



HAL
open science

Développement et application de méthodes de chromatographie liquide couplées à la spectrométrie de masse à haute résolution pour les analyses métabolomiques et lipidomiques de larges cohortes

Samia Boudah

► **To cite this version:**

Samia Boudah. Développement et application de méthodes de chromatographie liquide couplées à la spectrométrie de masse à haute résolution pour les analyses métabolomiques et lipidomiques de larges cohortes. Chimie analytique. Université Pierre et Marie Curie - Paris VI, 2014. Français. NNT : 2014PA066281 . tel-01374866

HAL Id: tel-01374866

<https://theses.hal.science/tel-01374866>

Submitted on 2 Oct 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de
DOCTEUR DE L'UNIVERSITÉ PARIS 6

PÔLE VIE & SANTÉ

École doctorale Interdisciplinaire du Vivant (iViv)

Présentée et soutenue publiquement

par

Samia BOUDAH

le 24 Septembre 2014

Développement et application de méthodes de chromatographie liquide couplées à la spectrométrie de masse à haute résolution pour les analyses métabolomiques et lipidomiques de larges cohortes

JURY

Pr. Pierre Chaminade	Université Paris Sud	Rapporteur
Pr. Gérard Hopfgartner	Université de Genève	Rapporteur
Dr. Anatol Kontush	Inserm/Université Pierre et Marie Curie - Paris 6	Examineur
Dr. Bertrand Rochat	Université de Lausanne	Examineur
Pr. Germain Trugnan	Université Pierre et Marie Curie - Paris 6	Président du jury
Dr. Christophe Junot	CEA Saclay	Directeur de thèse
Dr. Etienne Thévenot	CEA Saclay	Membre invité

THÈSE

Pour obtenir le grade de
DOCTEUR DE L'UNIVERSITÉ PARIS 6

PÔLE VIE & SANTÉ

École doctorale Interdisciplinaire du Vivant (iViv)

Présentée et soutenue publiquement

par

Samia BOUDAH

le 24 Septembre 2014

Développement et application de méthodes de chromatographie liquide couplées à la spectrométrie de masse à haute résolution pour les analyses métabolomiques et lipidomiques de larges cohortes

JURY

Pr. Pierre Chaminade	Université Paris Sud	Rapporteur
Pr. Gérard Hopfgartner	Université de Genève	Rapporteur
Dr. Anatol Kontush	Inserm/Université Pierre et Marie Curie - Paris 6	Examineur
Dr. Bertrand Rochat	Université de Lausanne	Examineur
Pr. Germain Trugnan	Université Pierre et Marie Curie - Paris 6	Président du jury
Dr. Christophe Junot	CEA Saclay	Directeur de thèse
Dr. Etienne Thévenot	CEA Saclay	Membre invité

REMERCIEMENTS

Ce travail de thèse a été réalisé au CEA de Saclay, au sein du Service de Pharmacologie et d'Immunologie, dans le laboratoire d'Etude du Métabolisme des Médicaments. Je souhaite donc remercier en premier lieu le Dr. Christophe CREMINON, chef du service, et le Dr. Christophe JUNOT, chef de laboratoire, de m'avoir accueillie au sein du laboratoire.

Je remercie également les laboratoires GlaxoSmithKline, et plus particulièrement le Dr Edwige Nicodème, ainsi que l'Association Nationale de Recherche et Technologie d'avoir financé ce travail de thèse.

Je remercie sincèrement le Dr. Christophe JUNOT d'avoir dirigé mes travaux de thèse durant ces 3 années toujours avec clairvoyance. J'adresse mes remerciements également au Dr. Benoit COLSCH pour avoir encadré une partie de mes travaux de recherche. Je tiens à remercier le Pr. Jean-Claude TABET pour son aide précieuse et sa disponibilité.

Ce travail étant l'aboutissement de nombreuses collaborations, j'adresse mes remerciements à toutes les personnes qui y ont participé et plus particulièrement au Dr. Etienne THEVENOT et au Dr. Etienne LARGER qui ont permis d'enrichir nos conversations tripartites.

Je tiens à exprimer mon profond respect aux Pr. Pierre CHAMINADE et Pr. Gérard HOPFGARTNER pour avoir accepté de juger ce travail. J'associe mes remerciements au Pr. Germain TRUGNAN ainsi qu'aux Dr. Anatol KONTUSH et Dr. Bertrand ROCHAT pour avoir accepté de siéger à ce jury de thèse.

J'exprime ma reconnaissance au Pr Arlette BAILLET-GUFFROY pour m'avoir aidée à trouver ma voie.

Enfin, je souhaite remercier chaleureusement l'ensemble du personnel du 136 pour son accueil chaleureux et sa disponibilité.

PREAMBULE

Le métabolome est l'ensemble des composés organiques de petite taille contenus dans une cellule, un tissu ou un organisme. Il représente l'une des ultimes réponses d'un organisme à une altération génétique, une pathologie, une exposition à un toxique ou tout autre facteur susceptible de perturber son fonctionnement. Son analyse, dite métabolomique, est une discipline récente qui s'inscrit en complément de la protéomique, de la transcriptomique et de la génomique. Elle permet de caractériser l'état biochimique d'un organisme, et également d'obtenir des informations sur les interactions individus-environnement. De nos jours, les analyses métabolomiques trouvent peu à peu des applications concrètes dans les domaines de la biologie clinique et de la médecine de laboratoire. Cependant, malgré les évolutions technologiques de ces dernières années réalisées dans les domaines de la chimie analytique et du traitement de données, il reste encore de nombreux obstacles à surmonter en vue de fiabiliser au maximum les empreintes acquises et faire de l'analyse métabolomique une approche de référence pour les biologistes.

En effet, le but de toute approche métabolomique globale est de détecter et de quantifier un maximum de métabolites en vue d'avoir la vue la plus exhaustive possible du métabolome à un instant donné. Cependant, alors que les profils LC-MS obtenus sont de plus en plus informatifs notamment depuis l'avènement de la spectrométrie de masse à haute résolution, leur exploitation biologique reste limitée. Parmi les milliers de signaux enregistrés, seule une faible partie d'entre eux est identifiée et sert *in fine* à l'analyse biologique des résultats. De même, la difficulté d'assurer une couverture complète du métabolome par une seule technique d'analyse et la difficulté de mettre en commun les informations issues des différentes techniques d'analyse que sont la résonance magnétique nucléaire NMR, chromatographie en phase gazeuse couplée à la spectrométrie de masse (GC-MS) et chromatographie en phase liquide couplée à la spectrométrie de masse (LC-MS), abaisse encore ce rendement d'identification. Dans ce contexte, le laboratoire a pris le parti d'entreprendre une étude qualitative compréhensive par matrice biologique via le développement de bases de données spectrales dédiées. Cette démarche vise à identifier le maximum de métabolites détectés et ce en utilisant une seule méthode analytique, la LC/MS, afin de faciliter l'intégration des données et leur interprétation biologique, mais avec le risque de limiter l'exhaustivité du profil métabolique obtenu.

Ainsi, le principal objectif de mes travaux de thèse a été de développer des outils de phénotypage métabolique basés sur l'utilisation de la chromatographie liquide couplée à la spectrométrie de masse haute résolution (LC-HRMS). Le but est démontrer la polyvalence de cet outil

en élargissant le spectre des molécules détectées aux lipides, mais aussi de combiner plusieurs méthodes chromatographiques afin de pallier à l'hétérogénéité physico-chimique des métabolites d'intérêt et ce dans des matrices biologiques telles que le sérum, le plasma et les extraits de globules rouges humains. Ces outils analytiques développés ont ensuite été appliqués à l'analyse de cohortes afin de proposer des biomarqueurs permettant de mieux caractériser des états physiopathologiques. L'approche pluridisciplinaire de cette étude a été l'objet de différentes collaborations mises en place avec les biostatisticiens de la direction de la recherche technologique (DRT) du CEA et de la société Profilomic, et avec des cliniciens du service de diabétologie du groupe hospitalier Hôtel-Dieu-Cochin (Assistance Publique des Hôpitaux de Paris). Ces travaux de thèse ont été financés par les laboratoires GlaxoSmithKline en vue des futures applications pharmacologiques et par l'Association Nationale de Recherche et Technologies (ANRT), dans le cadre d'un contrat doctoral CIFRE.

Notre manuscrit de thèse débutera par une introduction dans laquelle seront présentées les biomarqueurs, les analyses métabolomiques et lipidomiques par spectrométrie de masse ainsi que les développements réalisés en matière de bioinformatique et de traitement de données, notamment en termes de normalisation, fusion, annotation des jeux de données et d'identification des métabolites. Nous présenterons ensuite notre démarche et les résultats que nous avons obtenus lors du développement d'approches globales de profilages métabolique et lipidomique par chromatographie liquide couplée à la spectrométrie de masse à haute résolution pour l'analyse du plasma et du sérum humains. Nous détaillerons enfin les applications biologiques qui ont servi de support à ces développements technologiques : analyse d'échantillons sériques provenant d'une cohorte de salariés du Commissariat à l'Energie Atomique et aux Energies Alternatives ainsi que les échantillons plasmatiques et érythrocytaire issus d'une cohorte de patients diabétiques.

Développement d'approches globales de profilage métabolique :

Notre objectif était d'utiliser la polyvalence de la chromatographie liquide couplée à la spectrométrie de masse à haute résolution (LC-HRMS) afin d'obtenir des empreintes métaboliques les plus informatives possible. Compte tenu de l'hétérogénéité des métabolites humains, aucune méthode LC-HRMS unique ne peut permettre de détecter l'ensemble des constituants du métabolome simultanément. Au sein du laboratoire, nous avons donc fait le choix de développer deux axes distincts d'analyses globales métabolomique et lipidomique. Pour chacun, une méthode d'extraction dédiée ainsi que différentes méthodes chromatographiques permettant la rétention et/ou la résolution du plus grand nombre possible de métabolites/lipides ont été mises au point. Une fois développées, ces méthodes ont d'abord servi à caractériser les métabolites et lipides détectés dans les biofluides et cellules humaines. Ceci a permis de développer des bases de données ESI-MS et ESI-MS/MS servant à l'annotation des jeux de données métabolomiques et lipidomiques de diverses matrices biologiques.

Une fois l'identification des métabolites/lipides réalisée, ces méthodes de profilage métabolomiques et lipidomiques ont été appliquées à l'analyse de cohortes.

Evaluation de l'impact de facteurs physiologiques sur le métabolome sérique :

Il est nécessaire de vérifier que les variations des concentrations de métabolites observées dans les empreintes métabolomiques ne sont pas affectées par des facteurs confondants d'ordre analytique ou physiologique ou par des sources de variabilité incontrôlées. En effet, certains paramètres biologiques (tel que l'âge), environnementaux (tel que le régime alimentaire) ou analytiques (tel que le type d'anticoagulant) peuvent avoir une influence sur les concentrations des métabolites et peuvent mener à des conclusions erronées. Nous avons de ce fait évalué cette variabilité biologique dans une population de 227 salariés du CEA-Saclay prélevés dans le cadre de leur suivi médical au service de santé au travail. Les échantillons sériques ont été analysés selon les protocoles développés et des analyses statistiques ont été réalisées afin d'établir des corrélations avec des sources de variabilité biologique incontrôlée dans notre cas l'âge, l'indice de masse corporelle et le genre. Cette démarche a déjà été initiée au sein du laboratoire sur les échantillons d'urines humaines de cette même cohorte (Thèse A. Roux, 2008-2011).

A la recherche de biomarqueurs de maladies multifactorielles : Vers une meilleure stratification des patients atteints de diabète de type 2

Les patients présentant une hyperglycémie chronique sans marqueurs d'auto-immunité sont appelés «diabétiques de type 2». Il est admis que plusieurs pathologies coexistent sous le qualificatif de diabète de type 2. Dans ce contexte, nous avons réalisé une étude préliminaire dont le but était de trouver des signatures métaboliques susceptibles d'expliquer l'hétérogénéité phénotypique des patients atteints de diabètes de type 1 et 2. Pour ce faire, des extraits plasmatiques de patients diabétiques ont été analysés et une démarche statistique a été développée afin de mettre en évidence des profils métaboliques permettant de distinguer les deux types de diabète, tout en essayant de nous affranchir de l'impact de facteurs physiologiques confondants tels que l'âge et l'indice de masse corporelle (IMC).

L'analyse métabolomique des globules rouges : une alternative aux analyses plasmatiques pour la découverte de biomarqueurs ?

La majorité des études métabolomiques réalisées dans le domaine de la médecine ou de la toxicologie sont basées sur des analyses plasmatiques ou urinaires. Bien que non ou faiblement invasifs, ces biofluides ne permettent pas toujours d'obtenir des résultats biologiquement exploitables, en raison de leur complexité et surtout des faibles concentrations circulantes de

métabolites marqueurs de situations pathologiques. Partant de ce constat, nous nous sommes orientés vers l'analyse du métabolome des globules rouges. Ces cellules sont un modèle d'étude simple puisqu'elles sont dépourvues de noyau, circulent dans tous les tissus et ont, chez l'homme, une durée de vie de 120 jours. Grâce au grand nombre de transporteurs de métabolites présents sur leur membrane, elles peuvent piéger et concentrer des métabolites qui pourraient rendre compte de l'état métabolique d'un patient sur la durée de vie du globule rouge. Cette propriété a déjà été utilisée depuis plusieurs décennies pour suivre la stabilité de la glycémie chez les patients diabétiques par dosage de l'hémoglobine glyquée (Hb A1c).

Afin d'évaluer l'intérêt de l'analyse métabolomique des globules rouges (GR) comme modèle biologique approprié et/ou alternatif aux milieux biologiques classiquement utilisés, nous avons analysé les extraits érythrocytaires provenant des patients diabétiques pour lesquels le métabolome plasmatique a été étudié précédemment. Ce travail a été réalisé en collaboration avec les équipes de Paul-Henri Roméo (DSV/iRCM, CEA-Fontenay-aux-Roses et GRéX : laboratoire d'excellence sur le globule rouge).

SOMMAIRE

REMERCIEMENTS.....	I
PREAMBULE.....	III
SOMMAIRE	VII
LISTE DES ABREVIATIONS	XI
INTRODUCTION GENERALE.....	1
CHAPITRE I. BIOMARQUEURS ET METABOLOME.....	5
1. Le concept de biomarqueur	5
1.1. Qu'est-ce qu'un biomarqueur ?	5
1.1. Les différents types de biomarqueurs et applications en biologie clinique.....	5
1.2. Méthodes de découverte de biomarqueurs	9
2. Le métabolome.....	12
2.1. Définition.....	12
2.2. Classification.....	13
CHAPITRE II. APPROCHE METABOLOMIQUE	15
1. Démarche expérimentale	16
2. Préparation des échantillons.....	16
2.1. Recueil des échantillons.....	17
2.2. Arrêt forcé du métabolisme «notion de quenching».....	18
2.3. Extraction des métabolites.....	19
2.4. Conservation et stabilité des échantillons	22
2.5. Evaluation du processus de préparation des échantillons :.....	23
3. Les méthodes de spectrométrie de masse en analyse métabolomique.....	24
3.1. Profilage métabolique par spectrométrie de masse.....	24
3.2. Identification des métabolites par spectrométrie de masse	35
3.3. Quantification des métabolites par spectrométrie de masse	42
4. Traitement des données en analyse métabolomique.....	46
4.1. Des données brutes aux données traitées.....	46
4.2. Des données traitées aux données biologiques.....	51
4.3. Outils informatiques pour le partage et l'échange des données	61
4.4. Fusion des données.....	62
CHAPITRE III. APPROCHE LIPIDIOMIQUE	65

1. Lipides : définition et classification	65
2. Analyse des lipides et analyses lipidomiques.....	68
2.1. Méthodes non spectrométriques :.....	69
2.2. Méthodes spectrométriques.....	71
3. Outils bioinformatiques pour l'analyse lipidomique.....	79
3.1. Banques de données en lipides.....	80
3.2. Logiciels pour l'annotation de données lipidomiques.....	83
4. Applications cliniques de l'analyse lipidomique.....	86
RESULTATS	91
A. DEVELOPPEMENT D'APPROCHES GLOBALES DE PROFILAGE METABOLIQUE PAR CHROMATOGRAPHIE LIQUIDE COUPLEE A LA SPECTROMETRIE DE MASSE A HAUTE RESOLUTION	93
CHAPITRE I. METHODE GLOBALE D'ANALYSE METABOLOMIQUE	93
1. Introduction.....	93
2. Matériel et méthodes.....	96
2.1. Réactifs et produits chimiques	96
2.2. Echantillons biologiques et préparation des échantillons.....	97
2.3. Analyses par chromatographie liquide couplée à la spectrométrie de masse.....	99
2.4. Traitement des données.....	100
3. Résultats et discussion :	101
3.1. Construction de la banque de données spectrales	101
3.2. Annotation du métabolome plasmatique humain	103
3.3. Annotation du métabolome érythrocytaire humain	116
4. Conclusion	119
CHAPITRE II. METHODE GLOBALE D'ANALYSE LIPIDOMIQUE.....	121
1. Introduction.....	121
2. Matériel et méthodes.....	123
2.1. Réactifs et produits chimiques	123
2.2. Instrumentation LC-MS	124
2.3. Préparation des échantillons.....	126
3. Résultats	127
3.1. Analyses par spectrométrie de masse d'extraits lipidiques	127
3.2. Construction d'une base de données de lipides des mammifères.....	144
3.3. Développement d'une approche lipidomique.....	155
3.4. Annotation du lipidome plasmatique humain.....	163
4. Conclusion et perspectives.....	167
B. APPLICATIONS DES APPROCHES GLOBALES AUX COHORTES	169

CHAPITRE I. ANALYSE DU METABOLOME ET LIPIDOME SERIQUES D'UNE COHORTE DE 223 SALARIES DU CEA-SACLAY	173
1. Matériels et méthodes	174
2. Echantillons biologiques	174
3. Traitement des données.....	175
4. Résultats	176
4.1. Démarche d'analyse.....	176
4.2. Analyse métabolomique	180
4.3. Analyse lipidomique	199
CHAPITRE II. ANALYSE DES METABOLOME ET LIPIDOME SERIQUES D'UNE COHORTE DE 78 PATIENTS DIABETIQUES	211
1. Enjeux et objectifs de l'étude	213
2. Matériel et méthodes.....	215
3. Echantillons biologiques	215
4. Traitement des données.....	216
5. Résultats et discussion	216
5.1. La démarche.....	216
5.2. Analyses statistiques des données plasmatiques.....	217
5.3. Analyse du métabolome érythrocytaire de 54 patients diabétiques.....	226
CONCLUSION ET PERSPECTIVES	237
LISTE DES FIGURES, TABLES ET GRAPHES	243
ANNEXE 1	251
RAPPELS THEORIQUES : LOGICIELS DE DETECTION AUTOMATIQUE DES SIGNAUX	251
ANNEXE 2	254
APPROCHE GLOBALE DE PROFILAGE METABOLOMIQUE	254
ANNEXE 3	269
ANALYSE LC-NH2 DES EXTRAITS LIPIDIQUES	269
QUANTIFICATION DES LIPIDES : ETUDE PRELIMINAIRE.....	270
ANNEXE 4	273
FUSION DES DONNEES METABOLOMIQUES ET LIPIDOMIQUES SERIQUES ET PLASMATIQUES DES SUJETS « TEMOINS » ET PATIENTS DIABETIQUES.....	273
1. Analyse Métabolomique	273
1.1. Matrices de données XCMS en analyse métabolomique.....	273
1.2. Matrices de données QuanBrowser en analyse métabolomique	283

2. Analyse lipidomique	286
2.1. lip-xcm-pos	286
2.2. lip-xcm-neg	287
RESULTATS DES TESTS DE L'ANALYSE DE LA COHORTE SACLAYSIENNE	287
ANNEXE 5	291
RESULTATS DES TESTS STATISTIQUES DE L'ANALYSE DE LA COHORTE DE PATIENTS DIABETIQUES – DONNEES PLASMATIQUES.....	291
1. QuanBrowser (PFPP-pos and HILIC-neg).....	291
1.1. diaVsac.....	291
1.2. di1Vsac	292
1.3. di2Vsac	293
1.4. di1Vdi2.....	294
2. Correction du signal en fonction des facteurs physiologiques confondants	296
2.1. Age.....	296
2.2. BMI	297
DONNEES ERYTHROCYTAIRES.....	297
1. Type de diabète.....	297
2. Empreinte érythrocytaire et BMI	298
PRODUCTIONS SCIENTIFIQUES.....	299
RÉFÉRENCES.....	301

LISTE DES ABBREVIATIONS

ACN	Acétonitrile
ALAT	Alanine aminotransférase
APCI	Ionisation chimique à pression atmosphérique (<i>de l'anglais "Atmospheric Pressure Chemical Ionization"</i>)
API	Ionisation à pression atmosphérique (<i>de l'anglais Atmospheric Pressure Ionization</i>)
APPI	Photoionisation à pression atmosphérique (<i>de l'anglais "Atmospheric Pressure Photoionization"</i>)
ASAT	Aspartate aminotransférase
BiGG	Base de connaissance biochimique, génétique et génomique (<i>de l'anglais "Biochemical Genetic and Genomic"</i>)
CE	Esters de cholestérol (<i>de l'anglais "Cholesteryl Ester"</i>)
CE-MS	Electrophorèse capillaire couplée à la spectrométrie de masse (<i>de l'anglais "Capillary Electrophoresis coupled to"</i>)
Cer	Céramides
CID	<i>De l'anglais "Collision Induced Dissociation"</i>
CV	Coefficient de variations
DDA	Mode d'acquisition données-dépendant (<i>de l'anglais "Data Dependant Analysis"</i>)
DESI	Desorption Ionization par electrospray (<i>de l'anglais "Desorption Electrospray Ionization"</i>)
DG	Diglycérides
DIA	Mode d'acquisition données-indépendant (<i>de l'anglais "Data Independant Analysis"</i>)
EDTA	Acide éthylène diamine tétraacétique
EI	Ionisation par impact électronique (<i>de l'anglais "Electronic Impact"</i>)
EIC	Courant ionique extrait (<i>de l'anglais "Extracted Ion Chromatogram"</i>)
ELT	Extrait Lipidique Total
ESI	Ionisation par electrospray (<i>de l'anglais "Electrospray Ionization"</i>)
EtOH	Ethanol
ExS	Etalon externe (<i>de l'anglais "External Standard"</i>)
FAB	<i>De l'anglais "Fast Atom Bombardment"</i>
FFA	Acides gras libres (<i>de l'anglais "Free Fatty Acids"</i>)
GalCer	Galactosylcéramides
GC-MS	Chromatographie en phase gazeuse couplée à la spectrométrie de masse (<i>de l'anglais "Gas Chromatography"</i>)
GluCer	Glucosylcéramides
GR	Globules rouges ou érythrocytes
Hb1Ac	Hémoglobine glyquée
HCD	<i>De l'anglais "Higher energy Collision induced Dissociation"</i>
HDL	<i>De l'anglais "High Density Lipoprotein"</i>
HexCer	Hexosylcéramides
HILIC	Chromatographie liquide d'interactions hydrophiles (<i>de l'anglais "Hydrophilic Interaction Liquid Chromatography"</i>)
HRMS	Spectrométrie de masse à haute résolution (<i>de l'anglais "High Resolution Mass Spectrometry"</i>)
HTA	Hypertension artérielle
IR	Spectroscopie Infra-rouge
IS	Etalon interne (<i>de l'anglais "Internal Standard"</i>)
LADA	Latent Auto-Immune Diabetes of Adults
LC	Chromatographie liquide (<i>de l'anglais "Liquid Chromatography"</i>)
LC-MS	Chromatographie en phase liquide couplée à la spectrométrie de masse (<i>de l'anglais "Liquid Chromatography"</i>)
LCR	Liquide céphalo-rachidien

LDL	De l'anglais "Low Density Lipoprotein"
LMCO	De l'anglais <u>L</u> ow <u>M</u> ass <u>C</u> ut <u>O</u> ff
LPA	Acides lyso-phosphatidiques (de l'anglais " <u>L</u> yo- <u>p</u> hosphatidic <u>A</u> cid)
LPC	<u>L</u> yo-Glycéro <u>P</u> hosphatidyl <u>C</u> holine
LPE	<u>L</u> yo-Glycéro <u>P</u> hosphatidyl <u>E</u> thanolamine
LPG	<u>L</u> yo-Glycéro <u>P</u> hosphatidyl <u>G</u> lycerol
LPI	<u>L</u> yo-Glycéro <u>P</u> hosphatidyl <u>I</u> nositol
LPS	<u>L</u> yo-Glycéro <u>P</u> hosphatidyl <u>S</u> érine
MALDI	De l'anglais " <u>M</u> atrix- <u>A</u> ssisted <u>L</u> aser <u>D</u> esorption <u>I</u> onization"
MeOH	Méthanol
MG	Monoglycérides
MS	Spectrométrie de masse (de l'anglais "Mass spectrometry")
NLS	Mode de balayage de perte de neutres (de l'anglais " <u>N</u> eutral <u>L</u> oss <u>S</u> can")
NMR	Résonance magnétique nucléaire (de l'anglais " <u>N</u> uclear <u>M</u> agnetic <u>R</u> esonance")
NP ou NPLC	Chromatographie liquide en phase normale (de l'anglais " <u>N</u> ormal <u>P</u> hase <u>L</u> iquid <u>C</u> hromatography")
OPLS	<u>O</u> rthogonal <u>P</u> rojection to <u>L</u> atent <u>S</u> tructures
PA	<u>A</u> cide <u>P</u> hosphatidique
PC	Glycéro <u>P</u> hosphatidyl <u>C</u> holine
PCA	Analyse en composantes principales (de l'anglais " <u>P</u> rincipal <u>C</u> omponent <u>A</u> nalysis")
PE	Glycéro <u>P</u> hosphatidyl <u>E</u> thanolamine
PFPP	<u>P</u> entafluorophenylpropyl
PG	Glycéro <u>P</u> hosphatidyl <u>G</u> lycerol
PI	Glycéro <u>P</u> hosphatidyl <u>I</u> nositol
PIS	Mode de balayage d'ion parent (de l'anglais " <u>P</u> recursor <u>I</u> on <u>S</u> can")
PS	Glycéro <u>P</u> hosphatidyl <u>S</u> érine
QC	Contrôle qualité (de l'anglais " <u>Q</u> uality <u>C</u> ontrol")
RP ou RPLC	Chromatographie liquide en phase inverse (de l'anglais " <u>R</u> everse <u>P</u> hase <u>L</u> iquid <u>C</u> hromatography")
SIMS	Spectrométrie de masse à ionisation secondaire (de l'anglais " <u>S</u> econdary <u>I</u> on <u>M</u> ass <u>S</u> pectrometry")
SL	De l'anglais " <u>S</u> hotgun <u>L</u> ipidomics"
SM	Sphingomyélines
SRM	De l'anglais " <u>S</u> electe <u>d</u> <u>R</u> eaction <u>M</u> onitoring"
Su	Sulfatides ou sulfogalactosylcéramides
T1D	Diabète de type 1
T2D	Diabète de type 2
TG	Triglycérides
TIC	Courant ionique total (de l'anglais " <u>T</u> otal <u>I</u> on <u>C</u> urrent")
TLC	Chromatographie sur couche mince (de l'anglais " <u>T</u> hin <u>L</u> ayer <u>C</u> hromatography")
TOF	Analyseur à temps de vol (de l'anglais " <u>T</u> ime <u>O</u> f <u>F</u> light")
TQ	Analyseur triple quadripolaire (de l'anglais " <u>T</u> riple <u>Q</u> uadripole")
UV	Spectroscopie <u>U</u> ltra- <u>v</u> iolet
VIP	Variable Importance in the Projection
ZICpHILIC	De l'anglais " <u>Z</u> witterionic stationary phase covalently attached to porous polymer beads suitable for <u>H</u> ydrophilic

**ON FAIT LA SCIENCE AVEC DES FAITS, COMME ON FAIT UNE MAISON AVEC DES
PIERRES, MAIS UNE ACCUMULATION DE FAITS N'EST PAS PLUS UNE SCIENCE QU'UN
TAS DE PIERRES N'EST UNE MAISON.»**

HENRI POINCARÉ (1854-1912).
LA SCIENCE ET L'HYPOTHESE.

INTRODUCTION GENERALE

CHAPITRE I. BIOMARQUEURS ET METABOLOME

1. Le concept de biomarqueur

1.1. Qu'est-ce qu'un biomarqueur ?

Les biomarqueurs sont des paramètres indicateurs et mesurables d'un état biologique donné. En 2001, l'Institut national américain de la santé (NIH) restreignait leur utilisation au domaine de la recherche et du développement pharmaceutique de médicaments en les définissant comme «des caractéristiques objectivement mesurées et évaluées comme indicateur de processus biologiques normaux, processus pathogènes, ou des procédés pharmacologiques suite à une intervention thérapeutique» (Groupe de travail Définitions biomarqueurs, 2001)(Biomarkers Definitions Working Group, 2001). Aujourd'hui, de plus en plus largement utilisés dans le domaine de la chimie clinique appliquée à la médecine ou à la toxicologie, les biomarqueurs sont considérés comme des indicateurs internes mesurables d'altérations moléculaires et/ou cellulaires qui peuvent apparaître dans un organisme au cours de ou après la progression d'une maladie ou d'une éventuelle exposition à un agent toxique (Bennett & Waters, 2000, Paustenbach, 2001). Bien que la notion de biomarqueur n'ait été formalisée que récemment, son utilisation remonte néanmoins à l'antiquité, notamment avec l'analyse organoleptique des urines (goût, couleur et odeur), ou avec la recherche de la présence éventuelle de sédiments dans le diagnostic du diabète sucré (Suruta et Charaka, 2000 ans avant J.C, Papyrus de Thèbes, 1500 ans avant J.C., Pharmaceutice rationalis T. Willis, 1674, (Armstrong, 2007, Haber, 1988)). Il faudra attendre 1878, pour qu'un physiologiste français, Claude Bernard, introduise la notion de «milieu intérieur» en se basant sur la capacité que possède un organisme vivant à conserver son équilibre de fonctionnement en dépit des contraintes qui lui sont imposées par le milieu extérieur (Bernard, 1865). Plus tard, dans son ouvrage intitulé « The wisdom of the body », Walter B. Cannon a considéré les êtres vivants supérieurs comme un système ouvert présentant de nombreuses relations avec l'environnement, ce dernier affectant directement le «milieu intérieur» qui tend à rester à l'équilibre (notion d'état d'équilibre, ou «steady state», en anglais). Il formalisa cette notion de *sagesse du corps* en introduisant le terme d'«homéostasie» (Walter B. Cannon, 1932).

1.1. Les différents types de biomarqueurs et applications en biologie clinique

Selon leur nature, les biomarqueurs peuvent être divisés en plusieurs catégories. Ils peuvent être de nature (i) biologique telle que la variation de la température corporelle dans le diagnostic d'infections bactériennes, (ii) physique telle que la mesure de la pression artérielle dans le suivi de maladies cardiovasculaires ou encore (iii) biochimique telle que la concentration d'un métabolite ou

d'une protéine donnée dans un milieu biologique donné (ex ; dosage du glucose dans le diagnostic du diabète, dosage de la protéine C-réactive dans le suivi de maladies inflammatoires).

Il faut noter qu'avec les développements de l'instrumentation en imagerie, de plus en plus de biomarqueurs mesurés en imagerie par résonance magnétique nucléaire (NMR) ou par tomographie par émission de positrons (PET) sont utilisés en médecine, notamment du fait de leur caractère non invasif. Pour exemple, l'injection du traceur radiomarqué [¹⁸F]-fluorodeoxyglucose lors d'analyse par PET permet de suivre la métabolisation du glucose *in situ* d'où son utilisation dans le suivi de la propagation de certains processus tumoraux.

Plus récemment, les **biomarqueurs moléculaires** ont vu le jour (Baker, 2005). En effet, les développements en chimie analytique, biologie moléculaire et informatique ont permis d'obtenir des signatures moléculaires multiparamétriques de pathologies comme le démontre une étude menée au centre néerlandais de recherche sur le cancer. Soixante-dix gènes humains incriminés dans le développement du cancer du sein ont été mis en évidence à l'aide de puces à ADN et d'analyse statistiques multivariées supervisées sur des cellules primaires cancéreuses ponctionnées sur 117 jeunes filles atteintes d'un cancer du sein. Réunis sous la forme d'une signature moléculaire, ils peuvent prédire la survenue de métastases et peuvent être utilisés pour adapter la thérapeutique (van, V et al; 2002).

1.1.1. Biomarqueurs en toxicologie

Les biomarqueurs sont largement utilisés en toxicologie (Timbrell, 1998). Ils sont subdivisés en 3 catégories : les biomarqueurs d'exposition, les biomarqueurs d'effet et les biomarqueurs de susceptibilité. Un biomarqueur d'exposition renseigne quant à l'importance de l'exposition à un toxique. La biosurveillance d'exposition est utilisée depuis longtemps en milieu professionnel (exemple des dosages du Plomb (Kehoe, Thamann, & Cholak, 1933) ou des métabolites du benzène (Pearce, Schrenk, & Yant, 1936) dans les urines ou le sang). Les biomarqueurs d'effet documentent l'interaction entre une substance toxique et sa cible biologique (exemple des substances cancérigènes et/ou mutagènes qui se lient aux macromolécules de type ADN ou encore de la présence d'un stress oxydant qui engendre la production de substances comme la 8-hydroxy-2'-desoxyguanosine ou qui perturbe la voie de biosynthèse du glutathion) (Angerer, Ewers, & Wilhelm, 2007). Enfin, les biomarqueurs de susceptibilité décrivent les différences inter-individuelles en réponse à des substances toxiques, des facteurs génétiques ou environnementaux (régime alimentaire, pathologies associées...). Par exemple, le polymorphisme des enzymes de détoxification et/ou d'activation a été identifié comme un facteur clé des variations entre individus en réponse à un toxique. Dans ce contexte, Haufroid et coll. ont démontré la relation entre l'excrétion urinaire

d'acides phenylhydroxyethylmercapturique et le polyphormisme génétique de l'enzyme glutathion S-transférase M1 (Haufroid et al; 2002).

1.1.2. Biomarqueurs en médecine

En médecine, les biomarqueurs sont couramment utilisés pour le diagnostic d'un état physiopathologique mais aussi le suivi thérapeutique des patients. Ce sont souvent des métabolites ou protéines dont la pertinence a été démontrée par des recherches en médecine expérimentale, génétique, biochimie, et grâce aux progrès réalisés au 20^{ème} siècle dans le domaine de la chimie analytique, notamment en immunoanalyse (Yalow & Berson, 1960b, Yalow & Berson, 1960a) et en physicochimie (Bader, Lehnert, & Angerer, 1994).

Le glucose est le biomarqueur le plus connu et le plus anciennement utilisé. Sa concentration sérique permet le diagnostic et le suivi du diabète. Aujourd'hui, de plus en plus de molécules peuvent être dosées en routine dans des laboratoires de biologie clinique et rendent ainsi compte d'un état pathologique donné. Nous citerons comme exemples la troponine C qui est une protéine sécrétée dans le sang par le muscle cardiaque suite à un infarctus du myocarde et la vitamine D dont le dosage dans le sang rend compte de l'état de calcification des os, en prévention du rachitisme chez l'enfant.

De nos jours, les méthodes séparatives couplées aux spectromètres de masse (LC-MS, GC-MS) sont de plus en plus présentes en milieu hospitalier. Les approches de quantification développées en GC-MS ou LC-MS sont devenues des méthodes de référence pour le dosage de petites molécules telles que les hormones stéroïdiennes, acylcarnitines et dérivés. C'est particulièrement le cas aux Etats-Unis d'Amérique où la spectrométrie de masse est largement utilisée dans le diagnostic précoce des erreurs innées du métabolisme chez le nouveau-né (Chace et al; 2009b, Chace et al; 2009a, Chace et al; 2009c). Par contre, pour le dosage de biomarqueurs protéiques, l'immunoanalyse reste la technique de référence tel que recommandé par la Food and Drug Administration (FDA) du fait de ses performances en termes de sensibilité et de sélectivité (Surinova et al; 2011). Néanmoins, ces tests sont très coûteux et peuvent être remplacés par des techniques basées sur la dilution isotopique et l'analyse en spectrométrie de masse, même si le plus souvent, ces techniques restent moins sensibles qu'un test immunologique (Carr & Anderson, 2008).

