9:1)</b>						
M304.2117T8.94	EXO	ESI Positive	-0.20	0.01	0.4	Fruit and vegetable
M304.2304T12.33	EXO	ESI Positive	-0.24	0.003	0.3	Fruit and vegetable
M309.207T14.52	EXO	ESI Negative	0.20	0.01	0.6	Fruit and vegetable
M310.9129T0.78	EXO	ESI Positive	-0.20	0.02	0.4	Fruit and vegetable
M312.8361T20.41	EXO	ESI Negative	0.18	0.03	0.6	Fruit and vegetable
M312.836T20.9	EXO	ESI Negative	0.18	0.03	0.6	Fruit and vegetable
M313.1546T8.83	EXO	ESI Positive	-0.17	0.04	0.5	Fruit and vegetable
M314.2325T10.91	EXO	ESI Positive	-0.18	0.03	0.4	Fruit and vegetable
M316.2483T11.33	EXO	ESI Positive	-0.19	0.02	0.4	Fruit and vegetable
M320.9669T0.86	ENDO	ESI Positive	-0.17	0.04	0.7	Fruit and vegetable
M325.1129T0.86	EXO	ESI Positive	-0.18	0.03	0.4	Fruit and vegetable
M325.1129T0.87	ENDO	ESI Positive	-0.26	0.001	0.1	Fruit and vegetable
M330.246T12.7	EXO	ESI Positive	-0.19	0.02	0.4	Fruit and vegetable
M330.2639T11.6						
<b>(4,8 dimethylnonanoyl carnitine / undecanoylcarnitine; acylcarnitine C11:1)</b>	EXO	ESI Positive	-0.28	0.0004	0.1	Fruit and vegetable
M332.2616T13.19	EXO	ESI Positive	-0.17	0.04	0.5	Fruit and vegetable
M342.2637T11.93	EXO	ESI Positive	-0.21	0.009	0.3	Fruit and vegetable
M343.267T11.79	EXO	ESI Positive	-0.16	0.049	0.6	Fruit and vegetable
M344.2794T12.27	EXO	ESI Positive	-0.26	0.001	0.2	Fruit and vegetable
M352.1654T9.04	EXO	ESI Positive	-0.19	0.02	0.4	Fruit and vegetable
M353.1686T9.04	EXO	ESI Positive	-0.20	0.01	0.4	Fruit and vegetable
M356.2794T12.22	EXO	ESI Positive	-0.17	0.04	0.5	Fruit and vegetable
M357.2787T14.6	EXO	ESI Positive	-0.23	0.005	0.3	Fruit and vegetable
M358.0349T11.87	ENDO	ESI Positive	0.17	0.04	0.7	Fruit and vegetable
M358.2951T12.6	ENDO	ESI Positive	0.19	0.02	0.6	Fruit and vegetable
M367.121T0.97	ENDO	ESI Positive	-0.21	0.008	0.3	Fruit and vegetable
M370.2951T12.69	EXO	ESI Positive	-0.23	0.004	0.3	Fruit and vegetable
M372.3107T13.16	EXO	ESI Positive	-0.23	0.005	0.3	Fruit and vegetable
M373.314T13.16	EXO	ESI Positive	-0.24	0.004	0.3	Fruit and vegetable
M383.116T0.87	ENDO	ESI Positive	-0.26	0.001	0.1	Fruit and vegetable
M386.2899T11.84	EXO	ESI Positive	-0.18	0.03	0.4	Fruit and vegetable
M390.0982T16.3	ENDO	ESI Negative	-0.16	0.049	0.9	Fruit and vegetable
M391.2853T14.6	EXO	ESI Negative	-0.17	0.04	0.7	Fruit and vegetable
M398.3264T13.43	EXO	ESI Positive	-0.17	0.03	0.5	Fruit and vegetable
M400.342T14.01	EXO	ESI Positive	-0.23	0.005	0.3	Fruit and vegetable
M401.3454T14.01	EXO	ESI Positive	-0.20	0.02	0.4	Fruit and vegetable
M425.3635T19.48	EXO	ESI Negative	-0.19	0.02	0.6	Fruit and vegetable
M426.3577T14.21	EXO	ESI Positive	-0.20	0.01	0.4	Fruit and vegetable
M427.3611T14.21	EXO	ESI Positive	-0.23	0.005	0.3	Fruit and vegetable
M428.3734T14.81	EXO	ESI Positive	-0.20	0.01	0.4	Fruit and vegetable

## ANNEXE 6

M429.3767T14.81	EXO	ESI Positive	-0.18	0.03	0.4	Fruit and vegetable
M436.3989T20.2	EXO	ESI Positive	-0.18	0.03	0.4	Fruit and vegetable
M454.389T14.93	EXO	ESI Positive	-0.17	0.04	0.5	Fruit and vegetable
M464.1913T10.3	EXO	ESI Positive	-0.18	0.03	0.4	Fruit and vegetable
M471.7373T8.73	ENDO	ESI Positive	-0.19	0.02	0.6	Fruit and vegetable
M471.7378T8.73	EXO	ESI Positive	-0.27	0.0009	0.1	Fruit and vegetable
M479.2838T24.53	EXO	ESI Negative	0.18	0.02	0.6	Fruit and vegetable
M490.2903T13.82	EXO	ESI Positive	-0.17	0.04	0.5	Fruit and vegetable
M498.9301T19.76	EXO	ESI Negative	0.19	0.02	0.6	Fruit and vegetable
M500.277T13.97	ENDO	ESI Positive	0.25	0.002	0.2	Fruit and vegetable
M506.7569T8.32	ENDO	ESI Positive	-0.16	0.046	0.7	Fruit and vegetable
M508.3395T15.27 (LysoPC (17:1))	EXO	ESI Positive	-0.21	0.01	0.3	Fruit and vegetable
M510.4363T17.54	ENDO	ESI Positive	-0.22	0.007	0.3	Fruit and vegetable
M529.4072T20.19	EXO	ESI Positive	-0.20	0.01	0.4	Fruit and vegetable
M548.3709T15.96	ENDO	ESI Positive	0.18	0.02	0.6	Fruit and vegetable
M566.3366T14.69	ENDO	ESI Negative	0.20	0.01	0.9	Fruit and vegetable
M598.2651T15.23	EXO	ESI Positive	0.18	0.02	0.4	Fruit and vegetable
M619.2889T14.68	EXO	ESI Negative	0.23	0.005	0.6	Fruit and vegetable
M652.8T11.28	ENDO	ESI Positive	-0.17	0.03	0.7	Fruit and vegetable
M676.3054T15.26	EXO	ESI Negative	0.16	0.04	0.7	Fruit and vegetable
M681.094T0.86	EXO	ESI Positive	-0.28	0.0007	0.1	Fruit and vegetable
M711.2696T15.62	EXO	ESI Negative	-0.18	0.03	0.6	Fruit and vegetable
M711.353T0.78	EXO	ESI Positive	-0.21	0.009	0.3	Fruit and vegetable
M759.1386T11.3	ENDO	ESI Positive	-0.21	0.009	0.3	Fruit and vegetable
M764.6931T11.3	ENDO	ESI Positive	-0.16	0.045	0.7	Fruit and vegetable
M798.5292T24.78	EXO	ESI Negative	0.18	0.02	0.6	Fruit and vegetable
M870.9506T0.76	EXO	ESI Positive	-0.20	0.02	0.4	Fruit and vegetable
M888.889T0.76	EXO	ESI Positive	-0.16	0.04	0.5	Fruit and vegetable
M915.3152T0.78	EXO	ESI Positive	-0.22	0.007	0.3	Fruit and vegetable
M1004.8108T0.78	ENDO	ESI Negative	0.16	0.047	0.9	Fruit and vegetable
M53.0184T0.68	EXO	ESI Positive	-0.16	0.049	0.99	Meat, poultry, seafood and eggs
M57.0697T9.73	EXO	ESI Positive	-0.21	0.009	0.96	Meat, poultry, seafood and eggs
M62.0428T0.9	EXO	ESI Positive	-0.17	0.04	0.99	Meat, poultry, seafood and eggs
M67.0542T16.82	ENDO	ESI Positive	0.18	0.03	0.5	Meat, poultry, seafood and eggs
M74.9373T15.06	EXO	ESI Positive	-0.19	0.02	0.99	Meat, poultry, seafood and eggs
M80.9166T0.9	EXO	ESI Negative	0.18	0.03	0.9	Meat, poultry, seafood and eggs
M104.9537T1.18	ENDO	ESI Negative	-0.21	0.01	0.998	Meat, poultry, seafood and eggs
M118.0649T7.39	EXO	ESI Positive	-0.17	0.04	0.99	Meat, poultry, seafood and eggs
M120.0806T2.6	ENDO	ESI Positive	0.18	0.03	0.5	Meat, poultry, seafood and eggs
M129.0657T0.91	ENDO	ESI Positive	-0.24	0.003	0.3	Meat, poultry, seafood and eggs
M132.0654T0.85	ENDO	ESI Positive	0.17	0.04	0.5	Meat, poultry, seafood and eggs

## ANNEXE 6

M134.019T0.83	ENDO	ESI Positive	0.21	0.01	0.3	Meat, poultry, seafood and eggs
M140.0681T18.11	EXO	ESI Positive	-0.19	0.02	0.99	Meat, poultry, seafood and eggs
M155.1065T9.03	ENDO	ESI Positive	0.17	0.04	0.5	Meat, poultry, seafood and eggs
M156.0767T0.79	ENDO	ESI Positive	-0.17	0.04	0.5	Meat, poultry, seafood and eggs
M159.0927T7.39	EXO	ESI Negative	-0.20	0.01	0.7	Meat, poultry, seafood and eggs
M160.1331T1.12	ENDO	ESI Positive	-0.17	0.04	0.5	Meat, poultry, seafood and eggs
M164.9276T0.68	ENDO	ESI Negative	-0.17	0.04	0.998	Meat, poultry, seafood and eggs
M177.0611T0.75	EXO	ESI Positive	0.18	0.03	0.99	Meat, poultry, seafood and eggs
M200.1644T10.82	ENDO	ESI Positive	0.17	0.04	0.5	Meat, poultry, seafood and eggs
M203.1502T0.89	EXO	ESI Positive	-0.26	0.001	0.6	Meat, poultry, seafood and eggs
M205.1151T7.38	EXO	ESI Positive	0.22	0.008	0.96	Meat, poultry, seafood and eggs
M216.1593T16.62	ENDO	ESI Positive	0.16	0.047	0.5	Meat, poultry, seafood and eggs
M218.9196T0.82	ENDO	ESI Positive	-0.22	0.007	0.3	Meat, poultry, seafood and eggs
M223.1339T13.31	ENDO	ESI Negative	0.17	0.04	0.998	Meat, poultry, seafood and eggs
M230.0957T0.97	EXO	ESI Positive	0.19	0.02	0.99	Meat, poultry, seafood and eggs
M230.0958T1.03	ENDO	ESI Positive	0.23	0.005	0.3	Meat, poultry, seafood and eggs
M242.5618T1.11	ENDO	ESI Positive	-0.18	0.03	0.5	Meat, poultry, seafood and eggs
M268.1754T1.16	ENDO	ESI Positive	0.17	0.03	0.5	Meat, poultry, seafood and eggs
M269.0879T1.17	EXO	ESI Positive	-0.20	0.02	0.99	Meat, poultry, seafood and eggs
M269.1383T13.29	ENDO	ESI Positive	0.20	0.01	0.3	Meat, poultry, seafood and eggs
M271.2054T12.85	EXO	ESI Positive	-0.16	0.04	0.99	Meat, poultry, seafood and eggs
M276.119T0.95	ENDO	ESI Positive	-0.17	0.04	0.5	Meat, poultry, seafood and eggs
M285.9886T0.81	ENDO	ESI Positive	0.24	0.003	0.3	Meat, poultry, seafood and eggs
M286.1437T13.51	ENDO	ESI Positive	0.20	0.01	0.3	Meat, poultry, seafood and eggs
M286.1437T13.64 (Piperine)	EXO	ESI Positive	0.16	0.04	0.99	Meat, poultry, seafood and eggs
M287.147T13.5	EXO	ESI Positive	0.17	0.04	0.99	Meat, poultry, seafood and eggs
M287.1606T7.61	ENDO	ESI Positive	0.18	0.02	0.5	Meat, poultry, seafood and eggs
M287.1609T7.61	EXO	ESI Positive	0.17	0.04	0.99	Meat, poultry, seafood and eggs
M289.0031T10.98	EXO	ESI Positive	-0.17	0.04	0.99	Meat, poultry, seafood and eggs
M291.6468T14.7	ENDO	ESI Positive	0.16	0.04	0.5	Meat, poultry, seafood and eggs
M292.2268T13.66	ENDO	ESI Positive	0.16	0.047	0.5	Meat, poultry, seafood and eggs
M292.6353T11.22	EXO	ESI Positive	-0.17	0.04	0.99	Meat, poultry, seafood and eggs
M294.2427T14.76	ENDO	ESI Positive	0.19	0.02	0.4	Meat, poultry, seafood and eggs

## ANNEXE 6

M299.6254T14.04	ENDO	ESI Positive	0.23	0.004	0.3	Meat, poultry, seafood and eggs
M301.1184T20.57	EXO	ESI Positive	-0.16	0.0499	0.99	Meat, poultry, seafood and eggs
M301.2173T17.26	EXO	ESI Negative	0.16	0.046	0.9	Meat, poultry, seafood and eggs
M302.2325T10.49 (2,6 dimethylheptanoyl carnitine/ nonanoyl-L-carnitine; acylcarnitine C9:1)	EXO	ESI Positive	-0.19	0.02	0.99	Meat, poultry, seafood and eggs
M303.2357T10.49 (2,6 dimethylheptanoyl carnitine/ nonanoyl-L-carnitine; acylcarnitine C9:1)	EXO	ESI Positive	-0.20	0.02	0.99	Meat, poultry, seafood and eggs
M304.2117T8.94	EXO	ESI Positive	-0.18	0.03	0.99	Meat, poultry, seafood and eggs
M311.2227T15.13	EXO	ESI Negative	-0.16	0.048	0.9	Meat, poultry, seafood and eggs
M312.6331T14.64	ENDO	ESI Positive	0.22	0.006	0.3	Meat, poultry, seafood and eggs
M313.16T15.97	ENDO	ESI Positive	0.18	0.03	0.5	Meat, poultry, seafood and eggs
M313.2383T15.84	EXO	ESI Negative	-0.20	0.01	0.6	Meat, poultry, seafood and eggs
M319.2435T15.4	ENDO	ESI Positive	0.18	0.03	0.5	Meat, poultry, seafood and eggs
M320.256T14.81	ENDO	ESI Positive	0.18	0.03	0.5	Meat, poultry, seafood and eggs
M327.2329T17.75	ENDO	ESI Negative	0.16	0.046	0.998	Meat, poultry, seafood and eggs
M332.243T10.23	EXO	ESI Positive	-0.17	0.04	0.99	Meat, poultry, seafood and eggs
M340.2481T11.51	EXO	ESI Positive	-0.19	0.02	0.99	Meat, poultry, seafood and eggs
<b>M342.2637T11.93</b>	EXO	ESI Positive	-0.31	0.0001	0.1	Meat, poultry, seafood and eggs
M355.2818T18.65	EXO	ESI Positive	0.16	0.0499	0.99	Meat, poultry, seafood and eggs
M368.9764T16	EXO	ESI Negative	0.22	0.006	0.5	Meat, poultry, seafood and eggs
M372.3107T13.2	ENDO	ESI Positive	0.17	0.04	0.5	Meat, poultry, seafood and eggs
M373.314T13.16	EXO	ESI Positive	-0.17	0.04	0.99	Meat, poultry, seafood and eggs
M376.2115T16.82	ENDO	ESI Positive	0.19	0.02	0.4	Meat, poultry, seafood and eggs
M394.8907T0.75	ENDO	ESI Negative	0.16	0.04	0.998	Meat, poultry, seafood and eggs
M396.3107T12.94	EXO	ESI Positive	-0.18	0.03	0.99	Meat, poultry, seafood and eggs
M398.3263T13.47	ENDO	ESI Positive	0.17	0.04	0.5	Meat, poultry, seafood and eggs
M400.7642T16.71	EXO	ESI Positive	-0.19	0.02	0.99	Meat, poultry, seafood and eggs
M407.7483T16.75	EXO	ESI Positive	-0.22	0.007	0.96	Meat, poultry, seafood and eggs
M411.308T19.03	EXO	ESI Positive	0.22	0.007	0.96	Meat, poultry, seafood and eggs
M438.2978T15.51	ENDO	ESI Positive	0.19	0.02	0.4	Meat, poultry, seafood and eggs
M450.3213T12.95	ENDO	ESI Positive	0.20	0.01	0.3	Meat, poultry, seafood and eggs
M452.7371T8.77	ENDO	ESI Positive	0.16	0.0498	0.5	Meat, poultry, seafood and eggs
M456.3318T12.03	EXO	ESI Positive	-0.18	0.03	0.99	Meat, poultry, seafood and eggs

## ANNEXE 6

M479.2838T24.53	EXO	ESI Negative	-0.16	0.048	0.9	Meat, poultry, seafood and eggs
M487.3741T11.48	EXO	ESI Positive	0.19	0.02	0.99	Meat, poultry, seafood and eggs
M498.2895T13.25	EXO	ESI Negative	0.18	0.03	0.9	Meat, poultry, seafood and eggs
M499.3603T19.03	EXO	ESI Positive	0.22	0.007	0.96	Meat, poultry, seafood and eggs
M500.277T13.97	ENDO	ESI Positive	0.21	0.008	0.3	Meat, poultry, seafood and eggs
M502.2938T15.05	EXO	ESI Negative	-0.22	0.008	0.5	Meat, poultry, seafood and eggs
M504.3085T15.06	EXO	ESI Positive	-0.20	0.02	0.99	Meat, poultry, seafood and eggs
M512.2686T14.67	EXO	ESI Negative	-0.17	0.04	0.9	Meat, poultry, seafood and eggs
M516.316T15.2	ENDO	ESI Positive	0.16	0.046	0.5	Meat, poultry, seafood and eggs
M528.3082T14.9	ENDO	ESI Positive	0.18	0.03	0.5	Meat, poultry, seafood and eggs
M541.3287T15.59	ENDO	ESI Positive	0.18	0.03	0.5	Meat, poultry, seafood and eggs
M542.3241T14.04	ENDO	ESI Positive	0.24	0.003	0.3	Meat, poultry, seafood and eggs
M542.324T14.02	EXO	ESI Positive	0.19	0.02	0.99	Meat, poultry, seafood and eggs
M548.8124T15.59	ENDO	ESI Positive	0.22	0.007	0.3	Meat, poultry, seafood and eggs
M549.3139T15.59	ENDO	ESI Positive	0.23	0.004	0.3	Meat, poultry, seafood and eggs
M550.3866T17.13	ENDO	ESI Positive	0.20	0.01	0.3	Meat, poultry, seafood and eggs
M558.796T14.69	ENDO	ESI Positive	0.17	0.04	0.5	Meat, poultry, seafood and eggs
M560.2993T14.69	ENDO	ESI Positive	0.17	0.03	0.5	Meat, poultry, seafood and eggs
M562.3267T16.82	ENDO	ESI Positive	0.16	0.04	0.5	Meat, poultry, seafood and eggs
M566.3217T14.7	ENDO	ESI Positive	0.17	0.03	0.5	Meat, poultry, seafood and eggs
M568.3397T14.63	EXO	ESI Positive	0.19	0.02	0.99	Meat, poultry, seafood and eggs
M568.3397T14.65	ENDO	ESI Positive	0.20	0.01	0.3	Meat, poultry, seafood and eggs
M570.3554T14.97	EXO	ESI Positive	0.17	0.04	0.99	Meat, poultry, seafood and eggs
M572.3709T15.72	ENDO	ESI Positive	0.21	0.009	0.3	Meat, poultry, seafood and eggs
M572.7596T15.22	EXO	ESI Positive	-0.17	0.04	0.99	Meat, poultry, seafood and eggs
M575.8765T0.79	ENDO	ESI Positive	0.23	0.004	0.3	Meat, poultry, seafood and eggs
M585.2702T12.45	EXO	ESI Positive	-0.22	0.007	0.96	Meat, poultry, seafood and eggs
M586.315T14.02	EXO	ESI Negative	0.22	0.006	0.5	Meat, poultry, seafood and eggs
M595.9149T0.81	ENDO	ESI Positive	0.22	0.005	0.3	Meat, poultry, seafood and eggs
M612.3305T14.64	EXO	ESI Negative	0.25	0.002	0.3	Meat, poultry, seafood and eggs
M614.3462T14.98	EXO	ESI Negative	0.17	0.04	0.9	Meat, poultry, seafood and eggs
M615.3955T0.81	ENDO	ESI Positive	0.21	0.009	0.3	Meat, poultry, seafood and eggs
M624.2807T15.59	ENDO	ESI Positive	0.16	0.046	0.5	Meat, poultry, seafood and eggs

## ANNEXE 6

M624.2922T15.24	EXO	ESI Negative	0.26	0.001	0.3	Meat, poultry, seafood and eggs
M626.2964T16.91	EXO	ESI Positive	0.17	0.04	0.99	Meat, poultry, seafood and eggs
M643.8639T0.78	ENDO	ESI Positive	0.17	0.04	0.5	Meat, poultry, seafood and eggs
M649.3892T0.81	ENDO	ESI Positive	0.23	0.004	0.3	Meat, poultry, seafood and eggs
M663.9023T0.81	ENDO	ESI Positive	0.28	0.0004	0.2	Meat, poultry, seafood and eggs
M667.4155T14.68	EXO	ESI Positive	-0.18	0.03	0.99	Meat, poultry, seafood and eggs
M677.8577T0.78	ENDO	ESI Positive	0.20	0.01	0.3	Meat, poultry, seafood and eggs
M684.2609T14.64	EXO	ESI Negative	0.21	0.008	0.5	Meat, poultry, seafood and eggs
M710.2435T15.26	EXO	ESI Negative	0.19	0.02	0.8	Meat, poultry, seafood and eggs
M731.8898T0.81	ENDO	ESI Positive	0.23	0.005	0.3	Meat, poultry, seafood and eggs
M758.8672T0.78	ENDO	ESI Negative	-0.19	0.02	0.998	Meat, poultry, seafood and eggs
M759.8697T0.78	ENDO	ESI Negative	-0.20	0.01	0.998	Meat, poultry, seafood and eggs
M788.5795T24.78	EXO	ESI Negative	-0.17	0.04	0.9	Meat, poultry, seafood and eggs
M828.8088T0.8	EXO	ESI Positive	0.24	0.004	0.96	Meat, poultry, seafood and eggs
M833.824T0.77	ENDO	ESI Positive	-0.20	0.01	0.3	Meat, poultry, seafood and eggs
M835.843T0.78	ENDO	ESI Negative	-0.16	0.049	0.998	Meat, poultry, seafood and eggs
M880.8196T0.78	EXO	ESI Positive	-0.18	0.03	0.99	Meat, poultry, seafood and eggs
M895.8448T0.78	ENDO	ESI Negative	-0.16	0.04	0.998	Meat, poultry, seafood and eggs
M978.8856T0.77	EXO	ESI Negative	0.16	0.0497	0.9	Meat, poultry, seafood and eggs
M996.8228T0.78	EXO	ESI Negative	0.16	0.045	0.9	Meat, poultry, seafood and eggs
M1002.4832T0.78	EXO	ESI Negative	-0.16	0.04	0.9	Meat, poultry, seafood and eggs
M69.0699T1.18	ENDO	ESI Positive	0.19	0.02	0.3	Milk and dairy products
M74.0237T1.16	ENDO	ESI Positive	0.17	0.03	0.3	Milk and dairy products
M78.9187T0.9	EXO	ESI Negative	0.24	0.003	0.4	Milk and dairy products
M80.9478T0.85	ENDO	ESI Positive	0.20	0.01	0.2	Milk and dairy products
M82.5371T0.67	ENDO	ESI Positive	0.17	0.04	0.3	Milk and dairy products
M84.0808T1.17 (Pipelic acid)	ENDO	ESI Positive	0.17	0.03	0.3	Milk and dairy products
M90.0548T0.85	ENDO	ESI Positive	-0.23	0.004	0.2	Milk and dairy products
M90.9765T0.79	ENDO	ESI Positive	0.21	0.008	0.2	Milk and dairy products
M92.5412T0.68	ENDO	ESI Positive	0.22	0.006	0.2	Milk and dairy products
M94.0453T0.68	EXO	ESI Positive	0.18	0.03	0.5	Milk and dairy products
M94.065T0.9	ENDO	ESI Positive	0.16	0.049	0.3	Milk and dairy products
M96.922T1.18	EXO	ESI Negative	-0.17	0.04	0.6	Milk and dairy products
M97.9912T0.68	EXO	ESI Positive	0.26	0.001	0.2	Milk and dairy products
M98.9842T1.08	ENDO	ESI Positive	0.19	0.02	0.3	Milk and dairy products
M102.9886T0.68	EXO	ESI Negative	0.22	0.006	0.4	Milk and dairy products
M104.1071T14.68	ENDO	ESI Positive	0.16	0.04	0.3	Milk and dairy products

## ANNEXE 6

M105.0337T8.29	ENDO	ESI Positive	0.19	0.02	0.3	Milk and dairy products
M106.0154T11.3	EXO	ESI Positive	-0.25	0.002	0.2	Milk and dairy products
M109.1014T9.03	ENDO	ESI Positive	0.19	0.02	0.3	Milk and dairy products
M110.0087T0.67	ENDO	ESI Positive	0.20	0.01	0.2	Milk and dairy products
M118.086T1.14	ENDO	ESI Positive	0.20	0.01	0.2	Milk and dairy products
M118.9271T11.85	EXO	ESI Negative	-0.16	0.04	0.6	Milk and dairy products
M120.0654T0.85	ENDO	ESI Positive	-0.18	0.03	0.3	Milk and dairy products
M120.0806T2.6	ENDO	ESI Positive	0.16	0.04	0.3	Milk and dairy products
M121.0395T1.17	ENDO	ESI Positive	-0.17	0.04	0.3	Milk and dairy products
M124.0242T0.67	ENDO	ESI Positive	0.23	0.005	0.2	Milk and dairy products
M127.0512T0.85	EXO	ESI Negative	0.22	0.006	0.4	Milk and dairy products
M130.065T7.39	EXO	ESI Positive	-0.20	0.01	0.5	Milk and dairy products
M130.0861T1.16 (Pipelic acid)	ENDO	ESI Positive	0.21	0.009	0.2	Milk and dairy products
M130.0862T0.72	ENDO	ESI Positive	0.19	0.02	0.3	Milk and dairy products
M131.0895T1.16 (Pipelic acid)	ENDO	ESI Positive	0.21	0.01	0.2	Milk and dairy products
M132.0654T0.85	ENDO	ESI Positive	-0.17	0.04	0.3	Milk and dairy products
M132.0807T7.39	EXO	ESI Positive	-0.17	0.04	0.5	Milk and dairy products
M132.1018T1.19	ENDO	ESI Positive	0.25	0.002	0.2	Milk and dairy products
M132.1018T2.06	EXO	ESI Positive	-0.17	0.03	0.5	Milk and dairy products
M134.019T0.83	ENDO	ESI Positive	-0.16	0.046	0.3	Milk and dairy products
M136.8774T0.94	ENDO	ESI Negative	0.24	0.003	0.3	Milk and dairy products
M138.0542T0.92	EXO	ESI Positive	0.17	0.04	0.5	Milk and dairy products
M138.8751T0.92	EXO	ESI Negative	0.25	0.002	0.3	Milk and dairy products
M142.0348T0.67	EXO	ESI Positive	0.17	0.04	0.6	Milk and dairy products
M144.982T21.82	EXO	ESI Positive	-0.18	0.03	0.5	Milk and dairy products
M146.0458T0.87	EXO	ESI Negative	-0.17	0.03	0.6	Milk and dairy products
M146.1174T0.92	ENDO	ESI Positive	0.19	0.02	0.3	Milk and dairy products
M148.0604T0.86	EXO	ESI Positive	-0.19	0.02	0.5	Milk and dairy products
M148.0967T1.16	ENDO	ESI Positive	0.25	0.002	0.2	Milk and dairy products
M153.0657T1.19	ENDO	ESI Positive	0.17	0.04	0.3	Milk and dairy products
M154.0622T0.82	EXO	ESI Negative	-0.16	0.04	0.6	Milk and dairy products
M155.0426T0.82	ENDO	ESI Positive	-0.18	0.03	0.3	Milk and dairy products
M157.0118T1.23	ENDO	ESI Negative	0.24	0.003	0.3	Milk and dairy products
M160.1331T1.12	ENDO	ESI Positive	0.27	0.0006	0.2	Milk and dairy products
M166.0862T2.77	ENDO	ESI Positive	0.18	0.02	0.3	Milk and dairy products
M166.0862T4.55	EXO	ESI Positive	-0.18	0.03	0.5	Milk and dairy products
M166.0862T5.82	EXO	ESI Positive	-0.19	0.02	0.5	Milk and dairy products
M169.0356T1.16	EXO	ESI Positive	0.17	0.04	0.6	Milk and dairy products
M169.0583T0.83	EXO	ESI Positive	0.19	0.02	0.5	Milk and dairy products
M172.0402T1.16	EXO	ESI Positive	0.18	0.03	0.5	Milk and dairy products
M174.956T0.7	EXO	ESI Negative	0.17	0.04	0.6	Milk and dairy products
M184.9855T0.63	ENDO	ESI Positive	-0.20	0.01	0.2	Milk and dairy products
M185.1284T7.08	ENDO	ESI Positive	-0.17	0.03	0.3	Milk and dairy products
M185.1284T7.11	EXO	ESI Positive	-0.22	0.008	0.3	Milk and dairy products
M191.0403T0.83	EXO	ESI Positive	0.21	0.01	0.4	Milk and dairy products



## ANNEXE 6

M192.0435T0.82	EXO	ESI Positive	0.18	0.03	0.5	Milk and dairy products
M203.0526T1.05	EXO	ESI Positive	-0.16	0.046	0.6	Milk and dairy products
M213.0225T11.41	ENDO	ESI Negative	-0.16	0.04	0.997	Milk and dairy products
M218.9196T0.82	ENDO	ESI Positive	0.18	0.02	0.3	Milk and dairy products
M223.1077T7.2	EXO	ESI Positive	-0.17	0.03	0.5	Milk and dairy products
M229.1547T1.18	ENDO	ESI Positive	0.28	0.0006	0.2	Milk and dairy products
M231.0839T1.16 (Ethyl-β-D-glucopyranoside)	EXO	ESI Positive	-0.20	0.02	0.5	Milk and dairy products
M231.1703T6.84	EXO	ESI Positive	-0.25	0.002	0.2	Milk and dairy products
M233.9467T11.85	EXO	ESI Negative	-0.19	0.02	0.6	Milk and dairy products
M241.1547T1.17	ENDO	ESI Positive	0.20	0.01	0.2	Milk and dairy products
M245.0768T1.17	ENDO	ESI Positive	0.21	0.009	0.2	Milk and dairy products
M248.1677T10.5	EXO	ESI Positive	-0.22	0.007	0.3	Milk and dairy products
M248.9989T0.84	ENDO	ESI Positive	0.22	0.007	0.2	Milk and dairy products
M261.0112T0.87	ENDO	ESI Positive	0.21	0.008	0.2	Milk and dairy products
M261.0112T1.05	EXO	ESI Positive	-0.28	0.0006	0.2	Milk and dairy products
M261.1439T7.47	EXO	ESI Positive	-0.18	0.03	0.5	Milk and dairy products
M263.1389T8.18	EXO	ESI Positive	-0.19	0.02	0.5	Milk and dairy products
M269.088T1.17	ENDO	ESI Positive	0.23	0.004	0.2	Milk and dairy products
M279.1702T8.61	EXO	ESI Positive	-0.20	0.01	0.4	Milk and dairy products
M279.1703T8.61	ENDO	ESI Positive	-0.17	0.04	0.3	Milk and dairy products
M281.2475T15.78	EXO	ESI Positive	-0.17	0.04	0.5	Milk and dairy products
M282.1911T1.18	ENDO	ESI Positive	0.18	0.03	0.3	Milk and dairy products
M282.845T7.53	EXO	ESI Positive	-0.26	0.002	0.2	Milk and dairy products
M285.9887T0.81	EXO	ESI Positive	0.25	0.002	0.2	Milk and dairy products
M286.9064T18.9	EXO	ESI Positive	-0.17	0.03	0.5	Milk and dairy products
M287.1606T7.61	ENDO	ESI Positive	-0.19	0.02	0.3	Milk and dairy products
M292.198T0.79	EXO	ESI Positive	0.16	0.047	0.6	Milk and dairy products
M292.6353T11.22	EXO	ESI Positive	-0.18	0.03	0.5	Milk and dairy products
M295.1287T7.84	EXO	ESI Positive	-0.24	0.003	0.2	Milk and dairy products
M300.1301T1.17	ENDO	ESI Positive	0.21	0.009	0.2	Milk and dairy products
M301.2173T17.26	EXO	ESI Negative	-0.18	0.03	0.6	Milk and dairy products
M302.2326T10.51	ENDO	ESI Positive	0.23	0.005	0.2	Milk and dairy products
M304.9136T0.69	EXO	ESI Negative	0.19	0.02	0.6	Milk and dairy products
M311.2227T15.13	EXO	ESI Negative	0.22	0.005	0.4	Milk and dairy products
<b>M312.2261T15.13</b>	EXO	ESI Negative	0.30	0.0002	0.07	Milk and dairy products
<b>M313.1546T8.83</b>	EXO	ESI Positive	-0.31	0.0002	0.1	milk and dairy products
M313.2383T15.84	EXO	ESI Negative	0.16	0.04	0.6	Milk and dairy products
M316.6477T14.68	ENDO	ESI Positive	0.19	0.02	0.3	Milk and dairy products
M318.9698T0.86	ENDO	ESI Positive	0.21	0.008	0.2	Milk and dairy products
M320.9669T0.86	ENDO	ESI Positive	0.19	0.02	0.3	Milk and dairy products
M332.2432T10.25	ENDO	ESI Positive	0.19	0.02	0.3	Milk and dairy products
M342.2639T11.84	ENDO	ESI Positive	0.16	0.047	0.3	Milk and dairy products
M350.1095T1.17	ENDO	ESI Positive	0.19	0.02	0.3	Milk and dairy products
M357.2787T14.6	EXO	ESI Positive	-0.17	0.03	0.5	Milk and dairy products
M361.1671T17.93	EXO	ESI Positive	-0.18	0.03	0.5	Milk and dairy products
M367.1584T12.78	ENDO	ESI Negative	0.19	0.02	0.6	Milk and dairy products

## ANNEXE 6

M368.1617T12.99	EXO	ESI Negative	0.20	0.02	0.5	Milk and dairy products
M368.9764T16	EXO	ESI Negative	-0.16	0.045	0.6	Milk and dairy products
M376.9283T0.89	EXO	ESI Positive	-0.18	0.02	0.5	Milk and dairy products
M376.9284T0.86	ENDO	ESI Positive	0.19	0.02	0.3	Milk and dairy products
M378.9003T0.78	EXO	ESI Positive	0.18	0.03	0.5	Milk and dairy products
M378.9254T0.86	ENDO	ESI Positive	0.20	0.02	0.3	Milk and dairy products
M380.2558T13.66	EXO	ESI Positive	-0.18	0.03	0.5	Milk and dairy products
M385.167T17.71	EXO	ESI Positive	-0.17	0.04	0.5	Milk and dairy products
M386.2899T11.84	EXO	ESI Positive	-0.18	0.03	0.5	Milk and dairy products
M391.2853T14.6	EXO	ESI Negative	-0.17	0.04	0.6	Milk and dairy products
M394.8742T0.79	EXO	ESI Positive	0.17	0.04	0.5	Milk and dairy products
M394.8907T0.75	ENDO	ESI Negative	-0.20	0.01	0.6	Milk and dairy products
M426.3577T14.21	EXO	ESI Positive	-0.20	0.01	0.5	Milk and dairy products
M427.3611T14.21	EXO	ESI Positive	-0.20	0.01	0.4	Milk and dairy products
M429.3767T14.81	EXO	ESI Positive	-0.17	0.04	0.6	Milk and dairy products
M443.0721T0.87	ENDO	ESI Positive	0.20	0.01	0.2	Milk and dairy products
M452.7371T8.77	ENDO	ESI Positive	-0.22	0.007	0.2	Milk and dairy products
M453.2929T17.56	EXO	ESI Positive	-0.17	0.04	0.5	Milk and dairy products
M459.9403T0.81	ENDO	ESI Positive	-0.20	0.01	0.2	Milk and dairy products
M462.8616T0.79	EXO	ESI Positive	0.21	0.009	0.4	Milk and dairy products
M464.3134T15.85	EXO	ESI Positive	-0.25	0.003	0.2	Milk and dairy products
M472.8285T17.51	EXO	ESI Positive	-0.18	0.03	0.5	Milk and dairy products
M478.2928T14.59	ENDO	ESI Positive	0.20	0.01	0.2	Milk and dairy products
M479.2838T24.53	EXO	ESI Negative	0.22	0.007	0.4	Milk and dairy products
M480.3084T15.47	ENDO	ESI Positive	0.16	0.047	0.3	Milk and dairy products
M482.3241T14.5	ENDO	ESI Positive	0.16	0.04	0.3	Milk and dairy products
M494.8418T17.46	EXO	ESI Positive	-0.19	0.02	0.5	Milk and dairy products
M508.3396T14.88	ENDO	ESI Positive	0.17	0.04	0.3	Milk and dairy products
M516.8549T17.41	EXO	ESI Positive	-0.18	0.03	0.5	Milk and dairy products
M518.3239T13.99	ENDO	ESI Positive	0.17	0.03	0.3	Milk and dairy products
M520.3397T14.68	ENDO	ESI Positive	0.20	0.01	0.2	Milk and dairy products
M538.8681T17.36	EXO	ESI Positive	-0.18	0.03	0.5	Milk and dairy products
M556.9917T0.87	ENDO	ESI Positive	0.17	0.03	0.3	Milk and dairy products
M558.989T0.87	ENDO	ESI Positive	0.22	0.007	0.2	Milk and dairy products
M560.8813T17.31	EXO	ESI Positive	-0.19	0.02	0.5	Milk and dairy products
M574.8763T0.78	EXO	ESI Positive	0.18	0.03	0.5	Milk and dairy products
M575.3781T0.78	EXO	ESI Positive	0.24	0.003	0.2	Milk and dairy products
M583.2548T12.41	ENDO	ESI Positive	0.19	0.02	0.3	Milk and dairy products
M596.7595T14.67	EXO	ESI Positive	0.27	0.001	0.2	Milk and dairy products
M598.2651T15.23	EXO	ESI Positive	0.20	0.02	0.5	Milk and dairy products
M609.3717T0.78	EXO	ESI Positive	0.18	0.03	0.5	Milk and dairy products
M615.3955T0.81	ENDO	ESI Positive	-0.18	0.02	0.3	Milk and dairy products
M616.8569T0.78	EXO	ESI Positive	0.24	0.003	0.2	Milk and dairy products
M619.2569T14.67	EXO	ESI Positive	0.23	0.004	0.2	Milk and dairy products
M632.0223T11.32	ENDO	ESI Positive	-0.18	0.03	0.3	Milk and dairy products
M643.8647T0.78	EXO	ESI Positive	0.17	0.04	0.5	Milk and dairy products

## ANNEXE 6

M663.4654T17.81	EXO	ESI Positive	-0.18	0.03	0.5	Milk and dairy products
M663.9023T0.81	ENDO	ESI Positive	-0.19	0.02	0.3	Milk and dairy products
M667.9075T11.28	EXO	ESI Positive	-0.20	0.02	0.5	Milk and dairy products
M676.8574T0.78	EXO	ESI Positive	0.20	0.01	0.5	Milk and dairy products
M677.3592T0.78	EXO	ESI Positive	0.25	0.002	0.2	Milk and dairy products
M677.8584T0.78	EXO	ESI Positive	0.23	0.005	0.2	Milk and dairy products
M682.1033T8.76	ENDO	ESI Positive	-0.16	0.049	0.3	Milk and dairy products
M683.3829T0.81	ENDO	ESI Positive	-0.17	0.04	0.3	Milk and dairy products
M684.8443T0.78	EXO	ESI Positive	0.23	0.004	0.2	Milk and dairy products
M684.926T0.76	ENDO	ESI Positive	0.16	0.04	0.3	Milk and dairy products
M685.3464T0.78	EXO	ESI Positive	0.21	0.009	0.4	Milk and dairy products
M701.3291T11.32	ENDO	ESI Positive	-0.17	0.04	0.3	Milk and dairy products
M714.5287T8.49	ENDO	ESI Positive	-0.24	0.002	0.2	Milk and dairy products
M731.8898T0.81	ENDO	ESI Positive	-0.16	0.04	0.3	Milk and dairy products
M752.8318T0.78	EXO	ESI Positive	0.25	0.002	0.2	Milk and dairy products
M759.1386T11.3	ENDO	ESI Positive	-0.22	0.006	0.2	Milk and dairy products
M759.8014T11.3	ENDO	ESI Positive	-0.20	0.01	0.2	Milk and dairy products
M791.9992T8.77	ENDO	ESI Positive	-0.18	0.03	0.3	Milk and dairy products
M798.5292T24.78	EXO	ESI Negative	0.20	0.02	0.5	Milk and dairy products
M806.462T14.68	ENDO	ESI Positive	0.16	0.047	0.3	Milk and dairy products
M813.8326T0.78	ENDO	ESI Positive	0.21	0.009	0.2	Milk and dairy products
M820.8193T0.78	EXO	ESI Positive	0.25	0.003	0.2	Milk and dairy products
M847.3277T0.78	EXO	ESI Positive	0.25	0.003	0.2	Milk and dairy products
M847.8262T0.78	ENDO	ESI Positive	0.20	0.01	0.2	Milk and dairy products
M847.8273T0.78	EXO	ESI Positive	0.17	0.03	0.5	Milk and dairy products
M857.5186T24.8	EXO	ESI Negative	-0.16	0.045	0.6	Milk and dairy products
M881.8495T11.29	ENDO	ESI Positive	0.16	0.045	0.3	Milk and dairy products
M881.9498T11.29	ENDO	ESI Positive	0.17	0.04	0.3	Milk and dairy products
M882.0499T11.29	ENDO	ESI Positive	0.17	0.03	0.3	Milk and dairy products
M888.8067T0.78	EXO	ESI Positive	0.29	0.0004	0.2	Milk and dairy products
M894.8418T0.78	EXO	ESI Negative	-0.20	0.01	0.5	Milk and dairy products
M895.3435T0.78	EXO	ESI Negative	-0.18	0.03	0.6	Milk and dairy products
M922.7997T0.78	EXO	ESI Positive	0.21	0.01	0.4	Milk and dairy products
M926.3706T0.78	EXO	ESI Negative	-0.17	0.04	0.6	Milk and dairy products
M928.8352T0.77	EXO	ESI Negative	-0.17	0.03	0.6	Milk and dairy products
M930.3399T0.78	EXO	ESI Negative	-0.20	0.01	0.5	Milk and dairy products
M949.8077T0.78	ENDO	ESI Positive	0.17	0.04	0.3	Milk and dairy products
M956.7945T0.78	EXO	ESI Positive	0.17	0.04	0.6	Milk and dairy products
M957.4776T0.78	EXO	ESI Positive	0.17	0.04	0.5	Milk and dairy products
M963.3313T0.78	EXO	ESI Negative	-0.16	0.04	0.6	Milk and dairy products
M972.1634T10.64	ENDO	ESI Positive	-0.19	0.02	0.3	Milk and dairy products
M974.1621T0.78	ENDO	ESI Negative	-0.16	0.047	0.997	Milk and dairy products
M979.3865T11.29	ENDO	ESI Positive	0.17	0.03	0.3	Milk and dairy products
M983.8026T0.78	EXO	ESI Positive	0.19	0.02	0.5	Milk and dairy products
M985.6004T11.29	ENDO	ESI Positive	0.18	0.03	0.3	Milk and dairy products
M985.7148T11.28	ENDO	ESI Positive	0.18	0.02	0.3	Milk and dairy products

## ANNEXE 6

M997.1595T0.78	ENDO	ESI Negative	0.17	0.03	0.997	Milk and dairy products
M997.8257T0.78	EXO	ESI Negative	-0.17	0.04	0.6	Milk and dairy products
M1004.8108T0.78	ENDO	ESI Negative	0.23	0.005	0.3	Milk and dairy products
M69.0335T0.88	ENDO	ESI Positive	-0.17	0.03	0.6	Salt
M80.9166T0.9	EXO	ESI Negative	-0.20	0.01	0.7	Salt
M84.0808T1.17 (Pipelic acid)	ENDO	ESI Positive	-0.20	0.01	0.5	Salt
M101.0243T0.87	ENDO	ESI Negative	0.18	0.03	0.9	Salt
M114.0659T0.88	EXO	ESI Positive	-0.18	0.03	0.99	Salt
M124.0868T20.39	EXO	ESI Positive	-0.24	0.003	0.99	Salt
M130.0861T1.16 (Pipelic acid)	ENDO	ESI Positive	-0.22	0.006	0.4	Salt
M131.0895T1.16 (Pipelic acid)	ENDO	ESI Positive	-0.19	0.02	0.5	Salt
M132.1018T1.19	ENDO	ESI Positive	-0.16	0.04	0.6	Salt
M143.9589T15.11	EXO	ESI Positive	-0.22	0.007	0.99	Salt
M153.0657T1.19	ENDO	ESI Positive	-0.17	0.04	0.6	Salt
M155.1065T9.03	ENDO	ESI Positive	-0.16	0.049	0.6	Salt
M157.087T17.71	EXO	ESI Negative	0.16	0.049	0.8	Salt
M158.8466T0.99	EXO	ESI Negative	-0.17	0.03	0.8	Salt
M161.9694T16.61	ENDO	ESI Positive	0.22	0.006	0.4	Salt
M164.0294T0.83	EXO	ESI Positive	0.20	0.02	0.99	Salt
M166.0862T2.77	ENDO	ESI Positive	-0.17	0.04	0.6	Salt
M172.1331T9.03	ENDO	ESI Positive	-0.20	0.01	0.5	Salt
M176.1278T0.96	EXO	ESI Positive	0.24	0.003	0.99	Salt
M177.0246T0.82	EXO	ESI Positive	0.18	0.03	0.99	Salt
M183.139T14.61	EXO	ESI Negative	-0.19	0.02	0.8	Salt
M189.0469T0.66	EXO	ESI Positive	0.18	0.03	0.99	Salt
M194.9242T14.96	EXO	ESI Negative	-0.16	0.0498	0.8	Salt
M196.1261T6.92	EXO	ESI Positive	-0.19	0.02	0.99	Salt
M203.0526T1.05	EXO	ESI Positive	-0.21	0.009	0.99	Salt
M227.2116T10.17	EXO	ESI Positive	0.20	0.02	0.99	Salt
M227.6268T18.39	EXO	ESI Positive	-0.19	0.02	0.99	Salt
M251.1277T13.29	ENDO	ESI Positive	-0.18	0.02	0.5	Salt
M251.2016T17.07	EXO	ESI Negative	-0.17	0.03	0.8	Salt
M257.9773T0.68	ENDO	ESI Positive	-0.18	0.02	0.5	Salt
M268.2184T20.19	EXO	ESI Positive	-0.20	0.01	0.99	Salt
M275.6254T14.22	ENDO	ESI Positive	-0.16	0.04	0.6	Salt
M280.6547T15.59	ENDO	ESI Positive	-0.23	0.005	0.4	Salt
M282.845T7.53	EXO	ESI Positive	0.17	0.04	0.99	Salt
M287.1609T7.61	EXO	ESI Positive	0.23	0.006	0.99	Salt
M289.6593T15.59	ENDO	ESI Positive	-0.18	0.03	0.5	Salt
M295.1584T7.34	EXO	ESI Positive	0.17	0.04	0.99	Salt
M297.1468T15.2	EXO	ESI Positive	-0.23	0.006	0.99	Salt
M299.6254T14.04	ENDO	ESI Positive	-0.16	0.046	0.6	Salt
M331.2265T18.08	EXO	ESI Positive	-0.17	0.04	0.99	Salt
M332.9217T0.78	ENDO	ESI Negative	-0.16	0.04	0.9	Salt
M341.3162T15.4	ENDO	ESI Positive	-0.19	0.02	0.5	Salt

## ANNEXE 6

M343.3318T16.15	ENDO	ESI Positive	-0.16	0.04	0.6	Salt
M352.1653T9.04	ENDO	ESI Positive	-0.18	0.03	0.5	Salt
M372.3472T15.91	ENDO	ESI Positive	-0.16	0.0499	0.6	Salt
M452.7375T8.77	EXO	ESI Positive	0.21	0.01	0.99	Salt
M454.2928T15.09	ENDO	ESI Positive	-0.20	0.01	0.5	Salt
M458.3474T12.43	EXO	ESI Positive	0.17	0.04	0.99	Salt
M460.7351T8.63	EXO	ESI Positive	0.24	0.004	0.99	Salt
M463.2325T13.17	ENDO	ESI Positive	-0.18	0.02	0.5	Salt
M468.3084T13.84	ENDO	ESI Positive	-0.19	0.02	0.5	Salt
M475.1651T11.22	EXO	ESI Positive	-0.20	0.01	0.99	Salt
M482.324T16.6	ENDO	ESI Positive	-0.18	0.02	0.5	Salt
M494.3241T14.23	ENDO	ESI Positive	-0.16	0.048	0.6	Salt
M494.3252T16.02	EXO	ESI Negative	-0.17	0.04	0.8	Salt
M496.3398T15.2	ENDO	ESI Positive	-0.20	0.01	0.5	Salt
M508.3396T14.88	ENDO	ESI Positive	-0.17	0.04	0.6	Salt
M510.3554T15.97	ENDO	ESI Positive	-0.21	0.008	0.4	Salt
M512.2993T13.83	EXO	ESI Negative	-0.21	0.009	0.7	Salt
M518.3239T13.99	ENDO	ESI Positive	-0.19	0.02	0.5	Salt
M522.3554T15.58	ENDO	ESI Positive	-0.19	0.02	0.5	Salt
M524.371T16.82	ENDO	ESI Positive	-0.21	0.008	0.4	Salt
M532.6178T15.22	EXO	ESI Positive	0.19	0.02	0.99	Salt
M539.9777T0.78	EXO	ESI Positive	0.18	0.03	0.99	Salt
M540.3306T14.97	EXO	ESI Negative	-0.18	0.03	0.8	Salt
M540.3306T15.61	EXO	ESI Negative	-0.20	0.02	0.7	Salt
M541.334T15.22	EXO	ESI Negative	-0.18	0.02	0.8	Salt
M546.3553T15.18	ENDO	ESI Positive	-0.18	0.03	0.5	Salt
M548.3709T15.96	ENDO	ESI Positive	-0.18	0.03	0.5	Salt
M554.3464T15.84	EXO	ESI Negative	-0.22	0.007	0.7	Salt
M568.3619T17	EXO	ESI Negative	-0.17	0.03	0.8	Salt
M568.3621T17.55	EXO	ESI Negative	-0.26	0.001	0.4	Salt
M598.2651T15.21	ENDO	ESI Positive	-0.20	0.01	0.5	Salt
M610.3727T15.62	EXO	ESI Negative	-0.17	0.04	0.8	Salt
M614.3462T14.98	EXO	ESI Negative	-0.17	0.03	0.8	Salt
M621.2727T15.59	ENDO	ESI Positive	-0.20	0.02	0.5	Salt
M664.5051T11.29	ENDO	ESI Positive	-0.18	0.02	0.5	Salt
M736.2589T15.62	EXO	ESI Negative	-0.21	0.01	0.7	Salt
M753.2407T15.25	EXO	ESI Negative	0.20	0.01	0.7	Salt
M806.5014T15.23	EXO	ESI Negative	-0.17	0.03	0.8	Salt
M816.1771T0.77	EXO	ESI Positive	0.17	0.04	0.99	Salt
M842.8285T0.78	EXO	ESI Negative	-0.20	0.01	0.7	Salt
M861.3502T0.78	ENDO	ESI Negative	0.20	0.01	0.9	Salt
M869.144T11.31	ENDO	ESI Positive	-0.25	0.002	0.4	Salt
M894.8422T0.78	ENDO	ESI Negative	0.17	0.04	0.9	Salt
M918.3358T10.11	ENDO	ESI Positive	-0.22	0.006	0.4	Salt
M954.1525T11.32	ENDO	ESI Positive	-0.24	0.003	0.4	Salt
M960.9475T0.75	ENDO	ESI Negative	0.18	0.03	0.9	Salt

## ANNEXE 6

M965.2692T11.3	ENDO	ESI Positive	-0.22	0.005	0.4	Salt
M966.1605T11.32	ENDO	ESI Positive	-0.23	0.004	0.4	Salt
M972.1613T11.27	ENDO	ESI Positive	-0.24	0.003	0.4	Salt
M972.1634T10.64	ENDO	ESI Positive	-0.20	0.02	0.5	Salt
M1002.4839T0.78	ENDO	ESI Negative	0.25	0.002	0.4	Salt
M61.0394T0.91	EXO	ESI Positive	-0.25	0.002	0.2	Seafood
M67.522T0.67	EXO	ESI Positive	-0.23	0.005	0.2	Seafood
M68.5261T0.67	EXO	ESI Positive	-0.22	0.008	0.3	Seafood
M68.9929T0.66	EXO	ESI Positive	-0.17	0.04	0.5	Seafood
M69.0335T0.88	ENDO	ESI Positive	-0.18	0.02	0.4	Seafood
M74.0237T1.16	ENDO	ESI Positive	0.16	0.049	0.4	Seafood
M81.0331T0.68	EXO	ESI Positive	-0.18	0.03	0.4	Seafood
M84.0808T1.17 (Pipelic acid)	ENDO	ESI Positive	0.17	0.04	0.4	Seafood
M93.9839T0.67	ENDO	ESI Positive	0.18	0.03	0.4	Seafood
M94.0453T0.68	EXO	ESI Positive	-0.21	0.01	0.3	Seafood
M97.0658T11.79	EXO	ESI Negative	-0.23	0.004	0.1	Seafood
M99.044T0.86	EXO	ESI Positive	-0.18	0.03	0.4	Seafood
M87.044T12.01	EXO	ESI Positive	0.28	0.0005	0.1	Seafood
M100.0758T5.97	ENDO	ESI Positive	-0.21	0.01	0.4	Seafood
M102.0223T0.68	EXO	ESI Positive	-0.21	0.01	0.3	Seafood
M103.9201T11.85	EXO	ESI Negative	0.16	0.048	0.3	Seafood
M108.5179T0.63	ENDO	ESI Positive	-0.16	0.04	0.4	Seafood
M108.5375T0.68	EXO	ESI Positive	-0.25	0.002	0.2	Seafood
M108.9875T0.78	EXO	ESI Positive	-0.18	0.03	0.4	Seafood
M110.0088T0.68	EXO	ESI Positive	-0.21	0.01	0.3	Seafood
M112.9855T15.28	EXO	ESI Negative	-0.18	0.03	0.2	Seafood
M112.9855T16.93	ENDO	ESI Negative	-0.19	0.02	0.7	Seafood
M114.0548T1.17	EXO	ESI Positive	-0.17	0.04	0.5	Seafood
M114.0659T0.88	EXO	ESI Positive	-0.21	0.01	0.3	Seafood
M114.0659T0.89	ENDO	ESI Positive	-0.16	0.048	0.4	Seafood
M118.0348T0.68	EXO	ESI Positive	-0.20	0.01	0.3	Seafood
M118.0649T7.39	EXO	ESI Positive	-0.20	0.01	0.3	Seafood
M120.0654T0.85	ENDO	ESI Positive	0.18	0.02	0.4	Seafood
M124.0242T0.67	EXO	ESI Positive	-0.21	0.01	0.3	Seafood
M129.0395T0.67	EXO	ESI Positive	-0.24	0.003	0.2	Seafood
M132.0807T7.39	EXO	ESI Positive	-0.20	0.01	0.4	Seafood
M132.1018T2.06	EXO	ESI Positive	-0.17	0.04	0.5	Seafood
M136.048T0.87	EXO	ESI Positive	-0.19	0.02	0.4	Seafood
M138.8751T0.92	EXO	ESI Negative	0.18	0.02	0.2	Seafood
M142.0348T0.67	EXO	ESI Positive	-0.20	0.01	0.3	Seafood
M144.8696T0.89	EXO	ESI Negative	-0.21	0.008	0.2	Seafood
M145.0494T0.87	ENDO	ESI Positive	-0.20	0.02	0.4	Seafood
M146.0458T0.87	ENDO	ESI Negative	-0.18	0.03	0.7	Seafood
M146.9385T0.66	EXO	ESI Negative	-0.16	0.048	0.3	Seafood
M150.0137T0.82	ENDO	ESI Positive	0.23	0.005	0.4	Seafood
M150.8865T0.87	ENDO	ESI Negative	-0.17	0.03	0.7	Seafood

## ANNEXE 6

M152.0318T0.83	ENDO	ESI Positive	0.18	0.02	0.4	Seafood
M155.079T0.76	EXO	ESI Positive	0.17	0.03	0.4	Seafood
M155.8665T0.99	EXO	ESI Negative	-0.16	0.049	0.3	Seafood
M156.014T0.67	EXO	ESI Positive	-0.17	0.04	0.5	Seafood
<b>M157.0118T23.44</b>	EXO	ESI Negative	0.31	0.0001	0.02	Seafood
M157.087T17.71	EXO	ESI Negative	-0.19	0.02	0.2	Seafood
M158.0029T0.57	EXO	ESI Positive	-0.23	0.005	0.2	Seafood
M158.1175T0.98	EXO	ESI Positive	-0.27	0.0009	0.2	Seafood
M158.1175T1.17	ENDO	ESI Positive	-0.19	0.02	0.4	Seafood
M158.1175T7.42	EXO	ESI Positive	0.17	0.04	0.5	Seafood
M158.8466T0.99	EXO	ESI Negative	-0.19	0.02	0.2	Seafood
M162.1124T0.86	EXO	ESI Positive	-0.18	0.03	0.4	Seafood
M162.1124T0.87	ENDO	ESI Positive	-0.18	0.02	0.4	Seafood
M164.0293T0.84	ENDO	ESI Positive	0.24	0.003	0.4	Seafood
M164.1173T0.86	ENDO	ESI Positive	-0.17	0.03	0.4	Seafood
M166.0862T5.82	EXO	ESI Positive	-0.16	0.0499	0.5	Seafood
M166.8603T0.87	EXO	ESI Negative	-0.19	0.02	0.2	Seafood
M169.0583T0.84	ENDO	ESI Positive	0.16	0.04	0.4	Seafood
M174.9379T0.79	ENDO	ESI Positive	-0.16	0.04	0.4	Seafood
M174.956T0.7	EXO	ESI Negative	-0.17	0.04	0.3	Seafood
M176.8602T0.98	EXO	ESI Negative	-0.28	0.0004	0.05	Seafood
M178.858T0.99	EXO	ESI Negative	-0.20	0.01	0.2	Seafood
M180.973T0.78	EXO	ESI Negative	0.24	0.003	0.1	Seafood
M181.072T6.96	ENDO	ESI Positive	-0.26	0.001	0.4	Seafood
M181.072T7	EXO	ESI Positive	-0.26	0.001	0.2	Seafood
M181.072T7.44	EXO	ESI Positive	-0.19	0.02	0.4	Seafood
M182.0811T2.06	EXO	ESI Positive	-0.24	0.003	0.2	Seafood
M182.0811T2.09	ENDO	ESI Positive	-0.17	0.03	0.4	Seafood
M186.1124T0.98	EXO	ESI Positive	-0.19	0.02	0.4	Seafood
M186.1124T1.16	ENDO	ESI Positive	-0.16	0.047	0.4	Seafood
M189.1234T1.15	ENDO	ESI Positive	-0.18	0.02	0.4	Seafood
M189.1597T0.79	EXO	ESI Positive	-0.21	0.01	0.3	Seafood
M189.1597T0.79	ENDO	ESI Positive	-0.20	0.01	0.4	Seafood
M197.1284T7.96	EXO	ESI Positive	-0.22	0.006	0.3	Seafood
M200.8588T22.65	EXO	ESI Negative	0.18	0.02	0.2	Seafood
M201.8019T0.99	EXO	ESI Negative	-0.20	0.01	0.2	Seafood
M202.858T20.4	EXO	ESI Negative	0.27	0.0007	0.06	Seafood
M205.0971T7.37	ENDO	ESI Positive	-0.17	0.03	0.4	Seafood
M205.0971T7.38	EXO	ESI Positive	-0.20	0.02	0.4	Seafood
M207.1591T9.32	EXO	ESI Positive	0.18	0.03	0.4	Seafood
M209.1899T13.82	EXO	ESI Positive	0.17	0.03	0.4	Seafood
M211.144T8.79	EXO	ESI Positive	-0.24	0.004	0.2	Seafood
M216.1593T16.62	ENDO	ESI Positive	0.23	0.005	0.4	Seafood
M221.0632T0.86	EXO	ESI Positive	0.19	0.02	0.4	Seafood
M226.1649T7.7	EXO	ESI Positive	0.23	0.005	0.2	Seafood
M227.1251T0.87	EXO	ESI Positive	-0.19	0.02	0.4	Seafood

## ANNEXE 6

M230.144T9.36	EXO	ESI Positive	0.21	0.009	0.3	Seafood
M232.7638T24.56	EXO	ESI Negative	0.17	0.04	0.3	Seafood
M233.9821T1.16	EXO	ESI Positive	-0.18	0.03	0.4	Seafood
M235.926T13.26	EXO	ESI Negative	-0.16	0.049	0.3	Seafood
<b>M241.1031T0.89</b>	EXO	ESI Positive	-0.32	0.0001	0.03	Seafood
M242.9254T0.78	EXO	ESI Positive	-0.19	0.02	0.4	Seafood
M242.9254T0.79	ENDO	ESI Positive	-0.21	0.01	0.4	Seafood
M244.227T14.22	EXO	ESI Positive	0.20	0.01	0.3	Seafood
M245.0768T1.17	ENDO	ESI Positive	-0.20	0.01	0.4	Seafood
M251.9927T1.16	EXO	ESI Positive	-0.24	0.003	0.2	Seafood
M255.2329T0.75	EXO	ESI Negative	-0.19	0.02	0.2	Seafood
M257.0995T8.29	EXO	ESI Positive	-0.18	0.03	0.4	Seafood
M258.1101T0.85	ENDO	ESI Positive	-0.17	0.04	0.4	Seafood
M261.0112T1.05	EXO	ESI Positive	-0.17	0.04	0.5	Seafood
M263.2216T13.86	EXO	ESI Positive	0.18	0.02	0.4	Seafood
M265.1479T21.41	EXO	ESI Negative	0.17	0.04	0.3	Seafood
M268.2184T20.19	EXO	ESI Positive	-0.20	0.01	0.3	Seafood
M269.2218T20.19	EXO	ESI Positive	-0.23	0.004	0.2	Seafood
M271.0137T8.24	EXO	ESI Positive	-0.17	0.04	0.5	Seafood
M275.6257T14.21	EXO	ESI Positive	-0.17	0.03	0.4	Seafood
M279.2329T18.16	ENDO	ESI Negative	0.16	0.045	0.7	Seafood
M280.2482T13.86	EXO	ESI Positive	0.18	0.03	0.4	Seafood
M283.2643T0.75	EXO	ESI Negative	-0.19	0.02	0.2	Seafood
M285.9886T0.81	ENDO	ESI Positive	0.19	0.02	0.4	Seafood
M286.132T14.67	EXO	ESI Positive	-0.19	0.02	0.4	Seafood
M290.6322T13.84	ENDO	ESI Positive	0.17	0.04	0.4	Seafood
M292.2268T13.66	ENDO	ESI Positive	0.23	0.005	0.4	Seafood
M292.6377T13.82	EXO	ESI Positive	-0.18	0.03	0.4	Seafood
M297.1468T15.2	EXO	ESI Positive	-0.16	0.046	0.5	Seafood
M299.6254T14.04	ENDO	ESI Positive	0.16	0.046	0.4	Seafood
M302.2325T10.49						
(2,6 dimethylheptanoyl carnitine/ nonanoyl-L-carnitine; acylcarnitine C9:1)	EXO	ESI Positive	-0.16	0.047	0.5	Seafood
M303.6472T14.62	EXO	ESI Positive	0.17	0.03	0.4	Seafood
M304.9136T0.69	EXO	ESI Negative	-0.23	0.005	0.2	Seafood
M305.1569T7.67	EXO	ESI Positive	-0.16	0.047	0.5	Seafood
M309.1531T24.5	EXO	ESI Negative	0.18	0.02	0.2	Seafood
M310.1553T15.6	EXO	ESI Positive	-0.21	0.01	0.3	Seafood
M310.9129T0.78	EXO	ESI Positive	-0.26	0.001	0.2	Seafood
M312.8361T20.41	EXO	ESI Negative	0.18	0.03	0.2	Seafood
M317.6606T15.22	EXO	ESI Positive	-0.28	0.0005	0.1	Seafood
<b>M318.1622T15.23</b>	EXO	ESI Positive	-0.35	0.00001	0.01	seafood
M322.9641T0.88	EXO	ESI Positive	-0.22	0.007	0.3	Seafood
M327.2329T17.75	ENDO	ESI Negative	0.19	0.02	0.7	Seafood
M328.2483T11.29	EXO	ESI Positive	-0.16	0.049	0.5	Seafood
M329.3162T12.67	EXO	ESI Positive	-0.20	0.01	0.3	Seafood