Aussi, depuis les années 1990 avec les développements technologiques réalisés en biologie moléculaire, chimie analytique et bio-informatique, les applications des approches « omiques » en médecine ont vu le jour et donnent une vue d'ensemble de la biologie des systèmes en allant du génotype au phénotype via l'analyse du génome, transcriptome, protéome et métabolome. Ce nouveau paradigme ouvre des perspectives nouvelles en termes de biomarquage de pathologies. En effet, les approches omiques permettent de générer des hypothèses biologiques dans des contextes où les raisonnements focalisés sur un phénomène biologique ou une voie métabolique particuliers

sont inefficaces, comme c'est le cas pour les pathologies multifactorielles telles que les maladies métaboliques (Cheng et al; 2012, Wang et al; 2011) et les cancers (Sreekumar et al; 2009). Un exemple concret est l'étude menée par le service de neurologie de l'hôpital «Pitié Salpêtrière», où l'analyse du métabolome du liquide céphalorachidien par NMR de 149 patients souffrants d'encéphalopathies non-expliquées a permis de mettre en évidence une nouvelle forme d'ataxie cérébelleuse caractérisée par une augmentation d'acide sialique dite CAFSA (pour Cerebellar Ataxia with Free Sialic Acid) (Mochel et al; 2009).

1.1.3. Biomarqueurs dans la recherche et développement pharmaceutique

La découverte de candidats médicaments débute par un criblage à haut débit sur un modèle cellulaire ou *in vitro* (le plus souvent un récepteur ou une enzyme) qui a pour but de mettre en évidence les composés actifs sur une cible thérapeutique (appelés «hits»). Parmi les molécules testées, les molécules actives sur des modèles animaux (appelées «leads») sont optimisées au point de vue de leur toxicité, propriétés pharmacocinétiques (augmenter la biodisponibilité) et pharmacodynamiques (augmenter l'efficacité et la sélectivité du produit pour la cible). Les quelques molécules issues de cette optimisation subissent des tests précliniques (chez l'animal), puis pour les plus concluants des tests cliniques (chez l'homme) après formulation galénique. Ce processus de recherche et développement est long et coûteux. A chaque étape, le taux d'échec est augmenté pour cause d'inefficacité ou toxicité du candidat médicament. Ainsi peu de «leads» passent les tests précliniques et cliniques et un seul dossier d'autorisation de mise sur le marché est soumis aux autorités sanitaires compétentes. Avoir recours aux approches «omiques» aux différentes étapes de ce processus facilitera la découverte de nouveaux médicaments (Baker, 2005)(Figure 1).

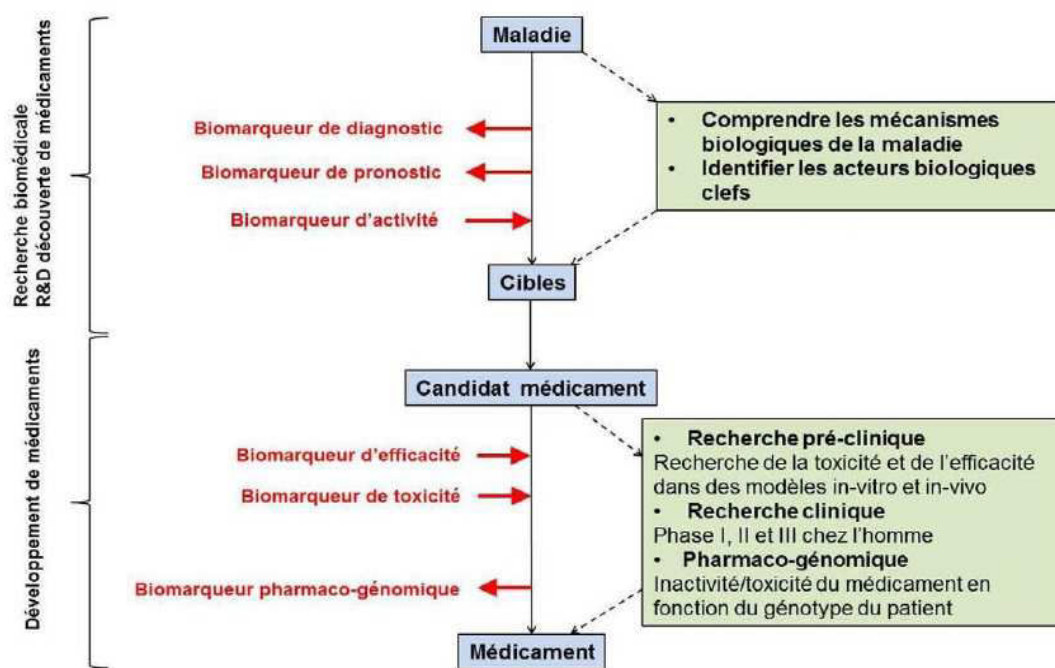


Figure 1. Apport de l'utilisation des biomarqueurs au processus de recherche et développement pharmaceutique

D'autre part, les biomarqueurs issus d'études pharmaco-génomiques sont également largement utilisés et permettent ainsi de mieux personnaliser les thérapies en mettant en exergue des génotypes sensibles ou résistants au traitement. Pour exemple, la protéine herceptin (HER2) est la cible du trastuzumab, un anticorps monoclonal anti tumeur du sein. La sur-expression de cette protéine chez 25% des patientes atteintes de cancer du sein peut aider à l'adaptation de la thérapie (Alymani et al; 2010). Ces biomarqueurs compagnons (c'est-à-dire vendus avec le médicament) trouvent des applications dans le domaine de la médecine personnalisée et intéressent de plus en plus l'industrie pharmaceutique qui commercialise des tests reposant sur leur mise en évidence (Trusheim et al; 2011). Citons par exemple le test herceptin/Herceptest™ de Roche, ou encore le test le gène KRAS/Erbitux® développé par Merck/Imclone qui sert à prédire la résistance d'un cancer colorectal au Cetuximab (Lievre et al; 2006).

1.2. Méthodes de découverte de biomarqueurs

Malgré les améliorations technologiques réalisées dans les domaines de la génomique, de la transcriptomique (gènes ou profils d'expression de gènes) et de la protéomique (protéines), seuls quelques biomarqueurs sont utilisés en clinique. En effet, la plupart des études présentent des listes de «candidats biomarqueurs» plutôt que des biomarqueurs validés. Ainsi, au terme de l'année 2009, la FDA n'avait approuvé que 109 protéines biomarqueurs (dont 80% ont été mesurées à l'aide de tests immunologiques). Sachant que 87% d'entre elles ont été introduites sur le marché avant 1993,

le taux moyen d'approbation de la FDA entre 1993 et 2009 est d'environ 1 biomarqueur de nature protéique par an (Anderson, 2010, Carr & Anderson, 2008, Rifai, Gillette, & Carr, 2006, Surinova et al; 2011, Whiteaker et al; 2011). Ce faible taux de réussite est imputé (i) à des limitations technologiques tel que le manque de techniques analytiques -autres qu'immunologiques- dans le criblage à haut débit et dans la validation sur un jeu de données important (ii) à un manque de cohortes de validation et (iii) à l'absence d'une procédure standardisée de recherche et développement de biomarqueurs.

Pepe et coll. ont été parmi les premiers auteurs à proposer une procédure à 5 étapes pour le développement de biomarqueurs, et ce dans le cadre de la détection précoce de cancers. La phase 1 consiste en une étude exploratoire pré-clinique qui utilise des profils d'immunohistochimie, Western blot ou d'expression des gènes. La phase 2 traite de la validation en essai clinique des candidats biomarqueurs mis en exergue lors des tests précliniques. La phase 3 consiste en l'analyse d'études rétrospectives longitudinales. Elle a pour objectif d'évaluer si les candidats biomarqueurs permettent d'améliorer la rapidité du diagnostic par rapport aux méthodes conventionnelles. En phase 4, l'étape ou la nature du cancer sont pris en compte dans le cadre d'études de dépistage et les sources de faux positifs sont identifiées. Enfin, la phase 5 a pour objectif de valider l'efficacité des candidats biomarqueurs: permettent-ils de réduire la prévalence des cancer (Pepe et al; 2001) ?

Plus récemment, le même genre de stratégie a été proposé pour la découverte et validation de biomarqueurs de nature protéique (Addona et al; 2011, Rifai, Gillette, & Carr, 2006, Surinova et al; 2011, Whiteaker et al; 2011) (voir Figure 2). La **phase de découverte** vise à identifier les protéines candidats biomarqueurs, notamment en utilisant la déplétion en protéines ou des étapes de fractionnement couplée à des approches protéomiques par chromatographie liquide couplée à la spectrométrie de masse à haute résolution, afin de détecter le plus grand nombre possibles de protéines. Au cours de la **phase de qualification**, les méthodes d'analyse ainsi que les méthodes de préparation d'échantillons sont ciblées de façon à vérifier la pertinence des candidats issus de la première étape. Des outils de protéomique ciblée comme ceux reposant sur l'utilisation combinée de la dilution isotopique et de la chromatographie liquide couplée à un spectromètre de masse de type triple quadripolaire fonctionnant en mode SRM (de l'anglais Selected Reaction Monitoring SRM) sont utilisés pour cela. Ensuite, lors de la **phase de vérification**, les performances cliniques (comme la spécificité ou la sensibilité) du candidat biomarqueur sont évaluées. Il y a alors optimisation du dosage de la protéine biomarqueur, et ce généralement via le développement d'un test d'immunoanalyse, qui répond à l'exigence d'un test de diagnostic *in vitro*. Enfin, le processus se termine par une phase de **validation de biomarqueurs** et de commercialisation (Rifai, Gillette, & Carr, 2006, Surinova et al; 2011).

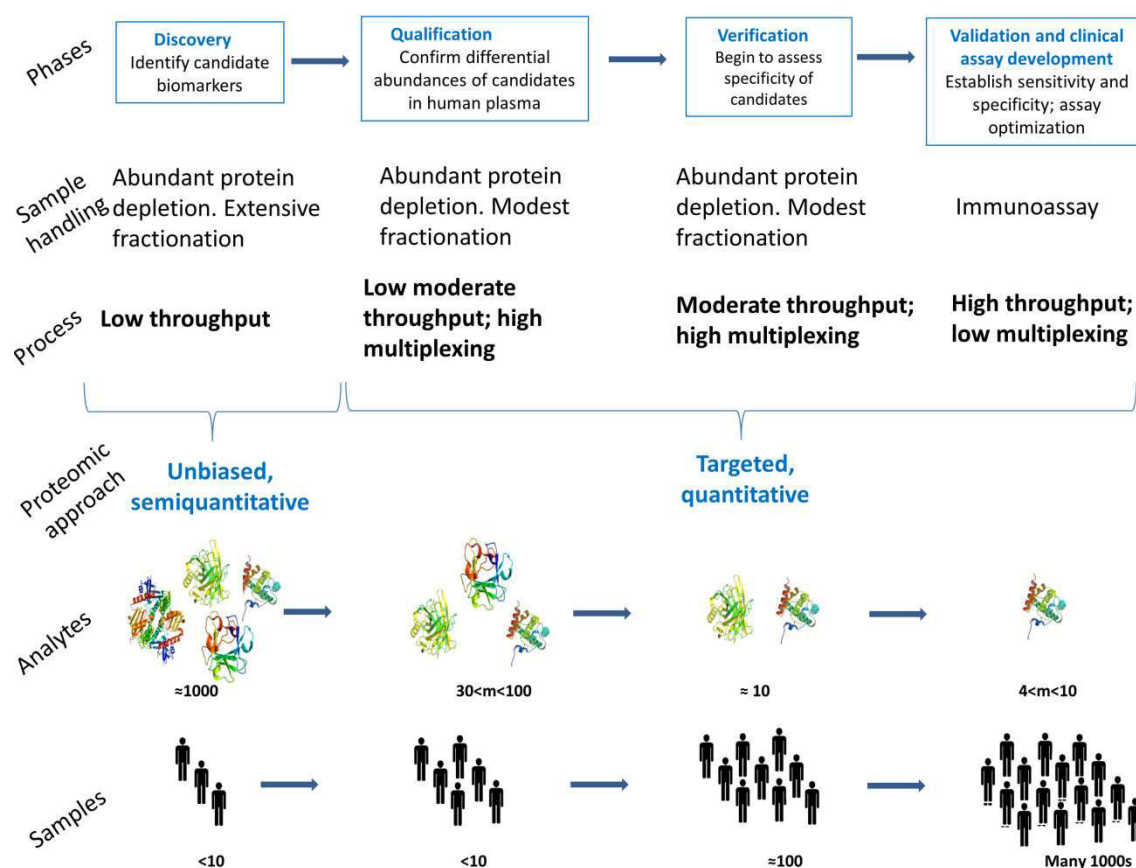


Figure 2. Stratégie de découverte de biomarqueurs protéiques (adaptée de (Rifai, Gillette, & Carr, 2006))

Une telle approche est illustrée par une étude publiée en 2011 visant à identifier des biomarqueurs protéiques de lésions myocardiques à partir d'échantillons de plasma veineux (Addona et al; 2011). La phase de découverte a été effectuée en utilisant le sang recueilli à partir des sinus coronariens de trois patients subissant un infarctus du myocarde thérapeutique provoqué dans le cadre d'une étude portant sur le traitement de la cardiomyopathie hypertrophique. Ces échantillons sanguins coronariens ont été obtenus au départ, à 10 puis à 60 minutes après l'infarctus du myocarde thérapeutique, et ils ont été analysés avec un spectromètre de masse type LTQ-Orbitrap. Après une étape de protéolyse de l'extrait de protéines, les masses liées aux peptides protéotypiques d'intérêt sont sélectionnées et intégrées à une liste d'inclusion. Dès que le peptide inclus dans la liste est détecté au bon rapport m/z, les spectres MS/MS sont acquis afin de confirmer la séquence peptidique (Jaffe et al; 2008). En procédant ainsi, 1105 protéines uniques ont été détectées, 70% d'entre elles étant observées chez les 3 patients. Les niveaux de concentration de 40 de ces protéines ont été plus que quintuplés par rapport aux concentrations témoins. De plus, 81 peptides protéotypiques de protéines d'intérêt ont également été détectées chez 2 des 3 patients, parmi lesquelles des marqueurs déjà connus tels que la myoglobine, la myéloperoxydase ou l'isoforme B de

la créatine kinase cardiaque. Une phase de qualification a ensuite été mise en œuvre en utilisant des échantillons de sang périphérique. Parmi les 121 protéines d'intérêt initialement observées dans des échantillons sanguins coronariens, l'approche AIMS a permis d'en détecter 83 dans des échantillons le sang périphérique de ces trois patients. Les concentrations de 53 de ces protéines suivent la même évolution dans le temps pour les 3 patients. Grâce à des expériences de LC-MS/MS (en mode MRM, voir Figure 8) réalisées sur les échantillons de ces mêmes 3 patients, 12 protéines candidats biomarqueurs (dont 8 nouvellement associées à la pathologie) ont été mises en évidence. Pour finir, la validation biologique de ces candidats biomarqueurs s'est effectuée sur une cohorte de 52 individus (26 témoins et 26 sujets ischémiques induits).

Des approches analogues à celles développées en analyses protéomiques ont vu le jour pour les analyses des molécules organiques de petite taille. Elles combinent des méthodes de chimie analytique et des outils de traitement de données. Ces approches sont dites «métabolomiques». Apparues à la fin des années 1990, elles ont déjà permis de mettre en évidence des biomarqueurs dans différents domaines telle la cancérologie (Dang et al; 2009, Sreekumar et al; 2009), de la reproduction (Brison et al; 2004) (Marhuenda-Egea et al; 2009), les maladies cardiovasculaires (Lewis et al; 2008) (Sabatine et al; 2005), ou encore les maladies neurologiques (Mochel et al; 2009) (Quinones & Kaddurah-Daouk, 2009).

2. Le métabolome

2.1. Définition

Apparu à la fin des années 90 en analogie aux notions de génome (ensemble des gènes d'un organisme), de transcriptome (ensemble des ARN messagers contenus dans un milieu biologique) et de protéome (ensemble des protéines contenues dans un milieu biologique), le métabolome fait référence à l'ensemble des métabolites contenus dans un système biologique donné (Tweeddale, Notley-McRobb, & Ferenci, 1998) (organismes, cellules, tissus, fluides biologiques). Le terme de «métabolite» inclut les molécules organiques de petite taille qui sont impliquées dans des processus métaboliques, qu'elles en soient les substrats, les produits ou nécessaires à leur bon fonctionnement (Dettmer & Hammock, 2004) (carbohydrates, acides aminés, acides organiques, lipides, ou encore co-facteurs de réactions enzymatiques...) (voir Figure 3). Nous détaillerons plus particulièrement les lipides et leur analyse dans le chapitre suivant dans la mesure où une partie de ce travail de thèse est consacrée.

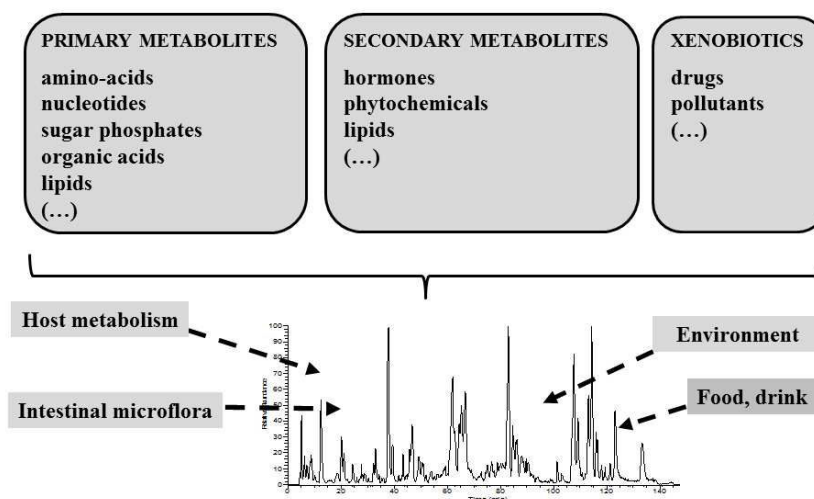


Figure 3. Métabolome et analyse métabolomique (Junot et al; 2013)

Le métabolome englobe l'ensemble des molécules de petite taille issues du métabolisme primaire et secondaire de l'organisme vivant ainsi que les xénométabolites. Chez l'Homme, son empreinte est impactée par l'ensemble des interactions hôte-environnement, le microbiote intestinal et l'apport nutritionnel.

2.2. Classification

Les métabolites sont le plus souvent classés selon leur distribution chez les êtres vivants au regard de la classification taxonomique. Dans ce sens, les métabolites ubiquitaires sont dits «primaires». Ils sont présents chez tous les organismes vivants eucaryotes et procaryotes (à savoir le règne des protistes, des mycètes, des végétaux, des animaux et des bactéries), et sont directement impliqués dans des processus physiologiques essentiels à la vie. Nous citerons comme exemple les acides aminés ou les métabolites impliqués dans la glycolyse. A l'opposé, les métabolites secondaires sont propres à chaque règne voire même à chaque espèce et sont synthétisés en vue d'une fonction biologique particulière. Citons les hormones stéroïdiennes chez l'homme, et les terpenoïdes chez les plantes. Enfin, en 2007, Holmes et coll. ont proposé le concept de «xénométabolome» (Holmes et al; 2007) pour faire référence aux xénotiques, composés qui ne sont pas synthétisés par un organisme vivant et qui sont apportés par l'alimentation, un traitement médical, ou l'exposition environnementale (polluants chimiques), ainsi qu'à leurs produits de biotransformation et leurs métabolites. Par exemple, la nicotine est un métabolite secondaire du plant de tabac, mais est considérée comme faisant partie du xénométabolome humain car apportée par la fumée de cigarette.

La classification s'est également historiquement structurée selon les critères chimiques tels que les méthodes d'extraction : les métabolites hydrosolubles pour lesquels des solvants d'extraction polaires sont utilisés et les lipides pour lesquels des solvants ou mélanges de solvants apolaires sont nécessaires à l'extraction. Cette classification reste insuffisante du fait de la grande diversité

structurale observée avec les métabolites et peut parfois mener à des conclusions erronées du fait de l'existence de lipides polaires (et de ce fait, extractibles par des solvants polaires). A cet effet, une classification selon les fonctions chimiques et substituants portés par la molécule peut être appropriée. Des bases de données biochimiques telles HMDB (Wishart et al; 2007) fournissent déjà cette information. Cependant, cette classification peut-être problématique pour les molécules hydrosolubles qui sont souvent pluri-fonctionnalisées et qui du coup ne peuvent pas appartenir à une seule catégorie. Par contre, cette démarche est largement applicable aux lipides et a fait l'objet d'un travail de concertation notamment à travers le consortium LIPID MAPS (Fahy et al; 2009, Fahy et al; 2005). Cette classification sera développée en PARTIE A.CHAPITRE III.1.

L'identification et la quantification ciblées d'un sous-ensemble des métabolites, ou encore les analyses dites globales visant à détecter le plus grand nombre possible de métabolites présents dans un système biologique donné définissent la «métabolomique» (Fiehn, 2002).

CHAPITRE II. APPROCHE METABOLOMIQUE

L'analyse métabolomique est une science récente qui consiste en l'analyse intégrative et fonctionnelle du phénotype métabolique d'un système biologique (Nicholson, Lindon, & Holmes, 1999, Oliver et al; 1998). Etant un des maillons finaux de la cascade des approches «omiques», elle reflète la réponse ultime d'un organisme à une altération génétique, une pathologie, une exposition à un toxique ou tout autre facteur susceptible de perturber son fonctionnement. Elle est une approche globale qui repose sur la comparaison d'empreintes métaboliques. Elle permet ainsi dans un premier temps de définir des classes d'échantillons, puis dans un deuxième temps de caractériser les métabolites discriminants ou biomarqueurs (dont les concentrations ou la présence varient d'une condition à l'autre).

Une analyse métabolomique commence par une phase de préparation des échantillons et aboutit à une liste de métabolites d'intérêt identifiés. Ce processus passe par une étape d'analyse spectrale (MS ou NMR, en général) des échantillons afin d'obtenir des empreintes métaboliques, une étape de traitement automatique des données permettant d'aboutir à une matrice de données qui pourra se prêter à des analyses statistiques, et par des analyses instrumentales complémentaires nécessaires pour identifier les métabolites. Cette pluridisciplinarité requiert des connaissances dans différents domaines : chimie analytique, chimie organique, statistiques, bio-informatique et biochimie (Figure 4).

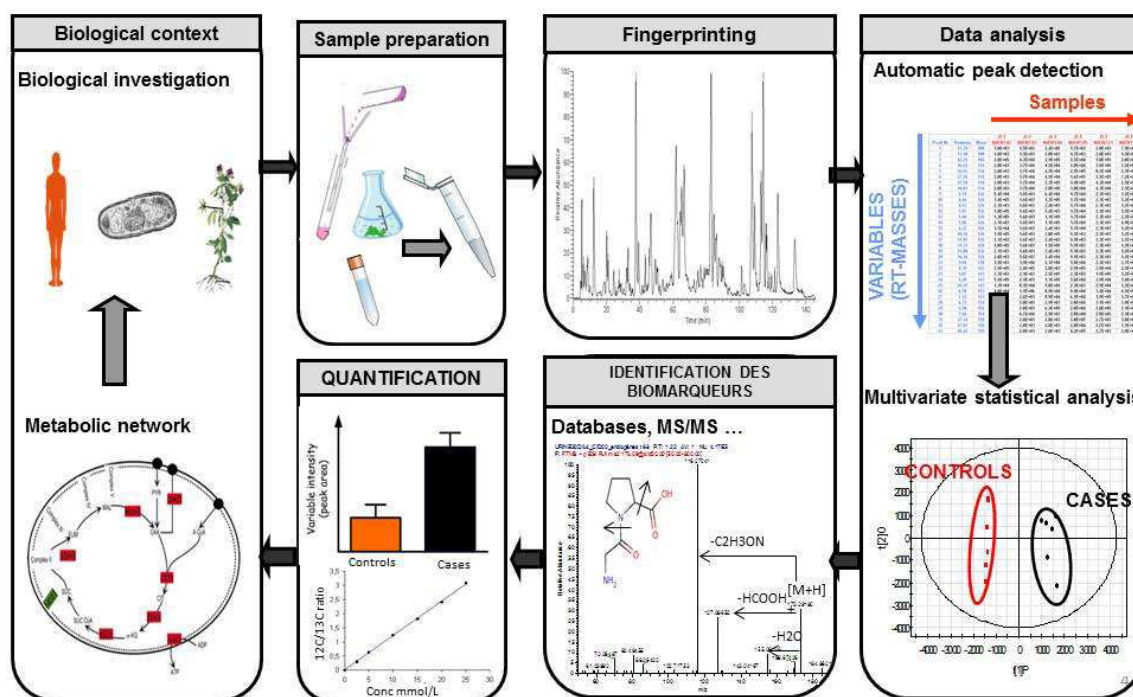


Figure 4. Déroulement d'une analyse métabolomique (Boudah, Paris, & Junot, 2013)

L'analyse métabolomique débute par la conception d'un protocole expérimental approprié. Après traitement des échantillons, s'en suit une analyse instrumentale (NMR ou MS). Les profils métaboliques sont ainsi obtenus (données brutes) puis traités par des logiciels spécialisés assurant la détection des signaux et les analyses statistiques. Enfin les signaux d'intérêt sont identifiés, et quantifiés. Leurs variations de concentrations sont interprétées biologiquement à partir de nos connaissances des voies métaboliques dans lesquelles ils sont impliqués

1. Démarche expérimentale

Toute analyse métabolomique est menée en vue de répondre à une question biologique. Tout d'abord, il faut s'assurer de bien concevoir son expérimentation afin d'empêcher, ou à minima de pouvoir gérer, la présence de facteurs confondants. Ces facteurs corrélerent avec des paramètres biologiques d'intérêt, conduisant à des interprétations biologiques erronées. Ils peuvent être d'origine analytique (par exemple, des variations de rendements de désorption d'une expérience à une autre ou au sein d'une même série d'analyse dû à l'encrassement de la source ESI), ou encore biologique (citons pour exemple l'indice de masse corporelle qui corréle avec la prévalence de maladies cardiovasculaires). Les facteurs confondants d'origine analytique et biologiques peuvent être gérés par la randomisation de l'ordre d'analyse des échantillons biologiques au cours d'une expérience, et également par des stratégies mathématiques qui mettent en œuvre des algorithmes de correction de l'intensité des signaux (Dunn et al; 2011). Reste la question des facteurs confondants non anticipés survenant dans le cadre d'une analyse métabolomique et qui est la plus délicate à résoudre.

2. Préparation des échantillons

La préparation des échantillons comprend différentes étapes : recueil des échantillons, l'arrêt du métabolisme et extraction des métabolites. En dépit de leur impact sur la qualité des résultats, les deux premières étapes sont souvent négligées alors que l'étape d'extraction proprement dite est bien documentée dans la littérature avec l'établissement de protocoles standards d'extraction de métabolites à partir des biofluides ou de cellules pour analyse LC-MS (Dunn et al; 2011,Want et al; 2010,Want et al; 2013).

Le succès d'une expérience métabolomique dépend non seulement des conditions LC-MS mais aussi implique des étapes en amont portant sur la manipulation de l'échantillon (par exemple, le choix de la matrice biologique adéquate, la collecte et l'acheminement des échantillons, l'extraction des métabolites et leur stabilité). Cela a été particulièrement souligné par Canelas et coll. dans leur étude dont l'objectif était de comparer deux conditions de croissance de levures au moyen de 5 méthodes différentes d'extraction. Ils ont observé que les métabolites détectés et leurs

concentrations dépendaient des protocoles d'extraction, ce qui pouvait conduire à des interprétations biologiques erronées (Canelas et al; 2009).

2.1. Recueil des échantillons

Différents types de fluides biologiques peuvent être analysés par une approche métabolomique, parmi lesquels le plasma, le sérum, l'urine ou le liquide céphalo-rachidien (LCR) sont les plus utilisés. Par exemple, l'étude du LCR est pertinente en neurologie en raison de sa proximité avec les cellules neuronales et gliales (Mochel et al; 2009, Sinclair et al; 2010), et les érythrocytes permettent d'étudier les altérations métaboliques chez les patients souffrant de maladies génétiques des cellules rouges du sang (Darghouth et al; 2011a, Darghouth et al; 2011b). Néanmoins, l'urine et le sang restent des liquides biologiques les plus largement utilisés.

L'urine est un liquide biologique composé d'une grande diversité de molécules organiques dont les taux sont amplifiés du fait de la concentration rénale. Son recueil est non invasif. Il peut être réalisé par des collectes ponctuelles ou sur des périodes allant de 24 à 48 heures. Dans ce dernier cas, l'utilisation de conservateurs est recommandée afin d'éviter toute contamination bactérienne pouvant modifier la composition métabolique des urines (Want et al; 2010). Or, des études de stabilité des urines humaines menées au laboratoire (Thèse Aurélie ROUX) ont démontré que bien que la contamination bactérienne soit limitée par les conservateurs, l'instabilité chimique de certains métabolites imposerait leur collecte dans une limite de 24h à +4°C.

Quant aux échantillons sanguins, ils sont centrifugés pour éliminer les cellules sanguines de manière à obtenir soit du plasma (par inactivation de la cascade de coagulation en utilisant des anticoagulants) ou du sérum (après coagulation du sang). Ces deux fluides biologiques semblent produire des résultats similaires dans les études cliniques et biologiques. De ce fait, ils peuvent être utilisés de façon interchangeable (Yu et al; 2011). En effet, une étude comparative des profils LC-MS menée par inspection visuelle et utilisation d'outils bioinformatiques a montré qu'il n'y avait pas de différences entre plasma et sérum (Denery, Nunes, & Dickerson, 2011).

Toutefois, quelques précautions restent à prendre. Concernant le sérum, en raison de l'activité enzymatique des globules rouges, des effets du temps de coagulation sur le métabolisme énergétique ont été mis en évidence par une analyse en composantes principales réalisée sur des spectres NMR (Teahan et al; 2006). Ces changements sont retardés à 4°C, suggérant à l'expérimentateur de manipuler le sérum sur de la glace. Par ailleurs, en ce qui concerne des échantillons de plasma, le choix de l'anticoagulant peut avoir un impact sur la détection de certains métabolites. Il n'existe toujours pas de consensus clair quant au choix de l'anticoagulant lors d'une étude de métabolomique. Par exemple, il est connu que l'héparine, qui contient du bisulfite de sodium comme agent de conservation, peut non seulement biaiser la quantification de la cystéine

et sulfocysteine due à la formation de sulfocysteine et homosulfocysteine (Dettmer et al; 2010, Parvy et al; 1989), mais également générer un bruit de fond susceptible de limiter la sensibilité des méthodes analytiques (Yin et al; 2013). De plus, Bando et coll. ont démontré en utilisant un GC-TOF que les pics produits par l'héparine interfèrent avec la détection de métabolites endogènes, contrairement à d'autres anticoagulants tels que l'EDTA (Bando et al; 2010). Ce dernier affecte notamment le profil protéique MS sans affecter le métabolome. L'utilisation de citrate n'est pas recommandée car ce composé est un métabolite impliqué dans le cycle de Krebs (Yin et al; 2013). Enfin, lorsqu'il est sélectionné, l'anticoagulant doit être maintenu le même dans une étude, et ce afin d'éviter des regroupements artefactuels d'échantillons dans le cadre d'analyses statistiques.

Les analyses métabolomiques peuvent également s'effectuer sur des cellules ou tissus. Leur préparation avant analyse consiste à isoler les cellules d'intérêt à partir de leur environnement biologique (milieu de culture et autres cellules) tout en préservant leur intégrité. Selon le type d'échantillon, la centrifugation ou la filtration sur des membranes appropriées peuvent permettre d'atteindre cet objectif. En ce qui concerne les échantillons microbiologique, des conditions stériles doivent être maintenues. Enfin, le lavage des cellules recueillies est recommandé pour éliminer le milieu ou composés extracellulaire (Meyer, Liebeke, & Lalk, 2010).

2.2. Arrêt forcé du métabolisme «notion de quenching»

Le «quenching» a pour but d'arrêter les réactions enzymatiques qui se produisent après la collecte de l'échantillon afin de mesurer la composition réelle du métabolome au moment de l'échantillonnage. Cette étape peut être ajoutée au processus de préparation d'échantillon car elle est particulièrement nécessaire pour les composés présentant un grand taux de renouvellement métabolique, comme c'est le cas avec les intermédiaires du métabolisme énergétique. L'utilisation de températures extrêmes (élevées ou basses, avec l'azote liquide), l'ajout d'acides ou le transfert de cellules en suspension dans un solvant organique froid peut arrêter l'activité enzymatique résiduelle par dénaturation protéique, mais ces techniques doivent être conformes au temps du processus métabolique d'intérêt (exemple turn-over) sans affecter la stabilité des métabolites (Vuckovic, 2012). Les procédures à base de solvants organiques froids sont les plus utilisées mais ont l'inconvénient de libérer des métabolites intracellulaires dans le milieu extracellulaire du fait de la lyse de la membrane cytoplasmique. Enfin, la question du quenching est régulièrement abordée en analyse métabolomique bactérienne ou encore chez les végétaux (Bolten et al; 2007) mais elle est souvent omise pour des approches globales réalisées dans les fluides biologiques. Ces biofluides requièrent pour leur collecte d'autres types de précautions telles que l'addition d'inhibiteurs de la glycolyse comme les ions fluorure qui inhibe l'enzyme de l'énolase et empêche ainsi la dégradation du glucose et la production de lactate (Mikesh & Bruns, 2008).

2.3. Extraction des métabolites

En raison du grand nombre et de la grande diversité chimique des métabolites détectés, l'expérimentateur doit définir un compromis entre la pureté de l'extrait qui entraîne une sensibilité optimale pour l'ionisation des molécules, et la couverture métabolique atteinte.

Bien que souvent négligée, cette étape conditionne la qualité des résultats obtenus notamment dans le cadre d'analyses ciblées où une extraction sélective des métabolites est de mise, en vue d'augmenter la sensibilité de détection (exemple de l'extraction de Folch pour l'analyse des lipides (Folch et al; 1951)). Cette sélectivité d'extraction est obligatoire pour la détection de certaines classes de métabolites présentant des propriétés physico-chimiques spécifiques, comme c'est par exemple le cas des lipides, ou pour la détection de métabolites présents à l'état de trace. En revanche, avec l'intérêt croissant pour les approches globales qui tentent de détecter le plus grand nombre de métabolites, le traitement de l'échantillon doit être réduit à son minimum en vue de limiter toute source de variabilité externe et surtout la perte d'information.

Cependant, il est toujours nécessaire d'éliminer les protéines qui pourraient détériorer la qualité de séparation chromatographique (colmatage de colonnes) mais aussi diminuer de la sensibilité de détection (majoration de l'impact de l'effet matrice et du bruit de fond). Pour ce faire, 4 catégories de techniques de traitement de l'échantillon sont envisageables : extraction sur phase solide, extraction liquide-liquide, précipitation par ajout de solvants organiques et les méthodes de filtration.

2.3.1. Extraction liquide-liquide

L'extraction liquide-liquide est basée sur les différences de solubilité des composés entre deux solvants non miscibles. Un partage a lieu entre le milieu biologique (considéré comme une phase aqueuse) et le solvant organique ajouté. Cette technique d'extraction est particulièrement utilisée dans les dosages ciblés dans des matrices complexes. Citons par exemple le dosage de médicaments tels que les benzodiazépines dans le sang total (Simonsen et al; 2010a, Simonsen et al; 2010b), d'agents de dopage dans les urines (Deventer et al; 2009) ou l'analyse des pesticides (Gilbert-Lopez et al; 2010). Souvent, un sel inorganique est ajouté à la phase aqueuse de façon à y diminuer la solubilité des composés organiques mais il faut s'assurer de la compatibilité avec la détection en spectrométrie de masse (Wu et al; 2008).

De plus, l'extraction liquide-liquide avec un mélange de chloroforme et de méthanol (BLIGH & DYER, 1959, Folch et al; 1951) reste largement utilisée en analyses lipidomiques. Elle génère un système biphasique dans lequel la phase inférieure contient les lipides totaux. Enfin, Eikel et coll ont réussi à combiner la micro-extraction depuis une surface solide avec une source nano-électrospray. Cette technique, dénommée "Liquide Extraction Surface Analyse–Mass Spectrometry" (LESA -MS), a

démonstré qu'elle était plus informative que l'imagerie MALDI pour les études de distribution et de biotransformation de médicaments sur des coupes de tissus de souris (Eikel et al; 2011). Un tel échantillonnage *in vivo* est intéressant, car il réduit les variations analytiques inhérentes aux procédés multi-étapes et laisse l'échantillon dans son environnement biologique.

2.3.2. Extraction en phase solide (SPE) et microextraction en phase solide (SPME)

L'extraction en phase solide consiste à adsorber les composés à extraire sur une phase stationnaire contenue dans une cartouche puis à les récupérer lors de l'élution après des lavages successifs permettant d'éliminer les composés gênants. Elle est un outil précieux pour la purification d'échantillon et la pré-concentration d'analytes notamment ceux présents à l'état de trace dans le cadre d'analyses métabolomiques ciblées. Ces approches semblent être moins appropriées à la métabolomique non ciblée car la couverture métabolique est limitée par la sélectivité qu'apporte le matériau adsorbant (Daykin et al; 2002). En fait, Vuckovic et coll. ont récemment évalué les performances d'extraction dans le plasma humain de 42 phases stationnaires de SPME pour l'analyse de 36 métabolites appartenant à différentes classes chimiques et de diverses polarités. La SPE et la SPME réduisent la suppression d'ionisation observée en détection par MS, et certaines phases permettent d'améliorer la couverture métabolique plasmatique (Vuckovic & Pawliszyn, 2011).

Une variante récente de la SPE dite «Turbulent Flow Chromatography» (TFC) a été développée pour des analyses à haut débit. Elle assure l'extraction de métabolites en ligne sur phase stationnaire à grosses particules sous flux turbulent depuis des matrices biologiques non traitées. En TFC, seules les petites molécules sont retenues puis éluées directement dans la colonne d'analyse. Des systèmes en ligne d'extraction à l'aide de TFC ont déjà été utilisés dans le dépistage toxicologique ciblé en LC-MS/MS dans les urines (Mueller et al; 2011). La TFC a également été utilisée avec succès pour une analyse globale en LC-MS d'échantillons de sérum et de plasma (Michopoulos et al; 2010a).

2.3.3. Précipitation protéique :

Le traitement des échantillons biologiques avec des solvants organiques, de fortes concentrations en sel, ou l'addition de bases ou d'acides forts induisent des altérations de la structure tridimensionnelle des protéines et donc leur précipitation du fait de leur faible solubilité. Want et coll. ont évalué la précipitation protéique dans le plasma au moyen d'acides, de la chaleur ainsi que différents solvants organiques dont le méthanol (MeOH), l'éthanol (EtOH), l'acétonitrile (ACN) et l'acétone. La méthode impliquant l'ACN était la moins reproductible, alors que le traitement des échantillons avec des acides forts et la chaleur permettait de détecter le plus grand nombre de molécules. En ce qui concerne la reproductibilité et le nombre de variables détectées en MS, la

précipitation des protéines au MeOH s'est avérée être l'approche la plus efficace (Want et al; 2006). Bruce et coll. ont évalué la quantité de protéines restantes après la précipitation aux solvants organiques en utilisant de l'électrophorèse sur gel. Ils ont conclu que la combinaison de MeOH/ Acétone/ ACN dans des volumes égaux était la plus appropriée pour éliminer les protéines et ce qui étendait la durée de vie de la colonne (Bruce et al; 2009). Or, certaines molécules peuvent ne pas être récupérées dans les surnageants organiques du fait qu'elles soient liées à des protéines ou encore qu'elles ne soient pas solubles dans le solvant organique utilisé (exemple des lipides neutres). Cette problématique n'a pas été évaluée par ces études, et seule l'équipe de l'Imperial College of London l'a abordée en comparant la distribution de la densité des variables selon leur temps d'élution en chromatographie à polarité de phase inversée et ce en considérant le caractère hydrophobe des lipides (Want et al; 2006). Cependant, la plupart des lipides polaires tels que lysophospholipides et gangliosides pourraient être extraits par des solvants organiques sans extraction adaptée.

L'échantillonnage par «goutte de sang séché» (Dried Blood Spot ou DBS, en anglais) est une technique émergente dans le domaine de la métabolomique. Elle est bien adaptée aux petits volumes, et a longtemps été utilisée aux Etats-Unis dans le dépistage des erreurs innées du métabolisme chez les nouveaux nés (Chace et al; 2009b, Chace et al; 2009a, Chace et al; 2009c). En bref, une goutte de sang total, de plasma ou d'urine est adsorbée sur du papier filtre puis séchée. La tâche qui en résulte est extraite par un solvant organique approprié. En utilisant une telle approche d'extraction, les empreintes LC-MS obtenues par Michopoulos et coll. étaient proches (sur le plan chimométrique) de celles obtenues par précipitation classique, sauf pour certains composés acides urinaires (Michopoulos et al; 2010b). Notons que les effets liés au bruit de fond induit par le papier utilisé et la stabilité des échantillons sont mal connus et doivent être abordés dans le cadre du développement de méthodes.

2.3.4. Ultrafiltration :

L'ultrafiltration est une technique qui utilise des membranes semi-perméables de diamètre prédéterminé pour retenir des macromolécules contenues dans un liquide. Cette technique peut être utilisée dans la préparation d'échantillons complexes notamment pour s'affranchir des protéines et lipoprotéines contenues dans un biofluide tel que démontré pour les protéines sériques en 1H – NMR (Daykin et al; 2002). Bien que cette technique permette d'assurer une plus grande sensibilité pour certaines classes de métabolites, certaines espèces hydrophobes peuvent être adsorbées sur la membrane de filtration. Pour résoudre ce problème, Tiziani et coll. proposent d'ajouter une étape de lavage de filtre et d'extraction au chloroforme après ultrafiltration afin de désorber les métabolites depuis le filtre (Tiziani et al; 2008).

Bien qu'il n'y ait pas de consensus concernant l'extraction des métabolites depuis les fluides biologiques, certaines tendances sont clairement observées. Dans la plupart des études, les échantillons d'urine sont traités par simples dilution dans l'eau (approche de type «dilute-and-shoot») (Plumb et al; 2003, Roux et al; 2012, Williams et al; 2005). En ce qui concerne le plasma et le sérum, les procédés de précipitation à base de solvants organiques sont les plus largement utilisés. L'acétonitrile réalise le meilleur taux de précipitation protéique, mais il a été rapporté qu'il présentait une mauvaise reproductibilité. Des mélanges contenant du méthanol ou le méthanol pur assurent une meilleure couverture métabolique (A J et al; 2005, Bruce et al; 2009, Want et al; 2006). Enfin, pour la métabolomique cellulaire, les protocoles sont généralement optimisés pour des études spécifiques et il n'existe pas de protocoles génériques (Merlo et al; 2011). A cela, peut s'ajouter une étape de pré-concentration. De plus, la rupture des membranes cellulaires est nécessaire pour les analyses du métabolome intracellulaire. Elle est obtenue soit mécaniquement à l'aide des billes ou bien avec des ultrasons (Meyer, Liebeke, & Lalk, 2010), ou encore à l'aide de cycles de congélation/décongélation, ou par transfert des cellules dans un milieu hypotonique comme de l'eau bouillante (Darghouth et al; 2011a, Darghouth et al; 2011b).

Une fois extraits, les échantillons (ou extraits) sont repris dans un solvant compatible avec le système analytique avant analyse. Plus en amont, la collecte même des échantillons doit respecter un protocole adapté à la thématique abordée et au milieu biologique étudié (choix d'un anti-coagulant pour des échantillons plasmatiques, ou ajout d'un conservateur pour des urines prélevées sur 24 heures, par exemple) et les conditions de conservation avant analyse LC-MS doivent être maîtrisées.

2.4. Conservation et stabilité des échantillons

Après l'extraction des métabolites à partir d'échantillons biologiques, les extraits sont souvent aliquotés (afin d'éviter les cycles de congélation-décongélation), séchés et conservés à -80 °C afin d'en minimiser la dégradation. Ils sont reconstitués extemporanément dans un solvant approprié, compatible avec les conditions d'analyse LC-MS, et qui assure la solubilisation des métabolites extraits. Il est essentiel d'évaluer les conditions de conservation et la stabilité des échantillons bruts et extraits avant l'analyse MS. Cependant, peu d'études de stabilité à long terme de métabolites dans les fluides biologiques sont disponibles. L'inspection visuelle des données LC-ESI/MS en PCA ont prouvé que les urines étaient stables sans conservateur après 1, voire même 6 mois de stockage à -20 °C ou -80 °C, mais ne le sont pas après 6 jours à 4°C (Gika et al; 2007, Gika, Theodoridis, & Wilson, 2008). Cependant, l'analyse en composantes principales (PCA) n'est qu'une représentation synthétique de l'ensemble de données, et certains métabolites, par exemple ceux présents à l'état de trace, peuvent présenter des variations de concentration ayant un impact limité

sur la structure d'une carte factorielle de PCA. Il est donc nécessaire d'examiner les questions de stabilité aux niveaux de métabolites annotés ou identifiés. Dans cette optique, Dunn et coll. ont procédé à l'analyse GC-TOF-MS d'extraits sériques et urinaires provenant de 40 volontaires sains et ayant été conservés à +4°C pendant 24h avant leur congélation à -80°C. Des analyses statistiques multivariées et univariées (test de Wilcoxon) ont été réalisées sur 700 variables attribuées aux 107 métabolites endogènes et exogènes identifiés (masse précise et index de temps de rétention équivalents à ceux répertoriés dans leur base de données interne). Les résultats ont prouvé que la variabilité inter-individuelle prédomine sur la variabilité intra-individuelle. De plus, après comparaison à des répliques analytiques, la variance due au temps de conservation à +4°C ne pouvait pas être distinguée de celle liée à la variabilité analytique globale.

De plus, la question de l'impact des étapes pré-analytiques telles que la conservation, la reproductibilité à long terme ou les cycles de congélation et de décongélation sur des échantillons plasmatiques a également été abordée récemment par analyses LC-MS (Yin et al; 2013) ou LC-MS/MS (Townsend et al; 2013, Yin et al; 2013). Ces études confirment la robustesse des profils plasmatiques et la possibilité de ce fait de recourir à des études prospectives dans un contexte d'analyse de cohortes.

2.5. Evaluation du processus de préparation des échantillons :

En analyse métabolomique globale, une méthode d'extraction idéale permet de récupérer le maximum de métabolites (tant quantitativement que qualitativement) dans un extrait purifié (délesté des macromolécules) tout en perturbant le moins possible le métabolisme basal de l'organisme étudié afin de rendre compte le plus fidèlement possible de l'état métabolique. Dans la plupart des publications, l'évaluation de l'efficacité d'extraction est restreinte au critère qualitatif de présence ou d'absence du ou des métabolites d'intérêt. Ceci se fait au niveau chimométrique, c'est à dire, en considérant le nombre de variables reproductibles détectées (qui sont des métabolites potentiels) (Bruce et al; 2009, Want et al; 2006), et/ou en incluant une liste non exhaustive de métabolites connus (A J et al; 2005, Daykin et al; 2002, Vuckovic & Pawliszyn, 2011) et/ou des étalons internes marqués (IS, de l'anglais «internal standard») ajoutés aux échantillons avant extraction (A J et al; 2005, Want et al; 2006). Ces paramètres ne peuvent pas rendre compte de la qualité globale des données du fait de leur manque de représentativité en termes de couverture métabolique. Néanmoins ils peuvent être considérés comme un premier niveau d'évaluation.

Il est nécessaire d'établir des procédures de contrôle qualité pour évaluer au mieux la préparation des échantillons. Comme proposé par Vuckovic et coll. les méthodes d'extraction peuvent être tout d'abord évaluées quant à la réduction de l'effet matrice qu'elles entraînent. Des tests d'analyse en introduction directe ou encore de dilutions en série des échantillons peuvent de ce

fait évaluer cette suppression d'ionisation (Vuckovic, 2012). En plus du calcul des rendements d'extraction des IS (Bruce et al; 2009), avoir la même démarche sur des métabolites endogènes serait un gage du bon déroulement de l'extraction au niveau global. Ceci est rendu possible par l'utilisation de matériaux biologiques de référence tels que le SRM 1950 pour le plasma humain (Simon-Manso et al; 2013). De plus, certains paramètres biologiques connus pour être stables devraient être vérifiés et considérés comme gage de qualité du processus de préparation d'échantillons. C'est par exemple le cas avec le métabolisme énergétique des bactéries qui offre une métrique connue sous le nom de charge énergétique («CE») et correspond à la formule suivante : $([ATP] + 0,5 [ADP]) / ([ATP] + [ADP] + [AMP])$. Elle indique l'énergie disponible dans la cellule étudiée et ses valeurs sont généralement situées autour de 0.9 (Vuckovic, 2012).

3. Les méthodes de spectrométrie de masse en analyse métabolomique

Le choix de la plateforme analytique utilisée au cours d'une analyse métabolomique est difficile. En effet, l'instrumentation sélectionnée doit permettre non seulement (i) la détection du plus grand nombre possible de métabolites dans un le milieu biologique étudié, mais aussi assurer (ii) leur identification et (iii) leur quantification.

3.1. Profilage métabolique par spectrométrie de masse

3.1.1. Introduction : les principales méthodes d'analyse du métabolome

La taille exacte du métabolome reste à ce jour indéterminée, et cela même pour les organismes les mieux connus. Néanmoins, certains auteurs l'estiment à plus de 3000 métabolites chez l'Homme (Duarte et al; 2007), et plus de 200000 dans le règne végétal (Fiehn, 2001). De plus, il est difficile d'intégrer l'ensemble des sources de variations susceptibles d'avoir des répercussions sur sa composition : rythme circadien, influence de l'environnement, variabilité inter-individuelle, pathologie et autres. Ceci sous-tend non seulement une grande diversité structurale impliquant des différences de propriétés physico-chimiques importantes (poids moléculaire, hydrophobicité, pKa..), mais aussi une large gamme dynamique de concentration allant très souvent du nanomolaire au millimolaire (6 ordres de grandeur), ce qui rend l'analyse simultanée de l'ensemble du métabolome très difficile (Dunn, Bailey, & Johnson, 2005) (voir Figure 5).

L'absence d'une méthode universelle d'analyse constitue la limitation majeure des analyses métabolomiques (Han et al; 2009) et explique la diversité des techniques analytiques mises en œuvre. Toutes fournissent une information chimiométrique en vue de classer les échantillons. Citons par exemple la spectroscopie infra-rouge (IR) (Wharfe et al; 2010), la résonance magnétique nucléaire (NMR) (Smith et al; 2009), la spectrométrie de masse non couplée (MS)(Fuhrer et al; 2011),

la chromatographie en phase liquide couplée à la spectrométrie de masse (LC-MS) (Veselkov et al; 2011), la chromatographie en phase gazeuse couplée à la spectrométrie de masse (GC-MS)(Qi et al; 2011), et l'électrophorèse capillaire couplée à la spectrométrie de masse (CE-MS) (Ramautar et al; 2012)). Certaines de ces techniques peuvent en plus apporter des informations structurales permettant l'identification des molécules responsables de cette classification chimométrique. C'est le cas des techniques basées sur (i) la spectroscopie par NMR et (ii) la MS.

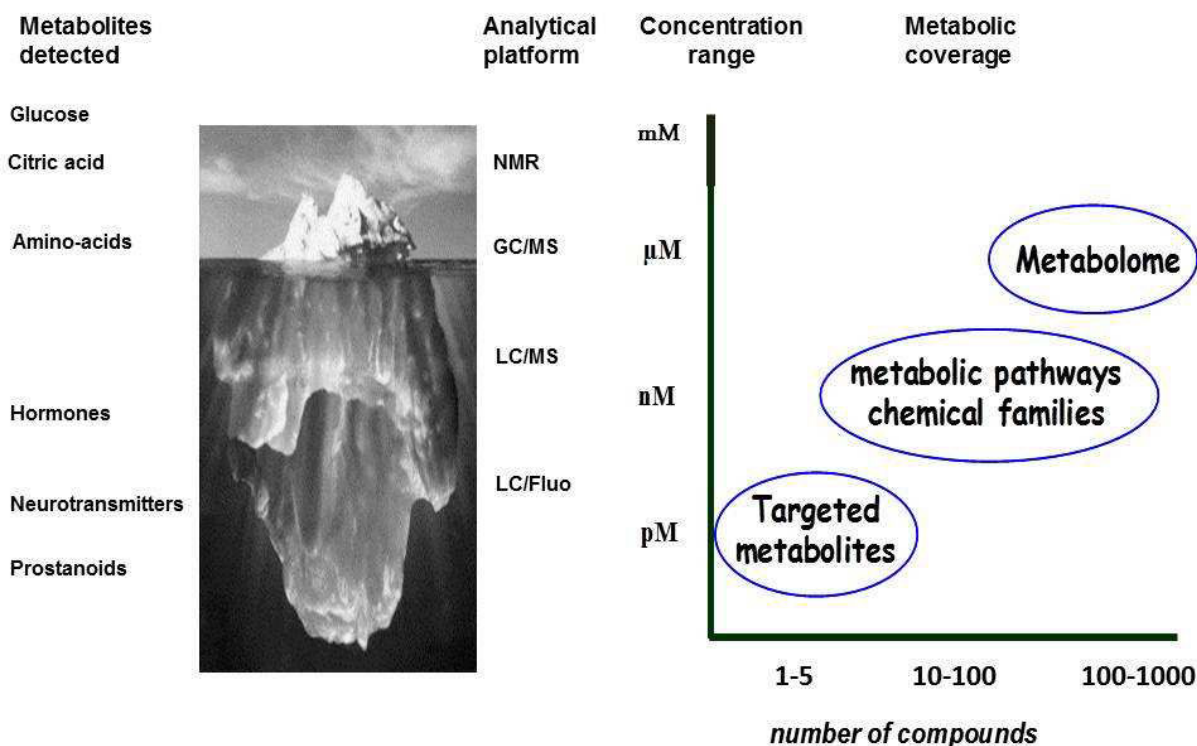


Figure 5. Enjeux technologiques en analyse métabolomique (adapté de Sumner LW et coll., 2003 (Sumner, Mendes, & Dixon, 2003))

La spectroscopie par NMR du proton ($^1\text{H-NMR}$) bénéficie d'une vingtaine d'années de recul dans le domaine de la toxicologie animale (Nicholson, Lindon, & Holmes, 1999), sa mise en œuvre est simple, rapide (requiert une préparation minimale des échantillons à savoir une simple dilution dans une solution tampon), non destructive et sa bonne reproductibilité est tout à fait adaptée aux analyses à haut débit. La $^1\text{H-NMR}$ et $^{13}\text{C-NMR}$ permettent non seulement la détection simultanée de différents métabolites dans un milieu, mais également la quantification et l'obtention de leurs structures moléculaires. Cependant, l'essor de cette méthode relativement limité à ce jour est attribuable à la fois aux difficultés d'accès à la technique de NMR (coût d'acquisition de l'équipement élevé), à son manque de sensibilité, mais aussi à la complexité inhérente à l'analyse des spectres. A titre d'exemple, l'exploration du métabolisme lipidique par $^1\text{H-NMR}$ reste délicate à cause de la formation de multiplets indiscernables. La spectroscopie $^{31}\text{P-NMR}$ présente les mêmes avantages et

inconvenients que les déclinaisons précédentes de la NMR, mais son application se restreint à l'analyse des composés phosphorylés tels que les phospholipides et les nucléotides.

Bien que la NMR ait longtemps été considérée comme la technique la plus aboutie pour mener des études métabolomiques, la MS, de par la multitude des sources d'ionisation et des analyseurs disponibles, a permis d'étendre la gamme des molécules détectées. Elle est désormais devenue une méthode de choix pour répondre à un grand nombre de questions biologiques. Dans ce qui suit, nous aborderons les principales avancées en analyses métabolomiques réalisées en LC-API/MS, nous discuterons des principales limitations qui restent à surmonter pour que cette approche puisse être utilisée à grande échelle en biologie et ferons le point sur les solutions pour y parvenir.

3.1.2. La spectrométrie de masse, une technologie polyvalente

La MS est une technique d'analyse qui permet la mesure de molécules après leur ionisation. Une fois qu'un échantillon a été introduit dans le spectromètre de masse (directement ou après séparation), les molécules sont ionisées et les ions résultants sont séparés en fonction de leur rapport de masse sur charge (m/z). Un spectromètre de masse est composé de 5 modules (Figure 6) : un système d'introduction de la substance à analyser, une source pour ioniser les analytes, un ou plusieurs analyseurs fonctionnant à un vide poussé pour séparer les différents ions générés dans la source en fonction de leur rapport m/z , un détecteur pour dénombrer les ions et enfin, un système de traitement des données permettant la visualisation et l'enregistrement des spectres de masse ainsi que le contrôle des paramètres d'acquisition.

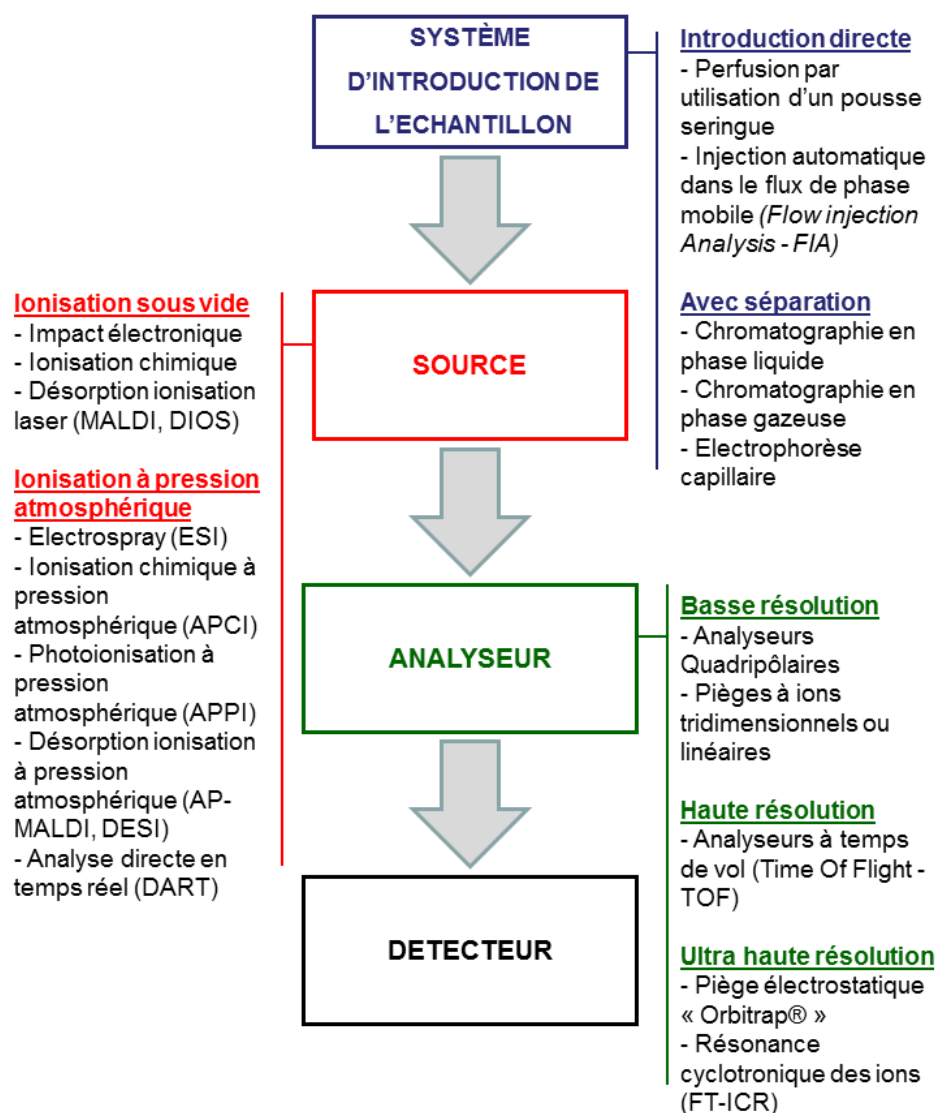


Figure 6. Schéma global d'un spectromètre de masse (adapté de (Boudah, Roux, & Junot, 2012))

Les sources utilisées en spectrométrie de masse sont basées sur l'ionisation en phase gazeuse ou sur la désorption-ionisation d'analytes. L'ionisation en phase gazeuse peut être réalisée sous vide à l'aide de sources à impact électronique (EI) ou à pression atmosphérique (API, de l'anglais « Atmospheric pressure ionization » : sources APCI (Horning et al; 1973), APPI (Marchi, Rudaz, & Veuthey, 2009, Robb, Covey, & Bruins, 2000) et DART (Cody, Laramée, & Durst, 2005)). La désorption-ionisation d'analytes peut, quant à elle, s'opérer sous vide (sources DIOS (Wei, Buriak, & Siuzdak, 1999)) ou encore sur des supports solides à l'aide de matrices (c'est le cas des sources MALDI (Tanaka et al; 1988)), ou à pression atmosphérique sur des nébulisats (sources électrospray ESI et, DESI, par exemple (Fenn et al; 1989)). Parmi ces sources, l'électrospray (ESI) est l'une des sources les plus utilisées.

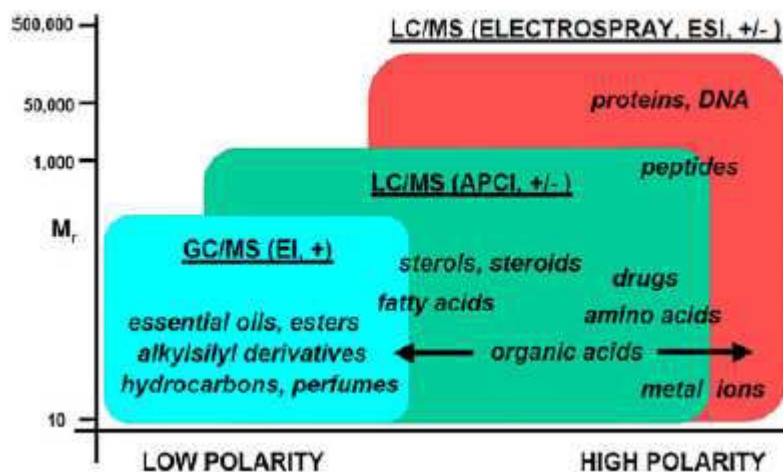


Figure 7. Sources d'ionisation et molécules détectées correspondantes (Kind & Fiehn, 2010)

Parmi ces sources d'ionisation, les spectromètres de masse à source d'ionisation par impact électronique sont utilisés depuis longtemps pour le profilage métabolique. Ils sont couplés à des chromatographes en phase gazeuse (GC-MS) (Horning & Horning, 1971) qui offrent une très bonne résolution chromatographique, une détection spécifique des analytes, et la possibilité de les quantifier avec une bonne sensibilité. L'énergie transmise aux analytes en ionisation électronique étant très importante (et présentant une large distribution), elle conduit à une fragmentation très reproductible, qui a permis de constituer des bibliothèques de spectres de masse universelles. L'existence de ces bases de données regroupant les spectres de plus de 100 000 composés (celle du *National Institute of Standards and Technology* (NIST) par exemple), ainsi que l'obtention directe d'information structurale, facilite l'identification des métabolites (Pasikanti, Ho, & Chan, 2008). Cependant, ces sources ne permettent pas toujours d'observer la masse du composé intact mais seulement les ions issus de sa fragmentation ce qui rend délicat toute identification de composés non décrits dans ces bases de données. De plus, l'analyse par GC-MS requiert que les analytes soient thermostables et volatiles. Ainsi, il est souvent nécessaire d'effectuer une étape de modification chimique pour améliorer la volatilité des composés polaires et peu volatiles tels que ceux dont le poids moléculaire dépasse les 400 Da. Ce traitement de l'échantillon est assez contraignant et peut conduire à des réactions de dégradation des analytes.

L'introduction de sources conçues pour fonctionner à pression atmosphérique comme l'ESI a étendu la gamme des molécules détectées aux peptides, protéines et lipides sans obligation de modification chimique. Elles assurent une ionisation douce, ceci permet la détection des espèces cationisées (par exemple, protonées $[M+H]^+$) ou anionisées (par exemple, déprotonnées $[M-H]^-$) donnant accès aux masses moléculaires des molécules intactes. Cependant, les ions produits dans les sources API fournissent des informations structurales limitées. Il est donc nécessaire d'effectuer des

fragmentations par activation des ions en mode MS/MS ou expériences séquentielles MSⁿ pour la caractérisation structurale des composés.

D'autre part, différentes technologies d'analyseurs peuvent être mises en œuvre en ESI-MS : les pièges à ions qui au cours du temps peuvent stocker des ions sélectionnés puis les fragmenter. Les ions fils produits pouvant ainsi à leur tour être fragmentés. Ces fragmentations par CID sont réalisées de manière temporelle et séquentielles (MSⁿ) permettent d'établir une filiation et remonter à la structure de la molécule en mode CID résonant. D'autres analyseurs fortement répandus notamment en analyses ciblées sont les analyseurs quadripolaires qui, montés en série, assurent quant à eux des expériences MS/MS dans l'espace (Werner et al; 2008b) en mode CID non-résonant. Ces analyseurs offrent la possibilité de fonctionner selon différents modes de balayage de type MS/MS à savoir le mode SRM, le balayage d'ions fragments, des ions précurseurs (PIS de l'anglais Precursor Ion Scanning) ou le balayage de pertes de neutres (NLS pour Neutral Loss Scan). Ces différentes configurations permettent à la fois d'obtenir des informations structurales et également de détecter sélectivement des familles métaboliques sur la base de groupements chimiques communs (Figure 8).

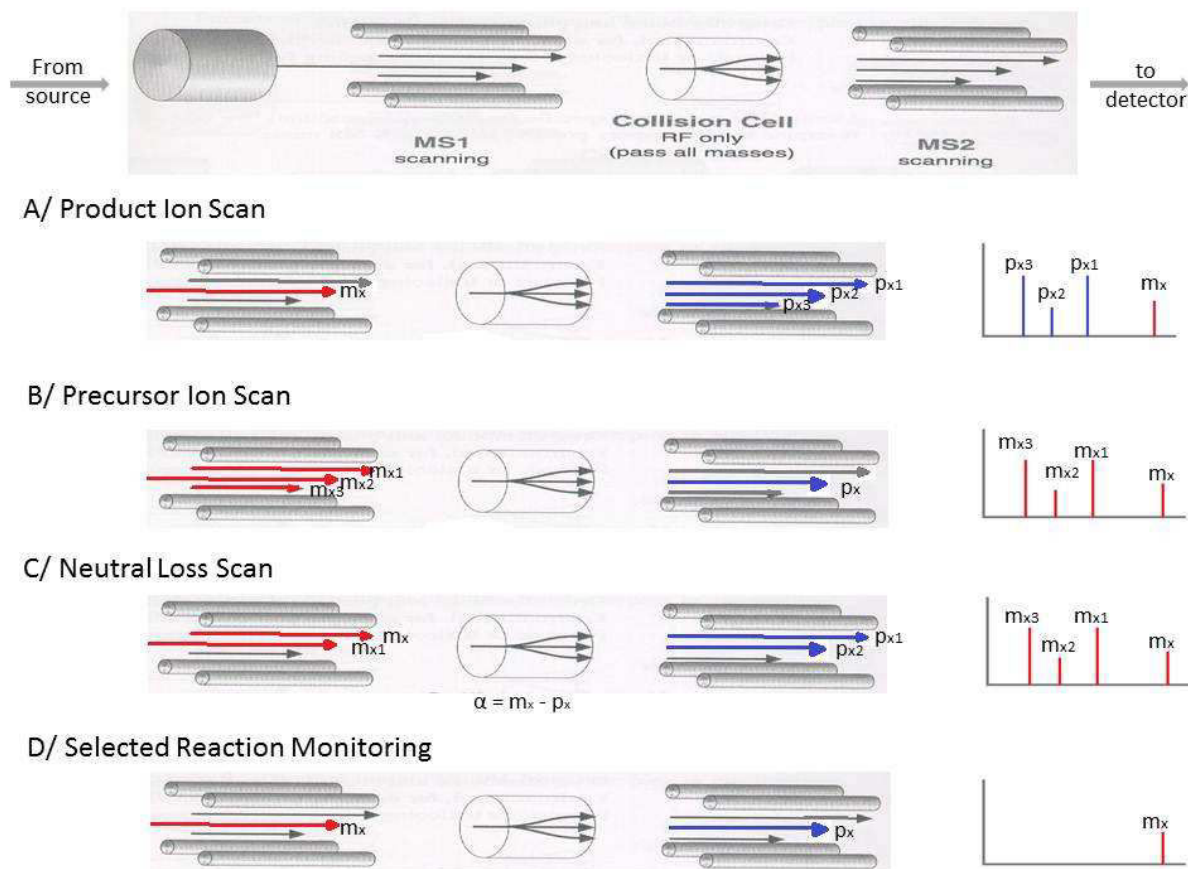


Figure 8. Différents modes d'acquisition en MS/MS sur des spectromètres de masse équipés d'un analyseur triple quadripolaire ((Boudah, Paris, & Junot, 2013)).

Le principe est d'isoler un ion donné (ion précurseur) dans le premier quadripôle et à le fragmenter par collisions dans le deuxième, afin d'obtenir le spectre de masse des « ions fils » dans le troisième quadripôle. Il existe plusieurs modes d'acquisitions de spectres CID dont les principaux sont l'acquisition (i) en mode balayage de fils (seul l'ion précurseur va se fragmenter et les ions fils sont séparés et analysés dans le troisième quadripôle Q) (ii) en mode balayage de parents (PIS) (la séparation a lieu cette fois-ci dans le premier Q1 qui permet de trouver le ou les précurseurs d'un ion fils donné) (iii) en mode balayage de perte de neutres (balayage en parallèle avec le premier et le troisième Q, mais en imposant entre les deux un décalage en m/z correspondant au neutre dont on veut détecter la perte. On détecte ainsi tous les ions, qui en se fragmentant, perdent un neutre de masse constante) et l'acquisition (iv) en mode de détection de réactions sélectionnées, encore appelée SRM (de l'anglais Selected Reaction Monitoring).

Les différentes combinaisons possibles entre les sources et les analyseurs permettent d'effectuer un large panel d'analyses tant ciblées que globales qui assurent la détection et/ou identification et/ou quantification de molécules diverses. En effet, les analyses peuvent être acquises sur une large gamme de rapports m/z (détection de type «Full scan», en anglais). Dans ce cas, tous les ions formés en source sont donc détectés, permettant ainsi une approche globale de profilage. Par contre, dans le cadre d'analyses ciblées, l'analyseur peut-être focalisé sur la détection de quelques ions d'intérêt (modes SIM, de l'anglais «Selected Ion Monitoring» ou SRM, dans le cas d'acquisitions de type MS/MS) tel qu'illustré en Figure 8. Grâce à cette polyvalence, la MS est devenue la méthode de choix dans le domaine de la métabolomique.

Néanmoins, l'efficacité de l'approche par MS a longtemps été limitée par la difficulté d'identifier rapidement et formellement les métabolites d'intérêt biologique. La raison principale de cette limitation résidait dans l'impossibilité de faire des mesures de masses suffisamment précises (la séparation des pics est de l'ordre de l'unité de masse pour les analyseurs à basse résolution). Ceci rendait complexe l'interrogation des bases de données biochimiques (plus d'une centaine de composés peuvent par exemple être référencés dans des bases de données chimiques pour une masse moléculaire de l'ordre de 350 Da). C'est alors que des approches semi-globales destinées à se focaliser sur certaines familles chimiques de métabolites (glucuronides, acylcarnitines, sulfoconjugués, phospholipides) se sont développées en analyse métabolomique. Ces méthodes sont basées sur l'utilisation de spectromètres de masse à analyseurs triple quadripôles fonctionnant en mode de détection de réactions sélectionnées (Bajad & Shulaev, 2007, Bajad et al; 2006, Brownsill et al; 1994, Kitteringham et al; 2009, Lafaye et al; 2004, McClellan, Quarmby, & Yost, 2002, Yoon, Kim, & Cho, 2003). Bajad et coll ont par exemple réussi à détecter 141 métabolites polaires d'*Escherichia coli* en LC-MS/MS et à en quantifier 61. Les approches globales avaient ainsi laissé place à des analyses ciblées et semi-globales plus sensibles, mais ne permettant pas d'accéder à d'autres métabolites ou familles chimiques que celles préalablement sélectionnées par l'utilisateur (Bajad et al; 2006).

Les analyseurs à haute résolution (HR) ont remis à l'ordre du jour les approches globales (Madalinski et al; 2008) telles qu'elles avaient été envisagées au début des années 2000 en les rendant plus informatives et plus sensibles. En effet, la HR permet d'accéder à des mesures de

masses précises avec des erreurs de l'ordre du ppm (analyseurs à temps de vol et à résonance cyclotronique des ions à transformée de Fourier, et plus récemment les pièges de type Orbitrap). Grâce à ces instruments, il est possible de détecter des métabolites de même masse nominale (dits métabolites isobares) dans des milieux biologiques complexes et également d'obtenir leur composition élémentaire différente, ce qui facilite l'interrogation des banques de données pour leur identification ultérieure (Aharoni et al; 2002).

Les analyses métabolomiques par spectrométrie de masse à source ESI peuvent être réalisées selon 2 approches : (i) en introduction directe et (ii) après couplage à des méthodes séparatives.