## ANNEXE 6

M331.2265T18.08	EXO	ESI Positive	-0.18	0.03	0.4	Seafood
M331.6761T16.91	EXO	ESI Positive	-0.17	0.04	0.5	Seafood
M332.0698T11.8	EXO	ESI Positive	0.16	0.04	0.5	Seafood
M340.3938T15.88	ENDO	ESI Positive	0.16	0.04	0.4	Seafood
M344.2796T12.31	ENDO	ESI Positive	0.16	0.047	0.4	Seafood
M347.2204T18.17	ENDO	ESI Negative	0.16	0.04	0.7	Seafood
M350.1254T21.46	EXO	ESI Negative	0.18	0.03	0.2	Seafood
M353.1425T17.9	ENDO	ESI Negative	0.17	0.04	0.7	Seafood
M353.1686T9.04	EXO	ESI Positive	-0.17	0.04	0.5	Seafood
M354.285T14.55	EXO	ESI Positive	0.21	0.01	0.3	Seafood
M355.3682T13.49	EXO	ESI Positive	0.18	0.02	0.4	Seafood
M368.2793T12.25	ENDO	ESI Positive	0.21	0.01	0.4	Seafood
M368.3156T16.46	ENDO	ESI Positive	0.17	0.04	0.4	Seafood
M370.2951T12.73	ENDO	ESI Positive	0.19	0.02	0.4	Seafood
M371.3346T17.46	ENDO	ESI Positive	0.20	0.01	0.4	Seafood
M375.4105T0.68	ENDO	ESI Positive	0.18	0.03	0.4	Seafood
M381.1738T18.97	EXO	ESI Negative	-0.18	0.02	0.2	Seafood
M381.2316T24.58	EXO	ESI Negative	0.23	0.004	0.1	Seafood
M384.0503T11.85	EXO	ESI Positive	-0.16	0.0497	0.5	Seafood
M385.167T17.71	EXO	ESI Positive	0.25	0.003	0.2	Seafood
M429.301T15.49	EXO	ESI Negative	-0.18	0.02	0.2	Seafood
M439.3529T12.8	EXO	ESI Positive	-0.17	0.03	0.4	Seafood
M450.3212T11.8	EXO	ESI Positive	-0.17	0.04	0.5	Seafood
M452.2771T14.11	EXO	ESI Positive	-0.17	0.04	0.5	Seafood
M462.8616T0.79	EXO	ESI Positive	-0.16	0.046	0.5	Seafood
M464.1913T10.3	EXO	ESI Positive	-0.25	0.002	0.2	Seafood
M468.3084T13.84	ENDO	ESI Positive	-0.17	0.04	0.4	Seafood
M478.2927T14.56	EXO	ESI Positive	-0.16	0.04	0.5	Seafood
M479.2838T24.53	EXO	ESI Negative	0.19	0.02	0.2	Seafood
M490.2903T13.82	EXO	ESI Positive	-0.17	0.04	0.5	Seafood
M498.9299T20.15	ENDO	ESI Negative	0.19	0.02	0.7	Seafood
M500.277T13.97	ENDO	ESI Positive	0.16	0.04	0.4	Seafood
M502.2938T15.05	EXO	ESI Negative	-0.19	0.02	0.2	Seafood
M504.3085T15.06	EXO	ESI Positive	-0.18	0.03	0.4	Seafood
M508.3397T14.88	EXO	ESI Positive	-0.19	0.02	0.4	Seafood
M512.2993T13.83	EXO	ESI Negative	-0.20	0.01	0.2	Seafood
M524.2781T14.53	EXO	ESI Negative	0.19	0.02	0.2	Seafood
M526.2927T14.57	ENDO	ESI Positive	0.18	0.02	0.4	Seafood
M528.3082T14.9	ENDO	ESI Positive	0.20	0.01	0.4	Seafood
M538.315T14.23	EXO	ESI Negative	-0.21	0.01	0.2	Seafood
M539.9777T0.78	EXO	ESI Positive	0.21	0.01	0.3	Seafood
M548.8124T15.59	ENDO	ESI Positive	0.18	0.03	0.4	Seafood
M550.3866T17.13	ENDO	ESI Positive	0.17	0.03	0.4	Seafood
M553.3339T14.95	EXO	ESI Negative	-0.19	0.02	0.2	Seafood
M566.3366T14.69	ENDO	ESI Negative	0.17	0.03	0.7	Seafood
M567.2254T13.82	EXO	ESI Positive	-0.22	0.008	0.3	Seafood

## ANNEXE 6

M568.3397T14.65	ENDO	ESI Positive	0.17	0.03	0.4	Seafood
M599.2684T15.21	EXO	ESI Positive	-0.25	0.002	0.2	Seafood
M609.3983T15.22	EXO	ESI Positive	-0.18	0.03	0.4	Seafood
M609.8708T0.78	EXO	ESI Positive	-0.18	0.03	0.4	Seafood
M612.3305T14.64	EXO	ESI Negative	0.25	0.002	0.09	Seafood
M612.3309T14.66	ENDO	ESI Negative	0.23	0.005	0.7	Seafood
M633.687T11.27	ENDO	ESI Positive	-0.19	0.02	0.4	Seafood
M642.8637T0.78	EXO	ESI Positive	-0.17	0.04	0.5	Seafood
M643.3655T0.78	EXO	ESI Positive	-0.17	0.04	0.5	Seafood
M644.2624T15.6	EXO	ESI Positive	-0.17	0.04	0.5	Seafood
M681.094T0.86	EXO	ESI Positive	-0.19	0.02	0.4	Seafood
M684.2609T14.64	EXO	ESI Negative	0.17	0.04	0.3	Seafood
M685.2535T15.23	EXO	ESI Negative	-0.19	0.02	0.2	Seafood
M724.8724T0.77	EXO	ESI Negative	0.19	0.02	0.2	Seafood
M734.2454T14.68	EXO	ESI Negative	-0.18	0.03	0.2	Seafood
M748.569T14.68	EXO	ESI Negative	0.17	0.03	0.2	Seafood
M759.8697T0.78	ENDO	ESI Negative	-0.17	0.04	0.7	Seafood
M798.5292T24.78	EXO	ESI Negative	0.21	0.008	0.2	Seafood
M801.3495T0.78	EXO	ESI Negative	0.17	0.03	0.3	Seafood
<b>M810.5654T24.77</b>	EXO	ESI Negative	0.31	0.0001	0.02	seafood
M827.857T0.78	EXO	ESI Negative	0.17	0.04	0.3	Seafood
M842.8288T0.78	ENDO	ESI Negative	-0.16	0.046	0.7	Seafood
M850.5604T15.22	EXO	ESI Negative	0.17	0.04	0.3	Seafood
M861.3499T0.78	EXO	ESI Negative	0.17	0.03	0.2	Seafood
M868.5027T10.07	ENDO	ESI Positive	-0.17	0.03	0.4	Seafood
M877.4057T0.77	ENDO	ESI Negative	0.18	0.03	0.7	Seafood
M881.548T11.29	ENDO	ESI Positive	-0.20	0.01	0.4	Seafood
M881.6492T11.29	ENDO	ESI Positive	-0.21	0.01	0.4	Seafood
M881.8495T11.29	ENDO	ESI Positive	-0.17	0.04	0.4	Seafood
M882.0499T11.29	ENDO	ESI Positive	-0.16	0.04	0.4	Seafood
M887.6517T11.25	ENDO	ESI Positive	-0.16	0.0496	0.4	Seafood
M894.8418T0.78	EXO	ESI Negative	0.17	0.04	0.3	Seafood
M895.3435T0.78	EXO	ESI Negative	0.28	0.0005	0.05	Seafood
M896.3457T0.78	EXO	ESI Negative	0.18	0.03	0.2	Seafood
M914.8134T0.78	EXO	ESI Positive	-0.18	0.03	0.4	Seafood
M922.7997T0.78	EXO	ESI Positive	-0.20	0.02	0.4	Seafood
M928.8352T0.77	EXO	ESI Negative	0.16	0.04	0.3	Seafood
M929.3373T0.78	EXO	ESI Negative	0.20	0.01	0.2	Seafood
M929.8382T0.77	EXO	ESI Negative	0.21	0.009	0.2	Seafood
M930.3399T0.78	EXO	ESI Negative	0.20	0.02	0.2	Seafood
M941.8226T0.77	EXO	ESI Positive	-0.17	0.03	0.4	Seafood
M952.1527T0.77	EXO	ESI Positive	-0.19	0.02	0.4	Seafood
M962.8295T0.78	EXO	ESI Negative	0.18	0.03	0.2	Seafood
M963.3313T0.78	EXO	ESI Negative	0.19	0.02	0.2	Seafood
M963.8319T0.78	EXO	ESI Negative	0.21	0.008	0.2	Seafood
M964.3333T0.78	EXO	ESI Negative	0.18	0.03	0.2	Seafood

## ANNEXE 6

M966.1605T11.32	ENDO	ESI Positive	-0.17	0.03	0.4	Seafood
M979.6094T11.29	ENDO	ESI Positive	-0.16	0.04	0.4	Seafood
M979.7207T11.29	ENDO	ESI Positive	-0.16	0.04	0.4	Seafood
M979.8322T11.29	ENDO	ESI Positive	-0.16	0.048	0.4	Seafood
M979.9436T11.29	ENDO	ESI Positive	-0.17	0.04	0.4	Seafood
M980.0549T11.29	ENDO	ESI Positive	-0.16	0.048	0.4	Seafood
M980.1663T11.29	ENDO	ESI Positive	-0.16	0.049	0.4	Seafood
M983.3032T0.78	EXO	ESI Positive	-0.17	0.04	0.5	Seafood
M983.8026T0.78	EXO	ESI Positive	-0.19	0.02	0.4	Seafood
M986.0555T11.26	ENDO	ESI Positive	-0.16	0.04	0.4	Seafood
M996.8228T0.78	EXO	ESI Negative	0.17	0.03	0.2	Seafood
M997.325T0.78	EXO	ESI Negative	0.22	0.005	0.2	Seafood
M997.8257T0.78	EXO	ESI Negative	0.25	0.002	0.09	Seafood
M1002.7992T0.78	ENDO	ESI Positive	-0.17	0.04	0.4	Seafood
M1003.1626T15.2	ENDO	ESI Positive	0.18	0.03	0.4	Seafood
M73.0284T1.02	ENDO	ESI Positive	-0.17	0.04	0.98	Sweetened beverages
M86.0964T1.18	EXO	ESI Positive	-0.17	0.04	0.995	Sweetened beverages
M92.9288T18.04	ENDO	ESI Negative	0.16	0.04	0.95	Sweetened beverages
M93.0115T0.68	EXO	ESI Positive	0.19	0.02	0.995	Sweetened beverages
M93.0344T10.36	EXO	ESI Negative	0.18	0.02	0.6	Sweetened beverages
M100.0757T1.17	ENDO	ESI Positive	-0.18	0.03	0.98	Sweetened beverages
M112.9855T15.63	EXO	ESI Negative	0.17	0.04	0.6	Sweetened beverages
M114.0548T1.17	ENDO	ESI Positive	-0.19	0.02	0.98	Sweetened beverages
M126.0219T0.84	EXO	ESI Positive	0.18	0.03	0.995	Sweetened beverages
M129.0395T0.67	EXO	ESI Positive	0.22	0.007	0.995	Sweetened beverages
M130.1589T1.18	EXO	ESI Positive	-0.22	0.008	0.995	Sweetened beverages
M138.8751T0.92	EXO	ESI Negative	-0.17	0.04	0.6	Sweetened beverages
M142.9289T21.76	EXO	ESI Negative	-0.18	0.03	0.6	Sweetened beverages
M144.1018T0.96	EXO	ESI Positive	0.20	0.02	0.995	Sweetened beverages
M144.1018T0.99	ENDO	ESI Positive	0.20	0.01	0.98	Sweetened beverages
M148.0968T1.09	EXO	ESI Positive	0.18	0.03	0.995	Sweetened beverages
M152.0705T7.13	EXO	ESI Positive	0.18	0.03	0.995	Sweetened beverages
M157.863T0.99	EXO	ESI Negative	-0.16	0.048	0.6	Sweetened beverages
M158.9785T0.7	EXO	ESI Negative	0.18	0.03	0.6	Sweetened beverages
M160.0348T0.92	EXO	ESI Positive	-0.17	0.04	0.995	Sweetened beverages
M160.0968T0.88	EXO	ESI Positive	0.21	0.01	0.995	Sweetened beverages
M161.9694T16.61	ENDO	ESI Positive	0.16	0.047	0.98	Sweetened beverages
M175.1189T0.95	EXO	ESI Positive	0.21	0.009	0.995	Sweetened beverages
M181.072T6.96	ENDO	ESI Positive	0.17	0.03	0.98	Sweetened beverages
M181.072T7	EXO	ESI Positive	0.17	0.04	0.995	Sweetened beverages
M192.9288T14.7	ENDO	ESI Negative	0.17	0.04	0.95	Sweetened beverages
M198.0849T0.86	EXO	ESI Positive	-0.16	0.048	0.995	Sweetened beverages
M198.9364T0.68	ENDO	ESI Negative	-0.18	0.02	0.95	Sweetened beverages
M200.0407T0.8	ENDO	ESI Positive	-0.16	0.04	0.98	Sweetened beverages
M202.9977T1	EXO	ESI Negative	0.25	0.002	0.2	Sweetened beverages
M209.1899T13.82	EXO	ESI Positive	-0.16	0.049	0.995	Sweetened beverages

## ANNEXE 6

M219.1418T1.18	EXO	ESI Positive	-0.17	0.03	0.995	Sweetened beverages
M229.1546T1.17	EXO	ESI Positive	-0.18	0.03	0.995	Sweetened beverages
M230.144T9.36	EXO	ESI Positive	-0.19	0.02	0.995	Sweetened beverages
M234.8199T19.71	EXO	ESI Negative	-0.17	0.04	0.6	Sweetened beverages
M246.17T7.87	ENDO	ESI Positive	-0.20	0.01	0.98	Sweetened beverages
M249.1557T0.9	EXO	ESI Positive	0.16	0.047	0.995	Sweetened beverages
M251.9927T1.16	EXO	ESI Positive	0.17	0.04	0.995	Sweetened beverages
M261.1445T1.17	EXO	ESI Positive	-0.22	0.007	0.995	Sweetened beverages
M267.1965T15.6	EXO	ESI Negative	0.20	0.01	0.5	Sweetened beverages
M274.1833T11.09	EXO	ESI Positive	-0.16	0.049	0.995	Sweetened beverages
M279.1702T8.61	EXO	ESI Positive	0.16	0.046	0.995	Sweetened beverages
M281.2486T18.96	EXO	ESI Negative	0.17	0.04	0.6	Sweetened beverages
M283.2643T0.75	EXO	ESI Negative	0.21	0.01	0.5	Sweetened beverages
M289.6593T15.59	ENDO	ESI Positive	0.20	0.01	0.98	Sweetened beverages
M304.2304T12.33	EXO	ESI Positive	-0.21	0.01	0.995	Sweetened beverages
M328.2483T11.29	EXO	ESI Positive	-0.16	0.048	0.995	Sweetened beverages
M328.2483T11.32	ENDO	ESI Positive	-0.16	0.047	0.98	Sweetened beverages
M329.2485T18.05	EXO	ESI Negative	0.19	0.02	0.6	Sweetened beverages
M335.1512T17.45	ENDO	ESI Positive	0.19	0.02	0.98	Sweetened beverages
M338.3417T21.15	ENDO	ESI Positive	0.19	0.02	0.98	Sweetened beverages
M340.3938T15.88	ENDO	ESI Positive	-0.20	0.02	0.98	Sweetened beverages
M369.174T13.75	EXO	ESI Negative	0.16	0.049	0.6	Sweetened beverages
M391.9529T0.81	ENDO	ESI Positive	0.17	0.04	0.98	Sweetened beverages
M402.9157T0.7	EXO	ESI Negative	0.18	0.03	0.6	Sweetened beverages
M429.0622T0.85	EXO	ESI Positive	0.18	0.02	0.995	Sweetened beverages
M453.3686T13.24	EXO	ESI Positive	0.19	0.02	0.995	Sweetened beverages
M456.3318T12.03	EXO	ESI Positive	-0.19	0.02	0.995	Sweetened beverages
M458.3474T12.43	EXO	ESI Positive	-0.18	0.03	0.995	Sweetened beverages
M459.9403T0.81	EXO	ESI Positive	0.19	0.02	0.995	Sweetened beverages
M479.2838T24.53	EXO	ESI Negative	-0.17	0.04	0.6	Sweetened beverages
M480.3095T16.61	ENDO	ESI Negative	0.18	0.02	0.95	Sweetened beverages
M524.2782T14.58	ENDO	ESI Negative	0.27	0.0008	0.2	Sweetened beverages
M614.6474T11.3	ENDO	ESI Positive	-0.18	0.02	0.98	Sweetened beverages
M615.3955T0.81	ENDO	ESI Positive	0.27	0.0007	0.4	Sweetened beverages
M628.1109T11.31	ENDO	ESI Positive	-0.17	0.04	0.98	Sweetened beverages
M663.9023T0.81	ENDO	ESI Positive	0.16	0.045	0.98	Sweetened beverages
M680.4231T15.6	EXO	ESI Positive	-0.16	0.049	0.995	Sweetened beverages
M681.2546T15.24	EXO	ESI Negative	0.17	0.04	0.6	Sweetened beverages
M683.3829T0.81	ENDO	ESI Positive	0.17	0.03	0.98	Sweetened beverages
M697.8492T0.77	ENDO	ESI Positive	-0.17	0.03	0.98	Sweetened beverages
M724.8724T0.77	EXO	ESI Negative	0.19	0.02	0.6	Sweetened beverages
M731.8898T0.81	ENDO	ESI Positive	0.18	0.03	0.98	Sweetened beverages
M736.2589T15.62	EXO	ESI Negative	0.18	0.03	0.6	Sweetened beverages
M767.3557T0.78	EXO	ESI Negative	0.18	0.02	0.6	Sweetened beverages
M792.8599T0.77	EXO	ESI Negative	0.29	0.0003	0.1	Sweetened beverages
M801.3495T0.78	EXO	ESI Negative	0.21	0.008	0.5	Sweetened beverages

## ANNEXE 6

M860.8476T0.77	EXO	ESI Negative	0.25	0.002	0.2	Sweetened beverages
M861.5023T0.77	EXO	ESI Positive	0.18	0.02	0.995	Sweetened beverages
M861.8508T0.77	EXO	ESI Negative	0.22	0.006	0.5	Sweetened beverages
M868.5027T10.07	ENDO	ESI Positive	0.17	0.04	0.98	Sweetened beverages
M893.4815T24.78	EXO	ESI Negative	0.20	0.02	0.6	Sweetened beverages
M929.8382T0.77	EXO	ESI Negative	0.20	0.01	0.5	Sweetened beverages
M946.9313T0.75	ENDO	ESI Negative	-0.25	0.002	0.2	Sweetened beverages
M952.8789T0.77	ENDO	ESI Negative	-0.16	0.047	0.95	Sweetened beverages
M962.8295T0.78	EXO	ESI Negative	0.22	0.007	0.5	Sweetened beverages
M963.3313T0.78	EXO	ESI Negative	0.17	0.04	0.6	Sweetened beverages
M978.8865T0.77	ENDO	ESI Negative	-0.18	0.02	0.95	Sweetened beverages
M996.8228T0.78	EXO	ESI Negative	0.17	0.03	0.6	Sweetened beverages
M997.8257T0.78	EXO	ESI Negative	0.18	0.02	0.6	Sweetened beverages
M68.5261T0.67	EXO	ESI Positive	-0.19	0.02	0.98	Sweetened food
M84.0808T1.17 (Pipelic acid)	ENDO	ESI Positive	-0.16	0.04	0.97	Sweetened food
M100.0759T20.06	EXO	ESI Positive	0.21	0.01	0.98	Sweetened food
M103.9201T8.29	EXO	ESI Negative	-0.18	0.03	0.99	Sweetened food
M104.1071T15.21	ENDO	ESI Positive	0.19	0.02	0.97	Sweetened food
M110.0087T0.67	ENDO	ESI Positive	-0.20	0.01	0.97	Sweetened food
M129.0556T8.28	EXO	ESI Negative	-0.21	0.01	0.99	Sweetened food
M130.0861T1.16 (Pipelic acid)	ENDO	ESI Positive	-0.17	0.04	0.97	Sweetened food
M131.0895T1.16 (Pipelic acid)	ENDO	ESI Positive	-0.19	0.02	0.97	Sweetened food
M136.8774T0.94	ENDO	ESI Negative	-0.16	0.04	0.99	Sweetened food
M138.8751T0.92	EXO	ESI Negative	-0.16	0.04	0.99	Sweetened food
M144.1018T0.96	EXO	ESI Positive	-0.19	0.02	0.98	Sweetened food
M144.1018T0.99	ENDO	ESI Positive	-0.21	0.009	0.97	Sweetened food
M148.0604T0.86	EXO	ESI Positive	0.21	0.009	0.98	Sweetened food
M148.0967T1.16	ENDO	ESI Positive	-0.18	0.02	0.97	Sweetened food
M156.8776T0.87	ENDO	ESI Negative	-0.18	0.03	0.99	Sweetened food
M157.0118T23.44	EXO	ESI Negative	-0.17	0.03	0.99	Sweetened food
M170.9413T9.44	EXO	ESI Negative	-0.18	0.02	0.99	Sweetened food
M173.9261T11.85	EXO	ESI Negative	0.16	0.04	0.99	Sweetened food
M180.0655T8.29	EXO	ESI Positive	-0.18	0.03	0.98	Sweetened food
M185.1284T7.08	ENDO	ESI Positive	0.19	0.02	0.97	Sweetened food
M185.1284T7.11	EXO	ESI Positive	0.24	0.003	0.98	Sweetened food
M194.9242T14.96	EXO	ESI Negative	0.22	0.007	0.99	Sweetened food
M199.1441T20.43	EXO	ESI Positive	0.23	0.005	0.98	Sweetened food
M214.1801T11.62	EXO	ESI Positive	-0.18	0.03	0.98	Sweetened food
M218.9823T0.68	ENDO	ESI Negative	-0.17	0.03	0.99	Sweetened food
M225.1861T16.68	ENDO	ESI Negative	-0.17	0.03	0.99	Sweetened food
M229.1182T0.97	EXO	ESI Positive	0.17	0.04	0.98	Sweetened food
M230.0958T1.03	ENDO	ESI Positive	-0.18	0.03	0.97	Sweetened food
M231.1452T0.89	ENDO	ESI Positive	-0.19	0.02	0.97	Sweetened food
M248.9989T0.84	ENDO	ESI Positive	-0.18	0.02	0.97	Sweetened food
M258.8993T0.79	EXO	ESI Positive	-0.17	0.04	0.98	Sweetened food

## ANNEXE 6

M261.0112T1.05	EXO	ESI Positive	0.20	0.02	0.98	Sweetened food
M262.1649T1.17	EXO	ESI Positive	-0.20	0.02	0.98	Sweetened food
M271.1517T9.22	EXO	ESI Positive	-0.18	0.03	0.98	Sweetened food
M295.1584T7.34	EXO	ESI Positive	-0.17	0.04	0.98	Sweetened food
M309.207T14.52	EXO	ESI Negative	-0.16	0.048	0.99	Sweetened food
M311.2227T15.13	EXO	ESI Negative	-0.18	0.02	0.99	Sweetened food
M312.2261T15.13	EXO	ESI Negative	-0.20	0.01	0.99	Sweetened food
M312.2533T13.07	ENDO	ESI Positive	0.20	0.01	0.97	Sweetened food
M316.1114T0.87	EXO	ESI Positive	0.18	0.03	0.98	Sweetened food
M332.2796T12.06	ENDO	ESI Positive	0.17	0.04	0.97	Sweetened food
M341.3162T15.4	ENDO	ESI Positive	0.19	0.02	0.97	Sweetened food
<b>M373.2736T12.81</b> <b>(Cholic acid)</b>	EXO	ESI Positive	-0.31	0.0001	0.09	Sweetened food
M440.8888T0.7	ENDO	ESI Negative	-0.18	0.03	0.99	Sweetened food
M450.3213T12.92	EXO	ESI Positive	-0.21	0.01	0.98	Sweetened food
M460.7351T8.63	EXO	ESI Positive	-0.19	0.02	0.98	Sweetened food
M466.3161T11.74	EXO	ESI Positive	-0.29	0.0003	0.1	Sweetened food
M475.1651T11.22	EXO	ESI Positive	-0.17	0.03	0.98	Sweetened food
M490.2903T13.82	EXO	ESI Positive	0.18	0.03	0.98	Sweetened food
M510.4363T17.54	ENDO	ESI Positive	0.21	0.01	0.97	Sweetened food
M518.3241T13.98	EXO	ESI Positive	-0.16	0.046	0.98	Sweetened food
M566.3366T14.69	ENDO	ESI Negative	-0.16	0.04	0.99	Sweetened food
M568.3397T14.63	EXO	ESI Positive	-0.17	0.04	0.98	Sweetened food
M576.26T14.67	EXO	ESI Positive	-0.20	0.01	0.98	Sweetened food
M578.2545T15.21	ENDO	ESI Positive	0.17	0.04	0.97	Sweetened food
M620.4364T15.65	EXO	ESI Positive	-0.19	0.02	0.98	Sweetened food
M622.8564T0.68	ENDO	ESI Negative	-0.16	0.046	0.99	Sweetened food
M626.2964T16.91	EXO	ESI Positive	-0.17	0.04	0.98	Sweetened food
M688.9975T0.75	EXO	ESI Negative	0.16	0.04	0.99	Sweetened food
M766.8543T0.78	ENDO	ESI Negative	0.17	0.04	0.99	Sweetened food
M779.8398T0.78	EXO	ESI Positive	0.18	0.03	0.98	Sweetened food
M861.5023T0.77	EXO	ESI Positive	0.17	0.03	0.98	Sweetened food
M868.5027T10.07	ENDO	ESI Positive	0.17	0.04	0.97	Sweetened food
M888.889T0.76	EXO	ESI Positive	0.21	0.01	0.98	Sweetened food
M918.3358T10.11	ENDO	ESI Positive	0.16	0.047	0.97	Sweetened food
M986.2788T11.25	ENDO	ESI Positive	0.18	0.02	0.97	Sweetened food
M79.9572T9.41	ENDO	ESI Negative	-0.18	0.03	0.96	Total added fat
M84.0444T0.85	ENDO	ESI Positive	0.19	0.02	0.9	Total added fat
M85.0285T0.88	ENDO	ESI Positive	0.18	0.03	0.9	Total added fat
M89.0022T15.11	EXO	ESI Positive	-0.16	0.049	0.996	Total added fat
M90.0548T0.85	ENDO	ESI Positive	0.16	0.04	0.9	Total added fat
M94.9249T1.18	EXO	ESI Negative	0.16	0.04	0.998	Total added fat
M97.0287T0.86	EXO	ESI Positive	0.16	0.046	0.996	Total added fat
M101.071T0.84	ENDO	ESI Positive	0.21	0.01	0.9	Total added fat
M102.0551T1.17	ENDO	ESI Positive	-0.18	0.03	0.9	Total added fat
M129.0395T0.67	EXO	ESI Positive	0.19	0.02	0.996	Total added fat
M130.0861T1.16	ENDO	ESI Positive	-0.18	0.03	0.9	Total added fat

## ANNEXE 6

(Pipelic acid)

M132.9812T0.68	ENDO	ESI Negative	-0.16	0.04	0.96	Total added fat
M137.0457T1.17	ENDO	ESI Positive	-0.17	0.04	0.9	Total added fat
M147.0763T0.85	ENDO	ESI Positive	0.17	0.03	0.9	Total added fat
M158.9612T14.36	EXO	ESI Positive	0.17	0.04	0.996	Total added fat
M165.0509T0.67	EXO	ESI Positive	0.16	0.0498	0.996	Total added fat
M166.1027T5.81	EXO	ESI Positive	0.20	0.01	0.996	Total added fat
M180.1593T9.72	EXO	ESI Positive	0.17	0.03	0.996	Total added fat
M184.9855T0.63	ENDO	ESI Positive	0.16	0.04	0.9	Total added fat
M215.1254T18.9	EXO	ESI Positive	-0.23	0.006	0.996	Total added fat
M239.9665T0.68	ENDO	ESI Positive	-0.16	0.049	0.9	Total added fat
M257.9773T0.68	ENDO	ESI Positive	-0.17	0.04	0.9	Total added fat
M279.1601T13.84	EXO	ESI Negative	-0.20	0.01	0.998	Total added fat
M283.0076T8.88	ENDO	ESI Positive	-0.17	0.04	0.9	Total added fat
M284.9108T14.58	EXO	ESI Positive	0.18	0.03	0.996	Total added fat
M324.2744T13.88	EXO	ESI Positive	0.18	0.03	0.996	Total added fat
M325.1129T0.86	EXO	ESI Positive	0.21	0.01	0.996	Total added fat
M325.1129T0.87	ENDO	ESI Positive	0.17	0.04	0.9	Total added fat
M327.1268T17.68	ENDO	ESI Negative	-0.16	0.046	0.96	Total added fat
M328.2483T11.29	EXO	ESI Positive	-0.18	0.03	0.996	Total added fat
M330.246T12.7	EXO	ESI Positive	0.23	0.005	0.996	Total added fat
M358.2774T13.44	EXO	ESI Positive	0.16	0.0497	0.996	Total added fat
M359.1513T17.27	EXO	ESI Positive	-0.16	0.047	0.996	Total added fat
M361.1251T16.91	EXO	ESI Positive	0.17	0.04	0.996	Total added fat
M383.116T0.87	ENDO	ESI Positive	0.18	0.02	0.9	Total added fat
M394.8907T0.75	ENDO	ESI Negative	0.19	0.02	0.96	Total added fat
M445.3674T20.1	EXO	ESI Positive	0.18	0.03	0.996	Total added fat
M452.2782T15.1	ENDO	ESI Negative	-0.17	0.04	0.96	Total added fat
M464.3134T15.85	EXO	ESI Positive	0.17	0.04	0.996	Total added fat
M467.3742T18.72	EXO	ESI Negative	0.17	0.03	0.998	Total added fat
M469.3887T18.73	EXO	ESI Positive	0.19	0.02	0.996	Total added fat
M479.2838T24.53	EXO	ESI Negative	-0.22	0.006	0.998	Total added fat
M482.3241T14.5	ENDO	ESI Positive	-0.17	0.04	0.9	Total added fat
M489.3586T16.39	EXO	ESI Negative	0.22	0.006	0.998	Total added fat
M508.3396T14.88	ENDO	ESI Positive	-0.18	0.03	0.9	Total added fat
M517.3897T17.32	EXO	ESI Negative	0.21	0.01	0.998	Total added fat
M518.3239T13.99	ENDO	ESI Positive	-0.17	0.04	0.9	Total added fat
M527.2976T15.2	ENDO	ESI Positive	0.20	0.01	0.9	Total added fat
M532.8573T1.17	ENDO	ESI Positive	-0.17	0.03	0.9	Total added fat
M540.366T15.45	EXO	ESI Positive	-0.17	0.04	0.996	Total added fat
M546.3553T15.18	ENDO	ESI Positive	-0.18	0.03	0.9	Total added fat
M548.3709T15.96	ENDO	ESI Positive	-0.17	0.04	0.9	Total added fat
M610.2453T14.22	EXO	ESI Negative	-0.16	0.046	0.998	Total added fat
M640.4916T11.26	EXO	ESI Positive	0.20	0.02	0.996	Total added fat
M643.8647T0.78	EXO	ESI Positive	-0.18	0.03	0.996	Total added fat
M677.3592T0.78	EXO	ESI Positive	-0.17	0.03	0.996	Total added fat

## ANNEXE 6

M685.2535T15.23	ENDO	ESI Negative	-0.20	0.01	0.96	Total added fat
M699.7319T15.22	EXO	ESI Positive	-0.26	0.001	0.996	Total added fat
M711.2696T15.62	EXO	ESI Negative	-0.17	0.03	0.998	Total added fat
M842.8285T0.78	EXO	ESI Negative	-0.20	0.01	0.998	Total added fat
M847.3277T0.78	EXO	ESI Positive	-0.18	0.03	0.996	Total added fat
M888.8067T0.78	EXO	ESI Positive	-0.17	0.04	0.996	Total added fat
M896.8464T0.78	EXO	ESI Negative	-0.19	0.02	0.998	Total added fat
M945.3942T0.77	ENDO	ESI Negative	-0.16	0.047	0.96	Total added fat
M975.8168T0.77	EXO	ESI Positive	-0.17	0.04	0.996	Total added fat
M980.5003T11.29	ENDO	ESI Positive	-0.18	0.03	0.9	Total added fat
M987.2941T0.8	EXO	ESI Negative	0.16	0.049	0.998	Total added fat
M61.0104T1.16	EXO	ESI Positive	-0.16	0.047	0.9	Vegetable added fat
M71.0489T0.87	EXO	ESI Positive	0.17	0.03	0.8	Vegetable added fat
M76.0391T1.15	EXO	ESI Positive	0.17	0.04	0.8	Vegetable added fat
M76.0393T1.17	ENDO	ESI Positive	0.19	0.02	0.5	Vegetable added fat
M82.9447T0.94	EXO	ESI Positive	-0.20	0.01	0.8	Vegetable added fat
M84.0807T0.98	EXO	ESI Positive	0.22	0.007	0.8	Vegetable added fat
M84.0808T1.17 (Pipelic acid)	ENDO	ESI Positive	0.20	0.01	0.5	Vegetable added fat
M89.0597T11.31	ENDO	ESI Positive	-0.18	0.02	0.5	Vegetable added fat
M97.0658T11.79	EXO	ESI Negative	-0.18	0.03	0.6	Vegetable added fat
M98.9842T1.08	ENDO	ESI Positive	-0.16	0.047	0.6	Vegetable added fat
M104.1072T0.83	ENDO	ESI Positive	-0.18	0.03	0.5	Vegetable added fat
M106.05T0.83	ENDO	ESI Positive	0.17	0.04	0.5	Vegetable added fat
M114.0659T0.88	EXO	ESI Positive	-0.19	0.02	0.8	Vegetable added fat
M118.0496T1.17	ENDO	ESI Positive	0.19	0.02	0.5	Vegetable added fat
M126.0219T0.84	EXO	ESI Positive	-0.19	0.02	0.8	Vegetable added fat
M130.0861T1.16 (Pipelic acid)	ENDO	ESI Positive	0.20	0.01	0.5	Vegetable added fat
M130.0862T0.98	EXO	ESI Positive	0.21	0.01	0.8	Vegetable added fat
M131.0895T0.98	EXO	ESI Positive	0.19	0.02	0.8	Vegetable added fat
M131.0895T1.16 (Pipelic acid)	ENDO	ESI Positive	0.21	0.01	0.5	Vegetable added fat
M147.1127T0.72	ENDO	ESI Positive	-0.18	0.03	0.5	Vegetable added fat
M157.087T17.71	EXO	ESI Negative	-0.23	0.005	0.3	Vegetable added fat
M158.1175T1.17	ENDO	ESI Positive	-0.23	0.005	0.5	Vegetable added fat
M159.0927T7.39	EXO	ESI Negative	-0.19	0.02	0.6	Vegetable added fat
M160.1331T1.12	ENDO	ESI Positive	-0.20	0.01	0.5	Vegetable added fat
M164.1173T0.86	ENDO	ESI Positive	-0.16	0.04	0.6	Vegetable added fat
M172.1331T9.04	EXO	ESI Positive	0.16	0.046	0.9	Vegetable added fat
M175.1189T0.8	ENDO	ESI Positive	-0.20	0.01	0.5	Vegetable added fat
M182.0811T1.18	ENDO	ESI Positive	-0.19	0.02	0.5	Vegetable added fat
M187.0577T0.95	EXO	ESI Positive	-0.19	0.02	0.8	Vegetable added fat
M187.0577T0.97 (2-deoxy-D-glucose)	ENDO	ESI Positive	-0.21	0.009	0.5	Vegetable added fat
M189.1234T1.15	ENDO	ESI Positive	-0.20	0.01	0.5	Vegetable added fat
M189.1597T0.79	ENDO	ESI Positive	-0.18	0.03	0.5	Vegetable added fat
M196.0604T9.19	EXO	ESI Positive	-0.17	0.04	0.8	Vegetable added fat



## ANNEXE 6

M197.1009T0.81	ENDO	ESI Positive	-0.21	0.009	0.5	Vegetable added fat
M203.139T1.16	ENDO	ESI Positive	-0.18	0.03	0.5	Vegetable added fat
M206.1057T17.51	EXO	ESI Positive	0.22	0.007	0.8	Vegetable added fat
M223.1078T7.19	ENDO	ESI Positive	-0.19	0.02	0.5	Vegetable added fat
M227.1251T0.87	EXO	ESI Positive	-0.19	0.02	0.8	Vegetable added fat
M229.1412T9.33	EXO	ESI Positive	-0.18	0.03	0.8	Vegetable added fat
M231.1451T0.88	EXO	ESI Positive	0.20	0.02	0.8	Vegetable added fat
M239.1652T16.29	EXO	ESI Negative	-0.23	0.005	0.3	Vegetable added fat
M248.1677T10.5	EXO	ESI Positive	0.17	0.04	0.8	Vegetable added fat
M252.0874T18.11	EXO	ESI Positive	0.21	0.01	0.8	Vegetable added fat
M268.1754T1.16	ENDO	ESI Positive	-0.19	0.02	0.5	Vegetable added fat
M269.2121T17.89	EXO	ESI Negative	-0.21	0.009	0.4	Vegetable added fat
M274.1222T14.21	EXO	ESI Positive	-0.18	0.03	0.8	Vegetable added fat
M275.6257T14.21	EXO	ESI Positive	-0.21	0.01	0.8	Vegetable added fat
M276.1264T8.82	ENDO	ESI Positive	-0.22	0.005	0.5	Vegetable added fat
M278.9822T10.34	EXO	ESI Positive	-0.17	0.04	0.8	Vegetable added fat
M282.9049T0.72	EXO	ESI Positive	-0.21	0.01	0.8	Vegetable added fat
M291.6097T14.58	ENDO	ESI Positive	-0.16	0.048	0.6	Vegetable added fat
M299.1448T16.6	ENDO	ESI Positive	-0.19	0.02	0.5	Vegetable added fat
M300.1992T11.56	EXO	ESI Positive	0.24	0.003	0.6	Vegetable added fat
M302.2325T10.49 (2,6 dimethylheptanoyl carnitine/ nonanoyl-l-carnitine; acylcarnitine C9:1)	EXO	ESI Positive	-0.27	0.001	0.3	Vegetable added fat
M302.3054T13.18	EXO	ESI Positive	-0.19	0.02	0.8	Vegetable added fat
M303.2357T10.49 (2,6 dimethylheptanoyl carnitine/ nonanoyl-l-carnitine; acylcarnitine C9:1)	EXO	ESI Positive	-0.28	0.0005	0.2	Vegetable added fat
M313.1546T8.83	EXO	ESI Positive	-0.21	0.01	0.8	Vegetable added fat
M313.2383T17.62	EXO	ESI Negative	0.17	0.04	0.6	Vegetable added fat
M315.6554T15.26	EXO	ESI Positive	0.18	0.03	0.8	Vegetable added fat
M327.0152T0.82	ENDO	ESI Positive	0.17	0.03	0.5	Vegetable added fat
M327.0516T0.76	ENDO	ESI Positive	-0.16	0.04	0.6	Vegetable added fat
M332.0698T11.8	EXO	ESI Positive	0.18	0.03	0.8	Vegetable added fat
M338.5964T14.68	EXO	ESI Positive	0.18	0.03	0.8	Vegetable added fat
M348.0646T11.7	EXO	ESI Positive	0.18	0.03	0.8	Vegetable added fat
M356.2618T12.94	EXO	ESI Positive	0.22	0.007	0.8	Vegetable added fat
M363.2166T10.81	ENDO	ESI Positive	-0.18	0.03	0.5	Vegetable added fat
M365.2697T16.47	EXO	ESI Negative	0.18	0.02	0.6	Vegetable added fat
M367.121T0.97	ENDO	ESI Positive	-0.17	0.04	0.6	Vegetable added fat
M379.2486T9.19	ENDO	ESI Positive	-0.16	0.049	0.6	Vegetable added fat
M391.2853T17.42	EXO	ESI Negative	0.18	0.03	0.6	Vegetable added fat
M396.3107T12.94	EXO	ESI Positive	0.17	0.03	0.8	Vegetable added fat
M397.2052T15.61	EXO	ESI Negative	0.18	0.03	0.6	Vegetable added fat
M399.1778T16.19	EXO	ESI Positive	-0.19	0.02	0.8	Vegetable added fat
M401.7632T9.27	ENDO	ESI Positive	-0.16	0.049	0.6	Vegetable added fat
M424.342T13.69	EXO	ESI Positive	0.16	0.049	0.9	Vegetable added fat
M429.0622T0.85	EXO	ESI Positive	0.19	0.02	0.8	Vegetable added fat

## ANNEXE 6

M445.3674T20.1	EXO	ESI Positive	0.17	0.04	0.8	Vegetable added fat
M452.2771T14.11	EXO	ESI Positive	-0.24	0.004	0.6	Vegetable added fat
M452.2782T15.06	EXO	ESI Negative	-0.18	0.03	0.6	Vegetable added fat
M456.3318T12.03	EXO	ESI Positive	0.18	0.03	0.8	Vegetable added fat
M463.2325T13.14	EXO	ESI Positive	0.25	0.002	0.4	Vegetable added fat
M468.3084T13.84	ENDO	ESI Positive	-0.16	0.04	0.6	Vegetable added fat
M479.2838T24.53	EXO	ESI Negative	0.26	0.001	0.3	Vegetable added fat
M482.3241T14.5	ENDO	ESI Positive	-0.19	0.02	0.5	Vegetable added fat
M482.324T14.49	EXO	ESI Positive	-0.17	0.04	0.8	Vegetable added fat
M494.3241T14.23	ENDO	ESI Positive	-0.24	0.003	0.5	Vegetable added fat
M495.3275T14.22	EXO	ESI Positive	-0.20	0.02	0.8	Vegetable added fat
M496.8398T0.91	EXO	ESI Positive	0.18	0.03	0.8	Vegetable added fat
<b>M508.3395T15.27 (LysoPC (17:1))</b>	EXO	ESI Positive	-0.30	0.0002	0.2	vegetable added fat
M508.3397T14.88	EXO	ESI Positive	-0.21	0.009	0.8	Vegetable added fat
M524.2782T14.58	ENDO	ESI Negative	-0.17	0.04	0.9	Vegetable added fat
M526.3149T14.5	EXO	ESI Negative	-0.18	0.03	0.6	Vegetable added fat
M538.315T14.23	EXO	ESI Negative	-0.29	0.0003	0.1	Vegetable added fat
M552.3306T14.9	EXO	ESI Negative	-0.23	0.004	0.3	Vegetable added fat
M552.3308T15.29	EXO	ESI Negative	-0.24	0.003	0.3	Vegetable added fat
M553.3339T14.95	EXO	ESI Negative	-0.21	0.008	0.4	Vegetable added fat
M554.3464T15.84	EXO	ESI Negative	-0.19	0.02	0.6	Vegetable added fat
M567.2254T13.82	EXO	ESI Positive	-0.17	0.03	0.8	Vegetable added fat
M570.3553T15.34	EXO	ESI Positive	-0.18	0.03	0.8	Vegetable added fat
M590.3463T15.2	EXO	ESI Negative	-0.17	0.03	0.6	Vegetable added fat
M595.9149T0.81	ENDO	ESI Positive	0.18	0.02	0.5	Vegetable added fat
M596.7595T14.67	EXO	ESI Positive	0.16	0.048	0.9	Vegetable added fat
M601.8855T0.77	EXO	ESI Positive	0.17	0.04	0.8	Vegetable added fat
M602.2229T15.26	EXO	ESI Positive	-0.17	0.04	0.8	Vegetable added fat
M616.3618T15.74	EXO	ESI Negative	-0.17	0.03	0.6	Vegetable added fat
M628.1109T11.31	ENDO	ESI Positive	-0.18	0.02	0.5	Vegetable added fat
M633.687T11.27	ENDO	ESI Positive	-0.20	0.01	0.5	Vegetable added fat
M644.1204T11.32	ENDO	ESI Positive	-0.16	0.04	0.6	Vegetable added fat
M652.4113T8.99	ENDO	ESI Positive	-0.16	0.04	0.6	Vegetable added fat
M652.8T11.28	ENDO	ESI Positive	-0.20	0.01	0.5	Vegetable added fat
M658.8738T0.73	EXO	ESI Positive	0.20	0.02	0.8	Vegetable added fat
M660.1299T11.34	ENDO	ESI Positive	-0.19	0.02	0.5	Vegetable added fat
M681.2546T15.24	EXO	ESI Negative	-0.19	0.02	0.6	Vegetable added fat
M683.6189T11.3	ENDO	ESI Positive	-0.19	0.02	0.5	Vegetable added fat
M692.1222T11.31	ENDO	ESI Positive	-0.21	0.01	0.5	Vegetable added fat
M692.3232T11.31	ENDO	ESI Positive	-0.20	0.01	0.5	Vegetable added fat
M696.4375T9.09	ENDO	ESI Positive	-0.17	0.03	0.5	Vegetable added fat
M704.8057T24.58	ENDO	ESI Negative	0.16	0.04	0.9	Vegetable added fat
M705.7322T11.32	ENDO	ESI Positive	-0.16	0.047	0.6	Vegetable added fat
M710.2435T15.26	EXO	ESI Negative	0.17	0.04	0.6	Vegetable added fat
M740.2331T11.29	ENDO	ESI Positive	-0.18	0.02	0.5	Vegetable added fat
M894.8418T0.78	EXO	ESI Negative	0.19	0.02	0.6	Vegetable added fat

## ANNEXE 6

M944.8925T0.77	ENDO	ESI Negative	-0.16	0.048	0.9	Vegetable added fat
M954.1525T11.32	ENDO	ESI Positive	-0.18	0.03	0.5	Vegetable added fat
M963.3313T0.78	EXO	ESI Negative	0.20	0.01	0.6	Vegetable added fat
M58.0649T0.91	EXO	ESI Positive	0.20	0.01	0.5	Whole grain food
M62.0428T0.9	EXO	ESI Positive	0.21	0.009	0.5	Whole grain food
M76.0391T1.15	EXO	ESI Positive	0.18	0.03	0.5	Whole grain food
M80.9478T0.85	ENDO	ESI Positive	-0.17	0.03	0.3	Whole grain food
M84.0444T0.85	ENDO	ESI Positive	0.22	0.005	0.2	Whole grain food
M84.0807T0.98	EXO	ESI Positive	0.23	0.005	0.5	Whole grain food
M84.0808T1.17 (Pipelic acid)	ENDO	ESI Positive	0.29	0.0004	0.2	Whole grain food
M89.0022T15.11	EXO	ESI Positive	-0.20	0.01	0.5	Whole grain food
M101.0243T0.99	EXO	ESI Negative	0.19	0.02	0.7	Whole grain food
M103.9201T11.85	EXO	ESI Negative	0.16	0.04	0.7	Whole grain food
M114.0659T0.88	EXO	ESI Positive	-0.25	0.002	0.4	Whole grain food
M115.0691T0.87	EXO	ESI Positive	-0.16	0.049	0.6	Whole grain food
M118.0497T1.15	EXO	ESI Positive	0.18	0.02	0.5	Whole grain food
M118.0861T11.29	EXO	ESI Positive	-0.17	0.04	0.6	Whole grain food
M118.0861T14.68	EXO	ESI Positive	-0.18	0.03	0.5	Whole grain food
M119.0491T1.17	ENDO	ESI Positive	-0.16	0.048	0.4	Whole grain food
M120.0654T0.85	ENDO	ESI Positive	0.19	0.02	0.2	Whole grain food
M122.9245T0.8	ENDO	ESI Positive	-0.19	0.02	0.3	Whole grain food
M130.0861T1.16 (Pipelic acid)	ENDO	ESI Positive	0.27	0.0007	0.2	Whole grain food
M130.0862T0.98	EXO	ESI Positive	0.25	0.002	0.4	Whole grain food
M131.0895T1.16 (Pipelic acid)	ENDO	ESI Positive	0.26	0.001	0.2	Whole grain food
M132.1018T1.19	ENDO	ESI Positive	-0.16	0.04	0.4	Whole grain food
M133.0971T0.72	ENDO	ESI Positive	0.17	0.04	0.4	Whole grain food
M134.0598T8.53	ENDO	ESI Positive	-0.19	0.02	0.2	Whole grain food
M134.8944T0.9	EXO	ESI Negative	0.24	0.003	0.7	Whole grain food
M144.982T21.82	EXO	ESI Positive	-0.26	0.001	0.4	Whole grain food
M148.0037T0.83	ENDO	ESI Positive	0.19	0.02	0.2	Whole grain food
M148.0426T1.17	ENDO	ESI Positive	-0.18	0.03	0.3	Whole grain food
M150.8865T0.87	EXO	ESI Negative	0.17	0.04	0.7	Whole grain food
M151.0477T0.83	EXO	ESI Positive	0.18	0.03	0.5	Whole grain food
M153.0657T1.19	ENDO	ESI Positive	-0.19	0.02	0.2	Whole grain food
M155.079T0.76	EXO	ESI Positive	0.19	0.02	0.5	Whole grain food
M157.0117T14.7	ENDO	ESI Negative	0.21	0.01	0.7	Whole grain food
M158.1175T1.17	ENDO	ESI Positive	-0.16	0.047	0.4	Whole grain food
M163.0989T8.08	ENDO	ESI Positive	-0.17	0.04	0.4	Whole grain food
M163.0992T8.1	EXO	ESI Positive	-0.17	0.04	0.6	Whole grain food
M169.0583T0.84	ENDO	ESI Positive	0.24	0.003	0.2	Whole grain food
M175.1076T1.01	ENDO	ESI Positive	0.18	0.02	0.3	Whole grain food
M180.973T17.01	EXO	ESI Negative	-0.16	0.048	0.7	Whole grain food
M188.088T7.39	EXO	ESI Positive	0.17	0.04	0.6	Whole grain food
M189.1234T1.15	ENDO	ESI Positive	-0.21	0.01	0.2	Whole grain food
M192.0325T1.11	ENDO	ESI Positive	-0.26	0.001	0.2	Whole grain food

## ANNEXE 6

M192.0435T0.82	EXO	ESI Positive	0.17	0.03	0.6	Whole grain food
M215.1652T17.21	EXO	ESI Negative	-0.17	0.03	0.7	Whole grain food
M223.1702T15.96	EXO	ESI Negative	0.17	0.03	0.7	Whole grain food
M227.1251T0.87	EXO	ESI Positive	-0.16	0.047	0.6	Whole grain food
M231.1451T0.88	EXO	ESI Positive	0.20	0.02	0.5	Whole grain food
M242.9254T0.79	ENDO	ESI Positive	-0.21	0.01	0.2	Whole grain food
M247.1441T7.91	ENDO	ESI Positive	0.18	0.03	0.3	Whole grain food
M247.1441T7.92	EXO	ESI Positive	0.20	0.01	0.5	Whole grain food
M249.1557T0.9	EXO	ESI Positive	0.20	0.01	0.5	Whole grain food
M249.1696T9.21	EXO	ESI Positive	0.22	0.006	0.5	Whole grain food
M258.8993T0.79	ENDO	ESI Positive	-0.21	0.009	0.2	Whole grain food
M267.1237T13.25	EXO	ESI Negative	0.21	0.01	0.7	Whole grain food
M268.1543T12.79	ENDO	ESI Positive	0.17	0.03	0.3	Whole grain food
M274.1222T14.21	EXO	ESI Positive	-0.17	0.04	0.6	Whole grain food
M275.6257T14.21	EXO	ESI Positive	-0.22	0.008	0.5	Whole grain food
M277.2172T17.4	EXO	ESI Negative	0.17	0.04	0.7	Whole grain food
M281.6274T13.82	EXO	ESI Positive	-0.18	0.02	0.5	Whole grain food
M282.1911T1.18	ENDO	ESI Positive	-0.20	0.01	0.2	Whole grain food
M282.9049T0.72	EXO	ESI Positive	-0.20	0.02	0.5	Whole grain food
M283.0076T8.88	ENDO	ESI Positive	-0.20	0.01	0.2	Whole grain food
M283.1087T20.55	EXO	ESI Negative	-0.18	0.02	0.7	Whole grain food
M285.0056T8.88	ENDO	ESI Positive	-0.20	0.01	0.2	Whole grain food
M285.9887T0.81	EXO	ESI Positive	0.20	0.01	0.5	Whole grain food
M287.0703T0.8	EXO	ESI Positive	0.22	0.007	0.5	Whole grain food
M289.1771T15.21	EXO	ESI Positive	-0.20	0.02	0.5	Whole grain food
M291.647T14.69	EXO	ESI Positive	-0.18	0.03	0.6	Whole grain food
M292.6353T11.22	EXO	ESI Positive	-0.20	0.01	0.5	Whole grain food
M295.2278T15.34	EXO	ESI Negative	0.19	0.02	0.7	Whole grain food
M297.1107T13.88	EXO	ESI Positive	0.17	0.04	0.6	Whole grain food
M297.2435T17.36	EXO	ESI Negative	0.19	0.02	0.7	Whole grain food
M299.6254T14.04	ENDO	ESI Positive	0.17	0.04	0.4	Whole grain food
M303.2357T10.49 (2,6 dimethylheptanoyl carnitine/ nonanoyl-l-carnitine; acylcarnitine C9:1)	EXO	ESI Positive	-0.19	0.02	0.5	Whole grain food
M310.3103T19.42	EXO	ESI Positive	0.20	0.02	0.5	Whole grain food
M313.0359T0.76	EXO	ESI Positive	0.20	0.02	0.5	Whole grain food
M315.6554T15.26	EXO	ESI Positive	0.17	0.03	0.6	Whole grain food
M320.9669T0.88	EXO	ESI Positive	0.19	0.02	0.5	Whole grain food
M327.0152T0.82	ENDO	ESI Positive	0.19	0.02	0.2	Whole grain food
M327.2641T12.81	EXO	ESI Positive	0.17	0.04	0.6	Whole grain food
M336.3259T19.67	EXO	ESI Positive	0.22	0.006	0.5	Whole grain food
M338.3415T21.17	EXO	ESI Positive	0.20	0.02	0.5	Whole grain food
M338.5964T14.68	EXO	ESI Positive	0.17	0.04	0.6	Whole grain food
M339.5956T14.69	EXO	ESI Positive	0.19	0.02	0.5	Whole grain food
M355.0578T0.8	EXO	ESI Positive	0.16	0.046	0.6	Whole grain food
M358.2951T12.6	ENDO	ESI Positive	0.17	0.04	0.4	Whole grain food

## ANNEXE 6

M363.2166T10.81	ENDO	ESI Positive	-0.17	0.04	0.4	Whole grain food
M368.2793T12.21	EXO	ESI Positive	0.18	0.03	0.6	Whole grain food
M412.7516T15.63	EXO	ESI Positive	0.19	0.02	0.5	Whole grain food
M431.3155T15.5	EXO	ESI Positive	-0.19	0.02	0.5	Whole grain food
M452.2771T14.11	EXO	ESI Positive	-0.28	0.0006	0.4	Whole grain food
M459.9403T0.81	ENDO	ESI Positive	0.16	0.04	0.4	Whole grain food
M466.3161T11.74	EXO	ESI Positive	0.20	0.02	0.5	Whole grain food
M468.3084T13.84	ENDO	ESI Positive	-0.23	0.005	0.2	Whole grain food
M482.3241T14.5	ENDO	ESI Positive	-0.22	0.005	0.2	Whole grain food
M494.3241T14.23	ENDO	ESI Positive	-0.22	0.005	0.2	Whole grain food
M495.3275T14.22	EXO	ESI Positive	-0.18	0.03	0.5	Whole grain food
M498.2895T13.25	EXO	ESI Negative	0.17	0.04	0.7	Whole grain food
M502.2927T14.62	ENDO	ESI Positive	-0.16	0.045	0.4	Whole grain food
M508.3395T15.27 (PC (17:1))	EXO	ESI Positive	-0.19	0.02	0.5	Whole grain food
M508.3397T14.88	EXO	ESI Positive	-0.16	0.048	0.6	Whole grain food
M532.6168T15.2	ENDO	ESI Positive	-0.16	0.0498	0.4	Whole grain food
M549.3071T15.18	ENDO	ESI Positive	-0.19	0.02	0.3	Whole grain food
M564.3307T14.69	ENDO	ESI Negative	0.16	0.047	0.99	Whole grain food
M564.8301T0.68	ENDO	ESI Negative	0.19	0.02	0.7	Whole grain food
M567.391T0.77	EXO	ESI Positive	0.22	0.006	0.5	Whole grain food
M572.7596T15.22	EXO	ESI Positive	0.22	0.006	0.5	Whole grain food
M574.2624T15.23	EXO	ESI Positive	0.17	0.04	0.6	Whole grain food
M582.9456T0.76	EXO	ESI Positive	0.18	0.03	0.5	Whole grain food
M584.26T20.56	EXO	ESI Positive	-0.17	0.04	0.6	Whole grain food
M595.9149T0.81	ENDO	ESI Positive	0.18	0.03	0.3	Whole grain food
M596.7595T14.67	EXO	ESI Positive	0.21	0.01	0.5	Whole grain food
M615.3955T0.81	ENDO	ESI Positive	0.19	0.02	0.2	Whole grain food
M622.8564T0.68	ENDO	ESI Negative	0.19	0.02	0.7	Whole grain food
M624.3806T0.74	EXO	ESI Positive	0.17	0.04	0.6	Whole grain food
M640.4916T11.26	EXO	ESI Positive	0.17	0.04	0.6	Whole grain food
M666.9885T0.76	EXO	ESI Positive	0.18	0.03	0.6	Whole grain food
M677.5142T11.3	ENDO	ESI Positive	-0.17	0.03	0.3	Whole grain food
M684.9268T0.76	EXO	ESI Positive	0.21	0.01	0.5	Whole grain food
M718.9206T0.76	EXO	ESI Positive	0.22	0.008	0.5	Whole grain food
M731.8898T0.81	ENDO	ESI Positive	0.19	0.02	0.3	Whole grain food
M732.8605T0.78	EXO	ESI Negative	-0.18	0.02	0.7	Whole grain food
M737.8604T0.77	EXO	ESI Positive	0.25	0.002	0.4	Whole grain food
M752.9143T0.76	EXO	ESI Positive	0.24	0.003	0.5	Whole grain food
M758.867T0.78	EXO	ESI Negative	0.23	0.005	0.7	Whole grain food
M773.9667T15.2	ENDO	ESI Positive	-0.18	0.02	0.3	Whole grain food
M786.908T0.76	EXO	ESI Positive	0.17	0.04	0.6	Whole grain food
M792.8599T0.77	EXO	ESI Negative	0.20	0.01	0.7	Whole grain food
M805.8478T0.77	EXO	ESI Positive	0.19	0.02	0.5	Whole grain food
M820.9017T0.76	EXO	ESI Positive	0.21	0.01	0.5	Whole grain food
M844.1646T0.78	EXO	ESI Positive	0.20	0.01	0.5	Whole grain food
M85.0283T0.87	EXO	ESI Positive	-0.17	0.04	0.6	Whole grain food

## ANNEXE 6

M861.3499T0.78	EXO	ESI Negative	0.17	0.04	0.7	Whole grain food
M870.9506T0.76	EXO	ESI Positive	0.21	0.01	0.5	Whole grain food
M873.8352T0.77	EXO	ESI Positive	0.18	0.03	0.5	Whole grain food
M881.548T11.29	ENDO	ESI Positive	-0.20	0.01	0.2	Whole grain food
M881.6492T11.29	ENDO	ESI Positive	-0.23	0.005	0.2	Whole grain food
M881.8495T11.29	ENDO	ESI Positive	-0.24	0.003	0.2	Whole grain food
M881.9498T11.29	ENDO	ESI Positive	-0.20	0.02	0.2	Whole grain food
M882.0499T11.29	ENDO	ESI Positive	-0.20	0.01	0.2	Whole grain food
M882.1503T11.29	ENDO	ESI Positive	-0.24	0.003	0.2	Whole grain food
M887.6517T11.25	ENDO	ESI Positive	-0.17	0.03	0.4	Whole grain food
M888.889T0.76	EXO	ESI Positive	0.16	0.04	0.6	Whole grain food
M889.4895T0.78	EXO	ESI Positive	0.17	0.04	0.6	Whole grain food
M892.9599T0.77	EXO	ESI Negative	0.18	0.03	0.7	Whole grain food
M907.1616T0.77	EXO	ESI Positive	0.23	0.005	0.5	Whole grain food
M922.8825T0.76	EXO	ESI Positive	0.18	0.03	0.6	Whole grain food
M945.3936T0.77	EXO	ESI Negative	0.17	0.03	0.7	Whole grain food
M962.8295T0.78	EXO	ESI Negative	0.16	0.04	0.7	Whole grain food
M963.3313T0.78	EXO	ESI Negative	0.21	0.01	0.7	Whole grain food
M963.8325T0.78	ENDO	ESI Negative	-0.20	0.01	0.7	Whole grain food
M965.2692T11.3	ENDO	ESI Positive	-0.17	0.04	0.4	Whole grain food
M975.8168T0.77	EXO	ESI Positive	0.17	0.04	0.6	Whole grain food
M979.3865T11.29	ENDO	ESI Positive	-0.23	0.005	0.2	Whole grain food
M979.498T11.29	ENDO	ESI Positive	-0.21	0.009	0.2	Whole grain food
M979.6094T11.29	ENDO	ESI Positive	-0.21	0.009	0.2	Whole grain food
M979.7207T11.29	ENDO	ESI Positive	-0.22	0.006	0.2	Whole grain food
M979.8322T11.29	ENDO	ESI Positive	-0.21	0.009	0.2	Whole grain food
M979.9436T11.29	ENDO	ESI Positive	-0.22	0.006	0.2	Whole grain food
M980.0549T11.29	ENDO	ESI Positive	-0.21	0.009	0.2	Whole grain food
M980.1663T11.29	ENDO	ESI Positive	-0.22	0.005	0.2	Whole grain food
M980.2777T11.29	ENDO	ESI Positive	-0.21	0.008	0.2	Whole grain food
M980.3889T11.29	ENDO	ESI Positive	-0.21	0.01	0.2	Whole grain food
M980.5003T11.29	ENDO	ESI Positive	-0.21	0.01	0.2	Whole grain food
M983.3677T0.78	ENDO	ESI Negative	0.22	0.007	0.7	Whole grain food
M985.7148T11.28	ENDO	ESI Positive	-0.21	0.009	0.2	Whole grain food
M985.9431T11.26	ENDO	ESI Positive	-0.20	0.01	0.2	Whole grain food
M986.0555T11.26	ENDO	ESI Positive	-0.21	0.009	0.2	Whole grain food
M986.2788T11.25	ENDO	ESI Positive	-0.18	0.02	0.3	Whole grain food
M986.3902T11.26	ENDO	ESI Positive	-0.22	0.006	0.2	Whole grain food
M986.6128T11.26	ENDO	ESI Positive	-0.24	0.003	0.2	Whole grain food

SU.VI.MAX, Supplémentation en Vitamines et Minéraux Antioxydants; R; correlation coefficient ; M, molecular mass ; T, retention time ; ENDO, endogenous protocol ; Exo, exogenous protocol  
Associations presented in **Table 3** in the manuscript are highlighted in bold.

<sup>1</sup> Partial spearman correlation adjusted for age (continuous), BMI (continuous), menopausal status (pre/post-menopause) and smoking status (current/former/never) at baseline, season at time of blood draw (a priori-defined periods: October/November, December/January/February, and March-April), number of 24h-dietary records and mean of daily energy intake (kcal) during the 2 years following the blood draw.

<sup>2</sup>160 samples for endogenous protocol and exogenous protocol in negative ionization mode and 156 samples for exogenous protocol in positive ionization mode were analyzed.