3.1.3. Introduction directe en spectrométrie de masse

L'introduction directe (DIMS de l'anglais «Direct Introduction Mass Spectrometry») permet des analyses d'extraits bruts, rapide et à haut débit. L'échantillon est introduit dans le spectromètre par une pompe à seringue (infusion MS) ou à l'aide de la pompe du système chromatographique («Flow Injection Analysis»). La DIMS a été utilisée dans les années 90 pour le profilage des acides aminés, des acides organiques et les dérivés de carnitine pour le diagnostic des erreurs innées du métabolisme chez les nouveaux nés (Chace, 2001), puis plus récemment, dans le domaine de la métabolomique végétale (Castrillo et al; 2003). De nos jours, elle est le plus souvent subordonnée à l'utilisation de spectromètres de masse à haute ou ultra haute résolution qui permettent la séparation de composés isobares (Madalinski et al; 2008). Cependant, cette approche d'introduction directe présente certaines limitations : (i) la présence d'un «effet matrice» qui provoque souvent une suppression du rendement d'ionisation et dont l'importance est molécule-dépendante, (ii) elle rend difficile l'attribution des ions fragments et/ou ions adduits en source en ne les distinguant pas des espèces moléculaires et (iii) elle ne permet pas de distinguer les isomères. Cela justifie la nécessité de disposer d'une dimension séparative supplémentaire amenée par la chromatographie liquide (couplage LC-MS). Notons que l'électrophorèse capillaire couplée à la spectrométrie de masse (CE-MS) représente une alternative intéressante pour l'analyse de métabolites très polaires non retenus dans des systèmes chromatographiques à polarité de phase inversée (Ramautar et al; 2011a, Ramautar et al; 2011b, Ramautar et al; 2012, Ramautar, Somsen, & de Jong, 2012, Robledo & Smyth, 2014). Elle est cependant pour l'instant moins sensible et moins robuste. Des couplages avec des cellules à mobilité ionique ont également été décrits (Henderson et al; 1999), mais ces dernières ne réduisent pas l'effet matrice et ne séparent que les isomères présentant une section efficace de mobilité différente, d'où leur popularité grandissante en analyse de lipides (Jackson et al; 2008, Kliman, May, & McLean, 2011).

3.1.4. Couplages chromatographie liquide et spectrométrie de masse

Les couplages entre la chromatographie en phase liquide et la spectrométrie de masse sont apparus plus tardivement que ceux avec la chromatographie en phase gazeuse en raison des obstacles technologiques liés à leur réalisation. La principale limitation était la nécessité d'évaporer la phase mobile chromatographique afin de faire passer les analytes en phase gazeuse avant leur introduction dans le spectromètre de masse. Ces problèmes ont été résolus avec le développement de méthodes d'ionisation à pression atmosphérique telles que l'ESI (Edmond de Hoffmann & Vincent Stroobant, 2007) qui présentent un positionnement de l'écrêteur permettant de séparer les ions du solvant de ceux d'intérêt.

La chromatographie en phase liquide complète le pouvoir résolutif amené par la spectrométrie de masse. Elle présente l'avantage d'être applicable à l'analyse de composés non volatiles et/ou thermosensibles sans qu'il soit nécessaire de mettre en œuvre des procédures fastidieuses de traitement de l'échantillon. Son principal intérêt est de limiter le phénomène de suppression d'ionisation observé avec les sources ESI et du à la complexité des milieux biologiques analysés. En outre, elle peut permettre de séparer des molécules isobares et éventuellement des isomères.

On distingue les systèmes chromatographique à haute (conventionnels) et à ultra-haute pression (UHPLC, de l'anglais «*ultra-high pressure liquid chromatography*»). D'apparition plus récente (depuis 2005), les systèmes de type UHPLC reposent sur l'utilisation de particules de phases stationnaire de diamètre réduit (<2 μ m) supportant de plus hautes pressions donc des débits plus élevés et des séparations plus rapides. Les pics chromatographiques élués sont alors plus fins et plus élevés, ce qui augmente le rapport signal/bruit, la capacité de pics et l'efficacité de la colonne (Wilson et al; 2005b, Wilson et al; 2005a). De plus, la possibilité d'utiliser des gradients avec des pentes plus importantes rend les séparations plus rapides (Leandro et al; 2006). Une analyse LC d'échantillons sériques a démontré que l'UPLC assurait une couverture métabolique plus étendue de 20% que les LC conventionnelles (Nordstrom et al; 2006). Cependant, la finesse des pics chromatographiques observés doit être compatible avec la vitesse d'acquisition des analyseurs. C'est le cas avec les instruments de type Q-TOF et LTQ-Orbitrap pour des résolutions en masse allant de 7500 à 30000 (exprimées sous la forme du rapport $m/\Delta m$ pour un ion de rapport m/z 400, Δm étant la largeur du pic à mi-hauteur). Par contre, le couplage de systèmes UHPLC avec des appareils de type FTICR peut-être problématique lorsque les spectres sont enregistrés à très haute résolution (>50000) en raison d'un temps d'acquisition des spectres de l'ordre de plusieurs secondes (Han et al; 2009, Metz et al; 2007).

En outre, différents types de mécanismes de rétention chromatographiques sont disponibles. La chromatographie en phase normale (NP, de l'anglais «*normal phase*») est bien adaptée à la

rétenion des composés polaires. Elle implique des phases stationnaires composées de silice ou d'alumine non modifiées et des phases mobiles non aqueuses qui sont incompatibles avec la détection en MS, limitant de ce fait son utilisation en analyse métabolomique par LC-MS. En ce qui concerne la chromatographie à polarité de phase inversée (RP de l'anglais «reverse phase»), la silice est fonctionnalisée, rendant la phase stationnaire hydrophobe et compatible avec des solvants aqueux et organiques. Les méthodes RP sont adaptées à la détection en API-MS. Elles sont hautement reproductibles et leurs mécanismes de séparation sont bien maîtrisés ((Kirkland, 2004, Stella et al; 2007, Vervoort et al; 2002). Toutefois, afin d'obtenir une bonne séparation en chromatographie de type RP, il est nécessaire que les analytes présentent des propriétés hydrophobes de sorte qu'ils puissent interagir avec la phase stationnaire. La chromatographie de type RP est donc adaptée pour les métabolites de polarité faible à intermédiaire.

Par contre, les métabolites très polaires sont mal retenus sur ces colonnes (élus dans le volume mort) et sont donc particulièrement sujets à l'effet matrice, ce qui limite leur sensibilité de détection. Des mécanismes de rétention alternatifs tels que la chromatographie de paires d'ions, la chromatographie en phase normale aqueuse (ANPC, de l'anglais «aqueous normal phase chromatography») ou la chromatographie d'interaction hydrophile (HILIC) ont été développés pour répondre à cette question clé. En ce qui concerne la chromatographie par paire d'ions, un contre-ion donnant des produits de décomposition volatils est additionné à la phase mobile pour former une paire d'ions avec les analytes. Les paires d'ions résultantes présentent une charge nulle, elles sont donc hydrophobes et retenues sur des colonnes RP. L'utilisation de tensioactifs anioniques et cationiques, et également d'agents d'appariement d'ions pour la détection et la quantification de métabolites basiques et acides est largement documentée (Dai et al; 2005, Dai & Carr, 2005, Dai & Carr, 2009, Shibue, Mant, & Hodges, 2005a, Shibue, Mant, & Hodges, 2005b, Shibue, Mant, & Hodges, 2005c). Cependant, les deux principaux inconvénients de cette approche sont la faible volatilité de la paire d'ions et/ou la trop forte affinité entre le réactif d'appariement d'ions et la substance à analyser qui est susceptible d'empêcher la dissociation de la paire d'ions dans la source électrospray induisant un phénomène de suppression d'ionisation (Apffel et al; 1995, Gustavsson et al; 2001, Lu, Bennett, & Rabinowitz, 2008, Sysi-Aho et al; 2007).

Une autre alternative est l'utilisation de phases mixtes présentant des sélectivités complémentaires tels que des supports chromatographiques présentant un greffage covalent avec des groupements cyano, diol, amino ou autres. Par exemple, les phases pentafluorophenylpropyl génèrent des interactions pi-pi et interactions stériques dues au noyau phényl, alors que les groupements fluorés assurent une interaction polaire. Ces phases sont bien adaptées pour la rétention de composés aromatiques azotés (Kinsley Narainsamy et al; 2011) et leur complémentarité

vis-à-vis des colonnes RP classiques a été démontrée pour les analytes basiques qui présentent une meilleure rétention sur ce type de phases stationnaires (Euerby & Petersson, 2003).

Récemment, une nouvelle technologie prometteuse dénommée ANPC a vu le jour (Pesek et al; 2008, Pesek, Matyska, & Fischer, 2011). Alors que la plupart des matériaux de silice utilisés pour la chromatographie ont une surface composée principalement de silanols (Si-OH-), les supports ANPC ont pour groupes terminaux des hydrures (Si-H-). Les phases mobiles utilisées sont des mélanges de solvants organiques (tels que le méthanol ou l'acétonitrile) contenant une petite quantité d'eau et qui sont compatibles avec la détection en ESI-MS. Ainsi, les solutés polaires (tels que les acides aliphatiques et les amines) sont efficacement retenus sur ces colonnes, avec des temps de rétention diminuant à mesure que la proportion d'eau dans la phase mobile augmente. Pesek et coll. ont démontré que le mode ANPC couplé à une détection MS assurait la rétention mais aussi la séparation des composés hydrophiles: sucres, acides aminés et acides carboxyliques dans des échantillons d'urine synthétique (un mélange d'étalons internes inspiré des métabolites urinaires endogènes) et d'urines humaines (Pesek et al; 2008, Pesek, Matyska, & Fischer, 2011).

La chromatographie liquide d'interaction hydrophile (HILIC) est devenue une approche alternative de plus en plus utilisée pour séparer efficacement les analytes polaires en utilisant des solvants d'élution compatibles avec la spectrométrie de masse. En fait, la rétention de molécules se produit grâce à une couche mince d'eau semi immobilisée sur la phase stationnaire de sorte que des liaisons hydrogène et les interactions dipôle-dipôle sont possibles et assurent une meilleure désolvatation des analytes. Depuis le début des années 2000, différentes applications ont été publiées en utilisant ce mode particulier de rétention. Citons l'analyse d'oligosaccharides, d'acides aminés et nucléotides depuis des extraits de feuilles de plante (Tolstikov & Fiehn, 2002), l'acquisition d'empreintes métaboliques d'urines de rats (Ildborg et al; 2005), la quantification de métabolites impliqués dans la voie de biosynthèse du glutathion dans la levure (Lafaye et al; 2005a), et la détection et la quantification de métabolites polaires à partir d'extraits cellulaires de E. coli (Bajad et al; 2006). De nouveaux supports chromatographiques HILIC tels que HILIC-amide ou les phases zwitterioniques ZIC-p-HILIC (Chirita et al; 2011) ont été développés pour améliorer la rétention des métabolites polaires présentant une large gamme de structures chimiques. En utilisant la phase stationnaire HILIC-amide (qui améliore la stabilité des liaisons hydrogène en considérant les atomes O et N de la fonction amide), plus de 90 métabolites polaires comprenant des acides aminés, des acides organiques, des sucres et des polyols ont été quantifiés dans des extraits de raisin (Gika et al; 2012c).

Les méthodes ANP et HILIC ont récemment été évaluées et comparées à la chromatographie RP par Zhang et coll dans des mélanges de composés de référence ainsi que des échantillons d'urines humaines (Zhang et al; 2012). Au vu de la forme des pics chromatographiques, des classes de

composés détectés ainsi que la résolution d'espèces isomères, la colonne ZICpHILIC utilisant des solvants d'élution basiques (pH>10) assurait la meilleure couverture métabolique.

Même si la LC reste associée à des valeurs de temps de rétention, contrairement à la GC (indice de rétention), de plus en plus de travaux s'intéressent à la prédiction de temps de rétention de composés sur la base de leurs propriétés physico-chimiques notamment pour des colonnes chromatographiques hors RP opérant en mode gradient (Boswell et al; 2011, Creek et al; 2011, Gika et al; 2012b, Gika et al; 2012a).

Cette polyvalence des sources d'ionisation, analyseurs et de supports chromatographiques a fait de la LC-MS une méthode de choix pour le profilage métabolique.

3.2. Identification des métabolites par spectrométrie de masse

Cette étape est une des plus délicates à mettre en œuvre. Elle nécessite le plus souvent l'utilisation conjointe de plusieurs techniques analytiques telles que les spectroscopies NMR, IR, UV et la MS. Cependant, compte tenu de la difficulté de purifier les métabolites d'intérêt, de leur présence fréquente à l'état de trace, et de l'impossibilité de certaines molécules à être détectées par certaines méthodes, une approche multidimensionnelle n'est pas toujours réalisable, et l'identification ne se réalise le plus souvent qu'avec une seule méthode d'analyse qui est celle utilisée pour l'acquisition de l'empreinte.

3.2.1. De l'annotation à l'identification en spectrométrie de masse

L'identification des variables d'intérêt débute par une étape d'annotation. En approche LC/API-MS, cette étape repose sur l'homologie de la masse précise détectée d'une variable à celle calculée d'un métabolite décrit dans des banques de données publiques chimiques, métabolomiques ou biochimiques. Les bases de données disponibles seront présentées en PARTIE A.CHAPITRE II.4.2.3 (voir Table 3). L'annotation peut aussi se faire à l'aide d'une base de données spectrales. Ces dernières sont le plus souvent des banques de données internes, développées dans les laboratoires qui répertorient les spectres de masse, et/ou MS/MS et/ou MSⁿ et/ou temps de rétention de métabolites d'intérêt. Une fois annotés, les variables doivent être identifiées à l'aide d'expériences de MS/MS. Les spectres obtenus peuvent être en accord ou au contraire permettent de réfuter certaines structures. Enfin, l'ultime étape de l'identification reste l'analyse comparative avec un composé de référence si disponible.

Cette étape critique d'identification a fait l'objet d'un travail de concertation en vue d'harmoniser les démarches et de pouvoir échanger les résultats à travers leur publication. A cet effet, le groupe de travail sur la standardisation des analyses chimiques (CAWG) de la société de métabolomique a défini les niveaux suivants d'identification des métabolites (Sumner et al; 2007) :

1. «Composés identifiés»: correspondent au niveau d'exigence le plus élevé. Celui-ci requiert un minimum de 2 critères d'identification orthogonaux : temps de rétention et spectre de masse, ou temps de rétention et spectre ^1H NMR, ou masse précise et spectre MS/MS, ou masse précise et amas isotopique correspondant, ou ^1H - et/ou ^{13}C -NMR avec le spectre NMR 2D. Ces informations sont obtenues à partir du produit chimique de référence analysé dans les mêmes conditions expérimentales. (Figure 9)

2. «Composés putativement annotés»: ce niveau d'identification concerne les composés qui ne sont pas commercialement disponibles ou difficilement synthétisables. L'identification du métabolite repose sur la similitude spectrale et/ou physico-chimique du composé avec les informations des banques de données spectrales.

3. «Composés putativement caractérisés» : correspondent aux composés pour lesquels il n'existe pas de proposition de structure dans les bases de données chimiques et spectrales mais qui présentent des similitudes spectrale et/ou physico-chimique avec des composés connus d'une classe chimique donnée.

4. « Composés inconnus» pour lesquels, aucune hypothèse structurale n'est émise.

Afin de respecter cette hiérarchisation dans la caractérisation des métabolites, l'identification des variables d'intérêt se fait dans un premier temps par interrogation de banques de données biochimiques telles KEGG (Kanehisa & Goto, 2000) et HMDB (Wishart et al; 2007). Il s'agit de comparer une masse précise expérimentale à une masse précise théorique répertoriée dans des bases de données. S'en suit la confirmation de l'hypothèse structurale alors émise par la concordance du temps de rétention expérimental et/ou de spectres MS/MS avec les propriétés physico-chimiques du composé proposé par les bases de données. Mais, encore une fois, seule la comparaison des propriétés physico-chimiques du composé à identifier avec celles d'un composé de référence permettent une identification formelle.

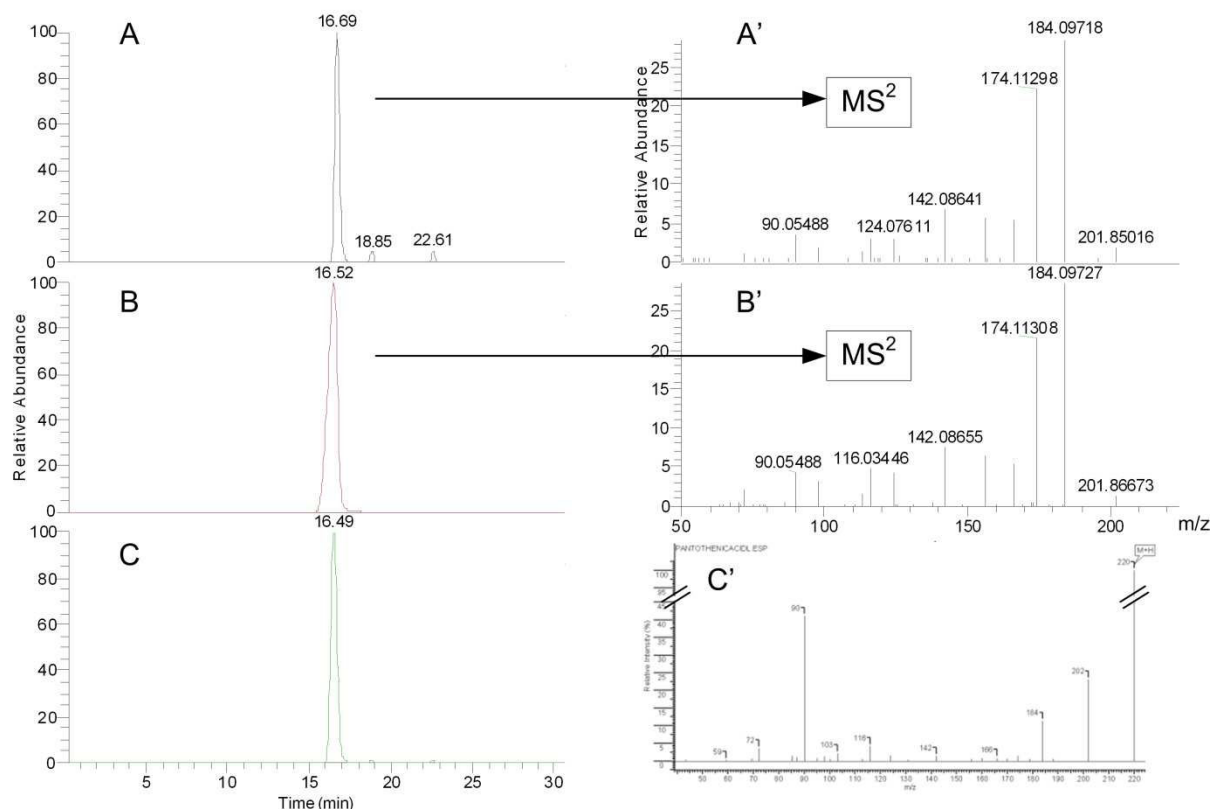


Figure 9. Exemple de métabolite formellement identifié selon les recommandations du CAWG

Identification de l'acide pantoïque dans les urines de rat. **(A)**: Chromatogramme du courant ionique de l'ion m/z 220.1180 (± 2 ppm) à identifier depuis l'empreinte d'urine de rat obtenue en LC/ESI(-)LTQ-Orbitrap, annoté comme étant «acide pantoïque». **(A')**: Spectre CID de l'ion d'intérêt $m/z=220.1180$. **(B)**: Chromatogramme du courant ionique de l'ion $m/z=220.1179$ (± 2 ppm) représentant la masse théorique de l'acide pantoïque protoné. **(B')**: Spectre de fragmentation par CID du composé de référence «acide pantoïque». **(C)**: Chromatogramme du courant ionique de l'ion $m/z=220.1180$ obtenu sur des urines humaines surchargées du composé de référence «acide pantoïque». **(C')**: Spectre CID de l'acide pantoïque selon HMDB, acquis sur un spectromètre de masse à analyseur triple quadripolaire (adapté à partir de (Werner et al; 2008b)).

3.2.1. Logiciels d'annotation et de regroupement des signaux en ESI/MS

L'ionisation ESI produit souvent plus d'un ion pour un métabolite donné, ces ions peuvent correspondre à des fragments produits en source, des adduits et leurs isotopes, ce qui entraîne une surestimation du nombre de métabolites (Figure 10). Leur présence peut cependant être utile dans la mesure où elle peut confirmer la pertinence des compositions élémentaires car elle permet plusieurs mesures de masses précises liées à la même molécule (Werner et al; 2008b, Werner et al; 2008a). Toutefois, si elle n'est pas prise en compte, cette redondance du signal complique le traitement des données, en effet un ion fragment peut être par exemple considéré à tort comme un ion mono-isotopique.

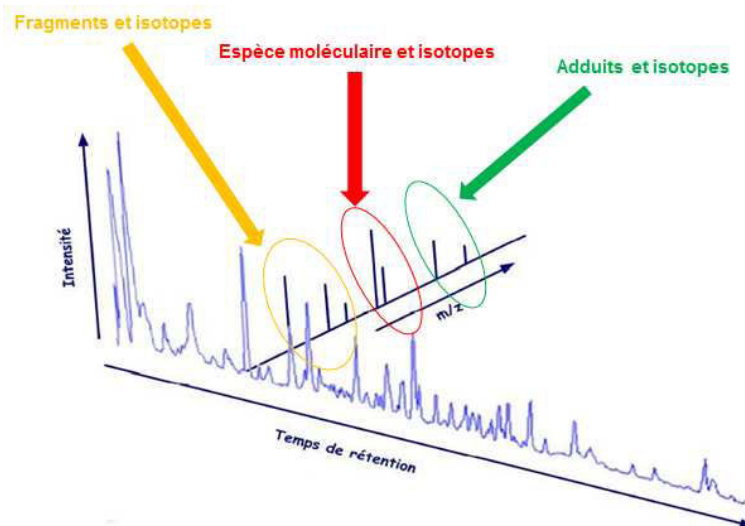


Figure 10. Redondance du signal en ESI/MS (Boudah, Roux, & Junot, 2012)

Tous ces ions générés pendant le processus de désorption-ionisation d'une molécule sont coélus avec l'espèce moléculaire ionisée et leurs intensités sont supposées corrélées avec celle de ce dernier (Werner et al; 2008a). Brown et coll ont de ce fait décrit les corrélations positives observées entre des ions coélus. Les différences de masse les plus fréquemment observées étaient celles entre les isotopomères $^{12}\text{C}/^{13}\text{C}$, $^{35}\text{Cl}/^{37}\text{Cl}$ et $^{32}\text{S}/^{34}\text{S}$. La détection d'ions adduits avec du sodium ou du formate d'ammonium ainsi que des pertes d'eau, d'ammoniac, d'acide formique, de monoxyde et de dioxyde de carbone ont été observées. De plus, la différence de masse la plus fréquemment rencontrée était de 0.3 Da. Celle-ci est attribuée aux pics d'épaulement correspondant aux pics artéfactuels générés par le calcul de la transformée de Fourier (Brown et al; 2009). Ceci met en exergue la complexité des matrices de données LC-API/MS et met l'accent à nouveau sur la nécessité de développer une base de données spectrales.

Certains outils ont de ce fait été développés pour structurer l'ensemble des variables issues de la même molécule. L'annotation différentielle peut être réalisée en faisant correspondre les différences de masse au sein d'un spectre de masse à des différences spécifiques couvrant isotopes connus, adduits ou pertes neutres. Le principe de ces méthodes est basé sur le fait que des signaux différents provenant d'un seul métabolite présenteraient un pic chromatographique de même forme avec les mêmes variations d'intensité d'une condition à une autre. Ainsi, la corrélation doit être calculée entre les paires d'ion soit à travers (i) plusieurs spectres dans un échantillon (corrélation spectrale) (Kuhl et al; 2012, Zeng et al; 2014) comme ceci est fait dans le programme CAMERA (Kuhl et al; 2012) ou sur (ii) l'ensemble des échantillons d'une analyse comme programmé dans AStream (Alonso et al; 2011) ou d'autres études en métabolomiques [35-38].

Différents auteurs ont développé des démarches de travail traitant de la corrélation d'intensité entre échantillons et/ou temps de rétention et/ou par différences m/z entre paires

appariées. Chen et coll. (Chen et al; 2008) ont calculé manuellement la corrélation entre courant ionique du pic d'intérêt et celui des pics co-élus. Werner et coll. ont quant à eux appliqué les matrices d'auto-corrélation sur l'ensemble des données LC-MS (Werner et al; 2008a). Ipsen et coll. ont utilisé le test de χ^2 pour mettre en évidence des paires d'ions parent-produit dans des données LC-TOF/MS (Ipsen et al; 2010).

A l'heure actuelle, certains logiciels statistiques sont disponibles pour aborder la question de la redondance d'information. IntelliXtract (ACD/IntelliXtract, Advanced Chemistry Development, Inc. www.acdlabs.com) développé par ACD/ Labs assigne les ions monoisotopiques de l'affectation $[M+H]^+$ ou $[M-H]^-$ en tenant compte des rapports $^{12}C/^{13}C$ attendus mais aussi les adduits, les multimères, les pertes neutres et les ions fragments (Ito et al; 2011).

Kuhl et coll ont développé une suite d'algorithmes nommé CAMERA, pour l'annotation de profils métaboliques (Kuhl et al; 2012). Ce programme est disponible chez Bioconductor et est implémenté en langage R, il est utilisé pour le post-traitement de listes provenant d'XCMS. En premier lieu, CAMERA regroupe les pics dans le domaine de temps de rétention dont la largeur est définie selon la méthode LC utilisée, il calcule alors la corrélation entre courants ioniques correspondants. Si la corrélation est supérieure à un seuil, les ions seront conservés dans un même groupe, autrement il y a création d'un nouveau groupe d'élution. Après cela, les différences de masses entre les ions d'un même groupe sont calculées et assignées aux isotopes naturels ou ions adduits connus. AStream est une autre suite logicielle écrite en langage R qui permet d'annoter les données LC-MS de métabolomique et pour laquelle la réduction de données repose sur la corrélation d'intensité inter-échantillons (Alonso et al; 2011).

Ces outils sont de plus en plus utilisés dans l'annotation des métabolomes et l'identification des métabolites dans des extraits biologiques. Récemment, notre équipe a proposé une démarche de traitement de données LC-MS utilisant le programme CAMERA, les matrices d'auto-corrélation et l'annotation des ions redondants avec une base de données interne ESI/MS pour l'annotation du métabolome urinaire humain (Roux et al; 2012). Nous avons identifié 384 métabolites qui représentent 659 et 825 ions annotés en mode d'ionisation positive et négative, respectivement. A l'aide de ces outils logiciels, il a été possible d'exclure l'information redondante générée par la détection ESI/MS afin de ne se concentrer que sur les ions d'intérêt (protonées ou déprotonées).

3.2.2. L'apport des bases de données spectrales

L'identification des métabolites est une étape limitante au développement des analyses métabolomiques en LC-API/MS (Werner et al; 2008b, Wishart, 2011). Lors d'une approche LC-MS, l'identification formelle d'un métabolite repose idéalement sur des homologues de temps de rétention chromatographiques, de masses précises et de spectres de MS/MS ou MS^n entre le

métabolite à identifier et un composé de référence d'une bibliothèque spectrale. Sinon, la caractérisation du métabolite n'est que partiellement atteinte et l'identification qui en découlera sera putative. Malheureusement, il arrive très souvent que le signal à caractériser ne soit pas décrit dans les banques de données biochimiques. Dans ces cas, l'interrogation de banques chimiques pourra aider à générer des hypothèses structurales dont la pertinence sera évaluée par l'interprétation des spectres MS/MS, et validée après la synthèse organique du composé (Werner et al; 2008b).

Cette démarche sous-tend la construction d'une base de données spectrales contenant à la fois les spectres de masse (MS et MS/MS) et temps de rétention dans le système chromatographique utilisé. Ces dernières ont été initialement construites en NMR et en GC-MS, qui sont des méthodes analytiques reproductibles fournissant des données fiables qui peuvent être partagées par différents laboratoires. Certaines de ces bases de données sont commerciales [Bruker (500 composés, 13 500 spectres NMR), Chemomx (315 composés, des spectres NMR), et la base de données du NIST (GC-MS), par exemple], alors que d'autres sont gratuites, telles que celle de l'institut Max Planck de Golm (Kopka et al; 2005), ou celle du projet HMDB (Wishart et al; 2007). La situation est différente avec la spectrométrie de masse à pression atmosphérique (API) pour laquelle les fragmentations sont peu reproductibles et par conséquent délicates à comparer (Oberacher et al; 2009a, Oberacher et al; 2009b, Palit & Mallard, 2009). Par ailleurs, contrairement à la GC-MS, les spectres de masse en API fournissent peu d'information structurale, ce qui nécessite l'utilisation complémentaire d'expériences de fragmentation en MS/MS. C'est la raison pour laquelle les banques spectrales d'API/MS, qui sont utiles pour gérer la redondance d'information et annoter des jeux de données (Figure 11) ne suffisent souvent pas à identifier des métabolites. Il faut alors recourir à des bases spectrales d'API-MS/MS (Courant et al; 2012, Roux et al; 2011, Roux et al; 2012).

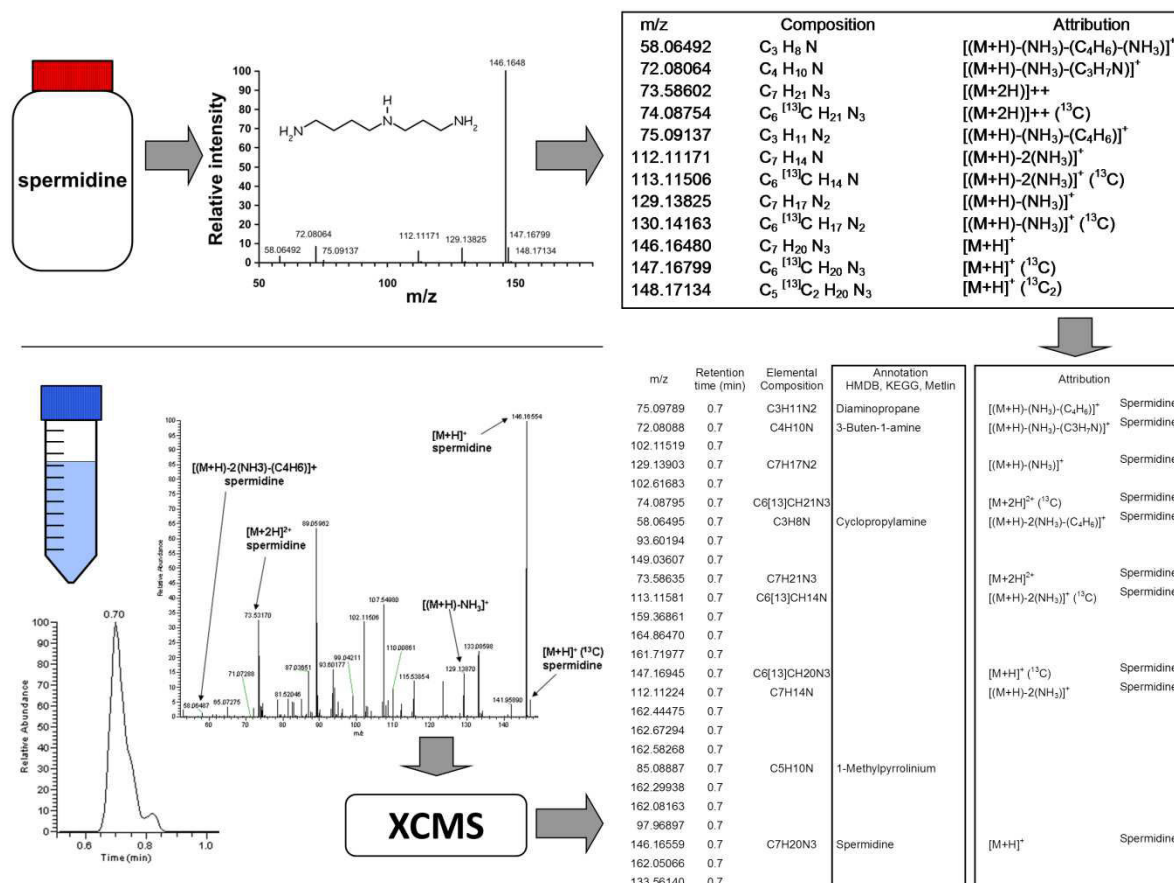


Figure 11. Constitution d'une banque de données spectrales

En haut : Une solution standard de spermidine est analysée par API/MS. Le spectre de masse est exporté sous forme d'une liste de signaux (avec la composition élémentaire et l'attribution). Cette liste, après interprétation, est importée dans la base de données spectrales du laboratoire.

En bas : Les extraits cellulaires sont analysés par UPLC-MS et les ions sont extraits à l'aide de logiciels de détection automatique des signaux. La liste des variables (temps de rétention, m/z) est d'abord annotée par comparaison des masses précises expérimentales aux masses théoriques répertoriées dans des bases de métabolites (KEGG (Kanehisa & Goto, 2000), HMDB (Wishart et al; 2007) et Metlin (Smith et al; 2005)). Les annotations indiquent la présence présumée de la spermidine, du diaminopropane, de la 3-butène-1-amine, du 1-méthylpyrrolinium et de la cyclopropylamine. L'annotation par la base de données spectrales du laboratoire a confirmé la présence de spermidine, mais a montré que 3 des autres annotations sont erronées. Ces ions mal annotés correspondent en fait à des ions fragments de spermidine générés dans la source electrospray pendant le processus de désolvatation. Cet exemple met en évidence la complexité des jeux de données générés par API-MS (adapté de (Roux et al; 2011))

Malgré leur faible reproductibilité inter-instrumentale, des bases de données de spectres MS/MS telles que celles de HMDB (Wishart et al; 2007), de Metlin (Smith et al; 2005), de metabolome.jp (Horai et al; 2007), ou de LIPID MAPS (Fahy et al; 2007b) commencent à se développer. Certaines d'entre elles présentent des spectres de collision en mode MS/MS avec différents niveaux d'énergie d'activation, ou des spectres obtenus avec différents types d'instruments afin d'être le plus universel possible.

Cependant, l'utilisation de ces spectres à des fins d'identification ou de comparaison peut conduire à des résultats erronés (Werner et al; 2008b). Ainsi, il y a un besoin urgent de normaliser les conditions expérimentales d'acquisition des spectres afin de promouvoir l'échange et le partage inter-laboratoire de composés inconnus ou tout de moins non disponibles commercialement. Une telle démarche a été initiée en 2010 par le Réseau Français de Métabolomique et de Fluxomique à travers une initiative regroupant 9 laboratoires, différents types de spectromètres de masse (pièges ioniques, Q-TOF et Orbitrap) et pilotée par Jean-Claude Tabet (Université Pierre et Marie Curie) et Christophe Junot (CEA). L'objectif de cette initiative, qui continue d'être développée au sein de l'infrastructure nationale de métabolomique (MetaboHUB, financée depuis 2013 par les investissements d'avenir), est de démontrer que le contrôle de la distribution d'énergie interne des ions permet de construire des bases de données de spectres CID partageables au sein de différents laboratoires (Ichou F et al; 2014, Ichou et al; 2013). A l'heure actuelle, à défaut de pouvoir constituer une base universelle de données spectrales, chaque utilisateur est contraint de constituer sa propre bibliothèque de spectres avec ses règles et modes opératoires.

3.3. Quantification des métabolites par spectrométrie de masse

Deux types de stratégies peuvent être mises en œuvre dans le cadre d'une quantification par spectrométrie de masse : (i) la quantification relative qui met en avant des différences de concentrations entre différents groupes d'échantillons (ii) la quantification absolue, qui permet la mesure de la concentration réelle des métabolites dans les échantillons. La première approche est moins contraignante à mettre en place. Elle permet entre autre de corriger des biais expérimentaux inhérents aux sources à pression atmosphérique, tandis que les approches de quantification absolue sont plus difficiles à développer mais permettent de donner une concentration réelle du métabolite dans l'échantillon ce qui facilite grandement le partage de résultats et le transfert de résultats aux cliniciens et biologistes.

3.3.1. Quantification relative

L'ESI-MS est sujette à de nombreuses dérives analytiques, notamment des effets de suppression ionique due à l'encrassement de la source qui modifie le rendement d'ionisation au sein d'une même analyse ou en d'expériences différentes. Il faut donc maîtriser et corriger ces fluctuations de rendement d'ionisation afin qu'elles n'entravent pas l'interprétation biologique des données. La quantification relative des métabolites peut apporter une solution à ce problème. Les fluctuations des signaux observés en API/MS peuvent par exemple être évaluées par des étalons internes ajoutés dans tous les échantillons. Ceux-ci doivent présenter une similitude structurale avec les composés d'intérêt (idéalement, métabolites marqués) (Guo, Ji, & Li, 2007, Guo & Li, 2009, Guo &

Li, 2010, Shortreed et al; 2006). Ces étalons internes peuvent être obtenus par synthèse chimique des métabolites marqués aux isotopes stables, mais cette procédure est coûteuse et longue. Ils peuvent également être obtenus par marquage métabolique uniforme d'organismes (Giavalisco et al; 2009, Kiefer, Portais, & Vorholt, 2008, Lafaye et al; 2005b, Mashego et al; 2004). Cette dernière approche est cependant limitée aux études de plantes ou des microorganismes, bien que le marquage métabolique de rongeurs ait déjà été décrit en protéomique (Wu et al; 2004) et pourrait être transposable à la métabolomique. Enfin, des molécules marquées aux isotopes stables («tags») peuvent être ajoutés aux métabolites à détecter grâce à des réactions chimiques spécifiques de certaines fonctions chimiques. Ce type d'approche a été décrit pour les métabolites contenant des fonctions acide carboxyliques et amines primaires (Guo, Ji, & Li, 2007, Guo & Li, 2009, Guo & Li, 2010). Ces deux dernières stratégies peuvent également être envisagées au cours d'une quantification absolue et posent le problème de la représentativité et l'exhaustivité du mélange choisi ainsi que son coût.

Avoir recours à des échantillons biologiques de contrôle qualité (QC) permet d'assurer la normalisation des données de métabolomique notamment la fusion d'analyse provenant de séries différentes. Ces QC sont un mélange des échantillons d'une expérience et sont donc représentatifs des échantillons à analyser. Ils sont injectés de façon répétée au cours de la séquence d'acquisition, ce qui permet de modéliser le biais analytique et de corriger le signal de chaque métabolite par régression et offre la possibilité d'intégrer des données provenant de plusieurs lots (Dunn et al; 2011).

L'injection d'échantillons de QC analysés à intervalles réguliers tout au long du déroulement de l'expérience est une solution intéressante mais qui reste néanmoins limitée aux seuls métabolites détectés dans ce type d'échantillon. De plus, cette approche impose également d'utiliser un aliquot d'un seul et même échantillon de QC, ce qui pose le problème de la disponibilité mais aussi de la stabilité à long terme des métabolites dans ces extraits biologiques.

Une alternative serait de disposer de matériel biologique de référence pour chaque matrice biologique. Dans ce cadre, le NIST (*National Institute of Standard and Technology* http://www.nist.gov/srm/program_info.cfm) en collaboration avec le NIH (*National Institute of Health*) ont développé un plasma humain de référence (SRM 1950) à partir d'un mélange de plasma provenant d'une cohorte d'hommes et de femmes âgés de 40 à 50 ans (Koal & Deigner, 2010, Simon-Manso et al; 2013). Le SRM 1950 (<http://srm1950.nist.gov>) est commercialement disponible, et la concentration de plus de 50 métabolites (vitamines, acides gras, par exemple) a été déterminée et l'identification de plus de 300 métabolites à l'aide d'une approche multiplateforme (GC-MS, LC-MS et NMR) et 250 lipides ont récemment été menées (Kind et al; 2013, Simon-Manso et al; 2013). Les SRM répondent au besoin de disponibilité, mais reste la question de leur stabilité à long terme qui peut

être source de biais analytiques. Développer des approches de quantification absolue à large échelle pourrait constituer une réponse à ce problème.

3.3.2. Quantification absolue

La quantification absolue a pour but de mesurer la concentration réelle (exprimée en molarité) de métabolites dans des échantillons. Les résultats rendus sont souvent des concentrations massiques (ng/mL) ou molaires (mmol/mL) dans le cadre d'analyses de biofluides ou encore rapportées à la quantité de tissu étudiée ($\mu\text{g}/\text{mg}$ de protéines). En plus de fournir des informations d'intérêt biologique, elle permet de standardiser efficacement les données de métabolomique à haut débit en vue de leur partage et de leur exploitation par les biologistes. Différentes concentrations de calibrants (molécules à quantifier) et des concentrations constantes d'étalons internes sont ajoutés dans le milieu biologique à analyser afin de construire une gamme d'étalonnage. Celle-ci exprime la relation entre le rapport des intensités des signaux des calibrants et des étalons interne en fonction des concentrations de calibrants (Figure 12).

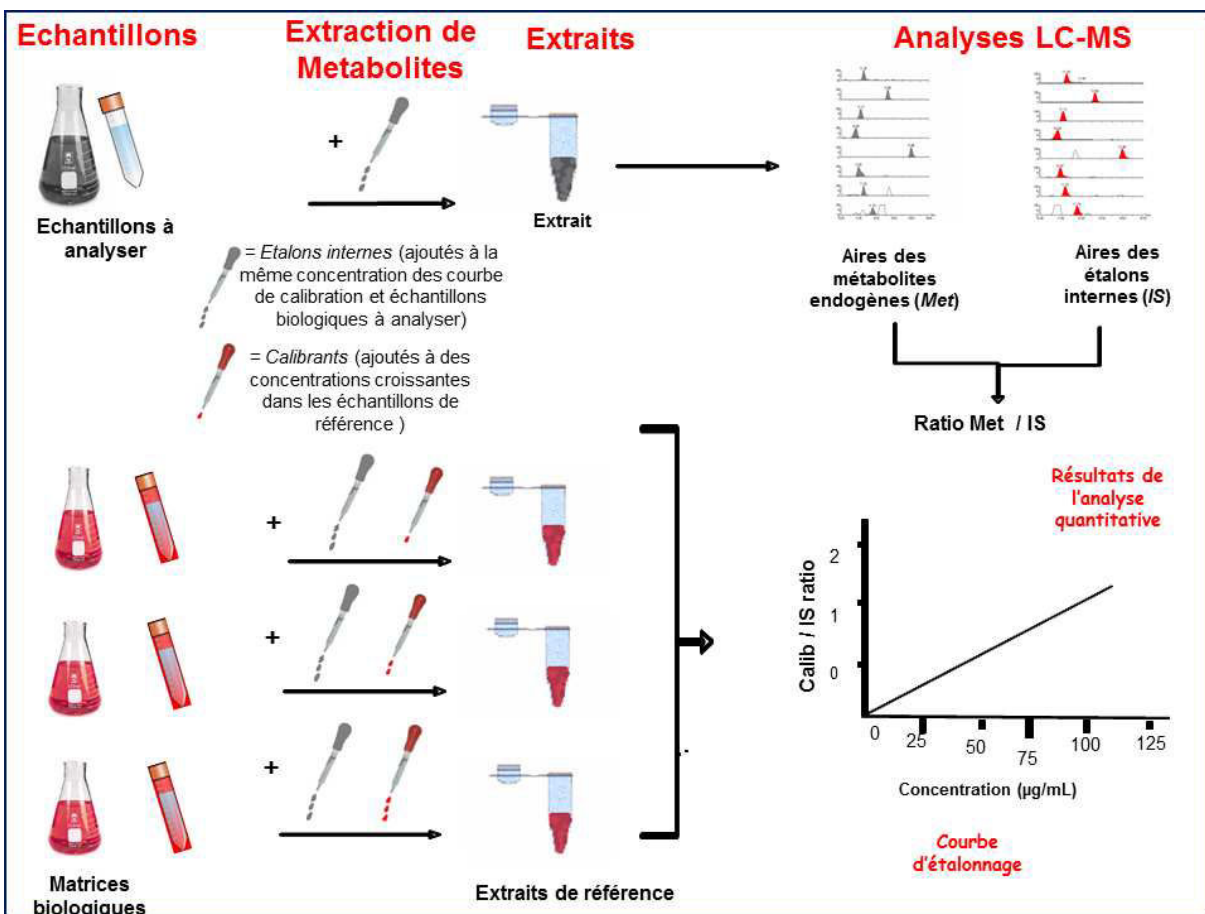


Figure 12. Démarche de quantification absolue de métabolites à l'aide de calibrants et étalons internes

Les calibrants sont ajoutés à la matrice biologique à une concentration constante alors que les étalons internes sont ajoutés à des concentrations croissantes afin de construire une courbe d'étalonnage sur la base des rapports d'intensités Calibrant/Etalon interne. Les aires des molécules endogènes à doser sont exprimées en rapport Calibrant/Molécule endogène et le résultat est donc exprimé en concentration massique ou molaire.

Une telle approche est largement utilisée en recherche et développement pharmaceutique afin de mesurer les concentrations de xénobiotiques dans les matrices biologiques. Elle est encadrée par des recommandations émises par les autorités sanitaires américaines (FDA) (Shah et al; 2000) et des conférences internationales d'harmonisation (ICH-Topic Q2B). Cependant, la quantification absolue des métabolites endogènes reste difficile. En effet, le grand nombre de métabolites présentant une grande diversité chimique nécessite plusieurs mélanges d'étalons internes obtenus par synthèse chimique, ce qui rend la validation des méthodes complexe et fastidieuse. D'autre part, la présence à l'état endogène des molécules à quantifier nécessite d'utiliser des calibrants marqués aux isotopes stables ou à défaut de réaliser des ajouts dosés. Enfin, la détection des métabolites à quantifier peut-être gênée par l'existence de signaux provenant d'isotopes. Cette dernière limitation est particulièrement présente dans le domaine de la lipidomique, compte tenu de l'intensité des ions provenant des signatures isotopiques de ^{13}C . Une alternative serait l'utilisation d'extraits biologiques obtenus à partir d'organismes métaboliquement marqués (Bennett et al; 2008, Giavalisco et al; 2009, Godat et al; 2010, Kim et al; 2005, Lafaye et al; 2005b, Mashego et al; 2004) mais des variations entre les lots et des problèmes de stabilité peuvent survenir.

L'utilisation en routine des méthodes de quantification absolue reste actuellement limitée par le manque de disponibilité commerciale des métabolites marqués avec des isotopes stables. Une solution serait de collaborer avec des équipes de chimistes organiciens pour réaliser des synthèses chimiques. Notons que des kits pour analyses quantitatives absolues par MS sont maintenant commercialement disponibles. Citons par exemple le «HMT metabolomics solution package-kit» (Human Metabolome Technologies, Inc, Japon humanmetabolome.com/) qui est basé sur l'utilisation du spectromètre de masse commercialisé par la société Agilent, ou encore le kit AbsoluteIDQ™ de la société Biocrates life sciences AG (biocrates.com). Ce dernier permet de quantifier par analyses MS/MS en introduction directe des acylcarnitines, des acides aminés, des glycerophospholipides, des sphingolipides et des hexoses, ou encore en LC-MS/MS des amines biogènes.

Enfin, bien que les instruments en tandem de type TQ soient toujours considérés comme les spectromètres de référence pour les études de quantification (Bajad et al; 2006, Wei, Li, & Seymour, 2010), les instruments hybrides à haute résolution commencent à émerger dans ce domaine, en particulier dans les études protéomiques, comme décrit récemment avec des instruments type FTICR (Collier et al; 2008), Orbitrap (Savitski et al; 2011) et Q-TOF (Garcia-Villalba et al; 2010). En fait, les

dernières améliorations apportées aux technologies Q-Exactive et Q-TOF permettent des analyses multipléées en MS et MS/MS de type PRM « Parallel Reaction Monitoring » en acquisition dépendante DDA (Gallien et al; 2012) et SWATH™ (de l'anglais « *sequential window acquisition of all theoretical fragment-ion spectra* ») en acquisition indépendante DIA (Gillet et al; 2012). Au vu de leur large spectre d'application (la possibilité qu'ils offrent de réaliser la détection de métabolite, l'identification et la quantification dans le même temps), il est envisageable que ces approches soient utilisées en analyse métabolomique dans un futur proche (Zhu, Chen, & Subramanian, 2014).

4. Traitement des données en analyse métabolomique

Le vif intérêt porté récemment aux analyses métabolomiques non ciblées à haut débit a conduit à la continuelle amélioration des techniques d'analyse. Aujourd'hui, les méthodes disponibles sont en mesure d'analyser plus de métabolites simultanément en utilisant des méthodes de préparation d'échantillons adéquates. De nouveaux spectromètres de masse type HRMS assurent une meilleure exactitude et sensibilité de mesure, et des chromatographes et colonnes type UHPLC dont la capacité de pics est améliorée permettent une meilleure séparation des composés. Dès lors, il n'est plus possible de gérer l'énorme quantité de données générées manuellement et il existe un réel besoin d'utiliser des outils de bioinformatique afin d'assurer l'exploitation des données le plus complètement et fidèlement possible.

La bioinformatique est de ce fait devenue une partie intégrante des analyses métabolomiques. Elle vise à convertir les données analytiques en des connaissances biologiques et ce en permettant de gérer le stockage et l'analyse des données biologiques à travers le développement d'algorithmes, programmes informatiques, des bases de données et des outils statistiques. La bioinformatique joue un rôle majeur à 2 niveaux : (i) le traitement du signal brut afin de générer des données numériques analytiquement pertinentes (traitement «de bas niveau») qui sera présenté ci-dessous, ainsi que (ii) l'organisation et visualisation de l'information contenue dans les jeux de données traitées pour leur exploitation biologique (traitement «de haut niveau») (Katajamaa & Oresic, 2007). Ceci inclura les analyses statistiques, les moyens de visualisation des données et les bases de données mises à la disposition de la métabolomique. Ces points seront détaillés en PARTIE A.CHAPITRE II.4.2. Pour finir, au vu de l'expansion des analyses métabolomiques, nous aborderons les questions relatives au partage (PARTIE A.CHAPITRE II.4.3) et la fusion des données métabolomiques (PARTIE A.CHAPITRE II.4.4).

4.1. Des données brutes aux données traitées

Le traitement des données brutes consiste en la détection automatique des informations spectrales et chromatographiques de l'ensemble des composés détectés dans l'intégralité des

échantillons analysés en LC-MS. Son objectif est de convertir les signaux bruts (représentation graphique) en un format numérique contenant uniquement les informations analytiques pertinentes. Les logiciels dédiés permettent la détection automatique des signaux analytiquement pertinents de chaque échantillon, et leur réalignement (dans les domaines spectral et chromatographique, dans le cas des analyses LC-MS). L'ensemble des signaux détectés est ensuite restitué sous une forme matricielle compatible avec les logiciels d'analyse statistique tel que présenté en Figure 4: à chaque ligne, correspond une variable, c'est-à-dire un signal caractérisé par un temps de rétention et un rapport m/z ; et à chaque colonne, l'intensité ou l'aire des pics chromatographiques de chaque variable dans un échantillon donné.

Différents logiciels libres ou commerciaux assurant la détection automatique des signaux sont actuellement disponibles. Parmi ces logiciels, MZmine (Katajamaa, Miettinen, & Oresic, 2006) et XCMS (Smith et al; 2006) "acronyme de diverses formes (X) de la chromatographie couplée à la spectrométrie de masse» sont les plus largement utilisés (Table 1). Ces logiciels incluent différents algorithmes assurant les différentes étapes de détection automatique des signaux tel qu'illustré en Figure 13 pour le logiciel XCMS. Leurs différentes étapes de fonctionnement sont décrites en ANNEXE1.

Différents paramètres peuvent être réglés et sont définis par l'utilisateur. Certains sont constants car ils sont dépendants de l'instrument (tels la largeur à mi-hauteur moyenne de chromatogramme $fwhm$, la précision en mesure de masse $step$ et $mzwid$), tandis que d'autres sont moins génériques et doivent être optimisés par des tests (tel que le rapport signal sur bruit). Une fois tous ces paramètres (algorithmes, paramètres) optimisés dans une plate-forme, ils peuvent être automatisés pour les futures utilisations.

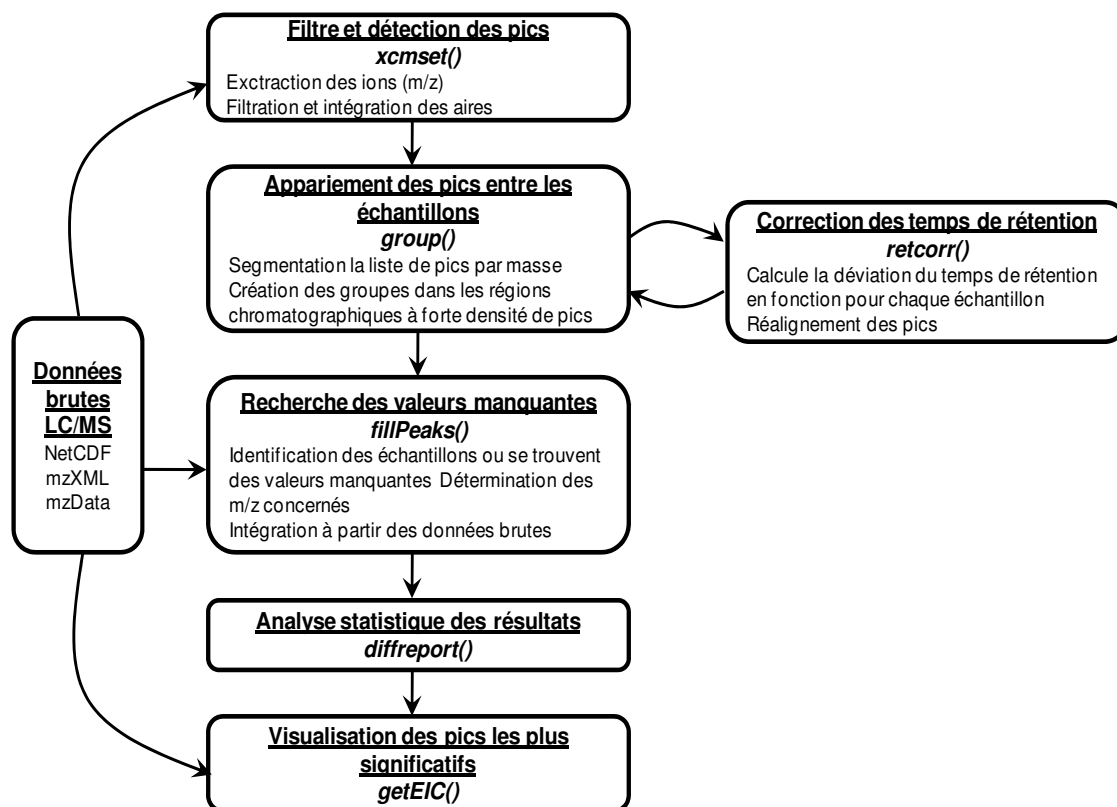


Figure 13. Stratégie de traitement des données mise en œuvre par XCMS

Une liste non-exhaustive des logiciels utilisés en analyses métabolomiques est présentée en Table 1. Notons que la plupart des logiciels commerciaux ne sont compatibles qu'avec les données acquises dans le format propriétaire du fournisseur du spectromètre de masse. Ils ne peuvent donc pas prendre en charge d'autres fichiers propriétaires ni même de nouveaux formats de données MS à venir. Pour exemple, MarkerLynx a été conçu pour fonctionner uniquement avec des données provenant des spectromètres de la société Waters (Guyancourt, France). A l'inverse, Refiner MS de la société Genedata AG (Bâle, Suisse) est compatible avec des données acquises sur des instruments provenant de plusieurs constructeurs.

De plus, certains logiciels libres tels que metAlign (Lommen, 2009) utilisent des codes propriétaires. Ceci entrave la vérification des étapes intermédiaires du traitement de données ainsi que l'amélioration de l'algorithme par d'autres équipes. Seuls les logiciels libres (et souvent à code source ouvert dits «*open source*») permettent la prise en charge de données provenant de différents types d'instruments sans restriction de marques (souvent après une conversion des fichiers) et la possibilité de faire évoluer l'algorithme au sein d'une communauté scientifique. Cette dernière option est d'un très grand intérêt étant donné qu'un logiciel générique de traitement de données LC-MS reste perfectible et qu'un ajustement selon les applications et les approches est souvent nécessaire.

LOGICIEL	ACCES	DONNEES BRUTES	TRAITEMENT DES DONNEES	NORMALISATION	ANALYSE STATISTIQUE	IDENTIFICATION
XCMS (Scripps Research Institute)	Package gratuit du logiciel gratuit R "Open-access"	Formats universels mzXML, NetCDF (conversion préalable des données)	Correction de la ligne de base et du bruit de fond, Détection des pics par déconvolution, Alignement des pics.	Non	Analyses univariées (T-test et ANOVA)	Connection à Metlin. Identification des adduits, fragments et isotopes (CAMERA R package).
Mzmine	Logiciel gratuit "Open-access" (Java)	Formats universels mzXML, NetCDF (conversion préalable des données) et Raw (format Thermo). plug-ins additionnels pour les autres formats constructeurs.	Correction de la ligne de base et du bruit de fond, Détection des pics par déconvolution, Alignement des pics.	Normalisation par l'utilisation de standards internes	Analyses multivariées (ACP, CDA, Sommon's plot)	Identification des adduits, fragments, isotopes et complexes. Connection aux principales bases de données en ligne (KEGG, PubChem, HMDB...) et recherche possible dans une base de données personnalisée.
MarkerLynx™XS (Waters)	Logiciel constructeur payant (Apex Track™ algorithme breveté)	Données acquises sur des Spectromètres de masse Waters uniquement.	Correction de la ligne de base et du bruit de fond, Détection des pics par déconvolution, Alignement des pics.	Normalisation par l'intensité totale des pics ou par l'utilisation de standards internes.	Analyses multivariées (ACP). Intégration directe avec Umetrics Simca-P (logiciel payant) pour les statistiques multivariées type PLS.	Connection aux principales bases de données ChemSpider et ChemFolder, et recherche possible dans une base de données personnalisée.
MetAlign (Plant Research International B.V)	Logiciel gratuit mais pas "Open-access"	Formats universels mzXML, NetCDF (conversion préalable des données)	Correction de la ligne de base et du bruit de fond, Détection des pics par déconvolution, Alignement des pics.	Normalisation par l'utilisation de standards internes	Analyses univariées (T-test)	Non
Refiner MS (Genedata)	Logiciel payant (module de Genedata Expressionist®)	Données acquises sur des Spectromètres de masse ABI, Agilent, Bruker, Ciphergen, Thermo et Waters. Supporte aussi les formats universels mzXML et NetCDF.	Correction de la ligne de base et du bruit de fond, Détection des pics par déconvolution, Alignement des pics.	Non	Module Genedata Analyst™ (T-test, ANOVA, PCA, PLS...)	Connection à SEQUEST, Mascot, Mass Frontier et HMDB.
MarkerView™ (AB SCIEX)	Logiciel constructeur payant	Données acquises sur des Spectromètres de masse AB SCIEX uniquement.	Détection des pics par déconvolution, Alignement des pics.	Normalisation en fonction de la taille de l'échantillon ou du volume d'injection.	Analyses univariées (T-test) et multivariées (ACP)	Non
MathDamp (Wolfram Research, Inc.)	Package gratuit de Mathematica (Logiciel propriétaire payant)	Formats universels mzXML, NetCDF (conversion préalable des données)	Détection des pics directement à partir des données brutes, Alignement des pics.	Normalisation par l'intensité totale des pics ou par l'utilisation de standards internes.	Analyses univariées (T-test, Z-score et ANOVA)	Non
Sieve® (ThermoFisher Scientific)	Logiciel constructeur payant	Données acquises sur des Spectromètres de masse Thermo-Fisher uniquement.	Détection des pics directement à partir des données brutes, Alignement des pics.	Non	Analyses univariées (T-Test et ANOVA)	Connection à ChemSpider et utilisation d'un outil de recherche SEQUEST.

Table 1. Principaux logiciels utilisés pour le traitement des données métabolomiques (Boudah, Roux, & Junot, 2012).

4.1.1. Avancées des logiciels

Depuis leur lancement, les logiciels ont constamment été sujets à des développements importants motivés par la nécessité de soutenir et suivre le progrès technologique des instruments d'analyse. Actuellement, les logiciels sont compatibles avec des données de MS précises et sensibles obtenues à partir d'analyseurs type à transformée de Fourier à haute résolution mais aussi des

données chromatographiques efficaces produites en séparation rapide, et ce en assurant la détection d'ions isobares, de composés faibles en abondances et de pics de coélution. Ceci s'est manifesté par le développement de nouvelles versions des logiciels XCMS et MZmine qui répondent mieux aux nouvelles attentes des équipes réalisant des analyses métabolomiques notamment sous UHPLC-HRMS.

Pour exemple, MetaXCMS s'appuie sur les algorithmes implémentés dans XCMS pour effectuer une méta-analyse métabolomique non ciblée englobant des séries d'analyses provenant d'une même plateforme analytique mais acquises à lors d'expériences différentes (Tautenhahn et al; 2011). Ce logiciel donne en plus de la matrice de sortie, un diagramme de Venn qui permet la visualisation directe du nombre de variables communes à certains/tous les groupes d'échantillons et celles spécifiques à un seul groupe. XCMS² peut être considéré comme une extension de XCMS car il offre la capacité supplémentaire de recherche automatique de spectres MS/MS dans la base de données METLIN (Benton et al; 2008). Récemment, l'outil convivial "XCMS en ligne", disponible sur le Web, a été mis en place pour assurer l'accessibilité de la métabolomique non ciblée aux scientifiques non bioinformaticiens (Tautenhahn et al; 2012). Une autre fonctionnalité a été introduite dans la dernière version de XCMS «X¹³CMS». Ce dernier logiciel identifie des isotopologues en considérant des expériences conduites parallèlement avec et sans marquage isotopique ouvrant des perspectives intéressante dans la découverte de flux métaboliques et nouvelles voies métaboliques (Huang et al; 2014).

Parallèlement, MZmine 2 a été remanié dans un cadre flexible assurant la modularité. Il traite maintenant de grands ensembles de données et permet d'intégrer facilement de nouveaux algorithmes qui pourraient effectuer des tâches supplémentaires que l'utilisateur pourra ajouter au besoin (Pluskal et al; 2010). MZmine qui était initialement dédié à la détection automatique et l'alignement des signaux comprend maintenant des modules qui gèrent plusieurs spectres du même échantillon (MS ou MSⁿ), permet une connectivité avec des bases de données assurant l'identification des métabolites et améliore la prise en charge des isotopes naturels.

4.1.2. Validation des logiciels d'extraction automatique des signaux

Contrairement à une intégration manuelle des pics d'intérêt dans l'ensemble des échantillons effectuée par les logiciels constructeurs des appareils utilisés, les logiciels de détection automatique des signaux permettent l'extraction sans a priori de l'intégralité –supposée– des pics contenus dans un jeu de données. Ceci demande une validation qui reste très peu abordée dans la littérature (Table 2) (Lange et al; 2008, Peters, van, & Janssen, 2009, Pluskal et al; 2010). Les raisons principales sont que la plupart des signaux présents dans les empreintes métaboliques restent non caractérisés et les résultats d'une procédure de détection automatique des métabolites peuvent être impactés à la fois

par le type d'instrument et le type de matrice biologique. En conséquence, les utilisateurs doivent évaluer le logiciel avec leurs propres critères afin de sélectionner celui ou celle la plus adapté à leurs instruments et à leur matrice biologique. Tautenhahn et coll. ont proposé une approche intéressante pour évaluer les logiciels d'extraction de signal. Elle est basée sur l'estimation de trois paramètres : le recouvrement, qui mesure la fraction de l'information pertinente extraite par l'algorithme, la précision, qui est le pourcentage des éléments pertinents par rapport aux faux positifs, et le temps d'analyse, qui est le temps nécessaire à l'algorithme pour traiter un jeu de données (Tautenhahn, Bottcher, & Neumann, 2008). En fait, de nombreux signaux artéfactuels sont présents dans les matrices de données. Ils ont été évalués à environ 400 pour 100 signaux pertinents (Tautenhahn, Bottcher, & Neumann, 2008). Une façon de résoudre ce problème consiste à effectuer des dilutions en série d'échantillons de contrôle qualité (QC) et à sélectionner les ions dont les niveaux de concentration sont corrélés avec le facteur de dilution (Croixmarie et al; 2009). Il est donc important de pouvoir distinguer les pics analytiquement pertinents des autres, et un des moyens envisageables est d'annoter un maximum d'ions dans ces échantillons. Ceci nous ramène à la nécessité de développer des bases spectrales ESI/MS d'annotation (Figure 11), aussi bien pour aider à l'interprétation biologique, que pour l'évaluation des performances des outils informatiques.

Data set	TracMass 2	msInspect ^a	MZmine ^a	OpenMS ^a	XAlign ^a	XCMS without RT correction ^a	XCMS with RT correction ^a
M1							
recall	0.88	0.27	0.89	0.87	0.88	0.98	0.94
precision	0.67	0.46	0.74	0.69	0.70	0.60	0.70
M2							
recall	0.93	0.23	0.98	0.93	0.93	0.97	0.98
precision	0.80	0.47	0.84	0.79	0.79	0.58	0.78

Table 2. validation des résultats d'alignement de différents logiciels de détection automatique des signaux (a : données de Lange et coll.(Lange et al; 2008)) (Tengstrand, Lindberg, & Aberg, 2014)

M1 et M2 représentent deux jeux de données différents acquis initialement par Lange et coll. (Lange et al; 2008) puis traités partiellement par Lange et coll. (Lange et al; 2008) et Tengstrand et coll. (Tengstrand, Lindberg, & Aberg, 2014).

4.2. Des données traitées aux données biologiques

Une analyse métabolomique aboutit à une analyse différentielle des empreintes, c'est-à-dire à une expression semi-quantitative des résultats. La difficulté est donc de gérer et de comparer ces empreintes qui contiennent plusieurs centaines à milliers de signaux. Pour ce faire des méthodes de traitement des données incluant des tests statistiques sont nécessaires pour faire ressortir des profils

métaboliques pertinents. Viennent alors des méthodes de visualisation des données pour assister à donner du sens biologique aux résultats obtenus.

4.2.1. Analyses statistiques en analyse métabolomique

En raison de la disparité en intensités mais aussi en variations inter-échantillons entre les variables étudiées (certains métabolites sont présents dans les milieux biologiques à des ordres de grandeur de concentrations allant jusqu'à 9 Log) et des fluctuations incontrôlées lors des analyses biologiques, les données exportées à partir des outils de détection automatique de pic ne peuvent être directement utilisées pour analyse statistique. Un procédé de prétraitement notamment précédant des analyses statistiques multivariées doit être appliqué en premier. Il convient de "centrer" les données (pour déplacer la distribution des valeurs autour de la valeur de 0 et non autour de la moyenne), d'effectuer une mise à l'échelle (appelée «scaling» en anglais) (van den Berg et al; 2006). Ceci se fait selon une mesure de distribution (par exemple en utilisant l'écart-type) ou une mesure de taille (comme l'utilisation de la moyenne ou du rang (Smilde et al; 2005) ou encore une «transformation» comme la transformation logarithmique. La procédure de normalisation la plus couramment utilisée en analyse métabolomique est le «centré réduit». Cette normalisation donne un poids égal aux signaux présentant des abondances très différentes. Cependant, cela peut conduire à une dilution de l'information biologique en surestimant les variables de faible intensité. Les méthodes qui reposent sur une transformation sont de plus en plus recommandées en analyse métabolomique, car elles ont un effet de «mise à l'échelle» tout en veillant à assurer l'homoscédasticité de distribution, afin de se conformer aux tests statistiques paramétriques.

4.2.1.1. Analyses univariées

Les analyses statistiques univariées sont des techniques classiquement utilisées en biologie. Elles permettent l'analyse d'un critère sans tenir compte des autres et ni des interactions existant entre eux. Le type de test utilisé dépend de la nature des variables à comparer : des calculs de corrélation sont réalisés pour les variables quantitatives alors que des analyses de variance sont utilisées pour les comparaisons entre variables quantitatives et variables qualitatives. De plus, la distribution de probabilité de la variable conditionne le choix du test. En effet, une distribution selon la loi normale autorise l'utilisation de tests paramétriques (test de Student ou corrélation de Spearman), alors que les variables dont la distribution au sein de la population étudiée ne suit pas une loi statistique sont étudiées à l'aide de tests non-paramétriques tels que le test de Mann-Whitney ou la corrélation de Pearson. Ces analyses univariées permettent de distinguer des signatures métaboliques.

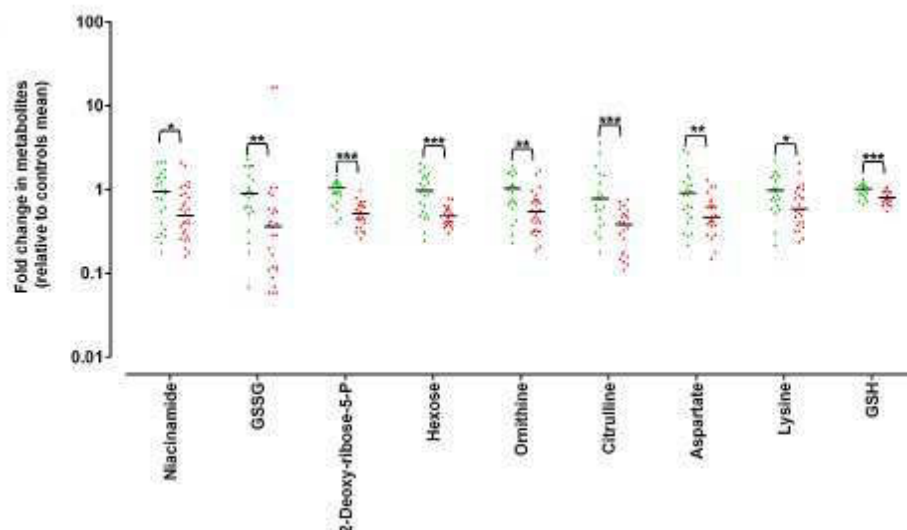


Figure 14. Application de tests statistiques univariés multiples (Darghouth et al; 2011b).

Comparaison des aires de métabolites érythrocytaires détectées en LC-HRMS à partir d'échantillons provenant de sujets sains (vert) et sujets drépanocytaires (rouge). La significativité du test statistique univarié est indiquée par le nombre d'étoiles [***] $p\text{value} < 0.001$, [**] $0.001 < p\text{value} < 0.01$, [*] $0.05 < p\text{value} < 0.01$]

Ces tests sont effectués indépendamment sur chaque variable de la matrice de données. Cette procédure est longue mais peut être automatisée en utilisant des logiciels écrits en langage R (Millot, 2009). Cependant, lors de l'application des tests multiples, le taux de tests faussement positifs augmente et une correction des valeurs obtenues est obligatoire. Des procédures de correction statistique des tests multiples existent mais certaines sont assez drastiques car elles ajustent le seuil de significativité au nombre de variables testées (exemple de la correction de Bonferroni (Daimon et al; 2012)). Dans le cadre d'analyses omiques où le nombre de variables est conséquent, en réduisant le taux de faux positifs, la correction des tests multiples augmente au contraire la probabilité de produire des faux négatifs et sous-estime donc le nombre de variables discriminantes. Par conséquent, la réduction de la dimensionnalité des données métabolomiques (matrices générées par les logiciels de détection automatique des signaux) avant de procéder aux tests statistiques multiples est fortement recommandée. Les analyses statistiques multivariées le permettent.

4.2.1.2. Analyses multivariées

Le recours aux analyses statistiques multivariées est justifié car ces dernières assurent la réduction de la dimensionnalité des données et permettent de résumer et de visualiser l'information biologiquement pertinente qui est contenue dans le jeu de données (Figure 15) (Trygg, Holmes, & Lundstedt, 2007). Il en existe deux types :

Les analyses statistiques descriptives, telles que l'analyse en composante principale (PCA, de l'anglais «principal component analysis»). Elles ne nécessitent pas d'information *a priori* sur la nature des

échantillons. Elles permettent de visualiser la répartition des échantillons dans un espace à une, deux ou trois dimensions. La PCA est souvent la première étape dans l'analyse des données. Grâce à elle, il est possible de voir comment le jeu de données se structure spontanément : selon un facteur confondant analytique tels que l'ordre d'analyse des échantillons, par exemple, ou selon le facteur biologique étudié. La PCA est suivie d'analyses dites explicatives qui facilitent la mise en évidence de signaux d'intérêt biologique.

Les analyses statistiques explicatives. Elles visent à expliquer une ou plusieurs réponses (Y). Dans ce cas, la nature des échantillons est renseignée et elle participe à la construction du modèle statistique. Parmi les méthodes explicatives, on trouve la régression PLS et également l'analyse discriminante PLS (PLS-DA). Ces méthodes peuvent également être utilisées dans un but prédictif. Un jeu de données initial ou une partie d'un jeu de données (jeu d'apprentissage) est alors utilisée pour construire un modèle explicatif. Ce modèle est ensuite appliqué à l'autre partie du jeu de données ou à d'autres jeux de données afin de prédire la classe d'appartenance des individus.