<sup>3</sup> P-trend after Benjamini Hochberg correction for multiple testing

### Supplemental file 7

**Supplementary Table S6** Associations ( $p < 0.05$ ) between ions from endogenous protocol (“ENDO”) and from exogenous protocol (“EXO”) and *a posteriori* dietary patterns derived by principal component analysis from spearman correlation analyses<sup>1</sup>, SU.VI.MAX cohort, France (1994-2007)

Ions Mass (M). retention time (T) (classified by mass)	Protocol <sup>2</sup>	Mode of detection	R	P-trend	P- corrected <sup>3</sup>	PCA-derived dietary pattern
M84.0444T1.18	ENDO	ESI Positive	-0.20	0.01	0.5	Factor 1 “Western diet”
M84.0807T0.98	EXO	ESI Positive	-0.25	0.002	0.4	Factor 1 “Western diet”
<b>M84.0808T1.17</b> (Pipelic acid)	ENDO	ESI Positive	-0.32	0.00005	0.009	Factor 1 “Western diet”
M101.0243T0.87	ENDO	ESI Negative	0.17	0.03	0.9	Factor 1 “Western diet”
M104.1071T14.68	ENDO	ESI Positive	-0.18	0.03	0.8	Factor 1 “Western diet”
M105.0337T8.29	ENDO	ESI Positive	-0.16	0.048	0.8	Factor 1 “Western diet”
M108.9018T0.87	EXO	ESI Negative	0.20	0.02	0.6	Factor 1 “Western diet”
M108.9875T0.78	EXO	ESI Positive	0.21	0.009	0.997	Factor 1 “Western diet”
M110.899T0.87	EXO	ESI Negative	0.21	0.008	0.6	Factor 1 “Western diet”
M118.0496T1.17	ENDO	ESI Positive	-0.16	0.045	0.8	Factor 1 “Western diet”
<b>M130.0861T1.16</b> (Pipelic acid)	ENDO	ESI Positive	-0.35	0.000008	0.002	Factor 1 “Western diet”
M130.0862T0.98	EXO	ESI Positive	-0.17	0.03	0.997	Factor 1 “Western diet”
M131.0895T0.98	EXO	ESI Positive	-0.18	0.03	0.997	Factor 1 “Western diet”
<b>M131.0895T1.16</b> (Pipelic acid)	ENDO	ESI Positive	-0.36	0.000006	0.002	Factor 1 “Western diet”
M132.0654T0.85	ENDO	ESI Positive	0.21	0.009	0.5	Factor 1 “Western diet”
M132.0655T0.84	EXO	ESI Positive	0.25	0.002	0.4	Factor 1 “Western diet”
M136.8774T0.94	ENDO	ESI Negative	-0.19	0.02	0.9	Factor 1 “Western diet”
M138.8751T0.92	EXO	ESI Negative	-0.20	0.01	0.6	Factor 1 “Western diet”
M146.0458T0.87	ENDO	ESI Negative	0.16	0.04	0.9	Factor 1 “Western diet”
M155.8666T0.92	ENDO	ESI Negative	-0.19	0.02	0.9	Factor 1 “Western diet”
M156.0767T0.79	EXO	ESI Positive	-0.18	0.03	0.997	Factor 1 “Western diet”
M157.0118T1.23	ENDO	ESI Negative	-0.17	0.03	0.9	Factor 1 “Western diet”
M157.087T17.71	EXO	ESI Negative	0.20	0.01	0.6	Factor 1 “Western diet”
M166.8603T0.87	EXO	ESI Negative	0.19	0.02	0.8	Factor 1 “Western diet”
M168.8501T1.02	EXO	ESI Negative	-0.17	0.04	0.9	Factor 1 “Western diet”
M170.0414T0.67	EXO	ESI Positive	-0.17	0.03	0.997	Factor 1 “Western diet”
M171.0916T11.04	EXO	ESI Positive	0.20	0.01	0.997	Factor 1 “Western diet”
M176.1278T0.96	EXO	ESI Positive	0.20	0.02	0.997	Factor 1 “Western diet”
M184.9855T0.63	ENDO	ESI Positive	0.17	0.04	0.8	Factor 1 “Western diet”
M188.1644T9.57	EXO	ESI Positive	-0.25	0.002	0.4	Factor 1 “Western diet”
M189.1597T0.79	ENDO	ESI Positive	0.22	0.007	0.5	Factor 1 “Western diet”
M189.1597T0.79	EXO	ESI Positive	0.18	0.03	0.997	Factor 1 “Western diet”
M209.1899T13.82	EXO	ESI Positive	-0.18	0.03	0.997	Factor 1 “Western diet”
M215.1652T17.21	EXO	ESI Negative	0.17	0.04	0.9	Factor 1 “Western diet”

## ANNEXE 6

M223.1077T7.2	EXO	ESI Positive	0.20	0.01	0.997	Factor 1 “Western diet”
M223.1078T7.19	ENDO	ESI Positive	0.17	0.04	0.8	Factor 1 “Western diet”
M241.1547T1.17	EXO	ESI Positive	-0.17	0.04	0.997	Factor 1 “Western diet”
M244.9081T14.25	EXO	ESI Negative	-0.25	0.002	0.4	Factor 1 “Western diet”
<b>M246.9051T14.25</b>	EXO	ESI Negative	-0.30	0.0002	0.08	Factor 1 “Western diet”
M251.1277T13.29	ENDO	ESI Positive	-0.17	0.04	0.8	Factor 1 “Western diet”
M260.2221T12.38	EXO	ESI Positive	-0.19	0.02	0.997	Factor 1 “Western diet”
M261.0112T1.05	EXO	ESI Positive	0.19	0.02	0.997	Factor 1 “Western diet”
M261.1439T7.47	EXO	ESI Positive	0.19	0.02	0.997	Factor 1 “Western diet”
M263.1389T8.18	EXO	ESI Positive	0.18	0.02	0.997	Factor 1 “Western diet”
M284.9108T14.58	EXO	ESI Positive	0.26	0.002	0.4	Factor 1 “Western diet”
M286.1437T13.51	ENDO	ESI Positive	0.20	0.01	0.5	Factor 1 “Western diet”
<b>M286.1437T13.64</b> <b>(Piperine)</b>	EXO	ESI Positive	0.26	0.001	0.4	Factor 1 “Western diet”
M287.147T13.5	EXO	ESI Positive	0.23	0.005	0.8	Factor 1 “Western diet”
M288.2533T13.95	ENDO	ESI Positive	-0.17	0.04	0.8	Factor 1 “Western diet”
M295.2278T15.34	EXO	ESI Negative	-0.17	0.03	0.8	Factor 1 “Western diet”
M301.2848T11.76	EXO	ESI Positive	-0.20	0.02	0.997	Factor 1 “Western diet”
M304.1657T7.39	EXO	ESI Positive	0.20	0.02	0.997	Factor 1 “Western diet”
M310.9129T0.78	EXO	ESI Positive	0.18	0.03	0.997	Factor 1 “Western diet”
M312.2261T15.13	EXO	ESI Negative	-0.16	0.047	0.9	Factor 1 “Western diet”
M327.2641T12.81	EXO	ESI Positive	-0.16	0.045	0.997	Factor 1 “Western diet”
M332.3159T13.18	EXO	ESI Positive	-0.19	0.02	0.997	Factor 1 “Western diet”
M338.3417T21.15	ENDO	ESI Positive	-0.16	0.04	0.8	Factor 1 “Western diet”
M339.1578T16.29	ENDO	ESI Positive	-0.17	0.03	0.8	Factor 1 “Western diet”
M353.1686T9.04	EXO	ESI Positive	0.17	0.04	0.997	Factor 1 “Western diet”
M359.7135T9.1	EXO	ESI Positive	-0.16	0.045	0.997	Factor 1 “Western diet”
M368.3156T16.46	ENDO	ESI Positive	-0.17	0.04	0.8	Factor 1 “Western diet”
M383.116T0.87	ENDO	ESI Positive	0.21	0.009	0.5	Factor 1 “Western diet”
M394.8907T0.75	ENDO	ESI Negative	0.16	0.04	0.9	Factor 1 “Western diet”
M395.2203T17.69	EXO	ESI Negative	-0.18	0.03	0.8	Factor 1 “Western diet”
M425.3635T19.48	EXO	ESI Negative	0.16	0.048	0.9	Factor 1 “Western diet”
M445.3674T20.1	EXO	ESI Positive	0.18	0.03	0.997	Factor 1 “Western diet”
M454.2928T15.09	ENDO	ESI Positive	-0.22	0.005	0.4	Factor 1 “Western diet”
M464.1913T10.3	EXO	ESI Positive	0.19	0.02	0.997	Factor 1 “Western diet”
M468.3083T15.8	EXO	ESI Positive	-0.18	0.03	0.997	Factor 1 “Western diet”
M478.2938T15.48	ENDO	ESI Negative	-0.17	0.04	0.9	Factor 1 “Western diet”
M479.2838T24.53	EXO	ESI Negative	-0.24	0.003	0.4	Factor 1 “Western diet”
M482.3241T14.5	ENDO	ESI Positive	-0.19	0.02	0.6	Factor 1 “Western diet”
M482.324T16.6	ENDO	ESI Positive	-0.19	0.02	0.6	Factor 1 “Western diet”
M494.3013T9.62	EXO	ESI Positive	-0.18	0.03	0.997	Factor 1 “Western diet”
M502.8144T9.62	EXO	ESI Positive	-0.20	0.02	0.997	Factor 1 “Western diet”
M510.3554T15.97	ENDO	ESI Positive	-0.21	0.01	0.5	Factor 1 “Western diet”
M518.3239T13.99	ENDO	ESI Positive	-0.20	0.01	0.5	Factor 1 “Western diet”
M524.2781T14.53	EXO	ESI Negative	-0.16	0.04	0.9	Factor 1 “Western diet”
M524.371T16.82	ENDO	ESI Positive	-0.17	0.04	0.8	Factor 1 “Western diet”
M526.2927T14.57	ENDO	ESI Positive	-0.24	0.003	0.4	Factor 1 “Western diet”



## ANNEXE 6

M526.3149T14.5	EXO	ESI Negative	-0.18	0.02	0.8	Factor 1 “Western diet”
M532.8573T1.17	ENDO	ESI Positive	-0.17	0.04	0.8	Factor 1 “Western diet”
M540.366T15.45	EXO	ESI Positive	-0.17	0.04	0.997	Factor 1 “Western diet”
M546.3553T15.18	ENDO	ESI Positive	-0.18	0.03	0.8	Factor 1 “Western diet”
M548.3709T15.96	ENDO	ESI Positive	-0.23	0.005	0.4	Factor 1 “Western diet”
M552.3308T15.29	EXO	ESI Negative	-0.17	0.03	0.8	Factor 1 “Western diet”
M556.3521T16.03	EXO	ESI Negative	-0.21	0.01	0.6	Factor 1 “Western diet”
M568.3397T14.65	ENDO	ESI Positive	-0.23	0.004	0.4	Factor 1 “Western diet”
M598.2651T15.21	ENDO	ESI Positive	-0.19	0.02	0.6	Factor 1 “Western diet”
M608.3851T8.88	ENDO	ESI Positive	-0.17	0.03	0.8	Factor 1 “Western diet”
M610.2622T14.69	EXO	ESI Positive	0.17	0.04	0.997	Factor 1 “Western diet”
M613.3404T8.89	EXO	ESI Positive	-0.22	0.008	0.997	Factor 1 “Western diet”
M618.2467T15.26	EXO	ESI Positive	0.17	0.04	0.997	Factor 1 “Western diet”
M621.2727T15.59	ENDO	ESI Positive	-0.19	0.02	0.6	Factor 1 “Western diet”
M689.5213T11.31	ENDO	ESI Positive	-0.21	0.009	0.5	Factor 1 “Western diet”
M732.8605T0.78	EXO	ESI Negative	0.19	0.02	0.8	Factor 1 “Western diet”
M737.8604T0.77	EXO	ESI Positive	-0.17	0.04	0.997	Factor 1 “Western diet”
M842.8285T0.78	EXO	ESI Negative	-0.21	0.01	0.6	Factor 1 “Western diet”
M873.8352T0.77	EXO	ESI Positive	-0.17	0.04	0.997	Factor 1 “Western diet”
M997.1595T0.78	ENDO	ESI Negative	-0.17	0.04	0.9	Factor 1 “Western diet”
M69.0335T0.88	ENDO	ESI Positive	-0.16	0.0497	0.5	Factor 2 “Healthy diet
M72.0806T1.15	EXO	ESI Positive	0.16	0.047	0.8	Factor 2 “Healthy diet
<b>M73.0284T1.02</b>	ENDO	ESI Positive	0.27	0.0006	0.07	Factor 2 “Healthy diet
M80.9166T0.9	EXO	ESI Negative	0.16	0.04	0.6	Factor 2 “Healthy diet
M84.0807T0.98	EXO	ESI Positive	0.25	0.002	0.6	Factor 2 “Healthy diet
<b>M84.0808T1.17 (Pipelic acid)</b>	ENDO	ESI Positive	0.32	0.0001	0.01	Factor 2 “Healthy diet
M85.0283T0.87	EXO	ESI Positive	-0.18	0.03	0.7	Factor 2 “Healthy diet
M86.0601T1.18	ENDO	ESI Positive	0.17	0.03	0.5	Factor 2 “Healthy diet
M101.0607T11.87	EXO	ESI Negative	0.17	0.03	0.6	Factor 2 “Healthy diet
M103.9201T8.29	EXO	ESI Negative	0.19	0.02	0.4	Factor 2 “Healthy diet
M104.1071T14.68	ENDO	ESI Positive	0.18	0.02	0.5	Factor 2 “Healthy diet
M105.0337T8.29	ENDO	ESI Positive	0.25	0.002	0.2	Factor 2 “Healthy diet
M112.9855T16.93	ENDO	ESI Negative	-0.17	0.04	0.96	Factor 2 “Healthy diet
M118.0348T0.68	EXO	ESI Positive	-0.24	0.003	0.6	Factor 2 “Healthy diet
M119.016T0.97	EXO	ESI Positive	0.21	0.01	0.6	Factor 2 “Healthy diet
<b>M119.016T1.03 (Dihydro4mercapto-3(2H) furanone)</b>	ENDO	ESI Positive	0.28	0.0004	0.05	Factor 2 “Healthy diet
M121.9662T18.9	EXO	ESI Positive	-0.20	0.01	0.6	Factor 2 “Healthy diet
M129.0556T8.28	EXO	ESI Negative	0.17	0.03	0.6	Factor 2 “Healthy diet
<b>M130.0861T1.16 (Pipelic acid)</b>	ENDO	ESI Positive	0.34	0.00002	0.006	Factor 2 “Healthy diet
M130.0862T0.98	EXO	ESI Positive	0.23	0.005	0.6	Factor 2 “Healthy diet
M131.0895T0.98	EXO	ESI Positive	0.23	0.005	0.6	Factor 2 “Healthy diet
<b>M131.0895T1.16 (Pipelic acid)</b>	ENDO	ESI Positive	0.36	0.000004	0.002	Factor 2 “Healthy diet
M132.0766T0.92	ENDO	ESI Positive	-0.16	0.04	0.5	Factor 2 “Healthy diet
M136.8774T0.94	ENDO	ESI Negative	0.20	0.01	0.96	Factor 2 “Healthy diet

## ANNEXE 6

M138.8751T0.92	EXO	ESI Negative	0.23	0.005	0.4	Factor 2 "Healthy diet
M144.1018T0.96	EXO	ESI Positive	0.18	0.03	0.7	Factor 2 "Healthy diet
M144.1018T0.99	ENDO	ESI Positive	0.21	0.009	0.3	Factor 2 "Healthy diet
M149.0596T5.81	EXO	ESI Positive	-0.18	0.03	0.7	Factor 2 "Healthy diet
M157.0117T14.7	ENDO	ESI Negative	0.19	0.02	0.96	Factor 2 "Healthy diet
M158.1175T7.42	EXO	ESI Positive	0.18	0.03	0.7	Factor 2 "Healthy diet
M166.8603T0.87	EXO	ESI Negative	-0.19	0.02	0.4	Factor 2 "Healthy diet
M174.9379T0.79	ENDO	ESI Positive	-0.17	0.03	0.5	Factor 2 "Healthy diet
M176.8602T0.98	EXO	ESI Negative	-0.19	0.02	0.4	Factor 2 "Healthy diet
M180.0655T8.29	EXO	ESI Positive	0.19	0.02	0.6	Factor 2 "Healthy diet
M182.0811T2.06	EXO	ESI Positive	-0.19	0.02	0.6	Factor 2 "Healthy diet
M184.9158T14.94	EXO	ESI Negative	0.17	0.04	0.6	Factor 2 "Healthy diet
M184.9855T18.9	EXO	ESI Positive	-0.18	0.03	0.7	Factor 2 "Healthy diet
M187.0577T0.95	EXO	ESI Positive	-0.21	0.01	0.6	Factor 2 "Healthy diet
M187.0577T0.97 (2-deoxy-D-glucose)	ENDO	ESI Positive	-0.23	0.004	0.2	Factor 2 "Healthy diet
M189.0784T10.98	EXO	ESI Positive	0.19	0.02	0.6	Factor 2 "Healthy diet
M189.1345T0.81	EXO	ESI Positive	0.17	0.03	0.8	Factor 2 "Healthy diet
M190.0859T10.97 (3-indolepropionic acid)	EXO	ESI Positive	0.19	0.02	0.6	Factor 2 "Healthy diet
M192.0325T1.11	ENDO	ESI Positive	-0.17	0.03	0.5	Factor 2 "Healthy diet
M197.1284T7.96	EXO	ESI Positive	-0.17	0.04	0.8	Factor 2 "Healthy diet
M200.8588T22.65	EXO	ESI Negative	0.16	0.046	0.7	Factor 2 "Healthy diet
M202.858T20.4	EXO	ESI Negative	0.19	0.02	0.4	Factor 2 "Healthy diet
M209.1899T13.82	EXO	ESI Positive	0.17	0.04	0.8	Factor 2 "Healthy diet
M223.1077T7.2	EXO	ESI Positive	-0.22	0.007	0.6	Factor 2 "Healthy diet
M223.1078T7.19	ENDO	ESI Positive	-0.19	0.02	0.5	Factor 2 "Healthy diet
M224.0998T11.87	ENDO	ESI Positive	0.19	0.02	0.5	Factor 2 "Healthy diet
M225.0345T0.87	EXO	ESI Positive	-0.17	0.04	0.8	Factor 2 "Healthy diet
M231.1451T0.88	EXO	ESI Positive	0.21	0.01	0.6	Factor 2 "Healthy diet
M231.1452T0.89	ENDO	ESI Positive	0.20	0.02	0.4	Factor 2 "Healthy diet
M242.9254T0.79	ENDO	ESI Positive	-0.18	0.03	0.5	Factor 2 "Healthy diet
M244.9081T14.25	EXO	ESI Negative	0.21	0.009	0.4	Factor 2 "Healthy diet
M245.0162T0.95	EXO	ESI Positive	-0.20	0.02	0.6	Factor 2 "Healthy diet
M246.9051T14.25	EXO	ESI Negative	0.23	0.004	0.4	Factor 2 "Healthy diet
M260.1856T8.88	EXO	ESI Positive	-0.18	0.03	0.7	Factor 2 "Healthy diet
M261.0112T1.05	EXO	ESI Positive	-0.24	0.004	0.6	Factor 2 "Healthy diet
M268.104T1.17	ENDO	ESI Positive	-0.16	0.047	0.5	Factor 2 "Healthy diet
M274.1222T14.21	EXO	ESI Positive	-0.17	0.04	0.8	Factor 2 "Healthy diet
M275.6257T14.21	EXO	ESI Positive	-0.19	0.02	0.6	Factor 2 "Healthy diet
M281.6274T13.82	EXO	ESI Positive	-0.16	0.045	0.8	Factor 2 "Healthy diet
M288.2168T10.25	EXO	ESI Positive	-0.19	0.02	0.6	Factor 2 "Healthy diet
M291.647T14.69	EXO	ESI Positive	-0.17	0.04	0.8	Factor 2 "Healthy diet
M292.6377T13.82	EXO	ESI Positive	-0.16	0.0496	0.8	Factor 2 "Healthy diet
M299.6254T14.04	ENDO	ESI Positive	0.18	0.03	0.5	Factor 2 "Healthy diet
M302.2325T10.49 (2,6 dimethylheptanoyl carnitine/ nonanoyl-L-carnitine; acylcarnitine C9:1)	EXO	ESI Positive	-0.21	0.01	0.6	Factor 2 "Healthy diet

## ANNEXE 6

<b>M303.2357T10.49</b>						
<b>(2,6 dimethylheptanoyl carnitine/ nonanoyl-L-carnitine; acylcarnitine C9:1)</b>	EXO	ESI Positive	-0.30	0.0002	0.2	Factor 2 "Healthy diet
M304.2304T12.33	EXO	ESI Positive	-0.16	0.04	0.8	Factor 2 "Healthy diet
M309.207T14.52	EXO	ESI Negative	0.24	0.003	0.4	Factor 2 "Healthy diet
M310.9129T0.78	EXO	ESI Positive	-0.24	0.003	0.6	Factor 2 "Healthy diet
M311.2227T15.13	EXO	ESI Negative	0.20	0.01	0.4	Factor 2 "Healthy diet
M312.2261T15.13	EXO	ESI Negative	0.19	0.02	0.4	Factor 2 "Healthy diet
M312.8361T20.41	EXO	ESI Negative	0.18	0.02	0.5	Factor 2 "Healthy diet
M312.836T20.9	EXO	ESI Negative	0.19	0.02	0.4	Factor 2 "Healthy diet
M313.1546T8.83	EXO	ESI Positive	-0.19	0.02	0.6	Factor 2 "Healthy diet
M313.16T15.97	ENDO	ESI Positive	0.16	0.048	0.5	Factor 2 "Healthy diet
M316.6477T14.68	ENDO	ESI Positive	0.18	0.03	0.5	Factor 2 "Healthy diet
M325.1129T0.87	ENDO	ESI Positive	-0.18	0.02	0.5	Factor 2 "Healthy diet
M330.2639T11.6						
<b>(4,8 dimethylnonanoyl carnitine / undecanoylcarnitine; acylcarnitine C11:1)</b>	EXO	ESI Positive	-0.19	0.02	0.6	Factor 2 "Healthy diet
M344.2794T12.27	EXO	ESI Positive	-0.16	0.048	0.8	Factor 2 "Healthy diet
M347.0969T14.68	ENDO	ESI Positive	0.16	0.048	0.5	Factor 2 "Healthy diet
M352.1654T9.04	EXO	ESI Positive	-0.18	0.03	0.8	Factor 2 "Healthy diet
M353.1686T9.04	EXO	ESI Positive	-0.21	0.01	0.6	Factor 2 "Healthy diet
M356.2794T12.26	ENDO	ESI Positive	0.19	0.02	0.5	Factor 2 "Healthy diet
M357.2787T14.6	EXO	ESI Positive	-0.26	0.002	0.6	Factor 2 "Healthy diet
M358.2951T12.6	ENDO	ESI Positive	0.20	0.01	0.4	Factor 2 "Healthy diet
M367.121T0.97	ENDO	ESI Positive	-0.23	0.005	0.2	Factor 2 "Healthy diet
M373.314T13.16	EXO	ESI Positive	-0.17	0.04	0.8	Factor 2 "Healthy diet
M383.116T0.87	ENDO	ESI Positive	-0.20	0.01	0.4	Factor 2 "Healthy diet
M391.2853T14.6	EXO	ESI Negative	-0.21	0.009	0.4	Factor 2 "Healthy diet
M452.2771T14.11	EXO	ESI Positive	-0.23	0.005	0.6	Factor 2 "Healthy diet
M463.2325T13.14	EXO	ESI Positive	0.17	0.03	0.8	Factor 2 "Healthy diet
M463.2325T13.17	ENDO	ESI Positive	0.16	0.049	0.5	Factor 2 "Healthy diet
M468.3084T13.82	EXO	ESI Positive	-0.20	0.01	0.6	Factor 2 "Healthy diet
M471.7378T8.73	EXO	ESI Positive	-0.16	0.046	0.8	Factor 2 "Healthy diet
M479.2838T24.53	EXO	ESI Negative	0.20	0.01	0.4	Factor 2 "Healthy diet
M495.3275T14.22	EXO	ESI Positive	-0.20	0.01	0.6	Factor 2 "Healthy diet
M500.277T13.97	ENDO	ESI Positive	0.18	0.02	0.5	Factor 2 "Healthy diet
M500.2782T14.57	EXO	ESI Negative	-0.18	0.03	0.5	Factor 2 "Healthy diet
M508.3395T15.27 (LysoPC (17:1))	EXO	ESI Positive	-0.21	0.01	0.6	Factor 2 "Healthy diet
M508.3397T14.88	EXO	ESI Positive	-0.18	0.03	0.7	Factor 2 "Healthy diet
M510.4363T17.54	ENDO	ESI Positive	-0.20	0.01	0.4	Factor 2 "Healthy diet
M512.2993T13.83	EXO	ESI Negative	-0.18	0.03	0.5	Factor 2 "Healthy diet
M520.3397T14.68	ENDO	ESI Positive	0.17	0.04	0.5	Factor 2 "Healthy diet
M529.4072T20.19	EXO	ESI Positive	-0.20	0.01	0.6	Factor 2 "Healthy diet
M538.315T14.23	EXO	ESI Negative	-0.25	0.002	0.4	Factor 2 "Healthy diet
M560.2993T14.69	ENDO	ESI Positive	0.19	0.02	0.5	Factor 2 "Healthy diet
M566.3366T14.69	ENDO	ESI Negative	0.18	0.03	0.96	Factor 2 "Healthy diet

## ANNEXE 6

M567.2254T13.82	EXO	ESI Positive	-0.21	0.01	0.6	Factor 2 “Healthy diet
M568.3397T14.65	ENDO	ESI Positive	0.17	0.04	0.5	Factor 2 “Healthy diet
M571.8671T0.77	ENDO	ESI Positive	0.16	0.04	0.5	Factor 2 “Healthy diet
M596.7595T14.67	EXO	ESI Positive	0.20	0.01	0.6	Factor 2 “Healthy diet
M612.3305T14.64	EXO	ESI Negative	0.19	0.02	0.4	Factor 2 “Healthy diet
M612.3309T14.66	ENDO	ESI Negative	0.19	0.02	0.96	Factor 2 “Healthy diet
M616.3618T15.74	EXO	ESI Negative	-0.21	0.01	0.4	Factor 2 “Healthy diet
M624.3806T0.74	EXO	ESI Positive	0.16	0.046	0.8	Factor 2 “Healthy diet
M677.5142T11.3	ENDO	ESI Positive	-0.18	0.02	0.5	Factor 2 “Healthy diet
M681.094T0.86	EXO	ESI Positive	-0.22	0.007	0.6	Factor 2 “Healthy diet
M711.353T0.78	EXO	ESI Positive	-0.20	0.01	0.6	Factor 2 “Healthy diet
M759.1386T11.3	ENDO	ESI Positive	-0.25	0.002	0.2	Factor 2 “Healthy diet
M764.6931T11.3	ENDO	ESI Positive	-0.23	0.005	0.2	Factor 2 “Healthy diet
M798.5292T24.78	EXO	ESI Negative	0.19	0.02	0.4	Factor 2 “Healthy diet
M868.5027T10.07	ENDO	ESI Positive	-0.24	0.003	0.2	Factor 2 “Healthy diet
M881.548T11.29	ENDO	ESI Positive	-0.23	0.005	0.2	Factor 2 “Healthy diet
M881.8495T11.29	ENDO	ESI Positive	-0.18	0.03	0.5	Factor 2 “Healthy diet
M895.8448T0.78	ENDO	ESI Negative	-0.17	0.04	0.96	Factor 2 “Healthy diet
<b>M918.3358T10.11</b>	ENDO	ESI Positive	-0.27	0.0009	0.09	Factor 2 “Healthy diet
M943.054T10.6	ENDO	ESI Positive	-0.24	0.003	0.2	Factor 2 “Healthy diet
M960.7834T0.78	ENDO	ESI Negative	-0.16	0.04	0.96	Factor 2 “Healthy diet
M960.9475T0.75	ENDO	ESI Negative	0.18	0.03	0.96	Factor 2 “Healthy diet
M979.498T11.29	ENDO	ESI Positive	-0.17	0.04	0.5	Factor 2 “Healthy diet
M979.6094T11.29	ENDO	ESI Positive	-0.16	0.04	0.5	Factor 2 “Healthy diet
M979.7207T11.29	ENDO	ESI Positive	-0.16	0.047	0.5	Factor 2 “Healthy diet
M979.8322T11.29	ENDO	ESI Positive	-0.17	0.04	0.5	Factor 2 “Healthy diet
M979.9436T11.29	ENDO	ESI Positive	-0.18	0.03	0.5	Factor 2 “Healthy diet
M980.0549T11.29	ENDO	ESI Positive	-0.17	0.03	0.5	Factor 2 “Healthy diet
M980.1663T11.29	ENDO	ESI Positive	-0.17	0.03	0.5	Factor 2 “Healthy diet
M980.3889T11.29	ENDO	ESI Positive	-0.17	0.04	0.5	Factor 2 “Healthy diet
M985.7148T11.28	ENDO	ESI Positive	-0.22	0.007	0.3	Factor 2 “Healthy diet
M986.0555T11.26	ENDO	ESI Positive	-0.17	0.03	0.5	Factor 2 “Healthy diet
M986.6128T11.26	ENDO	ESI Positive	-0.19	0.02	0.5	Factor 2 “Healthy diet
M60.0806T0.85	EXO	ESI Positive	-0.18	0.03	0.998	Factor 3 “Sweetened products”
M73.0284T1.02	ENDO	ESI Positive	-0.16	0.046	0.6	Factor 3 “Sweetened products”
M98.9842T1.08	ENDO	ESI Positive	0.20	0.02	0.6	Factor 3 “Sweetened products”
M100.0758T18.11	EXO	ESI Positive	-0.19	0.02	0.998	Factor 3 “Sweetened products”
M104.0708T0.91	ENDO	ESI Positive	-0.18	0.02	0.6	Factor 3 “Sweetened products”
M109.0287T0.86	EXO	ESI Positive	-0.18	0.03	0.998	Factor 3 “Sweetened products”
M112.9855T16.93	ENDO	ESI Negative	0.18	0.03	0.9	Factor 3 “Sweetened products”
M130.0861T1.16 (Pipelic acid)	ENDO	ESI Positive	0.16	0.04	0.6	Factor 3 “Sweetened products”
M134.019T0.83	ENDO	ESI Positive	-0.18	0.03	0.6	Factor 3 “Sweetened products”
M152.0705T7.13	EXO	ESI Positive	0.17	0.03	0.998	Factor 3 “Sweetened products”
M152.9254T11.85	EXO	ESI Negative	0.18	0.03	0.8	Factor 3 “Sweetened products”
M153.8783T24.34	EXO	ESI Negative	0.23	0.004	0.4	Factor 3 “Sweetened products”
M156.1018T7.93	ENDO	ESI Positive	-0.19	0.02	0.6	Factor 3 “Sweetened products”

## ANNEXE 6

M160.1331T1.12	ENDO	ESI Positive	0.16	0.04	0.6	Factor 3 “Sweetened products”
M166.8603T0.87	ENDO	ESI Negative	-0.20	0.01	0.9	Factor 3 “Sweetened products”
M176.0704T10.22	EXO	ESI Positive	-0.18	0.03	0.998	Factor 3 “Sweetened products”
M181.0859T11.87	ENDO	ESI Positive	-0.16	0.046	0.6	Factor 3 “Sweetened products”
M184.9158T14.94	EXO	ESI Negative	-0.24	0.003	0.4	Factor 3 “Sweetened products”
M184.9855T14.58	EXO	ESI Positive	0.17	0.04	0.998	Factor 3 “Sweetened products”
M185.0304T1.17	ENDO	ESI Positive	-0.18	0.02	0.6	Factor 3 “Sweetened products”
M186.1124T1.16	ENDO	ESI Positive	0.17	0.04	0.6	Factor 3 “Sweetened products”
M189.1234T1.15	ENDO	ESI Positive	0.17	0.03	0.6	Factor 3 “Sweetened products”
M193.1546T0.86	EXO	ESI Positive	-0.18	0.03	0.998	Factor 3 “Sweetened products”
M199.1441T20.43	EXO	ESI Positive	0.25	0.002	0.998	Factor 3 “Sweetened products”
M199.1703T16.25	EXO	ESI Negative	0.17	0.03	0.8	Factor 3 “Sweetened products”
M216.1593T16.62	ENDO	ESI Positive	-0.17	0.04	0.6	Factor 3 “Sweetened products”
<b>M224.0998T11.87</b>	ENDO	ESI Positive	-0.30	0.0002	0.09	Factor 3 “Sweetened products”
M230.0958T1.03	ENDO	ESI Positive	-0.20	0.02	0.6	Factor 3 “Sweetened products”
M244.2634T12.09	ENDO	ESI Positive	-0.18	0.02	0.6	Factor 3 “Sweetened products”
M245.0768T1.17	ENDO	ESI Positive	0.19	0.02	0.6	Factor 3 “Sweetened products”
M248.1677T10.5	EXO	ESI Positive	-0.23	0.006	0.998	Factor 3 “Sweetened products”
M269.2121T17.89	EXO	ESI Negative	0.23	0.005	0.4	Factor 3 “Sweetened products”
M270.191T8.04	EXO	ESI Positive	-0.17	0.04	0.998	Factor 3 “Sweetened products”
M279.1601T13.84	EXO	ESI Negative	0.19	0.02	0.8	Factor 3 “Sweetened products”
M290.6322T13.84	ENDO	ESI Positive	-0.24	0.002	0.4	Factor 3 “Sweetened products”
M291.6097T14.58	ENDO	ESI Positive	-0.18	0.03	0.6	Factor 3 “Sweetened products”
M292.2268T13.66	ENDO	ESI Positive	-0.17	0.04	0.6	Factor 3 “Sweetened products”
M299.6254T14.04	ENDO	ESI Positive	-0.19	0.02	0.6	Factor 3 “Sweetened products”
M301.2848T11.76	EXO	ESI Positive	0.17	0.04	0.998	Factor 3 “Sweetened products”
M302.2325T10.49						
(2,6 dimethylheptanoyl carnitine/ nonanoyl-l-carnitine; acylcarnitine C9:1)	EXO	ESI Positive	0.18	0.03	0.998	Factor 3 “Sweetened products”
M303.2357T10.49						
(2,6 dimethylheptanoyl carnitine/ nonanoyl-l-carnitine; acylcarnitine C9:1)	EXO	ESI Positive	0.22	0.007	0.998	Factor 3 “Sweetened products”
M303.6468T14.64	ENDO	ESI Positive	-0.17	0.04	0.6	Factor 3 “Sweetened products”
M327.0152T0.82	ENDO	ESI Positive	-0.18	0.03	0.6	Factor 3 “Sweetened products”
M327.2641T12.81	EXO	ESI Positive	0.18	0.03	0.998	Factor 3 “Sweetened products”
M332.9217T0.78	ENDO	ESI Negative	0.17	0.03	0.9	Factor 3 “Sweetened products”
<b>M340.0238T11.87</b>	ENDO	ESI Positive	-0.26	0.001	0.3	Factor 3 “Sweetened products”
M349.2361T18.97	EXO	ESI Negative	0.16	0.04	0.8	Factor 3 “Sweetened products”
M355.9588T24.49	EXO	ESI Negative	0.17	0.04	0.8	Factor 3 “Sweetened products”
M356.2618T12.94	EXO	ESI Positive	-0.19	0.02	0.998	Factor 3 “Sweetened products”
M357.2787T14.6	EXO	ESI Positive	-0.18	0.03	0.998	Factor 3 “Sweetened products”
M358.0349T11.87	ENDO	ESI Positive	-0.22	0.007	0.6	Factor 3 “Sweetened products”
M368.2793T12.25	ENDO	ESI Positive	-0.17	0.04	0.6	Factor 3 “Sweetened products”
M371.3346T17.46	ENDO	ESI Positive	-0.17	0.03	0.6	Factor 3 “Sweetened products”
M373.2736T12.81 (Cholic acid)	EXO	ESI Positive	-0.17	0.04	0.998	Factor 3 “Sweetened products”
M375.2891T14.4	EXO	ESI Positive	-0.19	0.02	0.998	Factor 3 “Sweetened products”

## ANNEXE 6

M385.1669T17.75	ENDO	ESI Positive	-0.18	0.03	0.6	Factor 3 “Sweetened products”
M391.2853T14.6	EXO	ESI Negative	-0.18	0.02	0.8	Factor 3 “Sweetened products”
M391.9529T0.81	ENDO	ESI Positive	-0.18	0.02	0.6	Factor 3 “Sweetened products”
M450.3213T12.92	EXO	ESI Positive	-0.18	0.03	0.998	Factor 3 “Sweetened products”
M450.3213T12.95	ENDO	ESI Positive	-0.20	0.02	0.6	Factor 3 “Sweetened products”
M473.33T17.55	ENDO	ESI Positive	-0.18	0.02	0.6	Factor 3 “Sweetened products”
M489.3586T16.39	EXO	ESI Negative	-0.18	0.02	0.8	Factor 3 “Sweetened products”
<b>M498.9301T19.76</b>	EXO	ESI Negative	-0.26	0.001	0.4	Factor 3 “Sweetened products”
M500.277T13.97	ENDO	ESI Positive	-0.21	0.008	0.6	Factor 3 “Sweetened products”
M508.3395T15.27 (LysoPC (17:1))	EXO	ESI Positive	0.17	0.04	0.998	Factor 3 “Sweetened products”
M526.3149T14.5	EXO	ESI Negative	0.16	0.04	0.8	Factor 3 “Sweetened products”
M527.2976T15.2	ENDO	ESI Positive	-0.17	0.04	0.6	Factor 3 “Sweetened products”
M529.4072T20.19	EXO	ESI Positive	0.18	0.03	0.998	Factor 3 “Sweetened products”
M548.8124T15.59	ENDO	ESI Positive	-0.16	0.047	0.6	Factor 3 “Sweetened products”
M549.3139T15.59	ENDO	ESI Positive	-0.17	0.03	0.6	Factor 3 “Sweetened products”
M550.3866T17.13	ENDO	ESI Positive	-0.22	0.006	0.6	Factor 3 “Sweetened products”
M553.0227T0.75	EXO	ESI Negative	0.19	0.02	0.8	Factor 3 “Sweetened products”
M572.2993T14.67	ENDO	ESI Positive	-0.17	0.03	0.6	Factor 3 “Sweetened products”
M572.3709T15.72	ENDO	ESI Positive	-0.19	0.02	0.6	Factor 3 “Sweetened products”
M598.837T0.79	EXO	ESI Positive	-0.17	0.04	0.998	Factor 3 “Sweetened products”
M658.8738T0.73	EXO	ESI Positive	-0.18	0.03	0.998	Factor 3 “Sweetened products”
M663.9023T0.81	ENDO	ESI Positive	-0.17	0.04	0.6	Factor 3 “Sweetened products”
M681.4312T16.91	EXO	ESI Positive	0.16	0.047	0.998	Factor 3 “Sweetened products”
M689.5213T11.31	ENDO	ESI Positive	0.17	0.04	0.6	Factor 3 “Sweetened products”
M716.5474T11.28	ENDO	ESI Positive	0.17	0.04	0.6	Factor 3 “Sweetened products”
M737.6199T10.44	ENDO	ESI Positive	0.20	0.01	0.6	Factor 3 “Sweetened products”
M835.843T0.78	ENDO	ESI Negative	0.19	0.02	0.9	Factor 3 “Sweetened products”
M869.144T11.31	ENDO	ESI Positive	0.17	0.04	0.6	Factor 3 “Sweetened products”
M881.548T11.29	ENDO	ESI Positive	0.16	0.046	0.6	Factor 3 “Sweetened products”
M895.8448T0.78	ENDO	ESI Negative	0.19	0.02	0.9	Factor 3 “Sweetened products”
M918.886T0.77	ENDO	ESI Negative	0.20	0.01	0.9	Factor 3 “Sweetened products”
M929.8382T0.77	EXO	ESI Negative	-0.17	0.04	0.8	Factor 3 “Sweetened products”
M963.3313T0.78	EXO	ESI Negative	-0.20	0.01	0.8	Factor 3 “Sweetened products”
M964.3333T0.78	EXO	ESI Negative	-0.16	0.049	0.8	Factor 3 “Sweetened products”
M972.32T0.78	ENDO	ESI Negative	0.18	0.02	0.9	Factor 3 “Sweetened products”
M996.8228T0.78	EXO	ESI Negative	-0.24	0.003	0.4	Factor 3 “Sweetened products”
M997.325T0.78	EXO	ESI Negative	-0.19	0.02	0.8	Factor 3 “Sweetened products”
M1003.1626T15.2	ENDO	ESI Positive	-0.16	0.04	0.6	Factor 3 “Sweetened products”

SU.VI.MAX, Supplémentation en Vitamines et Minéraux Antioxydants; R; correlation coefficient ; M, molecular mass ; T, retention time ; ENDO, endogenous protocol ; EXO, exogenous protocol

Associations presented in **Table 4** in the manuscript are highlighted in bold.

<sup>1</sup> Partial spearman correlation adjusted for age (continuous), BMI (continuous), menopausal status (pre/post-menopause) and smoking status (current/former/never) at baseline, season at time of blood draw (a priori-defined periods: October/November, December/January/February, and March-April), number of 24h-dietary records and mean of daily energy intake (kcal) during the 2 years following the blood draw.

<sup>2</sup> 160 samples for endogenous protocol and exogenous protocol in negative ionization mode and 156 samples for exogenous protocol in positive ionization mode were analyzed.

## ANNEXE 6

<sup>3</sup> P-trend after Benjamini Hochberg correction for multiple testing

## Annexe 7 Ions associés à la consommation de groupes d'aliments spécifiques chez les femmes (analyses en LC-MS)

Ions Masse (M). Temps de rétention (T)	Identification <sup>2</sup>	Protocole/ mode d'ionisation <sup>3</sup>	R	P	P-corrigée	Groupe d'aliments
M144.1018T0.96	Proline bêtaïne (niveau 1)	Exogène ESI+	0,53	<0,0001	<0,0001	Agrumes
M144.1018T0.99	Proline bêtaïne (niveau 1)	Endogène ESI+	0,54	<0,0001	<0,0001	Agrumes
M160.0968T0.88	Hydroxyproline bêtaïne (niveau 1)	Exogène ESI+	0,37	<0,0001	0,001	Agrumes
M847.8273T0.78	(niveau 4 : non identifié)	Exogène ESI+	0,31	0,0001	0,03	Agrumes
M245.0768T1.16	(niveau 4 : non identifié)	Exogène ESI+	0,27	0,0009	0,05	Autres féculents (châtaigne, tapioca, plantain)
M266.1511T21.59	(niveau 4 : non identifié)	Exogène ESI-	0,26	0,001	0,08	Autres féculents (châtaigne, tapioca, plantain)
M312.8361T20.41	(niveau 4 : non identifié)	Exogène ESI-	0,29	0,0002	0,05	Autres féculents (châtaigne, tapioca, plantain)
M312.836T21.67	(niveau 4 : non identifié)	Exogène ESI-	0,28	0,0005	0,05	Autres féculents (châtaigne, tapioca, plantain)
M381.2316T24.58	(niveau 4 : non identifié)	Exogène ESI-	0,28	0,0005	0,05	Autres féculents (châtaigne, tapioca, plantain)
M540.8819T0.78	(niveau 4 : non identifié)	Endogène ESI+	0,30	0,0002	0,09	Autres féculents (châtaigne, tapioca, plantain)
M97.9688T0.63	(niveau 4 : non identifié)	Exogène ESI+	0,27	0,0007	0,05	Autres féculents (châtaigne, tapioca, plantain)
M78.9187T0.9	(niveau 4 : non identifié)	Exogène ESI-	0,30	0,0002	0,06	Baies
M138.0542T0.92	Trigonelline (niveau 1)	Exogène ESI+	0,47	<0,0001	<0,0001	Café
M138.0545T0.94	Trigonelline (niveau 1)	Endogène ESI+	0,51	<0,0001	<0,0001	Café
M179.0561T0.92	(niveau 4 : non identifié)	Exogène ESI-	0,30	0,0002	0,08	Café
M181.072T7.41	Paraxanthine (niveau 1)	Endogène ESI+	0,40	<0,0001	<0,0001	Café
M181.072T7.44	Paraxanthine (niveau 1)	Exogène ESI+	0,36	<0,0001	0,001	Café
M211.144T8.79	Cyclo(isoleucyl-prolyl) (niveau 1)	Exogène ESI+	0,39	<0,0001	0,0002	Café
M475.1651T11.22	(niveau 4 : non identifié)	Exogène ESI+	0,43	<0,0001	<0,0001	Café
M497.238T10.12	Atractyligenin glucuronide (niveau 1)	Exogène ESI+	0,56	<0,0001	<0,0001	Café
M96.0444T1.17	(niveau 4: formule brute C5 H6 O N)	Exogène ESI+	0,36	<0,0001	0,001	Café
M189.1597T0.79	N6,N6,N6-Trimethyl-L-lysine (niveau 1)	Exogène ESI+	0,34	<0,0001	0,02	Charcuterie
M153.8783T24.34	(niveau 4 : non identifié)	Exogène ESI-	0,29	0,0003	0,06	chocolat
M399.221T13.62	(niveau 4 : non identifié)	Exogène ESI-	0,30	0,0001	0,05	Choux
M130.0862T0.72	Glycokinase (niveau 1)	Endogène ESI+	0,34	<0,0001	0,01	Lait
M160.1331T1.12	(niveau 4 : non identifié)	Endogène ESI+	0,30	0,0002	0,04	Lait
M186.1124T1.16	(niveau 4 : formule brute C9 H16 O3 N)	Endogène ESI+	0,30	0,0002	0,04	Lait
M73.0284T1.02	(niveau 4 : non identifié)	Endogène ESI+	0,28	0,0005	0,09	Légumes-feuilles
M338.5964T14.68	(niveau 4 : non identifié)	Exogène ESI+	0,31	<0,0001	0,09	Légumes-fruits
M445.3674T20.1	(niveau 4 : non identifié)	Exogène ESI+	0,35	<0,0001	0,02	Matière grasse végétale (sauce huile d'olive et soja)
M279.2329T18.16	(niveau 4 : non identifié)	Endogène ESI-	0,31	<0,0001	0,02	Pâtes et céréales



## ANNEXE 7

M341.1538T16.26	(niveau 4 : non identifié)	Exogène ESI+	0,32	<0,0001	0,07	Poissons gras
M395.2203T17.69	(niveau 4 : non identifié)	Exogène ESI-	0,32	<0,0001	0,03	Poissons gras
M391.2853T14.6	(niveau 4 : non identifié)	Exogène ESI-	0,30	0,0002	0,07	Viande rouge
M231.0839T1.16	(niveau 4 : formule brute C6H11O4N6 ou C8H16O6,Na)	Exogène ESI+	0,40	<0,0001	0,0003	Vin rouge
M104.1071T14.68	(niveau 4 : non identifié)	Endogène ESI+	0,31	0,0001	0,07	Yaourts
M244.9081T14.25	(niveau 4 : non identifié)	Exogène ESI-	0,28	0,0004	0,07	Yaourts
M246.9051T14.25	(niveau 4 : non identifié)	Exogène ESI-	0,28	0,0004	0,07	Yaourts
M518.3239T13.99	(niveau 4 : non identifié)	Endogène ESI+	0,26	0,001	0,10	Yaourts
M520.3397T14.68	(niveau 4 : non identifié)	Endogène ESI+	0,28	0,0005	0,09	Yaourts
M542.3241T14.04	(niveau 4 : non identifié)	Endogène ESI+	0,27	0,0006	0,09	Yaourts
M548.3709T15.96	(niveau 4 : non identifié)	Endogène ESI+	0,27	0,0008	0,09	Yaourts
M560.2993T14.69	(niveau 4 : non identifié)	Endogène ESI+	0,26	0,001	0,10	Yaourts
M568.3397T14.65	(niveau 4 : non identifié)	Endogène ESI+	0,28	0,0004	0,09	Yaourts

<sup>1</sup> Les corrélations de Spearman ont été ajustées sur les caractéristiques suivantes : âge, IMC, statut tabagique, statut ménopausique, saison de prélèvement sanguin, nombre d'enregistrements alimentaires sur 24 heures et moyenne de l'apport énergétique quotidien pendant les 2 années suivant le prélèvement sanguin. Associations présentées :  $r > 0.2$ ,  $P$ -corrigée  $< 0.1$ .

<sup>2</sup> Niveaux de confiance de l'identification : niveau 1 = composé identifié et validé par standard, niveau 2 = composé identifié de manière présumée, niveau 3 = classe du composé annotée de manière présumée et niveau 4 = composé inconnu.

<sup>3</sup> 160 individus ont été analysés pour le protocole endogène et le protocole exogène en mode d'ionisation négative et 156 individus pour le protocole exogène en mode ionisation positive

## Annexe 8 Associations entre les signaux RMN et les composantes alimentaires du mPNNS-GS chez les femmes

Déplacement chimique (ppm) (Assignations)	Séquence <sup>1</sup>	R <sup>2</sup>	P	P-corrigée	Composantes mPNNS-GS
0,7663	CPMG	0,16	0,04	0,9	Alcool
1,1663	NOESY1D	0,17	0,03	1,0	Alcool
1,1663	CPMG	0,16	0,04	0,9	Alcool
1,6303	CPMG	0,16	0,049	0,9	Alcool
1,6363	CPMG	0,17	0,03	0,9	Alcool
2,706	CPMG	0,20	0,01	0,9	Alcool
2,71	NOESY1D	0,20	0,02	1,0	Alcool
2,8333	NOESY1D	0,18	0,03	1,0	Alcool
2,8383	CPMG	0,16	0,04	0,9	Alcool
6,526	CPMG	-0,17	0,04	0,9	Alcool
6,906	NOESY1D	-0,16	0,049	1,0	Alcool
6,908	CPMG	-0,21	0,008	0,9	Alcool
0,2353	CPMG	-0,19	0,02	0,6	Boissons sucrées
0,3083	CPMG	-0,16	0,046	0,6	Boissons sucrées
0,3373	CPMG	-0,18	0,03	0,6	Boissons sucrées
2,1503	CPMG	0,17	0,04	0,6	Boissons sucrées
2,1653	CPMG	-0,19	0,02	0,6	Boissons sucrées
3,5312	CPMG	0,18	0,03	0,6	Boissons sucrées
3,8602	NOESY1D	0,17	0,03	0,7	Boissons sucrées
3,8602	CPMG	0,17	0,04	0,6	Boissons sucrées
6,108	NOESY1D	-0,16	0,0498	0,7	Boissons sucrées
6,286	NOESY1D	-0,18	0,03	0,7	Boissons sucrées
7,4299	CPMG	-0,17	0,03	0,6	Boissons sucrées
8,0469	CPMG	-0,21	0,01	0,6	Boissons sucrées
8,27	CPMG	-0,18	0,02	0,6	Boissons sucrées
8,308	CPMG	-0,18	0,03	0,6	Boissons sucrées
8,4261	CPMG	-0,23	0,005	0,6	Boissons sucrées
8,4641	CPMG	-0,18	0,03	0,6	Boissons sucrées
8,6241	CPMG	-0,18	0,03	0,6	Boissons sucrées
8,9903	CPMG	-0,19	0,02	0,6	Boissons sucrées
9,9637	CPMG	-0,17	0,04	0,6	Boissons sucrées
0,7263	NOESY1D	-0,16	0,045	0,1	Fruits et légumes
0,7663	NOESY1D	-0,21	0,009	0,06	Fruits et légumes
1,1663	NOESY1D	-0,20	0,01	0,07	Fruits et légumes
1,1663	CPMG	-0,18	0,03	0,3	Fruits et légumes
1,3103	NOESY1D	-0,17	0,04	0,1	Fruits et légumes
1,4323	CPMG	-0,16	0,048	0,4	Fruits et légumes
1,6303	NOESY1D	-0,19	0,02	0,09	Fruits et légumes
1,6363	NOESY1D	-0,21	0,01	0,07	Fruits et légumes
1,6463	CPMG	-0,17	0,03	0,3	Fruits et légumes
1,6463	NOESY1D	-0,27	0,0009	0,03	Fruits et légumes
1,9263	CPMG	-0,17	0,03	0,3	Fruits et légumes
2,018	NOESY1D	-0,18	0,02	0,1	Fruits et légumes
2,0633	CPMG	-0,25	0,002	0,1	Fruits et légumes
2,0633	NOESY1D	-0,24	0,003	0,03	Fruits et légumes
2,5013	CPMG	0,16	0,049	0,4	Fruits et légumes
2,5803	NOESY1D	-0,18	0,03	0,1	Fruits et légumes
2,6273	NOESY1D	-0,17	0,04	0,1	Fruits et légumes
2,646	CPMG	-0,22	0,007	0,2	Fruits et légumes
<b>2,8863</b>	CPMG	0,27	0,0008	0,09	Fruits et légumes
3,028	NOESY1D	-0,17	0,03	0,1	Fruits et légumes

## ANNEXE 8

3,028	CPMG	-0,17	0,03	0,3	Fruits et légumes
3,0702	NOESY1D	0,19	0,02	0,09	Fruits et légumes
3,3682	NOESY1D	-0,18	0,03	0,1	Fruits et légumes
3,4372	NOESY1D	-0,17	0,03	0,1	Fruits et légumes
3,4972	NOESY1D	-0,19	0,02	0,1	Fruits et légumes
3,5312	NOESY1D	-0,27	0,001	0,03	Fruits et légumes
3,5312	CPMG	-0,25	0,002	0,1	Fruits et légumes
3,6162	NOESY1D	-0,24	0,003	0,03	Fruits et légumes
3,6162	CPMG	-0,23	0,004	0,1	Fruits et légumes
3,7132	NOESY1D	-0,23	0,005	0,04	Fruits et légumes
3,7132	CPMG	-0,17	0,04	0,3	Fruits et légumes
3,7902	NOESY1D	-0,20	0,01	0,09	Fruits et légumes
3,7902	CPMG	-0,18	0,03	0,3	Fruits et légumes
3,7932	NOESY1D	-0,23	0,004	0,04	Fruits et légumes
3,7932	CPMG	-0,21	0,009	0,2	Fruits et légumes
3,8062	NOESY1D	-0,24	0,003	0,03	Fruits et légumes
3,8062	CPMG	-0,19	0,02	0,3	Fruits et légumes
3,8602	NOESY1D	-0,18	0,03	0,1	Fruits et légumes
3,8602	CPMG	-0,17	0,04	0,3	Fruits et légumes
3,9522	NOESY1D	-0,18	0,03	0,1	Fruits et légumes
5,1261	NOESY1D	-0,22	0,006	0,04	Fruits et légumes
5,1661	NOESY1D	-0,18	0,02	0,1	Fruits et légumes
5,2061	NOESY1D	-0,20	0,01	0,07	Fruits et légumes
5,4701	NOESY1D	-0,19	0,02	0,1	Fruits et légumes
5,5101	NOESY1D	-0,17	0,03	0,1	Fruits et légumes
5,5491	NOESY1D	-0,18	0,02	0,1	Fruits et légumes
5,5831	NOESY1D	-0,18	0,02	0,1	Fruits et légumes
5,6401	NOESY1D	-0,20	0,02	0,09	Fruits et légumes
5,6461	NOESY1D	-0,17	0,04	0,1	Fruits et légumes
6,166	NOESY1D	-0,17	0,03	0,1	Fruits et légumes
6,213	NOESY1D	-0,25	0,002	0,03	Fruits et légumes
6,215	CPMG	-0,23	0,005	0,1	Fruits et légumes
6,246	NOESY1D	-0,16	0,049	0,2	Fruits et légumes
6,286	NOESY1D	0,18	0,03	0,1	Fruits et légumes
6,686	CPMG	0,23	0,004	0,1	Fruits et légumes
<b>6,726</b>	CPMG	0,28	0,0005	0,09	Fruits et légumes
6,726	NOESY1D	0,23	0,004	0,03	Fruits et légumes
6,766	NOESY1D	0,26	0,001	0,03	Fruits et légumes
6,766	CPMG	0,21	0,01	0,2	Fruits et légumes
6,906	NOESY1D	0,21	0,01	0,07	Fruits et légumes
6,908	CPMG	0,17	0,03	0,3	Fruits et légumes
6,926	NOESY1D	0,26	0,001	0,03	Fruits et légumes
7,006	NOESY1D	0,24	0,002	0,03	Fruits et légumes
7,016	CPMG	0,23	0,005	0,1	Fruits et légumes
7,1099	NOESY1D	0,17	0,03	0,1	Fruits et légumes
7,1229	CPMG	0,17	0,04	0,3	Fruits et légumes
7,3499	NOESY1D	0,19	0,02	0,09	Fruits et légumes
<b>7,3659 (Phénylalanine)</b>	NOESY1D	0,27	0,0007	0,03	Fruits et légumes
7,3699	CPMG	0,23	0,005	0,1	Fruits et légumes
7,4299	NOESY1D	0,25	0,002	0,03	Fruits et légumes
7,5459	NOESY1D	0,19	0,02	0,10	Fruits et légumes
7,5659	NOESY1D	0,24	0,003	0,03	Fruits et légumes
7,6099	NOESY1D	0,21	0,008	0,06	Fruits et légumes
7,6459	NOESY1D	0,24	0,004	0,03	Fruits et légumes
7,8299	CPMG	0,17	0,04	0,4	Fruits et légumes
7,8699	NOESY1D	0,17	0,03	0,1	Fruits et légumes
7,8699	CPMG	0,16	0,04	0,4	Fruits et légumes
7,9659	NOESY1D	0,20	0,02	0,1	Fruits et légumes

## ANNEXE 8

8,0059	NOESY1D	0,23	0,004	0,04	Fruits et légumes
<b>8,0459</b>	NOESY1D	0,27	0,0008	0,03	Fruits et légumes
8,0469	CPMG	0,18	0,03	0,3	Fruits et légumes
8,0859	NOESY1D	0,27	0,0009	0,03	Fruits et légumes
8,1539	NOESY1D	0,24	0,003	0,03	Fruits et légumes
8,23	NOESY1D	0,26	0,001	0,03	Fruits et légumes
8,27	NOESY1D	0,22	0,006	0,04	Fruits et légumes
8,39	NOESY1D	0,17	0,03	0,1	Fruits et légumes
8,4301	NOESY1D	0,18	0,02	0,1	Fruits et légumes
8,7092	NOESY1D	0,17	0,04	0,1	Fruits et légumes
8,7502	NOESY1D	0,17	0,04	0,1	Fruits et légumes
8,7902	NOESY1D	0,19	0,02	0,1	Fruits et légumes
8,8302	NOESY1D	0,23	0,005	0,04	Fruits et légumes
8,8702	NOESY1D	0,27	0,0009	0,03	Fruits et légumes
<b>8,9103</b>	NOESY1D	0,29	0,0002	0,03	Fruits et légumes
8,9503	NOESY1D	0,24	0,003	0,03	Fruits et légumes
8,9903	NOESY1D	0,26	0,001	0,03	Fruits et légumes
9,0303	NOESY1D	0,25	0,002	0,03	Fruits et légumes
9,0703	NOESY1D	0,24	0,003	0,03	Fruits et légumes
9,1103	NOESY1D	0,21	0,009	0,1	Fruits et légumes
9,2304	NOESY1D	0,16	0,046	0,1	Fruits et légumes
0,0337	NOESY1D	0,18	0,02	1,0	Lait et produits laitiers
0,0463	NOESY1D	0,18	0,03	1,0	Lait et produits laitiers
0,0863	NOESY1D	0,19	0,02	1,0	Lait et produits laitiers
0,1263	NOESY1D	0,22	0,007	1,0	Lait et produits laitiers
0,1663	NOESY1D	0,19	0,02	1,0	Lait et produits laitiers
0,2063	NOESY1D	0,16	0,04	1,0	Lait et produits laitiers
0,2463	NOESY1D	0,17	0,03	1,0	Lait et produits laitiers
0,7663	CPMG	-0,17	0,03	0,8	Lait et produits laitiers
1,1663	NOESY1D	-0,17	0,03	1,0	Lait et produits laitiers
4,8571	CPMG	-0,18	0,02	0,8	Lait et produits laitiers
4,9501	CPMG	-0,17	0,04	0,8	Lait et produits laitiers
4,9901	CPMG	-0,16	0,047	0,8	Lait et produits laitiers
5,0391	CPMG	-0,17	0,04	0,8	Lait et produits laitiers
5,0761	CPMG	-0,18	0,02	0,8	Lait et produits laitiers
6,488	CPMG	0,17	0,04	0,8	Lait et produits laitiers
6,526	CPMG	0,22	0,007	0,8	Lait et produits laitiers
6,941	CPMG	0,16	0,04	0,8	Lait et produits laitiers
2,1103	NOESY1D	0,17	0,04	1,0	Matière grasse ajoutée totale
2,277	CPMG	-0,17	0,03	0,5	Matière grasse ajoutée totale
2,6273	NOESY1D	0,18	0,03	1,0	Matière grasse ajoutée totale
2,8463	CPMG	-0,17	0,03	0,5	Matière grasse ajoutée totale
3,3682	NOESY1D	0,19	0,02	0,9	Matière grasse ajoutée totale
3,3682	CPMG	0,18	0,03	0,5	Matière grasse ajoutée totale
3,4372	CPMG	0,19	0,02	0,5	Matière grasse ajoutée totale
3,4372	NOESY1D	0,19	0,02	0,9	Matière grasse ajoutée totale
3,4972	NOESY1D	0,16	0,04	1,0	Matière grasse ajoutée totale
3,4982	CPMG	0,20	0,02	0,5	Matière grasse ajoutée totale
3,7132	CPMG	0,17	0,04	0,5	Matière grasse ajoutée totale
3,8602	CPMG	0,21	0,01	0,5	Matière grasse ajoutée totale
3,8602	NOESY1D	0,20	0,01	0,9	Matière grasse ajoutée totale
3,918	NOESY1D	-0,19	0,02	0,9	Matière grasse ajoutée totale
3,92	CPMG	-0,18	0,03	0,5	Matière grasse ajoutée totale
4,8571	CPMG	0,16	0,04	0,5	Matière grasse ajoutée totale
5,1261	CPMG	0,16	0,04	0,5	Matière grasse ajoutée totale
<b>5,2061 (Glucose)</b>	CPMG	0,29	0,0003	0,1	Matière grasse ajoutée totale
<b>5,2061 (Glucose)</b>	NOESY1D	0,28	0,0006	0,1	Matière grasse ajoutée totale
6,128	CPMG	-0,17	0,03	0,5	Matière grasse ajoutée totale

## ANNEXE 8

6,526	CPMG	-0,18	0,02	0,5	Matière grasse ajoutée totale
6,567	CPMG	-0,16	0,04	0,5	Matière grasse ajoutée totale
6,941	CPMG	-0,19	0,02	0,5	Matière grasse ajoutée totale
7,5409	CPMG	-0,16	0,05	0,5	Matière grasse ajoutée totale
7,5459	NOESY1D	-0,17	0,03	1,0	Matière grasse ajoutée totale
7,6309	CPMG	-0,21	0,01	0,5	Matière grasse ajoutée totale
7,6459	CPMG	-0,25	0,002	0,2	Matière grasse ajoutée totale
7,6859	CPMG	-0,16	0,04	0,5	Matière grasse ajoutée totale
9,2134	CPMG	-0,17	0,04	0,5	Matière grasse ajoutée totale
2,264	NOESY1D	0,21	0,01	1,0	Matière grasse végétale ajoutée
2,267	CPMG	0,18	0,03	1,0	Matière grasse végétale ajoutée
5,864	CPMG	-0,21	0,009	1,0	Matière grasse végétale ajoutée
7,9259	CPMG	-0,17	0,03	1,0	Matière grasse végétale ajoutée
0,2353	CPMG	0,19	0,02	0,3	Pain, céréales, pommes de terre et légumineuses
0,2863	NOESY1D	0,18	0,03	0,2	Pain, céréales, pommes de terre et légumineuses
0,3263	NOESY1D	0,16	0,04	0,3	Pain, céréales, pommes de terre et légumineuses
0,6463	NOESY1D	-0,23	0,005	0,06	Pain, céréales, pommes de terre et légumineuses
0,6463	CPMG	-0,18	0,03	0,4	Pain, céréales, pommes de terre et légumineuses
0,6863	NOESY1D	-0,18	0,03	0,2	Pain, céréales, pommes de terre et légumineuses
<b>0,8063 (Lipoprotéines (CH3))</b>	NOESY1D	-0,27	0,0008	0,03	Pain, céréales, pommes de terre et légumineuses
0,8063	CPMG	-0,26	0,001	0,1	Pain, céréales, pommes de terre et légumineuses
0,969	CPMG	-0,16	0,049	0,4	Pain, céréales, pommes de terre et légumineuses
<b>1,0213 (Valine)</b>	NOESY1D	-0,27	0,0007	0,03	Pain, céréales, pommes de terre et légumineuses
1,0533	NOESY1D	-0,23	0,005	0,06	Pain, céréales, pommes de terre et légumineuses
1,0823	NOESY1D	-0,25	0,002	0,04	Pain, céréales, pommes de terre et légumineuses
1,0893	NOESY1D	-0,26	0,0010	0,03	Pain, céréales, pommes de terre et légumineuses
<b>1,1263 (Isobutyrate)</b>	NOESY1D	-0,27	0,0006	0,03	Pain, céréales, pommes de terre et légumineuses
1,1663	NOESY1D	-0,22	0,006	0,07	Pain, céréales, pommes de terre et légumineuses
1,2063	NOESY1D	-0,24	0,003	0,05	Pain, céréales, pommes de terre et légumineuses
1,2063	CPMG	-0,19	0,02	0,3	Pain, céréales, pommes de terre et légumineuses
1,3503	NOESY1D	-0,22	0,007	0,07	Pain, céréales, pommes de terre et légumineuses
1,3843	NOESY1D	-0,26	0,001	0,03	Pain, céréales, pommes de terre et légumineuses
1,3903	CPMG	-0,16	0,04	0,4	Pain, céréales, pommes de terre et légumineuses
1,4293	NOESY1D	-0,25	0,002	0,04	Pain, céréales, pommes de terre et légumineuses
1,4303	CPMG	-0,19	0,02	0,3	Pain, céréales, pommes de terre et légumineuses
1,4323	CPMG	-0,17	0,04	0,4	Pain, céréales, pommes de terre et légumineuses
1,4463	NOESY1D	-0,23	0,004	0,06	Pain, céréales, pommes de terre et légumineuses
1,5503	NOESY1D	-0,18	0,03	0,2	Pain, céréales, pommes de terre et légumineuses
1,5903	NOESY1D	-0,27	0,0009	0,03	Pain, céréales, pommes de terre et légumineuses
1,6303	NOESY1D	-0,26	0,0010	0,03	Pain, céréales, pommes de terre et légumineuses
1,6303	CPMG	-0,23	0,005	0,2	Pain, céréales, pommes de terre et légumineuses
1,6363	NOESY1D	-0,25	0,002	0,04	Pain, céréales, pommes de terre et légumineuses
1,6363	CPMG	-0,22	0,006	0,2	Pain, céréales, pommes de terre et légumineuses
1,7903	NOESY1D	-0,19	0,02	0,2	Pain, céréales, pommes de terre et légumineuses
1,8303	NOESY1D	-0,21	0,01	0,1	Pain, céréales, pommes de terre et légumineuses
1,9263	CPMG	-0,17	0,04	0,4	Pain, céréales, pommes de terre et légumineuses
2,1103	NOESY1D	-0,16	0,05	0,3	Pain, céréales, pommes de terre et légumineuses
2,1503	NOESY1D	-0,25	0,002	0,04	Pain, céréales, pommes de terre et légumineuses
2,1653	CPMG	0,19	0,02	0,3	Pain, céréales, pommes de terre et légumineuses
2,1753	NOESY1D	-0,17	0,04	0,2	Pain, céréales, pommes de terre et légumineuses
2,9792	NOESY1D	0,19	0,02	0,2	Pain, céréales, pommes de terre et légumineuses
3,1662	NOESY1D	-0,21	0,01	0,1	Pain, céréales, pommes de terre et légumineuses
3,1662	CPMG	-0,19	0,02	0,3	Pain, céréales, pommes de terre et légumineuses
3,3122	CPMG	0,17	0,03	0,4	Pain, céréales, pommes de terre et légumineuses
3,7132	CPMG	0,18	0,03	0,3	Pain, céréales, pommes de terre et légumineuses
3,8062	CPMG	0,17	0,03	0,4	Pain, céréales, pommes de terre et légumineuses
3,8602	NOESY1D	0,23	0,004	0,06	Pain, céréales, pommes de terre et légumineuses
3,8602	CPMG	0,22	0,006	0,2	Pain, céréales, pommes de terre et légumineuses

## ANNEXE 8

3,948	CPMG	0,16	0,05	0,4	Pain, céréales, pommes de terre et légumineuses
3,9502	CPMG	0,23	0,005	0,2	Pain, céréales, pommes de terre et légumineuses
4,4691	NOESY1D	-0,20	0,02	0,1	Pain, céréales, pommes de terre et légumineuses
5,2461	NOESY1D	-0,26	0,001	0,04	Pain, céréales, pommes de terre et légumineuses
5,2461	CPMG	-0,24	0,003	0,2	Pain, céréales, pommes de terre et légumineuses
5,3501	NOESY1D	-0,21	0,01	0,1	Pain, céréales, pommes de terre et légumineuses
5,3501	CPMG	-0,18	0,03	0,3	Pain, céréales, pommes de terre et légumineuses
5,3891	CPMG	-0,21	0,008	0,2	Pain, céréales, pommes de terre et légumineuses
5,4301	CPMG	-0,17	0,04	0,4	Pain, céréales, pommes de terre et légumineuses
6,246	CPMG	0,16	0,05	0,4	Pain, céréales, pommes de terre et légumineuses
6,286	NOESY1D	0,20	0,01	0,1	Pain, céréales, pommes de terre et légumineuses
<b>6,331</b>	NOESY1D	0,30	0,0002	0,03	Pain, céréales, pommes de terre et légumineuses
6,486	NOESY1D	0,18	0,03	0,2	Pain, céréales, pommes de terre et légumineuses
6,726	CPMG	0,16	0,04	0,4	Pain, céréales, pommes de terre et légumineuses
<b>6,86 (Tyrosine)</b>	CPMG	0,27	0,0009	0,1	Pain, céréales, pommes de terre et légumineuses
6,864	NOESY1D	0,23	0,004	0,06	Pain, céréales, pommes de terre et légumineuses
6,906	NOESY1D	0,19	0,02	0,2	Pain, céréales, pommes de terre et légumineuses
6,908	CPMG	0,23	0,005	0,2	Pain, céréales, pommes de terre et légumineuses
7,016	CPMG	0,18	0,02	0,3	Pain, céréales, pommes de terre et légumineuses
7,3699	CPMG	0,16	0,05	0,4	Pain, céréales, pommes de terre et légumineuses
8,9903	NOESY1D	0,17	0,04	0,3	Pain, céréales, pommes de terre et légumineuses
9,0303	NOESY1D	0,19	0,02	0,2	Pain, céréales, pommes de terre et légumineuses
9,0703	NOESY1D	0,17	0,03	0,2	Pain, céréales, pommes de terre et légumineuses
9,9107	NOESY1D	0,17	0,04	0,2	Pain, céréales, pommes de terre et légumineuses
0,91	NOESY1D	0,21	0,01	1,0	Produits céréaliers complets
0,9103	CPMG	0,21	0,01	0,5	Produits céréaliers complets
1,3093	CPMG	-0,19	0,02	0,5	Produits céréaliers complets
1,3103	NOESY1D	-0,16	0,046	1,0	Produits céréaliers complets
1,5903	CPMG	-0,17	0,04	0,7	Produits céréaliers complets
<b>1,6463 (Arginine, Lysine)</b>	CPMG	0,30	0,0002	0,04	Produits céréaliers complets
2,926	CPMG	-0,18	0,02	0,6	Produits céréaliers complets
5,864	CPMG	-0,17	0,04	0,7	Produits céréaliers complets
5,891	CPMG	-0,16	0,046	0,7	Produits céréaliers complets
6,105	CPMG	-0,20	0,02	0,5	Produits céréaliers complets
6,128	CPMG	-0,20	0,01	0,5	Produits céréaliers complets
6,215	CPMG	-0,20	0,01	0,5	Produits céréaliers complets
6,246	CPMG	-0,22	0,007	0,5	Produits céréaliers complets
6,567	CPMG	-0,16	0,05	0,7	Produits céréaliers complets
0,2063	NOESY1D	0,20	0,01	0,8	Produits de la mer
0,2353	CPMG	0,16	0,04	0,8	Produits de la mer
0,2463	NOESY1D	0,21	0,008	0,8	Produits de la mer
0,2863	NOESY1D	0,19	0,02	0,8	Produits de la mer
0,3083	CPMG	0,16	0,048	0,8	Produits de la mer
0,6463	NOESY1D	0,21	0,01	0,8	Produits de la mer
<b>0,91 (Leucine)</b>	NOESY1D	0,30	0,0002	0,04	Produits de la mer
<b>0,9103 (Leucine)</b>	CPMG	0,30	0,0001	0,03	Produits de la mer
1,3093	CPMG	-0,22	0,007	0,5	Produits de la mer
1,3103	NOESY1D	-0,17	0,03	0,8	Produits de la mer
1,3503	CPMG	-0,16	0,046	0,8	Produits de la mer
1,4323	CPMG	-0,19	0,02	0,8	Produits de la mer
2,1663	CPMG	-0,18	0,02	0,8	Produits de la mer
2,264	NOESY1D	0,16	0,04	0,8	Produits de la mer
4,4691	NOESY1D	0,16	0,049	0,8	Produits de la mer
5,726	CPMG	-0,19	0,02	0,8	Produits de la mer
5,83	CPMG	-0,24	0,003	0,4	Produits de la mer
5,864	CPMG	-0,16	0,04	0,8	Produits de la mer
6,686	NOESY1D	-0,18	0,03	0,8	Produits de la mer
7,9099	NOESY1D	-0,19	0,02	0,8	Produits de la mer

ANNEXE 8

0,2463	NOESY1D	-0,20	0,02	0,7	Produits sucrés
0,3263	NOESY1D	-0,19	0,02	0,7	Produits sucrés
0,7663	NOESY1D	0,16	0,04	0,7	Produits sucrés
1,6463	NOESY1D	0,17	0,04	0,7	Produits sucrés
1,8523	NOESY1D	0,18	0,03	0,7	Produits sucrés
2,1663	CPMG	0,17	0,03	0,9	Produits sucrés
2,264	NOESY1D	-0,17	0,04	0,7	Produits sucrés
2,267	CPMG	-0,22	0,007	0,7	Produits sucrés
2,4703	CPMG	-0,16	0,04	0,9	Produits sucrés
2,5013	CPMG	-0,17	0,03	0,9	Produits sucrés
3,948	CPMG	-0,22	0,006	0,7	Produits sucrés
4,035	NOESY1D	0,20	0,01	0,7	Produits sucrés
4,0352	CPMG	0,21	0,010	0,7	Produits sucrés
5,5441	CPMG	0,20	0,02	0,8	Produits sucrés
7,8299	CPMG	-0,18	0,03	0,9	Produits sucrés
0,3263	NOESY1D	0,18	0,02	0,6	Sel
0,7663	CPMG	-0,22	0,005	0,6	Sel
1,1663	NOESY1D	-0,18	0,03	0,6	Sel
1,6463	CPMG	0,18	0,03	0,8	Sel
2,4703	CPMG	0,23	0,004	0,6	Sel
2,6113	CPMG	0,21	0,01	0,6	Sel
3,8602	CPMG	0,17	0,04	0,8	Sel
3,8602	NOESY1D	0,16	0,04	0,6	Sel
3,92	CPMG	-0,17	0,04	0,8	Sel
3,9502	CPMG	0,21	0,008	0,6	Sel
4,7971	NOESY1D	0,17	0,03	0,6	Sel
4,8701	NOESY1D	0,21	0,009	0,6	Sel
4,8911	NOESY1D	0,21	0,009	0,6	Sel
4,9881	NOESY1D	0,19	0,02	0,6	Sel
6,286	NOESY1D	0,17	0,04	0,6	Sel
6,331	NOESY1D	0,17	0,03	0,6	Sel
6,86	CPMG	0,19	0,02	0,8	Sel
6,908	CPMG	0,18	0,03	0,8	Sel
7,6309	CPMG	-0,17	0,03	0,8	Sel
7,6859	NOESY1D	-0,16	0,05	0,6	Sel
9,6686	NOESY1D	0,16	0,05	0,6	Sel
0,969	CPMG	0,18	0,03	1,0	Viande, volaille, produits de la mer, oeufs
0,97	NOESY1D	0,16	0,05	1,0	Viande, volaille, produits de la mer, oeufs
3,948	CPMG	0,16	0,05	1,0	Viande, volaille, produits de la mer, oeufs
4,8571	CPMG	0,16	0,05	1,0	Viande, volaille, produits de la mer, oeufs
5,5101	CPMG	-0,18	0,03	1,0	Viande, volaille, produits de la mer, oeufs
6,567	CPMG	-0,17	0,04	1,0	Viande, volaille, produits de la mer, oeufs

R; correlation coefficient

Les associations présentées dans le Tableau 23 sont surlignées en gras. La validation 2D des signaux en italique est en cours sur la plateforme RMN.

<sup>1</sup> 159 individus ont été analysés.

<sup>2</sup> Les corrélations de Spearman ont été ajustées sur les caractéristiques suivantes : âge, IMC, statut tabagique, statut ménopausique, saison de prélèvement sanguin, nombre d'enregistrements alimentaires sur 24 heures et moyenne de l'apport énergétique quotidien pendant les 2 années suivant le prélèvement sanguin.

## Annexe 9 Associations entre les signaux RMN et les profils alimentaires obtenus par ACP chez les femmes

Déplacement chimique (ppm) (assignations)	Séquence <sup>1</sup>	R <sup>2</sup>	P	P-corrigée	Profils alimentaires obtenus par ACP
0,0337	NOESY1D	-0,18	0,03	0,9	Composante 1 Profil alimentaire "occidental"
0,2463	NOESY1D	-0,20	0,02	0,9	Composante 1 Profil alimentaire "occidental"
0,4093	CPMG	-0,16	0,048	0,9	Composante 1 Profil alimentaire "occidental"
<b>0,91 (Leucine)</b>	NOESY1D	-0,22	0,006	0,9	Composante 1 Profil alimentaire "occidental"
<b>0,9103 (Leucine)</b>	CPMG	-0,23	0,005	0,6	Composante 1 Profil alimentaire "occidental"
2,3863	CPMG	-0,18	0,03	0,9	Composante 1 Profil alimentaire "occidental"
2,4703	CPMG	0,18	0,03	0,9	Composante 1 Profil alimentaire "occidental"
2,4703	NOESY1D	0,17	0,04	0,9	Composante 1 Profil alimentaire "occidental"
2,6113	CPMG	0,17	0,03	0,9	Composante 1 Profil alimentaire "occidental"
3,3682	NOESY1D	0,17	0,04	0,9	Composante 1 Profil alimentaire "occidental"
3,4372	NOESY1D	0,18	0,03	0,9	Composante 1 Profil alimentaire "occidental"
3,4372	CPMG	0,16	0,047	0,9	Composante 1 Profil alimentaire "occidental"
3,918	NOESY1D	-0,18	0,02	0,9	Composante 1 Profil alimentaire "occidental"
3,92	CPMG	-0,18	0,03	0,9	Composante 1 Profil alimentaire "occidental"
3,9782	CPMG	-0,16	0,048	0,9	Composante 1 Profil alimentaire "occidental"
5,2061	CPMG	0,20	0,01	0,9	Composante 1 Profil alimentaire "occidental"
5,2061	NOESY1D	0,16	0,047	0,9	Composante 1 Profil alimentaire "occidental"
5,966	NOESY1D	0,20	0,01	0,9	Composante 1 Profil alimentaire "occidental"
6,286	NOESY1D	0,17	0,04	0,9	Composante 1 Profil alimentaire "occidental"
<b>7,6459</b>	CPMG	-0,23	0,005	0,6	Composante 1 Profil alimentaire "occidental"
7,6859	CPMG	-0,17	0,04	0,9	Composante 1 Profil alimentaire "occidental"
7,7729	NOESY1D	0,18	0,02	0,9	Composante 1 Profil alimentaire "occidental"
0,1263	NOESY1D	0,20	0,01	0,4	Composante 2 Profil alimentaire "sain"
0,2353	CPMG	0,16	0,046	0,5	Composante 2 Profil alimentaire "sain"
0,2463	NOESY1D	0,16	0,049	0,4	Composante 2 Profil alimentaire "sain"
0,91	NOESY1D	0,21	0,01	0,4	Composante 2 Profil alimentaire "sain"
0,9103	CPMG	0,17	0,04	0,5	Composante 2 Profil alimentaire "sain"
<b>1,1663 (Lipoprotéines (CH2))</b>	NOESY1D	-0,26	0,001	0,3	Composante 2 Profil alimentaire "sain"
<b>1,1663 (Lipoprotéines (CH2))</b>	CPMG	-0,23	0,005	0,3	Composante 2 Profil alimentaire "sain"
1,2063	CPMG	-0,19	0,02	0,4	Composante 2 Profil alimentaire "sain"
1,2063	NOESY1D	-0,17	0,04	0,4	Composante 2 Profil alimentaire "sain"
1,3093	CPMG	-0,18	0,03	0,5	Composante 2 Profil alimentaire "sain"
1,3103	NOESY1D	-0,20	0,02	0,4	Composante 2 Profil alimentaire "sain"
1,5223	CPMG	-0,16	0,04	0,5	Composante 2 Profil alimentaire "sain"
1,5503	NOESY1D	-0,17	0,04	0,4	Composante 2 Profil alimentaire "sain"
1,5903	CPMG	-0,20	0,02	0,4	Composante 2 Profil alimentaire "sain"
1,5903	NOESY1D	-0,17	0,03	0,4	Composante 2 Profil alimentaire "sain"
1,6303	NOESY1D	-0,17	0,04	0,4	Composante 2 Profil alimentaire "sain"
1,6363	NOESY1D	-0,17	0,03	0,4	Composante 2 Profil alimentaire "sain"
1,6463	NOESY1D	-0,19	0,02	0,4	Composante 2 Profil alimentaire "sain"
<b>1,9263 (Lipides CH2-C=C)</b>	CPMG	-0,22	0,006	0,3	Composante 2 Profil alimentaire "sain"
2,0633	NOESY1D	-0,16	0,048	0,4	Composante 2 Profil alimentaire "sain"
2,1663	CPMG	-0,20	0,01	0,4	Composante 2 Profil alimentaire "sain"
2,1753	NOESY1D	-0,19	0,02	0,4	Composante 2 Profil alimentaire "sain"
2,264	NOESY1D	0,21	0,008	0,4	Composante 2 Profil alimentaire "sain"
2,267	CPMG	0,17	0,03	0,5	Composante 2 Profil alimentaire "sain"
2,3863	CPMG	0,21	0,01	0,4	Composante 2 Profil alimentaire "sain"
2,5013	CPMG	0,20	0,01	0,4	Composante 2 Profil alimentaire "sain"
2,5033	NOESY1D	0,18	0,03	0,4	Composante 2 Profil alimentaire "sain"
2,817	NOESY1D	-0,21	0,008	0,4	Composante 2 Profil alimentaire "sain"
<b>2,8863</b>	CPMG	0,28	0,0004	0,08	Composante 2 Profil alimentaire "sain"
3,0702	NOESY1D	0,18	0,02	0,4	Composante 2 Profil alimentaire "sain"
4,2152	NOESY1D	-0,20	0,01	0,4	Composante 2 Profil alimentaire "sain"
5,1661	NOESY1D	-0,19	0,02	0,4	Composante 2 Profil alimentaire "sain"
5,2061	NOESY1D	-0,18	0,03	0,4	Composante 2 Profil alimentaire "sain"



## ANNEXE 9

6,128	CPMG	-0,16	0,045	0,5	Composante 2 Profil alimentaire "sain"
<b>6,215</b>	CPMG	-0,23	0,004	0,3	Composante 2 Profil alimentaire "sain"
6,686	CPMG	0,18	0,02	0,5	Composante 2 Profil alimentaire "sain"
6,726	CPMG	0,18	0,03	0,5	Composante 2 Profil alimentaire "sain"
6,766	NOESY1D	0,18	0,03	0,4	Composante 2 Profil alimentaire "sain"
6,766	CPMG	0,16	0,045	0,5	Composante 2 Profil alimentaire "sain"
6,926	NOESY1D	0,18	0,03	0,4	Composante 2 Profil alimentaire "sain"
6,941	CPMG	0,16	0,049	0,5	Composante 2 Profil alimentaire "sain"
7,3659	NOESY1D	0,21	0,01	0,4	Composante 2 Profil alimentaire "sain"
7,3699	CPMG	0,21	0,009	0,4	Composante 2 Profil alimentaire "sain"
8,9103	NOESY1D	0,19	0,02	0,4	Composante 2 Profil alimentaire "sain"
8,9903	NOESY1D	0,17	0,03	0,4	Composante 2 Profil alimentaire "sain"
0,0017	CPMG	-0,16	0,047	0,4	Composante 3 Régime à base de produits sucrés
1,8233	CPMG	0,18	0,03	0,3	Composante 3 Régime à base de produits sucrés
1,8493	CPMG	0,22	0,008	0,2	Composante 3 Régime à base de produits sucrés
2,277	CPMG	0,19	0,02	0,2	Composante 3 Régime à base de produits sucrés
2,4703	CPMG	-0,18	0,02	0,2	Composante 3 Régime à base de produits sucrés
<b>2,5693</b>	CPMG	0,25	0,002	0,1	Composante 3 Régime à base de produits sucrés
<b>2,646 (Citrate)</b>	NOESY1D	0,26	0,001	0,3	Composante 3 Régime à base de produits sucrés
2,646	CPMG	0,22	0,007	0,2	Composante 3 Régime à base de produits sucrés
4,8571	CPMG	-0,18	0,02	0,2	Composante 3 Régime à base de produits sucrés
4,9901	CPMG	-0,16	0,048	0,4	Composante 3 Régime à base de produits sucrés
5,0391	CPMG	-0,17	0,04	0,3	Composante 3 Régime à base de produits sucrés
5,0761	CPMG	-0,19	0,02	0,2	Composante 3 Régime à base de produits sucrés
5,1261	CPMG	-0,18	0,03	0,3	Composante 3 Régime à base de produits sucrés
5,5441	CPMG	0,21	0,01	0,2	Composante 3 Régime à base de produits sucrés
5,726	NOESY1D	-0,19	0,02	1,0	Composante 3 Régime à base de produits sucrés
5,83	NOESY1D	-0,20	0,01	1,0	Composante 3 Régime à base de produits sucrés
6,07	NOESY1D	-0,17	0,04	1,0	Composante 3 Régime à base de produits sucrés
<b>6,128</b>	CPMG	0,26	0,001	0,1	Composante 3 Régime à base de produits sucrés
<b>6,215</b>	CPMG	0,25	0,002	0,1	Composante 3 Régime à base de produits sucrés
<b>6,246</b>	CPMG	0,23	0,005	0,2	Composante 3 Régime à base de produits sucrés
6,286	NOESY1D	-0,17	0,04	1,0	Composante 3 Régime à base de produits sucrés
6,331	NOESY1D	-0,19	0,02	1,0	Composante 3 Régime à base de produits sucrés
6,941	CPMG	0,20	0,01	0,2	Composante 3 Régime à base de produits sucrés
<b>6,966</b>	CPMG	0,24	0,003	0,1	Composante 3 Régime à base de produits sucrés
6,966	NOESY1D	0,16	0,04	1,0	Composante 3 Régime à base de produits sucrés
7,6309	CPMG	0,20	0,02	0,2	Composante 3 Régime à base de produits sucrés
7,6459	CPMG	0,21	0,01	0,2	Composante 3 Régime à base de produits sucrés
<b>7,6859</b>	CPMG	0,25	0,002	0,1	Composante 3 Régime à base de produits sucrés
7,6859	NOESY1D	0,16	0,0499	1,0	Composante 3 Régime à base de produits sucrés
7,8299	CPMG	-0,21	0,01	0,2	Composante 3 Régime à base de produits sucrés
7,8699	CPMG	-0,20	0,01	0,2	Composante 3 Régime à base de produits sucrés
7,9089	CPMG	-0,21	0,01	0,2	Composante 3 Régime à base de produits sucrés
7,9709	CPMG	-0,21	0,01	0,2	Composante 3 Régime à base de produits sucrés
8,6411	CPMG	-0,20	0,01	0,2	Composante 3 Régime à base de produits sucrés
8,8662	CPMG	-0,20	0,02	0,2	Composante 3 Régime à base de produits sucrés
8,9403	CPMG	-0,17	0,03	0,3	Composante 3 Régime à base de produits sucrés

R; correlation coefficient

Les associations présentées dans le Tableau 24 sont surlignées en gras. La validation 2D des signaux en italique est en cours sur la plateforme RMN.

<sup>1</sup> 159 individus ont été analysés.

<sup>2</sup> Les corrélations de Spearman ont été ajustées sur les caractéristiques suivantes : âge, IMC, statut tabagique, statut ménopausique, saison de prélèvement sanguin, nombre d'enregistrements alimentaires sur 24 heures et moyenne de l'apport énergétique quotidien pendant les 2 années suivant le prélèvement sanguin.

## Annexe 10 Associations entre les signaux RMN et les composantes alimentaires du mPNNS-GS chez les hommes

Déplacement chimique (ppm) (assignations)	Séquence <sup>1</sup>	R <sup>2</sup>	P	P-corrigée	Composantes mPNNS-GS
0,0871	NOESY1D	-0,23	0,009	0,3	Alcool
0,1271	NOESY1D	-0,19	0,03	0,5	Alcool
0,1651	CPMG	-0,21	0,01	0,4	Alcool
0,6471	NOESY1D	0,18	0,04	0,5	Alcool
1,021	NOESY1D	-0,18	0,04	0,5	Alcool
1,051	CPMG	0,19	0,03	0,5	Alcool
1,0661	NOESY1D	0,20	0,02	0,5	Alcool
1,0891	NOESY1D	0,17	0,048	0,5	Alcool
<b>1,1271 (Propylène glycol)</b>	CPMG	0,33	0,00008	0,009	Alcool
<b>1,145</b>	CPMG	-0,35	0,00003	0,008	Alcool
1,1671	NOESY1D	0,19	0,03	0,5	Alcool
1,1671	CPMG	0,19	0,03	0,5	Alcool
1,7511	NOESY1D	0,24	0,005	0,3	Alcool
1,7911	CPMG	-0,21	0,02	0,4	Alcool
1,7911	NOESY1D	0,18	0,04	0,5	Alcool
1,8221	CPMG	-0,26	0,002	0,1	Alcool
1,927	CPMG	0,24	0,005	0,2	Alcool
<b>1,9271 (Lipides (CH<sub>2</sub>CH<sub>2</sub>COOC))</b>	NOESY1D	0,30	0,0004	0,05	Alcool
2,017	CPMG	-0,24	0,005	0,2	Alcool
2,3111	NOESY1D	0,27	0,002	0,1	Alcool
2,3511	NOESY1D	0,21	0,02	0,4	Alcool
2,3871	NOESY1D	0,18	0,04	0,5	Alcool
2,647	NOESY1D	-0,23	0,008	0,3	Alcool
<b>3,167 (Choline, glucose)</b>	NOESY1D	0,32	0,0002	0,03	Alcool
<b>3,167 (Choline, glucose)</b>	CPMG	0,29	0,0006	0,04	Alcool
3,269	CPMG	0,21	0,02	0,4	Alcool
3,918	NOESY1D	-0,18	0,04	0,5	Alcool
4,191	NOESY1D	0,22	0,01	0,4	Alcool
4,311	NOESY1D	0,18	0,03	0,5	Alcool
5,2069	CPMG	0,17	0,047	0,5	Alcool
5,5289	CPMG	-0,18	0,04	0,5	Alcool
5,5509	NOESY1D	-0,19	0,03	0,5	Alcool
7,1218	CPMG	0,18	0,04	0,5	Alcool
7,6467	NOESY1D	-0,17	0,048	0,5	Alcool
8,4309	CPMG	0,18	0,04	0,5	Alcool
9,0231	CPMG	0,19	0,03	0,5	Alcool
9,6654	CPMG	0,19	0,03	0,5	Alcool
9,7794	CPMG	0,18	0,04	0,5	Alcool
0,1271	NOESY1D	-0,18	0,04	1,0	Boissons sucrées
4,075	NOESY1D	0,18	0,04	1,0	Boissons sucrées
7,023	CPMG	0,17	0,047	1,0	Boissons sucrées
0,0871	NOESY1D	0,24	0,005	0,4	Fruits et légumes
0,1271	NOESY1D	0,28	0,001	0,3	Fruits et légumes
0,2071	NOESY1D	0,22	0,01	0,6	Fruits et légumes
0,2471	NOESY1D	0,25	0,004	0,4	Fruits et légumes
1,1271	CPMG	-0,17	0,0495	0,8	Fruits et légumes
1,1671	CPMG	-0,18	0,04	0,8	Fruits et légumes
1,7911	CPMG	0,18	0,04	0,8	Fruits et légumes
1,899	CPMG	0,22	0,01	0,6	Fruits et légumes
2,8471	NOESY1D	0,20	0,02	0,8	Fruits et légumes
2,8871	NOESY1D	0,21	0,02	0,7	Fruits et légumes
2,9271	CPMG	0,18	0,03	0,8	Fruits et légumes
3,07	CPMG	0,22	0,01	0,6	Fruits et légumes
3,167	NOESY1D	-0,18	0,04	1,0	Fruits et légumes
4,215	NOESY1D	-0,20	0,02	0,8	Fruits et légumes
5,5099	CPMG	-0,18	0,04	0,8	Fruits et légumes
5,8308	CPMG	-0,22	0,01	0,6	Fruits et légumes
5,8668	CPMG	-0,21	0,02	0,6	Fruits et légumes
6,2068	CPMG	-0,27	0,002	0,4	Fruits et légumes
6,2468	NOESY1D	-0,17	0,046	1,0	Fruits et légumes
6,2478	CPMG	-0,21	0,02	0,6	Fruits et légumes
7,5097	CPMG	-0,17	0,0497	0,8	Fruits et légumes
9,2002	CPMG	-0,17	0,047	0,8	Fruits et légumes
0,0003	CPMG	-0,23	0,007	0,8	Lait et produits laitiers

## ANNEXE 10

0,3811	CPMG	-0,17	0,0499	0,8	Lait et produits laitiers
2,6121	CPMG	0,18	0,04	0,8	Lait et produits laitiers
3,314	NOESY1D	-0,19	0,03	1,0	Lait et produits laitiers
<b>3,366 (Glucose)</b>	NOESY1D	0,30	0,0006	0,1	Lait et produits laitiers
3,437	CPMG	0,20	0,02	0,8	Lait et produits laitiers
3,437	NOESY1D	0,19	0,03	1,0	Lait et produits laitiers
3,714	CPMG	0,18	0,03	0,8	Lait et produits laitiers
3,807	CPMG	0,19	0,02	0,8	Lait et produits laitiers
3,86	CPMG	0,20	0,02	0,8	Lait et produits laitiers
5,4709	NOESY1D	-0,20	0,02	1,0	Lait et produits laitiers
5,9668	CPMG	-0,19	0,03	0,8	Lait et produits laitiers
8,9111	NOESY1D	-0,19	0,03	1,0	Lait et produits laitiers
0,9997	CPMG	0,20	0,02	0,9	Matière grasse ajoutée totale
1,021	CPMG	-0,18	0,04	0,9	Matière grasse ajoutée totale
1,051	NOESY1D	0,23	0,009	0,9	Matière grasse ajoutée totale
1,051	CPMG	0,20	0,02	0,9	Matière grasse ajoutée totale
1,6471	CPMG	-0,18	0,04	0,9	Matière grasse ajoutée totale
3,588	NOESY1D	-0,20	0,02	0,9	Matière grasse ajoutée totale
4,311	NOESY1D	-0,20	0,02	0,9	Matière grasse ajoutée totale
4,351	NOESY1D	-0,20	0,02	0,9	Matière grasse ajoutée totale
5,8288	NOESY1D	-0,17	0,048	1,0	Matière grasse ajoutée totale
6,8608	CPMG	-0,21	0,02	0,9	Matière grasse ajoutée totale
7,1508	NOESY1D	-0,20	0,02	0,9	Matière grasse ajoutée totale
<b>7,159 (Tyrosine)</b>	CPMG	-0,28	0,0009	0,2	Matière grasse ajoutée totale
7,1828	NOESY1D	-0,20	0,02	0,9	Matière grasse ajoutée totale
7,2777	CPMG	-0,17	0,0498	0,9	Matière grasse ajoutée totale
8,2698	CPMG	-0,18	0,03	0,9	Matière grasse ajoutée totale
8,4309	CPMG	-0,18	0,04	0,9	Matière grasse ajoutée totale
9,0361	CPMG	-0,17	0,049	0,9	Matière grasse ajoutée totale
9,2002	CPMG	-0,20	0,02	0,9	Matière grasse ajoutée totale
1,1671	CPMG	-0,19	0,03	0,8	Matière grasse végétale ajoutée
2,017	CPMG	0,17	0,045	0,8	Matière grasse végétale ajoutée
2,6271	NOESY1D	0,22	0,01	0,9	Matière grasse végétale ajoutée
2,705	CPMG	0,19	0,03	0,8	Matière grasse végétale ajoutée
5,8668	CPMG	-0,19	0,03	0,8	Matière grasse végétale ajoutée
5,8878	NOESY1D	-0,17	0,045	1,0	Matière grasse végétale ajoutée
6,1278	NOESY1D	-0,21	0,01	0,9	Matière grasse végétale ajoutée
6,1668	CPMG	-0,21	0,01	0,8	Matière grasse végétale ajoutée
6,2068	CPMG	-0,25	0,003	0,5	Matière grasse végétale ajoutée
6,2128	NOESY1D	-0,21	0,02	0,9	Matière grasse végétale ajoutée
6,2468	NOESY1D	-0,27	0,002	0,4	Matière grasse végétale ajoutée
6,2478	CPMG	-0,23	0,008	0,6	Matière grasse végétale ajoutée
6,9088	CPMG	0,19	0,03	0,8	Matière grasse végétale ajoutée
7,8307	CPMG	0,18	0,04	0,8	Matière grasse végétale ajoutée
8,1448	CPMG	0,18	0,04	0,8	Matière grasse végétale ajoutée
8,3508	CPMG	0,19	0,03	0,8	Matière grasse végétale ajoutée
8,4309	CPMG	0,24	0,005	0,5	Matière grasse végétale ajoutée
0,7671	CPMG	-0,23	0,009	0,1	Pain, céréales, pommes de terre et légumineuses
0,9697	NOESY1D	-0,22	0,01	0,3	Pain, céréales, pommes de terre et légumineuses
0,9997	CPMG	0,17	0,046	0,3	Pain, céréales, pommes de terre et légumineuses
0,9997	NOESY1D	0,18	0,03	0,4	Pain, céréales, pommes de terre et légumineuses
1,2071	CPMG	0,24	0,004	0,09	Pain, céréales, pommes de terre et légumineuses
1,2071	NOESY1D	0,20	0,02	0,3	Pain, céréales, pommes de terre et légumineuses
1,3511	CPMG	0,26	0,003	0,08	Pain, céréales, pommes de terre et légumineuses
1,3911	CPMG	0,20	0,02	0,2	Pain, céréales, pommes de terre et légumineuses
1,4311	CPMG	0,18	0,03	0,2	Pain, céréales, pommes de terre et légumineuses
1,5211	CPMG	0,20	0,02	0,2	Pain, céréales, pommes de terre et légumineuses
1,5221	NOESY1D	0,23	0,009	0,2	Pain, céréales, pommes de terre et légumineuses
1,5271	CPMG	0,28	0,001	0,04	Pain, céréales, pommes de terre et légumineuses
1,6311	CPMG	0,21	0,01	0,1	Pain, céréales, pommes de terre et légumineuses
1,6361	CPMG	0,18	0,04	0,2	Pain, céréales, pommes de terre et légumineuses
1,6471	CPMG	-0,20	0,02	0,2	Pain, céréales, pommes de terre et légumineuses
1,6471	NOESY1D	-0,22	0,01	0,3	Pain, céréales, pommes de terre et légumineuses
1,7511	NOESY1D	-0,19	0,03	0,4	Pain, céréales, pommes de terre et légumineuses
1,7911	CPMG	0,21	0,02	0,1	Pain, céréales, pommes de terre et légumineuses
1,8221	CPMG	0,25	0,004	0,08	Pain, céréales, pommes de terre et légumineuses
<b>1,927 (Lipides (CH2-CH))</b>	CPMG	-0,29	0,0008	0,04	Pain, céréales, pommes de terre et légumineuses
1,9271	NOESY1D	-0,27	0,002	0,08	Pain, céréales, pommes de terre et légumineuses
2,057	NOESY1D	-0,21	0,01	0,3	Pain, céréales, pommes de terre et légumineuses
2,0961	CPMG	0,22	0,01	0,1	Pain, céréales, pommes de terre et légumineuses
2,1511	CPMG	0,25	0,003	0,08	Pain, céréales, pommes de terre et légumineuses

## ANNEXE 10

2,1671	CPMG	0,28	0,001	0,04	Pain, céréales, pommes de terre et légumineuses
2,1741	NOESY1D	0,26	0,003	0,1	Pain, céréales, pommes de terre et légumineuses
<b>2,3511 (Pyruvate)</b>	NOESY1D	-0,29	0,0009	0,06	Pain, céréales, pommes de terre et légumineuses
2,708	NOESY1D	0,25	0,004	0,1	Pain, céréales, pommes de terre et légumineuses
2,715	CPMG	0,24	0,006	0,09	Pain, céréales, pommes de terre et légumineuses
2,9271	CPMG	-0,17	0,047	0,3	Pain, céréales, pommes de terre et légumineuses
2,9771	CPMG	-0,22	0,01	0,1	Pain, céréales, pommes de terre et légumineuses
3,129	CPMG	-0,19	0,03	0,2	Pain, céréales, pommes de terre et légumineuses
<b>3,167 (Choline, glucose)</b>	CPMG	-0,28	0,0009	0,04	Pain, céréales, pommes de terre et légumineuses
3,167	NOESY1D	-0,19	0,03	0,4	Pain, céréales, pommes de terre et légumineuses
3,312	CPMG	0,22	0,01	0,1	Pain, céréales, pommes de terre et légumineuses
3,314	NOESY1D	0,27	0,002	0,08	Pain, céréales, pommes de terre et légumineuses
3,368	CPMG	-0,22	0,01	0,1	Pain, céréales, pommes de terre et légumineuses
3,437	CPMG	-0,24	0,005	0,09	Pain, céréales, pommes de terre et légumineuses
3,496	CPMG	-0,19	0,03	0,2	Pain, céréales, pommes de terre et légumineuses
3,616	CPMG	-0,27	0,002	0,06	Pain, céréales, pommes de terre et légumineuses
3,714	CPMG	-0,22	0,01	0,1	Pain, céréales, pommes de terre et légumineuses
3,807	CPMG	-0,20	0,02	0,2	Pain, céréales, pommes de terre et légumineuses
3,951	CPMG	0,19	0,02	0,2	Pain, céréales, pommes de terre et légumineuses
<b>4,031 (Créatine, glycéryl)</b>	NOESY1D	0,33	0,0001	0,02	Pain, céréales, pommes de terre et légumineuses
<b>4,055 (Choline)</b>	CPMG	0,37	0,00001	0,003	Pain, céréales, pommes de terre et légumineuses
<b>4,075 (Lactate)</b>	NOESY1D	-0,30	0,0004	0,04	Pain, céréales, pommes de terre et légumineuses
4,209	CPMG	0,22	0,01	0,1	Pain, céréales, pommes de terre et légumineuses
4,215	NOESY1D	0,18	0,04	0,4	Pain, céréales, pommes de terre et légumineuses
5,1669	CPMG	0,25	0,004	0,08	Pain, céréales, pommes de terre et légumineuses
5,1669	NOESY1D	0,25	0,004	0,1	Pain, céréales, pommes de terre et légumineuses
5,2469	CPMG	0,22	0,01	0,1	Pain, céréales, pommes de terre et légumineuses
5,2469	NOESY1D	0,20	0,02	0,3	Pain, céréales, pommes de terre et légumineuses
5,3509	CPMG	0,20	0,02	0,2	Pain, céréales, pommes de terre et légumineuses
5,4309	CPMG	0,20	0,02	0,2	Pain, céréales, pommes de terre et légumineuses
6,2068	CPMG	0,21	0,01	0,1	Pain, céréales, pommes de terre et légumineuses
6,2468	NOESY1D	0,19	0,03	0,4	Pain, céréales, pommes de terre et légumineuses
6,8608	CPMG	-0,20	0,02	0,2	Pain, céréales, pommes de terre et légumineuses
7,1218	CPMG	-0,18	0,04	0,2	Pain, céréales, pommes de terre et légumineuses
0,9107	CPMG	0,18	0,04	0,5	Produits céréaliers complets
0,9111	NOESY1D	0,19	0,03	1,0	Produits céréaliers complets
1,7911	CPMG	0,20	0,02	0,4	Produits céréaliers complets
2,017	CPMG	0,24	0,006	0,3	Produits céréaliers complets
2,8481	CPMG	-0,19	0,03	0,4	Produits céréaliers complets
3,028	CPMG	-0,18	0,03	0,5	Produits céréaliers complets
3,167	NOESY1D	-0,18	0,04	1,0	Produits céréaliers complets
3,311	NOESY1D	-0,18	0,04	1,0	Produits céréaliers complets
5,8908	CPMG	-0,17	0,046	0,6	Produits céréaliers complets
5,9668	CPMG	-0,20	0,02	0,4	Produits céréaliers complets
<b>6,1058</b>	CPMG	-0,30	0,0005	0,1	Produits céréaliers complets
6,1328	CPMG	-0,25	0,003	0,2	Produits céréaliers complets
6,1668	CPMG	-0,21	0,01	0,4	Produits céréaliers complets
6,2478	CPMG	-0,25	0,003	0,2	Produits céréaliers complets
6,3398	CPMG	-0,24	0,005	0,3	Produits céréaliers complets
6,3668	CPMG	-0,21	0,01	0,4	Produits céréaliers complets
6,4068	CPMG	-0,21	0,02	0,4	Produits céréaliers complets
6,4468	CPMG	-0,20	0,02	0,4	Produits céréaliers complets
6,5068	CPMG	-0,19	0,02	0,4	Produits céréaliers complets
9,4983	CPMG	-0,18	0,04	0,5	Produits céréaliers complets
0,1271	NOESY1D	0,18	0,04	0,9	Produits de la mer
0,3271	NOESY1D	0,18	0,04	0,9	Produits de la mer
0,9107	CPMG	0,28	0,001	0,2	Produits de la mer
2,0631	CPMG	-0,17	0,04	0,9	Produits de la mer
2,647	NOESY1D	-0,22	0,01	0,8	Produits de la mer
2,6471	CPMG	-0,21	0,02	0,7	Produits de la mer
2,873	CPMG	0,19	0,03	0,7	Produits de la mer
3,366	NOESY1D	0,22	0,01	0,8	Produits de la mer
3,367	NOESY1D	-0,19	0,03	0,9	Produits de la mer
3,497	NOESY1D	-0,23	0,009	0,8	Produits de la mer
3,791	NOESY1D	-0,20	0,02	0,9	Produits de la mer
3,793	NOESY1D	-0,21	0,01	0,8	Produits de la mer
3,807	NOESY1D	-0,18	0,03	0,9	Produits de la mer
3,86	NOESY1D	-0,17	0,049	0,9	Produits de la mer
3,952	NOESY1D	-0,18	0,04	0,9	Produits de la mer
4,215	NOESY1D	-0,18	0,03	0,9	Produits de la mer
6,2068	CPMG	-0,22	0,01	0,7	Produits de la mer

## ANNEXE 10

7,023	CPMG	0,19	0,03	0,7	Produits de la mer
7,4307	CPMG	0,19	0,03	0,7	Produits de la mer
7,7807	CPMG	0,24	0,006	0,7	Produits de la mer
7,8307	CPMG	0,20	0,02	0,7	Produits de la mer
7,8707	CPMG	0,19	0,03	0,7	Produits de la mer
1,767	CPMG	-0,18	0,04	0,9	Produits sucrés
1,8501	CPMG	0,21	0,02	0,9	Produits sucrés
1,8511	NOESY1D	0,18	0,04	0,8	Produits sucrés
2,354	CPMG	0,17	0,045	0,9	Produits sucrés
2,372	CPMG	-0,23	0,008	0,9	Produits sucrés
2,411	CPMG	0,23	0,007	0,9	Produits sucrés
2,4711	CPMG	-0,17	0,04	0,9	Produits sucrés
3,044	CPMG	-0,21	0,02	0,9	Produits sucrés
5,5109	NOESY1D	0,18	0,04	0,8	Produits sucrés
6,1058	CPMG	0,18	0,04	0,9	Produits sucrés
6,1668	CPMG	0,17	0,04	0,9	Produits sucrés
6,2478	CPMG	0,19	0,03	0,9	Produits sucrés
6,4068	NOESY1D	-0,18	0,04	0,8	Produits sucrés
6,5668	NOESY1D	-0,17	0,046	0,8	Produits sucrés
7,8707	CPMG	-0,17	0,048	0,9	Produits sucrés
0,0871	NOESY1D	0,21	0,01	0,4	Sel
2,3511	NOESY1D	-0,25	0,004	0,4	Sel
2,354	CPMG	-0,17	0,048	0,9	Sel
2,6121	CPMG	-0,18	0,04	0,9	Sel
2,6271	NOESY1D	-0,22	0,01	0,4	Sel
3,311	NOESY1D	0,20	0,02	0,5	Sel
3,61	NOESY1D	-0,18	0,03	0,5	Sel
3,616	CPMG	-0,19	0,02	0,9	Sel
3,951	NOESY1D	-0,19	0,03	0,5	Sel
4,021	CPMG	-0,20	0,02	0,9	Sel
4,031	NOESY1D	0,21	0,02	0,5	Sel
4,055	CPMG	0,25	0,004	0,7	Sel
<b>4,075 (Lactate)</b>	NOESY1D	-0,32	0,0002	0,04	Sel
5,5109	NOESY1D	-0,20	0,02	0,5	Sel
6,1058	CPMG	-0,19	0,02	0,9	Sel
6,1328	CPMG	-0,23	0,007	0,7	Sel
6,5268	NOESY1D	0,18	0,03	0,5	Sel
6,6868	NOESY1D	0,21	0,01	0,4	Sel
6,7668	NOESY1D	0,19	0,03	0,5	Sel
6,9268	NOESY1D	0,22	0,01	0,4	Sel
6,9668	NOESY1D	0,22	0,01	0,4	Sel
7,1508	NOESY1D	0,18	0,04	0,5	Sel
7,3667	NOESY1D	0,20	0,02	0,5	Sel
9,3102	NOESY1D	-0,19	0,03	0,5	Sel
3,314	NOESY1D	-0,18	0,04	1,0	Viande, volaille, produits de la mer, oeufs
6,5068	CPMG	-0,18	0,04	1,0	Viande, volaille, produits de la mer, oeufs
6,5268	CPMG	-0,19	0,03	1,0	Viande, volaille, produits de la mer, oeufs
7,9817	CPMG	-0,18	0,03	1,0	Viande, volaille, produits de la mer, oeufs

R; correlation coefficient

Les associations présentées dans le Tableau 27 sont surlignées en gras. La validation 2D des signaux en italique est en cours sur la plateforme RMN.

<sup>1</sup> 140 individus ont été analysés pour la séquence CPMG et 138 pour la séquence NOESY1D.

<sup>2</sup> Les corrélations de Spearman ont été ajustées sur les caractéristiques suivantes : âge, IMC, statut tabagique, saison de prélèvement sanguin, nombre d'enregistrements alimentaires sur 24 heures et moyenne de l'apport énergétique quotidien pendant les 2 années suivant le prélèvement sanguin.

## Annexe 11 Associations entre les signaux RMN et les profils alimentaires obtenus par ACP chez les hommes

Déplacement chimique (ppm) (assignations)	Séquence <sup>1</sup>	R <sup>2</sup>	P	P-corrigée	Profils alimentaires obtenus par ACP
0,0003	CPMG	0,17	0,04	1,0	Composante 1 Profil alimentaire "occidental"
1,6471	CPMG	- 0,18	0,03	1,0	Composante 1 Profil alimentaire "occidental"
2,98	CPMG	- 0,19	0,02	1,0	Composante 1 Profil alimentaire "occidental"
<b>3,028 (Créatine)</b>	CPMG	0,29	0,0008	0,2	Composante 1 Profil alimentaire "occidental"
3,07	NOESY1D	- 0,18	0,04	0,8	Composante 1 Profil alimentaire "occidental"
3,366	NOESY1D	- 0,19	0,03	0,8	Composante 1 Profil alimentaire "occidental"
3,588	NOESY1D	- 0,19	0,03	0,8	Composante 1 Profil alimentaire "occidental"
5,0729	NOESY1D	- 0,18	0,04	0,8	Composante 1 Profil alimentaire "occidental"
5,5099	CPMG	0,18	0,04	1,0	Composante 1 Profil alimentaire "occidental"
7,159	CPMG	- 0,21	0,02	1,0	Composante 1 Profil alimentaire "occidental"
<b>7,5097</b>	CPMG	0,21	0,02	1,0	Composante 1 Profil alimentaire "occidental"
7,8707	NOESY1D	0,20	0,02	0,8	Composante 1 Profil alimentaire "occidental"
7,9077	NOESY1D	0,18	0,04	0,8	Composante 1 Profil alimentaire "occidental"
7,9267	NOESY1D	0,19	0,03	0,8	Composante 1 Profil alimentaire "occidental"
8,3108	NOESY1D	0,19	0,03	0,8	Composante 1 Profil alimentaire "occidental"
8,3508	NOESY1D	0,20	0,02	0,8	Composante 1 Profil alimentaire "occidental"
8,5909	NOESY1D	0,18	0,04	0,8	Composante 1 Profil alimentaire "occidental"
8,6309	NOESY1D	0,20	0,02	0,8	Composante 1 Profil alimentaire "occidental"
8,671	NOESY1D	0,19	0,03	0,8	Composante 1 Profil alimentaire "occidental"
8,87	CPMG	- 0,18	0,04	1,0	Composante 1 Profil alimentaire "occidental"
0,0471	NOESY1D	0,17	0,046	0,9	Composante 2 Profil alimentaire "sain"
<b>0,0871</b>	NOESY1D	0,24	0,005	0,4	Composante 2 Profil alimentaire "sain"
<b>0,1271</b>	NOESY1D	0,28	0,001	0,1	Composante 2 Profil alimentaire "sain"
<b>0,2071</b>	NOESY1D	0,22	0,01	0,4	Composante 2 Profil alimentaire "sain"
<b>0,2471</b>	NOESY1D	0,28	0,001	0,1	Composante 2 Profil alimentaire "sain"
0,9107	CPMG	0,18	0,04	0,7	Composante 2 Profil alimentaire "sain"
0,9111	NOESY1D	0,18	0,04	0,9	Composante 2 Profil alimentaire "sain"
1,0651	CPMG	- 0,17	0,048	0,7	Composante 2 Profil alimentaire "sain"
<b>1,1671 (Lipoprotéines (CH2))</b>	CPMG	- 0,21	0,01	0,5	Composante 2 Profil alimentaire "sain"
1,7911	CPMG	0,20	0,02	0,5	Composante 2 Profil alimentaire "sain"
1,8221	CPMG	0,19	0,03	0,5	Composante 2 Profil alimentaire "sain"
1,899	CPMG	0,20	0,02	0,5	Composante 2 Profil alimentaire "sain"
2,017	CPMG	0,21	0,02	0,5	Composante 2 Profil alimentaire "sain"
2,8471	NOESY1D	0,20	0,02	0,8	Composante 2 Profil alimentaire "sain"
2,8871	NOESY1D	0,18	0,04	0,9	Composante 2 Profil alimentaire "sain"
<b>3,167 (Choline, glucose)</b>	NOESY1D	- 0,23	0,009	0,4	Composante 2 Profil alimentaire "sain"
3,918	NOESY1D	0,18	0,04	0,9	Composante 2 Profil alimentaire "sain"
5,0729	NOESY1D	0,18	0,04	0,9	Composante 2 Profil alimentaire "sain"
5,5099	CPMG	- 0,20	0,02	0,5	Composante 2 Profil alimentaire "sain"
5,8308	CPMG	- 0,20	0,02	0,5	Composante 2 Profil alimentaire "sain"
5,8668	CPMG	- 0,20	0,02	0,5	Composante 2 Profil alimentaire "sain"
6,1088	NOESY1D	- 0,17	0,0495	0,9	Composante 2 Profil alimentaire "sain"
6,1668	CPMG	- 0,18	0,04	0,7	Composante 2 Profil alimentaire "sain"
<b>6,2068</b>	CPMG	- 0,24	0,005	0,5	Composante 2 Profil alimentaire "sain"

## ANNEXE 11

<b>6,2478</b>	CPMG	- 0,22	0,01	0,5	Composante 2 Profil alimentaire "sain"
<b>6,3398</b>	CPMG	- 0,21	0,02	0,5	Composante 2 Profil alimentaire "sain"
6,5068	CPMG	- 0,18	0,04	0,7	Composante 2 Profil alimentaire "sain"
0,0581	CPMG	0,20	0,02	0,6	Composante 3 Régime à base de produits sucrés
1,1271	CPMG	- 0,18	0,04	0,6	Composante 3 Régime à base de produits sucrés
<b>1,6311</b>	CPMG	- 0,21	0,01	0,6	Composante 3 Régime à base de produits sucrés
1,677	NOESY1D	- 0,18	0,04	1,0	Composante 3 Régime à base de produits sucrés
2,411	CPMG	0,19	0,03	0,6	Composante 3 Régime à base de produits sucrés
2,705	CPMG	0,18	0,04	0,6	Composante 3 Régime à base de produits sucrés
3,028	CPMG	- 0,19	0,03	0,6	Composante 3 Régime à base de produits sucrés
6,1328	CPMG	0,20	0,02	0,6	Composante 3 Régime à base de produits sucrés
<b>6,1668</b>	CPMG	0,23	0,006	0,6	Composante 3 Régime à base de produits sucrés
6,2478	CPMG	0,19	0,03	0,6	Composante 3 Régime à base de produits sucrés
<b>7,2707</b>	NOESY1D	- 0,27	0,002	0,4	Composante 3 Régime à base de produits sucrés
7,4707	CPMG	- 0,19	0,03	0,6	Composante 3 Régime à base de produits sucrés
7,9347	CPMG	- 0,18	0,03	0,6	Composante 3 Régime à base de produits sucrés
8,3078	CPMG	- 0,19	0,03	0,6	Composante 3 Régime à base de produits sucrés
8,87	NOESY1D	0,21	0,02	1,0	Composante 3 Régime à base de produits sucrés
8,9111	NOESY1D	0,18	0,04	1,0	Composante 3 Régime à base de produits sucrés
<b>8,9511</b>	NOESY1D	0,25	0,004	0,4	Composante 3 Régime à base de produits sucrés
9,7494	CPMG	- 0,19	0,03	0,6	Composante 3 Régime à base de produits sucrés
9,8184	CPMG	- 0,19	0,03	0,6	Composante 3 Régime à base de produits sucrés

R; correlation coefficient

Les associations présentées dans le Tableau 28 sont surlignées en gras. La validation 2D des signaux en italique est en cours sur la plateforme RMN.

<sup>1</sup> 140 individus ont été analysés pour la séquence CPMG et 138 pour la séquence NOESY1D.

<sup>2</sup> Les corrélations de Spearman ont été ajustées sur les caractéristiques suivantes : âge, IMC, statut tabagique, saison de prélèvement sanguin, nombre d'enregistrements alimentaires sur 24 heures et moyenne de l'apport énergétique quotidien pendant les 2 années suivant le prélèvement sanguin.

Annexe 12 Texte intégral: “Diet-related metabolomic signature of long-term breast cancer risk using penalized regression: an exploratory study in the SU.VI.MAX cohort”

## **Diet-related metabolomic signature of long-term breast cancer risk using penalized regression: an exploratory study in the SU.VI.MAX cohort**

Lucie Lécuyer<sup>1</sup>, Céline Dalle<sup>2,3</sup>, Sophie Lefevre-Arbogast<sup>4</sup>, Pierre Micheau<sup>2</sup>, Bernard Lyan<sup>3</sup>, Adrien Rossary<sup>5</sup>, Aicha Demidem<sup>5</sup>, Mélanie Petera<sup>3</sup>, Marie Lagree<sup>6</sup>, Delphine Centeno<sup>3</sup>, Pilar Galan<sup>1</sup>, Serge Hercberg<sup>1,7</sup>, Cecilia Samieri<sup>4</sup>, Nada Assi<sup>8</sup>, Pietro Ferrari<sup>8</sup>, Vivian Viallon<sup>8</sup>, Mélanie Deschasaux<sup>1</sup>, Valentin Partula<sup>1</sup>, Bernard Srour<sup>1</sup>, Paule Latino-Martel<sup>1</sup>, Emmanuelle Kesse-Guyot<sup>1</sup>, Nathalie Druesne-Pecollo<sup>1</sup>, Marie-Paule Vasson<sup>5,9</sup>, Stéphanie Durand<sup>3</sup>, Estelle Pujos-Guillot<sup>3</sup>, Claudine Manach<sup>2</sup>, Mathilde Touvier<sup>1</sup>

### **Authors’ Affiliations:**

<sup>1</sup> Center of Research of Epidemiology and Statistics (CRESS), French National Institute of Health and Medical Research (Inserm) U1153, French National Institute for Agricultural Research (Inra) U1125, French National Conservatory of Arts and Crafts (Cnam), Paris 13 University, Nutritional Epidemiology Research Team (EREN), 93017 Bobigny, France.

<sup>2</sup> Clermont Auvergne University, INRA, UMR 1019, Human Nutrition Unit (UNH), CRNH Auvergne, Micronutriments et Santé cardiovasculaire (MicroCard), F-63000 Clermont-Ferrand, France.

<sup>3</sup> Clermont Auvergne University, INRA, UNH, Plateforme d’Exploration du Métabolisme, MetaboHUB Clermont, F-63000 Clermont-Ferrand, France.

<sup>4</sup> University of Bordeaux, Inserm, Bordeaux Population Health Research Center, UMR 1219, F-33000 Bordeaux, France.

<sup>5</sup> Clermont Auvergne University, INRA, UMR 1019, Human Nutrition Unit (UNH), CRNH Auvergne, Cellular Micro-Environment, Immunomodulation and Nutrition (ECREIN), F-63000 Clermont-Ferrand, France.

<sup>6</sup> Clermont Auvergne University, Institut de Chimie de Clermont-Ferrand, Plateforme d’Exploration du Métabolisme, MetaboHUB-Clermont, BP 80026, F-63171 Aubière, France.

<sup>7</sup> Public Health Department, Avicenne Hospital, Bobigny, France.

<sup>8</sup> International Agency for Research on Cancer, Section of Nutrition and Metabolism, Nutritional Methodology and Biostatistics Group, 69372 Lyon, France.

<sup>9</sup> Anticancer Center Jean-Perrin, CHU Clermont-Ferrand, France.

### **Corresponding author contact information:**

Lucie Lécuyer

EREN, Inserm U1153, SMBH Paris 13, 74 rue Marcel Cachin, F-93017 Bobigny Cedex, France

Telephone number: +33 1 48 38 89 54

Fax number: +33 1 48 38 89 31

e-mail: l.lecuyer@eren.smbh.univ-paris13.fr

**Running title:** Dietary metabolomic signatures of breast cancer risk

**Number of figures:** 2 / **Number of tables:** 2

**Conflicts of interest:** The authors have no conflict of interest to disclose.

**Funding**



This work was supported by the French National Cancer Institute [grant number INCa\_8085 for the project, and PhD grant number INCa\_11323 for L. Lecuyer]; the Federative Institute for Biomedical Research IFRB Paris 13; and the Cancéropôle Ile-de-France/Région Ile de France [PhD grant for M. Deschasaux]. The funders had no role in the design, analysis, or writing of this article.

**Abbreviations:**

SU.VI.MAX, Supplémentation en Vitamines et Minéraux Antioxydants; LC-MS, Liquid Chromatography Mass Spectrometry; ER, Estrogen Receptor; BMI, Body Mass Index; mPNNS-GS modified Programme National Nutrition Santé – Guideline Score; DQI-I, Diet Quality Index-International; PCA, Principal Component Analysis; FDR, False Discovery Rate BH, Benjamini Hochberg; OR, Odds Ratio; LASSO, Least Absolute Shrinkage and Selection Operator; ATBC, AcetylTributylCitrate; HCC, HepatoCellular Carcinoma; GC-MS, Gaz Chromatography Mass Spectrometry; DHEA, dehydroepiandrosterone; SXR, Steroid and Xenobiotic Receptor

**Authors' contributions**

The author's responsibilities were as follow – LL and MT: designed the research and conducted the research; SH, PG, MT, EKG: conducted the SU.VI.MAX cohort and provided essential data and samples; CD, ML, DC, SD, BL and EPG: performed metabolomic analyses; LL: performed statistical analysis; MT, SLA, NA, PF, VV: supervised statistical analysis; LL, MT: wrote the paper; LL, PM, SLA, CD, BL, AR, AD, MP, ML, DC, PG, SH, CS, NA, PF, VV, MD, VP, BS, PLM, EKG, NDP, MPV, SD, EPG, CM, MT: contributed to the data interpretation and revised each draft for important intellectual content; and MT had primary responsibility for final content. All authors read and approved the final manuscript.

**Acknowledgments**

The authors thank Younes Esseddik, Frédéric Coffinieres, Thi Hong Van Duong, Paul Flanzy, Régis Gatibelza, Jagatjit Mohinder and Maithyly Sivapalan (computer scientists), Rachida Mehroug and Frédérique Ferrat (logistic assistants), Nathalie Arnault, Véronique Gourlet, PhD, Fabien Szabo, PhD, Julien Allegre, and Laurent Bourhis (data-manager/statisticians) and Cédric Agaesse (dietitian) for their technical contribution to the SU.VI.MAX study. We also thank Nathalie Druesne-Pecollo, PhD (operational coordination) as well as all participants of the SU.VI.MAX study. This work was conducted in the framework of the French network for Nutrition And Cancer Research (NACRe network), [www.inra.fr/nacre](http://www.inra.fr/nacre) and received the NACRe Partnership Label. Metabolomic analysis were performed within the metaboHUB French infrastructure (ANR-INBS-0010).

**Abstract**

**Background:** Diet is a modifiable risk factor for breast cancer. Highlighting predictive diet-related biomarkers would be of great public health relevance for the identification of at-risk subjects. The aim of this exploratory study was to select diet-related metabolites which discriminate women at higher risk of breast cancer using untargeted metabolomics.

**Methods:** Plasma samples from a nested-case control study of the SU.VI.MAX cohort involving 200 incident breast cancer cases and matched controls were analysed by untargeted LC-MS. Metabolites related to diet were identified by partial correlation with dietary exposures, then best predictors of breast cancer risk were selected by Elastic Net penalized regression. The stability of selection was assessed by bootstrap resampling.

**Results:** 595 ions were selected as candidate diet-related metabolites. 14 of them were selected by Elastic Net regression as breast cancer risk discriminant ions. A lower level of piperine (a compound from pepper) and higher levels of AcetylTributylCitrate (an alternative plasticizer to phthalates), pregnene-triol sulfate (a steroid sulfate) and 2-amino-4-cyano butanoic acid (a metabolite linked to microbiota metabolism) were observed in plasma from women who subsequently developed breast cancer. This metabolomic signature was related to several dietary exposures such as a “Western” dietary pattern and higher alcohol and coffee intakes.

**Conclusion:** Our study suggested a diet-related plasma metabolic signature involving exogenous, steroid metabolites and microbiota-related compounds associated with long-term breast cancer risk that should be confirmed in large-scale independent studies. These results could help to identify healthy women at higher risk of breast cancer and improve the understanding of nutrition and health relationship.

## INTRODUCTION

Cancer is a multifactorial disease (1). Although a large part of cancers are explained by intrinsic factors, between 30 and 50 per cent of all cancer cases are estimated to be preventable (2). In France in 2015, over 30% of diagnosed incident breast cancers - the first female cancer in the world in terms of incidence (3) - are attributable to nutrition-related factors, including alcohol, physical activity and weight status (4,5). Beyond the three latter factors, a protective role of other dietary factors in breast carcinogenesis has been suggested for non-starchy vegetables (Negative Estrogen-Receptor (ER-) breast cancer) (3), dairy products (in premenopausal women) (3), foods containing carotenoids (3), foods high in calcium (3) or dietary fibre (6). In contrast, increased risk has been postulated for some types of lipids such as plasma levels of trans-fatty acids produced by industrial processing (ER- breast cancer) (7) and saturated fatty acids intake (8). Diet impacts both the endogenous metabolome as well as the food metabolome (i.e. the metabolites derived from ingested foods and their subsequent metabolism in the human body (9)). The identification of a set of metabolites that are both i) influenced by diet (and thus on which it is possible to act through dietary interventions) and ii) related to increased or decreased breast cancer risk would open new perspectives in terms of prevention and potentially provide insights into its underlying mechanisms.