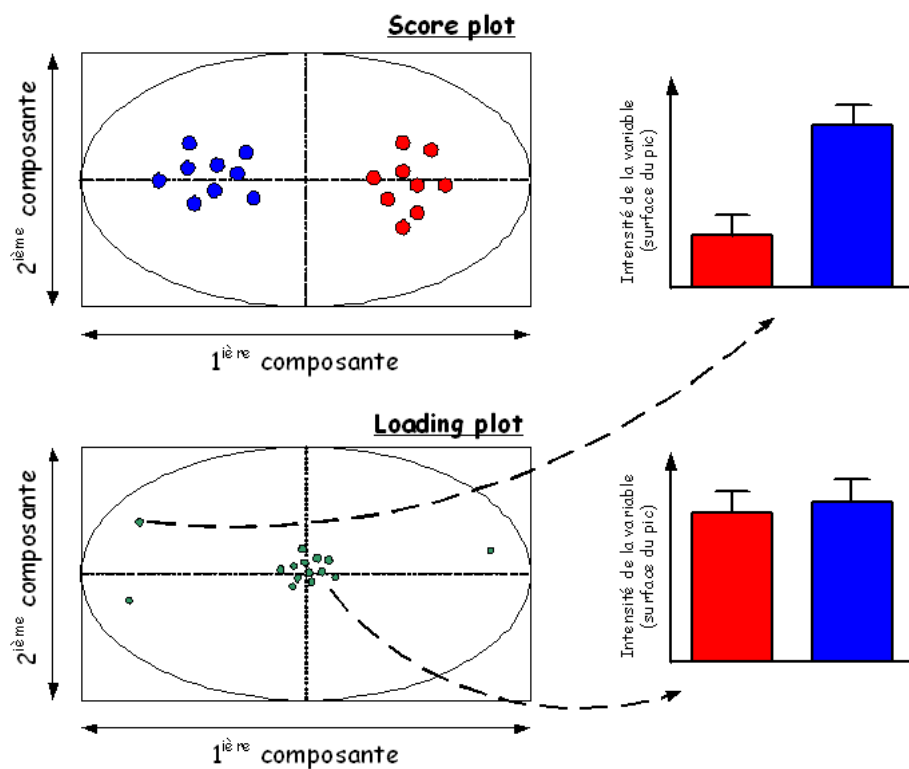


Figure 15. Analyses statistiques multivariées (adaptée de (Roux et al; 2011))

L'analyse statistique multivariée permet de résumer l'information contenue dans le jeu de données original dans un espace de dimension inférieure. Pour cela, un changement de coordonnées est effectué afin de définir de nouveaux axes, appelés **composantes principales**. Ces composantes latentes sont des combinaisons linéaires des variables d'origine calculées de façon à extraire le maximum de variance de façon itérative. En effet, la première composante extrait la plus grande source de variance, la seconde composante, orthogonale à la première, extrait la plus grande variance restante, et ainsi de suite, si bien que les dernières composantes contiennent surtout le bruit aléatoire. En pratique, les deux ou trois premières composantes permettent souvent d'expliquer plus de 90 % de la variance totale. Les échantillons sont

présentées sous forme de cartes factorielles des individus ou «*score plots*», grâce à leurs coordonnées dans le nouveau repère. Les échantillons appartenant à un même groupe sont proches les uns des autres. Parallèlement, une carte factorielle du poids des variables ou «*loading plot* » est établie. Elle représente la projection des variables initiales (*m/z*, *tr*) sur les composantes latentes. Les variables responsables de la séparation des deux groupes se répartissent de part et d'autre du graphique par analogie aux échantillons qui les sur-expriment ou sous-expriment, permettant ainsi de mettre en exergue des variables discriminantes.

Il existe des logiciels qui traitent les données en sortie des logiciels de détection automatique des signaux en réalisant annotations et analyses statistiques tel Matabolyzer (Mak et al; 2014). De plus, des suites de logiciels plus complètes sont également développées. Leur but est d'arriver à prendre en charge une analyse métabolomique depuis l'importation des données brutes à la mise en évidence et identification des métabolites d'intérêt. Cela a été entrepris avec des logiciels tels que MetaboAnalyst (Xia et al; 2009, Xia & Wishart, 2011), BIOJeti (Margaria, Kubczak, & Steffen, 2008) et MeltDB (Neuweger et al; 2008) qui sont des ensembles de modules multi-configurables dont les logiciels de détection automatique est XCMS, et qui proposent en plus une annotation et des outils d'analyses statistiques et de visualisation des résultats de la matrice de sortie.

4.2.2. Visualisation des données biologiques : Les réseaux métaboliques

La production de données d'analyse métabolomique biologiquement exploitables passe par le développement de méthodes d'exploration des résultats biologiques telle l'utilisation de graphiques avec des nœuds (ou métabolites) et des arêtes (ou réactions biochimiques) pour représenter les métabolites endogènes et les interactions existant entre eux et qui permettent de manière simple de visualiser des réseaux métaboliques (Kohl, Wiese, & Warscheid, 2011).

Historiquement, ces réseaux ont été étudiés de façon indépendante en se concentrant sur des sous-ensembles de molécules liées par des réactions biochimiques depuis le substrat de la première enzyme au produit final (Shlomi et al; 2007). Ces «voies métaboliques» ont d'abord été considérées comme des processus distincts tels que décrites dans les cartes métaboliques des bases de données biochimiques telles que KEGG (Kanehisa & Goto, 2000) ou MetaCyc (Caspi et al; 2014, Karp et al; 2000, Karp et al; 2002, Krieger et al; 2004). Bien que cette approche statique génère de l'information biochimiquement pertinente, elle donne une vision erronée du métabolisme puisque la plupart des métabolites sont en fait connectés car ils sont impliqués dans plusieurs voies en parallèle (par exemple, les produits intermédiaires ou finaux pourraient être les substrats initiaux d'une autre voie anabolique ou catabolique –voir Figure 16-).

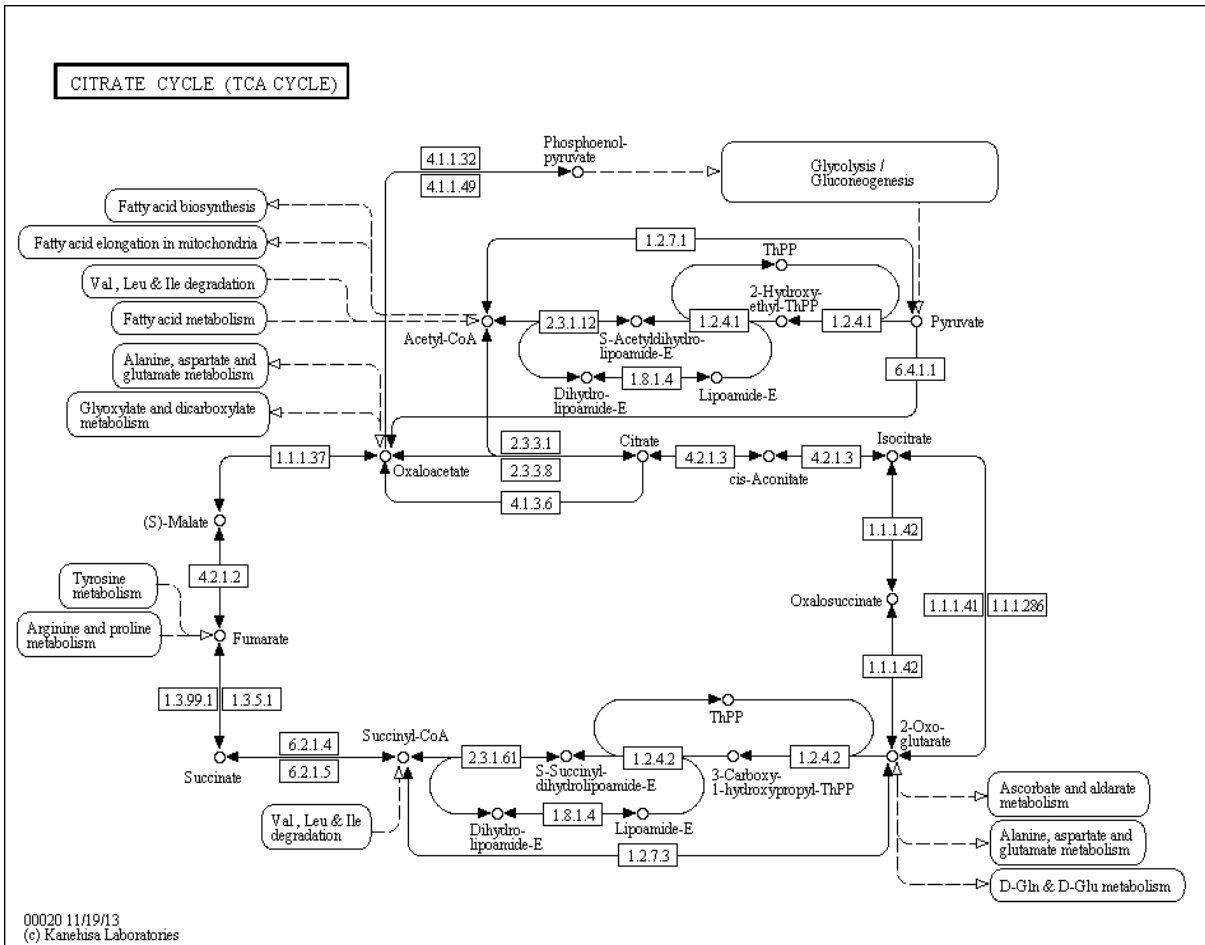


Figure 16. Voies métaboliques des acides tricarboxyliques selon KEGG (Cycle de Krebs)

Chaque nœud représente un métabolite. Les arêtes illustrent des réactions biochimiques. Les flèches indiquent le sens de la réaction donc définissent les substrats et produits de chaque réaction. Une arête ayant 2 flèches serait de ce fait une réaction réversible. Les rectangles sont les enzymes. Il existe des connexions entre le cycle de acides tricarboxyliques et d'autres voies métaboliques (qui sont indiquées mais non représentées). Cette représentation statique ne permet donc pas de visualiser l'implication d'un métabolite dans plusieurs voies métaboliques différentes.

Dans une perspective de biologie des systèmes, le métabolisme est aujourd'hui analysé au niveau global en intégrant le plus grand nombre possible de voies métaboliques en un seul réseau (Jourdan et al; 2010). En effet, grâce à l'annotation fonctionnelle de génomes complets et aux progrès des approches «omiques», un plus grand nombre de voies métaboliques peut être inclus dans un réseau théorique unique offrant la possibilité d'étudier un organisme dans sa globalité, comme en témoigne la reconstitution faite pour *Escherichia coli* (Feist et al; 2007). Des bases de données relationnelles (appelées BiGG, de l'anglais «*Biochemical Genetic and Genomic knowledgebase*») répertorient l'ensemble des connaissances biochimiques et génomiques d'organismes tels que ceux de la souris (Sigurdsson et al; 2010), des bovins (Seo & Lewin, 2009) et de l'homme (Duarte et al; 2007). Cependant, l'extrême complexité de la tâche de reconstitution soulève la question de la modélisation de l'ensemble des processus cellulaires chez les eucaryotes du fait de

la taille importante et de la configuration complexe de leur génome, en plus de la difficulté d'affectation d'un compartiment cellulaire. C'est pourquoi, seules certaines espèces modèles possèdent une BiGG.

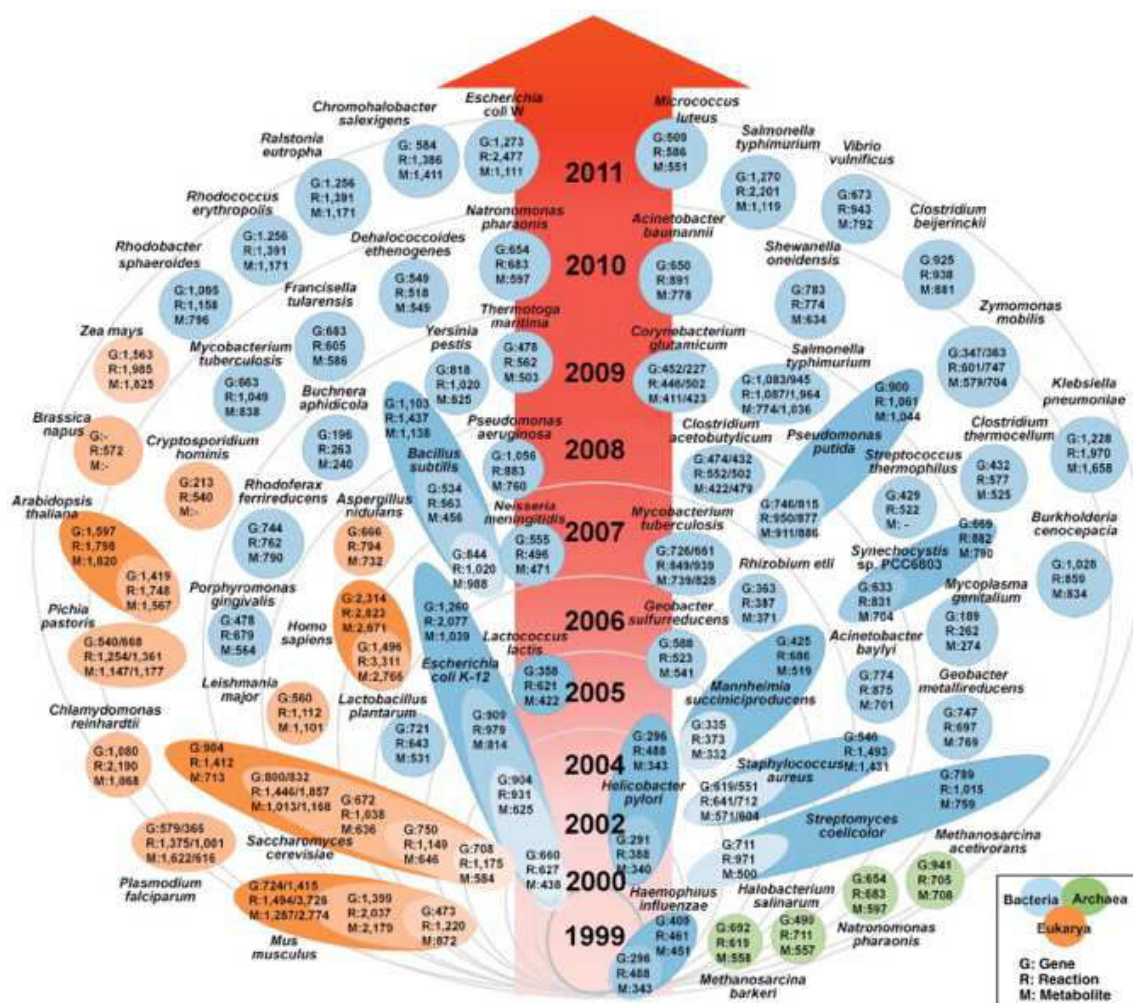


Figure 17. Calendrier des reconstructions des voies métaboliques à l'échelle du génome (Kim et al; 2012)

Ce calendrier indique les modèles métaboliques reconstruits depuis 1999 à l'échelle du génome répertoriés par Kim et coll (Kim et al; 2012). Les modèles sont représentés par des ellipses placées de façon à indiquer de sa publication. En orange les espèces eucaryotes, en bleu les procaryotes et en vert les archéobactéries pour lesquels les modèles métaboliques à l'échelle du génome ont été reconstruits. Les lettres G, R, et M représentent le nombre de gènes, de réactions et de métabolites, respectivement indiqué dans le modèle. Cette figure montre le nombre croissant de modèles de reconstruction métabolique à l'échelle du génome avec 30% de modèles reconstruits depuis 2010

Contrairement à la vision statique, l'avantage de cette approche est de permettre d'étudier les interactions entre les voies métaboliques, et de faciliter la découverte de voies alternatives. Dans ce contexte, la bioinformatique et les mathématiques apportent là encore une contribution indispensable à l'analyse globale des réseaux métaboliques grâce à de nouveaux algorithmes et à leur reconstruction automatique sous forme de graphes.

Comment construire un réseau métabolique ?

Des modèles de métabolisme à l'échelle du génome établissent un lien entre les informations issues du séquençage des génomes (base de données GOLD, NCBI, Plant Genome Central) et les analyses subséquentes jusqu'au phénotype lié à ce métabolisme. Ces reconstructions globales sont réalisables par des (i) approches haut-débit qui mettent en évidence des interactions ou plus fréquemment à l'aide de (ii) modèles *in silico* à partir des données génomiques et bibliographiques (Thiele & Palsson, 2010).

(i) Les données métaboliques acquises à haute résolution MS permettent l'inférence *ab initio* de relations métaboliques depuis des différences de masses précises entre métabolites calculées directement à partir du spectre de masse (Breitling, Pitt, & Barrett, 2006, Jourdan et al; 2008). Par exemple, relier une réaction biochimique d'hydrogénation et/ou de déshydrogénation à une différence de masse typique de 2,01565 Da entre le substrat et le produit détectés. Ainsi, une affectation automatique de connectivité (réaction chimique) nécessite une liste de transformations biochimiques avec la liste des différences de masses correspondantes.

(ii) La mise en place d'un réseau métabolique à niveau de l'échelle du génome quant à elle repose sur les réponses aux questions suivantes pour chaque enzyme répertoriée dans la base de connaissances BiGG: quels sont les substrats et produits de l'enzyme listée? Quels sont les coefficients stoechiométriques de la réaction biochimique? La réaction décrite est-elle réversible? Dans quel compartiment se déroule la réaction? (Feist et al; 2009).

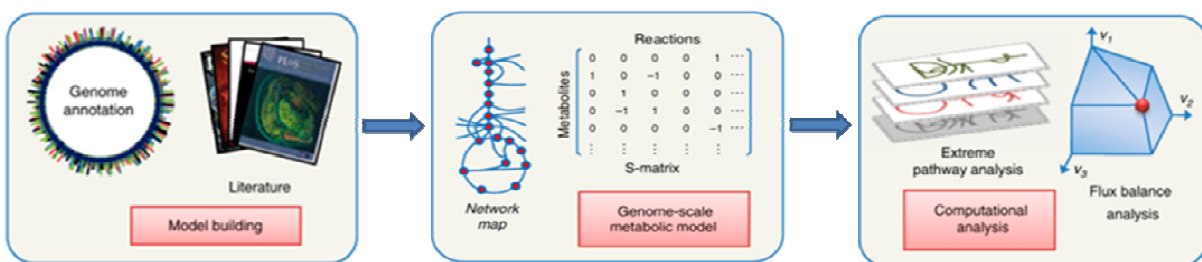


Figure 18. Etapes de reconstruction initiale du métabolome à l'échelle du génome

Le point de départ est la reconstruction automatisée à partir du génome, qui consiste en la collecte d'informations concernant tous les gènes qui codent pour les fonctions métaboliques dans les organismes ciblés ou phylogénétiquement proches (voir Figure 18). L'annotation du génome génère une liste d'enzymes codées par des numéros de classification (Enzyme Commission) communs à toutes les bases de données biochimiques. Ces enzymes apportent à leur tour des hypothèses de fonctions métaboliques. Cependant, les activités enzymatiques peuvent varier selon l'organisme et la plupart des bases de données ne sont pas spécifiques à un organisme, c'est pourquoi des annotations manquantes ou encore fausses peuvent se glisser dans cette étape de reconstruction des réseaux. Il convient alors d'affiner le modèle prédit et ce, en impliquant des experts spécifiques de l'organisme d'intérêt afin d'atténuer ce risque. Cette BiGG est convertie ensuite en un modèle de calcul ajusté à la réalité biologique.

Ces modèles *in silico* permettent, lors d'utilisations prospectives, la découverte *de novo* de voies métaboliques mais aussi d'interactions nouvelles (Shlomi et al; 2007) (Gevorgyan, Poolman, & Fell, 2008). Ceci a été particulièrement souligné par Rolfsson et coll. lors du calcul des voies manquantes à l'aide du 1^{er} réseau humain de reconstitution métabolique RECON1 pour la voie de l'acide iduronique (Duarte 2007). Ils ont démontré que la complétion à l'aide de SMILEY permettait d'émettre l'hypothèse de nouvelles fonctions dans le métabolisme humain (Rolfsson, Palsson, & Thiele, 2011).

Grâce à ces approches, les réseaux métaboliques de certains organismes tels que *Saccharomyces cerevisiae* sont complètement reconstitués. Aujourd'hui, les métabolites sont considérés dans leur contexte. Toutefois, cela ne suffit pas toujours à comprendre l'équilibre spatial et temporel qui ne peut être atteint qu'à travers des données expérimentales *in vivo* qui peuvent gérer les flux métaboliques (Likic et al; 2010).

4.2.3. Les bases de données métaboliques

L'objectif premier des bases de données est l'annotation des signaux obtenus en MS tel que présenté en PARTIE A.CHAPITRE II.3.2. Cela nécessite des requêtes par masse moléculaire ou par composition élémentaire. Les bases de données peuvent également contenir des informations biologiques afin d'assurer l'attribution de fonctions biologiques. A cet effet, plusieurs types de banques de données peuvent être distingués : (i) les bases de données chimiques générales (par exemple, PubChem) qui englobent des composés synthétiques et / ou composés naturels, (ii) des bases de données métaboliques qui traitent les voies métaboliques annotées telles que KEGG, (iii) les bases de données de métabolomiques spécifiques (par exemple, LIPIDMAPS en lipidomique) et (iv) les bases de données de spectre de masse (par exemple NIST). En fait, certaines peuvent combiner plusieurs de ces fonctions. Il est important de garder à l'esprit que toutes les bases de données sont en cours de développement, avec des implémentations et corrections apportées constantes. Par conséquent, des fonctions supplémentaires peuvent être envisageables dans un avenir proche. La Table 3 ci-dessous, résume de façon non-exhaustive les bases de données les plus utilisées en analyse métabolomique.

Databases	Conception	Accès	Données spectrales	Website (URL)	Référence
-----------	------------	-------	--------------------	---------------	-----------

Base de données chimique

ChEBI	European Bioinformatics Inst. (UK) / European Molecular Biology Lab.	●		www.ebi.ac.uk/chebi/	(Degtyarenko, 2008)
Chemfinder	Cambridge Soft (USA)	●		chemfinder.CambridgeSoft.com	(Brecher, 1998)
CHEMnetBASE (Dict. Nat. Prod.)	Chapman & Hall / CRC	○		www.chemnetbase.com/	
CSLS	CADD Lab. Med. Chem NCI, NIH (USA)	●		http://129.43.27.140/cgi-bin/lookup/search	
Enhanced NCI Database Browser	CADD Lab. Med. Chem NCI, NIH (USA), U of Erlangen-Nuremberg (Germ.)	●		http://129.43.27.140/ncidb2/-	(Nicklaus, 2001)
Madison Metabolomics Consortium Database	National Magnetic Resonance Facility, University of Wisconsin-Madison	●		http://mmcd.nmrfa.wisc.edu/	(Cui, 2008)
Merck index	Merck publishing	○			
NIST	National institute for standard and technology (USA)	⊙	●	www.nist.gov/srd/nist1a.htm	
PubChem	National Center for Biotechnology Information (USA)	●		www.ncbi.nlm.nih.gov/entrez/query.fcgi?db	
SDBS	National Institute of Advanced Industrial Science and Technology	●	●	riodb01.ibase.aist.go.jp/sdbs/	(Yamamoto, 1988)
Scifinder	American Chemical Society (USA)	○			

Base de données métaboliques ou métabolomiques

BiGG	University of California (USA)	●		bigg.ucsd.edu/	(Duarte, 2007)
Biocyc (humancyc, metacyc)	SRI International (USA)	●		biocyc.org/	(Romero, 2005; Caspi, 2008)
Fiehn Library	Fiehn Laboratory Univ California Davis – Genome center	●		http://fiehnlab.ucdavis.edu/Metabolite-	
Golm	Max Planck Institute for Molecular Plant Physiology (Germany)	●		csbdb.mpimp-golm.mpg.de	(Kopka, 2005)
HMDB	Department of Computing Science, University of Alberta (Canada)	●	●	www.hmdb.ca/extrIndex.htm	(Wishart, 2007)
KEGG ligand database	Kyoto University Bioinformatics center (Japan)	●		www.genome.jp/kegg/ligand.html	(Kanehisa, 2000)
Knapsack	RIKEN Plant Science Center (Japan)	●		http://kanaya.naist.jp/KNAPsACK/KNAPsACK.	(Shinbo, 2006)
Lipidmaps	LIPID MAPS Bioinformatics Core (USA)	●		www.lipidmaps.org/data/index.html	(Fahy, 2007)
Lipidbank	Japanese Conference on the Biochemistry of Lipids (Japan)	●		lipidbank.jp/.	(Watanabe, 2000)
Massbank	Keio university, university of Tokyo, Kyoto university, RIKEN plant Science center (Japan) and others	●	●	www.massbank.jp	(Horai, 2007)
Metabolome.JP	University of Tokyo, RIKEN Plant Science Center, Keio University, Kyoto University and NAIST	●		www.metabolome.jp/	(Arita, 2003)
Metlin	Scripps Center for Mass Spectrometry	●	●	metlin.scripps.edu	(Smith, 2005)
MoTo	Wageningen University,	●		appliedbioinformatics.wur.nl	(Moco, 2006)
MSlib	University of Alberta	●	●	http://www.ualberta.ca/~gjones/mslib.htm	

Table 3. Bases de données utilisées en métabolomique (adapté de Werner et coll.(Werner et al; 2008b))

Accès : (●) libre (éventuellement après identification), (⊙) limité, (○) payant, (■) données spectrales présentes.

4.3. Outils informatiques pour le partage et l'échange des données

Si l'accès et l'échange de données brutes non traitées a été facilité ces dernières années grâce au développement de formats universels (mzXML, mzData, mzML, netCDF) compatibles avec de nombreux logiciels commerciaux, mais aussi libres et gratuits, il n'en est pas de même pour les données traitées. Il n'existe pas encore de format reconnu et accepté par la communauté pour partager facilement les grandes quantités de données métabolomiques d'analyses LC-MS issues du traitement par des logiciels différents. Un certain nombre d'initiatives visant à uniformiser le stockage des données traitées de LC-MS ont déjà été publiées mais elles n'étaient jusqu'à maintenant que peu adaptées aux données de métabolomique. Citons par exemple le format FeatureXML (Sturm & Kohlbacher, 2009), bien adapté pour les données LC-MS, mais plutôt destiné à des applications dans le domaine de la protéomique. Récemment, un format plus complet, PeakML, développé spécifiquement pour les données de métabolomique, a été proposé par Scheltema et coll. (Scheltema et al; 2011). Ce format est compatible avec la plupart des logiciels existants et permet de stocker à la fois les informations sur l'acquisition (mode d'ionisation, séquence, données brutes...) et le traitement des données (mesure d'intensité, annotations et identifications...). De plus, le développement de modèles formels de déroulement d'analyse ainsi qu'une standardisation des méthodes d'analyse et de traitement de l'information au sein de la communauté devraient faciliter l'utilisation de ce type de données par les biologistes.

L'harmonisation de la dénomination des métabolites est également un sujet à aborder. Alors que la nomenclature IUPAC offre sans équivoque un seul nom par métabolite, son utilisation reste restreinte en métabolomique du fait de sa complexité notamment s'agissant de molécules pluri-fonctionnalisées et ramifiées (par exemple l'acide aminé tryptophane a pour nom IUPAC : acide- 2-amino-3-(1H-indol-3-yl)propanoïque). Ainsi, le nom commun est celui qui est le plus couramment utilisé. Or, plusieurs noms communs peuvent exister pour un même métabolite incluant différentes déclinaisons (alpha, cis...), et dont certains ne sont pas intuitivement affiliés à la même molécule (pour exemple, 14 synonymes sont répertoriés pour l'acide pyruvique dans la base de données HMDB dont le «pyruvate», «acide pyroracémique» ou encore «acide alpha kétopropionique»).

Dans ce contexte, les bioinformaticiens de l'infrastructure MetaboHUB (projet d'infrastructure française financé par les investissements d'avenir) propose d'aborder le problème de la dénomination des métabolites de la façon suivante :

1. Considérer d'abord le nom commun associé à un identifiant « ID » tel que décrit dans les bases de données publiques si disponible (d'abord recherche dans HMDB, puis PubChem...) ou bien le nom IUPAC, tout en s'assurant de mettre les préfixes en dernier (par exemple pour «l'alpha-pinene» écrire «pinene-alpha»). Le numéro CAS, qui est propriétaire et non retrouvé dans toutes les bases de données publiques n'est pas pris en compte.

2. joindre la représentation bidimensionnelle de la structure InChelKey (International Chemical Identifier).

Des initiatives comme celles-ci devraient se multiplier et être partagées dans le WorldWideWeb pour permettre son évolution et propagation.

4.4. Fusion des données

De plus en plus d'analyses métabolomiques incluent des données obtenues à partir de différents types de technologies (mêmes échantillons analysés dans des plateformes analytiques différentes) ou de plusieurs expériences indépendantes acquises avec le même instrument. Combiner ces données permet d'élargir la couverture métabolique mais aussi assure une meilleure identification du métabolite si ce dernier est détecté dans plus d'une plateforme. Les données multi expériences répondent quant à elle à la nécessité de diviser des séquences d'analyse parfois trop longues. Dès lors, des variables peuvent être communes alors que d'autres ne seront spécifiques qu'à un seul jeu de données. Il est donc important d'intégrer l'ensemble des données souvent complémentaires tout en s'assurant de ne pas créer de biais lors de la fusion. Un prétraitement des matrices à fusionner est primordial (des variables ou variables latentes) –voir PARTIE A.CHAPITRE II.4.2.1-.

Différents niveaux de fusion de données ont été décrits :

La **fusion de bas niveau** se fait à partir des matrices de données directement issues du traitement des données brutes par les logiciels de détection automatiques et d'alignement des signaux. En théorie, elle consiste en une simple concaténation de matrices de données. Par conséquent, toutes les variables mesurées sur les échantillons sont simplement mises côte à côte nécessitant que les fichiers de sortie de chacune des expériences soient au même format. Des procédures plus complexes peuvent également être mises en œuvre en vue de regrouper les variables communes tels que la modélisation des temps de rétention des ions de même masse précise d'une plateforme à l'autre (Vaughan et al; 2012) en plus d'un calcul de corrélation afin de les jumeler.

Dans tous les cas, la fusion à bas niveau génère donc une matrice de données avec un rapport échantillons/variables très défavorable aux analyses statistiques. Toutes ces variables ne sont cependant pas pertinentes. Elles peuvent en effet correspondre à des artéfacts générés au cours du processus informatique de détection des signaux et des outils statistiques tels la PLS peuvent être utilisés afin de réaliser une sélection préliminaire des variables à fusionner. Les matrices de données épurées sont ensuite concaténées pour l'obtention de la matrice finale. Cette procédure constitue la **fusion de niveau intermédiaire**. Des méthodes telles que «consensus-PCA», «consensus OPLS-DA» ou «MultiBlock-PLS» peuvent répondre à ce type d'attente (Smilde et al; 2005) (Boccard & Rutledge,

2013). Une application intéressante serait la fusion de données de protéomiques et de métabolomiques (Thomas Moyon et al; 2012), ou encore la fusion de données de métabolomiques acquises dans les deux modes d'ionisation (Boccard & Rutledge, 2013). D'autres approches ont également permis de relier des données métabolomiques obtenues par NMR et par GC-MS (Smolinska et al; 2012a) ou encore données de NMR obtenues à partir de deux matrices biologiques différentes (plasma et LCR) (Smolinska et al; 2012b).

Enfin, il est possible de construire des modèles statistiques à partir de jeux de données préalablement filtrés et annotés et issus de différents types d'analyses. La fusion de données réalisée à ce niveau est dite **fusion de haut niveau**. Il s'agit du niveau de fusion le plus abouti pour réaliser l'interprétation biologique des données dans la mesure où les variables prises en compte sont bien caractérisées.

La fusion des données ouvre des perspectives intéressantes sur un plan analytique (fusion de différentes séries d'analyse, fusion de jeux de données métabolomiques acquis par des techniques différentes...) mais aussi biologique (fusion de données métabolomiques de matrices biologiques différentes appartenant aux mêmes individus,...). Cependant, les résultats générés doivent être validés et vérifiés avec précaution. Ceci nécessite un retour aux données brutes soit aux variables d'origine ce qui est souvent plus compliqué.

CHAPITRE III. APPROCHE LIPIDOMIQUE

1. Lipides : définition et classification

La « Lipidomique » consiste en l'analyse du lipidome et de son interaction avec son milieu (Watson, 2006). Ce terme a été utilisé la première fois en 2001 (Kishimoto et al; 2001) pour faire référence à l'ensemble des lipides d'une cellule, un tissu et/ou d'un organisme (Dennis, 2009, Wenk, 2005).

Les lipides assurent de multiples fonctions biologiques. En effet, ils sont les principaux constituants des membranes cellulaires dont l'intégrité est primordiale au fonctionnement physiologique de la cellule et aux échanges entre différents compartiments cellulaires. En outre, l'hydrophobicité qu'ils apportent est importante aux fonctions et interactions des protéines membranaires (phospholipides, sphingolipides et glycolipides). Les lipides représentent aussi une réserve d'énergie et certains d'entre eux génèrent des 2^{nds} messagers tels les diacylglycérols, les eicosanoïdes et céramides, qui sont des médiateurs dans les voies de signalisation cellulaire.

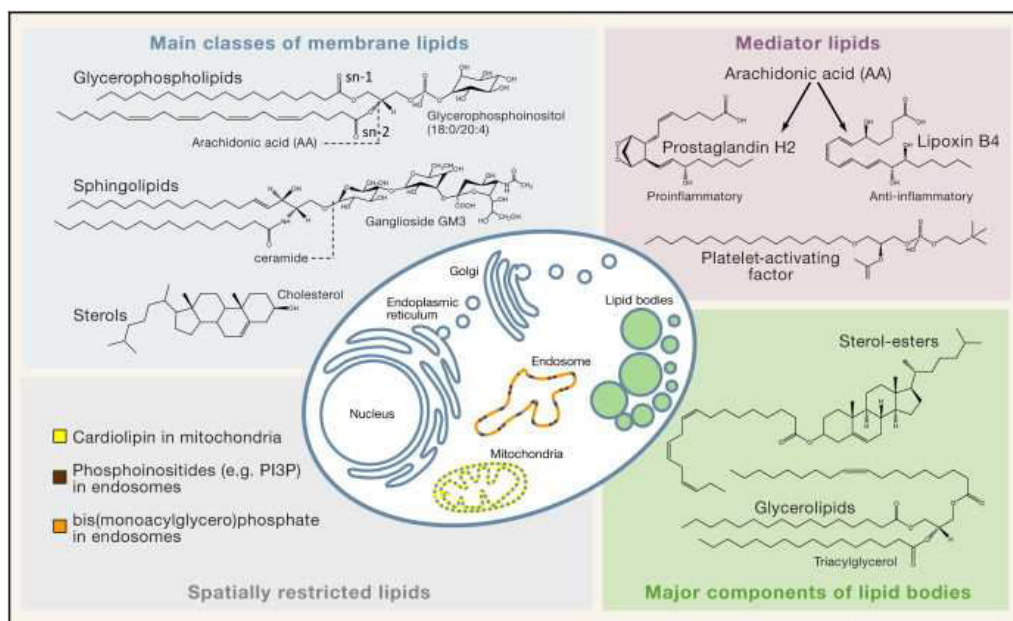


Figure 19. Lipides et fonctions biologiques (Wenk, 2005)

Une classification des lipides présents chez les espèces mammifères a été établie en 2005 par «l'International Lipid Classification and Nomenclature Committee (ILCNC)» dont une mise à jour a assuré l'étendue aux lipides présents chez les plantes, bactéries et champignons (Fahy et al; 2009). Selon cette dernière classification, il existe 8 grandes familles de lipides : acides gras, glycérolipides, glycérophospholipides, sphingolipides, stérols, prénols, saccharolipides et polycétides. Dans ce qui suit, nous présenterons les familles les plus communes de lipides présents chez les mammifères à

savoir (i) lipides neutres incluant la famille des esters de cholestérol et des glycérolipides (ii) lipides polaires qui incluent la famille des glycérophospholipides et sphingolipides (iii) et leurs métabolites

- (i) Les **lipides neutres** correspondent principalement au *cholestérol* (qui est fortement présent au niveau de la couche laminaire des membranes plasmiques), à ses esters ainsi qu'aux *acylglycérols mono, di ou tri* substitués (MG, DG et TG) qui constituent les lipoprotéines.