High-throughput technologies allow for the exploration of thousands of molecules resulting from complex system-wide biological interactions. An in-depth untargeted investigation which is not hypothesis-driven holds promise for the discovery of new biomarkers. In comparison to approaches focused on a restricted number of a priori selected biomarkers, untargeted metabolomics allows highlighting combinations of metabolites (potentially acting with synergistic and antagonist effects) associated with disease risk which could allow a better sensitivity and specificity of predictive models (10). This application to the field of nutrition may highlight breast cancer-related metabolomic signatures, combining exogenous metabolites resulting directly from dietary exposure and endogenous or microbial metabolites, and therefore reflecting the impact of diet on host and its microbiota metabolism. So far, to our knowledge, only one previous study used nutritional semi-untargeted metabolomics to examine diet-related serum metabolite associations with long term breast cancer risk (11). This study highlighted 22 diet-related metabolites - mainly related to alcohol, vitamin E and animal fat intakes - associated with breast cancer risk. However, this study was conducted in postmenopausal women only, thus these results cannot be generalized in premenopausal women. Moreover, the unknown compounds were not taken into account in statistical analyses, limiting the potential discoveries of new biomarkers.

The aim of our study was to select a small subset of diet-related metabolites which best predicted long term breast cancer risk using up-to-date statistical method, particularly adapted to large scale omic data (Elastic Net regressions) by analysing untargeted mass spectrometry metabolomic data.

## METHODS

### Population study

The present study involved participants from the SU.VI.MAX prospective cohort (clinicaltrials.gov; NCT00272428), initially aimed at investigating the effect of a daily supplementation of nutritional doses of antioxidants on the incidence of cardiovascular diseases and cancers. This population-based, double-blinded, placebo-controlled, randomized trial was conducted over 8 years, and observational follow-up of health events was subsequently maintained during 5 years (13 years of follow-up). The study design and methods have been detailed previously (12,13). 13,017 participants were recruited in 1994-1995 and gave their written informed consent. The trial was approved by the Ethics Committee for Studies with Human Subjects of Paris-Cochin Hospital (CCPPRB 706/2364) and the 'Commission Nationale de l'Informatique et des Libertés' (CNIL 334641/907094), and was conducted according to the Declaration of Helsinki guidelines. The present work focused on a nested-case control study including SU.VI.MAX participants with a first incident invasive breast cancer diagnosed after baseline (N=215) matched with controls (1:1 ratio, using the density sampling method (14)) for the following characteristics at baseline : age, menopausal

status, BMI, intervention group of the trial, smoking status, season of blood draw. Details about the nested-case control study are presented in **Additional file 1** and the flow-chart is presented in **Figure 1**.

### **Baseline data collection and case ascertainment**

Baseline data collection and case ascertainment are described in **Additional file 1**. Briefly, at enrollment, participants filled in self-administered questionnaires about sociodemographic characteristics, smoking status, medication use, health status and family history of cancer, and underwent anthropometric measurements as well as a fasting blood draw. During the trial phase, participants were asked to complete computerized 24-h dietary records every 2 months including >990 food items (15). Dietary habits at baseline were estimated by averaging intakes from all dietary records collected during the first 2 years of participation in the SU.VI.MAX study. Self-reported health events were reviewed by an independent physician expert committee. Pathological reports were used to validate cases, and cancers were classified by using the International Chronic Diseases Classification, 10th Revision, Clinical Modification (16).

### **Metabolomic analyses**

Untargeted metabolomics was performed on plasma samples (N=430) following a slightly modified version of the procedure described by Pereira et al. (17). Profiling was conducted using high performance liquid chromatography/mass spectroscopy with the Metabolic profiler® platform (Bruker Daltonique, Wissembourg, France). The stability of the analytical system was monitored using pooled samples as quality control, analyzed in the same sequence than participant samples (injected one time at the beginning of each sequence and then after each set of 12 samples). Details on chemicals and reagents, biological samples preparation, metabolic profiling, raw data extraction, quality controls, and metabolite identification are available in **Additional file 2**. Data were processed under the Galaxy web-based platform Workflow4Metabolomics (18) using first XCMS module for peak detection, followed by quality checks and signal drift correction to yield a data matrix containing variables (retention times, masses) and peak intensities that have been corrected for batch effects. After a fast overview of all chromatograms, some individuals were excluded due to problems during sample preparation, too much ions with null intensity or pollution on chromatograms, leaving 200 cases and 200 controls samples for both positive and negative mode analyses. Highly correlated ions (correlation threshold at 0.9) from the same metabolite within the same retention time cluster were removed using the Metabolite Correlation Analysis (MCA) Galaxy module. After these processing steps, 528 and 690 ions (detected in positive and negative ionization modes respectively) from plasma metabolome remained in datasets for statistical analyses. This present metabolomic discovery approach allows semi quantitative measurements, representing relative ion intensities in arbitrary unit. However, even if it is not absolute quantitation, there are linear relations between ion intensity for each metabolite and its concentration level in the samples.

### **Statistical analysis**

Participants' baseline characteristics were compared between cases and controls using conditional logistic regressions.

In order to cover different aspects of diet, dietary exposure was assessed using 4 complementary methods. First, two dietary scores were computed: the mPNNS-GS (modified 'Programme National Nutrition Santé - Guideline Score') (19), reflecting adherence to 2001 French nutritional recommendations and including 12 components, eight of which referred to food-serving recommendations and four referred to moderation in consumption; and the DQI-I (Diet Quality Index-International) (20), comprising four components (variety, adequacy, moderation and overall balance) with adapted the cutoffs corresponding to French recommendations (21); higher scores representing a better quality of the diet. Then, we computed the average daily intake of 74 specific food groups (g/day). Finally, a Principal Component Analysis (PCA) with varimax rotated factors by orthogonal transformation was performed. This PCA highlighted two dietary

patterns representing a “Western diet” (mainly represented by higher intakes of alcohol, bread, processed and red meat, animal fat, cheese and potatoes) and a “Healthy diet” (with higher intakes of fruits, vegetables, whole grain, yoghurt, vegetable oil). Details of these dietary exposures computation are presented in **Additional file 3**.

#### *Identification of ions potentially associated with diet*

In a first step, we estimated correlations between the 1,218 ions (both positive and negative modes) and dietary exposures using partial Spearman correlations adjusted for age, BMI, menopausal status, smoking status, season at time of blood draw, number of 24h-dietary records and mean of daily energy intake during the 2 years following the blood draw. In this exploratory step, ions potentially associated with at least one dietary exposure ( $p$ -value  $<0.05$  and absolute correlation coefficient  $>0.15$ ) were selected for further analysis. In order to handle the False Discovery Rate (FDR) due to multiple testing, a Benjamini-Hochberg (BH) correction (22) was tested.

#### *Subset selection of diet-related breast cancer risk discriminant ions*

Among the ions potentially associated with diet highlighted in the first step, the best subset of predictors of breast cancer risk were selected by using penalized logistic regression. This method allows variable selection by forcing the coefficients of the less predictive variables to be exactly zero. As metabolomic data are highly correlated, we used the Elastic Net method (23) implemented in the R glmnet package (24). Our model considered all diet-related ions simultaneously (unit-variance scaling intensity), and the following confounding factors were included as unpenalized explanatory variables: age, BMI, season, menopausal status, smoking status, height, physical activity, education level, alcohol intake, use of hormone replacement therapy for menopause, number of children and family history of breast cancer at blood draw (baseline) and intervention group of the initial SU.VI.MAX trial. The Elastic Net penalization parameters  $\alpha$  and  $\lambda$  ( $\alpha=0$  implies no variable selection;  $\alpha=1$  is equivalent to LASSO regression;  $\lambda$  defines the strength of regularization) were optimized by using 5-fold cross-validation repeated 100 times, to account for the variance of cross-validation. Details on these steps are given in footnote of **Figure 2**.

Spearman correlation matrix of the selected metabolites was computed. Associations between selected ions and dietary exposure were checked by plotting (ion intensity) x (dietary exposure). In order to determine the stability of the selected ions, we applied a Bootstrap resampling Elastic Net method (25) by repeating the selection process 1,000 times on re-sampled data and recording the percentage of times where each ion was selected. Direction of the associations (OR) for the selected ions was estimated using logistic regression with all the selected ions in the same model and adjusted for the confounding factors cited above. Although  $p$ -values are generated by this model, they cannot be used for inference given the prior Elastic Net-based selection. Complementary analyses were performed using logistic regression models for each selected ion adjusted for the same confounding factors in order to check the associations between each individual ion and breast cancer risk.

A flow chart of the different statistical steps is shown in **Additional File 4**.

Due to the fully untargeted approach, only ions of interest were annotated according to the procedure described in **Additional file 2**. As proposed by Sumner et al (26), the metabolites were classified according to levels of confidence in the identification process: identified (level 1), putatively annotated (level 2), putatively characterized compound classes (level 3) and unknown compound (level 4).

Analyses were performed using SAS (v9.3, Cary, NC) and R (v3.5.2) softwares.

## **RESULTS**

Baseline characteristics of breast cancer cases and controls in the studied population are summarized in **Table 1**. Among the 1,218 detected ions, 595 were selected as candidate diet-related ions and were taken into account in the penalized logistic regression analyses (associations between these ions and dietary exposure are presented in **Additional File 5**, N=1085).

Fourteen ions resulted from the Elastic Net penalized regression. Among these, 2 were identified with a high level of confidence in annotation: piperine (M286T989) and AcetylTributylCitrate (ATBC) (M425T1158), two were putatively annotated: 2-amino-cyano-butanoic acid (M153T116) and pregnene-triol sulfate (M413T967) and the other were unknown compounds (M192T181, M265T186, M335T864, M364T125, M97T134, M166T144, M201T1091, M415T1344, M475T122, and M587T121). The Spearman correlation matrix of these ions is provided in **Additional file 6**. 2-amino-cyano-butanoic acid was highly correlated ( $p < 0.0001$ ) with 3 unknown compounds including 2 NaCOOH adducts.

In particular, piperine was positively associated with alcoholic drinks intake ( $r = 0.19$ ) and with a “Western” dietary pattern ( $r = 0.22$ ), ATBC was positively associated with coffee intake ( $r = 0.17$ ), 2-amino-cyano-butanoic acid was negatively associated with cake and biscuits intakes ( $r = -0.16$ ) and pregnene-triol sulfate was positively associated with alcoholic drinks intakes ( $r = 0.17$ ). The unknown compounds were associated with several dietary exposures such as pasta and cereals, salty products, processed meat, tomatoes, citrus fruit and pressed cooked cheese intakes (see **Additional File 5**). **Table 2** presents the associations between the 14 selected ions and breast cancer risk from adjusted logistic regression including the 14 ions altogether or one ion at a time. Lower levels of piperine and 6 unknown compounds (M335T864, M364T125, M166T144, M415T1344, M475T122 and M587T121) and higher levels of 2-amino-4-cyano butanoic acid, ATBC, pregnene-triol sulfate and 4 unknown compounds (M192T181, M265T186, M97T134 and M201T1091) were found in plasma from women who have subsequently developed breast cancer during follow-up.

Concerning the stability of the selected 14 ions on original data, their selection frequencies were  $\geq 50\%$  on the 1,000 bootstraps except for 2-amino-cyano-butanoic acid (27%). It was even  $> 70\%$  for piperine (M286T989), ATBC (M425T1158) and pregnene-triol sulfate (M413T967) and 8 unknown compounds (see **Figure 2** presenting the frequency of selection of ions  $\geq 40\%$  over 1000 bootstraps). Further investigation was carried out to annotate ions with selection frequencies  $> 70\%$  but not selected on the original data set (i.e. M114T115, M498T1040, M230T171, M423T120, M454T1064, M196T953, M485T1437, M491T120 and M581T123); however none of them were identified. The associations of these ions with dietary exposures are available in **Additional file 5** and with breast cancer risk (from adjusted logistic regression) in **Additional File 7**.

A summary of the results is shown in **Additional File 4**.

## DISCUSSION

This exploratory study used untargeted metabolomics coupled with multivariable penalized regression in order to screen for a limited set of ions potentially associated with various dietary exposures and maximized breast cancer risk discrimination.

Few studies have investigated the associations between diet-related metabolomics signatures and cancer risk. They observed associations between metabolites and, on the one hand, nutritional exposures (in particular coffee, alcohol, fibres, vitamin E and fried foods intake, BMI, physical activity), and on the other hand, risk of HCC (27–29), colorectal cancer (30,31) and only two dealing with breast cancer (11,32). In particular, most of these studies found metabolites related to alcohol intake, in particular, associated with breast cancer (11) or HCC (27–29) risks. In our study, among the 14 selected ions; 3 were positively associated to alcohol intake (M286T989, M335T864, M413T967), including piperine and pregnene-triol sulfate. Moreover, the ion M230T171, frequently selected during the bootstrap resampling but not selected on the original data set, was also associated with alcoholic drinks intake. Compared to the literature, some associations were newly identified in the present study (such as the positive association between ATBC, coffee intake and breast cancer risk), while several others were not replicated. These differences across studies are probably explained by differences in analytical techniques, study design, and statistical approaches, as well as population study and cancer locations.

In the present study, piperine, an exogenous active alkaloid with no endogenous origin reported so far, was highlighted as potential predictor of breast cancer risk (inverse association) in this study, with high stability across penalized models (selection frequencies >90% on the 1,000 bootstraps). Piperine is present in black and long pepper (33) and is used as feed additive in animal feed, in particular for poultry (34,35). The human exposure to piperine via this latter source has been estimated at 0.93µg/metabolic body weight per day (35). Several animal or cellular studies suggested a promising spectrum of properties for piperine such as anti-oxidant (36,37), anti-inflammatory (33,36,37), immunomodulatory, bioavailability and absorption promoter for many active molecules (33,38), anti-asthmatic, anti-convulsing, anti-mutagenic, anti-mycobacterial and anti-cancer activities (33) (chemopreventive properties including inhibition of angiogenesis and increased cell apoptosis) – specially in breast cancer models (39–42). In our study, piperine was in particular positively associated with alcohol intake and with a “Western” dietary pattern. These associations are probably explained by the fact that either several foods (e.g. processed meat, sauces, industrial cheese, poultry) containing piperine as feed additive or via pepper are either part of a Western diet or consumed in association with alcoholic drinks. The association with alcohol intake could also be explained by higher alkaloid solubility in ethanol (43). Consistent with our findings, Playdon et al. found a positive association between piperine and liquor intake and a decreased risk of breast cancer (11). Direct association was also found between piperine and wine intake in blood of female Twins (44). However, origine of circulating piperine is not restricted to Western-type foods and can also be the results of adding black pepper into food (e.g. for salad or fish seasoning). Unfortunately the level of detail of the SU.VI.MAX dietary questionnaire did not allow us to estimate pepper intake.

ATBC is an alternative plasticizer to phthalates (45) commonly used in polyvinyl resins and permitted as a food additive and food contact substance (46). Migration of ATBC from food packaging material into food has been observed for cheese, wrapped cake, microwaved soup, and microwaved peanut-containing cookies (46,47) and its leaching rate from medical equipment was found ten times faster than the potent endocrine disruptor di-2-ethylhexyl phthalate (46,48–50). In our study, we found a positive association between plasma ATBC and coffee intake. This association should be confirmed in independent human observational studies and in vivo or in vitro animal intervention studies; however it may reflect contaminant migration from plastic glass into coffee. Milk added in coffee may facilitate this migration since one study suggested that ATBC is prone to migrate into protein liquids, such as aqueous skim milk solution (51). Its increase in plasma from women who have developed breast cancer could also come from other exposure that we were unable to detect in this study. Recent studies found potential biological activity of ATBC on tissue growth (52) and a potential disruption of ovarian function in female mice due exposure to ATBC at low dosage imitating human exposure (50).

Currently, alcohol intake is the only established dietary risk factor for breast cancer risk with strong evidence (3). Some underlying mechanisms are already described, however, other remained to elucidate (3). In our study, pregnene-triol sulfate, a steroid sulfate hormone belonging to progestative family, was positively associated with both alcohol intake and breast cancer risk. Consistently with our results, a recent metabolomic study found positive correlations between alcohol intake and several serum steroids, notably pregnene-diol sulfate, which were associated with an increased breast cancer risk (11). Moreover, increased levels of sex steroids seems strongly associated with risk of postmenopausal breast cancers (54). Alcohol intake may increase circulating levels of steroid hormone, which could affect susceptibility to transform or promote cancer growth (55). An interventional study in postmenopausal women showed that alcohol consumption increased serum level of dehydroepiandrosterone (DHEA) sulfate (56), a precursor to androstenedione, testosterone, and, ultimately, estrone and estradiol. Furthermore, the pregnene-triol sulfate, found in our study, may come from the 17-hydroxy-pregnenolone, which losing its side chain can produce DHEA. Other factors could influence the level of sex steroids as BMI and lactation (3). Furthermore, it has been shown that ATBC strongly activated human and rat Steroid and Xenobiotic Receptor (SXR) and may alter metabolism of endogenous steroid hormones (45).

In our study, increased plasma level of 2-amino-4-cyano butanoic acid was found in plasma from women who have subsequently developed a breast cancer during follow-up. This metabolite, also called alpha-amino-gamma-cyanobutanoic acid, is a non-proteinogenic alpha-amino acid (2-aminobutanoic acid) substituted at position 4 by a cyano group and an aliphatic nitrile. It derives from a butyric acid (57). Butyrate is a short-chain fatty acid synthesized by the fermentation of non-digestible fibers by colon bacteria (58). Butyrate has recently been the focus of attention for its beneficial effects on intestinal homeostasis and energy metabolism. With its anti-inflammatory properties, butyrate improves intestinal barrier function and mucosal immunity (59). In our study, we found an inverse association between plasma 2-amino-4-cyano butanoic acid and cake and biscuits intakes. To our knowledge, at this time no direct association between plasma 2-amino-cyano butanoic acid and diet exposure has been established. The production and effects of butyrate appear to be related to diet, including the type of dietary fibres and fat consumed respectively (60). Butanoic acid is present in various types of foods including vanilla, oats, peanut and several fruits (61). Having no data available on the plasma butyrate level in our study, we cannot conclude on a possible link between the observed variations and a disturbance of butyrate metabolism and therefore of the microbiota. Our results need to be replicated in other independent observational studies, and intervention studies on animal models would provide a better understanding of the origin of its variations and links to food. Its association with the risk of breast cancer and a potential causal relationship could be studied in cellular mechanistic studies.

Unfortunately, due to the fully untargeted approach, the other metabolites selected by Elastic Net regression for breast cancer risk analyses were not identified. However, the results concerning these metabolites could be useful in the future, with the development of this type of approaches and with the increased annotation efforts in metabolomic databases.

The major strengths of this study pertained to the very sensitive untargeted MS metabolomic analysis, the prospective design and the choice of statistical design using Elastic Net penalized regression; a suitable method to high-dimensional correlated data. Penalized regression techniques were applied to control the variance of estimates (that increases in the presence of many predictors or of multicollinearity) by shrinking coefficients toward zero (62). Some penalized regression techniques such as the Least Absolute Shrinkage and Selection Operator (LASSO) (62) and Elastic Net (23) simultaneously perform automatic variable selection by shrinking the irrelevant predictor coefficients to exactly zero. The latter tends to better select important variables when high correlations are present as in metabolomic data. Several studies applied these methods in multiple fields (as youth violence, genetics, cardiovascular disease risk) (25,63–65), however to our knowledge, no previous study focusing on nutritional metabolomics and cancer prevention used penalized techniques. However, several limitations should be acknowledged for this study. First, despite the processes of cross validation, iterations and bootstraps, these results need to be validated in an independent study sample. Unbiased predictive performance could not be investigated in this study due to the lack of an independent data set. The study of performance gain when adding a metabolic signature related to diet to a model including established breast cancer risk factors would be useful to improve discrimination of women at higher risk of breast cancer. However to get a positive impact in terms of prevention this signature must be modifiable following a change of diet. This issue, as well as several others (replication, quantification and so on) should be investigated before considering an application in public health. Second, associations may have been missed due to lack of power, self-reporting bias or to the analytical protocol (LC-MS) that did not allow detecting all categories of metabolites. Complementary GC-MS analyses could be useful to provide additional types of metabolites. Moreover, a larger sample size would have allowed a stratification of the population according to several parameters such as time prior to cancer diagnosis and menopausal status. Third, the possibility of residual or unmeasured confounding cannot be ruled out in this observational study. However, many potential confounders were accounted for. Finally, our study was based on a single blood draw, which limits the investigation of metabolomic profiles stability across time. Nevertheless,



several studies have showed a good reproducibility of metabolomic measurements for most of metabolites (66,67). Several blood draws during follow-up would allow a finer detection of metabolic pathways that are disrupted during carcinogenesis.

In conclusion, this exploratory prospective study identified a plasma diet-related metabolic signature of long-term breast cancer risk involving exogenous, steroid and microbiota-related metabolites. The hypotheses highlighted in this study should be further investigated in future large-scale independent studies. In the future, such signatures could help to better understand the aetiology of nutrition and breast cancer and to identify key metabolites both associated to modifiable nutritional behaviour and to breast cancer risk.

VERSION SOUMISE

## REFERENCES

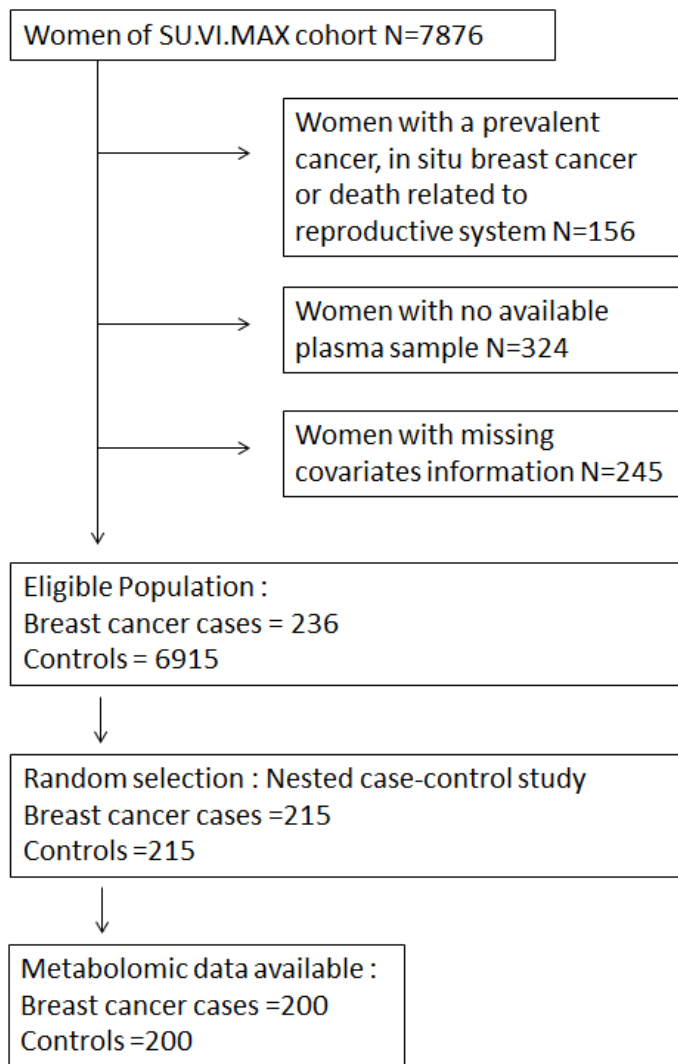
1. Wu S, Zhu W, Thompson P, Hannun YA. Evaluating intrinsic and non-intrinsic cancer risk factors. *Nat Commun.* 2018 28;9(1):3490.
2. WHO | Cancer prevention [Internet]. WHO. [cited 2019 Feb 18]. Available from: <http://www.who.int/cancer/prevention/en/>
3. World Cancer Research Fund/American Institute for Cancer Research. Continuous Update Project Expert Report 2018. Diet, nutrition, physical activity and breast cancer [Internet]. Available from: [dietandcancerreport.org](http://dietandcancerreport.org)
4. IARC. Les cancers attribuables au mode de vie et à l'environnement en France métropolitaine. [Internet]. Lyon: International Agency for Research on Cancer; 2018. Available from: [http://gco.iarc.fr/includes/PAF/PAF\\_FR\\_report.pdf](http://gco.iarc.fr/includes/PAF/PAF_FR_report.pdf)
5. Shield KD, Freisling H, Boutron-Ruault M-C, Touvier M, Marant Micallef C, Jenab M, et al. New cancer cases attributable to diet among adults aged 30-84 years in France in 2015. *Br J Nutr.* 2018 Nov;120(10):1171–80.
6. Reynolds A, Mann J, Cummings J, Winter N, Mete E, Te Morenga L. Carbohydrate quality and human health: a series of systematic reviews and meta-analyses. *Lancet Lond Engl.* 2019 02;393(10170):434–45.
7. Chajès V, Assi N, Biessy C, Ferrari P, Rinaldi S, Slimani N, et al. A prospective evaluation of plasma phospholipid fatty acids and breast cancer risk in the EPIC study. *Ann Oncol.* 2017 Nov 1;28(11):2836–42.
8. Sellem L, Srour B, Guéraud F, Pierre F, Kesse-Guyot E, Fiolet T, et al. Saturated, mono- and polyunsaturated fatty acid intake and cancer risk: results from the French prospective cohort NutriNet-Santé. *Eur J Nutr.* 2018 Apr 3;1–13.
9. Scalbert A, Brennan L, Manach C, Andres-Lacueva C, Dragsted LO, Draper J, et al. The food metabolome: a window over dietary exposure. *AmJClinNutr.* 2014 Jun;99(6):1286–308.
10. Davis CD, Milner JA. Biomarkers for diet and cancer prevention research: potentials and challenges. *Acta PharmacolSin.* 2007 Sep;28:1262–73.
11. Playdon MC, Ziegler RG, Sampson JN, Stolzenberg-Solomon R, Thompson HJ, Irwin ML, et al. Nutritional metabolomics and breast cancer risk in a prospective study. *AmJClinNutr.* 2017 Aug;106:637–49.
12. Hercberg S, Galan P, Preziosi P, Bertrais S, Mennen L, Malvy D, et al. The SU.VI.MAX study: a randomized, placebo-controlled trial of the health effects of antioxidant vitamins and minerals. *ArchInternMed.* 2004 Nov 22;164:2335–42.
13. Hercberg S, Preziosi P, Briancon S, Galan P, Triol I, Malvy D, et al. A primary prevention trial using nutritional doses of antioxidant vitamins and minerals in cardiovascular diseases and cancers in a general population: the SU.VI.MAX study--design, methods, and participant characteristics. *Supplementation en Vitamines et Minéraux AntioXydants. Control ClinTrials.* 1998 Aug;19:336–51.

14. Vandembroucke JP, Pearce N. Case-control studies: basic concepts. *Int J Epidemiol*. 2012 Oct 1;41(5):1480–9.
15. Le Moullec N, Deheeger M, Preziosi P, Montero P, Valeix P, Rolland-Cachera MF, et al. Validation du manuel photos utilisé pour l'enquête alimentaire de l'étude SU.VI.MAX (Validation of the food portion size booklet used in the SU.VI.MAX study). *CahNutrDiet*. 1996;31:158–64.
16. WHO. ICD-10, International Classification of Diseases and Related Health Problems. 10th revision. Geneva: World Health Organization; 1993.
17. Pereira F, Martin JF, Joly C, Sébédio JL, Pujos-Guillot E. Development and validation of a UPLC/MS method for a nutritional metabolomic study of human plasma. *Metabolomics*. 2010 Feb 19;6(2):207–18.
18. Giacomoni F, Le Corguille G, Monsoor M, Landi M, Pericard P, Petera M, et al. Workflow4Metabolomics: a collaborative research infrastructure for computational metabolomics. *Bioinformatics*. 2015 May 1;31(9):1493–5.
19. Estaquio C, Kesse-Guyot E, Deschamps V, Bertrais S, Dauchet L, Galan P, et al. Adherence to the French Programme National Nutrition Sante Guideline Score is associated with better nutrient intake and nutritional status. *JAmDiet Assoc*. 2009 Jun;109:1031–41.
20. Kim S, Haines PS, Siega-Riz AM, Popkin BM. The Diet Quality Index-International (DQI-I) provides an effective tool for cross-national comparison of diet quality as illustrated by China and the United States. *J Nutr*. 2003 Nov;133(11):3476–84.
21. Martin A. Apports nutritionnels conseillés pour la population française. Tec & Doc Lavoisier/AFSSA: Paris edn. 2001.
22. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *J R Stat Soc B*. 1995;57(1):289–300.
23. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Ser B Stat Methodol*. 2005 Apr;67(2):301–20.
24. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw*. 2010;33(1):1–22.
25. Bunea F, She Y, Ombao H, Gongvatana A, Devlin K, Cohen R. Penalized least squares regression methods and applications to neuroimaging. *NeuroImage*. 2011 Apr;55(4):1519–27.
26. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, et al. Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics*. 2007 Sep 19;3(3):211–21.
27. Assi N. A statistical framework to model the meeting-in-the-middle principle using metabolomic data: application to hepatocellular carcinoma in the EPIC study. *Mutagenesis*. 2015 Nov;30(6):743–53.
28. Assi N, Thomas DC, Leitzman M, Stepien M, Chajes V, Philip T, et al. Are metabolic signatures mediating the relationship between lifestyle factors and hepatocellular carcinoma risk? Results from a nested case-control study in EPIC. *Cancer Epidemiol Biomarkers Prev*. 2018 Mar 21;27(5):531–40.

29. Assi N, Gunter MJ, Thomas DC, Leitzmann M, Stepien M, Chajès V, et al. Metabolic signature of healthy lifestyle and its relation with risk of hepatocellular carcinoma in a large European cohort. *Am J Clin Nutr*. 2018 Jul 1;108(1):117–26.
30. Chadeau-Hyam M, Athersuch TJ, Keun HC, De Iorio M, Ebbels TMD, Jenab M, et al. Meeting-in-the-middle using metabolic profiling - a strategy for the identification of intermediate biomarkers in cohort studies. *Biomark Biochem Indic Expo Response Susceptibility Chem*. 2011 Feb;16(1):83–8.
31. Guertin KA, Loftfield E, Boca SM, Sampson JN, Moore SC, Xiao Q, et al. Serum biomarkers of habitual coffee consumption may provide insight into the mechanism underlying the association between coffee consumption and colorectal cancer. *Am J Clin Nutr*. 2015 May;101:1000–11.
32. Moore SC, Playdon MC, Sampson JN, Hoover RN, Trabert B, Matthews CE, et al. A Metabolomics Analysis of Body Mass Index and Postmenopausal Breast Cancer Risk. *JNCI J Natl Cancer Inst*. 2018 Jan 9;110(6):dix244.
33. Zakerali T, Shahbazi S. Rational Druggability Investigation Toward Selection of Lead Molecules: Impact of the Commonly Used Spices on Inflammatory Diseases. *ASSAY Drug Dev Technol*. 2018 Oct;16(7):397–407.
34. European Union. Register of Feed Additives, pursuant to Regulation (EC) No 1831/2003, Annex I : List of additives. 2019.
35. EFSA, Panel on Additives and Products or Substances used in Animal Feed. Safety and efficacy of pyridine and pyrrole derivatives belonging to chemical group 28 when used as flavourings for all animal species. *EFSA Journal*. 2016;14(2).
36. Diwan V, Poudyal H, Brown L. Piperine Attenuates Cardiovascular, Liver and Metabolic Changes in High Carbohydrate, High Fat-Fed Rats. *Cell Biochem Biophys*. 2013 Nov;67(2):297–304.
37. Rather RA, Bhagat M. Cancer Chemoprevention and Piperine: Molecular Mechanisms and Therapeutic Opportunities. *Front Cell Dev Biol* [Internet]. 2018 Feb 15 [cited 2019 Feb 28];6. Available from: <http://journal.frontiersin.org/article/10.3389/fcell.2018.00010/full>
38. Ajazuddin null, Alexander A, Qureshi A, Kumari L, Vaishnav P, Sharma M, et al. Role of herbal bioactives as a potential bioavailability enhancer for Active Pharmaceutical Ingredients. *Fitoterapia*. 2014 Sep;97:1–14.
39. Greenshields AL, Doucette CD, Sutton KM, Madera L, Annan H, Yaffe PB, et al. Piperine inhibits the growth and motility of triple-negative breast cancer cells. *Cancer Lett*. 2015 Feb;357(1):129–40.
40. Do MT, Kim HG, Choi JH, Khanal T, Park BH, Tran TP, et al. Antitumor efficacy of piperine in the treatment of human HER2-overexpressing breast cancer cells. *Food Chem*. 2013 Dec;141(3):2591–9.
41. Doucette CD, Hilchie AL, Liwski R, Hoskin DW. Piperine, a dietary phytochemical, inhibits angiogenesis. *J Nutr Biochem*. 2013 Jan;24(1):231–9.
42. Zheng J, Zhou Y, Li Y, Xu D-P, Li S, Li H-B. Spices for Prevention and Treatment of Cancers. *Nutrients*. 2016 Aug 12;8(8):495.
43. Alkaloids [Internet]. [cited 2019 May 24]. Available from: [https://www.chemicalbook.com/ProductCatalog\\_EN/2322.htm](https://www.chemicalbook.com/ProductCatalog_EN/2322.htm)

44. Pallister T, Jennings A, Mohney RP, Yarand D, Mangino M, Cassidy A, et al. Characterizing Blood Metabolomics Profiles Associated with Self-Reported Food Intakes in Female Twins. *PLoS One*. 2016;11(6):e0158568.
45. Takeshita A, Igarashi-Migitaka J, Nishiyama K, Takahashi H, Takeuchi Y, Koibuchi N. Acetyl tributyl citrate, the most widely used phthalate substitute plasticizer, induces cytochrome p450 3a through steroid and xenobiotic receptor. *Toxicol Sci Off J Soc Toxicol*. 2011 Oct;123(2):460–70.
46. United States Consumer Product Safety Commission. Review of Exposure and Toxicity Data for Phthalate Substitutes. 2010.
47. Sheftel VO. Indirect food additives and polymers: migration and toxicology [Internet]. 2000 [cited 2019 May 24]. Available from: [https://nls.ldls.org.uk/welcome.html?ark:/81055/vdc\\_100056086042.0x000001](https://nls.ldls.org.uk/welcome.html?ark:/81055/vdc_100056086042.0x000001)
48. Welle F, Wolz G, Franz R. Migration of plasticizers from PVC tubes into enteral feeding solutions. *Pharma International*. 2005;(3):17–21.
49. Testai E, Ms Scientific Committee SCENIHR. Electronic address: SANTE-C2-SCENIHR@ec.europa.eu, Hartemann P, Rastogi SC, Bernauer U, Piersma A, et al. The safety of medical devices containing DEHP plasticized PVC or other plasticizers on neonates and other groups possibly at risk (2015 update). *Regul Toxicol Pharmacol RTP*. 2016 Apr;76:209–10.
50. Rasmussen LM, Sen N, Liu X, Craig ZR. Effects of oral exposure to the phthalate substitute acetyl tributyl citrate on female reproduction in mice. *J Appl Toxicol JAT*. 2017 Jun;37(6):668–75.
51. Nara K, Nishiyama K, Natsugari H, Takeshita A, Takahashi H. Leaching of the Plasticizer, Acetyl Tributyl Citrate: (ATBC) from Plastic Kitchen Wrap. *J Health Sci*. 2009;55(2):281–4.
52. Rasmussen LM, Sen N, Vera JC, Liu X, Craig ZR. Effects of in vitro exposure to dibutyl phthalate, mono-butyl phthalate, and acetyl tributyl citrate on ovarian antral follicle growth and viability. *Biol Reprod*. 2017 May 1;96(5):1105–17.
53. Jung S, Wang M, Anderson K, Baglietto L, Bergkvist L, Bernstein L, et al. Alcohol consumption and breast cancer risk by estrogen receptor status: in a pooled analysis of 20 studies. *Int J Epidemiol*. 2016;45(3):916–28.
54. Key T, Appleby P, Barnes I, Reeves G, Endogenous Hormones and Breast Cancer Collaborative Group. Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. *J Natl Cancer Inst*. 2002 Apr 17;94(8):606–16.
55. Singletary KW, Gapstur SM. Alcohol and breast cancer: review of epidemiologic and experimental evidence and potential mechanisms. *JAMA*. 2001 Nov 7;286(17):2143–51.
56. Dorgan JF, Baer DJ, Albert PS, Judd JT, Brown ED, Corle DK, et al. Serum hormones and the alcohol-breast cancer association in postmenopausal women. *J Natl Cancer Inst*. 2001 May 2;93(9):710–5.
57. PubChem. 2-Amino-4-cyanobutanoic acid [Internet]. [cited 2019 Jun 26]. Available from: <https://pubchem.ncbi.nlm.nih.gov/compound/440770>
58. Bourassa MW, Alim I, Bultman SJ, Ratan RR. Butyrate, neuroepigenetics and the gut microbiome: Can a high fiber diet improve brain health? *Neurosci Lett*. 2016 20;625:56–63.

59. Liu H, Wang J, He T, Becker S, Zhang G, Li D, et al. Butyrate: A Double-Edged Sword for Health? *Adv Nutr*. 2018 Jan 1;9(1):21–9.
60. Lupton JR. Microbial Degradation Products Influence Colon Cancer Risk: the Butyrate Controversy. *J Nutr*. 2004 Feb 1;134(2):479–82.
61. Api AM, Belmonte F, Belsito D, Botelho D, Bruze M, Burton GA, et al. RIFM fragrance ingredient safety assessment, butyric acid, CAS Registry Number 107-92-6. *Food Chem Toxicol* [Internet]. 2019 Jan [cited 2019 Mar 21]; Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0278691519300420>
62. Tibshirani R. Regression Shrinkage and Selection via the Lasso. *JRSSB*. 1996;58:267–88.
63. Goldstick JE, Carter PM, Walton MA, Dahlberg LL, Sumner SA, Zimmerman MA, et al. Development of the SaFETy Score: A Clinical Screening Tool for Predicting Future Firearm Violence Risk. *Ann Intern Med*. 2017 May 16;166(10):707–14.
64. Frost HR, Amos CI. Gene set selection via LASSO penalized regression (SLPR). *Nucleic Acids Res*. 2017 Jul 7;45(12):e114.
65. Stegemann C, Pechlaner R, Willeit P, Langley SR, Mangino M, Mayr U, et al. Lipidomics profiling and risk of cardiovascular disease in the prospective population-based Bruneck study. *Circulation*. 2014 May 6;129(18):1821–31.
66. Carayol M, Licaj I, Achaintre D, Sacerdote C, Vineis P, Key TJ, et al. Reliability of Serum Metabolites over a Two-Year Period: A Targeted Metabolomic Approach in Fasting and Non-Fasting Samples from EPIC. *PLoSOne*. 2015;10:e0135437.
67. Floegel A, Drohan D, Wang-Sattler R, Prehn C, Illig T, Adamski J, et al. Reliability of serum metabolite concentrations over a 4-month period using a targeted metabolomic approach. *PLoSOne*. 2011;6:e21103.

**Figure 1:** Participant flowchart, selection of the SU.VI.MAX study

**Table 2** Baseline characteristics of breast cancer cases and controls<sup>1</sup>, SU.VI.MAX cohort, France (1994-2007)

	Breast cancer cases (N=200)	Controls (N=200)	P-value <sup>2</sup>
Age at baseline (y)	48.8±5.8	48.7±5.9	0.4
BMI (kg/m <sup>2</sup> )	23.1±3.9	23.5±4.2	0.08
			Not applicable
<18.5 kg/m <sup>2</sup> (underweight)	8 (4)	6 (3)	
≥18.5 to <25 kg/m <sup>2</sup> (normal weight)	143 (71.5)	145 (72.5)	
≥25 kg/m <sup>2</sup> (overweight)	49 (24.5)	49 (24.5)	
Height (cm)	162.9±6.1	160.9±5.9	0.001
Intervention group			Not applicable
Placebo	99 (49.5)	99 (49.5)	
Antioxidants	101 (50.5)	101 (50.5)	
Smoking status			Not applicable
Never and former	159 (79.5)	159 (79.5)	
Current smoker	41 (20.5)	41 (20.5)	
Physical activity			0.8
Irregular	64 (32)	58 (29)	
< 1h/d walking equivalent	63 (31.5)	67 (33.5)	
≥ 1h/d walking equivalent	73 (36.5)	75 (37.5)	
Educational level			0.1
Primary	35 (17.5)	43 (21.5)	
Secondary	75 (37.5)	86 (43)	
Superior	90 (45)	71 (35.5)	
Number of biological children	1.9±1.2	2±1.2	0.3
Hormonal treatment for menopause (yes)	69 (34.5)	70 (35)	0.9
Menopausal status at baseline			Not applicable
Premenopausal	127 (63.5)	127 (63.5)	
Postmenopausal	73 (36.5)	73 (36.5)	
Menopausal status at diagnosis			Not applicable
Premenopausal	77 (38.5)	77 (38.5)	
Postmenopausal	123 (61.5)	123 (61.5)	
Family history of breast cancer <sup>3</sup> (yes)	34 (17)	21 (10.5)	0.05
Alcohol intake (g/d)	10.8±11.2	11.6±13.3	0.5
Month of blood draw			Not applicable
March-April-May	74 (37)	74 (37)	
October-November	30 (15)	31 (15.5)	
December-January-February	96 (48)	95 (47.5)	

SU.VI.MAX, Supplémentation en Vitamines et Minéraux Antioxydants; BMI, body mass index; SD, standard deviation

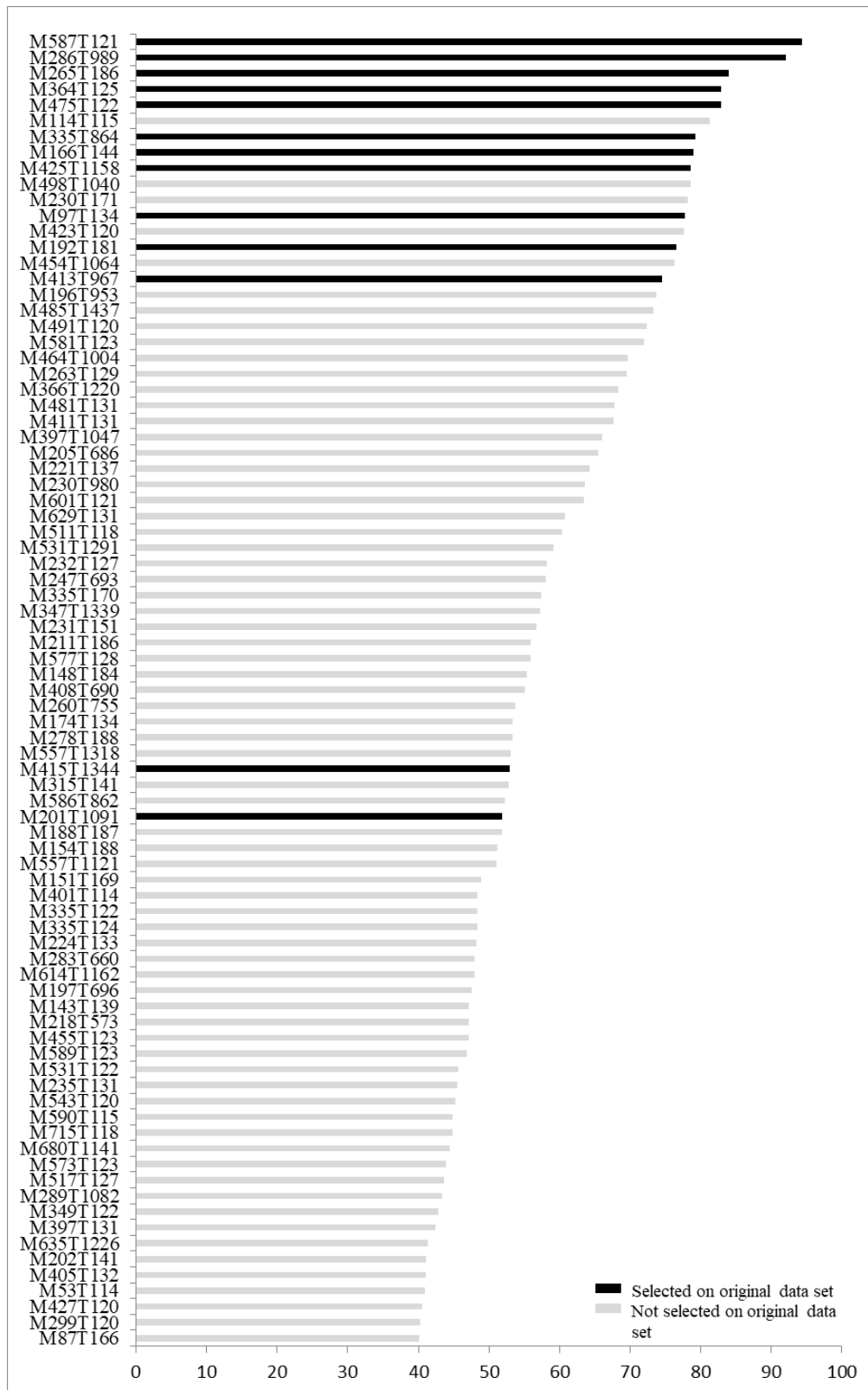
<sup>1</sup> Values are means ± SDs or n (%).

<sup>2</sup> P-value for the comparison between breast cancer cases and controls using conditional logistic regression. Not applicable for matching factors except for more precise variables (e.g. age and BMI).

<sup>3</sup> Among first degree female relatives.



**Figure 2** Percentage of the ions' selection frequency  $\geq 40\%$  over 1000 bootstraps via Elastic Net regression<sup>1</sup>



<sup>1</sup> Elastic Net regression model considered all diet-related ions simultaneously, and the following confounding factors were included as unpenalized explanatory variables: age (continuous), BMI (continuous), season (a priori-defined periods: October-November/December-January-February/March-

April-May), menopausal status (pre/post-menopause status), smoking status (current, former and non-smokers), height (continuous), physical activity (low, moderate, intense), education level (primary, secondary, superior), alcohol intake (continuous), use of hormone replacement therapy for menopause, number of children and family history of breast cancer at blood draw (baseline) and intervention group of the initial SU.VI.MAX trial (placebo/supplemented). The Elastic Net parameters  $\alpha$  and  $\lambda$  were optimized by using 5-fold cross-validation (assigning the case/control pairs in the same fold) repeated 100 times, to account for the variance of cross-validation. After  $\alpha$  had been optimized over a grid of 0.5 to 0.9 ( $\alpha=0$  implies no variable selection;  $\alpha=1$  is equivalent to LASSO regression) in order to get enough sparsity and still control for multicollinearity,  $\lambda$  (defining the strength of regularization) were also optimized by minimizing the mean deviance. Over the cross validations of the Elastic Net regression models, the most frequent optimal  $\alpha$  value was 0.5 and the corresponding average optimum  $\lambda$  value was 0.09. In order to determine the stability of ions selected on the original data set, we applied a bootstrap resampling Elastic Net method(264) by repeating the selection process 1000 times on re-sampled data and recording the percentage of times where each ion was selected (using the optimized parameter  $\alpha=0.5$ ).

VERSION SOUMISE

**Table 2** Direction of the 14 selected ions variations from adjusted logistic regression models<sup>1</sup> - SU.VI.MAX cohort, France (1994-2007)

Ions (annotation <sup>2</sup> )	Mass/retention time	Mode of detection	All selected ions together		Selected ions one by one	
			OR [95%CI]	P-value <sup>3</sup>	OR [95%CI]	P-value <sup>3</sup>
M153T116 (level 2: <b>2-amino-cyano-butanoic acid</b> )	153.0162/1.93	ESI Positive	1.23 [0.93-1.62]	0.1	1.08 [1.03-1.14]	0.002
M192T181 (level 4: molecular formula ; C9H6O4N)	192.0305/3.01	ESI Positive	1.37 [1.05-1.79]	0.02	1.08 [1.03-1.13]	0.003
M265T186 (level 4 : unknown)	265.3932/3.1	ESI Positive	1.28 [0.93-1.76]	0.1	1.09 [1.04-1.14]	0.0008
M286T989 (level 1: <b>piperine</b> )	286.143/16.48	ESI Positive	0.76 [0.58-0.99]	0.04	0.94 [0.89-0.99]	0.01
M335T864 (level 4 : unknown)	335.2404/14.4	ESI Positive	0.77 [0.61-0.99]	0.04	0.94 [0.89-0.98]	0.009
M364T125 (level 4 : NaCOOH adduct)	363.9292/2.09	ESI Positive	0.74 [0.58-0.94]	0.01	0.93 [0.89-0.98]	0.007
M425T1158 (level 1 : <b>Acetyl Tributyl Citrate ATBC</b> )	425.2172/19.31	ESI Positive	1.39 [1.08-1.79]	0.01	1.07 [1.02-1.12]	0.006
M97T134 (level 4 : NaCOOH adduct)	96.9218/2.24	ESI Positive	1.23 [0.94-1.60]	0.1	1.10 [1.04-1.15]	0.0003
M166T144 (level 4 : unknown)	166.0387/2.4	ESI Negative	0.80 [0.62-1.03]	0.09	0.94 [0.89-0.98]	0.01
M201T1091 (level 4 : unknown)	201.149/18.18	ESI Negative	1.21 [0.94-1.57]	0.1	1.10 [1.04-1.15]	0.0003
M413T967 (level 2 : <b>Pregnene-triol Sulfate</b> )	413.2002/16.11	ESI Negative	1.40 [1.08-1.84]	0.01	1.08 [1.03-1.14]	0.004
M415T1344 (level 4 : unknown)	415.2078/22.4	ESI Negative	0.83 [0.64-1.07]	0.2	0.92 [0.87-0.96]	0.0006
M475T122 (level 4 : unknown)	474.7265/2.04	ESI Negative	0.89 [0.70-1.14]	0.4	0.93 [0.88-0.97]	0.003
M587T121 (level 4 : unknown)	586.6522/2.02	ESI Negative	0.69 [0.54-0.89]	0.004	0.93 [0.89-0.98]	0.005

SU.VI.MAX, Supplémentation en Vitamines et Minéraux Antioxydants; OR, odds ratio; CI, confidence interval ;

<sup>1</sup> The principal logistic regression model considered all the 14 selected ions at the same time. Logistic regression models from complementary analyses considered separately the selected ions (one model for each of the 14 ions). All these models were adjusted for age (continuous), BMI (continuous), season (a priori-defined periods: October-November/December-January-February/March-April-May), menopausal status (pre/post-menopause status), smoking status (current, former and non-smokers), height (continuous), physical activity (low, moderate, intense), education level (primary, secondary, superior), alcohol intake (continuous), use of hormone replacement therapy for menopause, number of children and family history of breast cancer at blood draw (baseline) and intervention group of the initial SU,VI,MAX trial (placebo/supplemented). Tests for linear trend were performed using the continuous variables. ORs were presented for a 1 SD-increase of the continuous variable (semi-quantification).

<sup>2</sup> Levels of confidence for every identification were given accordingly to Sumner et al. (26): level 1; formally identified compound (confirmed with analysis of authentic standard), level 2; putatively identified compound (based upon spectral similarity with public/commercial spectral libraries or reference compound in the literature and/or physicochemical properties), level 3; putatively characterized compound classes and level 4; unknown compound.

<sup>3</sup> Although p-values are generated by this model, they cannot be used for inference given the prior Elastic Net-based selection.

**Additional File 1: Supplementary methods for case ascertainment, baseline data collection, the nested-case control study and dietary data computation in the SU.VI.MAX cohort.**

**Case ascertainment**

Participants self-reported their health events that occurred during follow-up (from 1994 to 2007). Participants, physicians, and hospitals gathered medical data and an independent physician expert committee reviewed them. Pathological reports were used to validate cases and to extract cancer characteristics. Cancers were classified by using the International Chronic Diseases Classification, 10th Revision, Clinical Modification (1).

**Baseline data collection**

At enrolment, participants filled-in self-administered questionnaires about sociodemographic characteristics, smoking status, medication use, health status and family history of cancer. During a baseline clinical examination, all participants underwent anthropometric measurements (height and weight) and a blood draw (occurring after a 12h-fasting period) by study nurses and physicians. These 35-ml venous blood samples were collected in sodium heparin Vacutainer tubes (Becton Dickinson, Rungis, France). After centrifugation at +4°C, plasma aliquots were immediately prepared and stored frozen at -20°C during less than two days and then stored in liquid nitrogen.

*Dietary data assessment*

During the trial phase, all participants were invited to complete a 24-h dietary record every 2 months, covering all days of the week and all seasons of the year, using computerised questionnaires. To indicate the consumed portion, participants were assisted by a validated instruction manual with photographs of >250 generic food items corresponding to 990 specific foods (2). A French published food composition table was used to calculate nutrient intakes (3). The usual food intakes at baseline were measured by the average intake calculated from the dietary records collected during the 2 years following the blood draw, thus accounting for intra-variability of intake.

**Nested case-control study**

The present study included 215 participants with a first incident invasive breast cancer matched 1:1 with controls for baseline age (35-39y/40-44y/45-49y/50-54y/55-59y/>60y), menopausal status (pre/post-menopause status at baseline and at diagnosis), BMI (<18.5kg/m<sup>2</sup>: underweight, ≥18.5 kg/m<sup>2</sup> to <25 kg/m<sup>2</sup>: normal weight and ≥25 kg/m<sup>2</sup>: overweight/obese), intervention group of the initial SU.VI.MAX trial (placebo/supplemented), smoking status (current smokers and non/former smokers at baseline) and season of blood draw (a priori-defined periods: October-November/December-January-February/March-April-May). Controls selection was based the classical density sampling method (4), i.e. every time a case was diagnosed, one control was selected from other members of the cohort who, at that time, were not diagnosed with breast cancer.

REFERENCES

1. WHO. ICD-10, International Classification of Diseases and Related Health Problems. 10th revision. Geneva: World Health Organization; 1993.
2. Le Moullec N, Deheeger M, Preziosi P, Montero P, Valeix P, Rolland-Cachera MF, et al. Validation of the portion size booklet used in the SU.VI.MAX study (article in French). *CahNutrDiet*. 1996;31:158-64.
3. Hercberg S. Table de Composition SU.VI.MAX des Aliments. Paris: Les éditions INSERM/Economica; 2005.
4. Vandembroucke JP, Pearce N. Case-control studies: basic concepts. *Int J Epidemiol*. 1 oct 2012;41(5):1480-9.

## **Additional file 2 Supplementary methods for MS metabolomic analysis**

### **Chemicals and reagents**

Methanol (Lichrosolv Hypergrade for LC-MS) was purchased from Merck. Acetonitrile (ULC-MS), water (ULC-MS) and formic acid (99%) were purchased from Biosolve. All standard compounds were purchased from Sigma-Aldrich (Saint Quentin Fallavier, France). All were analytical standard grade ( $\geq 98\%$ ).

### **Biological samples preparation**

Untargeted semi-quantitative metabolomics was performed on plasma samples following a slightly modified procedure described in Pereira et al. (1). In brief, plasma samples (200  $\mu\text{L}$ ) were deproteinized using cold methanol; after centrifugation and evaporation under nitrogen, the dry residues were redissolved in 50/50 (v/v) acetonitrile/water containing 0.1% formic acid. Pooled quality control samples were prepared by mixing 20  $\mu\text{L}$  from each of the plasma samples and prepared similarly.

### **UPLC-MS analyses for metabolic profiling**

Metabolic profiles were determined using the Metabolic profiler® platform (Bruker Daltonique, Wissembourg, France). Briefly, six  $\mu\text{L}$  of each sample were injected into a fast LC system (Agilent 1200 series) coupled to a micrOTOF mass spectrometer equipped with electrospray ionization source (Bruker). Chromatographic separation was performed using a 150 x 2.1mm Acquity HSS T3 C18 column (Waters, Saint Quentin en Yvelines, France), with a particle size of 1.8 $\mu\text{m}$ . The flow rate was set at 0.2 mL/min with mobile phase A (100% water) and mobile phase B (100% acetonitrile) both containing 0.1% formic acid. The elution gradient was carried out as follows: 0-1min, 100% A; 1-13min linear gradient to 100% B; 13-20min 100% B; 20-26 min 100% A for re equilibration. The column temperature was set at 30°C. Samples were kept at 4°C during analyses. Data were acquired in positive and negative ion modes with a scan range from 50 to 1000 mass-to-charge ratio (m/z).

Samples were randomized within the analytical sequence based on a Williams Latin Square strategy defined according to the main factors of the study. The stability of the analytical system was monitored using pooled samples as quality control, injected one time at the beginning of each sequence and then after each set of 12 samples.

### **Data extraction**

Raw data files were processed and converted to NetCDF format using the Bruker Data Analysis software. Data were processed under the Galaxy web-based platform Workflow4metabolomics (2), using first XCMS, followed by quality checks and signal drift correction according to the algorithm described by Van der Kloet et al (3), to yield a data matrix containing retention times, masses and peak intensities that have been corrected for batch effects. After a fast overview of all chromatograms, some individuals were excluded due to problems during sample preparation, too much ions with null intensity or pollution on chromatograms, leaving 200 cases and 200 controls samples for both positive and negative mode analyses. These steps include noise filtering, automatic peak detection and chromatographic alignment allowing the appropriate comparison of multiple samples by further processing methods. Analytical correlations were filtered using the 'Metabolites Correlation Analysis' tool (Galaxy Version 1.0.1) with a correlation coefficient threshold of 0.9. Low intensity features were then filtered using a threshold of 1,500 a.u. for the 90<sup>th</sup> higher percentile. After data processing, 528 and 690 ions (for positive and negative modes respectively) remained for statistical analysis.

### **Metabolite identification**

Identification experiment were performed on a ThermoScientific LTQ Orbitrap Velos hybrid mass spectrometer (Thermo Fisher Scientific, San José, CA, USA) using high resolution, full MS (m/z 70–1,000) at 100,000 resolving power, with acquisitions carried out for a limited selection of samples of interest.

Metabolites were first identified using an in-house database containing the reference spectra of more than 1,000 authentic standard compounds analyzed in the same analytical conditions. Then, the remaining unknown compounds were identified on the basis of their exact masses which were compared to those registered in Metlin (<https://metlin.scripps.edu>), in the Human Metabolome Database (HMDB; [www.hmdb.ca](http://www.hmdb.ca)) (4), or in Kyoto Encyclopedia of Genes and Genomes (KEGG) database (<http://www.genome.jp/kegg/>). Database queries were performed with a mass error of 0.005 Da and a retention time difference of 0.1 min (for the in-house database). Database results were confirmed using appropriate standards when available, isotopic patterns and mass fragmentation analyses using collision induced dissociation MS2 data acquired on the same samples. For unidentified ions, the number of plausible elemental compositions are restricted to a small number (or uniquely identified) with the support of additional chemical information-i.e. the molecular formula of the parent and knowledge of possible metabolic pathways. Metabolites were classified accordingly to Sumner et al (5) concerning the levels of confidence in the identification process: identified (confirmed by standard=level 1), putatively annotated (based upon physicochemical properties and/or spectral similarity with public/commercial spectral libraries=level 2), putatively characterized compound classes (level 3) and unknown compound (level 4).

#### REFERENCES

1. Pereira F, Martin JF, Joly C, Sébédio JL, Pujos-Guillot E. Development and validation of a UPLC/MS method for a nutritional metabolomic study of human plasma. *Metabolomics*. 2010 Feb 19;6:207-218.
2. Giacomoni F, Le Corguille G, Monsoor M, Landi M, Pericard P, Petera M, et al. Workflow4Metabolomics: a collaborative research infrastructure for computational metabolomics. *Bioinformatics*. 2015 May 1;31(9):1493-5.
3. van der Kloet FM, Bobeldijk I, Verheij ER, Jellema RH. Analytical error reduction using single point calibration for accurate and precise metabolomic phenotyping. *JProteomeRes*. 2009 Nov;8:5132-41.
4. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, et al. HMDB 3.0--The Human Metabolome Database in 2013. *Nucleic Acids Res*. 2013 Jan;41:D801-7.
5. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, et al. Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics*. 2007 Sep 19;3(3):211-21.

**Additional File 3: Supplementary methods for dietary data computation.**

On the base of the 990 specific foods we computed 2 dietary scores (mPNNS-GS and DQI-I), we estimated 74 food groups (list in **Supplementary Table S1**) and performed a Principal Component Analysis in order to highlight patterns of overall diet.

*mPNNS-GS (modified Programme National Nutrition Santé – Guideline Score)*

The PNNS-GS is computed in compliance with the 2001 recommendations of the French ‘Programme National Nutrition Santé’, reflecting the quality of the overall diet. The score computation, including food grouping, serving sizes, scoring, cut-off values and penalties, has previously been described in detail (1). Briefly, the score includes 13 components and has a range of 0–15 points: intakes of fruits and vegetables; bread, cereals, potatoes and legumes; whole grain food; milk and dairy products; meat and poultry, seafood and eggs; seafood; vegetable added fats; total added fats; sweetened foods; sweetened beverages; salt; alcohol; and adherence to physical activity recommendations. Points are deducted for elevated energy intakes. Several publications (2–4) have used a modified score, the mPNNS-GS, which contains only the 12 dietary items and has a maximum value of 13.5, in order to specifically investigate dietary quality (and not a combination of dietary quality and physical activity). For the present study, we used this modified score.

*DQI-I (Diet Quality Index-International)*

The computation of the DQI-I is based on the US dietary guidelines and has been extensively described elsewhere (5). Briefly, the score has a range of 0–100 points and includes a component of overall food group variety (0-15 points; meat/poultry/fish/eggs, dairy products/beans, grain, fruit, vegetable) and within-group variety for protein source (0-5 points; meat, poultry, fish, dairy products, beans, eggs), a component of adequacy (vegetables, fruits, cereals, fiber, protein, Fe, Ca and vitamin C; 0-5 points each), a component of moderation (total fat, saturated fat, cholesterol, Na and empty energy foods; 0-6 points each) and a component of overall balance (macronutrient ratio, 0-6 points; carbohydrate-protein-fat and fatty acid ratio, 0-4 points; PUFA-MUFA-SFA). The cut-offs used for adequacy and moderation were those corresponding to the French recommendations (6).

*PCA-derived patterns of overall diet*

Patterns of overall diet were obtained using a principal component analysis (PCA) with varimax rotated factors by orthogonal transformation on 29 food categories (labelled “overall” in **Table S1**). The two first factors (=“dietary patterns”) were retained according to the plot of the total variance explained (Cattel’s scree plots) (7) and explained >17% of total variability. Loading values are displayed in **Supplementary Table S2**. For interpreting the data, we considered food categories with a factor loading over 0.25. The first factor was characterized by higher intake of alcoholic drinks, animal fat, bread, cheese, pasta, rice and cereals, potatoes and tubers, processed meat, red meat and offal, salty products, sugar confectionery, sauces, vegetable fat and poultry. We considered this pattern as a “Western diet”. The second factor was characterized by higher intake of fruits, dried fruits, vegetables, breakfast cereals, vegetable fat, whole grain and yoghurts. We considered this pattern as a “healthy diet”.

68 participants were excluded for food groups estimation and PCA analysis due to missing dietary data and an additional 4 participants for DQI-I computation and 72 for mPNNS-GS computation were excluded due to other missing data as physical activity.

**Supplementary Table S1** List of the 74 food groups

---

Alcoholic drinks (overall)  
Allium  
Animal fat (overall)  
Aperitifs  
Apple and pear  
Banana  
Beers  
Berries  
Bread (overall)  
Breakfast cereals (overall)  
Cabbage  
Cakes and biscuits  
Cheese (overall)  
Chocolate  
Citrus fruits  
Coffee  
Cola  
Condiment and cooking aid  
Dairy desserts  
Dried fruits (overall except nuts)  
Eggs (overall)  
Exotic fruits  
Fatty fish  
Fish and seafood (overall)  
Fish eggs, crisps, salted biscuits and meat appetizers  
Fruit juice (overall)  
Fruits (overall)  
Fruit vegetable  
Grapes  
Green beans  
Infusion, chicory and artificially sweetened drinks  
Lean fish  
Leafy vegetable  
Legumes (overall)  
Milk (overall)  
Nuts (overall)  
Offal  
Olive  
Other alcohols (cider, champagne and so on)  
Other cheeses  
Other starchy food (chestnut, tapioca, plantain)  
Other type of food (overall)  
Other vegetables (avocado, asparagus, artichoke, mushroom, celery)  
Pasta and cereals  
Pasta, rice and cereals (overall)  
Pastries and sweetened cereal-based food (overall)  
Poultry (overall)  
Potatoes  
Potatoes and tubers (overall)  
Pressed cooked cheese  
Processed meat (overall)



Red meat  
Red meat and offal (overall)  
Red wine  
Rhubarb  
Rice  
Roots and tubers  
Salty products (overall)  
Sauces (overall)  
Shellfish and seafood  
Soya-based food  
Stone fruits  
Sugar confectionery (overall)  
Sweetened drinks (overall)  
Sweetened products (honey, candies, jam and so on)  
Tea  
Tomatoes  
Unsweetened drinks (overall)  
Vegetable fat (overall)  
Vegetable fat except olive and soya oils  
Vegetables (overall)  
White and rosé wines  
Whole grain (overall)  
Yoghurts (overall)

---

**Supplementary Table S2:** Loadings values of the *a posteriori* 2 dietary patterns derived by Principal Component Analyses<sup>1</sup> - SU.VI.MAX cohort, France (1994-2007)

Food groups	Factor 1 "Western diet"	Factor 2 "Healthy diet"
- Alcoholic drinks	<b>0.55</b>	-0.22
- Animal fat	<b>0.48</b>	0.15
- Bread	<b>0.62</b>	0.09
- Breakfast cereals	-0.21	<b>0.29</b>
- Cheese	<b>0.44</b>	0.23
- Dried fruits	-0.00	<b>0.28</b>
- Eggs	0.12	0.14
- Fish and seafood	0.12	0.24
- Fruits	-0.01	<b>0.65</b>
- Fruit juice	0.06	0.03
- Legumes	0.21	0.13
- Milk	0.05	0.09
- Nuts	0.11	0.22
- Other type of food	0.07	0.22
- Pasta, rice and cereals	<b>0.31</b>	0.09
- Pastries and sweetened cereal-based food	0.22	-0.00
- Poultry	<b>0.36</b>	0.08
- Potatoes and tubers	<b>0.44</b>	0.09
- Processed meat	<b>0.55</b>	-0.07
- Red meat and offal	<b>0.40</b>	-0.13
- Salty products	<b>0.34</b>	0.02
- Sauces	<b>0.33</b>	0.00
- Sugar confectionery	<b>0.35</b>	0.20
- Sweetened drinks	0.15	-0.07
- Unsweetened drinks	-0.09	<b>0.25</b>
- Vegetables	0.07	<b>0.67</b>
- Vegetable fat	<b>0.27</b>	<b>0.53</b>
- Whole grain	-0.13	<b>0.42</b>
- Yoghurts	-0.39	<b>0.43</b>

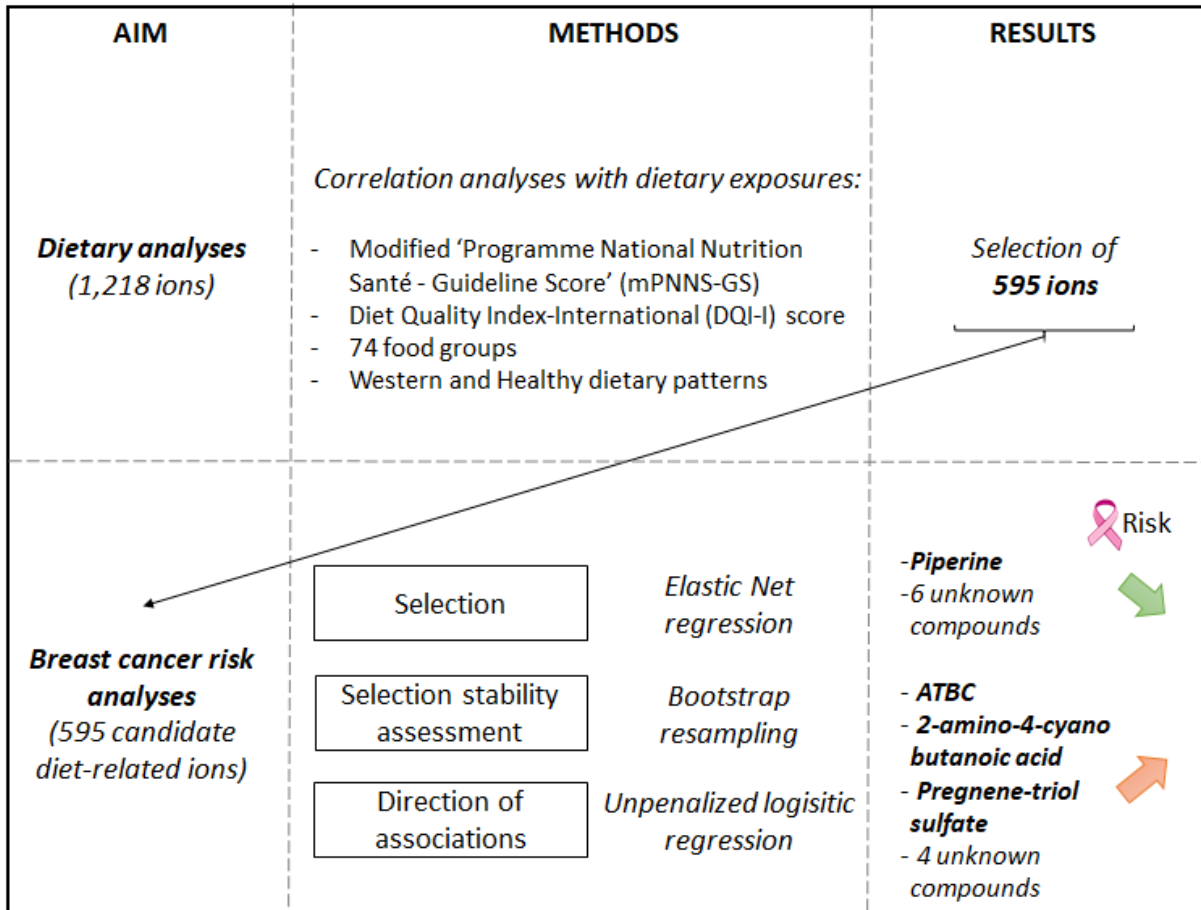
<sup>1</sup> Patterns of overall diet were obtained using a principal component analysis (PCA) with varimax rotated factors by orthogonal transformation on 29 food categories on the general SU.VI.MAX population having at least 3 validated 24h-dietary records during the 2 years following baseline (N=8112).

## REFERENCES

1. Estaquio C, Kesse-Guyot E, Deschamps V, Bertrais S, Dauchet L, Galan P, et al. Adherence to the French Programme National Nutrition Sante Guideline Score is associated with better nutrient intake and nutritional status. *JAmDiet Assoc.* 2009 Jun;109:1031–41.
2. Lelong H, Blacher J, Menai M, Galan P, Fezeu L, Hercberg S, et al. Association Between Blood Pressure and Adherence to French Dietary Guidelines. *Am J Hypertens.* 2016 Aug;29(8):948–58.
3. Lassale C, Galan P, Castetbon K, Péneau S, Méjean C, Hercberg S, et al. Differential association between adherence to nutritional recommendations and body weight status across educational levels: a cross-sectional study. *Prev Med.* 2013 Nov;57(5):488–93.
4. Assmann KE, Lassale C, Galan P, Hercberg S, Kesse-Guyot E. Dietary quality and 6-year anthropometric changes in a sample of French middle-aged overweight and obese adults. *PLoSOne.* 2014;9:e87083-.
5. Kim S, Haines PS, Siega-Riz AM, Popkin BM. The Diet Quality Index-International (DQI-I) provides an effective tool for cross-national comparison of diet quality as illustrated by China and the United States. *J Nutr.* 2003 Nov;133(11):3476–84.
6. Martin A. Apports nutritionnels conseillés pour la population française. Tec & Doc Lavoisier/AFSSA: Paris edn. 2001.
7. Cattell RB. The scree test for the number of factors. *Multivar BehavRes.* 1966 Mar;1:245–76.

Additional File 4

Supplementary Figure S1: Flow chart of statistical analyses and summary of results



VERSIC

## Additional File 5

**Supplementary Table S3** Associations ( $p < 0.05$  and  $r > |0.15|$ ) between ions and dietary exposure; DQI-I score, mPNNS-GS, Western dietary pattern, Healthy dietary pattern and food groups (daily intakes) from Spearman correlation analyses<sup>1</sup> - SU.VI.MAX cohort, France (1994-2007)

Ions (M: mass/T: retention time in second; classified by mass)	Mode of detection	Dietary exposure	R	P	P- corrected <sup>2</sup>
M103T302	ESI Negative	Alcoholic drinks (overall)	0.20	0.0002	0.0222
M117T691	ESI Negative	Alcoholic drinks (overall)	0.28	<0.0001	<0.0001
M118T691	ESI Negative	Alcoholic drinks (overall)	0.23	<0.0001	0.0045
M127T152	ESI Positive	Alcoholic drinks (overall)	0.15	0.0046	0.2207
M131T824	ESI Negative	Alcoholic drinks (overall)	0.18	0.0009	0.0591
M157T130	ESI Negative	Alcoholic drinks (overall)	0.16	0.0039	0.1287
M168T170	ESI Negative	Alcoholic drinks (overall)	0.16	0.0037	0.1272
M169T186	ESI Positive	Alcoholic drinks (overall)	0.22	<0.0001	0.0274
M179T673	ESI Negative	Alcoholic drinks (overall)	0.17	0.0014	0.0647
M181T115	ESI Negative	Alcoholic drinks (overall)	0.18	0.0007	0.0564
M185T690	ESI Negative	Alcoholic drinks (overall)	0.23	<0.0001	0.0045
M230T171	ESI Positive	Alcoholic drinks (overall)	0.17	0.0016	0.1062
M231T187	ESI Positive	Alcoholic drinks (overall)	0.20	0.0003	0.0359
M261T142	ESI Positive	Alcoholic drinks (overall)	0.18	0.0008	0.0580
<b>M286T989</b>	ESI Positive	Alcoholic drinks (overall)	0.19	0.0004	0.0359
M297T1290	ESI Negative	Alcoholic drinks (overall)	0.18	0.0010	0.0595
M325T150	ESI Positive	Alcoholic drinks (overall)	0.17	0.0018	0.1062
M367T1038	ESI Negative	Alcoholic drinks (overall)	0.17	0.0021	0.0869
M369T1053	ESI Negative	Alcoholic drinks (overall)	0.27	<0.0001	0.0001
M369T1080	ESI Negative	Alcoholic drinks (overall)	0.21	<0.0001	0.0097
M369T972	ESI Negative	Alcoholic drinks (overall)	0.19	0.0006	0.0552
M371T1080	ESI Negative	Alcoholic drinks (overall)	0.18	0.0006	0.0552
M403T186	ESI Positive	Alcoholic drinks (overall)	0.15	0.0039	0.2080
M405T122	ESI Negative	Alcoholic drinks (overall)	-0.18	0.0011	0.0595
<b>M413T967</b>	ESI Negative	Alcoholic drinks (overall)	0.17	0.0017	0.0742
M437T1080	ESI Negative	Alcoholic drinks (overall)	0.16	0.0032	0.1152
M437T1278	ESI Negative	Alcoholic drinks (overall)	0.15	0.0052	0.1560
M481T1080	ESI Negative	Alcoholic drinks (overall)	0.18	0.0011	0.0595
M482T1026	ESI Positive	Alcoholic drinks (overall)	-0.19	0.0004	0.0359
M483T1026	ESI Positive	Alcoholic drinks (overall)	-0.21	0.0001	0.0307
M485T1437	ESI Negative	Alcoholic drinks (overall)	-0.15	0.0052	0.1560
M501T143	ESI Positive	Alcoholic drinks (overall)	0.19	0.0004	0.0359
M509T132	ESI Negative	Alcoholic drinks (overall)	0.16	0.0027	0.1022
M526T1145	ESI Negative	Alcoholic drinks (overall)	-0.17	0.0012	0.0615
M81T131	ESI Negative	Allium	0.16	0.0036	0.2427
M97T126	ESI Negative	Allium	0.17	0.0022	0.2427
M183T1056	ESI Negative	Allium	-0.15	0.0043	0.2427

## ANNEXE 12

M185T690	ESI Negative	Allium	-0.17	0.0014	0.2201
<b>M201T1091</b>	ESI Negative	Allium	-0.16	0.0039	0.2427
M230T980	ESI Positive	Allium	-0.15	0.0042	0.3298
M260T755	ESI Positive	Allium	-0.17	0.0016	0.2153
M267T126	ESI Negative	Allium	0.15	0.0043	0.2427
M281T1189	ESI Positive	Allium	-0.18	0.0010	0.1851
M443T128	ESI Negative	Allium	0.15	0.0056	0.2533
M480T1089	ESI Positive	Allium	-0.18	0.0011	0.1851
M544T1096	ESI Positive	Allium	-0.15	0.0048	0.3298
M557T143	ESI Positive	Allium	-0.15	0.0043	0.3298
M559T143	ESI Positive	Allium	-0.18	0.0006	0.1851
M566T1201	ESI Negative	Allium	-0.19	0.0003	0.1151
M566T1234	ESI Negative	Allium	-0.15	0.0046	0.2427
M588T1148	ESI Negative	Allium	-0.17	0.0016	0.2201
M590T1181	ESI Negative	Allium	-0.16	0.0039	0.2427
M614T1162	ESI Negative	Allium	-0.19	0.0003	0.1151
M615T1162	ESI Negative	Allium	-0.16	0.0027	0.2427
M629T131	ESI Negative	Allium	-0.18	0.0009	0.2176
M169T186	ESI Positive	Animal fat (overall)	0.18	0.0010	0.1743
M224T133	ESI Positive	Animal fat (overall)	0.17	0.0015	0.2031
M231T151	ESI Positive	Animal fat (overall)	0.16	0.0036	0.2707
M283T660	ESI Positive	Animal fat (overall)	0.18	0.0006	0.1558
M359T185	ESI Positive	Animal fat (overall)	0.21	<0.0001	0.0322
M403T186	ESI Positive	Animal fat (overall)	0.16	0.0036	0.2707
M432T125	ESI Positive	Animal fat (overall)	0.16	0.0030	0.2707
M467T122	ESI Negative	Animal fat (overall)	-0.15	0.0044	0.6459
M651T113	ESI Negative	Animal fat (overall)	-0.16	0.0028	0.6459
M130T178	ESI Negative	Aperitifs	-0.16	0.0040	0.3144
M213T979	ESI Negative	Aperitifs	0.15	0.0049	0.3144
M221T626	ESI Positive	Aperitifs	-0.15	0.0049	0.6766
M231T187	ESI Positive	Aperitifs	0.15	0.0046	0.6766
M235T131	ESI Positive	Aperitifs	-0.17	0.0016	0.6766
M299T120	ESI Negative	Aperitifs	0.18	0.0010	0.3144
M339T122	ESI Negative	Aperitifs	-0.15	0.0044	0.3144
M369T1053	ESI Negative	Aperitifs	0.16	0.0039	0.3144
M383T926	ESI Negative	Aperitifs	0.16	0.0033	0.3144
M405T122	ESI Negative	Aperitifs	-0.19	0.0006	0.3144
M449T817	ESI Positive	Aperitifs	0.15	0.0051	0.6766
M485T1437	ESI Negative	Aperitifs	-0.17	0.0021	0.3144
M561T1410	ESI Negative	Aperitifs	0.16	0.0025	0.3144
M606T1130	ESI Negative	Aperitifs	0.17	0.0021	0.3144
M115T170	ESI Negative	Apple and pear	0.18	0.0007	0.1159
M140T137	ESI Positive	Apple and pear	0.16	0.0036	0.7283
M187T116	ESI Negative	Apple and pear	-0.16	0.0036	0.1784
M257T1082	ESI Negative	Apple and pear	0.24	<0.0001	0.0045
<b>M286T989</b>	ESI Positive	Apple and pear	-0.19	0.0005	0.2511

## ANNEXE 12

M365T170	ESI Negative	Apple and pear	0.19	0.0006	0.1159
M452T1182	ESI Negative	Apple and pear	0.15	0.0051	0.1949
M498T1110	ESI Negative	Apple and pear	0.16	0.0024	0.1488
M524T1138	ESI Negative	Apple and pear	0.17	0.0021	0.1488
M525T1139	ESI Negative	Apple and pear	0.16	0.0033	0.1749
M526T1145	ESI Negative	Apple and pear	0.17	0.0020	0.1488
M539T121	ESI Negative	Apple and pear	0.17	0.0012	0.1429
M556T1121	ESI Negative	Apple and pear	0.19	0.0006	0.1159
M557T1121	ESI Negative	Apple and pear	0.16	0.0023	0.1488
M586T1113	ESI Negative	Apple and pear	0.18	0.0008	0.1159
M592T1138	ESI Negative	Apple and pear	0.16	0.0029	0.1672
M612T1142	ESI Negative	Apple and pear	0.17	0.0023	0.1488
M625T131	ESI Negative	Apple and pear	-0.15	0.0043	0.1949
<b>M166T144</b>	ESI Negative	Banana	-0.17	0.0014	0.2604
M325T134	ESI Positive	Banana	0.19	0.0005	0.2790
M343T905	ESI Positive	Banana	0.15	0.0047	0.7395
M379T995	ESI Negative	Banana	-0.16	0.0030	0.2604
M380T995	ESI Negative	Banana	-0.16	0.0028	0.2604
M401T995	ESI Negative	Banana	-0.16	0.0028	0.2604
M417T121	ESI Negative	Banana	-0.16	0.0023	0.2604
M447T113	ESI Negative	Banana	0.17	0.0015	0.2604
M523T123	ESI Negative	Banana	0.18	0.0008	0.2604
M529T121	ESI Negative	Banana	0.16	0.0036	0.2726
M583T113	ESI Negative	Banana	0.17	0.0021	0.2604
M101T172	ESI Negative	Beers	0.18	0.0010	0.3540
M103T302	ESI Negative	Beers	0.18	0.0009	0.3540
M117T691	ESI Negative	Beers	0.15	0.0055	0.6121
M131T779	ESI Negative	Beers	0.15	0.0046	0.6121
M197T814	ESI Negative	Beers	0.16	0.0039	0.6121
<b>M286T989</b>	ESI Positive	Beers	0.21	0.0001	0.0600
M295T1234	ESI Negative	Beers	0.15	0.0056	0.6121
<b>M335T864</b>	ESI Positive	Beers	0.15	0.0051	0.4815
M79T132	ESI Negative	Berries	0.17	0.0022	0.4760
M213T131	ESI Negative	Berries	-0.16	0.0034	0.4760
M405T122	ESI Negative	Berries	-0.16	0.0034	0.4760
<b>M415T1344</b>	ESI Negative	Berries	-0.19	0.0005	0.3682
M745T131	ESI Negative	Berries	-0.17	0.0018	0.4760
M115T170	ESI Negative	Bread (overall)	-0.15	0.0043	0.5557
M119T167	ESI Negative	Bread (overall)	-0.16	0.0026	0.5557
M199T147	ESI Negative	Bread (overall)	0.19	0.0004	0.3027
M209T149	ESI Negative	Bread (overall)	0.17	0.0021	0.5557
M247T693	ESI Positive	Bread (overall)	-0.17	0.0019	0.5638
M367T157	ESI Positive	Bread (overall)	0.16	0.0021	0.5638
M389T1322	ESI Negative	Bread (overall)	0.15	0.0048	0.5557
M629T131	ESI Negative	Bread (overall)	0.16	0.0039	0.5557
M61T115	ESI Positive	Breakfast cereals (overall)	0.19	0.0004	0.1071

## ANNEXE 12

M75T114.2	ESI Positive	Breakfast cereals (overall)	0.20	0.0002	0.0940
M82T114	ESI Positive	Breakfast cereals (overall)	0.16	0.0036	0.3762
<b>M97T134</b>	ESI Positive	Breakfast cereals (overall)	0.16	0.0026	0.3455
M131T114	ESI Positive	Breakfast cereals (overall)	0.15	0.0052	0.4563
M243T111	ESI Negative	Breakfast cereals (overall)	0.15	0.0043	0.3487
M280T794	ESI Negative	Breakfast cereals (overall)	-0.16	0.0037	0.3487
M289T104	ESI Negative	Breakfast cereals (overall)	0.18	0.0012	0.1933
M289T122	ESI Positive	Breakfast cereals (overall)	0.17	0.0013	0.2299
M339T105	ESI Negative	Breakfast cereals (overall)	0.18	0.0006	0.1933
M447T113	ESI Negative	Breakfast cereals (overall)	0.19	0.0006	0.1933
M583T113	ESI Negative	Breakfast cereals (overall)	0.17	0.0016	0.1933
M629T131	ESI Negative	Breakfast cereals (overall)	-0.17	0.0017	0.1933
M651T113	ESI Negative	Breakfast cereals (overall)	0.17	0.0014	0.1933
M241T119	ESI Negative	Cabbage	0.19	0.0004	0.2983
M267T1082	ESI Positive	Cabbage	0.19	0.0005	0.1232
M289T1082	ESI Positive	Cabbage	0.21	<0.0001	0.0436
M290T1035	ESI Positive	Cabbage	0.17	0.0012	0.2031
M309T1141	ESI Negative	Cabbage	0.16	0.0031	0.6261
M321T852	ESI Positive	Cabbage	0.16	0.0025	0.3339
M395T134	ESI Positive	Cabbage	-0.15	0.0046	0.4813
M53T114	ESI Positive	Cakes and biscuits	-0.17	0.0014	0.1526
M75T114.2	ESI Positive	Cakes and biscuits	-0.16	0.0033	0.2176
M82T114	ESI Positive	Cakes and biscuits	-0.17	0.0014	0.1526
M112T115	ESI Positive	Cakes and biscuits	-0.16	0.0026	0.2176
M126T115	ESI Positive	Cakes and biscuits	-0.18	0.0011	0.1526
M131T114	ESI Positive	Cakes and biscuits	-0.15	0.0043	0.2187
<b>M153T116</b>	ESI Positive	Cakes and biscuits	-0.16	0.0030	0.2176
M153T570	ESI Positive	Cakes and biscuits	-0.15	0.0040	0.2187
M205T122	ESI Positive	Cakes and biscuits	-0.15	0.0050	0.2187
M263T142	ESI Positive	Cakes and biscuits	-0.20	0.0002	0.0416
M341T123	ESI Positive	Cakes and biscuits	-0.15	0.0047	0.2187
M567T125	ESI Positive	Cakes and biscuits	0.22	<0.0001	0.0151
M62T114	ESI Positive	Cheese (overall)	-0.17	0.0020	0.4286
M148T184	ESI Positive	Cheese (overall)	0.16	0.0034	0.4286
M243T170	ESI Negative	Cheese (overall)	0.15	0.0042	0.9927
M336T863	ESI Positive	Cheese (overall)	-0.17	0.0012	0.4286
M337T864	ESI Positive	Cheese (overall)	-0.15	0.0039	0.4286
M341T124	ESI Negative	Cheese (overall)	0.17	0.0016	0.9927
M423T129	ESI Positive	Cheese (overall)	0.15	0.0041	0.4286
M197T696	ESI Positive	Chocolate	0.16	0.0034	0.4130
M251T971	ESI Positive	Chocolate	0.15	0.0042	0.4130
M269T971	ESI Positive	Chocolate	0.16	0.0028	0.4130
M291T971	ESI Positive	Chocolate	0.17	0.0020	0.4130
M358T940	ESI Positive	Chocolate	0.15	0.0042	0.4130
M377T152	ESI Negative	Chocolate	0.17	0.0020	0.1751
M391T1276	ESI Negative	Chocolate	-0.17	0.0013	0.1627



## ANNEXE 12

M437T1278	ESI Negative	Chocolate	-0.16	0.0041	0.2214
M445T1357	ESI Negative	Chocolate	-0.16	0.0029	0.2033
M450T1125	ESI Negative	Chocolate	-0.19	0.0004	0.1586
M451T123	ESI Negative	Chocolate	0.18	0.0007	0.1586
M453T1181	ESI Negative	Chocolate	-0.18	0.0011	0.1627
M467T1360	ESI Negative	Chocolate	-0.19	0.0005	0.1586
M476T1146	ESI Negative	Chocolate	-0.17	0.0017	0.1642
M489T1214	ESI Negative	Chocolate	-0.15	0.0055	0.2214
M502T1172	ESI Negative	Chocolate	-0.17	0.0014	0.1627
M544T1145	ESI Negative	Chocolate	-0.16	0.0042	0.2214
M569T124	ESI Negative	Chocolate	0.15	0.0049	0.2214
M591T1181	ESI Negative	Chocolate	-0.16	0.0024	0.1802
M144T163	ESI Positive	Citrus fruits	0.33	<0.0001	<0.0001
M145T156	ESI Positive	Citrus fruits	0.26	<0.0001	0.0002
M146T158	ESI Positive	Citrus fruits	-0.16	0.0038	0.4027
M158T186	ESI Positive	Citrus fruits	-0.18	0.0007	0.1267
M175T170	ESI Negative	Citrus fruits	0.15	0.0051	0.7105
M283T124	ESI Negative	Citrus fruits	0.19	0.0003	0.2092
M365T170	ESI Negative	Citrus fruits	0.16	0.0038	0.6587
M379T142	ESI Positive	Citrus fruits	-0.16	0.0038	0.4027
M448T1071	ESI Negative	Citrus fruits	-0.16	0.0028	0.6377
<b>M475T122</b>	ESI Negative	Citrus fruits	0.19	0.0006	0.2092
M53T114	ESI Positive	Coffee	-0.16	0.0025	0.0469
M85T151	ESI Positive	Coffee	-0.15	0.0042	0.0720
M94T154	ESI Positive	Coffee	0.42	<0.0001	<0.0001
M116T114	ESI Positive	Coffee	-0.18	0.0007	0.0233
M119T184	ESI Positive	Coffee	-0.16	0.0032	0.0576
M124T128	ESI Negative	Coffee	0.18	0.0008	0.0255
M138T153	ESI Positive	Coffee	0.43	<0.0001	<0.0001
M147T119	ESI Positive	Coffee	-0.19	0.0005	0.0179
M149T149	ESI Negative	Coffee	0.27	<0.0001	<0.0001
M157T130	ESI Negative	Coffee	0.15	0.0047	0.1077
M171T879	ESI Negative	Coffee	-0.15	0.0043	0.1017
M174T171	ESI Negative	Coffee	0.33	<0.0001	<0.0001
M179T147	ESI Negative	Coffee	0.19	0.0005	0.0168
M179T673	ESI Negative	Coffee	0.28	<0.0001	<0.0001
M180T1073	ESI Negative	Coffee	0.32	<0.0001	<0.0001
M181T651	ESI Positive	Coffee	0.34	<0.0001	<0.0001
M182T547	ESI Positive	Coffee	-0.16	0.0022	0.0425
M189T717	ESI Negative	Coffee	0.31	<0.0001	<0.0001
M191T146	ESI Negative	Coffee	0.54	<0.0001	<0.0001
M195T699	ESI Positive	Coffee	0.27	<0.0001	<0.0001
M197T171	ESI Negative	Coffee	0.28	<0.0001	<0.0001
M197T696	ESI Positive	Coffee	0.28	<0.0001	<0.0001
M199T914	ESI Negative	Coffee	0.23	<0.0001	0.0008
M203T769	ESI Negative	Coffee	0.28	<0.0001	<0.0001

## ANNEXE 12

M205T686	ESI Negative	Coffee	0.19	0.0004	0.0143
M259T131	ESI Negative	Coffee	-0.18	0.0007	0.0234
M288T1010	ESI Positive	Coffee	0.15	0.0051	0.0809
M289T1035	ESI Positive	Coffee	0.18	0.0010	0.0291
M289T104	ESI Negative	Coffee	-0.18	0.0007	0.0222
M290T1035	ESI Positive	Coffee	0.16	0.0037	0.0656
M317T873	ESI Positive	Coffee	0.20	0.0002	0.0088
M333T995	ESI Negative	Coffee	0.49	<0.0001	<0.0001
M371T145	ESI Negative	Coffee	0.49	<0.0001	<0.0001
M371T688	ESI Positive	Coffee	0.17	0.0020	0.0406
M379T995	ESI Negative	Coffee	0.49	<0.0001	<0.0001
M380T995	ESI Negative	Coffee	0.43	<0.0001	<0.0001
M400T1012	ESI Positive	Coffee	0.15	0.0047	0.0778
M401T129	ESI Negative	Coffee	-0.15	0.0053	0.1133
M401T995	ESI Negative	Coffee	0.45	<0.0001	<0.0001
M419T106	ESI Negative	Coffee	-0.17	0.0012	0.0346
<b>M425T1158</b>	ESI Positive	Coffee	0.17	0.0012	0.0309
M428T1048	ESI Positive	Coffee	0.22	<0.0001	0.0038
M429T809	ESI Negative	Coffee	0.58	<0.0001	<0.0001
M429T995	ESI Negative	Coffee	0.41	<0.0001	<0.0001
M437T1265	ESI Negative	Coffee	-0.15	0.0052	0.1133
M447T113	ESI Negative	Coffee	-0.16	0.0032	0.0778
M463T1291	ESI Negative	Coffee	-0.20	0.0002	0.0079
M478T1030	ESI Positive	Coffee	0.17	0.0015	0.0342
M479T1030	ESI Positive	Coffee	0.17	0.0016	0.0355
M495T903	ESI Negative	Coffee	0.35	<0.0001	<0.0001
M496T1071	ESI Positive	Coffee	0.20	0.0002	0.0080
M502T1030	ESI Positive	Coffee	0.17	0.0012	0.0309
M517T127	ESI Negative	Coffee	-0.17	0.0023	0.0586
M518T1072	ESI Positive	Coffee	0.22	<0.0001	0.0031
M520T1034	ESI Positive	Coffee	0.19	0.0004	0.0155
M521T1034	ESI Positive	Coffee	0.20	0.0002	0.0080
M522T1095	ESI Positive	Coffee	0.18	0.0010	0.0290
M523T1095	ESI Positive	Coffee	0.17	0.0015	0.0342
M542T1034	ESI Positive	Coffee	0.21	0.0001	0.0067
M543T1034	ESI Positive	Coffee	0.20	0.0001	0.0079
M544T1096	ESI Positive	Coffee	0.19	0.0004	0.0155
M566T1034	ESI Positive	Coffee	0.17	0.0017	0.0365
M583T113	ESI Negative	Coffee	-0.17	0.0013	0.0346
M651T113	ESI Negative	Coffee	-0.17	0.0022	0.0574
M992T1071	ESI Positive	Coffee	0.18	0.0007	0.0233
M81T131	ESI Negative	Cola	0.16	0.0042	0.9730
M174T134	ESI Negative	Cola	-0.18	0.0007	0.5059
M405T132	ESI Positive	Cola	0.15	0.0045	0.8090
M451T123	ESI Negative	Cola	0.15	0.0052	0.9730
M104T139	ESI Positive	Condiment and cooking aid	0.15	0.0041	0.4328

## ANNEXE 12

M146T131	ESI Negative	Condiment and cooking aid	0.16	0.0027	0.3855
M155T1041	ESI Negative	Condiment and cooking aid	-0.17	0.0020	0.3855
M157T180	ESI Negative	Condiment and cooking aid	-0.15	0.0048	0.4032
M169T139	ESI Positive	Condiment and cooking aid	-0.16	0.0028	0.4328
M187T1088	ESI Negative	Condiment and cooking aid	-0.16	0.0039	0.3855
M199T1228	ESI Negative	Condiment and cooking aid	-0.17	0.0017	0.3855
M199T914	ESI Negative	Condiment and cooking aid	-0.16	0.0034	0.3855
M219T185	ESI Positive	Condiment and cooking aid	-0.16	0.0034	0.4328
M321T128	ESI Negative	Condiment and cooking aid	-0.16	0.0037	0.3855
M349T141	ESI Positive	Condiment and cooking aid	-0.16	0.0030	0.4328
M433T141	ESI Positive	Condiment and cooking aid	0.17	0.0016	0.4328
M443T128	ESI Negative	Condiment and cooking aid	0.15	0.0053	0.4032
M579T125	ESI Negative	Condiment and cooking aid	0.17	0.0012	0.3855
M72T185	ESI Positive	Dairy desserts	-0.17	0.0012	0.3155
M129T772	ESI Negative	Dairy desserts	-0.17	0.0016	0.5261
M129T814	ESI Negative	Dairy desserts	-0.16	0.0024	0.5261
M153T570	ESI Positive	Dairy desserts	-0.15	0.0046	0.3445
M169T132	ESI Negative	Dairy desserts	-0.17	0.0017	0.5261
M196T136	ESI Positive	Dairy desserts	-0.20	0.0002	0.0937
M229T186	ESI Positive	Dairy desserts	-0.16	0.0024	0.3445
M309T1341	ESI Negative	Dairy desserts	0.16	0.0030	0.5261
M336T863	ESI Positive	Dairy desserts	0.16	0.0029	0.3445
M345T1118	ESI Positive	Dairy desserts	-0.15	0.0043	0.3445
M439T1337	ESI Negative	Dairy desserts	0.15	0.0052	0.6291
M114T146	ESI Negative	DQI-I	-0.16	0.0038	0.3304
M117T154	ESI Positive	DQI-I	-0.17	0.0014	0.1484
M144T163	ESI Positive	DQI-I	0.20	0.0003	0.0673
M191T171	ESI Negative	DQI-I	-0.17	0.0020	0.3304
M197T696	ESI Positive	DQI-I	-0.19	0.0004	0.0708
M257T1082	ESI Negative	DQI-I	0.26	<0.0001	0.0009
M259T131	ESI Negative	DQI-I	0.16	0.0031	0.3304
M365T119	ESI Negative	DQI-I	0.19	0.0006	0.1908
M365T170	ESI Negative	DQI-I	0.16	0.0044	0.3304
M375T1336	ESI Negative	DQI-I	0.15	0.0052	0.3304
M401T995	ESI Negative	DQI-I	0.16	0.0035	0.3304
M455T123	ESI Negative	DQI-I	-0.16	0.0041	0.3304
M527T132	ESI Positive	DQI-I	-0.20	0.0002	0.0673
M605T115	ESI Negative	DQI-I	0.16	0.0039	0.3304
M646T816	ESI Positive	DQI-I	-0.18	0.0010	0.1322
M860T141	ESI Positive	DQI-I	-0.16	0.0026	0.2260
M115T170	ESI Negative	Dried fruits (overall except nuts)	0.15	0.0054	0.3737
M123T626	ESI Positive	Dried fruits (overall except nuts)	-0.16	0.0021	0.2783
M130T772	ESI Negative	Dried fruits (overall except nuts)	0.19	0.0004	0.0970
M197T772	ESI Negative	Dried fruits (overall except nuts)	0.17	0.0023	0.2712
M211T186	ESI Positive	Dried fruits (overall except nuts)	0.19	0.0004	0.1934
M217T615	ESI Positive	Dried fruits (overall except nuts)	0.15	0.0041	0.3067

## ANNEXE 12

M253T1155	ESI Negative	Dried fruits (overall except nuts)	0.19	0.0003	0.0970
M254T1156	ESI Negative	Dried fruits (overall except nuts)	0.19	0.0004	0.0970
M290T1035	ESI Positive	Dried fruits (overall except nuts)	0.16	0.0031	0.2783
M292T171	ESI Negative	Dried fruits (overall except nuts)	0.16	0.0034	0.2712
M292T173	ESI Negative	Dried fruits (overall except nuts)	0.18	0.0011	0.1922
M401T1367	ESI Negative	Dried fruits (overall except nuts)	0.16	0.0027	0.2712
M512T999	ESI Positive	Dried fruits (overall except nuts)	0.16	0.0030	0.2783
M520T1034	ESI Positive	Dried fruits (overall except nuts)	0.16	0.0032	0.2783
M521T1034	ESI Positive	Dried fruits (overall except nuts)	0.15	0.0048	0.3160
M541T121	ESI Negative	Dried fruits (overall except nuts)	-0.16	0.0028	0.2712
M568T1028	ESI Positive	Dried fruits (overall except nuts)	0.16	0.0031	0.2783
M612T1142	ESI Negative	Dried fruits (overall except nuts)	0.16	0.0035	0.2712
M197T696	ESI Positive	Eggs (overall)	0.16	0.0037	0.4987
M261T642	ESI Positive	Eggs (overall)	0.15	0.0047	0.4987
<b>M286T989</b>	ESI Positive	Eggs (overall)	0.17	0.0016	0.4987
M287T659	ESI Positive	Eggs (overall)	0.17	0.0021	0.4987
M339T105	ESI Negative	Eggs (overall)	-0.17	0.0019	0.6393
M350T838	ESI Negative	Eggs (overall)	0.15	0.0055	0.6393
M433T119	ESI Negative	Eggs (overall)	0.16	0.0028	0.6393
M475T1114	ESI Negative	Eggs (overall)	0.16	0.0035	0.6393
M481T131	ESI Positive	Eggs (overall)	0.15	0.0047	0.4987
M209T609	ESI Positive	Exotic fruits	-0.16	0.0036	0.9567
M239T776	ESI Positive	Exotic fruits	0.19	0.0003	0.1469
M79T132	ESI Negative	Factor 1 "Western diet"	-0.28	<0.0001	0.0002
M130T145	ESI Negative	Factor 1 "Western diet"	0.16	0.0028	0.3266
M181T903	ESI Positive	Factor 1 "Western diet"	0.15	0.0039	0.2977
M196T953	ESI Positive	Factor 1 "Western diet"	0.16	0.0027	0.2873
M225T995	ESI Negative	Factor 1 "Western diet"	0.15	0.0046	0.4532
M257T1082	ESI Negative	Factor 1 "Western diet"	-0.20	0.0002	0.0474
M259T131	ESI Negative	Factor 1 "Western diet"	-0.24	<0.0001	0.0024
M261T130	ESI Negative	Factor 1 "Western diet"	-0.17	0.0014	0.2478
M264T904	ESI Positive	Factor 1 "Western diet"	0.16	0.0034	0.2977
<b>M286T989</b>	ESI Positive	Factor 1 "Western diet"	0.22	<0.0001	0.0166
M359T185	ESI Positive	Factor 1 "Western diet"	0.17	0.0021	0.2873
M395T134	ESI Positive	Factor 1 "Western diet"	0.16	0.0022	0.2873
M449T817	ESI Positive	Factor 1 "Western diet"	0.17	0.0021	0.2873
M481T1080	ESI Negative	Factor 1 "Western diet"	0.17	0.0021	0.2851
M79T132	ESI Negative	Factor 2 "Healthy diet"	0.16	0.0037	0.1511
M87T166	ESI Negative	Factor 2 "Healthy diet"	-0.15	0.0051	0.1808
M113T725	ESI Positive	Factor 2 "Healthy diet"	-0.16	0.0021	0.1859
M115T170	ESI Negative	Factor 2 "Healthy diet"	0.17	0.0013	0.1124
M115T590	ESI Negative	Factor 2 "Healthy diet"	-0.17	0.0017	0.1305
M129T772	ESI Negative	Factor 2 "Healthy diet"	-0.18	0.0008	0.1091
M130T185	ESI Positive	Factor 2 "Healthy diet"	0.20	0.0002	0.1173
M131T779	ESI Negative	Factor 2 "Healthy diet"	-0.16	0.0027	0.1511
M157T130	ESI Negative	Factor 2 "Healthy diet"	-0.19	0.0005	0.1091

## ANNEXE 12

M179T673	ESI Negative	Factor 2 "Healthy diet"	-0.15	0.0052	0.1808
M196T953	ESI Positive	Factor 2 "Healthy diet"	-0.17	0.0014	0.1851
M197T696	ESI Positive	Factor 2 "Healthy diet"	-0.16	0.0025	0.1859
M197T814	ESI Negative	Factor 2 "Healthy diet"	-0.18	0.0009	0.1091
M199T147	ESI Negative	Factor 2 "Healthy diet"	-0.18	0.0008	0.1091
M209T149	ESI Negative	Factor 2 "Healthy diet"	-0.18	0.0008	0.1091
M225T975	ESI Positive	Factor 2 "Healthy diet"	-0.16	0.0028	0.1859
M231T848	ESI Negative	Factor 2 "Healthy diet"	0.16	0.0038	0.1511
M255T861	ESI Negative	Factor 2 "Healthy diet"	0.16	0.0028	0.1511
M257T1082	ESI Negative	Factor 2 "Healthy diet"	0.25	<0.0001	0.0022
M261T142	ESI Positive	Factor 2 "Healthy diet"	-0.17	0.0011	0.1851
<b>M286T989</b>	ESI Positive	Factor 2 "Healthy diet"	-0.18	0.0008	0.1851
M333T132	ESI Negative	Factor 2 "Healthy diet"	-0.16	0.0033	0.1511
M333T171	ESI Negative	Factor 2 "Healthy diet"	0.16	0.0037	0.1511
M367T157	ESI Positive	Factor 2 "Healthy diet"	-0.15	0.0048	0.2717
M370T945	ESI Positive	Factor 2 "Healthy diet"	-0.15	0.0051	0.2717
M395T131	ESI Negative	Factor 2 "Healthy diet"	-0.16	0.0039	0.1511
M401T129	ESI Negative	Factor 2 "Healthy diet"	0.16	0.0032	0.1511
M625T131	ESI Negative	Factor 2 "Healthy diet"	-0.17	0.0013	0.1124
M639T123.2	ESI Negative	Factor 2 "Healthy diet"	-0.16	0.0037	0.1511
M646T816	ESI Positive	Factor 2 "Healthy diet"	-0.16	0.0022	0.1859
M93T126	ESI Negative	Fatty fish	0.19	0.0004	0.0296
M94T128	ESI Positive	Fatty fish	-0.15	0.0046	0.2794
M100T625	ESI Positive	Fatty fish	-0.19	0.0004	0.1952
M117T154	ESI Positive	Fatty fish	-0.16	0.0025	0.2202
M117T647	ESI Negative	Fatty fish	0.18	0.0012	0.0296
M120T154	ESI Positive	Fatty fish	-0.17	0.0016	0.2202
M130T447	ESI Negative	Fatty fish	0.17	0.0023	0.0356
M159T127	ESI Positive	Fatty fish	-0.16	0.0035	0.2671
M164T559	ESI Negative	Fatty fish	0.17	0.0019	0.0339
M164T605	ESI Negative	Fatty fish	0.17	0.0016	0.0312
M180T1073	ESI Negative	Fatty fish	-0.15	0.0050	0.0481
M187T116	ESI Negative	Fatty fish	0.18	0.0012	0.0296
M195T1164	ESI Negative	Fatty fish	-0.16	0.0038	0.0419
M203T689	ESI Negative	Fatty fish	0.16	0.0024	0.0356
M219T122	ESI Negative	Fatty fish	0.16	0.0036	0.0407
M239T1215	ESI Negative	Fatty fish	-0.16	0.0028	0.0378
M241T119	ESI Negative	Fatty fish	-0.15	0.0051	0.0486
M249T692	ESI Negative	Fatty fish	0.18	0.0012	0.0296
M255T116	ESI Negative	Fatty fish	0.15	0.0053	0.0486
M281T1423	ESI Negative	Fatty fish	-0.16	0.0035	0.0407
M286T794	ESI Positive	Fatty fish	0.18	0.0007	0.1952
M295T1234	ESI Negative	Fatty fish	-0.15	0.0045	0.0440
M299T1409	ESI Negative	Fatty fish	-0.19	0.0003	0.0296
M307T119	ESI Negative	Fatty fish	-0.16	0.0031	0.0404
M313T1367	ESI Negative	Fatty fish	-0.17	0.0014	0.0299

## ANNEXE 12

M323T115	ESI Negative	Fatty fish	0.18	0.0010	0.0296
M331T1384	ESI Negative	Fatty fish	-0.19	0.0006	0.0296
M349T690	ESI Negative	Fatty fish	0.18	0.0007	0.0296
M353T1437	ESI Negative	Fatty fish	-0.16	0.0024	0.0356
M358T940	ESI Positive	Fatty fish	0.16	0.0024	0.2202
M359T120	ESI Negative	Fatty fish	-0.16	0.0034	0.0407
M365T1220	ESI Negative	Fatty fish	-0.20	0.0002	0.0296
M366T1220	ESI Negative	Fatty fish	-0.20	0.0002	0.0296
M385T115	ESI Negative	Fatty fish	-0.19	0.0006	0.0296
M391T115	ESI Negative	Fatty fish	0.17	0.0019	0.0339
M391T1276	ESI Negative	Fatty fish	-0.19	0.0006	0.0296
M393T123	ESI Negative	Fatty fish	0.18	0.0012	0.0296
M397T131	ESI Negative	Fatty fish	0.18	0.0010	0.0296
M399T1384	ESI Negative	Fatty fish	-0.16	0.0039	0.0423
M407T690	ESI Negative	Fatty fish	0.18	0.0011	0.0296
M408T690	ESI Negative	Fatty fish	0.17	0.0013	0.0299
M415T945	ESI Negative	Fatty fish	0.17	0.0013	0.0299
M421T1335	ESI Negative	Fatty fish	-0.17	0.0015	0.0299
M423T120	ESI Negative	Fatty fish	-0.18	0.0007	0.0296
M429T690	ESI Negative	Fatty fish	0.16	0.0036	0.0407
M445T1338	ESI Negative	Fatty fish	-0.18	0.0009	0.0296
M445T1357	ESI Negative	Fatty fish	-0.19	0.0005	0.0296
M459T115	ESI Negative	Fatty fish	0.15	0.0043	0.0438
M463T1337	ESI Negative	Fatty fish	-0.16	0.0041	0.0429
M465T1324	ESI Negative	Fatty fish	-0.16	0.0040	0.0429
M467T122	ESI Negative	Fatty fish	0.17	0.0014	0.0299
M467T1360	ESI Negative	Fatty fish	-0.18	0.0010	0.0296
M467T1402	ESI Negative	Fatty fish	-0.16	0.0027	0.0377
M468T1402	ESI Negative	Fatty fish	-0.16	0.0041	0.0429
M473T1411	ESI Negative	Fatty fish	-0.16	0.0032	0.0407
M476T1146	ESI Negative	Fatty fish	-0.18	0.0011	0.0296
M489T1214	ESI Negative	Fatty fish	-0.18	0.0011	0.0296
M489T1340	ESI Negative	Fatty fish	-0.17	0.0022	0.0356
M489T1402	ESI Negative	Fatty fish	-0.16	0.0036	0.0407
M502T1172	ESI Negative	Fatty fish	-0.16	0.0033	0.0407
M511T1411	ESI Negative	Fatty fish	-0.15	0.0047	0.0457
M513T123	ESI Negative	Fatty fish	0.15	0.0043	0.0438
M517T1268	ESI Negative	Fatty fish	-0.17	0.0020	0.0339
M522T115	ESI Negative	Fatty fish	-0.16	0.0029	0.0378
M525T122	ESI Negative	Fatty fish	0.17	0.0022	0.0356
M526T1025	ESI Positive	Fatty fish	0.16	0.0024	0.2202
M537T1402	ESI Negative	Fatty fish	-0.17	0.0019	0.0339
M551T121	ESI Negative	Fatty fish	-0.17	0.0015	0.0299
M559T1402	ESI Negative	Fatty fish	-0.16	0.0024	0.0356
M560T1402	ESI Negative	Fatty fish	-0.15	0.0045	0.0440
M571T123	ESI Negative	Fatty fish	0.20	0.0002	0.0296

## ANNEXE 12

M573T123	ESI Negative	Fatty fish	0.17	0.0022	0.0356
M573T1415	ESI Negative	Fatty fish	-0.19	0.0006	0.0296
M579T125	ESI Negative	Fatty fish	0.18	0.0007	0.0296
M581T123	ESI Negative	Fatty fish	0.15	0.0053	0.0486
M590T115	ESI Negative	Fatty fish	-0.18	0.0008	0.0296
M590T1181	ESI Negative	Fatty fish	-0.16	0.0028	0.0378
M591T1181	ESI Negative	Fatty fish	-0.16	0.0025	0.0357
M591T1431	ESI Negative	Fatty fish	-0.19	0.0004	0.0296
M611T122	ESI Negative	Fatty fish	-0.20	0.0003	0.0296
M635T1226	ESI Negative	Fatty fish	-0.17	0.0016	0.0312
M658T1181	ESI Negative	Fatty fish	-0.18	0.0010	0.0296
M715T118	ESI Negative	Fatty fish	-0.16	0.0036	0.0407
M100T625	ESI Positive	Fish and seafood (overall)	-0.15	0.0051	0.4522
M164T559	ESI Negative	Fish and seafood (overall)	0.18	0.0010	0.1811
M180T1073	ESI Negative	Fish and seafood (overall)	-0.17	0.0020	0.1811
M182T115	ESI Negative	Fish and seafood (overall)	-0.16	0.0024	0.1811
M187T158	ESI Positive	Fish and seafood (overall)	-0.16	0.0029	0.4522
M197T696	ESI Positive	Fish and seafood (overall)	-0.15	0.0050	0.4522
M299T1409	ESI Negative	Fish and seafood (overall)	-0.16	0.0035	0.1861
M331T1384	ESI Negative	Fish and seafood (overall)	-0.17	0.0023	0.1811
M353T1437	ESI Negative	Fish and seafood (overall)	-0.16	0.0029	0.1849
M367T157	ESI Positive	Fish and seafood (overall)	-0.16	0.0036	0.4522
M399T1384	ESI Negative	Fish and seafood (overall)	-0.16	0.0033	0.1861
M408T690	ESI Negative	Fish and seafood (overall)	0.18	0.0009	0.1811
M415T945	ESI Negative	Fish and seafood (overall)	0.17	0.0020	0.1811
M429T690	ESI Negative	Fish and seafood (overall)	0.17	0.0013	0.1811
M500T1144	ESI Negative	Fish and seafood (overall)	-0.16	0.0028	0.1849
M502T1172	ESI Negative	Fish and seafood (overall)	-0.18	0.0011	0.1811
M568T1028	ESI Positive	Fish and seafood (overall)	0.16	0.0026	0.4522
M569T1028	ESI Positive	Fish and seafood (overall)	0.17	0.0011	0.4522
M586T1113	ESI Negative	Fish and seafood (overall)	0.17	0.0015	0.1811
M97T126	ESI Negative	Fish eggs, crisps, salted biscuits and meat appetizers	0.16	0.0023	0.4364
M101T172	ESI Negative	Fish eggs, crisps, salted biscuits and meat appetizers	0.16	0.0041	0.4364
M188T992	ESI Negative	Fish eggs, crisps, salted biscuits and meat appetizers	-0.15	0.0043	0.4364
M209T217	ESI Negative	Fish eggs, crisps, salted biscuits and meat appetizers	0.17	0.0023	0.4364
M235T131	ESI Positive	Fish eggs, crisps, salted biscuits and meat appetizers	-0.16	0.0035	0.7942
M517T127	ESI Negative	Fish eggs, crisps, salted biscuits and meat appetizers	-0.15	0.0046	0.4364
M577T128	ESI Negative	Fish eggs, crisps, salted biscuits and meat appetizers	-0.19	0.0005	0.3440
M726T116	ESI Negative	Fish eggs, crisps, salted biscuits and meat appetizers	-0.16	0.0029	0.4364
M144T163	ESI Positive	Fruit juice (overall)	0.24	<0.0001	0.0032
M145T156	ESI Positive	Fruit juice (overall)	0.19	0.0004	0.1093
M437T1067	ESI Negative	Fruit juice (overall)	-0.15	0.0043	0.6680
M647T117	ESI Negative	Fruit juice (overall)	-0.19	0.0005	0.3617
M715T118	ESI Negative	Fruit juice (overall)	-0.16	0.0042	0.6680
M601T121	ESI Negative	Fruit vegetable	-0.19	0.0005	0.3473
M115T170	ESI Negative	Fruits (overall)	0.19	0.0004	0.0687

## ANNEXE 12

M130T185	ESI Positive	Fruits (overall)	0.21	<0.0001	0.0517
M144T163	ESI Positive	Fruits (overall)	0.18	0.0006	0.1597
M145T156	ESI Positive	Fruits (overall)	0.16	0.0038	0.5694
M175T170	ESI Negative	Fruits (overall)	0.17	0.0021	0.2076
M255T861	ESI Negative	Fruits (overall)	0.18	0.0008	0.1102
M257T1082	ESI Negative	Fruits (overall)	0.26	<0.0001	0.0006
M333T132	ESI Negative	Fruits (overall)	-0.19	0.0003	0.0687
M333T171	ESI Negative	Fruits (overall)	0.15	0.0048	0.2632
M365T170	ESI Negative	Fruits (overall)	0.19	0.0004	0.0687
M367T120	ESI Negative	Fruits (overall)	0.15	0.0056	0.2632
M452T1182	ESI Negative	Fruits (overall)	0.16	0.0042	0.2632
M526T1145	ESI Negative	Fruits (overall)	0.15	0.0044	0.2632
M557T1121	ESI Negative	Fruits (overall)	0.18	0.0011	0.1258
M586T1113	ESI Negative	Fruits (overall)	0.15	0.0052	0.2632
M625T131	ESI Negative	Fruits (overall)	-0.16	0.0041	0.2632
M95T186	ESI Positive	Grapes	0.15	0.0051	0.9993
M188T992	ESI Negative	Grapes	0.15	0.0045	0.7780
M239T776	ESI Positive	Grapes	0.16	0.0028	0.9993
M255T861	ESI Negative	Grapes	0.17	0.0019	0.7780
M309T1341	ESI Negative	Grapes	0.16	0.0032	0.7780
M415T945	ESI Negative	Grapes	0.16	0.0035	0.7780
M202T141	ESI Negative	Green beans	-0.16	0.0028	0.9552
M287T659	ESI Positive	Green beans	0.16	0.0025	0.9322
M679T127	ESI Negative	Green beans	-0.18	0.0010	0.6580
M181T1425	ESI Negative	Infusion, chicory and artificially sweetened drinks	-0.16	0.0027	0.6103
M188T992	ESI Negative	Infusion, chicory and artificially sweetened drinks	0.19	0.0003	0.1092
M190T869	ESI Positive	Infusion, chicory and artificially sweetened drinks	0.23	<0.0001	0.0115
M231T848	ESI Negative	Infusion, chicory and artificially sweetened drinks	0.22	<0.0001	0.0431
M411T131	ESI Positive	Infusion, chicory and artificially sweetened drinks	0.18	0.0010	0.2609
M523T123	ESI Negative	Infusion, chicory and artificially sweetened drinks	-0.15	0.0047	0.8136
M151T169	ESI Negative	Leafy vegetable	-0.15	0.0042	0.9937
M232T639	ESI Positive	Leafy vegetable	0.16	0.0027	0.4693
M294T1037	ESI Positive	Leafy vegetable	-0.18	0.0007	0.1719
M367T157	ESI Positive	Leafy vegetable	-0.16	0.0037	0.4746
M428T1048	ESI Positive	Leafy vegetable	-0.19	0.0005	0.1719
M860T141	ESI Positive	Leafy vegetable	-0.15	0.0045	0.4746
M123T626	ESI Positive	Lean fish	-0.18	0.0008	0.3614
M197T696	ESI Positive	Lean fish	-0.17	0.0018	0.3614
M569T1028	ESI Positive	Lean fish	0.17	0.0021	0.3614
M680T1141	ESI Negative	Lean fish	0.17	0.0022	0.9952
M65T138	ESI Positive	Legumes (overall)	0.21	<0.0001	0.0323
M391T1276	ESI Negative	Legumes (overall)	0.16	0.0034	0.2510
M408T690	ESI Negative	Legumes (overall)	-0.16	0.0026	0.2510
M467T1360	ESI Negative	Legumes (overall)	0.16	0.0027	0.2510
M467T1402	ESI Negative	Legumes (overall)	0.15	0.0046	0.2663
M489T1402	ESI Negative	Legumes (overall)	0.15	0.0053	0.2828



## ANNEXE 12

M500T1144	ESI Negative	Legumes (overall)	0.16	0.0032	0.2510
M517T1268	ESI Negative	Legumes (overall)	0.16	0.0039	0.2510
M519T1320	ESI Negative	Legumes (overall)	0.16	0.0033	0.2510
M557T1318	ESI Negative	Legumes (overall)	0.18	0.0009	0.2062
M571T123	ESI Negative	Legumes (overall)	-0.19	0.0006	0.2062
M573T1415	ESI Negative	Legumes (overall)	0.16	0.0040	0.2510
M591T1431	ESI Negative	Legumes (overall)	0.17	0.0013	0.2269
M667T120	ESI Negative	Legumes (overall)	0.18	0.0006	0.2062
M160T184	ESI Positive	Milk (overall)	0.19	0.0004	0.1448
<b>M286T989</b>	ESI Positive	Milk (overall)	-0.18	0.0006	0.1448
M453T1242	ESI Positive	Milk (overall)	-0.15	0.0048	0.3745
M471T131	ESI Positive	Milk (overall)	-0.16	0.0036	0.3745
M527T1025	ESI Positive	Milk (overall)	-0.18	0.0008	0.1448
M551T121	ESI Negative	Milk (overall)	-0.17	0.0017	0.9646
M586T862	ESI Positive	Milk (overall)	-0.15	0.0041	0.3745
M587T815	ESI Positive	Milk (overall)	-0.15	0.0050	0.3745
M95T186	ESI Positive	mPNNS-GS	0.15	0.0133	0.2512
M101T172	ESI Negative	mPNNS-GS	-0.18	0.0040	0.1202
M105T733	ESI Positive	mPNNS-GS	0.20	0.0009	0.0821
M115T590	ESI Negative	mPNNS-GS	-0.22	0.0002	0.0569
M129T772	ESI Negative	mPNNS-GS	-0.17	0.0061	0.1572
M129T814	ESI Negative	mPNNS-GS	-0.16	0.0102	0.1961
M132T445	ESI Positive	mPNNS-GS	-0.17	0.0045	0.1956
M149T132	ESI Positive	mPNNS-GS	-0.16	0.0079	0.2194
M153T739	ESI Positive	mPNNS-GS	-0.17	0.0064	0.2103
M157T180	ESI Negative	mPNNS-GS	0.15	0.0130	0.2136
M161T107	ESI Negative	mPNNS-GS	-0.19	0.0023	0.0976
M178T792	ESI Negative	mPNNS-GS	0.18	0.0029	0.1004
M180T733	ESI Positive	mPNNS-GS	0.21	0.0005	0.0575
M187T116	ESI Negative	mPNNS-GS	-0.16	0.0111	0.2076
M188T992	ESI Negative	mPNNS-GS	0.19	0.0025	0.0976
M190T869	ESI Positive	mPNNS-GS	0.22	0.0002	0.0378
M191T171	ESI Negative	mPNNS-GS	-0.21	0.0008	0.0809
M197T814	ESI Negative	mPNNS-GS	-0.21	0.0007	0.0809
M202T733	ESI Positive	mPNNS-GS	0.17	0.0045	0.1956
M203T140	ESI Positive	mPNNS-GS	-0.18	0.0035	0.1956
M204T140	ESI Positive	mPNNS-GS	-0.17	0.0048	0.1956
M204T891	ESI Negative	mPNNS-GS	0.15	0.0145	0.2230
M215T131	ESI Negative	mPNNS-GS	-0.16	0.0102	0.1961
M221T1057	ESI Negative	mPNNS-GS	0.15	0.0133	0.2136
M225T136	ESI Positive	mPNNS-GS	-0.15	0.0140	0.2552
M226T895	ESI Negative	mPNNS-GS	0.20	0.0014	0.0927
M243T127	ESI Positive	mPNNS-GS	0.15	0.0126	0.2512
M246T792	ESI Negative	mPNNS-GS	0.19	0.0020	0.0975
M254T1156	ESI Negative	mPNNS-GS	-0.16	0.0085	0.1906
M257T1082	ESI Negative	mPNNS-GS	0.28	<0.0001	0.0033

## ANNEXE 12

M261T142	ESI Positive	mPNNS-GS	-0.24	<0.0001	0.0217
M262T142	ESI Positive	mPNNS-GS	-0.17	0.0048	0.1956
M275T131	ESI Negative	mPNNS-GS	-0.17	0.0047	0.1362
M275T132	ESI Positive	mPNNS-GS	-0.24	<0.0001	0.0217
M280T122	ESI Negative	mPNNS-GS	-0.19	0.0025	0.0976
M313T171	ESI Negative	mPNNS-GS	-0.16	0.0093	0.1942
M319T142	ESI Positive	mPNNS-GS	-0.18	0.0027	0.1956
M321T142	ESI Positive	mPNNS-GS	-0.16	0.0089	0.2231
M329T136	ESI Positive	mPNNS-GS	-0.18	0.0030	0.1956
M333T995	ESI Negative	mPNNS-GS	0.16	0.0086	0.1906
M339T105	ESI Negative	mPNNS-GS	0.17	0.0067	0.1603
M355T130	ESI Positive	mPNNS-GS	-0.16	0.0094	0.2258
M357T792	ESI Negative	mPNNS-GS	0.19	0.0016	0.0927
M373T107	ESI Negative	mPNNS-GS	0.18	0.0029	0.1004
M377T142	ESI Positive	mPNNS-GS	-0.16	0.0110	0.2390
M379T112	ESI Negative	mPNNS-GS	0.15	0.0123	0.2136
M379T792	ESI Negative	mPNNS-GS	0.20	0.0009	0.0809
M379T995	ESI Negative	mPNNS-GS	0.15	0.0129	0.2136
M380T995	ESI Negative	mPNNS-GS	0.18	0.0034	0.1126
M383T140	ESI Positive	mPNNS-GS	-0.16	0.0101	0.2308
M385T855	ESI Negative	mPNNS-GS	-0.15	0.0127	0.2136
M395T131	ESI Negative	mPNNS-GS	-0.19	0.0018	0.0964
M397T131	ESI Negative	mPNNS-GS	-0.20	0.0009	0.0809
M401T995	ESI Negative	mPNNS-GS	0.19	0.0015	0.0927
M427T120	ESI Negative	mPNNS-GS	0.18	0.0039	0.1202
M429T995	ESI Negative	mPNNS-GS	0.17	0.0064	0.1582
M441T110	ESI Negative	mPNNS-GS	0.19	0.0016	0.0927
M441T143	ESI Positive	mPNNS-GS	-0.17	0.0059	0.2103
M452T1182	ESI Negative	mPNNS-GS	0.24	0.0001	0.0365
M454T1064	ESI Positive	mPNNS-GS	0.16	0.0080	0.2194
M455T123	ESI Negative	mPNNS-GS	-0.15	0.0130	0.2136
M495T142	ESI Positive	mPNNS-GS	-0.15	0.0133	0.2512
M499T143	ESI Positive	mPNNS-GS	-0.21	0.0005	0.0575
M501T143	ESI Positive	mPNNS-GS	-0.16	0.0083	0.2194
M524T1138	ESI Negative	mPNNS-GS	0.19	0.0024	0.0976
M525T1139	ESI Negative	mPNNS-GS	0.17	0.0061	0.1572
M527T132	ESI Positive	mPNNS-GS	-0.17	0.0064	0.2103
M543T120	ESI Negative	mPNNS-GS	0.15	0.0142	0.2227
M549T120	ESI Negative	mPNNS-GS	0.17	0.0054	0.1492
M557T143	ESI Positive	mPNNS-GS	-0.15	0.0113	0.2390
M612T1142	ESI Negative	mPNNS-GS	0.16	0.0097	0.1961
M625T131	ESI Negative	mPNNS-GS	-0.21	0.0006	0.0809
M689T132	ESI Negative	mPNNS-GS	-0.16	0.0089	0.1926
M860T141	ESI Positive	mPNNS-GS	-0.16	0.0071	0.2194
M187T1062	ESI Negative	Nuts (overall)	-0.15	0.0049	0.5603
M199T914	ESI Negative	Nuts (overall)	0.16	0.0033	0.5603