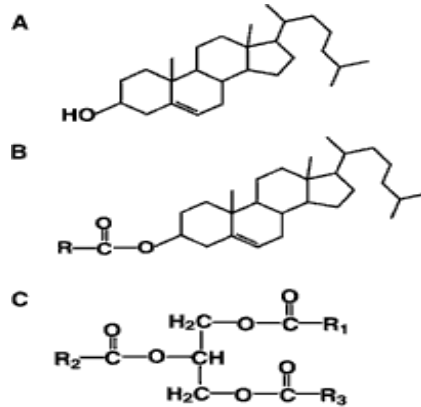


Figure 20. Structures de lipides neutres (Han & Gross, 2005)

- (ii) Les **lipides polaires** dont on distingue 2 familles importantes :

Les **glycérophospholipides** pour lesquels la position *sn*-3 du glycérol est estérifiée par un groupement phosphate, l'une ou les deux autres positions étant estérifiées par des acides gras principalement. Au groupement phosphate en position *sn*-3 viennent se lier différents groupements polaires de type choline, éthanolamine, glycérol, myo-inositol ou une sérine et qui à leur tour déterminent les classes suivantes: *glycérophosphatidylcholine*, *glycérophosphatidyléthanolamine*, *glycérophosphatidylglycérol*, *glycérophosphatidylinositol*, *glycérophosphatidylsérine* respectivement. La classe des acides phosphatidiques quant à elle correspond aux molécules dans lesquelles le groupement phosphate reste non estérifié.

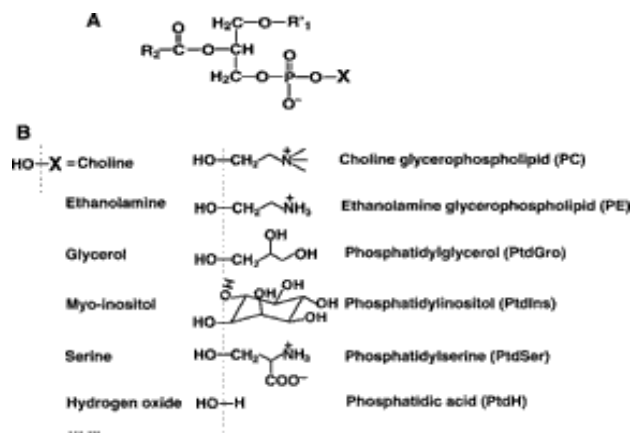


Figure 21. Structures des différentes classes de la famille des glycérophospholipides (Han & Gross, 2005)

De plus, au sein des classes de glycérophospholipides, la nature de la liaison covalente entre la chaîne acyl et la position *sn*-1 du glycérol peuvent être de 3 types et définissent de ce fait 3 sous-classes : les *phosphatidyl* ou *acyl* (liaison ester), *plasmeryl* ou *plasmalogène* (liaison vinyl ether) et *plasmanyl* (liaison alkyl ether). Ces liaisons sont le plus souvent rencontrées en position *sn*-1 d'espèces glycérophosphatidylcholine et glycérophosphatidylethanolamine.

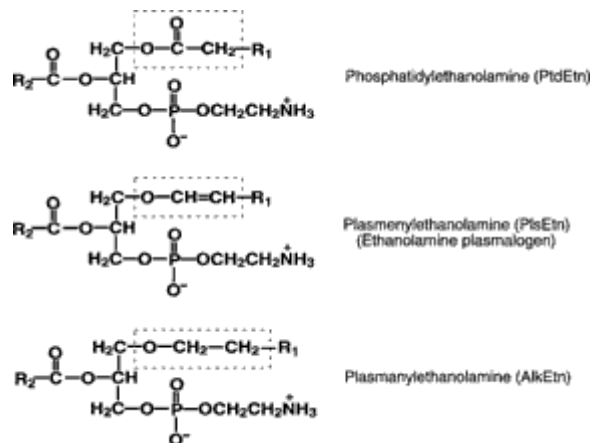


Figure 22. Structures des sous-classes de glycérophospholipides (Han & Gross, 2005)

En plus des classes et sous-classes, la présence d'un ou deux acides gras, leur position en *sn*-1 ou *sn*-2, leur nature avec longueur de chaîne, le degré d'insaturation et la localisation des doubles liaisons génère théoriquement plus de 1000 espèces de glycérophospholipides (dont 500 ont déjà été décrites dans des cellules murines cancéreuses de myélome NS-1 (Taguchi et al; 2000).

L'unité biochimique sur laquelle se bâtissent les **sphingolipides** est le *céramide*. Il est composé d'un acide gras qui peut être saturé, insaturé ou encore même hydroxylé, de longueur de chaîne variable (16 à 26 atomes de carbone) et fixé par son groupement carboxylique à la fonction amine d'un amino-alcool à longue chaîne (de 18 à 20 atomes de carbone), appelé base sphingoïde. Ces bases sphingoïdes existent sous plusieurs formes : (i) la sphinganine qui est un amino-alcool saturé à 18 atomes de carbone d18 :0, (ii) la sphingosine qui présente une insaturation d18 :1, (iii) sphingadiénine qui présente deux insaturations d18 :2 (iv) eicosasphingosine qui est un amino-alcool à 20 atomes de carbone avec une insaturation d20 :1, (v) ainsi que des formes hydroxylées exemple la 4-hydroxy sphinganine communément appelée phytosphingosine.

De cette unité céramide dérivent les glycosphingolipides constitués d'un élément saccharidique «*cérébrosides*» ou d'une chaîne oligosaccharidique fixée sur la fonction alcool primaire de la base sphingoïde. La première série est celle des *glucosylcéramides* pour lesquels le premier sucre fixé est un glucose. S'en suit la fixation obligatoire d'un galactose formant la classe des «*lactosylcéramides*». Viennent se fixer des galactoses –après perte d'une molécule d'eau- et molécules d'acide sialique formant la classe des «*gangliosides*».

Par contre, si le premier sucre fixé au céramide est un galactose, il y a formation de la série *galactosylcéramide*, qui peuvent être sulfatés en position 3 à leur tour et former les « *sulfogalactosylcéramides* » couramment appelés « *sulfatides* ». La classe des « *sphingomyelines* » provient elle aussi des céramides qui se lient via une liaison phosphodiester à une tête polaire choline.

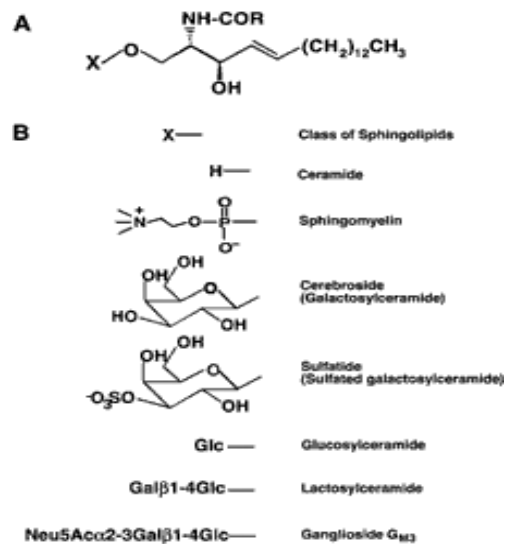


Figure 23. Structure des sphingolipides (Han & Gross, 2005)

(iii) Un dernier groupe associé aux lipides est celui de leurs métabolites qui sont leur produit de catabolisme ou encore leur précurseurs. Citons les *acides gras libres* non estérifiés, les *acylcarnitines*, les *acylCoAs*, la sphingosine-1-phosphate et autres lipides bioactifs qui sont impliqués dans des processus physiologiques et pathologiques (Han & Gross, 1991).

Au vu de cette diversité structurale, différentes méthodes d'analyses ont été utilisées pour l'investigation des lipides.

2. Analyse des lipides et analyses lipidomiques

Bien que l'analyse lipidomique intégrative au sens « omique » du terme soit une discipline récente, l'intérêt pour l'analyse des lipides a toujours été présent du fait de leur implication dans différents processus biologiques. En effet, des méthodes biochimiques reposant sur des tests enzymatiques, immunologiques ou encore le marquage métabolique des lipides par l'utilisation de précurseurs radioactifs (^3H , ^{14}C et ^{32}P) ou fluorescents (NBD) des sources de carbone ou des têtes polaires des phospholipides ont été adaptées à l'analyse des lipides. Des approches analytiques traditionnelles ont également servi à analyser des extraits lipidiques, telle l'électrophorèse sur gel (Vitic, Nikolic, & Nikolic, 1981) ou sur papier (MICHALEC, 1963) ou la chromatographie sur plaque

(Bodennec et al; 2000). Ces approches ont évolué avec les avancées technologiques en chimie analytique. Ainsi, des méthodes analytiques de pointe (NMR, HPLC, MS) combinées à des moyens bioinformatiques ont été appliquées avec succès à l'analyse de lipides (Wenk, 2005) (voir Table 4)

Table 1 Experimental approaches using extracts in lipidomics research			
Experimental approach	Lipid classes covered	Advantages	Disadvantages
Chromatography			
Thin-layer chromatography (TLC)	Solvent systems established for most lipid classes	Very established technique; technically relatively easy; does not require sophisticated instrumentation; spot chromatograms allow for rapid screening of mutant extract libraries.	Low resolution and sensitivity limits many lipidomic applications; detection of lipid by iodine vapour and (class-specific) dyes and radioactivity.
High-performance liquid chromatography (HPLC)	Many lipids, including sterols, GP, TG, DG, FA and lipid headgroup derivatives.	Well established with worked out reverse- and normal-phase conditions available; ease of automation; very quantitative.	Detection by refractive index or mass detector (lipids, as defined here, in general do not absorb visible and UV light effectively); medium sensitivities in general.
Gas chromatography (GC)	Non-polar compounds such as TG; derivatized FA and sterols.	Very widely used for determination of fatty-acid composition, detection generally by MS.	Requires volatile compounds or derivatization of polar lipids.
Mass spectrometry			
ESI	Polar compounds such as GP (more apolar compounds can be analysed by APCI)	Direct detection by <i>m/z</i> ; high sensitivity and resolution; direct profiling of complex lipid mixtures; ease of automation; compatible with upfront LC separation.	Suppression of ionization, in particular in the case of crude extracts and when low-abundance species are to be analysed; absolute quantification requires considerable efforts (for example, class and mass dependent internal standards).
MALDI	Many lipids including complex glycolipids.	Direct detection by <i>m/z</i> ; buffer and salt contaminants generally well tolerated; can be combined with prior TLC separation.	Suppression of ionization, in particular in the case of crude extracts and when low-abundance species are to be analysed; matrix backgrounds.
NMR			
³¹ P	Phospholipids.	Direct measurement; non-destructive; quantitative.	Line broadening of lipids in aqueous solutions; low sensitivity.
¹ H	All lipids.	Direct measurement; non-destructive; powerful technique for structural analysis of purified compounds.	low sensitivity, spectra dominated by very abundant lipids (cholesterol, PC).
Biochemistry			
Assays using immobilized lipids (for example, monolayer adsorption, langmuir blot/spot films, immobilized lipid biosensors, lipid blots/beads)	Many lipids both in pure form or mixtures.	Sensitive approaches which allow determination of interaction of ligands with lipids.	Functional immobilization of lipids difficult; technically tedious; automation and throughput generally limited.
Assays using lipid in solution/suspension (for example, optical, calorimetric, radiometric approaches using liposomes and micelles)	Many lipids both in pure form or mixtures.	Solution conditions; binding studies; often quantitative; enzyme assays, some which of high throughput.	Often experimentally challenging; optimization of conditions can require significant effort.
Lipid antibodies	Very few.	Cell biological studies (for example, subcellular localization) possible.	Specificity of antibodies.
Reactive lipids (for example, photoactivatable)	Few.	Identification of lipid-binding proteins.	Specificity; probes limited.

DG, diacylglycerol; ESI, electrospray ionization; FA, fatty acid; GP, glycerophospholipids; LC, liquid chromatography; MALDI, matrix-assisted laser desorption/ionization; MS, mass spectrometry; TG, triacylglycerol.

Table 4. Approches d'analyse de lipides (Wenk, 2005)

2.1. Méthodes non spectrométriques :

La chromatographie sur couche mince (CCM ou TLC de l'anglais «*Thin-Layer Chromatography*») est l'une des méthodes les plus anciennement utilisées du fait de sa rapidité de mise en œuvre pour l'analyse d'extraits lipidiques (White et al; 1998). Des méthodes de CCM ont ainsi été développées pour la plupart des classes lipidiques, le support le plus couramment utilisé étant le gel de silice. Ces méthodes ne requièrent pas de matériel sophistiqué mais posent le problème d'instabilité des espèces lipidiques suite à leur exposition prolongée à l'air et aux solvants

de migration notamment pour les espèces présentant des temps de migrations longs. De plus, la CCM souffre d'un manque de spécificité et de résolution tout comme l'utilisation de colonnes d'extraction sur phase solide SPE (de l'anglais : **solid-phase extraction**)(Bodennec et al; 2000,Fauland et al; 2013).

D'autres méthodes chromatographiques plus adaptées ont ainsi été utilisées, à commencer par **la chromatographie en phase gazeuse (CG)**, qui est appropriée pour l'identification des acides gras libres (Gutnikov, 1995). Des détecteurs comme celui à ionisation de flamme (FID) y sont souvent couplés (Albert et al; 2009). Bien que la séparation en CG soit efficace, le rendement de dérivation peut être faible, mais aussi les hautes températures requises pour l'élution en phase gazeuse peuvent engendrer l'isomérisation des espèces insaturées mais surtout la décomposition des lipides complexes libérant des AG dont l'attribution aux molécules parents est délicate.

Dans ce contexte, **la chromatographie en phase liquide (LC)** est apparue plus intéressante notamment du fait de sa polyvalence. En effet, des rétentions en phase normale permettent une résolution selon les classes lipidiques (tête polaire) alors que la phase inverse sépare plutôt des composés selon la longueur de chaînes, mais aussi le nombre d'insaturations. Des colonnes de chromatographie liquide haute performance (HPLC de l'anglais High Performance Liquid Chromatography) ont de ce fait été couplées avec succès à des détecteurs UV (Kuksis et al; 2009), ou encore détecteur évaporatif à diffusion de lumière DEDL (Imbert et al; 2012b,Kuksis et al; 2009,Yunoki et al; 2009) pour l'analyse d'extraits lipidiques.

Après l'électrophorèse classique (gel, papier), **l'électrophorèse capillaire (CE)** de zone avec une détection UV indirecte est la plus utilisée en analyse de lipides (Otieno & Mwangela, 2008). Or, du fait du manque de solubilisation des lipides dans les tampons aqueux utilisés, l'utilisation de la CE est limitée par des problèmes d'agrégation et précipitation. De plus, ces méthodes décrites ne fournissent pas à elles seules d'informations structurales. Des analyses de lipides et analyses lipidomiques basées sur la spectroscopie NMR et la MS ont donc vu le jour.

La spectroscopie $^1\text{H-NMR}$ peut en effet être utilisée pour réaliser des analyses lipidomiques. Citons pour exemple l'analyse des lipides érythrocytaires menée par Adosraku et coll. et qui a démontré l'existence d'une grande proportion d'espèces plasmalogènes PE dont les acides gras présentent un haut degré d'insaturation, en plus de la prédominance de la sphingosine comme base sphingoïde des sphingolipides (Adosraku et al; 1994). De plus, ses variantes $^{13}\text{C-NMR}$ (Mahrous, Lee, & Lee, 2008) ou encore $^{31}\text{P-NMR}$ pour l'analyse des phospholipides (Schiller & Arnold, 2002) sont également informatives. Néanmoins, en plus du manque de sensibilité, les spectres NMR sont dominés par les espèces abondantes type glycérophosphatidylcholine et cholestérol, et les spectres $^1\text{H-NMR}$ sont souvent complexes et ne permettent donc pas la discrimination entre les résonances des différents signaux ce qui complique l'interprétation des spectres.

Enfin, des études récentes traitent également de l'application de la spectroscopie Raman (Michael-Jubeli et al; 2011) en analyse de lipides. Une récente application *in vivo* sur les microalgues a démontré l'intérêt de la spectroscopie Raman dans la détermination du degré d'insaturation des lipides (Wu et al; 2011). Dans ce contexte, la MS s'est révélée être une méthode de choix pour l'analyse de lipides, que ce soit en introduction directe (analyses dites «shotgun») ou couplée à des méthodes séparatives de type chromatographiques.

2.2. Méthodes spectrométriques

Du fait de sa polyvalence, la MS est devenue une méthode de référence en analyse lipidomique. En effet, ses différentes sources d'ionisation et différents analyseurs permettent la détection, l'identification à travers les différentes analyses structurales, la quantification mais aussi la localisation tissulaire des espèces lipidiques.

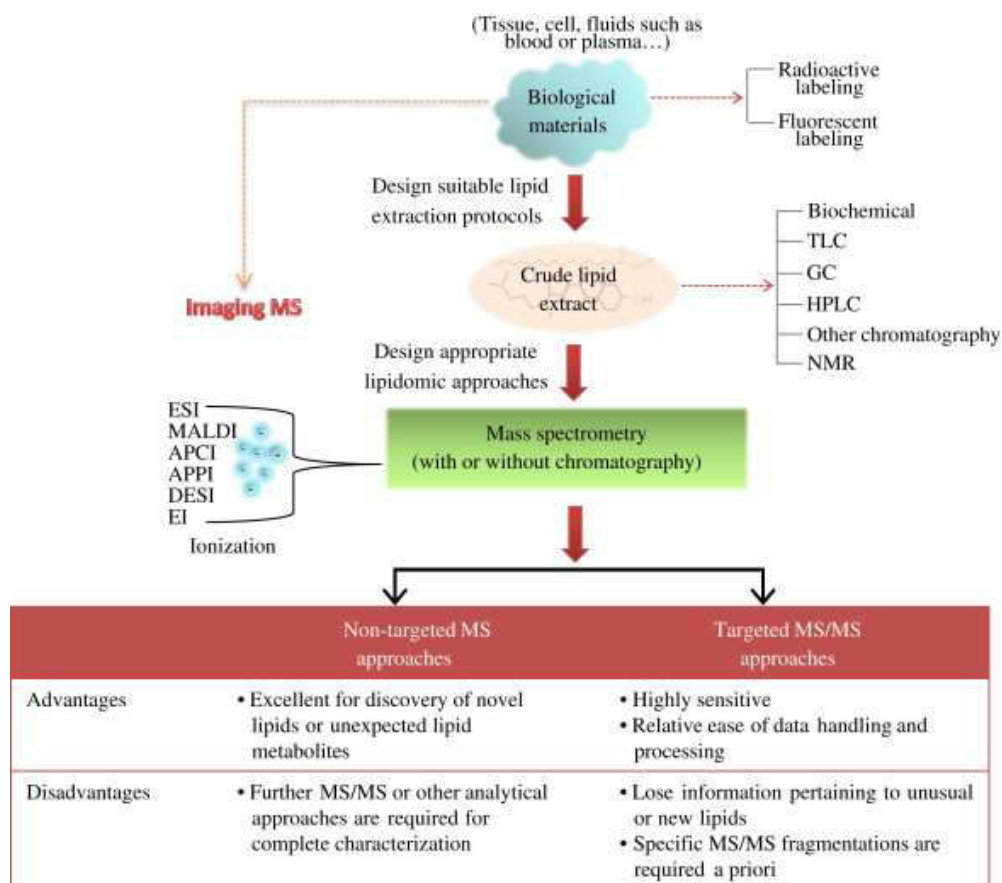


Figure 24. Analyse de lipides par spectrométrie de masse (adaptée de (Lam & Shui, 2013))

Tout d'abord comme décrit en section précédente (PARTIE A.CHAPITRE III.2.1), **l'ionisation par impact électronique (EI)**, le plus souvent couplée à la GC, requiert de fortes énergies pouvant faire migrer des doubles liaisons mais aussi induire des fragmentations, rendant l'identification mais aussi l'attribution des AG délicate. Des sources d'ionisation plus douces ont de ce fait été utilisées pour l'analyse des lipides. Faisant appel à des gaz réactants les rendant moins énergétiques, **l'APCI**

(Byrdwell, 2001, Byrdwell, 2005) et l'APPI (Imbert et al; 2012a) sont utiles en analyse de lipides notamment dans la détection des PC (Delobel, Touboul, & Laprevote, 2005), sphingolipides (Delobel et al; 2006), AG et glycérolipides (Cai & Syage, 2006).

Les sources **MALDI** (Tanaka et al; 1988) sont également couramment utilisées en lipidomique. Elles requièrent un choix critique de la matrice, l'une des plus couramment utilisée est l'acide 2,5 dihydrobenzoïque (DHB) (Preiano et al; 2012). La source MALDI est facilement couplable à des analyseurs de type TOF (Al-Saad et al; 2003), ou encore des pièges ioniques (Marto et al; 1995). Des couplages avec des méthodes chromatographiques hors ligne notamment TLC-MALDI (Guittard, Hronowski, & Costello, 1999) ont également été décrits.

Le second avantage de ces sources MALDI est qu'elles permettent l'analyse directement depuis une phase solide assurant la détermination de la distribution des lipides directement sur des coupes de tissu ouvrant lieu à des applications en imagerie par spectrométrie de masse (Jackson et al; 2005). L'imagerie des lipides par MS sera présentée en PARTIE A.CHAPITRE III.2.2.5.

Malgré la multitude de sources déjà décrites en analyse de lipides, l'**électrospray (ESI)** reste la plus utilisée en profilage d'espèces de lipides (Koeberle et al; 2013, Pulfer & Murphy, 2003, Singh & Prasad, 2011). A ces différentes sources d'ionisation, différents analyseurs sont couplés : les pièges ioniques (Jonasdottir et al; 2013), triples quadripôles (Shiva et al; 2013) ainsi que leurs variantes QqLIT (Spickett, Reis, & Pitt, 2011) et QqTOF (Ekroos et al; 2002), les TOF (TOF/TOF (Gidden et al; 2009) et QIT/TOF (Hayasaka et al; 2008), les orbitraps (Schuhmann et al; 2011, Schwudke et al; 2007), (Sokol et al; 2013) mais aussi les FT-ICR (Ivanova et al; 2001).

Plusieurs démarches ont été décrites faisant appel à de l'introduction directe avec expériences MS/MS appelée « shotgun lipidomics (SL) » ou « couplage avec la chromatographie liquide LC-MS ou LC-MS/MS ». L'évolution des approches lipidomiques notamment en SL n'a été rendue possible que grâce au développement des analyseurs MS et des outils bioinformatiques associés. Pour une meilleure compréhension, les méthodologies analytiques et outils bioinformatiques correspondants seront présentés à la suite ayant été développés pour répondre aux mêmes impératifs.

2.2.1. Approches lipidomiques par introduction directe (approches «Shotgun»)

Le profilage d'espèces lipidiques depuis des échantillons biologiques par introduction directe dans un spectromètre de masse équipé d'une source ESI a été décrit pour la première fois en ESI par Han et Gross en 1994 (Han & Gross, 1994). Le temps d'acquisition relativement long permet d'effectuer différents types d'expériences successives à différentes énergies de collision, mener différents types de fragmentations ou encore effectuer un changement de polarité (Schuhmann et al; 2012).

Classiquement réalisées avec des spectromètres de masse à analyseurs triple quadripolaires, les approches «shotgun» permettent de détecter et d'identifier des classes prédéterminées grâce aux modes MS/MS (i) de balayages d'ions parents (PIS), tel que démontré par Hsu et coll. pour la détection de sulfatides (Hsu, Bohrer, & Turk, 1998), (ii) de pertes de neutres (NLS) (Han et al; 2006), ou encore (iii) de détection de réactions sélectionnées (SRM) (Taguchi, Nishijima, & Shimizu, 2007) (voir Table 5). De plus, la régiospécificité (position de l'AG sur le glycérol) peut également être déterminée à travers l'intensité des fragments correspondants (Bou et al; 2010, Hsu & Turk, 2000). La méthode «shotgun» représente encore aujourd'hui l'approche la plus couramment utilisée en analyse lipidomique car elle permet de détecter et de quantifier un grand nombre d'espèces lipidiques dans de grandes séries d'échantillons.

D'autres types d'analyseurs peuvent également être utilisés. Ainsi, bien que les pièges ioniques utilisés en mode CID dans des conditions d'excitation résonante soient limités par l'existence du phénomène de «low mass cut-off» qui empêche la détection d'ions produits de masse inférieure à 20% de celle de l'ion parent, ils présentent l'avantage de permettre la réalisation d'expériences de type MSⁿ qui sont bien utiles pour l'élucidation structurale des espèces lipidiques (Hsu & Turk, 2008b, Hsu & Turk, 2008a, Larsen et al; 2001). D'autres modes de fragmentation non résonants HCD sont également couramment utilisés en analyse de lipides (Bird et al; 2011). De plus, d'autres modes d'activation tel que l'ECD (electron capture dissociation) (McFarland et al; 2005) ont été utilisés pour l'analyse d'ions multichargés de GM1 donnant accès à des informations structurales sur la base sphingoïde, l'AG, la partie céramide mais aussi sur les sucres. Ces techniques ne sont généralement pas envisageables sur les instruments à piégeage d'ions conventionnels mais plutôt dans les cellules ICR (Syka et al; 2004).

Classes lipidiques	Références bibliographiques	Ion à identifier	Scans pour identifier les classes	Scans pour identifier les acides gras
PC	Yang et al, 2009	[M+Li]+	NLS 189, 35eV NLS 193, 35eV	NLS (59+FA), 40eV
lyso PC	Yang et al, 2009	[M+Na]+	NLS 59, 22eV NLS 205, 34eV	PIS 104, 34eV PIS 147, 34eV
PE et lysoPE	Han et al, 2005	[M-H]-	PIS 196, 50eV NLS 222, 30eV	PIS (FA-H), 30eV
PI et lysoPI	Han et al, 2004	[M-H]-	PIS 241, 50eV	PIS (FA-H), 47eV
PS et lysoPS	Han et al, 2004	[M-H]-	NLS 87, 24eV	PIS (FA-H), 30eV
PA et lysoPA	Han et al, 2004	[M-H]-	PIS 153, 35eV	PIS (FA-H), 30eV
PG et lysoPG	Han et al, 2004	[M-H]-	PIS 153, 35eV	PIS (FA-H), 30eV
Triacylglycerol	Han et al, 2001	[M+Li]+		NLS (FA), 35eV
Sphingomyelin	Yang et al, 2009	[M+Li]+	NLS 213, 50eV	NLS (base sphingoïde)
Céramide	Han et al, 2002	[M-H]-	NLS 256 (d18:1), 32eV	
Héxosylceramide	Hsu et al, 2001 Han et al, 2005	[M+Li]+	NLS 162, 50eV	NLS (base sphingoïde)
Sulfatides	Hsu et al, 1998	[M-H]-	PIS 97, 65eV	NLS (base sphingoïde)

Table 5. Acquisitions en PIS et NLS pour l'identification des classes et sous-classes de lipides : adaptée (Han, Yang, & Gross, 2012)

Différents balayages permettent l'identification des classes (tête polaire) ou des espèces (acides gras). Les cations Na e Li favorisent les clivages

L'utilisation de spectromètres de masse à haute résolution et à temps d'acquisition courts tels les Q-TOF a permis de faire évoluer les analyses «shotgun» en des approches réalisant des balayages multiples d'ions précurseurs (MPIS pour Multiple Precursor Ion Scattering) (Ekroos et al; 2002, Schwudke et al; 2006) (Ejsing et al; 2006, Stahlman et al; 2009). Ce mode MPIS introduit par l'équipe de Shevchenko permet alors la détection en mode PIS des têtes polaires et des chaînes d'acide gras simultanément, assurant de ce fait la séparation des ions isobares et isomères. Pour exemple, si l'on considère l'identification de la PE36 :2, détectée en mode d'ionisation négative à la masse de l'ion précurseur m/z 742.6, deux isomères sont possibles (sans considérer la régiosélectivité et la position des doubles liaisons). Pour la PE dissymétrique, ce même précurseur (à savoir m/z 742.6) serait détecté dans 2 spectres PIS correspondant au FA18 :0 et FA18 :2. Pour la PE symétrique, le précurseur n'est détecté que pour le spectre correspondant au FA18 :1. Un spectre supplémentaire de confirmation de la tête polaire (PIS m/z 196) permet alors de confirmer l'identification de la PE 36 :2.

La principale limitation de l'approche MPIS restait la non acquisition simultanée de spectres NLS. Pour y pallier le groupe a pensé une nouvelle stratégie : réaliser en 2 secondes une analyse MS dite «survey scan» servant à identifier de façon prospective l'ion précurseur d'intérêt suivie de 2 acquisitions MS/MS consécutives PIS et NLS réalisées à partir de la liste d'inclusion.

Dans une démarche similaire et en se basant sur l'étroite relation existant entre les techniques de spectrométrie de masse en tandem (c'est-à-dire, l'analyse des ions produits, balayage de perte de neutres, et balayage des ions précurseurs), Han et Gross ont récemment développé l'approche multi-dimensionnelle par spectrométrie de masse pour l'analyse des lipides dite MDSM-SL pour Multi-Dimensional Mass Spectrometry-based Shotgun Lipidomics (Han, Yang, & Gross, 2012). Cette approche tend à maximiser l'exploitation des données shotgun en assurant de multiples PIS et NLS ciblant les différentes classes de lipides tel que présenté en Figure 25.

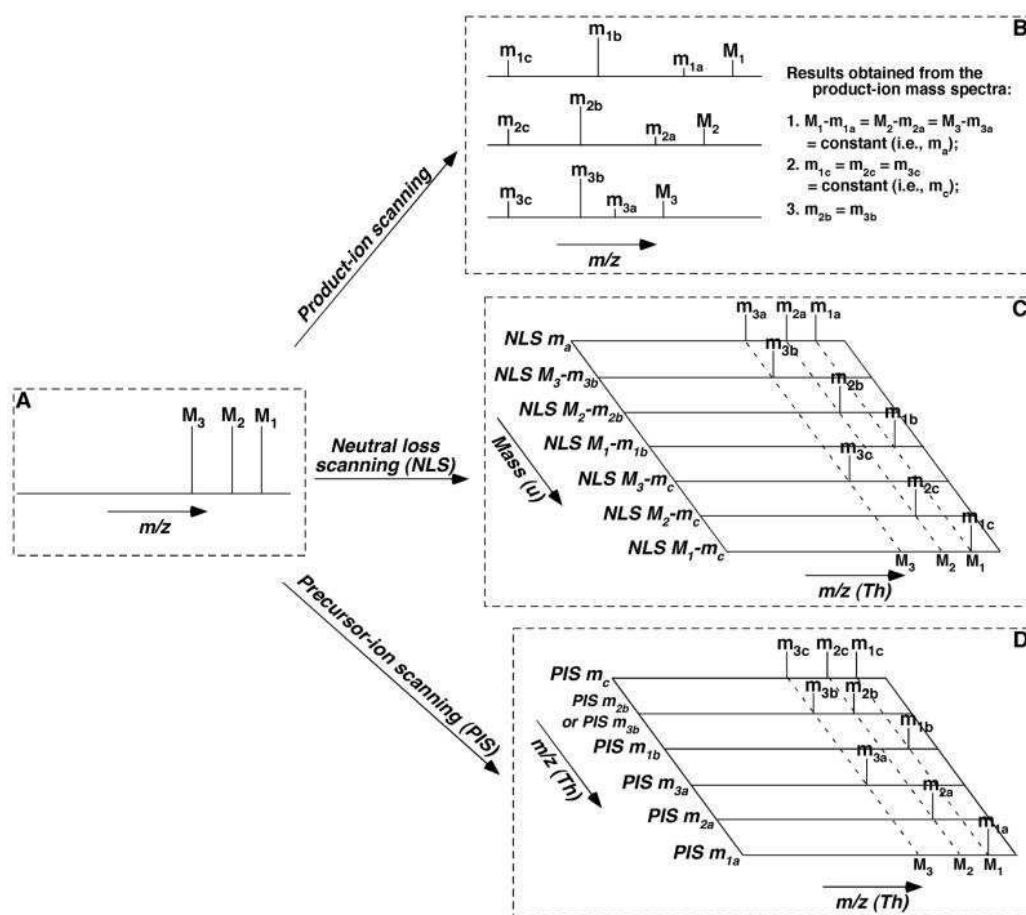


Figure 25. Principe des approches MDMS (Han, Yang, & Gross, 2012)

Dans cet exemple, chacune des trois molécules présente un rapport m/z différent mais avec un même schéma de fragmentation du fait qu'elles appartiennent à une même classe. Sont donc présentés 3 types d'ions produits (figure 3B). Tout d'abord, ces molécules produisent des ions résultant de la perte d'un fragment neutre commun d'une masse de m_a. Ceci donne lieu aux ions fragments m_{1a}, m_{2a}, m_{3a} depuis les molécules M₁, M₂, et M₃, respectivement, où m_a = M₁ - m_{1a} = M₂ - m_{2a} = M₃ - m_{3a}. Ensuite, ces ions moléculaires donnent un autre ion fragment commun m_c (c'est-à-dire m_{1c} = m_{2c} = m_{3c} = m_c). Ces fragments communs sont dus à la tête polaire commune. Enfin, chaque molécule donne naissance à un ion fragment spécifique du à une chaîne d'acide gras (les ions produits m_{1b}, m_{2b}, m_{3b} résultant des molécules M₁, M₂, et M₃ respectivement). L'ensemble des ions produits à partir d'une molécule peut être détecté dans le mode de balayage de perte de neutre (NLS) à des différences de masse dans un ordre croissant : m_a, M₃ - m_{3b}, M₂ - m_{2b}, M₁ - m_{1b}, M₃ - m_c, M₂ - m_c, et M₁ - m_c (Fig. 3C). De même, que tous les ions produits peuvent être détectés

dans le mode PIS par balayage à m/z dans un ordre ascendant de m_c/z , m_{2_b}/z , m_{3_b}/z , m_{1_b}/z , m_{3_a}/z , m_{2_a}/z , et m_{1_a}/z (Fig. 3D). Ainsi, si les séries de spectres de masse NLS et PIS sont mis en correspondance, des spectres de masse d'ions produits de la molécule correspondante peuvent être déduits.

Ces différents modes d'acquisition sont très utilisés en analyse lipidomique SL de matrices biologiques complexes. Ceci est notamment dû au développement concomitant de logiciels de reconstruction de spectres et d'identification dédiés à ces approches tel que présenté en PARTIE A.CHAPITRE III.3.2. Ces outils bioinformatiques seront présentés en PARTIE A.CHAPITRE III.3. Néanmoins, l'identification parfois délicate des molécules isobares et la difficulté de quantifier des espèces mineures liée à la présence d'effet de suppression d'ionisation sont à l'origine des approches LC-MS en analyse lipidomique.

2.2.2. Approches lipidomiques par LC-HRMS et LC-MS/MS

Les analyses LC-MS assurent la séparation des ions isobares parfois même isomères ainsi que la quantification de molécules présentes à l'état de traces dans les milieux biologiques. Cette dernière application reste limitée de façon générale car elle nécessite l'élaboration d'une courbe d'étalonnage par espèce du fait que le rendement d'ionisation diffère en fonction du temps de rétention. Néanmoins, des études de quantification à large spectre en LC-MS ont déjà été décrites notamment avec les travaux de Hermansson et coll. (Hermansson et al; 2005) qui ont assuré la quantification absolue de 100 espèces lipidiques après séparation sur une colonne diOL en mode d'élution isocratique.

Les approches ont d'abord été développées telles que décrites pour le SL, c'est-à-dire, le couplage d'une méthode chromatographique à des spectromètres en tandem de type TQ assurant des analyses MS/MS de classes de lipides exemple profilage des sphingolipides (Merrill, Jr. et al; 2005, Sullards et al; 2003). Grâce aux appareils à haut pouvoir résolutif, des analyses globales de lipides se sont développées. Ces dernières permettent l'acquisition d'un spectre de masse sur une large gamme de masses détectant de ce fait simultanément différentes classes de molécules avec la possibilité de reconstituer le chromatogramme de l'ion d'intérêt. Ces techniques assurent le profilage à haute résolution des différentes espèces lipidiques en présence. Des couplages NP-MS (Lesnefsky et al; 2000), RP-MS (Bird et al; 2011, Hu et al; 2008) ou encore des séparation en deux dimensions combinant NP et RP ont également été décrites (Sommer et al; 2006) notamment en ligne via une interface d'évaporation de la phase mobile (Nie et al; 2010). L'utilisation de colonnes telles que l'HILIC (Ikeda & Taguchi, 2010) (Bure et al; 2013) commence également à se populariser.

Le recours à des expériences MS/MS est fortement recommandé afin de confirmer les annotations et identifications réalisées grâce aux mesures de masse de temps de rétention chromatographiques. Cependant, aujourd'hui encore, peu d'études menées en LC-MS/MS sur des

instruments à haute résolution ont été rapportées. L'élucidation structurale des ions les plus abondants en analyse de données indépendantes (DIA) ou des ions anticipés depuis une liste d'inclusion en analyse données dépendante (DDA) peuvent être envisagées. Cette démarche nécessite d'avoir des cycles d'acquisition courts et dans ce sens, les spectromètres de masse de type Q-TOF sont les plus performants.

Notons que d'autres méthodes séparatives couplées à la MS sont également utilisées en analyse lipidomique notamment la chromatographie en phase supercritique (SFC) qui permet une meilleure résolution du fait de la faible viscosité du gaz (Bamba et al; 2012) et la mobilité ionique qui répond au besoin de résolution notamment des isomères mais ne réduit pas l'effet matrice, la séparation ayant lieu post-désorption/ionisation (Kliman, May, & McLean, 2011).