## ANNEXE 12

M297T1199	ESI Negative	Nuts (overall)	-0.15	0.0045	0.5603
M321T1321	ESI Negative	Nuts (overall)	-0.16	0.0032	0.5603
M351T1319	ESI Negative	Nuts (overall)	-0.16	0.0033	0.5603
M400T1012	ESI Positive	Nuts (overall)	-0.17	0.0011	0.5767
M426T1020	ESI Positive	Nuts (overall)	-0.16	0.0022	0.5767
M471T122	ESI Negative	Nuts (overall)	-0.16	0.0034	0.5603
M76T145	ESI Positive	Offal	0.16	0.0031	0.7841
M369T1053	ESI Negative	Offal	-0.16	0.0042	0.9934
M437T1080	ESI Negative	Offal	-0.17	0.0021	0.9934
M450T1125	ESI Negative	Olive	-0.17	0.0016	0.5841
M512T1109	ESI Negative	Olive	-0.17	0.0017	0.5841
M601T121	ESI Negative	Olive	-0.16	0.0034	0.5841
M667T120	ESI Negative	Olive	-0.16	0.0028	0.5841
M131T607	ESI Positive	Other alcohols	-0.15	0.0049	0.5145
M188T187	ESI Positive	Other alcohols	-0.16	0.0026	0.5145
M278T188	ESI Positive	Other alcohols	-0.15	0.0043	0.5145
M403T186	ESI Positive	Other alcohols	0.17	0.0013	0.5145
M547T130	ESI Positive	Other alcohols	0.16	0.0034	0.5145
M118T182	ESI Positive	Other cheeses	0.16	0.0036	0.5099
M119T184	ESI Positive	Other cheeses	0.16	0.0032	0.5099
M279T947	ESI Positive	Other cheeses	-0.15	0.0040	0.5099
M336T863	ESI Positive	Other cheeses	-0.16	0.0032	0.5099
M221T122	ESI Positive	Other starchy food	-0.18	0.0008	0.4472
M104T139	ESI Positive	Other type of food (overall)	0.15	0.0045	0.4730
M146T131	ESI Negative	Other type of food (overall)	0.16	0.0039	0.4462
M155T1041	ESI Negative	Other type of food (overall)	-0.16	0.0030	0.4462
M169T139	ESI Positive	Other type of food (overall)	-0.16	0.0027	0.4730
M199T1228	ESI Negative	Other type of food (overall)	-0.17	0.0015	0.4462
M199T914	ESI Negative	Other type of food (overall)	-0.16	0.0033	0.4462
M219T185	ESI Positive	Other type of food (overall)	-0.16	0.0026	0.4730
M263T129	ESI Negative	Other type of food (overall)	-0.15	0.0051	0.4935
M321T128	ESI Negative	Other type of food (overall)	-0.16	0.0034	0.4462
M349T141	ESI Positive	Other type of food (overall)	-0.15	0.0039	0.4730
M433T141	ESI Positive	Other type of food (overall)	0.16	0.0023	0.4730
M579T125	ESI Negative	Other type of food (overall)	0.18	0.0012	0.4462
M116T154	ESI Positive	Other vegetables	-0.19	0.0005	0.1698
M133T114	ESI Positive	Other vegetables	0.17	0.0013	0.1698
M168T170	ESI Negative	Other vegetables	-0.21	<0.0001	0.0569
M199T147	ESI Negative	Other vegetables	-0.16	0.0026	0.2541
M207T662	ESI Positive	Other vegetables	-0.16	0.0034	0.1992
M209T149	ESI Negative	Other vegetables	-0.19	0.0003	0.0764
M209T609	ESI Positive	Other vegetables	-0.16	0.0023	0.1759
M213T979	ESI Negative	Other vegetables	-0.16	0.0038	0.2882
M218T573	ESI Positive	Other vegetables	-0.17	0.0012	0.1698
M219T185	ESI Positive	Other vegetables	-0.17	0.0016	0.1698
M221T137	ESI Positive	Other vegetables	-0.16	0.0030	0.1984

## ANNEXE 12

M335T170	ESI Negative	Other vegetables	-0.19	0.0003	0.0764
M357T170	ESI Negative	Other vegetables	-0.17	0.0020	0.2457
M358T170	ESI Negative	Other vegetables	-0.16	0.0030	0.2597
M397T1047	ESI Negative	Other vegetables	-0.17	0.0018	0.2457
M403T107	ESI Negative	Other vegetables	0.15	0.0053	0.3025
M437T1080	ESI Negative	Other vegetables	-0.15	0.0044	0.3013
M450T946	ESI Positive	Other vegetables	-0.17	0.0016	0.1698
M506T125	ESI Positive	Other vegetables	-0.17	0.0019	0.1698
M611T122	ESI Negative	Other vegetables	-0.17	0.0021	0.2457
<b>M265T186</b>	ESI Positive	Pasta and cereals	0.15	0.0043	0.6358
M464T1004	ESI Negative	Pasta and cereals	0.15	0.0047	0.9530
M531T1291	ESI Negative	Pasta and cereals	-0.18	0.0009	0.5931
M130T145	ESI Negative	Pasta, rice and cereals (overall)	0.15	0.0047	0.2488
M164T605	ESI Negative	Pasta, rice and cereals (overall)	0.19	0.0004	0.1351
<b>M166T144</b>	ESI Negative	Pasta, rice and cereals (overall)	0.16	0.0038	0.2440
M203T689	ESI Negative	Pasta, rice and cereals (overall)	0.16	0.0028	0.2440
M217T131	ESI Positive	Pasta, rice and cereals (overall)	0.16	0.0023	0.6475
M254T1156	ESI Negative	Pasta, rice and cereals (overall)	-0.16	0.0039	0.2440
M299T690	ESI Negative	Pasta, rice and cereals (overall)	0.16	0.0023	0.2440
M347T1339	ESI Negative	Pasta, rice and cereals (overall)	-0.15	0.0056	0.2593
M379T1339	ESI Negative	Pasta, rice and cereals (overall)	-0.15	0.0043	0.2461
M407T690	ESI Negative	Pasta, rice and cereals (overall)	0.16	0.0034	0.2440
M439T1314	ESI Negative	Pasta, rice and cereals (overall)	-0.15	0.0051	0.2494
M457T123	ESI Negative	Pasta, rice and cereals (overall)	0.17	0.0019	0.2440
M464T1004	ESI Negative	Pasta, rice and cereals (overall)	0.19	0.0006	0.1351
M475T1302	ESI Negative	Pasta, rice and cereals (overall)	-0.16	0.0036	0.2440
M498T1040	ESI Negative	Pasta, rice and cereals (overall)	0.16	0.0033	0.2440
M506T125	ESI Positive	Pasta, rice and cereals (overall)	0.16	0.0030	0.6475
M531T1291	ESI Negative	Pasta, rice and cereals (overall)	-0.22	<0.0001	0.0419
M65T138	ESI Positive	Pastries and sweetened cereal-based food (overall)	0.16	0.0033	0.2581
M263T142	ESI Positive	Pastries and sweetened cereal-based food (overall)	-0.19	0.0003	0.0836
M289T1035	ESI Positive	Pastries and sweetened cereal-based food (overall)	0.15	0.0045	0.2581
M290T1035	ESI Positive	Pastries and sweetened cereal-based food (overall)	0.17	0.0015	0.2581
M437T1067	ESI Negative	Pastries and sweetened cereal-based food (overall)	-0.16	0.0028	0.9984
M520T1034	ESI Positive	Pastries and sweetened cereal-based food (overall)	0.15	0.0050	0.2581
M542T1034	ESI Positive	Pastries and sweetened cereal-based food (overall)	0.17	0.0020	0.2581
M543T1034	ESI Positive	Pastries and sweetened cereal-based food (overall)	0.15	0.0043	0.2581
M567T125	ESI Positive	Pastries and sweetened cereal-based food (overall)	0.19	0.0003	0.0836
M79T132	ESI Negative	Potatoes	-0.18	0.0011	0.7522
M259T131	ESI Negative	Potatoes	-0.15	0.0053	0.7957
M261T130	ESI Negative	Potatoes	-0.15	0.0056	0.7957
M463T132	ESI Positive	Potatoes	0.15	0.0040	0.3951
M471T131	ESI Positive	Potatoes	0.15	0.0040	0.3951
M526T1025	ESI Positive	Potatoes	-0.16	0.0032	0.3951
M542T1034	ESI Positive	Potatoes	-0.16	0.0039	0.3951
M543T1034	ESI Positive	Potatoes	-0.15	0.0045	0.3951

## ANNEXE 12

M544T1096	ESI Positive	Potatoes	-0.17	0.0016	0.3951
M79T132	ESI Negative	Potatoes and tubers (overall)	-0.17	0.0015	0.5983
M187T158	ESI Positive	Potatoes and tubers (overall)	0.16	0.0024	0.4159
M367T157	ESI Positive	Potatoes and tubers (overall)	0.15	0.0050	0.4411
M448T1161	ESI Positive	Potatoes and tubers (overall)	-0.15	0.0047	0.4411
M463T132	ESI Positive	Potatoes and tubers (overall)	0.17	0.0019	0.4159
M526T1025	ESI Positive	Potatoes and tubers (overall)	-0.16	0.0023	0.4159
M527T1025	ESI Positive	Potatoes and tubers (overall)	-0.15	0.0040	0.4411
M103T152	ESI Negative	Pressed cooked cheese	0.15	0.0054	0.4143
M154T188	ESI Positive	Pressed cooked cheese	0.15	0.0042	0.6532
M216T131	ESI Negative	Pressed cooked cheese	0.18	0.0007	0.4143
M243T170	ESI Negative	Pressed cooked cheese	0.16	0.0041	0.4143
M489T1340	ESI Negative	Pressed cooked cheese	-0.16	0.0032	0.4143
M513T123	ESI Negative	Pressed cooked cheese	0.15	0.0054	0.4143
M557T143	ESI Positive	Pressed cooked cheese	0.16	0.0021	0.6532
M559T1402	ESI Negative	Pressed cooked cheese	-0.16	0.0031	0.4143
M560T1402	ESI Negative	Pressed cooked cheese	-0.15	0.0049	0.4143
<b>M587T121</b>	ESI Negative	Pressed cooked cheese	0.15	0.0056	0.4143
M79T132	ESI Negative	Processed meat (overall)	-0.17	0.0022	0.7785
M104T186	ESI Positive	Processed meat (overall)	0.16	0.0023	0.3067
<b>M166T144</b>	ESI Negative	Processed meat (overall)	0.15	0.0047	0.7785
M169T186	ESI Positive	Processed meat (overall)	0.15	0.0046	0.4743
M371T688	ESI Positive	Processed meat (overall)	0.17	0.0020	0.3067
M393T687	ESI Positive	Processed meat (overall)	0.17	0.0014	0.3067
M531T122	ESI Negative	Processed meat (overall)	-0.15	0.0055	0.7785
M613T816	ESI Positive	Processed meat (overall)	0.17	0.0017	0.3067
M107T128	ESI Positive	Red meat	-0.16	0.0029	0.4892
M468T986	ESI Positive	Red meat	-0.15	0.0044	0.4892
M490T986	ESI Positive	Red meat	-0.16	0.0030	0.4892
M512T1109	ESI Negative	Red meat	-0.16	0.0031	0.9366
M562T1115	ESI Negative	Red meat	-0.15	0.0051	0.9366
M107T128	ESI Positive	Red meat and offal (overall)	-0.18	0.0009	0.2760
M276T1010	ESI Positive	Red meat and offal (overall)	-0.15	0.0051	0.2760
M277T132	ESI Positive	Red meat and offal (overall)	0.16	0.0022	0.2760
M468T986	ESI Positive	Red meat and offal (overall)	-0.16	0.0024	0.2760
M490T986	ESI Positive	Red meat and offal (overall)	-0.16	0.0027	0.2760
M512T1109	ESI Negative	Red meat and offal (overall)	-0.17	0.0023	0.9399
M545T123	ESI Positive	Red meat and offal (overall)	0.16	0.0023	0.2760
M103T302	ESI Negative	Red wine	0.17	0.0014	0.1233
M117T691	ESI Negative	Red wine	0.22	<0.0001	0.0219
M118T691	ESI Negative	Red wine	0.19	0.0006	0.0813
M127T152	ESI Positive	Red wine	0.18	0.0007	0.1000
M129T171	ESI Negative	Red wine	0.16	0.0037	0.1640
M131T779	ESI Negative	Red wine	0.16	0.0036	0.1640
M131T824	ESI Negative	Red wine	0.16	0.0028	0.1640
M181T115	ESI Negative	Red wine	0.16	0.0029	0.1640

## ANNEXE 12

M185T690	ESI Negative	Red wine	0.19	0.0004	0.0707
M230T171	ESI Positive	Red wine	0.16	0.0027	0.2067
M231T187	ESI Positive	Red wine	0.16	0.0037	0.2191
M261T142	ESI Positive	Red wine	0.16	0.0033	0.2169
<b>M286T989</b>	ESI Positive	Red wine	0.18	0.0008	0.1000
M297T1290	ESI Negative	Red wine	0.18	0.0009	0.0842
M302T904	ESI Positive	Red wine	0.17	0.0017	0.1512
M309T1141	ESI Negative	Red wine	0.15	0.0055	0.1667
M366T1220	ESI Negative	Red wine	0.18	0.0008	0.0842
M369T1053	ESI Negative	Red wine	0.22	<0.0001	0.0219
M369T972	ESI Negative	Red wine	0.15	0.0056	0.1667
M378T1100	ESI Negative	Red wine	0.15	0.0046	0.1667
M437T1278	ESI Negative	Red wine	0.17	0.0019	0.1439
M443T1011	ESI Negative	Red wine	0.16	0.0024	0.1640
M445T1357	ESI Negative	Red wine	0.16	0.0038	0.1640
M463T1337	ESI Negative	Red wine	0.15	0.0043	0.1667
M473T1411	ESI Negative	Red wine	0.15	0.0047	0.1667
M481T1080	ESI Negative	Red wine	0.20	0.0002	0.0507
M482T1026	ESI Positive	Red wine	-0.17	0.0011	0.1172
M483T1026	ESI Positive	Red wine	-0.19	0.0004	0.1000
M485T1437	ESI Negative	Red wine	-0.15	0.0054	0.1667
M501T143	ESI Positive	Red wine	0.19	0.0003	0.1000
M589T123	ESI Negative	Red wine	-0.15	0.0049	0.1667
M591T123	ESI Negative	Red wine	-0.16	0.0037	0.1640
M69T103	ESI Positive	Rhubarb	0.17	0.0015	0.1407
M113T186	ESI Positive	Rhubarb	0.16	0.0021	0.1590
M113T725	ESI Positive	Rhubarb	0.20	0.0001	0.0719
M115T170	ESI Negative	Rhubarb	0.17	0.0017	0.1743
M116T114	ESI Positive	Rhubarb	0.19	0.0005	0.0888
M129T772	ESI Negative	Rhubarb	0.17	0.0018	0.1743
M129T814	ESI Negative	Rhubarb	0.16	0.0026	0.1768
M130T772	ESI Negative	Rhubarb	0.17	0.0021	0.1743
M131T607	ESI Positive	Rhubarb	0.15	0.0041	0.2050
M137T186	ESI Positive	Rhubarb	0.17	0.0016	0.1407
M143T139	ESI Positive	Rhubarb	0.16	0.0025	0.1676
M149T132	ESI Positive	Rhubarb	0.15	0.0048	0.2050
M189T186	ESI Positive	Rhubarb	0.19	0.0003	0.0828
<b>M192T181</b>	ESI Positive	Rhubarb	0.17	0.0012	0.1407
M197T141	ESI Positive	Rhubarb	0.15	0.0044	0.2050
M197T772	ESI Negative	Rhubarb	0.19	0.0005	0.1743
M197T814	ESI Negative	Rhubarb	0.15	0.0054	0.2259
M227T126	ESI Positive	Rhubarb	0.15	0.0040	0.2050
M343T131	ESI Positive	Rhubarb	0.15	0.0050	0.2050
M378T1100	ESI Negative	Rhubarb	0.16	0.0035	0.2130
M417T121	ESI Negative	Rhubarb	-0.19	0.0006	0.1743
M417T1433	ESI Negative	Rhubarb	-0.15	0.0054	0.2259

## ANNEXE 12

M485T1437	ESI Negative	Rhubarb	-0.16	0.0037	0.2130
M526T1145	ESI Negative	Rhubarb	0.18	0.0008	0.1743
M556T1121	ESI Negative	Rhubarb	0.18	0.0011	0.1743
M564T1152	ESI Negative	Rhubarb	0.17	0.0023	0.1743
M612T1142	ESI Negative	Rhubarb	0.17	0.0016	0.1743
M614T1162	ESI Negative	Rhubarb	0.15	0.0054	0.2259
M680T1141	ESI Negative	Rhubarb	0.16	0.0040	0.2131
M65T138	ESI Positive	Rice	0.20	0.0002	0.0967
M199T147	ESI Negative	Rice	0.16	0.0041	0.6488
M232T127	ESI Positive	Rice	0.15	0.0049	0.9920
M254T1156	ESI Negative	Rice	-0.15	0.0051	0.6488
M475T1302	ESI Negative	Rice	-0.18	0.0011	0.6488
M87T166	ESI Negative	Roots and tubers	-0.16	0.0031	0.3372
M103T152	ESI Negative	Roots and tubers	-0.16	0.0028	0.3372
M115T590	ESI Negative	Roots and tubers	-0.15	0.0055	0.3372
M127T152	ESI Positive	Roots and tubers	-0.19	0.0003	0.0922
M133T168	ESI Negative	Roots and tubers	-0.16	0.0041	0.3372
M156T136	ESI Positive	Roots and tubers	0.15	0.0043	0.4785
M247T693	ESI Positive	Roots and tubers	0.21	0.0001	0.0620
M277T122	ESI Negative	Roots and tubers	-0.16	0.0039	0.3372
M280T1189	ESI Positive	Roots and tubers	-0.15	0.0048	0.4785
M281T1189	ESI Positive	Roots and tubers	-0.15	0.0044	0.4785
M395T131	ESI Negative	Roots and tubers	-0.15	0.0049	0.3372
M431T113	ESI Negative	Roots and tubers	-0.18	0.0010	0.3372
M455T123	ESI Negative	Roots and tubers	-0.17	0.0014	0.3372
M635T113	ESI Negative	Roots and tubers	-0.17	0.0015	0.3372
M101T172	ESI Negative	Salty products (overall)	0.16	0.0029	0.6702
M209T217	ESI Negative	Salty products (overall)	0.16	0.0025	0.6702
<b>M364T125</b>	ESI Positive	Salty products (overall)	0.15	0.0041	0.9579
M577T128	ESI Negative	Salty products (overall)	-0.19	0.0006	0.3908
M146T158	ESI Positive	Sauces (overall)	-0.15	0.0048	0.7565
M175T107	ESI Negative	Sauces (overall)	0.16	0.0031	0.2934
M212T793	ESI Negative	Sauces (overall)	0.16	0.0037	0.2934
M213T793	ESI Negative	Sauces (overall)	0.16	0.0034	0.2934
M217T615	ESI Positive	Sauces (overall)	0.15	0.0045	0.7565
M243T111	ESI Negative	Sauces (overall)	0.18	0.0007	0.2934
M308T1031	ESI Negative	Sauces (overall)	0.16	0.0025	0.2934
M359T120	ESI Negative	Sauces (overall)	-0.15	0.0055	0.3781
M481T1080	ESI Negative	Sauces (overall)	0.16	0.0032	0.2934
M491T120	ESI Negative	Sauces (overall)	-0.16	0.0026	0.2934
M583T113	ESI Negative	Sauces (overall)	0.16	0.0038	0.2934
M651T113	ESI Negative	Sauces (overall)	0.17	0.0021	0.2934
M79T132	ESI Negative	Shellfish and seafood	0.15	0.0052	0.9678
M204T186	ESI Positive	Shellfish and seafood	-0.17	0.0021	0.5528
M230T171	ESI Positive	Shellfish and seafood	0.17	0.0017	0.5528
M214T133	ESI Positive	Soya-based food	0.16	0.0025	0.6673

## ANNEXE 12

M320T139	ESI Positive	Soya-based food	0.15	0.0048	0.6673
M325T128	ESI Negative	Soya-based food	-0.15	0.0056	0.7234
M505T128	ESI Negative	Soya-based food	-0.17	0.0015	0.7234
M515T126	ESI Positive	Soya-based food	0.15	0.0052	0.6673
M225T150	ESI Negative	Stone fruits	-0.22	<0.0001	0.0352
M240T188	ESI Positive	Stone fruits	-0.16	0.0031	0.3270
M315T141	ESI Positive	Stone fruits	0.17	0.0019	0.3270
M325T150	ESI Positive	Stone fruits	-0.15	0.0051	0.3460
M333T132	ESI Negative	Stone fruits	-0.17	0.0014	0.3282
M431T141	ESI Positive	Stone fruits	0.16	0.0029	0.3270
M433T141	ESI Positive	Stone fruits	0.17	0.0020	0.3270
M441T143	ESI Positive	Stone fruits	-0.16	0.0029	0.3270
M687T132	ESI Negative	Stone fruits	-0.15	0.0046	0.6409
M689T132	ESI Negative	Stone fruits	-0.17	0.0013	0.3282
M129T814	ESI Negative	Sugar confectionery (overall)	-0.16	0.0039	0.5077
M131T824	ESI Negative	Sugar confectionery (overall)	-0.18	0.0011	0.5077
M169T132	ESI Negative	Sugar confectionery (overall)	-0.16	0.0040	0.5077
M196T136	ESI Positive	Sugar confectionery (overall)	-0.21	<0.0001	0.0350
M221T122	ESI Negative	Sugar confectionery (overall)	0.16	0.0034	0.5077
M231T632	ESI Positive	Sugar confectionery (overall)	-0.20	0.0002	0.0350
M246T697	ESI Positive	Sugar confectionery (overall)	-0.17	0.0014	0.1814
M283T644	ESI Positive	Sugar confectionery (overall)	-0.20	0.0002	0.0350
M287T659	ESI Positive	Sugar confectionery (overall)	-0.16	0.0023	0.2443
M339T122	ESI Negative	Sugar confectionery (overall)	0.16	0.0042	0.5077
M437T1080	ESI Negative	Sugar confectionery (overall)	-0.15	0.0048	0.5077
M97T118	ESI Negative	Sweetened drinks (overall)	0.15	0.0043	0.1798
M99T159	ESI Negative	Sweetened drinks (overall)	0.15	0.0046	0.1798
M120T154	ESI Positive	Sweetened drinks (overall)	0.17	0.0015	0.5135
M129T171	ESI Negative	Sweetened drinks (overall)	0.19	0.0005	0.1576
M129T772	ESI Negative	Sweetened drinks (overall)	0.16	0.0032	0.1798
M130T772	ESI Negative	Sweetened drinks (overall)	0.16	0.0033	0.1798
M132T445	ESI Positive	Sweetened drinks (overall)	0.16	0.0025	0.5135
M133T168	ESI Negative	Sweetened drinks (overall)	0.16	0.0028	0.1798
M174T134	ESI Negative	Sweetened drinks (overall)	-0.16	0.0025	0.1798
M187T116	ESI Negative	Sweetened drinks (overall)	0.15	0.0047	0.1798
M197T772	ESI Negative	Sweetened drinks (overall)	0.18	0.0008	0.1576
M221T626	ESI Positive	Sweetened drinks (overall)	-0.15	0.0049	0.5135
M227T131	ESI Negative	Sweetened drinks (overall)	-0.15	0.0044	0.1798
M255T116	ESI Negative	Sweetened drinks (overall)	0.15	0.0046	0.1798
M265T170	ESI Negative	Sweetened drinks (overall)	0.16	0.0033	0.1798
M301T169	ESI Negative	Sweetened drinks (overall)	0.17	0.0015	0.1798
M319T142	ESI Positive	Sweetened drinks (overall)	0.15	0.0048	0.5135
M391T115	ESI Negative	Sweetened drinks (overall)	0.16	0.0027	0.1798
M401T129	ESI Negative	Sweetened drinks (overall)	-0.20	0.0003	0.1576
M417T121	ESI Negative	Sweetened drinks (overall)	-0.16	0.0028	0.1798
M445T118	ESI Negative	Sweetened drinks (overall)	-0.18	0.0009	0.1576



## ANNEXE 12

M469T115	ESI Negative	Sweetened drinks (overall)	-0.16	0.0042	0.1798
M613T816	ESI Positive	Sweetened drinks (overall)	0.16	0.0037	0.5135
M97T126	ESI Negative	Sweetened products	0.16	0.0037	0.1908
M114T115	ESI Negative	Sweetened products	-0.17	0.0015	0.1649
M231T632	ESI Positive	Sweetened products	-0.16	0.0029	0.2507
M249T115	ESI Negative	Sweetened products	-0.17	0.0017	0.1649
M281T122	ESI Negative	Sweetened products	0.19	0.0005	0.1585
M335T122	ESI Negative	Sweetened products	0.15	0.0047	0.1908
M335T124	ESI Negative	Sweetened products	0.15	0.0047	0.1908
M339T122	ESI Negative	Sweetened products	0.16	0.0039	0.1908
M393T123	ESI Negative	Sweetened products	0.16	0.0034	0.1908
M401T114	ESI Negative	Sweetened products	-0.18	0.0012	0.1649
M429T995	ESI Negative	Sweetened products	-0.15	0.0046	0.1908
M445T118	ESI Negative	Sweetened products	-0.15	0.0043	0.1908
M451T123	ESI Negative	Sweetened products	0.18	0.0011	0.1649
M455T1064	ESI Positive	Sweetened products	0.16	0.0023	0.2507
M469T115	ESI Negative	Sweetened products	-0.16	0.0026	0.1908
M469T131	ESI Positive	Sweetened products	0.16	0.0030	0.2507
M482T1026	ESI Positive	Sweetened products	0.17	0.0020	0.2507
M482T1107	ESI Positive	Sweetened products	0.18	0.0010	0.2507
M483T1026	ESI Positive	Sweetened products	0.17	0.0020	0.2507
M510T1120	ESI Positive	Sweetened products	0.16	0.0033	0.2507
M511T118	ESI Negative	Sweetened products	-0.16	0.0030	0.1908
M537T115	ESI Negative	Sweetened products	-0.17	0.0017	0.1649
M573T123	ESI Negative	Sweetened products	0.20	0.0003	0.1585
M601T121	ESI Negative	Sweetened products	-0.17	0.0019	0.1649
M94T154	ESI Positive	Tea	-0.22	<0.0001	0.0117
M116T114	ESI Positive	Tea	0.16	0.0028	0.3734
M138T153	ESI Positive	Tea	-0.21	<0.0001	0.0117
M173T159	ESI Negative	Tea	0.17	0.0018	0.1139
M174T171	ESI Negative	Tea	-0.19	0.0004	0.0499
M175T170	ESI Negative	Tea	0.16	0.0032	0.1370
M175T172	ESI Positive	Tea	0.15	0.0047	0.4919
M180T1073	ESI Negative	Tea	-0.19	0.0005	0.0551
M191T146	ESI Negative	Tea	-0.25	<0.0001	0.0019
M227T159	ESI Negative	Tea	0.18	0.0009	0.0800
M259T131	ESI Negative	Tea	0.16	0.0025	0.1331
M333T995	ESI Negative	Tea	-0.18	0.0010	0.0800
M360T1203	ESI Negative	Tea	-0.16	0.0030	0.1370
M365T119	ESI Negative	Tea	0.19	0.0003	0.0499
M365T170	ESI Negative	Tea	0.16	0.0037	0.1370
M371T145	ESI Negative	Tea	-0.24	<0.0001	0.0020
M371T688	ESI Positive	Tea	-0.21	<0.0001	0.0117
M379T995	ESI Negative	Tea	-0.17	0.0021	0.1197
M380T995	ESI Negative	Tea	-0.18	0.0012	0.0800
M401T129	ESI Negative	Tea	0.18	0.0011	0.0800

## ANNEXE 12

M401T995	ESI Negative	Tea	-0.15	0.0047	0.1474
M429T809	ESI Negative	Tea	-0.24	<0.0001	0.0020
M429T995	ESI Negative	Tea	-0.16	0.0031	0.1370
M463T1291	ESI Negative	Tea	0.16	0.0036	0.1370
M507T128	ESI Negative	Tea	-0.15	0.0044	0.1454
M517T127	ESI Negative	Tea	0.16	0.0042	0.1452
M539T901	ESI Negative	Tea	-0.15	0.0056	0.1678
M592T1138	ESI Negative	Tea	0.16	0.0038	0.1370
M357T792	ESI Negative	Tomatoes	-0.16	0.0039	0.5690
<b>M415T1344</b>	ESI Negative	Tomatoes	-0.16	0.0040	0.5690
M501T128	ESI Negative	Tomatoes	0.15	0.0046	0.5690
M612T1142	ESI Negative	Tomatoes	0.15	0.0053	0.5690
M677T127	ESI Negative	Tomatoes	0.17	0.0017	0.5690
M105T733	ESI Positive	Unsweetened drinks (overall)	0.20	0.0002	0.0336
M173T159	ESI Negative	Unsweetened drinks (overall)	0.15	0.0052	0.4513
M178T792	ESI Negative	Unsweetened drinks (overall)	0.19	0.0006	0.1185
M180T733	ESI Positive	Unsweetened drinks (overall)	0.20	0.0002	0.0336
M181T651	ESI Positive	Unsweetened drinks (overall)	0.15	0.0051	0.1535
M182T547	ESI Positive	Unsweetened drinks (overall)	-0.15	0.0049	0.1535
M189T717	ESI Negative	Unsweetened drinks (overall)	0.19	0.0005	0.1185
M195T699	ESI Positive	Unsweetened drinks (overall)	0.17	0.0020	0.1029
M202T733	ESI Positive	Unsweetened drinks (overall)	0.19	0.0005	0.0482
M203T769	ESI Negative	Unsweetened drinks (overall)	0.18	0.0009	0.1186
M246T792	ESI Negative	Unsweetened drinks (overall)	0.17	0.0013	0.1545
<b>M286T989</b>	ESI Positive	Unsweetened drinks (overall)	-0.15	0.0052	0.1535
M357T792	ESI Negative	Unsweetened drinks (overall)	0.22	<0.0001	0.0260
M379T792	ESI Negative	Unsweetened drinks (overall)	0.18	0.0007	0.1185
M423T129	ESI Positive	Unsweetened drinks (overall)	0.16	0.0028	0.1108
M428T1048	ESI Positive	Unsweetened drinks (overall)	0.16	0.0021	0.1029
M454T1064	ESI Positive	Unsweetened drinks (overall)	0.20	0.0003	0.0336
M455T1064	ESI Positive	Unsweetened drinks (overall)	0.21	0.0001	0.0336
M480T1089	ESI Positive	Unsweetened drinks (overall)	0.17	0.0013	0.1029
M503T1030	ESI Positive	Unsweetened drinks (overall)	0.17	0.0021	0.1029
M522T1095	ESI Positive	Unsweetened drinks (overall)	0.17	0.0018	0.1029
M523T1095	ESI Positive	Unsweetened drinks (overall)	0.16	0.0024	0.1050
M526T1025	ESI Positive	Unsweetened drinks (overall)	0.15	0.0042	0.1487
M544T1096	ESI Positive	Unsweetened drinks (overall)	0.16	0.0029	0.1108
M570T1049	ESI Positive	Unsweetened drinks (overall)	0.17	0.0018	0.1029
M627T130	ESI Negative	Unsweetened drinks (overall)	-0.16	0.0040	0.3960
M187T158	ESI Positive	Vegetable fat (overall)	-0.16	0.0036	0.3766
M199T147	ESI Negative	Vegetable fat (overall)	-0.16	0.0024	0.5497
M209T149	ESI Negative	Vegetable fat (overall)	-0.18	0.0011	0.5497
M276T1010	ESI Positive	Vegetable fat (overall)	-0.18	0.0007	0.3458
M454T1073	ESI Negative	Vegetable fat (overall)	0.16	0.0041	0.5497
M494T1010	ESI Positive	Vegetable fat (overall)	-0.16	0.0026	0.3458
M495T1010	ESI Positive	Vegetable fat (overall)	-0.16	0.0026	0.3458

## ANNEXE 12

M516T1009	ESI Positive	Vegetable fat (overall)	-0.15	0.0047	0.3910
M525T714	ESI Positive	Vegetable fat (overall)	0.16	0.0025	0.3458
M187T158	ESI Positive	Vegetable fat except olive and soya oils	-0.18	0.0008	0.1373
M199T147	ESI Negative	Vegetable fat except olive and soya oils	-0.16	0.0040	0.1841
M209T149	ESI Negative	Vegetable fat except olive and soya oils	-0.18	0.0011	0.0820
M213T131	ESI Negative	Vegetable fat except olive and soya oils	0.16	0.0026	0.1460
M276T1010	ESI Positive	Vegetable fat except olive and soya oils	-0.18	0.0009	0.1373
M279T1338	ESI Negative	Vegetable fat except olive and soya oils	0.16	0.0024	0.1460
M295T1234	ESI Negative	Vegetable fat except olive and soya oils	0.16	0.0027	0.1460
M318T730	ESI Positive	Vegetable fat except olive and soya oils	0.17	0.0018	0.1387
M340T735	ESI Positive	Vegetable fat except olive and soya oils	0.17	0.0011	0.1373
M351T736	ESI Positive	Vegetable fat except olive and soya oils	0.16	0.0034	0.1995
M362T739	ESI Positive	Vegetable fat except olive and soya oils	0.16	0.0022	0.1462
M365T1220	ESI Negative	Vegetable fat except olive and soya oils	0.18	0.0006	0.0627
M373T742	ESI Positive	Vegetable fat except olive and soya oils	0.15	0.0039	0.1995
M384T743	ESI Positive	Vegetable fat except olive and soya oils	0.18	0.0007	0.1373
M414T1309	ESI Negative	Vegetable fat except olive and soya oils	0.15	0.0056	0.2163
M445T1338	ESI Negative	Vegetable fat except olive and soya oils	0.20	0.0002	0.0561
M445T1357	ESI Negative	Vegetable fat except olive and soya oils	0.19	0.0005	0.0581
M453T1073	ESI Negative	Vegetable fat except olive and soya oils	0.15	0.0043	0.1841
M454T1073	ESI Negative	Vegetable fat except olive and soya oils	0.19	0.0004	0.0581
M465T1275	ESI Negative	Vegetable fat except olive and soya oils	0.16	0.0039	0.1841
M467T1360	ESI Negative	Vegetable fat except olive and soya oils	0.21	0.0001	0.0560
M489T1214	ESI Negative	Vegetable fat except olive and soya oils	0.15	0.0055	0.2163
M489T1340	ESI Negative	Vegetable fat except olive and soya oils	0.20	0.0002	0.0560
M495T1010	ESI Positive	Vegetable fat except olive and soya oils	-0.15	0.0042	0.1995
M517T1268	ESI Negative	Vegetable fat except olive and soya oils	0.19	0.0005	0.0581
M525T714	ESI Positive	Vegetable fat except olive and soya oils	0.17	0.0013	0.1373
M564T1152	ESI Negative	Vegetable fat except olive and soya oils	0.17	0.0015	0.1001
M569T720	ESI Positive	Vegetable fat except olive and soya oils	0.17	0.0018	0.1387
M573T1415	ESI Negative	Vegetable fat except olive and soya oils	0.18	0.0009	0.0817
M87T166	ESI Negative	Vegetables (overall)	-0.15	0.0047	0.3612
M103T152	ESI Negative	Vegetables (overall)	-0.15	0.0051	0.3612
M113T725	ESI Positive	Vegetables (overall)	-0.16	0.0032	0.4281
M129T772	ESI Negative	Vegetables (overall)	-0.17	0.0020	0.3612
M129T814	ESI Negative	Vegetables (overall)	-0.17	0.0012	0.3612
M133T168	ESI Negative	Vegetables (overall)	-0.16	0.0038	0.3612
M197T696	ESI Positive	Vegetables (overall)	-0.17	0.0011	0.3344
M197T814	ESI Negative	Vegetables (overall)	-0.17	0.0019	0.3612
M267T742	ESI Positive	Vegetables (overall)	-0.16	0.0024	0.4189
M559T128	ESI Positive	Vegetables (overall)	0.17	0.0013	0.3344
M567T131	ESI Negative	Vegetables (overall)	-0.16	0.0041	0.3612
M639T123.2	ESI Negative	Vegetables (overall)	-0.17	0.0022	0.3612
M169T186	ESI Positive	White and rosé wines	0.23	<0.0001	0.0064
M230T171	ESI Positive	White and rosé wines	0.15	0.0043	0.4753
M359T185	ESI Positive	White and rosé wines	0.17	0.0018	0.3222

## ANNEXE 12

M403T186	ESI Positive	White and rosé wines	0.19	0.0004	0.1131
M509T132	ESI Negative	White and rosé wines	0.17	0.0015	0.7209
M87T166	ESI Negative	Poultry (overall)	-0.16	0.0032	0.2785
M144T163	ESI Positive	Poultry (overall)	-0.16	0.0030	0.5250
M144T935	ESI Negative	Poultry (overall)	0.17	0.0013	0.1473
M219T122	ESI Negative	Poultry (overall)	-0.22	<0.0001	0.0230
M321T142	ESI Positive	Poultry (overall)	-0.15	0.0050	0.5250
M333T995	ESI Negative	Poultry (overall)	0.18	0.0007	0.1371
M349T122	ESI Negative	Poultry (overall)	-0.15	0.0056	0.3369
M379T995	ESI Negative	Poultry (overall)	0.18	0.0010	0.1371
M380T995	ESI Negative	Poultry (overall)	0.18	0.0009	0.1371
M401T995	ESI Negative	Poultry (overall)	0.18	0.0007	0.1371
M429T809	ESI Negative	Poultry (overall)	0.15	0.0047	0.3369
M429T995	ESI Negative	Poultry (overall)	0.16	0.0024	0.2397
M453T1073	ESI Negative	Whole grain (overall)	-0.16	0.0029	0.8552
M454T1064	ESI Positive	Whole grain (overall)	0.15	0.0042	0.7390
M498T984	ESI Positive	Whole grain (overall)	-0.16	0.0029	0.7390
M79T132	ESI Negative	Yoghurts (overall)	0.20	0.0002	0.1005
M113T725	ESI Positive	Yoghurts (overall)	-0.16	0.0025	0.3946
M114T146	ESI Negative	Yoghurts (overall)	-0.15	0.0053	0.2622
M115T556	ESI Negative	Yoghurts (overall)	-0.15	0.0047	0.2622
M115T590	ESI Negative	Yoghurts (overall)	-0.18	0.0009	0.1393
M117T114	ESI Positive	Yoghurts (overall)	-0.16	0.0037	0.3946
M117T691	ESI Negative	Yoghurts (overall)	-0.18	0.0011	0.1393
M118T691	ESI Negative	Yoghurts (overall)	-0.17	0.0016	0.1619
M121T145	ESI Positive	Yoghurts (overall)	0.16	0.0036	0.3946
M129T772	ESI Negative	Yoghurts (overall)	-0.18	0.0009	0.1393
M157T130	ESI Negative	Yoghurts (overall)	-0.20	0.0003	0.1005
M199T147	ESI Negative	Yoghurts (overall)	-0.18	0.0012	0.1393
M209T149	ESI Negative	Yoghurts (overall)	-0.15	0.0054	0.2622
M257T1082	ESI Negative	Yoghurts (overall)	0.16	0.0031	0.2154
M259T131	ESI Negative	Yoghurts (overall)	0.16	0.0039	0.2426
<b>M286T989</b>	ESI Positive	Yoghurts (overall)	-0.18	0.0006	0.1463
M401T1367	ESI Negative	Yoghurts (overall)	0.16	0.0026	0.2002
M449T817	ESI Positive	Yoghurts (overall)	-0.20	0.0002	0.1249
M465T122	ESI Negative	Yoghurts (overall)	-0.16	0.0024	0.2002

SU.VI.MAX, Supplémentation en Vitamines et Minéraux Antioxydants ; R; correlation coefficient ; P, P-value

The 14 ions selected in breast cancer analyses are highlighted in bold.

<sup>1</sup> Partial spearman correlation adjusted for age, BMI, menopausal status (pre/post-menopause), smoking status (current/former/never) and season at time of blood draw (a priori-defined periods: October-November/December-January-February/March-April-May), number of 24h-dietary records and mean of daily energy intake during the 2 years following the blood draw (kcal). Correlations regarding alcohol specific food groups were adjusted for mean of daily energy intake calculated without the alcohol component.

<sup>2</sup> P-trend after Benjamini Hochberg correction for multiple testing

ANNEXE 12  
Additional file 6

Supplementary Table S4 Spearman correlation matrix of the identified associated metabolites

Ions	M153T116 (2-amino-cyano-butanoic acid)		M192T181		M265T186		M286T989 (piperine)		M335T864		M364T125		M425T1158 (ATBC)		M97T134		M166T144		M201T1091		M413T967 (pregnene-triol sulfate)		M415T1344		M475T122		
	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	
M192T181	<b>0.25</b>	<b>&lt;0.0001</b>																									
M265T186	-0.02	0.7	<b>0.15</b>	<b>0.002</b>																							
M286T989 (piperine)	0.00	0.9	-0.01	0.9	-0.07	0.1																					
M335T864	<b>-0.14</b>	<b>0.004</b>	0.00	1	-0.02	0.7	0.06	0.2																			
M364T125	<b>-0.20</b>	<b>&lt;0.0001</b>	-0.10	0.06	0.01	0.8	0.04	0.4	<b>0.12</b>	<b>0.02</b>																	
M425T1158 (ATBC)	-0.09	0.07	0.01	0.8	0.07	0.2	0.08	0.1	-0.08	0.1	0.10	0.06															
M97T134	<b>0.36</b>	<b>&lt;0.0001</b>	0.09	0.09	<b>0.11</b>	<b>0.02</b>	<b>-0.10</b>	<b>0.04</b>	-0.09	0.06	<b>-0.11</b>	<b>0.02</b>	-0.01	0.9													
M166T144	0.04	0.4	0.05	0.3	<b>-0.15</b>	<b>0.002</b>	<b>0.15</b>	<b>0.004</b>	0.02	0.7	-0.03	0.6	-0.07	0.2	0.03	0.5											
M201T1091	0.04	0.4	-0.04	0.5	0.05	0.3	-0.07	0.1	0.01	0.8	-0.02	0.7	0.06	0.3	0.10	0.05	<b>-0.12</b>	<b>0.01</b>									
M413T967 (pregnene-triol sulfate)	0.06	0.2	-0.04	0.5	<b>0.12</b>	<b>0.01</b>	<b>0.11</b>	<b>0.03</b>	0.02	0.7	0.03	0.6	0.06	0.2	-0.08	0.1	0.02	0.7	<b>0.18</b>	<b>0.0004</b>							
M415T1344	-0.03	0.5	-0.06	0.3	<b>-0.13</b>	<b>0.009</b>	-0.03	0.5	-0.01	0.9	-0.06	0.3	-0.03	0.5	-0.06	0.3	<b>0.15</b>	<b>0.003</b>	<b>-0.32</b>	<b>&lt;0.0001</b>	<b>-0.16</b>	<b>0.002</b>					
M475T122	<b>-0.13</b>	<b>0.007</b>	-0.04	0.4	<b>-0.11</b>	<b>0.03</b>	0.02	0.6	-0.01	0.9	0.07	0.2	-0.04	0.4	-0.08	0.1	<b>0.19</b>	<b>0.0001</b>	<b>-0.11</b>	<b>0.03</b>	-0.04	0.4	<b>0.19</b>	<b>0.0002</b>			
M587T121	<b>0.13</b>	<b>0.01</b>	0.07	0.2	0.00	1	-0.06	0.3	-0.08	0.1	-0.06	0.3	-0.01	0.9	-0.05	0.3	0.01	0.8	<b>-0.13</b>	<b>0.009</b>	-0.05	0.3	<b>0.16</b>	<b>0.001</b>	<b>0.16</b>	<b>0.002</b>	

R, correlation coefficient

## Additional file 7

**Supplementary Table S5** Direction of the variations of the ions with selection frequencies >70% from adjusted logistic regression models<sup>1</sup> - SU.VI.MAX cohort, France (1994-2007)

Ions (annotation <sup>2</sup> )	Mass/retention time	Mode of detection	OR [95%CI]	P-value <sup>3</sup>
M114T115 (level 4 : unknown)	113.9899/1.92	ESI Negative	0.79 [0.63-0.97]	0.03
M498T1040 (level 4 : unknown)	498.289/17.33	ESI Negative	0.74 [0.57-0.97]	0.03
M230T171 (level 4 : unknown)	230.0971/2.86	ESI Positive	1.22 [0.97-1.54]	0.1
M423T120 (level 4 : unknown)	422.8358/2.01	ESI Negative	1.24 [0.99-1.56]	0.06
M454T1064 (level 4 : unknown)	454.2939/17.73	ESI Positive	0.80 [0.64-1.00]	0.05
M196T953 (level 4 : unknown)	196.1696/15.88	ESI Positive	1.17 [0.95-1.46]	0.1
M485T1437 (level 4 : unknown)	485.2107/23.95	ESI Negative	0.70 [0.55-0.88]	0.003
M491T120 (level 4 : unknown)	490.8212/1.99	ESI Negative	0.86 [0.69-1.07]	0.2
M581T123 (level 4 : unknown)	580.6127/2.06	ESI Negative	1.11 [0.90-1.38]	0.3

SU.VI.MAX, Supplémentation en Vitamines et Minéraux Antioxydants; OR, odds ratio; CI, confidence interval ;

<sup>1</sup> Logistic regression models considered separately the ions. These models were adjusted for age (continuous), BMI (continuous), season (a priori-defined periods: October-November/December-January-February/March-April-May), menopausal status (pre/post-menopause status), smoking status (current, former and non-smokers), height (continuous), physical activity (low, moderate, intense), education level (primary, secondary, superior), alcohol intake (continuous), use of hormone replacement therapy for menopause, number of children and family history of breast cancer at blood draw (baseline) and intervention group of the initial SU,VI,MAX trial (placebo/supplemented). Tests for linear trend were performed using the continuous variables. ORs were presented for a 1 SD-increase of the continuous variable (semi-quantification).

<sup>2</sup> Levels of confidence for every identification were given accordingly to Sumner et al.(1): level 1; formally identified compound (confirmed with analysis of authentic standard), level 2; putatively identified compound (based upon spectral similarity with public/commercial spectral libraries or reference compound in the literature and/or physicochemical properties), level 3; putatively characterized compound classes and level 4; unknown compound.

<sup>3</sup> Although p-values are generated by this model, they cannot be used for inference given the prior Elastic Net-based selection

## REFERENCE

1. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, et al. Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics*. 2007 Sep 19;3(3):211–21.

Annexe 13 Signaux RMN associés à l'alimentation (score DQI-I, mPNNS-GS, profil alimentaire « occidental », profil alimentaire « sain », groupes d'aliments) chez les femmes

Déplacement chimique (ppm) (assignations)	Séquence	R <sup>1</sup>	P	P-corrigée	Exposition
0,1671	NOESY1D	-0,16	0,0001	0,003	Autres légumes (avocat, asperge, artichaut, champignon, céleri)
0,4471	NOESY1D	-0,18	<0,0001	0,0007	Autres légumes (avocat, asperge, artichaut, champignon, céleri)
0,4871	NOESY1D	-0,18	<0,0001	0,001	Autres légumes (avocat, asperge, artichaut, champignon, céleri)
3,269	NOESY1D	-0,16	0,0002	0,004	Autres légumes (avocat, asperge, artichaut, champignon, céleri)
3,311	NOESY1D	-0,19	<0,0001	0,0007	Autres légumes (avocat, asperge, artichaut, champignon, céleri)
0,0003	NOESY1D	-0,15	0,0003	0,006	Boissons alcoolisées (totales)
0,2071	NOESY1D	-0,17	<0,0001	0,002	Boissons alcoolisées (totales)
0,2471	NOESY1D	-0,17	<0,0001	0,002	Boissons alcoolisées (totales)
0,2871	NOESY1D	-0,16	0,0001	0,003	Boissons alcoolisées (totales)
1,1671	NOESY1D	0,15	0,0003	0,006	Boissons alcoolisées (totales)
5,2069	NOESY1D	0,18	<0,0001	0,002	Boissons alcoolisées (totales)
2,3881	NOESY1D	0,15	0,0003	0,02	Café
2,3315	CPMG	-0,16	0,0003	0,006	Composante 2 Profil alimentaire "sain"
2,369	CPMG	-0,17	<0,0001	0,003	Composante 2 Profil alimentaire "sain"
2,6295	CPMG	-0,17	<0,0001	0,003	Composante 2 Profil alimentaire "sain"
2,6675	CPMG	-0,17	0,0001	0,003	Composante 2 Profil alimentaire "sain"
3,5425	CPMG	-0,15	0,0004	0,007	Composante 2 Profil alimentaire "sain"
0,0871	NOESY1D	0,17	<0,0001	0,01	Fruits (totaux)
0,0003	NOESY1D	0,17	<0,0001	0,002	Pomme et poire
0,0471	NOESY1D	0,16	0,0001	0,002	Pomme et poire
0,0871	NOESY1D	0,18	<0,0001	0,002	Pomme et poire
0,1271	NOESY1D	0,16	0,0001	0,002	Pomme et poire
0,1671	NOESY1D	0,16	0,0001	0,002	Pomme et poire
0,2071	NOESY1D	0,17	<0,0001	0,002	Pomme et poire
0,2471	NOESY1D	0,17	<0,0001	0,002	Pomme et poire
0,0003	NOESY1D	0,15	0,0002	0,005	Produits à base de sucre/sucreries (miel, bonbons, confitures, etc.)
0,2071	NOESY1D	0,18	<0,0001	0,001	Produits à base de sucre/sucreries (miel, bonbons, confitures, etc.)
0,2471	NOESY1D	0,18	<0,0001	0,002	Produits à base de sucre/sucreries (miel, bonbons, confitures, etc.)
0,4071	NOESY1D	0,17	<0,0001	0,002	Produits à base de sucre/sucreries (miel, bonbons, confitures, etc.)
0,4471	NOESY1D	0,16	0,0002	0,005	Produits à base de sucre/sucreries (miel, bonbons, confitures, etc.)
2,7911	NOESY1D	0,19	<0,0001	0,0007	Produits de la mer (totaux)
2,8311	NOESY1D	0,16	0,0002	0,009	Produits de la mer (totaux)
5,2069	NOESY1D	0,16	0,0002	0,02	Vin rouge
0,1671	NOESY1D	-0,15	0,0003	0,02	Vins blancs et rosés
0,2071	NOESY1D	-0,17	<0,0001	0,005	Vins blancs et rosés
2,5915	CPMG	-0,15	0,0003	0,03	Vins blancs et rosés

R; correlation coefficient

<sup>1</sup> Les corrélations de Spearman ont été ajustées sur les caractéristiques suivantes : âge, IMC, statut tabagique, statut ménopausique, saison de prélèvement sanguin, nombre d'enregistrements alimentaires sur 24 heures et moyenne de l'apport énergétique quotidien pendant les 2 années suivant le prélèvement sanguin.

Annexe 14 Signaux RMN associés à l'alimentation (score DQI-I, mPNNS-GS, profil alimentaire « occidental », profil alimentaire « sain », groupes d'aliments) chez les hommes

Déplacement chimique (ppm) (assignations)	Séquence	R <sup>1</sup>	P	P-corrigée	Exposition
7,8307	CPMG	-0,15	0,01	1,0	Abats
2,277	CPMG	-0,16	0,009	0,8	Agrumes
2,873	CPMG	-0,17	0,006	0,8	Agrumes
3,269	CPMG	0,18	0,004	0,8	Aliments à base de soja
2,411	CPMG	0,19	0,002	0,4	Allium
1,899	CPMG	-0,23	0,0001	0,02	Apéritifs (alcool)
2,705	CPMG	0,18	0,004	0,3	Apéritifs (alcool)
6,1698	CPMG	0,16	0,009	0,5	Apéritifs (alcool)
1,1271	CPMG	0,17	0,007	1,0	Autres alcools (cidre, champagne, etc.)
1,707	NOESY1D	-0,16	0,008	1,0	Autres alcools (cidre, champagne, etc.)
0,9107	CPMG	0,15	0,01	0,9	Autres féculents (châtaigne, tapioca, plantain)
2,873	CPMG	0,15	0,01	0,9	Autres féculents (châtaigne, tapioca, plantain)
2,374	NOESY1D	-0,16	0,009	0,4	Autres légumes (avocat, asperge, artichaut, champignon, céleri)
3,1666	NOESY1D	0,16	0,009	0,4	Autres légumes (avocat, asperge, artichaut, champignon, céleri)
3,2696	NOESY1D	0,17	0,006	0,4	Autres légumes (avocat, asperge, artichaut, champignon, céleri)
0,4867	NOESY1D	-0,23	0,0001	0,01	Autres types d'aliments (condiments, produits énergétique, à base de soja etc)
0,5267	NOESY1D	-0,23	0,0002	0,01	Autres types d'aliments (condiments, produits énergétique, à base de soja etc)
0,5667	NOESY1D	-0,21	0,0007	0,04	Autres types d'aliments (condiments, produits énergétique, à base de soja etc)
0,7267	NOESY1D	-0,16	0,01	0,2	Autres types d'aliments (condiments, produits énergétique, à base de soja etc)
0,7667	NOESY1D	-0,19	0,002	0,06	Autres types d'aliments (condiments, produits énergétique, à base de soja etc)
2,4711	CPMG	0,15	0,01	0,3	Autres types d'aliments (condiments, produits énergétique, à base de soja etc)
3,3106	NOESY1D	-0,16	0,01	0,2	Autres types d'aliments (condiments, produits énergétique, à base de soja etc)
6,6064	NOESY1D	0,15	0,01	0,2	Autres types d'aliments (condiments, produits énergétique, à base de soja etc)
6,6464	NOESY1D	0,19	0,002	0,07	Autres types d'aliments (condiments, produits énergétique, à base de soja etc)
6,6864	NOESY1D	0,16	0,008	0,2	Autres types d'aliments (condiments, produits énergétique, à base de soja etc)
6,7264	NOESY1D	0,18	0,003	0,07	Autres types d'aliments (condiments, produits énergétique, à base de soja etc)
6,912	CPMG	0,16	0,009	0,3	Autres types d'aliments (condiments, produits énergétique, à base de soja etc)
7,1104	NOESY1D	0,18	0,003	0,07	Autres types d'aliments (condiments, produits énergétique, à base de soja etc)
7,4703	NOESY1D	0,16	0,01	0,2	Autres types d'aliments (condiments, produits énergétique, à base de soja etc)
7,6867	CPMG	-0,16	0,007	0,3	Autres types d'aliments (condiments, produits énergétique, à base de soja etc)
7,8307	CPMG	0,19	0,002	0,1	Autres types d'aliments (condiments, produits énergétique, à base de soja etc)
7,8707	CPMG	0,20	0,001	0,1	Autres types d'aliments (condiments, produits énergétique, à base de soja etc)
7,9107	CPMG	0,17	0,007	0,3	Autres types d'aliments (condiments, produits énergétique, à base de soja etc)
7,9207	CPMG	0,16	0,007	0,3	Autres types d'aliments (condiments, produits énergétique, à base de soja etc)
3,367	CPMG	-0,18	0,003	0,1	Baies
3,437	CPMG	-0,18	0,003	0,1	Baies
3,496	CPMG	-0,20	0,0009	0,05	Baies
3,7066	NOESY1D	-0,16	0,007	0,8	Baies
3,713	CPMG	-0,21	0,0006	0,05	Baies
3,795	CPMG	-0,17	0,006	0,2	Baies
3,86	CPMG	-0,21	0,0007	0,05	Baies
1,759	CPMG	-0,17	0,006	1,0	Banane
0,6067	NOESY1D	0,19	0,002	0,02	Boissons alcoolisées (totales)
0,6467	NOESY1D	0,21	0,0006	0,007	Boissons alcoolisées (totales)
0,7267	NOESY1D	0,18	0,003	0,02	Boissons alcoolisées (totales)
0,7667	NOESY1D	0,17	0,005	0,03	Boissons alcoolisées (totales)
0,8067	NOESY1D	0,16	0,007	0,04	Boissons alcoolisées (totales)
0,9107	NOESY1D	-0,16	0,01	0,06	Boissons alcoolisées (totales)
1,051	NOESY1D	0,18	0,004	0,03	Boissons alcoolisées (totales)
1,051	CPMG	0,16	0,01	0,1	Boissons alcoolisées (totales)
1,0667	NOESY1D	0,26	<0,0001	0,002	Boissons alcoolisées (totales)
1,0887	NOESY1D	0,25	<0,0001	0,002	Boissons alcoolisées (totales)
1,1267	NOESY1D	0,24	<0,0001	0,003	Boissons alcoolisées (totales)
1,1271	CPMG	0,27	<0,0001	0,002	Boissons alcoolisées (totales)
1,145	CPMG	-0,20	0,0009	0,03	Boissons alcoolisées (totales)
1,1667	NOESY1D	0,25	<0,0001	0,002	Boissons alcoolisées (totales)
1,1671	CPMG	0,15	0,01	0,1	Boissons alcoolisées (totales)
1,3907	NOESY1D	0,19	0,001	0,01	Boissons alcoolisées (totales)
1,4287	NOESY1D	0,23	0,0001	0,003	Boissons alcoolisées (totales)
1,4467	NOESY1D	0,22	0,0003	0,005	Boissons alcoolisées (totales)
1,5907	NOESY1D	0,15	0,01	0,06	Boissons alcoolisées (totales)



## ANNEXE 14

1,6307	NOESY1D	0,24	<0,0001	0,003	Boissons alcoolisées (totales)
1,6357	NOESY1D	0,23	0,0001	0,003	Boissons alcoolisées (totales)
1,707	CPMG	-0,22	0,0004	0,02	Boissons alcoolisées (totales)
1,7507	NOESY1D	0,21	0,0005	0,006	Boissons alcoolisées (totales)
1,759	CPMG	0,17	0,004	0,08	Boissons alcoolisées (totales)
1,7907	NOESY1D	0,23	0,0001	0,003	Boissons alcoolisées (totales)
1,8307	NOESY1D	0,19	0,002	0,02	Boissons alcoolisées (totales)
1,899	CPMG	-0,23	0,0002	0,01	Boissons alcoolisées (totales)
1,927	NOESY1D	0,18	0,004	0,03	Boissons alcoolisées (totales)
1,927	CPMG	0,16	0,01	0,1	Boissons alcoolisées (totales)
2,0941	CPMG	-0,25	<0,0001	0,003	Boissons alcoolisées (totales)
2,1507	NOESY1D	0,17	0,005	0,03	Boissons alcoolisées (totales)
2,2637	NOESY1D	0,21	0,0004	0,005	Boissons alcoolisées (totales)
2,3107	NOESY1D	0,22	0,0003	0,005	Boissons alcoolisées (totales)
2,411	CPMG	-0,20	0,0009	0,03	Boissons alcoolisées (totales)
2,647	NOESY1D	-0,18	0,003	0,02	Boissons alcoolisées (totales)
2,6471	CPMG	-0,17	0,005	0,08	Boissons alcoolisées (totales)
2,705	CPMG	0,17	0,005	0,08	Boissons alcoolisées (totales)
2,708	NOESY1D	0,15	0,01	0,06	Boissons alcoolisées (totales)
2,8481	CPMG	-0,17	0,005	0,08	Boissons alcoolisées (totales)
3,1666	NOESY1D	0,23	0,0001	0,003	Boissons alcoolisées (totales)
3,167	CPMG	0,18	0,003	0,08	Boissons alcoolisées (totales)
5,2065	NOESY1D	0,19	0,001	0,01	Boissons alcoolisées (totales)
6,2864	NOESY1D	-0,15	0,01	0,06	Boissons alcoolisées (totales)
6,3274	NOESY1D	-0,15	0,01	0,07	Boissons alcoolisées (totales)
6,8634	NOESY1D	-0,20	0,001	0,01	Boissons alcoolisées (totales)
6,9054	NOESY1D	-0,17	0,006	0,03	Boissons alcoolisées (totales)
7,1078	CPMG	-0,16	0,008	0,1	Boissons alcoolisées (totales)
7,2787	CPMG	-0,17	0,006	0,09	Boissons alcoolisées (totales)
7,3437	CPMG	-0,17	0,004	0,08	Boissons alcoolisées (totales)
0,9697	CPMG	-0,15	0,01	0,2	Boissons non sucrées (totales)
0,9997	CPMG	0,16	0,01	0,2	Boissons non sucrées (totales)
1,021	CPMG	-0,16	0,008	0,2	Boissons non sucrées (totales)
1,0811	CPMG	0,17	0,005	0,1	Boissons non sucrées (totales)
1,0881	CPMG	0,18	0,003	0,1	Boissons non sucrées (totales)
1,1271	CPMG	0,20	0,001	0,09	Boissons non sucrées (totales)
1,707	CPMG	-0,18	0,003	0,1	Boissons non sucrées (totales)
1,707	NOESY1D	-0,16	0,009	0,7	Boissons non sucrées (totales)
3,496	CPMG	-0,19	0,001	0,09	Boissons non sucrées (totales)
3,713	CPMG	-0,20	0,001	0,09	Boissons non sucrées (totales)
3,791	CPMG	-0,16	0,008	0,2	Boissons non sucrées (totales)
3,795	CPMG	-0,18	0,004	0,1	Boissons non sucrées (totales)
4,2136	NOESY1D	0,16	0,007	0,7	Boissons non sucrées (totales)
0,8067	NOESY1D	0,16	0,01	0,3	Boissons sucrées (totales)
0,8071	CPMG	0,17	0,006	0,1	Boissons sucrées (totales)
1,0811	CPMG	0,18	0,003	0,1	Boissons sucrées (totales)
1,0881	CPMG	0,18	0,003	0,1	Boissons sucrées (totales)
1,2067	NOESY1D	0,15	0,01	0,3	Boissons sucrées (totales)
1,2071	CPMG	0,15	0,01	0,1	Boissons sucrées (totales)
1,8211	CPMG	0,19	0,002	0,09	Boissons sucrées (totales)
2,0941	CPMG	0,16	0,01	0,1	Boissons sucrées (totales)
2,277	CPMG	-0,17	0,005	0,1	Boissons sucrées (totales)
2,4707	NOESY1D	-0,18	0,003	0,3	Boissons sucrées (totales)
2,4711	CPMG	-0,16	0,009	0,1	Boissons sucrées (totales)
3,951	CPMG	-0,16	0,007	0,1	Boissons sucrées (totales)
4,151	CPMG	-0,16	0,007	0,1	Boissons sucrées (totales)
4,2136	NOESY1D	0,17	0,005	0,3	Boissons sucrées (totales)
5,1669	CPMG	0,15	0,01	0,1	Boissons sucrées (totales)
5,2065	NOESY1D	0,16	0,01	0,3	Boissons sucrées (totales)
5,4309	CPMG	0,15	0,01	0,1	Boissons sucrées (totales)
6,2864	NOESY1D	-0,15	0,01	0,3	Boissons sucrées (totales)
7,0708	CPMG	-0,16	0,008	0,1	Boissons sucrées (totales)
7,1198	CPMG	-0,21	0,0007	0,09	Boissons sucrées (totales)
7,3437	CPMG	-0,16	0,008	0,1	Boissons sucrées (totales)
7,3747	CPMG	-0,17	0,006	0,1	Boissons sucrées (totales)
7,4307	CPMG	-0,19	0,002	0,09	Boissons sucrées (totales)
7,8307	CPMG	-0,19	0,002	0,09	Boissons sucrées (totales)
0,0003	CPMG	0,17	0,005	0,2	Café
0,0581	CPMG	0,16	0,01	0,2	Café
0,4091	CPMG	0,16	0,01	0,2	Café
0,6067	NOESY1D	0,16	0,01	0,1	Café

## ANNEXE 14

0,6071	CPMG	0,17	0,007	0,2	Café
0,6467	NOESY1D	0,17	0,005	0,1	Café
0,7131	CPMG	0,17	0,006	0,2	Café
0,7271	CPMG	0,16	0,01	0,2	Café
0,7671	CPMG	0,16	0,008	0,2	Café
1,051	CPMG	0,15	0,01	0,2	Café
1,0667	NOESY1D	0,18	0,004	0,1	Café
1,0887	NOESY1D	0,17	0,004	0,1	Café
1,1267	NOESY1D	0,15	0,01	0,2	Café
1,1667	NOESY1D	0,17	0,006	0,1	Café
1,1671	CPMG	0,15	0,01	0,2	Café
1,8867	NOESY1D	0,18	0,003	0,1	Café
1,927	NOESY1D	0,16	0,007	0,1	Café
2,3107	NOESY1D	0,19	0,001	0,1	Café
2,705	CPMG	0,15	0,01	0,2	Café
3,1666	NOESY1D	0,21	0,0006	0,09	Café
3,167	CPMG	0,19	0,002	0,2	Café
4,1906	NOESY1D	0,16	0,009	0,1	Café
4,3106	NOESY1D	0,16	0,009	0,1	Café
1,5211	CPMG	0,15	0,01	0,5	Céréales pour petit-déjeuner
2,6267	NOESY1D	-0,19	0,002	0,3	Céréales pour petit-déjeuner
3,496	CPMG	-0,17	0,007	0,5	Céréales pour petit-déjeuner
3,587	CPMG	-0,17	0,006	0,5	Céréales pour petit-déjeuner
3,86	CPMG	-0,15	0,01	0,5	Céréales pour petit-déjeuner
4,031	NOESY1D	0,17	0,007	0,5	Céréales pour petit-déjeuner
0,0003	CPMG	0,20	0,0009	0,05	Charcuterie
0,0581	CPMG	0,19	0,002	0,06	Charcuterie
0,1721	CPMG	0,18	0,004	0,1	Charcuterie
0,2321	CPMG	0,16	0,009	0,1	Charcuterie
0,3181	CPMG	0,20	0,001	0,06	Charcuterie
0,4091	CPMG	0,16	0,009	0,1	Charcuterie
2,0941	CPMG	-0,16	0,007	0,1	Charcuterie
5,0069	CPMG	-0,17	0,005	0,1	Charcuterie
6,7278	CPMG	-0,16	0,01	0,1	Charcuterie
6,7668	CPMG	-0,16	0,01	0,1	Charcuterie
7,0708	CPMG	-0,21	0,0006	0,05	Charcuterie
7,1078	CPMG	-0,15	0,01	0,2	Charcuterie
7,3437	CPMG	-0,17	0,007	0,1	Charcuterie
7,3747	CPMG	-0,17	0,005	0,1	Charcuterie
7,4307	CPMG	-0,24	<0,0001	0,01	Charcuterie
1,849	CPMG	-0,16	0,007	0,9	Chocolat
3,028	CPMG	-0,15	0,01	0,9	Chocolat
1,051	CPMG	-0,17	0,006	1,0	Choux
0,9997	CPMG	0,22	0,0002	0,01	Cola
0,9997	NOESY1D	0,18	0,002	0,2	Cola
1,051	NOESY1D	0,25	<0,0001	0,006	Cola
1,051	CPMG	0,23	0,0001	0,01	Cola
1,707	CPMG	-0,17	0,005	0,09	Cola
2,277	CPMG	-0,15	0,01	0,2	Cola
2,5711	CPMG	0,17	0,007	0,1	Cola
2,5797	NOESY1D	0,15	0,01	0,3	Cola
2,8481	CPMG	0,16	0,01	0,1	Cola
3,3656	NOESY1D	-0,16	0,007	0,2	Cola
3,7926	NOESY1D	0,16	0,007	0,2	Cola
3,9506	NOESY1D	-0,16	0,007	0,2	Cola
5,7265	NOESY1D	-0,17	0,005	0,2	Cola
5,7269	CPMG	-0,16	0,008	0,1	Cola
5,8308	CPMG	-0,17	0,007	0,1	Cola
6,968	CPMG	0,19	0,002	0,04	Cola
7,023	CPMG	-0,16	0,008	0,1	Cola
7,128	CPMG	-0,15	0,01	0,2	Cola
7,3437	CPMG	-0,22	0,0004	0,02	Cola
7,3747	CPMG	-0,22	0,0002	0,01	Cola
7,4307	CPMG	-0,19	0,002	0,05	Cola
7,8307	CPMG	-0,20	0,001	0,04	Cola
7,8707	CPMG	-0,17	0,004	0,09	Cola
0,4867	NOESY1D	-0,22	0,0002	0,02	Condiments et aides culinaires
0,5267	NOESY1D	-0,22	0,0003	0,02	Condiments et aides culinaires
0,5667	NOESY1D	-0,20	0,0009	0,05	Condiments et aides culinaires
0,7267	NOESY1D	-0,15	0,01	0,2	Condiments et aides culinaires
0,7667	NOESY1D	-0,19	0,002	0,07	Condiments et aides culinaires

## ANNEXE 14

2,4711	CPMG	0,16	0,009	0,3	Condiments et aides culinaires
3,3106	NOESY1D	-0,16	0,01	0,2	Condiments et aides culinaires
6,6064	NOESY1D	0,16	0,01	0,2	Condiments et aides culinaires
6,6464	NOESY1D	0,18	0,002	0,07	Condiments et aides culinaires
6,6864	NOESY1D	0,16	0,008	0,2	Condiments et aides culinaires
6,7264	NOESY1D	0,18	0,003	0,07	Condiments et aides culinaires
6,912	CPMG	0,16	0,01	0,3	Condiments et aides culinaires
7,1104	NOESY1D	0,18	0,003	0,07	Condiments et aides culinaires
7,4703	NOESY1D	0,16	0,01	0,2	Condiments et aides culinaires
7,6867	CPMG	-0,15	0,01	0,3	Condiments et aides culinaires
7,8307	CPMG	0,19	0,002	0,2	Condiments et aides culinaires
7,8707	CPMG	0,20	0,001	0,2	Condiments et aides culinaires
7,9107	CPMG	0,17	0,004	0,2	Condiments et aides culinaires
7,9207	CPMG	0,17	0,005	0,2	Condiments et aides culinaires
7,9837	CPMG	0,16	0,01	0,3	Condiments et aides culinaires
2,017	CPMG	-0,15	0,01	0,9	Crustacés et fruits de mer
2,018	NOESY1D	-0,18	0,003	0,5	Crustacés et fruits de mer
0,4867	NOESY1D	0,16	0,007	0,2	Desserts lactés
0,5267	NOESY1D	0,17	0,006	0,2	Desserts lactés
0,5667	NOESY1D	0,15	0,01	0,2	Desserts lactés
4,075	NOESY1D	-0,16	0,008	0,2	Desserts lactés
5,6869	CPMG	-0,16	0,01	0,6	Desserts lactés
5,7269	CPMG	-0,17	0,006	0,5	Desserts lactés
5,8308	CPMG	-0,18	0,004	0,5	Desserts lactés
6,6064	NOESY1D	-0,15	0,01	0,2	Desserts lactés
7,2093	NOESY1D	-0,15	0,01	0,2	Desserts lactés
7,2703	NOESY1D	-0,17	0,005	0,2	Desserts lactés
7,4703	NOESY1D	-0,17	0,006	0,2	Desserts lactés
7,5103	NOESY1D	-0,16	0,007	0,2	Desserts lactés
4,1906	NOESY1D	-0,15	0,01	1,0	DQI-I
5,9654	NOESY1D	-0,17	0,006	1,0	DQI-I
0,8071	CPMG	0,15	0,01	0,2	Fromage à pâte pressée cuite
0,9107	CPMG	-0,16	0,01	0,2	Fromage à pâte pressée cuite
1,2071	CPMG	0,17	0,004	0,2	Fromage à pâte pressée cuite
1,3511	CPMG	0,15	0,01	0,2	Fromage à pâte pressée cuite
1,5271	CPMG	0,15	0,01	0,2	Fromage à pâte pressée cuite
2,6121	CPMG	-0,15	0,01	0,2	Fromage à pâte pressée cuite
2,9271	CPMG	-0,16	0,01	0,2	Fromage à pâte pressée cuite
4,2136	NOESY1D	0,17	0,007	1,0	Fromage à pâte pressée cuite
5,3509	CPMG	0,16	0,009	0,2	Fromage à pâte pressée cuite
5,8638	CPMG	0,15	0,01	0,2	Fromage à pâte pressée cuite
6,857	CPMG	0,17	0,005	0,2	Fromage à pâte pressée cuite
7,159	CPMG	0,18	0,004	0,2	Fromage à pâte pressée cuite
0,5667	NOESY1D	-0,18	0,003	0,08	Fromages (sauf à pâte pressée cuite)
0,9107	CPMG	-0,17	0,005	0,4	Fromages (sauf à pâte pressée cuite)
2,017	CPMG	-0,16	0,008	0,4	Fromages (sauf à pâte pressée cuite)
2,2641	CPMG	-0,16	0,01	0,4	Fromages (sauf à pâte pressée cuite)
2,4707	NOESY1D	0,18	0,004	0,08	Fromages (sauf à pâte pressée cuite)
4,1506	NOESY1D	0,17	0,006	0,09	Fromages (sauf à pâte pressée cuite)
5,0465	NOESY1D	-0,15	0,01	0,1	Fromages (sauf à pâte pressée cuite)
5,0865	NOESY1D	-0,23	0,0002	0,03	Fromages (sauf à pâte pressée cuite)
5,1265	NOESY1D	-0,20	0,001	0,08	Fromages (sauf à pâte pressée cuite)
5,3905	NOESY1D	-0,18	0,003	0,08	Fromages (sauf à pâte pressée cuite)
5,4305	NOESY1D	-0,19	0,002	0,08	Fromages (sauf à pâte pressée cuite)
5,4705	NOESY1D	-0,18	0,002	0,08	Fromages (sauf à pâte pressée cuite)
5,5105	NOESY1D	-0,15	0,01	0,1	Fromages (sauf à pâte pressée cuite)
5,5905	NOESY1D	-0,17	0,006	0,09	Fromages (sauf à pâte pressée cuite)
6,7664	NOESY1D	0,15	0,01	0,1	Fromages (sauf à pâte pressée cuite)
7,0074	NOESY1D	0,17	0,005	0,09	Fromages (sauf à pâte pressée cuite)
7,1104	NOESY1D	0,15	0,01	0,1	Fromages (sauf à pâte pressée cuite)
7,7717	CPMG	0,16	0,008	0,4	Fromages (sauf à pâte pressée cuite)
7,7723	NOESY1D	0,17	0,005	0,09	Fromages (sauf à pâte pressée cuite)
0,9107	CPMG	-0,22	0,0003	0,05	Fromages (totaux)
0,9107	NOESY1D	-0,17	0,005	0,08	Fromages (totaux)
0,9697	NOESY1D	0,16	0,008	0,09	Fromages (totaux)
2,017	CPMG	-0,18	0,004	0,3	Fromages (totaux)
2,018	NOESY1D	-0,15	0,01	0,1	Fromages (totaux)
2,4707	NOESY1D	0,16	0,008	0,09	Fromages (totaux)
4,1506	NOESY1D	0,18	0,004	0,08	Fromages (totaux)
5,0465	NOESY1D	-0,16	0,009	0,09	Fromages (totaux)
5,0865	NOESY1D	-0,23	0,0001	0,02	Fromages (totaux)

## ANNEXE 14

5,1265	NOESY1D	-0,20	0,0008	0,06	Fromages (totaux)
5,3905	NOESY1D	-0,17	0,005	0,08	Fromages (totaux)
5,4305	NOESY1D	-0,20	0,001	0,06	Fromages (totaux)
5,4705	NOESY1D	-0,19	0,002	0,06	Fromages (totaux)
5,5105	NOESY1D	-0,15	0,01	0,1	Fromages (totaux)
5,5905	NOESY1D	-0,17	0,007	0,09	Fromages (totaux)
5,8638	CPMG	0,15	0,01	0,5	Fromages (totaux)
6,1184	NOESY1D	-0,17	0,005	0,08	Fromages (totaux)
7,0074	NOESY1D	0,17	0,005	0,08	Fromages (totaux)
7,1503	NOESY1D	0,19	0,002	0,07	Fromages (totaux)
7,159	CPMG	0,17	0,005	0,3	Fromages (totaux)
7,1833	NOESY1D	0,18	0,003	0,07	Fromages (totaux)
7,3673	NOESY1D	0,16	0,01	0,1	Fromages (totaux)
1,707	NOESY1D	-0,16	0,008	0,6	Fruits (totaux)
2,5797	NOESY1D	-0,16	0,008	0,6	Fruits (totaux)
3,587	CPMG	-0,18	0,004	0,3	Fruits (totaux)
3,616	CPMG	-0,19	0,002	0,3	Fruits (totaux)
1,707	NOESY1D	-0,15	0,01	0,4	Fruits à noyau
3,616	CPMG	-0,18	0,004	0,4	Fruits à noyau
0,4867	NOESY1D	-0,16	0,008	0,08	Fruits exotiques
0,5267	NOESY1D	-0,16	0,01	0,1	Fruits exotiques
1,6307	NOESY1D	-0,15	0,01	0,1	Fruits exotiques
1,6357	NOESY1D	-0,16	0,01	0,1	Fruits exotiques
1,849	CPMG	-0,21	0,0006	0,06	Fruits exotiques
2,5711	CPMG	-0,20	0,001	0,06	Fruits exotiques
5,8614	NOESY1D	0,16	0,01	0,1	Fruits exotiques
6,6064	NOESY1D	0,18	0,003	0,05	Fruits exotiques
6,6464	NOESY1D	0,18	0,003	0,05	Fruits exotiques
6,6864	NOESY1D	0,17	0,006	0,07	Fruits exotiques
6,7264	NOESY1D	0,17	0,005	0,07	Fruits exotiques
6,9054	NOESY1D	0,15	0,01	0,1	Fruits exotiques
6,968	CPMG	-0,16	0,009	0,2	Fruits exotiques
7,4303	NOESY1D	0,20	0,001	0,04	Fruits exotiques
7,4307	CPMG	0,17	0,005	0,1	Fruits exotiques
7,4703	NOESY1D	0,20	0,001	0,04	Fruits exotiques
7,4707	CPMG	0,18	0,003	0,08	Fruits exotiques
7,5077	CPMG	0,20	0,001	0,06	Fruits exotiques
7,5103	NOESY1D	0,22	0,0002	0,04	Fruits exotiques
7,6063	NOESY1D	0,19	0,002	0,05	Fruits exotiques
7,8303	NOESY1D	0,17	0,005	0,07	Fruits exotiques
7,8307	CPMG	0,18	0,003	0,08	Fruits exotiques
7,8703	NOESY1D	0,18	0,002	0,05	Fruits exotiques
7,8707	CPMG	0,20	0,001	0,06	Fruits exotiques
7,9083	NOESY1D	0,16	0,007	0,08	Fruits exotiques
7,9107	CPMG	0,19	0,002	0,08	Fruits exotiques
7,9173	NOESY1D	0,19	0,002	0,04	Fruits exotiques
7,9207	CPMG	0,15	0,01	0,2	Fruits exotiques
7,9263	NOESY1D	0,21	0,0005	0,04	Fruits exotiques
7,9663	NOESY1D	0,20	0,0009	0,04	Fruits exotiques
3,3666	NOESY1D	-0,18	0,003	0,1	Fruits secs (hors fruits à coque)
3,437	CPMG	-0,15	0,01	0,6	Fruits secs (hors fruits à coque)
3,4376	NOESY1D	-0,19	0,002	0,09	Fruits secs (hors fruits à coque)
3,4966	NOESY1D	-0,22	0,0004	0,04	Fruits secs (hors fruits à coque)
3,6106	NOESY1D	-0,17	0,004	0,1	Fruits secs (hors fruits à coque)
3,616	CPMG	-0,17	0,007	0,5	Fruits secs (hors fruits à coque)
3,7066	NOESY1D	-0,20	0,001	0,06	Fruits secs (hors fruits à coque)
3,713	CPMG	-0,16	0,009	0,5	Fruits secs (hors fruits à coque)
3,795	CPMG	-0,17	0,006	0,5	Fruits secs (hors fruits à coque)
3,8066	NOESY1D	-0,21	0,0005	0,04	Fruits secs (hors fruits à coque)
3,9516	NOESY1D	-0,16	0,009	0,2	Fruits secs (hors fruits à coque)
2,372	CPMG	0,17	0,005	0,2	Gâteaux et biscuits
2,4137	NOESY1D	-0,16	0,01	0,3	Gâteaux et biscuits
2,5507	NOESY1D	-0,18	0,003	0,3	Gâteaux et biscuits
3,3106	NOESY1D	-0,16	0,007	0,3	Gâteaux et biscuits
3,3666	NOESY1D	-0,17	0,005	0,3	Gâteaux et biscuits
3,367	CPMG	-0,23	0,0001	0,01	Gâteaux et biscuits
3,437	CPMG	-0,23	0,0001	0,01	Gâteaux et biscuits
3,4376	NOESY1D	-0,16	0,009	0,3	Gâteaux et biscuits
3,713	CPMG	-0,21	0,0007	0,03	Gâteaux et biscuits
3,795	CPMG	-0,21	0,0008	0,03	Gâteaux et biscuits
5,2069	CPMG	-0,17	0,007	0,2	Gâteaux et biscuits

## ANNEXE 14

7,054	CPMG	0,16	0,01	0,3	Gâteaux et biscuits
0,6067	NOESY1D	-0,16	0,01	0,2	Haricots verts
1,0811	CPMG	-0,16	0,008	0,5	Haricots verts
1,927	NOESY1D	-0,18	0,003	0,09	Haricots verts
1,927	CPMG	-0,18	0,003	0,3	Haricots verts
3,1666	NOESY1D	-0,21	0,0004	0,04	Haricots verts
3,167	CPMG	-0,20	0,001	0,2	Haricots verts
3,2696	NOESY1D	-0,17	0,006	0,2	Haricots verts
3,3136	NOESY1D	-0,21	0,0007	0,04	Haricots verts
3,6106	NOESY1D	-0,19	0,002	0,06	Haricots verts
3,918	NOESY1D	0,15	0,01	0,2	Haricots verts
3,9516	NOESY1D	-0,16	0,007	0,2	Haricots verts
4,3106	NOESY1D	-0,21	0,0005	0,04	Haricots verts
5,6339	CPMG	-0,15	0,01	1,0	Infusion, chicorée et boissons sucrées artificiellement
7,5663	NOESY1D	0,17	0,005	0,6	Infusion, chicorée et boissons sucrées artificiellement
2,1661	CPMG	-0,17	0,006	1,0	Jus de fruits
2,873	CPMG	-0,16	0,01	1,0	Jus de fruits
3,1666	NOESY1D	-0,16	0,007	0,4	Lait
3,269	CPMG	-0,17	0,007	0,7	Lait
3,2696	NOESY1D	-0,17	0,005	0,4	Lait
3,6106	NOESY1D	-0,18	0,004	0,4	Lait
3,9516	NOESY1D	-0,15	0,01	0,4	Lait
6,4468	CPMG	-0,15	0,01	0,7	Lait
3,496	CPMG	-0,18	0,004	0,2	Légumes (totaux)
3,587	CPMG	-0,22	0,0003	0,03	Légumes (totaux)
3,5876	NOESY1D	-0,17	0,004	0,4	Légumes (totaux)
3,6106	NOESY1D	-0,17	0,005	0,4	Légumes (totaux)
3,616	CPMG	-0,23	0,0001	0,02	Légumes (totaux)
3,86	CPMG	-0,21	0,0006	0,04	Légumes (totaux)
0,5671	CPMG	0,15	0,01	0,1	Légumes-feuilles
1,3911	CPMG	-0,18	0,003	0,06	Légumes-feuilles
1,4311	CPMG	-0,17	0,005	0,07	Légumes-feuilles
1,4371	CPMG	-0,18	0,004	0,07	Légumes-feuilles
1,6311	CPMG	-0,20	0,001	0,05	Légumes-feuilles
1,6361	CPMG	-0,21	0,0006	0,05	Légumes-feuilles
2,0631	CPMG	-0,20	0,001	0,05	Légumes-feuilles
3,616	CPMG	-0,16	0,008	0,09	Légumes-feuilles
6,4468	CPMG	0,16	0,007	0,09	Légumes-feuilles
6,5668	CPMG	0,19	0,002	0,06	Légumes-feuilles
6,6478	CPMG	0,18	0,003	0,06	Légumes-feuilles
6,7278	CPMG	0,19	0,002	0,06	Légumes-feuilles
7,0074	NOESY1D	0,16	0,007	0,4	Légumes-feuilles
7,0708	CPMG	0,16	0,01	0,1	Légumes-feuilles
7,1078	CPMG	0,17	0,005	0,07	Légumes-feuilles
7,1198	CPMG	0,20	0,0009	0,05	Légumes-feuilles
7,128	CPMG	0,17	0,005	0,07	Légumes-feuilles
7,4307	CPMG	0,19	0,002	0,06	Légumes-feuilles
7,5077	CPMG	0,17	0,007	0,08	Légumes-feuilles
1,0811	CPMG	0,17	0,007	0,6	Légumes-fruits
2,6267	NOESY1D	-0,21	0,0005	0,09	Légumes-fruits
3,86	CPMG	-0,21	0,0005	0,08	Légumes-fruits
3,8606	NOESY1D	-0,16	0,009	0,7	Légumes-fruits
1,6467	NOESY1D	-0,20	0,001	0,2	Légumineuses
1,6471	CPMG	-0,18	0,004	0,2	Légumineuses
2,0941	CPMG	0,16	0,01	0,3	Légumineuses
2,1511	CPMG	0,15	0,01	0,3	Légumineuses
2,5017	NOESY1D	-0,16	0,008	0,4	Légumineuses
2,5507	NOESY1D	-0,15	0,01	0,4	Légumineuses
2,9271	CPMG	-0,17	0,007	0,3	Légumineuses
5,2465	NOESY1D	0,15	0,01	0,4	Légumineuses
5,2469	CPMG	0,16	0,01	0,3	Légumineuses
7,2107	CPMG	-0,18	0,003	0,2	Légumineuses
7,6467	CPMG	-0,18	0,003	0,2	Légumineuses
1,0811	CPMG	-0,15	0,01	0,4	Matière grasse végétale (sauce huile d'olive et soja)
2,017	CPMG	0,16	0,008	0,4	Matière grasse végétale (sauce huile d'olive et soja)
3,8066	NOESY1D	-0,16	0,008	0,7	Matière grasse végétale (sauce huile d'olive et soja)
7,0708	CPMG	0,16	0,01	0,4	Matière grasse végétale (sauce huile d'olive et soja)
7,8307	CPMG	0,16	0,01	0,4	Matière grasse végétale (sauce huile d'olive et soja)
7,8707	CPMG	0,20	0,0009	0,2	Matière grasse végétale (sauce huile d'olive et soja)
3,6106	NOESY1D	-0,16	0,01	0,8	Matière grasse végétale (totale)
3,8066	NOESY1D	-0,16	0,008	0,8	Matière grasse végétale (totale)

## ANNEXE 14

7,8307	CPMG	0,17	0,007	0,5	Matière grasse végétale (totale)
7,8707	CPMG	0,22	0,0002	0,04	Matière grasse végétale (totale)
0,4867	NOESY1D	-0,26	<0,0001	0,01	mPNNs-GS
0,5267	NOESY1D	-0,18	0,006	0,2	mPNNs-GS
0,6067	NOESY1D	-0,15	0,02	0,3	mPNNs-GS
1,6467	NOESY1D	-0,18	0,007	0,2	mPNNs-GS
1,707	CPMG	0,16	0,02	0,2	mPNNs-GS
2,5711	CPMG	-0,27	<0,0001	0,01	mPNNs-GS
2,5797	NOESY1D	-0,19	0,005	0,2	mPNNs-GS
2,9671	CPMG	-0,20	0,003	0,1	mPNNs-GS
3,028	CPMG	-0,18	0,007	0,2	mPNNs-GS
3,111	CPMG	-0,16	0,02	0,2	mPNNs-GS
3,129	CPMG	-0,16	0,02	0,2	mPNNs-GS
3,7906	NOESY1D	-0,18	0,008	0,2	mPNNs-GS
4,184	CPMG	-0,17	0,01	0,2	mPNNs-GS
4,198	CPMG	-0,18	0,007	0,2	mPNNs-GS
4,311	CPMG	-0,17	0,01	0,2	mPNNs-GS
4,351	CPMG	-0,16	0,02	0,2	mPNNs-GS
4,39	CPMG	-0,16	0,02	0,2	mPNNs-GS
5,8614	NOESY1D	0,15	0,02	0,3	mPNNs-GS
6,2084	NOESY1D	-0,20	0,002	0,2	mPNNs-GS
6,5268	CPMG	-0,16	0,02	0,2	mPNNs-GS
6,6464	NOESY1D	0,16	0,02	0,3	mPNNs-GS
6,6864	NOESY1D	0,19	0,005	0,2	mPNNs-GS
6,7264	NOESY1D	0,18	0,006	0,2	mPNNs-GS
6,8634	NOESY1D	0,16	0,02	0,3	mPNNs-GS
7,128	CPMG	0,15	0,02	0,2	mPNNs-GS
7,159	CPMG	0,17	0,01	0,2	mPNNs-GS
7,2107	CPMG	-0,18	0,008	0,2	mPNNs-GS
7,2697	CPMG	-0,15	0,02	0,2	mPNNs-GS
7,6867	CPMG	-0,19	0,004	0,1	mPNNs-GS
7,8307	CPMG	0,19	0,004	0,1	mPNNs-GS
7,8707	CPMG	0,21	0,002	0,1	mPNNs-GS
7,9207	CPMG	0,18	0,006	0,2	mPNNs-GS
7,9663	NOESY1D	0,17	0,01	0,2	mPNNs-GS
1,145	CPMG	0,16	0,01	0,3	Noix/fruits à coque
1,927	NOESY1D	-0,16	0,01	0,4	Noix/fruits à coque
3,3666	NOESY1D	-0,15	0,01	0,4	Noix/fruits à coque
3,367	CPMG	-0,16	0,009	0,3	Noix/fruits à coque
3,437	CPMG	-0,16	0,009	0,3	Noix/fruits à coque
3,4376	NOESY1D	-0,16	0,01	0,4	Noix/fruits à coque
3,4966	NOESY1D	-0,16	0,01	0,4	Noix/fruits à coque
3,616	CPMG	-0,17	0,006	0,3	Noix/fruits à coque
3,713	CPMG	-0,16	0,009	0,3	Noix/fruits à coque
4,034	CPMG	0,16	0,01	0,3	Noix/fruits à coque
4,075	NOESY1D	-0,16	0,008	0,4	Noix/fruits à coque
7,159	CPMG	-0,15	0,01	0,3	Noix/fruits à coque
3,86	CPMG	0,16	0,009	0,6	Œufs
5,2069	CPMG	0,15	0,01	0,6	Œufs
1,021	NOESY1D	0,17	0,006	0,5	Olive
2,017	CPMG	-0,16	0,01	0,9	Olive
2,018	NOESY1D	-0,17	0,007	0,5	Olive
2,6121	CPMG	-0,18	0,003	0,5	Olive
0,6471	CPMG	-0,15	0,01	0,2	Pain
0,6867	NOESY1D	-0,15	0,01	0,2	Pain
0,7267	NOESY1D	-0,16	0,007	0,2	Pain
0,7667	NOESY1D	-0,15	0,01	0,2	Pain
1,0667	NOESY1D	-0,15	0,01	0,2	Pain
1,0887	NOESY1D	-0,16	0,008	0,2	Pain
1,1267	NOESY1D	-0,17	0,005	0,2	Pain
1,1667	NOESY1D	-0,16	0,007	0,2	Pain
1,1671	CPMG	-0,16	0,007	0,2	Pain
1,4287	NOESY1D	-0,16	0,01	0,2	Pain
1,8867	NOESY1D	-0,17	0,005	0,2	Pain
2,411	CPMG	0,18	0,003	0,1	Pain
3,312	CPMG	0,17	0,005	0,2	Pain
3,3136	NOESY1D	0,18	0,003	0,2	Pain
3,951	CPMG	0,20	0,001	0,1	Pain
6,1328	CPMG	0,20	0,001	0,1	Pain
6,1698	CPMG	0,18	0,003	0,1	Pain
6,3274	NOESY1D	0,15	0,01	0,2	Pain

## ANNEXE 14

7,3503	NOESY1D	0,15	0,01	0,2	Pain
7,5663	NOESY1D	0,16	0,01	0,2	Pain
0,9997	NOESY1D	0,15	0,01	0,2	Pâtes et céréales
1,6467	NOESY1D	-0,15	0,01	0,2	Pâtes et céréales
1,8211	CPMG	0,20	0,001	0,2	Pâtes et céréales
2,4707	NOESY1D	-0,18	0,003	0,1	Pâtes et céréales
3,0706	NOESY1D	-0,16	0,01	0,2	Pâtes et céréales
3,918	NOESY1D	0,21	0,0006	0,09	Pâtes et céréales
5,6385	NOESY1D	0,17	0,005	0,2	Pâtes et céréales
5,9688	CPMG	-0,17	0,004	0,4	Pâtes et céréales
6,1104	NOESY1D	0,19	0,002	0,09	Pâtes et céréales
6,1184	NOESY1D	0,19	0,002	0,09	Pâtes et céréales
7,1198	CPMG	-0,15	0,01	0,5	Pâtes et céréales
1,145	CPMG	0,17	0,007	0,3	Pâtes, riz et céréales (total)
1,6471	CPMG	-0,16	0,01	0,3	Pâtes, riz et céréales (total)
1,8211	CPMG	0,18	0,004	0,3	Pâtes, riz et céréales (total)
3,311	CPMG	-0,15	0,01	0,3	Pâtes, riz et céréales (total)
3,3656	NOESY1D	-0,16	0,009	0,3	Pâtes, riz et céréales (total)
3,3666	NOESY1D	-0,16	0,009	0,3	Pâtes, riz et céréales (total)
3,367	CPMG	-0,16	0,01	0,3	Pâtes, riz et céréales (total)
3,437	CPMG	-0,17	0,006	0,3	Pâtes, riz et céréales (total)
3,4376	NOESY1D	-0,19	0,002	0,1	Pâtes, riz et céréales (total)
3,918	NOESY1D	0,23	0,0001	0,02	Pâtes, riz et céréales (total)
6,1184	NOESY1D	0,19	0,002	0,1	Pâtes, riz et céréales (total)
7,1198	CPMG	-0,16	0,01	0,3	Pâtes, riz et céréales (total)
0,3181	CPMG	0,16	0,01	0,3	Pâtisseries et aliments sucrés à base de céréales (total)
2,372	CPMG	0,19	0,001	0,05	Pâtisseries et aliments sucrés à base de céréales (total)
2,374	NOESY1D	0,17	0,006	0,2	Pâtisseries et aliments sucrés à base de céréales (total)
2,5507	NOESY1D	-0,18	0,004	0,2	Pâtisseries et aliments sucrés à base de céréales (total)
3,3106	NOESY1D	-0,17	0,005	0,2	Pâtisseries et aliments sucrés à base de céréales (total)
3,3666	NOESY1D	-0,19	0,002	0,2	Pâtisseries et aliments sucrés à base de céréales (total)
3,367	CPMG	-0,24	<0,0001	0,007	Pâtisseries et aliments sucrés à base de céréales (total)
3,437	CPMG	-0,24	<0,0001	0,007	Pâtisseries et aliments sucrés à base de céréales (total)
3,4376	NOESY1D	-0,18	0,003	0,2	Pâtisseries et aliments sucrés à base de céréales (total)
3,4966	NOESY1D	-0,15	0,01	0,3	Pâtisseries et aliments sucrés à base de céréales (total)
3,713	CPMG	-0,22	0,0003	0,02	Pâtisseries et aliments sucrés à base de céréales (total)
3,795	CPMG	-0,22	0,0004	0,02	Pâtisseries et aliments sucrés à base de céréales (total)
4,055	CPMG	0,15	0,01	0,3	Pâtisseries et aliments sucrés à base de céréales (total)
4,075	NOESY1D	-0,15	0,01	0,3	Pâtisseries et aliments sucrés à base de céréales (total)
5,2069	CPMG	-0,16	0,008	0,2	Pâtisseries et aliments sucrés à base de céréales (total)
7,054	CPMG	0,15	0,01	0,3	Pâtisseries et aliments sucrés à base de céréales (total)
0,6471	CPMG	-0,22	0,0004	0,03	Poissons gras
0,6867	NOESY1D	-0,24	<0,0001	0,007	Poissons gras
0,7131	CPMG	-0,19	0,002	0,04	Poissons gras
0,7267	NOESY1D	-0,22	0,0003	0,01	Poissons gras
0,7271	CPMG	-0,20	0,001	0,04	Poissons gras
0,7667	NOESY1D	-0,16	0,01	0,07	Poissons gras
0,9697	NOESY1D	0,15	0,01	0,07	Poissons gras
0,9997	NOESY1D	-0,27	<0,0001	0,002	Poissons gras
0,9997	CPMG	-0,20	0,001	0,04	Poissons gras
1,051	NOESY1D	-0,20	0,001	0,02	Poissons gras
1,051	CPMG	-0,16	0,009	0,1	Poissons gras
1,0531	CPMG	-0,20	0,001	0,04	Poissons gras
1,0811	CPMG	-0,22	0,0003	0,03	Poissons gras
1,0881	CPMG	-0,19	0,001	0,04	Poissons gras
1,3507	NOESY1D	-0,19	0,002	0,03	Poissons gras
1,3511	CPMG	-0,16	0,01	0,1	Poissons gras
1,3907	NOESY1D	-0,17	0,005	0,04	Poissons gras
1,4287	NOESY1D	-0,17	0,005	0,04	Poissons gras
1,5217	NOESY1D	-0,17	0,004	0,04	Poissons gras
1,5907	NOESY1D	-0,20	0,001	0,02	Poissons gras
1,6307	NOESY1D	-0,22	0,0003	0,01	Poissons gras
1,6357	NOESY1D	-0,21	0,0006	0,01	Poissons gras
1,7507	NOESY1D	-0,15	0,01	0,07	Poissons gras
1,7907	NOESY1D	-0,18	0,003	0,03	Poissons gras
1,8307	NOESY1D	-0,20	0,001	0,02	Poissons gras
1,8507	NOESY1D	-0,21	0,0004	0,01	Poissons gras
1,8867	NOESY1D	-0,19	0,002	0,02	Poissons gras
2,1507	NOESY1D	-0,18	0,003	0,03	Poissons gras
2,1737	NOESY1D	-0,16	0,009	0,06	Poissons gras
2,354	CPMG	0,15	0,01	0,2	Poissons gras

## ANNEXE 14

2,4711	CPMG	0,17	0,006	0,1	Poissons gras
2,5797	NOESY1D	-0,18	0,003	0,03	Poissons gras
2,8467	NOESY1D	-0,16	0,007	0,05	Poissons gras
3,111	CPMG	0,20	0,0009	0,04	Poissons gras
3,2696	NOESY1D	-0,20	0,001	0,02	Poissons gras
3,3106	NOESY1D	-0,22	0,0003	0,01	Poissons gras
3,587	CPMG	0,16	0,01	0,1	Poissons gras
3,918	NOESY1D	-0,15	0,01	0,07	Poissons gras
4,021	CPMG	0,17	0,006	0,1	Poissons gras
6,2068	CPMG	-0,16	0,01	0,1	Poissons gras
6,2084	NOESY1D	-0,18	0,004	0,03	Poissons gras
6,2464	NOESY1D	-0,15	0,01	0,07	Poissons gras
6,912	CPMG	0,17	0,006	0,1	Poissons gras
7,5663	NOESY1D	0,18	0,003	0,03	Poissons gras
3,6106	NOESY1D	-0,16	0,01	0,6	Poissons maigres
3,7906	NOESY1D	-0,17	0,004	0,6	Poissons maigres
3,9516	NOESY1D	-0,15	0,01	0,6	Poissons maigres
1,5211	CPMG	0,16	0,007	0,3	Pomme et poire
2,0941	CPMG	0,15	0,01	0,4	Pomme et poire
2,1511	CPMG	0,16	0,007	0,3	Pomme et poire
2,5797	NOESY1D	-0,16	0,01	0,8	Pomme et poire
3,028	CPMG	-0,17	0,005	0,3	Pomme et poire
3,587	CPMG	-0,15	0,01	0,4	Pomme et poire
6,968	CPMG	-0,17	0,006	0,3	Pomme et poire
2,018	NOESY1D	0,16	0,008	1,0	Pommes de terre
3,616	CPMG	0,15	0,01	0,7	Pommes de terre
1,707	NOESY1D	0,15	0,01	0,8	Pommes de terre, racines et tubercules
2,017	CPMG	0,16	0,01	0,5	Pommes de terre, racines et tubercules
2,018	NOESY1D	0,17	0,004	0,6	Pommes de terre, racines et tubercules
3,496	CPMG	0,16	0,01	0,5	Pommes de terre, racines et tubercules
3,587	CPMG	0,16	0,01	0,5	Pommes de terre, racines et tubercules
3,616	CPMG	0,20	0,001	0,2	Pommes de terre, racines et tubercules
0,8071	CPMG	0,15	0,01	0,2	Produits à base de céréales complètes
1,1271	CPMG	0,16	0,01	0,2	Produits à base de céréales complètes
1,5211	CPMG	0,15	0,01	0,2	Produits à base de céréales complètes
1,677	NOESY1D	-0,18	0,003	0,3	Produits à base de céréales complètes
2,057	NOESY1D	-0,15	0,01	0,4	Produits à base de céréales complètes
2,9671	CPMG	-0,17	0,005	0,2	Produits à base de céréales complètes
3,496	CPMG	-0,17	0,006	0,2	Produits à base de céréales complètes
3,7906	NOESY1D	-0,18	0,004	0,3	Produits à base de céréales complètes
3,791	CPMG	-0,16	0,008	0,2	Produits à base de céréales complètes
3,9516	NOESY1D	-0,17	0,005	0,3	Produits à base de céréales complètes
4,055	CPMG	0,17	0,006	0,2	Produits à base de céréales complètes
4,3506	NOESY1D	-0,15	0,01	0,4	Produits à base de céréales complètes
5,1669	CPMG	0,15	0,01	0,2	Produits à base de céréales complètes
5,2469	CPMG	0,17	0,005	0,2	Produits à base de céréales complètes
6,2068	CPMG	0,19	0,002	0,2	Produits à base de céréales complètes
7,2107	CPMG	-0,15	0,01	0,2	Produits à base de céréales complètes
0,9107	CPMG	0,20	0,001	0,06	Produits à base de sucre/sucreries (miel, bonbons, confitures, etc.)
0,9107	NOESY1D	0,17	0,006	0,07	Produits à base de sucre/sucreries (miel, bonbons, confitures, etc.)
1,1267	NOESY1D	-0,16	0,008	0,07	Produits à base de sucre/sucreries (miel, bonbons, confitures, etc.)
1,1667	NOESY1D	-0,23	0,0001	0,004	Produits à base de sucre/sucreries (miel, bonbons, confitures, etc.)
1,1671	CPMG	-0,15	0,01	0,4	Produits à base de sucre/sucreries (miel, bonbons, confitures, etc.)
1,4287	NOESY1D	-0,16	0,009	0,08	Produits à base de sucre/sucreries (miel, bonbons, confitures, etc.)
1,4467	NOESY1D	-0,18	0,003	0,05	Produits à base de sucre/sucreries (miel, bonbons, confitures, etc.)
1,7507	NOESY1D	-0,15	0,01	0,1	Produits à base de sucre/sucreries (miel, bonbons, confitures, etc.)
1,7907	NOESY1D	-0,17	0,004	0,06	Produits à base de sucre/sucreries (miel, bonbons, confitures, etc.)
1,8307	NOESY1D	-0,16	0,007	0,07	Produits à base de sucre/sucreries (miel, bonbons, confitures, etc.)
1,8867	NOESY1D	-0,16	0,009	0,08	Produits à base de sucre/sucreries (miel, bonbons, confitures, etc.)
2,1507	NOESY1D	-0,18	0,003	0,05	Produits à base de sucre/sucreries (miel, bonbons, confitures, etc.)
2,2637	NOESY1D	-0,20	0,001	0,02	Produits à base de sucre/sucreries (miel, bonbons, confitures, etc.)
2,3107	NOESY1D	-0,17	0,005	0,06	Produits à base de sucre/sucreries (miel, bonbons, confitures, etc.)
3,1666	NOESY1D	-0,19	0,001	0,03	Produits à base de sucre/sucreries (miel, bonbons, confitures, etc.)
3,3666	NOESY1D	-0,27	<0,0001	0,0006	Produits à base de sucre/sucreries (miel, bonbons, confitures, etc.)
3,367	CPMG	-0,20	0,001	0,06	Produits à base de sucre/sucreries (miel, bonbons, confitures, etc.)
3,437	CPMG	-0,20	0,001	0,06	Produits à base de sucre/sucreries (miel, bonbons, confitures, etc.)
3,4376	NOESY1D	-0,25	<0,0001	0,002	Produits à base de sucre/sucreries (miel, bonbons, confitures, etc.)
3,4966	NOESY1D	-0,24	<0,0001	0,002	Produits à base de sucre/sucreries (miel, bonbons, confitures, etc.)
3,7066	NOESY1D	-0,17	0,006	0,07	Produits à base de sucre/sucreries (miel, bonbons, confitures, etc.)
3,8066	NOESY1D	-0,19	0,001	0,03	Produits à base de sucre/sucreries (miel, bonbons, confitures, etc.)
5,2065	NOESY1D	-0,28	<0,0001	0,0006	Produits à base de sucre/sucreries (miel, bonbons, confitures, etc.)



## ANNEXE 14

5,2069	CPMG	-0,25	<0,0001	0,005	Produits à base de sucre/sucreries (miel, bonbons, confitures, etc.)
5,5905	NOESY1D	0,15	0,01	0,1	Produits à base de sucre/sucreries (miel, bonbons, confitures, etc.)
5,9664	NOESY1D	0,22	0,0003	0,008	Produits à base de sucre/sucreries (miel, bonbons, confitures, etc.)
0,9997	CPMG	-0,17	0,007	0,4	Produits de la mer (totaux)
0,9997	NOESY1D	-0,17	0,005	0,8	Produits de la mer (totaux)
3,071	CPMG	0,17	0,006	0,4	Produits de la mer (totaux)
3,111	CPMG	0,18	0,003	0,4	Produits de la mer (totaux)
0,4867	NOESY1D	0,20	0,001	0,01	Produits salés (totaux)
0,5267	NOESY1D	0,23	0,0002	0,006	Produits salés (totaux)
0,5667	NOESY1D	0,22	0,0002	0,006	Produits salés (totaux)
0,6067	NOESY1D	0,20	0,001	0,01	Produits salés (totaux)
0,6467	NOESY1D	0,21	0,0006	0,01	Produits salés (totaux)
0,6867	NOESY1D	0,18	0,003	0,02	Produits salés (totaux)
0,7267	NOESY1D	0,17	0,004	0,02	Produits salés (totaux)
0,8067	NOESY1D	0,18	0,003	0,02	Produits salés (totaux)
0,8071	CPMG	0,15	0,01	0,4	Produits salés (totaux)
1,0667	NOESY1D	0,23	0,0001	0,006	Produits salés (totaux)
1,0887	NOESY1D	0,23	0,0002	0,006	Produits salés (totaux)
1,1267	NOESY1D	0,21	0,0005	0,009	Produits salés (totaux)
1,1667	NOESY1D	0,16	0,008	0,03	Produits salés (totaux)
1,3907	NOESY1D	0,20	0,001	0,01	Produits salés (totaux)
1,4287	NOESY1D	0,20	0,0008	0,01	Produits salés (totaux)
1,4467	NOESY1D	0,20	0,0008	0,01	Produits salés (totaux)
2,017	CPMG	-0,17	0,007	0,4	Produits salés (totaux)
3,1666	NOESY1D	0,18	0,003	0,02	Produits salés (totaux)
3,2696	NOESY1D	0,19	0,001	0,01	Produits salés (totaux)
5,0865	NOESY1D	0,15	0,01	0,04	Produits salés (totaux)
5,1265	NOESY1D	0,18	0,004	0,02	Produits salés (totaux)
5,3505	NOESY1D	0,17	0,004	0,02	Produits salés (totaux)
5,3905	NOESY1D	0,24	<0,0001	0,006	Produits salés (totaux)
5,4305	NOESY1D	0,22	0,0003	0,006	Produits salés (totaux)
5,4705	NOESY1D	0,17	0,005	0,02	Produits salés (totaux)
5,5105	NOESY1D	0,16	0,009	0,03	Produits salés (totaux)
5,6385	NOESY1D	0,17	0,007	0,03	Produits salés (totaux)
5,6869	CPMG	-0,16	0,01	0,4	Produits salés (totaux)
5,7269	CPMG	-0,16	0,009	0,4	Produits salés (totaux)
6,2864	NOESY1D	-0,18	0,003	0,02	Produits salés (totaux)
6,3274	NOESY1D	-0,17	0,005	0,02	Produits salés (totaux)
6,4064	NOESY1D	-0,16	0,007	0,03	Produits salés (totaux)
6,6464	NOESY1D	-0,15	0,01	0,04	Produits salés (totaux)
6,6864	NOESY1D	-0,17	0,005	0,02	Produits salés (totaux)
6,7264	NOESY1D	-0,19	0,001	0,01	Produits salés (totaux)
6,7664	NOESY1D	-0,18	0,004	0,02	Produits salés (totaux)
6,8634	NOESY1D	-0,15	0,01	0,04	Produits salés (totaux)
6,9054	NOESY1D	-0,17	0,006	0,02	Produits salés (totaux)
7,0074	NOESY1D	-0,20	0,001	0,01	Produits salés (totaux)
7,0708	CPMG	-0,16	0,01	0,4	Produits salés (totaux)
7,1104	NOESY1D	-0,18	0,004	0,02	Produits salés (totaux)
7,128	CPMG	-0,17	0,006	0,4	Produits salés (totaux)
7,1833	NOESY1D	-0,17	0,004	0,02	Produits salés (totaux)
7,2093	NOESY1D	-0,16	0,009	0,03	Produits salés (totaux)
7,2763	NOESY1D	-0,16	0,008	0,03	Produits salés (totaux)
7,3503	NOESY1D	-0,16	0,009	0,03	Produits salés (totaux)
7,3673	NOESY1D	-0,15	0,01	0,04	Produits salés (totaux)
7,4303	NOESY1D	-0,21	0,0006	0,01	Produits salés (totaux)
7,4703	NOESY1D	-0,16	0,007	0,03	Produits salés (totaux)
7,5103	NOESY1D	-0,17	0,005	0,02	Produits salés (totaux)
7,5663	NOESY1D	-0,22	0,0002	0,006	Produits salés (totaux)
7,6063	NOESY1D	-0,17	0,005	0,02	Produits salés (totaux)
7,6463	NOESY1D	-0,19	0,002	0,01	Produits salés (totaux)
7,7723	NOESY1D	-0,18	0,002	0,02	Produits salés (totaux)
7,9263	NOESY1D	-0,15	0,01	0,04	Produits salés (totaux)
7,9663	NOESY1D	-0,18	0,004	0,02	Produits salés (totaux)
0,4867	NOESY1D	0,17	0,005	0,05	Produits salés appétitifs : oeufs de poisson, chips, biscuits salés et amuse-gueule à la viande
0,5267	NOESY1D	0,19	0,002	0,04	Produits salés appétitifs : oeufs de poisson, chips, biscuits salés et amuse-gueule à la viande
0,5667	NOESY1D	0,17	0,005	0,05	Produits salés appétitifs : oeufs de poisson, chips, biscuits salés et amuse-gueule à la viande
0,6467	NOESY1D	0,15	0,01	0,06	Produits salés appétitifs : oeufs de poisson, chips, biscuits salés et amuse-gueule à la viande

## ANNEXE 14

0,6867	NOESY1D	0,19	0,002	0,04	Produits salés appétitifs : oeufs de poisson, chips, biscuits salés et amuse-gueule à la viande
0,7267	NOESY1D	0,17	0,005	0,05	Produits salés appétitifs : oeufs de poisson, chips, biscuits salés et amuse-gueule à la viande
0,8067	NOESY1D	0,15	0,01	0,06	Produits salés appétitifs : oeufs de poisson, chips, biscuits salés et amuse-gueule à la viande
1,0667	NOESY1D	0,17	0,006	0,05	Produits salés appétitifs : oeufs de poisson, chips, biscuits salés et amuse-gueule à la viande
1,0887	NOESY1D	0,17	0,006	0,05	Produits salés appétitifs : oeufs de poisson, chips, biscuits salés et amuse-gueule à la viande
1,1267	NOESY1D	0,17	0,006	0,05	Produits salés appétitifs : oeufs de poisson, chips, biscuits salés et amuse-gueule à la viande
1,3907	NOESY1D	0,18	0,003	0,04	Produits salés appétitifs : oeufs de poisson, chips, biscuits salés et amuse-gueule à la viande
1,4287	NOESY1D	0,18	0,004	0,05	Produits salés appétitifs : oeufs de poisson, chips, biscuits salés et amuse-gueule à la viande
2,9777	NOESY1D	-0,16	0,01	0,06	Produits salés appétitifs : oeufs de poisson, chips, biscuits salés et amuse-gueule à la viande
5,0865	NOESY1D	0,16	0,008	0,05	Produits salés appétitifs : oeufs de poisson, chips, biscuits salés et amuse-gueule à la viande
5,1265	NOESY1D	0,19	0,002	0,04	Produits salés appétitifs : oeufs de poisson, chips, biscuits salés et amuse-gueule à la viande
5,3505	NOESY1D	0,19	0,002	0,04	Produits salés appétitifs : oeufs de poisson, chips, biscuits salés et amuse-gueule à la viande
5,3905	NOESY1D	0,22	0,0002	0,03	Produits salés appétitifs : oeufs de poisson, chips, biscuits salés et amuse-gueule à la viande
5,4305	NOESY1D	0,22	0,0003	0,03	Produits salés appétitifs : oeufs de poisson, chips, biscuits salés et amuse-gueule à la viande
5,4705	NOESY1D	0,18	0,003	0,04	Produits salés appétitifs : oeufs de poisson, chips, biscuits salés et amuse-gueule à la viande
5,5105	NOESY1D	0,17	0,006	0,05	Produits salés appétitifs : oeufs de poisson, chips, biscuits salés et amuse-gueule à la viande
5,6385	NOESY1D	0,16	0,009	0,05	Produits salés appétitifs : oeufs de poisson, chips, biscuits salés et amuse-gueule à la viande
6,7264	NOESY1D	-0,16	0,007	0,05	Produits salés appétitifs : oeufs de poisson, chips, biscuits salés et amuse-gueule à la viande
6,7664	NOESY1D	-0,16	0,007	0,05	Produits salés appétitifs : oeufs de poisson, chips, biscuits salés et amuse-gueule à la viande
6,8634	NOESY1D	-0,15	0,01	0,06	Produits salés appétitifs : oeufs de poisson, chips, biscuits salés et amuse-gueule à la viande
7,0074	NOESY1D	-0,18	0,002	0,04	Produits salés appétitifs : oeufs de poisson, chips, biscuits salés et amuse-gueule à la viande
7,1104	NOESY1D	-0,15	0,01	0,06	Produits salés appétitifs : oeufs de poisson, chips, biscuits salés et amuse-gueule à la viande
7,128	CPMG	-0,16	0,01	0,7	Produits salés appétitifs : oeufs de poisson, chips, biscuits salés et amuse-gueule à la viande
7,1833	NOESY1D	-0,17	0,004	0,05	Produits salés appétitifs : oeufs de poisson, chips, biscuits salés et amuse-gueule à la viande
7,3503	NOESY1D	-0,15	0,01	0,06	Produits salés appétitifs : oeufs de poisson, chips, biscuits salés et amuse-gueule à la viande
7,3673	NOESY1D	-0,15	0,01	0,06	Produits salés appétitifs : oeufs de poisson, chips, biscuits salés et amuse-gueule à la viande
7,4303	NOESY1D	-0,19	0,002	0,04	Produits salés appétitifs : oeufs de poisson, chips, biscuits salés et amuse-gueule à la viande
7,5663	NOESY1D	-0,19	0,002	0,04	Produits salés appétitifs : oeufs de poisson, chips, biscuits salés et amuse-gueule à la viande
7,6463	NOESY1D	-0,19	0,002	0,04	Produits salés appétitifs : oeufs de poisson, chips, biscuits salés et amuse-gueule à la viande
7,7723	NOESY1D	-0,16	0,009	0,06	Produits salés appétitifs : oeufs de poisson, chips, biscuits salés et amuse-gueule à la viande
0,9107	CPMG	0,16	0,01	0,9	Produits sucrés (totaux)
3,496	CPMG	-0,16	0,009	0,5	Racines et tubercules
3,5316	NOESY1D	-0,16	0,01	0,6	Racines et tubercules
3,587	CPMG	-0,19	0,002	0,2	Racines et tubercules
3,5876	NOESY1D	-0,16	0,007	0,6	Racines et tubercules
3,6106	NOESY1D	-0,16	0,01	0,6	Racines et tubercules
3,616	CPMG	-0,19	0,002	0,2	Racines et tubercules
3,028	CPMG	-0,16	0,01	0,4	Raisin
3,269	CPMG	0,15	0,01	0,4	Raisin
3,2696	NOESY1D	0,19	0,002	0,2	Raisin
5,8888	CPMG	-0,19	0,001	0,3	Raisin

## ANNEXE 14

6,1664	NOESY1D	-0,19	0,002	0,2	Raisin
7,2787	CPMG	-0,17	0,006	0,4	Raisin
3,312	CPMG	0,18	0,003	0,1	Profil alimentaire "occidental"
3,496	CPMG	0,21	0,0006	0,04	Profil alimentaire "occidental"
3,587	CPMG	0,24	<0,0001	0,007	Profil alimentaire "occidental"
3,5876	NOESY1D	0,15	0,01	0,5	Profil alimentaire "occidental"
3,6106	NOESY1D	0,18	0,003	0,5	Profil alimentaire "occidental"
3,616	CPMG	0,24	<0,0001	0,007	Profil alimentaire "occidental"
3,7906	NOESY1D	0,17	0,006	0,5	Profil alimentaire "occidental"
2,0631	CPMG	-0,16	0,01	0,1	Profil alimentaire "sain"
2,5711	CPMG	-0,16	0,009	0,1	Profil alimentaire "sain"
3,367	CPMG	-0,17	0,005	0,1	Profil alimentaire "sain"
3,437	CPMG	-0,18	0,003	0,1	Profil alimentaire "sain"
3,4376	NOESY1D	-0,16	0,01	0,2	Profil alimentaire "sain"
3,496	CPMG	-0,25	<0,0001	0,002	Profil alimentaire "sain"
3,4966	NOESY1D	-0,17	0,004	0,07	Profil alimentaire "sain"
3,5316	NOESY1D	-0,20	0,001	0,02	Profil alimentaire "sain"
3,587	CPMG	-0,26	<0,0001	0,002	Profil alimentaire "sain"
3,5876	NOESY1D	-0,23	0,0001	0,007	Profil alimentaire "sain"
3,6106	NOESY1D	-0,26	<0,0001	0,003	Profil alimentaire "sain"
3,616	CPMG	-0,28	<0,0001	0,0006	Profil alimentaire "sain"
3,7066	NOESY1D	-0,20	0,001	0,03	Profil alimentaire "sain"
3,713	CPMG	-0,16	0,007	0,1	Profil alimentaire "sain"
3,7906	NOESY1D	-0,21	0,0004	0,02	Profil alimentaire "sain"
3,7926	NOESY1D	-0,20	0,001	0,02	Profil alimentaire "sain"
3,795	CPMG	-0,20	0,0009	0,03	Profil alimentaire "sain"
3,8066	NOESY1D	-0,23	0,0001	0,007	Profil alimentaire "sain"
3,86	CPMG	-0,22	0,0003	0,01	Profil alimentaire "sain"
3,9516	NOESY1D	-0,20	0,0009	0,02	Profil alimentaire "sain"
4,1906	NOESY1D	-0,16	0,008	0,1	Profil alimentaire "sain"
7,8307	CPMG	0,16	0,009	0,1	Profil alimentaire "sain"
7,8707	CPMG	0,17	0,006	0,1	Profil alimentaire "sain"
7,9107	CPMG	0,16	0,01	0,1	Profil alimentaire "sain"
3,028	CPMG	-0,15	0,01	0,4	Rhubarbe
3,3666	NOESY1D	-0,19	0,002	0,1	Rhubarbe
3,367	CPMG	-0,16	0,008	0,3	Rhubarbe
3,437	CPMG	-0,19	0,002	0,2	Rhubarbe
3,4376	NOESY1D	-0,20	0,001	0,1	Rhubarbe
3,4966	NOESY1D	-0,18	0,003	0,1	Rhubarbe
3,713	CPMG	-0,18	0,003	0,2	Rhubarbe
3,795	CPMG	-0,19	0,001	0,2	Rhubarbe
3,8066	NOESY1D	-0,18	0,003	0,1	Rhubarbe
3,3666	NOESY1D	-0,16	0,01	0,4	Riz
3,4376	NOESY1D	-0,15	0,01	0,4	Riz
3,4966	NOESY1D	-0,17	0,005	0,4	Riz
3,8066	NOESY1D	-0,15	0,01	0,4	Riz
7,2763	NOESY1D	0,17	0,004	0,4	Riz
2,017	CPMG	0,20	0,001	0,2	Sauces
2,018	NOESY1D	0,15	0,01	0,5	Sauces
2,6471	CPMG	0,17	0,005	0,4	Sauces
3,587	CPMG	0,15	0,01	0,7	Sauces
1,145	CPMG	0,17	0,007	0,7	Thé
0,9997	CPMG	-0,16	0,009	0,2	Viande blanche
1,021	CPMG	0,16	0,007	0,2	Viande blanche
1,051	CPMG	-0,17	0,005	0,2	Viande blanche
1,051	NOESY1D	-0,17	0,004	0,6	Viande blanche
1,3511	CPMG	-0,16	0,01	0,2	Viande blanche
1,3911	CPMG	-0,19	0,002	0,2	Viande blanche
1,4311	CPMG	-0,17	0,006	0,2	Viande blanche
1,5211	CPMG	-0,16	0,008	0,2	Viande blanche
1,707	NOESY1D	0,16	0,01	0,8	Viande blanche
2,2641	CPMG	0,17	0,007	0,2	Viande blanche
3,587	CPMG	0,16	0,01	0,8	Viande rouge
3,5876	NOESY1D	0,15	0,01	0,8	Viande rouge
1,8211	CPMG	-0,16	0,01	0,7	Viandes rouges et abats
2,98	CPMG	-0,16	0,01	0,7	Viandes rouges et abats
3,028	CPMG	0,16	0,01	0,7	Viandes rouges et abats
0,9107	NOESY1D	-0,17	0,006	0,07	Vin rouge
1,0667	NOESY1D	0,20	0,001	0,04	Vin rouge
1,0887	NOESY1D	0,17	0,005	0,06	Vin rouge
1,1267	NOESY1D	0,17	0,005	0,06	Vin rouge

ANNEXE 14

1,1271	CPMG	0,24	0,0001	0,006	Vin rouge
1,145	CPMG	-0,16	0,008	0,2	Vin rouge
1,1667	NOESY1D	0,20	0,0009	0,04	Vin rouge
1,4287	NOESY1D	0,18	0,003	0,05	Vin rouge
1,4467	NOESY1D	0,18	0,003	0,05	Vin rouge
1,6307	NOESY1D	0,19	0,002	0,05	Vin rouge
1,6357	NOESY1D	0,18	0,003	0,05	Vin rouge
1,707	CPMG	-0,17	0,007	0,2	Vin rouge
1,7507	NOESY1D	0,21	0,0006	0,04	Vin rouge
1,7907	NOESY1D	0,18	0,002	0,05	Vin rouge
1,8307	NOESY1D	0,15	0,01	0,09	Vin rouge
1,899	CPMG	-0,24	<0,0001	0,006	Vin rouge
1,927	NOESY1D	0,15	0,01	0,09	Vin rouge
2,0941	CPMG	-0,25	<0,0001	0,006	Vin rouge
2,2637	NOESY1D	0,18	0,003	0,05	Vin rouge
2,3107	NOESY1D	0,21	0,0005	0,04	Vin rouge
2,411	CPMG	-0,17	0,007	0,2	Vin rouge
2,5507	NOESY1D	0,16	0,008	0,08	Vin rouge
2,6471	CPMG	-0,16	0,009	0,2	Vin rouge
2,8481	CPMG	-0,16	0,01	0,2	Vin rouge
3,1666	NOESY1D	0,20	0,001	0,04	Vin rouge
3,3106	NOESY1D	0,16	0,009	0,09	Vin rouge
5,2065	NOESY1D	0,17	0,006	0,07	Vin rouge
6,8634	NOESY1D	-0,15	0,01	0,09	Vin rouge
7,2787	CPMG	-0,19	0,002	0,1	Vin rouge
7,3437	CPMG	-0,16	0,009	0,2	Vin rouge
0,6467	NOESY1D	0,17	0,006	0,2	Vins blancs et rosés
1,0667	NOESY1D	0,16	0,008	0,2	Vins blancs et rosés
1,0887	NOESY1D	0,17	0,006	0,2	Vins blancs et rosés
1,1271	CPMG	0,18	0,003	0,4	Vins blancs et rosés
1,707	CPMG	-0,17	0,005	0,4	Vins blancs et rosés
1,759	CPMG	0,15	0,01	0,4	Vins blancs et rosés
1,899	CPMG	-0,16	0,008	0,4	Vins blancs et rosés
2,9777	NOESY1D	-0,16	0,01	0,2	Vins blancs et rosés
3,1666	NOESY1D	0,19	0,002	0,2	Vins blancs et rosés
3,918	NOESY1D	-0,19	0,002	0,2	Vins blancs et rosés
5,2065	NOESY1D	0,16	0,008	0,2	Vins blancs et rosés
5,3905	NOESY1D	0,15	0,01	0,2	Vins blancs et rosés
6,8634	NOESY1D	-0,17	0,005	0,2	Vins blancs et rosés
7,1833	NOESY1D	-0,15	0,01	0,2	Vins blancs et rosés
7,2093	NOESY1D	-0,17	0,005	0,2	Vins blancs et rosés
7,2703	NOESY1D	-0,16	0,01	0,2	Vins blancs et rosés
2,715	CPMG	0,16	0,009	0,7	Yaourts

R; correlation coefficient

<sup>1</sup> Les corrélations de Spearman ont été ajustées sur les caractéristiques suivantes : âge, IMC, statut tabagique, saison de prélèvement sanguin, nombre d'enregistrements alimentaires sur 24 heures et moyenne de l'apport énergétique quotidien pendant les 2 années suivant le prélèvement sanguin.

## Résumé

---

Les cancers du sein et de la prostate sont parmi les cancers ayant la plus forte incidence dans le monde, notamment dans les pays occidentaux. Les principaux défis actuels sont d'améliorer la compréhension des relations nutrition/santé et l'identification des personnes à plus haut risque bien avant l'apparition du cancer afin de mettre en place des actions de préventions. De nombreux facteurs influencent la mise en place et la progression du cancer. Parmi eux, la nutrition apparaît comme un facteur clé, puisqu'il s'agit d'un facteur modifiable sur lequel il est possible d'agir *via* des interventions, il est donc essentiel d'évaluer sa contribution. Pour cela, une mesure précise de l'apport nutritionnel est nécessaire. La métabolomique permettant l'identification de potentiels biomarqueurs endogènes, exogènes et microbiens ouvre donc de nouvelles perspectives en épidémiologie nutritionnelle. A ce jour, encore très peu d'études ont investigué l'impact de l'alimentation globale sur le métabolisme et le risque de cancer du sein et de la prostate par profilage métabolomique.

Dans le cadre de cette thèse, nous avons donc conduit des études cas-témoins nichées et transversales dans la cohorte SU.VI.MAX afin de mettre en évidence des signatures plasmatiques du risque de cancers du sein et de la prostate et de l'alimentation globale. Les échantillons plasmatiques ont été collectés à l'inclusion dans la cohorte et analysés par deux méthodes complémentaires : la spectrométrie de masse couplée à la chromatographie liquide et la résonance magnétique nucléaire du proton. Les habitudes alimentaires des participants ont été estimées grâce à des enregistrements alimentaires de 24h répétés et les données sociodémographiques et de mode de vie ont été obtenues grâce à des questionnaires autodéclarés.

Ces recherches ont permis de mettre en évidence des métabolites endogènes et issus du métabolisme microbien associés à l'alimentation globale et également des biomarqueurs candidats d'une exposition alimentaire spécifique. Nous avons également identifié des métabolites associés au risque de cancers du sein et de la prostate, endogènes, exogènes et microbiens suggérant une perturbation métabolique jusqu'à 13 ans avant le diagnostic du cancer. Par ailleurs, l'alimentation semble jouer un rôle dans la variation des taux plasmatiques de certains métabolites permettant de discriminer les individus à plus haut risque de développer un cancer du sein ou de la prostate. Ces résultats devront être répliqués dans d'autres études indépendantes d'observation et d'intervention.

A terme, l'identification de signatures métaboliques robustes du risque de cancers du sein et de la prostate, de l'impact de l'alimentation sur le métabolisme et la cancérogenèse et de l'apport alimentaire pourraient permettre de contribuer à l'amélioration de la compréhension des relations entre environnement et santé, de l'évaluation de l'exposition nutritionnelle voire à la mise en place de nouvelles recommandations en matière de santé publique en vue de la diminution de l'incidence de ces pathologies.

## Abstract

---

Breast and prostate cancers are among the cancers with the highest incidence worldwide and notably in Western countries. The main current challenges lie in the improvement of understanding of nutrition/health relationships and in the identification of individuals at higher risk long before the development of overt cancer to set up prevention actions. A variety of factors exert an impact on the onset and progression of cancer. Among these, nutrition appears as a key factor, in that it can be modified and acted upon through interventions. It is therefore crucial to assess its contribution. For this purpose, detailed and accurate assessment of nutritional intake is essential. Metabolomics, allowing the identification of endogenous, exogenous and microbial biomarkers, opens new perspectives in nutritional epidemiology. So far, few have studies investigated the impact of overall diet on metabolism and risk of breast and prostate cancer through metabolomic profiling.

As part of this thesis, we conducted nested case-controls and cross-sectional studies within the SU.VI.MAX cohort to highlight plasma signatures of breast and prostate cancer risks and of overall diet. Plasma samples were collected at baseline and were analysed using two complementary methods: mass spectrometry coupled with liquid chromatography and proton nuclear magnetic resonance. Participants dietary habits were estimated using repeated 24h dietary records and socio-demographic and lifestyle data were collected from self-administered questionnaires.

These investigations highlighted endogenous and microbial metabolites associated with overall diet as well as candidate biomarkers of specific dietary exposures. We also identified endogenous, exogenous and microbial metabolites associated with breast and prostate cancers risk suggesting a metabolic disruption up to 13 years before cancer diagnostic. Furthermore, diet appears to be implicated in the variation in plasma levels of some metabolites discriminating individuals at higher risk of developing breast or prostate cancers. These results need to be replicated in future independent observational and interventional studies.

In the future, the identification of robust metabolic signatures of breast and prostate cancers risk, of the impact of diet on metabolism and carcinogenesis, and food intake would contribute to better understand health and environment relationships, to better estimate nutritional exposure or even to contribute to the set-up of new public health recommendations in order to reduce the incidence of these pathologies.