2.2.3. Analyse structurale des lipides insaturés : Détermination des doubles liaisons

Les acides gras estérifiés ou libres présents chez les mammifères présentent une importante diversité structurale notamment la position de la double liaison qui reste souvent inconnue même après des expériences de MS/MS.

La méthode la plus anciennement utilisée pour leur analyse structurale est le couplage GC/EI-MS mais comme discuté précédemment, les hautes énergies et températures sollicitées engendrent la migration des doubles liaisons en plus de n'être applicable qu'aux molécules volatiles (Mossoba et al; 1994). Actuellement, la méthode la plus couramment utilisée est l'*ozonolyse*. L'ozone agit sur les carbones sp^2 induisant la formation d'un pont peroxyde. Après des expériences MS/MS, un clivage homolytique suivi d'un réarrangement permet de détecter un aldéhyde et un acide carboxylique (Harrison, Clay, & Murphy, 1999). Ce procédé peut se faire hors ligne lors de la préparation de l'échantillon. Or, cette stratégie a été récemment mise en œuvre en ligne soit (i) dans la source avec l'utilisation de l'oxygène comme gaz de nébulisation («OzESI-MS» pour ozone electrospray ionization mass spectrometry) (Thomas, Mitchell, & Blanksby, 2006) (ii) mais aussi en mode OzID-MS (pour ozone induced dissociation) (Thomas et al; 2008) où un mélange ozone/oxygène est introduit dans le piège. D'autres solutions ont été décrites :

(i) l'*APCI* avec l'utilisation de l'acétonitrile comme agent réactant. Cette méthode nommée APCACI pour Atmospheric pressure Covalent adduct Chemical Ionization couplé à de l'HPLC a permis de localiser les doubles liaisons et la position des AG sur le glycerol des TG (Xu & Brenna, 2007).

(ii) l'*hydroxylation vicinale* peut se faire à l'aide de tetroxide d'osmium (OsO_4) (Moe & Jensen, 2004)

(iii) l'ionisation FAB (Ji et al; 2007a, Ji et al; 2007b), les processus de charge piégée (en anglais « charge remote dissociation ») réalisés sur des analyseurs de type TOF/TOF, la formation de complexes avec le $Cu(II)$ (Afonso et al; 2005) ou encore avec l'acide 4-iodo-benzoïque (Pham et al; 2012) peuvent être utilisées pour produire des *clivages allyliques*.

2.2.4. Quantification absolue des lipides par ESI/MS

Contrairement aux analyses métabolomiques, la quantification absolue des lipides est un domaine émergent et complexe. En effet, il existe différentes espèces au sein de différentes classes de lipides ce qui rend les rendements d'ionisation et fragmentations différents. De plus, les rendements d'ionisation dépendent de l'instrument (Koivusalo et al; 2001), et des paramètres utilisés (DeLong et al; 2001), et enfin, il s'avère impossible de synthétiser chimiquement toutes les espèces à quantifier afin d'obtenir des calibrants et des étalons internes. D'autre part, contrairement à la quantification absolue en analyse métabolomique, il est indispensable de corriger les aires obtenues des isotopologues M+1 et M+2. En effet, les isotopologues M+2 sont isobares aux ions monoisotopiques des espèces saturées correspondantes pouvant de ce fait surestimer la concentration de ces derniers (Ceci sera illustré dans un contexte d'annotation en chapitre de résultats PARTIE A.CHAPITRE II.3.2.5 voir Figure 64).

Dans ce contexte, des équipes ont réalisé des dosages en ajouts dosés (Liebisch et al; 2002, Whitehead et al; 2007), ou encore à l'aide de standards homologues non présents naturellement, ou bien marqués aux isotopes stables (Ejsing et al; 2009), ou encore modifiés chimiquement (Berry & Murphy, 2005). Dans ce dernier cas, l'utilisation de la N-Methyl piperazine qui réagit sur des espèces de type PS et PE induit des formes plus lourdes de 144Da donnant des fragments caractéristiques à m/z 117 en CID.

Enfin, des stratégies reposant sur le marquage de milieux de culture avec des précurseurs de lipides marqués aux isotopes stables peuvent être un moyen intéressant d'obtenir une bonne représentativité des espèces à doser. Citons par exemple le dosage des PC à l'aide de précurseurs marqués de type D₄-ethanolamine et D₉-choline (DeLong et al; 1999), ou celui des PI et LPI grâce au ³H-inositol (Wenk et al; 2003).

2.2.5. Imagerie des lipides par spectrométrie de masse

Avoir accès à la cartographie moléculaire des lipides dans un tissu est un enjeu majeur pour la compréhension des processus biologiques locaux. Alors que les techniques *in vivo* -bien que non invasives- manquent de sensibilité, la spectrométrie de masse représente une alternative intéressante. Ainsi, l'imagerie de lipide par spectrométrie de masse a connu un essor considérable. Différentes techniques d'ionisation permettent de réaliser des expériences d'imagerie par spectrométrie de masse:

(i) La **spectrométrie de masse des ions secondaires** «SIMS» (de l'anglais «Secondary Ion Mass Spectrometry») assure l'analyse de lipides depuis des surfaces solides ou films (Handa et al; 1983). Elle consiste en un bombardement à haute énergie de cations (Ar⁺ ou Xe⁺) suivi de l'analyse des ions

secondaires produits. Cette méthode offre une résolution spatiale de l'ordre du μm , autorisant l'analyse de cellules uniques (Lanekoff, Sjovall, & Ewing, 2011).

(ii) Le **FAB: Fast Atom bombardment Ionization** (Barber et al; 1981) présente le même principe de fonctionnement sous vide. Avec cette fois-ci un bombardement d'atomes (exp Ar^+/Xe^+) sur une matrice liquide. Mais ces méthodes ne permettent pas toujours de remonter à la structure de l'ion parent.

(iii) L'imagerie **MALDI** (Caprioli, Farmer, & Gile, 1997, Jackson et al; 2005, Touboul et al; 2004) présentent l'avantage de donner accès à la masse moléculaire de l'ion parent et de leur distribution mais les dépôts non homogènes, la résolution spatiale reste limitante ainsi que le bruit de fond souvent lié à l'ionisation de la matrice. Des matrices présentant un minimum d'interférences et de faibles lasers sont actuellement utilisés pour optimiser l'utilisation.

(iv) La **desorption electrospray ionization (DESI)** permet de réaliser un échantillonnage *in situ*, à pression atmosphérique (Takats et al; 2004). Cette méthode d'ionisation donne une information spatiale et moléculaire. Elle a déjà été utilisée pour des études descriptives visant à détecter des espèces de glycérophospholipides et sphingolipides dans des échantillons biologiques de cerveau de porc (Manicke et al; 2008), mais aussi pour des applications biologiques notamment en cancérologie (Dill et al; 2009, Dill et al; 2010, Eberlin et al; 2012). Citons par exemple les travaux de Dill et coll qui ont porté sur la distribution mais aussi l'intensité d'espèces PI et PS en cas de carcinome canalaire infiltrant (Dill et al; 2009).

3. Outils bioinformatiques pour l'analyse lipidomique

Alors que l'essentiel des développements en analyse de lipides a été effectué sur un plan technique, peu de développements sur un plan bioinformatique ont été réalisés avant l'année 2006. A l'ère des approches «omiques» la démarche lipidomique a peu à peu évolué. Aujourd'hui, elle est une approche intégrée alliant chimie analytique (anciennement analyse de lipides) et outils bioinformatiques pour la découverte de biomarqueurs lipidiques.

Certains outils bioinformatiques utilisés en analyse lipidomiques peuvent être partagés avec ceux de l'analyse métabolomique tels que les outils de bas niveau PARTIE A.CHAPITRE II.4.1 ainsi que les analyses statistiques PARTIE A.CHAPITRE II.4.2.1. Par ailleurs, bien que les bases de données en métabolomiques répertorient les espèces lipidiques, des bases de données dédiées aux analyses lipidomiques existent et seront présentées. De plus, du fait que certaines approches analytiques aient été spécialement décrites qu'en lipidomique (PARTIE A.CHAPITRE III.2.2), des outils bioinformatiques dédiés notamment ceux assurant l'aide à l'annotation et à l'identification seront présentés dans ce chapitre.

3.1. Banques de données en lipides

Les lipides détectés par des approches MS sont annotés à l'aide de banques de données dédiées (Table 6). La base de données la plus adaptée en lipidomique est celle créée par le consortium «LIPID Metabolites And Pathways Strategy» (**LIPID MAPS**) (Fahy et al; 2009, Fahy et al; 2005). Cette dernière propose une classification et une ontologie des espèces lipidiques. Plusieurs fonctionnalités y sont proposées. La plus utilisée est «lipid database search» (LMSD) qui est un outil permettant l'annotation sur la base d'une masse précise (d'un ion pseudomoléculaire ou ion adduit) ou d'une formule brute, mais permet aussi la recherche de composés de référence (Fahy et al; 2007b). La fonctionnalité LMPD pour «LIPID MAPS Proteome Database» recense quant à elle les protéines et gènes associés aux lipides (Sud et al; 2007) dans une optique de biologie des systèmes.

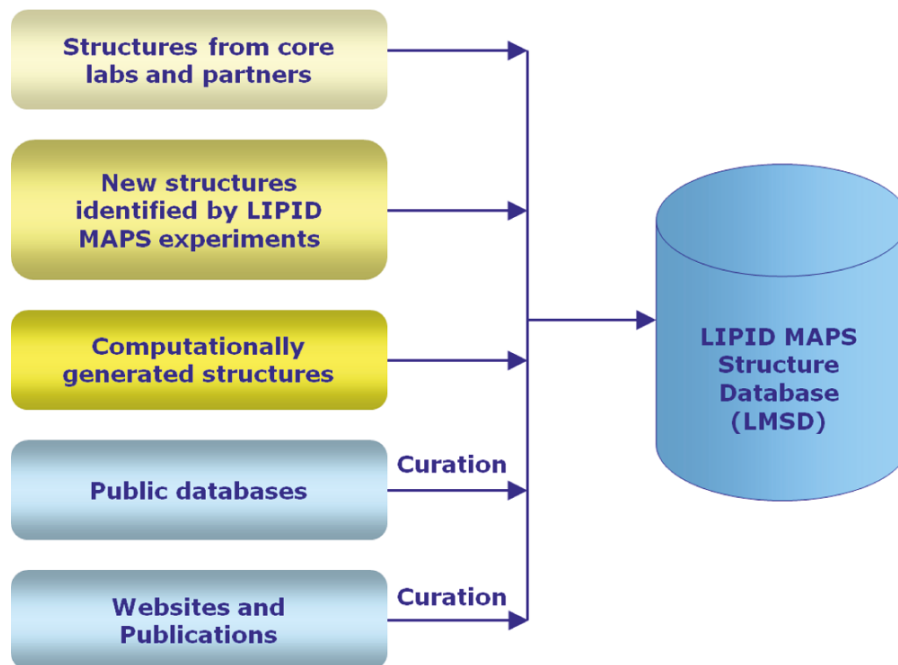


Figure 26. Création et implémentation de la base de données LMSD (Subramaniam et al; 2011)

Structure database (LMSD)

Structure (points to the chemical structure)

Identifiant LIPID MAPS (points to LM ID: LMFA03010002)

Noms (points to Common Name: PGF2α and Systematic Name: 9S,11R-15S-lithydroxy-5Z,13E-prostadienoic acid)

Calculateur de m/z (points to Exact Mass: 354.2406 (neutral) and Formula: C₂₀H₃₄O₅)

Classification (points to Category: Fatty Acids (FA) and Main Class: Eicosanoids (FA03))

Références autres bases de données (cross-références) (points to PubChem Substance ID (SID): 4205953, METABOLOMICS ID: -, KEGG ID: C00639, HMDB ID: HMDB01139, CHEBI ID: 15553)

Identifiant InChIKey (points to InChIKey: PKCPLTODNUGFLYNNPMYKQSA-N)

Spectres MS/MS (points to MS Standard: View lipid standard)

Propriétés physiochimiques (points to Calculated physicochemical properties table)

Autres formats de structure (points to Download file: MDLMOL)

Calculated physicochemical properties					
Heavy Atoms	25	Rings	1	Aromatic Rings	0
van der Waals	378.23	Topological Polar Surface Area	97.99	Rotatable Bonds	12
Molecular Volume		Hydrogen Bond Donors	4	Hydrogen Bond Acceptors	5
logP	3.27	Molar Refractivity	107.08		

Figure 27. Vue détaillée d'une structure implémentée dans la base de données LMSD

Une autre banque de données intéressante est la «**GOLM** metabolome database» (Kopka et al; 2005). Dédicée aux données de GC-MS, elle inclut des spectres de masse, des temps de rétention et une liste de marqueurs chimiques « tags ». Bien que destinée aux analyses métabolomiques, les annotations de lipides répertoriées sont pertinentes et elles sont reliées à d'autres banques plus générales telles que **KEGG** (<http://www.genome.jp/kegg/>) (Kanehisa & Goto, 2000) ou **ChemSpider** (www.chemspider.com). Une autre base de données dédiées aux lipides est **LipidBank** (<http://www.lipidbank.jp/>) (Watanabe, Yasugi, & Oshima, 2000, Yasugi & Watanabe, 2002, Yetukuri et al; 2008), développée par des équipes japonaises (Japanese Conference on the biochemistry of lipids). Celle-ci fournit des données spectrales (NMR, MS, UV, IR) ainsi que des informations sur la synthèse chimique et activité biologique. D'autres bases de données ont été développées et sont consultables en ligne. Citons par exemple, «**Lipid Library**» pour les acides gras, «**Cyberlipid Center**» (<http://www.cyberlipid.org/>) dont le champ d'application est plus large, «**LIPIDAT**» (Caffrey & Hogan, 1992) pour les phospholipides, ou encore «**LIPIDAG**». Cependant, pour la plupart, les mises à jour ne sont pas faites et leur contenu est déjà répertorié dans LIPID MAPS.

Il existe peu de bases de données spectrales adaptées aux besoins des analyses lipidomiques. La plupart permettent l'exploitation des données MS et pour certaines la visualisation de spectres MS/MS, mais cela ne répond pas au besoin d'identification à haut débit *via* des expériences MS/MS.

Compte tenu de l'impossibilité de synthétiser tous les composés de référence, il est nécessaire d'avoir recours à des méthodes *in silico* analogues à celles utilisées en protéomique pour la construction d'une base de données MS/MS. Dans cette optique, une base de données spectrales appelée «**LipidBlast**» a été développée par le laboratoire de Fiehn (Kind et al; 2013). Ces auteurs ont généré *in silico* des molécules à l'aide d'algorithmes de chimie combinatoire tels «ChemAxonReactor» et «SmiLib» pour atteindre 119200 composés couvrant 26 classes de lipides glycerophospholipides et glycerolipides. Après interprétation de 500 spectres MS/MS par différentes plateformes en plus de spectres publiés, des règles de fragmentation ont été établies et appliquées aux molécules présentes dans la base de données. Les données spectrales générées ont été validées en interne par une recherche inversée, puis une recherche sur le programme NIST MS search, puis à l'aide de 134 expériences de MS/MS choisies depuis la littérature et finalement de 27 composés synthétiques de référence. Grâce à cette base de données, 264 lipides ont été identifiés dans un échantillon plasmatique de référence commercialisé par le NIST à savoir le SRM1950, à l'aide d'analyses de type «shotgun».

Comme nous venons de le voir, la plupart des bases de données de lipides sont des bases de données dédiées à faciliter l'identification de lipides dans les milieux biologiques. L'interprétation biologique des résultats nécessite cependant de disposer d'autres types d'outils, et notamment des outils de visualisation. Or, il n'existe qu'un nombre limité d'outils informatiques permettant de visualiser les réseaux métaboliques de lipides. Le logiciel «**SphinGOMAP**» (<http://www.sphingomap.org/>) (Fahy et al; 2007a) décrit par exemple des voies métaboliques spécifiques à la biosynthèse des sphingolipides. La base de données «**LipidDB**» (Yetukuri et al; 2007) est l'outil le plus abouti, qui combine données structurales et données biochimiques. Il assure en effet l'identification des lipides à l'aide de données MS (et/ou MS/MS) et temps de rétention chromatographiques, puis les retranscrit en format SMILES (Simplified Molecular Line Entry Specification) avant de les inclure dans les voies métaboliques décrites dans KEGG et de rendre un résultat graphique.

Resource	Pays	Commentaire	URL
LIPID MAPS	Etats Unis	Classification des lipides	http://www.lipidmaps.org/
LipidBank	Japon	Couvre le règne animal et végétal	http://lipidbank.jp/
LIPIDAT	Etats Unis	Données thermodynamiques des phospholipides	http://www.lipidat.tcd.ie/
LMSD	Etats Unis	Structures et annotations de lipides	http://www.lmsd.tcd.ie
LMPD	Etats Unis	Proteines et gènes associés aux lipides	
Cyberlipid	France	Informations diverses sur les lipides (cours) et sur la communauté scientifique oeuvrant dans le domaine (calendrier des conférences)	http://www.cyberlipid.org
SphinGOMAP	Etats Unis	Voies métaboliques des sphingolipides	
Lipid Library	Etats Unis	Information sur la chimie, la biologie et l'analyse des lipides	http://lipidlibrary.aocs.org/
KEGG	Japon	Voies métaboliques notamment AG, stérols et phospholipides	
Genome OnLine Database GOLD	Australie	Genome et données associées	http://gold.uni-graz.at/index.html
LIPID DAG			http://www.lipidag.tcd.ie/

Table 6. Ressources et bases de données biochimiques en ligne pour la classification des lipides

3.2. Logiciels pour l'annotation de données lipidomiques

Le manque de composés de référence de nature lipidique fait de l'identification en analyse lipidomique une approche souvent *de novo*. Nous pouvons classer les outils actuellement disponibles dédiés à l'identification des lipides en trois types (Hartler et al; 2013):

- (i) les outils utilisant des bibliothèques de spectres de référence avec une recherche des similarités,
- (ii) les outils pour l'analyse shotgun,
- (iii) les outils pour l'analyse de données obtenues en LC-MS ou LC-MS/MS.

3.2.1. Outils de similarité spectrale

Ces logiciels font correspondre des spectres expérimentaux de molécules inconnues à ceux décrits dans une base de données spectrales. Un score est alors rendu selon le degré de similarité des spectres (en termes de présence ou d'absence des ions fragments et de leurs intensités relatives). Ces approches ressemblent à celles décrites en protéomique, à la différence près que les bases de données de protéomique sont construites *in silico* à partir de séquences peptidiques, alors que la plupart des bases de données de lipidomique contiennent des données expérimentales, exception faite de LipidBlast qui est disponible depuis juin 2013 (Kind et al; 2013).

Des constructeurs commencent également à se positionner sur ce créneau, comme par exemple la société Thermo Fischer Scientifics avec le logiciel «LipidSearch™». D'autres logiciels et banques de données sont disponibles en libre-service. Citons, XCMS² (Benton et al; 2008), qui grâce à l'algorithme «shared peak account», calcule une matrice de similitude entre spectres observés et spectres MS/MS de composés de référence obtenus avec un instrument de type Q-TOF MS/MS et référencés dans la base de données «Metlin» (Smith et al; 2005). La limitation majeure de ce type d'outil est la restriction à un type de spectromètre de masse. La base de données «HMDB» contient

des spectres acquis sur différents analyseurs (FT-ICR, QTOF, TQ) et même des spectres NMR) (Wishart et al; 2007), mais elle est restreinte aux métabolome et lipidome humains. Bien que moins complète, «MassBank» (Horai et al; 2010), quant à elle, fonctionne selon un mode collaboratif et permet ainsi d'ajouter des spectres de masse et de MS/MS, mais elle ne permet pas l'annotation de lipides inconnus.

Bien que d'utilisation aisée, tous ces logiciels présentent les limitations d'être restreints à un type d'instrument et d'être dépendants de paramètres expérimentaux précis tels les adduits considérés ou des énergies de fragmentation déposées, ce qui peut donner des valeurs de score erronées. Par exemple, Hsu et Turk ont démontré que des fragmentations de phospholipides obtenues sur un TQ ou un piège étaient différentes (Hsu & Turk, 2009). De plus, bien que permettant d'avoir une identification formelle, ces logiciels ne permettent l'annotation que de lipides répertoriés dans les bases de données.

3.2.2. Outils pour l'analyse shotgun

Des logiciels tels que «Lipid Qualitative/Quantitative Analysis» (LipidQA) traitent des données PIS (Song et al; 2007, Song, Ladenson, & Turk, 2009). Lipid QA a été développé pour l'identification de sphingolipides et phospholipides à partir de données acquises en DDA sur différents types d'analyseurs (LIT, Q-TOF, TQ). Bien que développé par des équipes très impliquées dans les modèles de fragmentation (Hsu, Bohrer, & Turk, 1998, Hsu & Turk, 2008a, Hsu & Turk, 2008b), la restriction de ce logiciel aux données PIS limite son intérêt pour l'identification d'espèces lipidiques dans les milieux biologiques (tel que présenté précédemment en PARTIE A.CHAPITRE III.2.2.1). Avec l'évolution des approches PIS en des approches MPIS, des logiciels plus élaborés ont été développés par l'équipe de Shevchenko en collaboration avec la société AB Sciex. Le logiciel «**LipidProfiler**» (devenu «**LipidView**») (Ejsing et al; 2006) gère par exemple des données MPIS parallèles pour l'identification et la quantification de lipides sur des gammes de concentration allant de 10nM à 100µM. Le logiciel «**LipidInspector**» (Schwudke et al; 2006) développé par la société Scionics Computer Innovations assure quant à lui le traitement de données lipidomiques contenant des spectres NLS acquises selon les modalités « survey scan + PIS +NLS » décrites en 72PARTIE A.CHAPITRE III.2.2.1. L'identification des espèces lipidiques en mode d'acquisition multidimensionnel MDMS-SL a été rendue possible grâce au logiciel «**AMDMS-SL**» qui assure la reconstruction post acquisition des spectres (Han & Gross, 2005, Han, Yang, & Gross, 2012, Yang et al; 2009). Ce dernier propose une bibliothèque extensible de lipides théoriques, constituée par combinaison linéaire de chaînes d'acide gras (de longueur de chaîne et nombre de doubles liaisons différents) et de «têtes polaires» (phosphocholine, galactose 3-sulfate...) reliés aux différents «squelettes» (soit glycérol, base sphingoïde et cholestérol)

et représentant de ce fait chacune des 4 familles lipidiques : stérol, sphingolipides, glycérolipides et glycérophospholipides.

Enfin, d'autres logiciels de traitement de données de type shotgun spécifiques à certaines classes lipidiques existent. Le logiciel «**MSPECTRA**» (Kurvinen et al; 2001) est par exemple dédié à l'analyse des TG par CID. Il peut en outre fournir des informations sur la régiosélectivité. De façon générale, ces logiciels ont tous des limites qui ne leur sont pas propres mais qui sont inhérentes à l'approche shotgun en elle-même.

3.2.3. Outils pour l'analyse de données obtenues en LC-MS ou LC-MS/MS

Peu de logiciels ont été développés pour l'analyse LC-MS des lipides. On citera «**Lipid Mass Spectrum Analysis**» (LIMSA) (Haimi et al; 2006) dont l'intérêt majeur est son module SECD (spectrum extraction from Chromatographic Data) qui assure la détection et l'alignement des ions et offre la possibilité d'incrémenter la base de données de référence contenant déjà 3000 composés. D'autre part, le logiciel «**LipID**» (Hubner, Crone, & Lindner, 2009) a été développé pour traiter les données HPLC-FTICR, ou encore le logiciel «**FAAT**» (Leavell & Leary, 2006) a été conçu pour l'analyse des acides gras libres. Ces logiciels sont peu utilisés car ils ne permettent d'annotation et d'identification efficaces. L'annotation n'est en effet basée que sur la masse précise étant donné que l'information sur les temps de rétention est difficilement transposable d'un système LC/MS à un autre.

Le premier logiciel décrit dans la littérature pour l'analyse de données LC-MS/MS est **Profiler-Merger-Viewer**. Celui-ci gère les données acquises sur des analyseurs à haute résolution selon le protocole suivant : une première acquisition d'un spectre de masse sur une large gamme de masses, suivie de 2 acquisitions en mode Single Ion Monitoring des 2 ions les plus abondants permettant ainsi d'effectuer des expériences type MS^2 et MS^n au cours d'une même analyse (DDA). L'identification se fait alors grâce à la valeur de m/z observée et aux éventuelles pertes de neutres caractéristiques. Ce logiciel ne traite que les données HR et nécessite la vérification manuelle des pics avant leur sélection pour une quantification.

Récemment, **LipidMiner** (Meng et al; 2014) a démontré sa capacité à identifier et quantifier en LC-MS/MS des lipides dans des milieux biologiques (Gao et al; 2012). Le logiciel opère en 3 étapes consécutives. D'abord le module «Ion Detection» génère une liste d'ions observés en MS et MS/MS, calcule les différences de masse entre les ions parents et fragments pour faire correspondre les pertes observées aux pertes de neutres caractéristiques prédéfinies et attribue des classes. Après alignement de l'ensemble des échantillons, les variables déjà assignées à des classes, ainsi que les variables non fragmentées ou dont la MS/MS n'a pas été jugée caractéristique, sont annotées depuis la base de données LIPID MAPS donnant ainsi accès aux espèces. Ce logiciel semble le plus complet car assure toutes les étapes du processus d'identification depuis l'import des données brutes à

l'export d'une liste d'ions identifiés avec différents statuts (Sumner et al; 2007) selon que la MS/MS ait été concluante ou pas.

Fonctionnalité\Logiciel	AMDMS-SL	Lipid QA	LIMS A	FAAT	LipID	LipidProfiler LipidView	LipidSearch	LipidXplorer	LipidMiner
MS	✓		✓	✓	✓	✓	✓	✓	✓
MS/MS	✓	✓				✓	✓	✓	✓
LC			✓	✓	✓		✓		✓
Banque de masses	✓		✓	✓	✓	✓	✓		✓
Banque de spectres	✓	✓							
Possibilité d'extension			✓	✓	✓	✓		✓	✓
Correction des isotopes		✓	✓			✓		✓	✓
Multi-plateforme		✓	✓	✓	✓		✓	✓	✓
Alignement des spectres				✓				✓	✓
Grouping des spectres						✓		✓	✓
Référence	Han, 2012	Song, 2009	Haimi, 2006	Leavell, 2006	Hubner, 2009	Ejsing, 2006		Herzog, 2011	Meng, 2014

Table 7. Principales propriétés des logiciels bioinformatiques en analyse lipidomique

Cette multitude de logiciels dédiés à l'analyse lipidomique reflète la diversité des approches analytiques mises à disposition et la difficulté à partager les données. A notre connaissance, aucune étude comparative n'a été menée et de ce fait pas de démarche consensus déjà établie. Cependant, à l'instar de l'analyse métabolomique, l'approche lipidomique peut être aujourd'hui envisagée comme une approche intégrée au service de la biologie des systèmes.

4. Applications cliniques de l'analyse lipidomique

Tel que présenté en introduction, les lipides jouent un rôle majeur dans des processus physiologiques vitaux tels que la signalisation cellulaire ou en tant que réserve énergétique (voir Figure 19) mais aussi dans des processus pathologiques. Citons les maladies métaboliques propres aux lipides (Kolter & Sandhoff, 2006, Lamari et al; 2013). Ces lipidoses représentent des désordres métaboliques entraînant une accumulation morbide de lipides dans les lysosomes pour cause de déficience de leurs enzymes de biosynthèse ou de dégradation (voir Figure 28). Ces maladies métaboliques sont largement étudiées par spectrométrie de masse car ces méthodes permettent de détecter, identifier et quantifier toutes les espèces impliquées dans les voies métaboliques impactées (Delobel et al; 2006, Touboul et al; 2005) (Colsch et al; 2008). Jusqu'à lors, ces études s'apparentaient à des analyses ciblées de lipides et non à des analyses lipidomiques, en ce sens qu'il n'y ait pas eu recours à des outils bioinformatiques ou analyses statistiques.

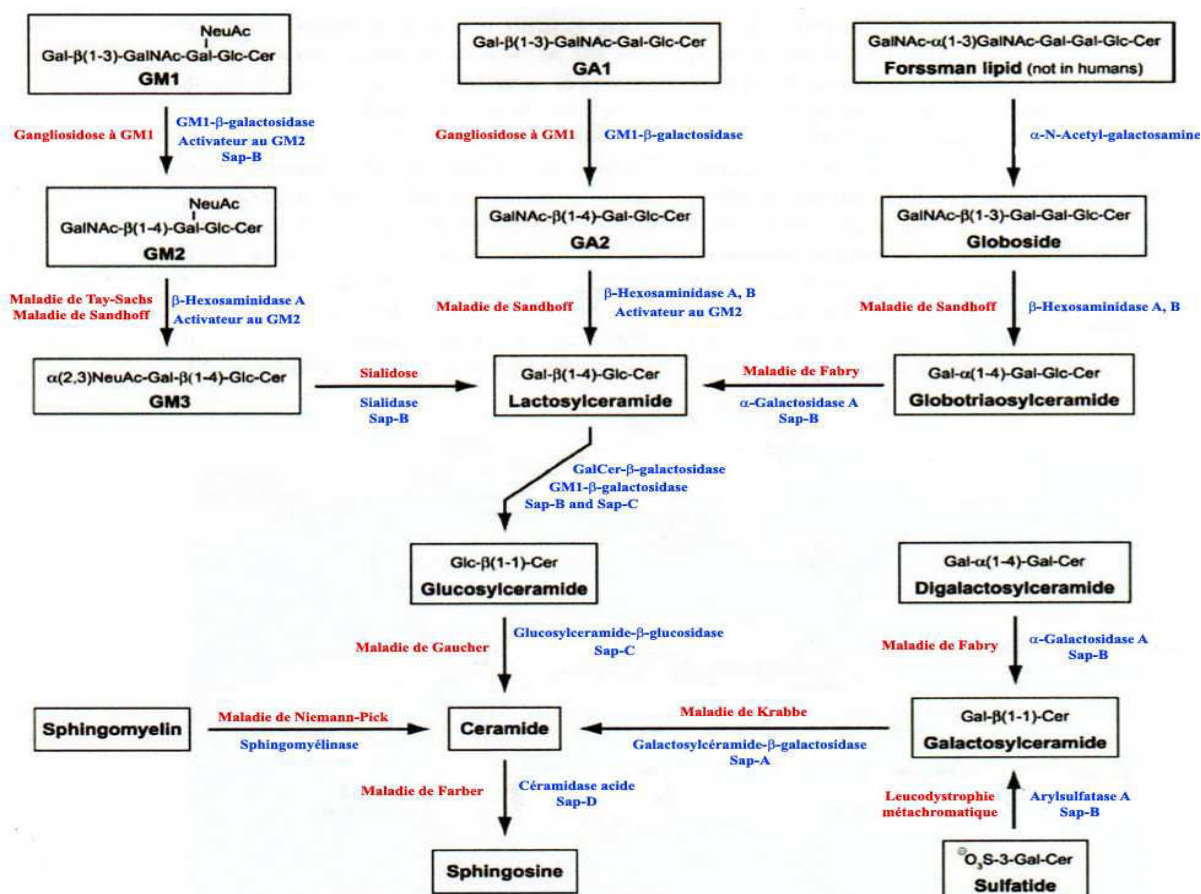


Figure 28. Sphingolipidoses : Enzymes, substrats et produits associés (Kolter & Sandhoff, 2006)

Récemment, analyses haut-débit et outils bioinformatiques aidant, les analyses lipidomiques telles qu'envisagées en analyse métabolomique ont été réalisées dans le cadre de diverses investigations biologiques quant à l'effet du stress oxydant (Yang et al; 2011), l'obésité (Barber et al; 2012), différents cancers ((Gorden et al; 2011,Hilvo et al; 2011,Zhou et al; 2012) et maladies neurodégénératives (Panchal et al; 2014) (voir Table 8).

Pendant, malgré ce vif intérêt porté aux analyses lipidomiques appliquées au diagnostic clinique, il existe peu de recouvrements entre les résultats de ces études et peu de possibilités de remonter aux voies métaboliques impactées. Ainsi, une validation des listes de biomarqueurs proposés tant sur un plan analytique, que biologique (Fahy et al; 2007a,Hadadi et al; 2014,Kanehisa & Goto, 2000) pourrait réduire le nombre de candidats biomarqueurs et permettre de ne se focaliser que sur les plus pertinents.

Domaine clinique	Maladie	Matrice biologique	Lipides investigués	Méthode utilisée	Statistiques	Références
Neurologie	Maladie d'Alzheimer	Cerveau humain	gangliosides	TLC/MALDI-TOF MS	\	Taki et al ; 2012
		Cerveau humain	MG, FA	RPLC-ESI-IT MS	t-Student, corrélation de Spearman	Jung et al ; 2012
		Cerveau humain et cerveau de rat	PC, LPC, pPC	FIA-MS/MS	t-Student	Grimm et al ; 2011
		Plasma humain	SM, Cer	QqQ MS	Test de Wilcoxon , FDR	Han et al ; 2011
Neurologie	Schizophrénie	Cerveau humain	GM3, LPC, LPCe, LPA, LPS, LPI, LPE, LPEp, PC, ePC, PA, PS, PI, PG, PE, pPE, SM,	NPLC-QqQ MS	t-Student	Chan et al ; 2012
		Plaques séniles	Cer à acide gras saturé	RPLC-ESI-QToF	ANOVA	Laprévote et al ; 2014
		Cortex visuel, amygdales, et le cortex cingulaire antérieur	SM, Cer, GM3, PA, PC, LPC, PE, LPE, PI, LPI, PS, DAG, TAG, CE, Chol	NPLC-QqQ MS; NPLC-IT MS, RPLC-Qtrap MS, RPLC-APCI-Qtrap MS	t-Student, FDR	Cheng et al ; 2011
		Plasma humain	plasmalogènes, FA	TLC, GC-FID	Corrélation de Spearman, test de Wilcoxon	Kadurah-Daouk et al ; 2012
Cancérologie	Cancer de la prostate Carcinome hépatique Cancer colorectal	Plasma et cerveau humains	LPC, TG, SM	RPLC-QToF MS	ANOVA	Oresic et al ; 2012
		Plasma humain	ePE, ePC, PE	ESI-QqQ MS	HCA, PCA, RandomForest	Zhou et al ; 2012
		Biopsies de foies humains et de souris	DAG	NPLC-ESI-QqQ MS	ANOVA, t-Student	Gorden et al ; 2011
		serum humain	FA, LPC	DI-ESI-FTICR	OPLS-DA	Li et al ; 2013
Endocrinologie (Diabétologie)	Diabète de type1 Diabète de type2	Plasma, foie et cœur de rats et de souris	FA, Cer, SM, Cérébroside	GC-FID, RPLC-ESI-MS/MS (Qtrap)	ANOVA	Fox et al ; 2011
		Plasma humain	CE, PE, PI, LPC	ESI-QqQ MS	HCA	Zhao et al ; 2013
		Plasma de rat	TG, HDLs, VLDLs	LC-IT-QToF MS	PCA, t-Student	Hu et al ; 2011
		Plasma humain	TG, FA, DAG, PC, LPC, Chol, Cer, CE	RPLC-Qtrap MS	ANOVA	Drew et al ; 2011
		Plasma humain	TAG, DAG, Sphingolipids, LPC	RPLC-Qtrap MS	t-Student et ratios	Barber et al ; 2012

Table 8. Applications des analyses lipidomiques

Enfin, l'étude combinée du métabolome et du lipidome a déjà été envisagée et devrait se développer dans les années à venir (Pflueger et al; 2011). En effet, ces approches partagent les mêmes exigences méthodologiques et ont des besoins bioinformatiques similaires. Seuls les protocoles de préparation d'échantillons diffèrent. De plus, leurs champs d'investigation biologique convergent du fait qu'ils se rejoignent dans des voies métaboliques communes (Figure 29). Citons par exemple le cas de l'acétyl CoA qui réagit avec l'oxaloacétate dans le cycle de Krebs et qui s'engage également dans la bêta oxydation mitochondriale des acides gras libres (considérés comme les métabolites des lipides complexes), ou encore le glycérol 3-phosphate squelette des glycérolipides et glycérophospholipides qui est produit au cours de la glycolyse. Ainsi, il ne fait aucun doute qu'il y aura à l'avenir de plus en plus d'études métabolomiques et lipidomiques menées parallèlement sur les mêmes thématiques biologiques assurant de ce fait une vue d'ensemble la plus générale possible des phénomènes biologiques sous-jacents. C'est d'ailleurs un des objectifs du travail de thèse qui va maintenant être présenté.

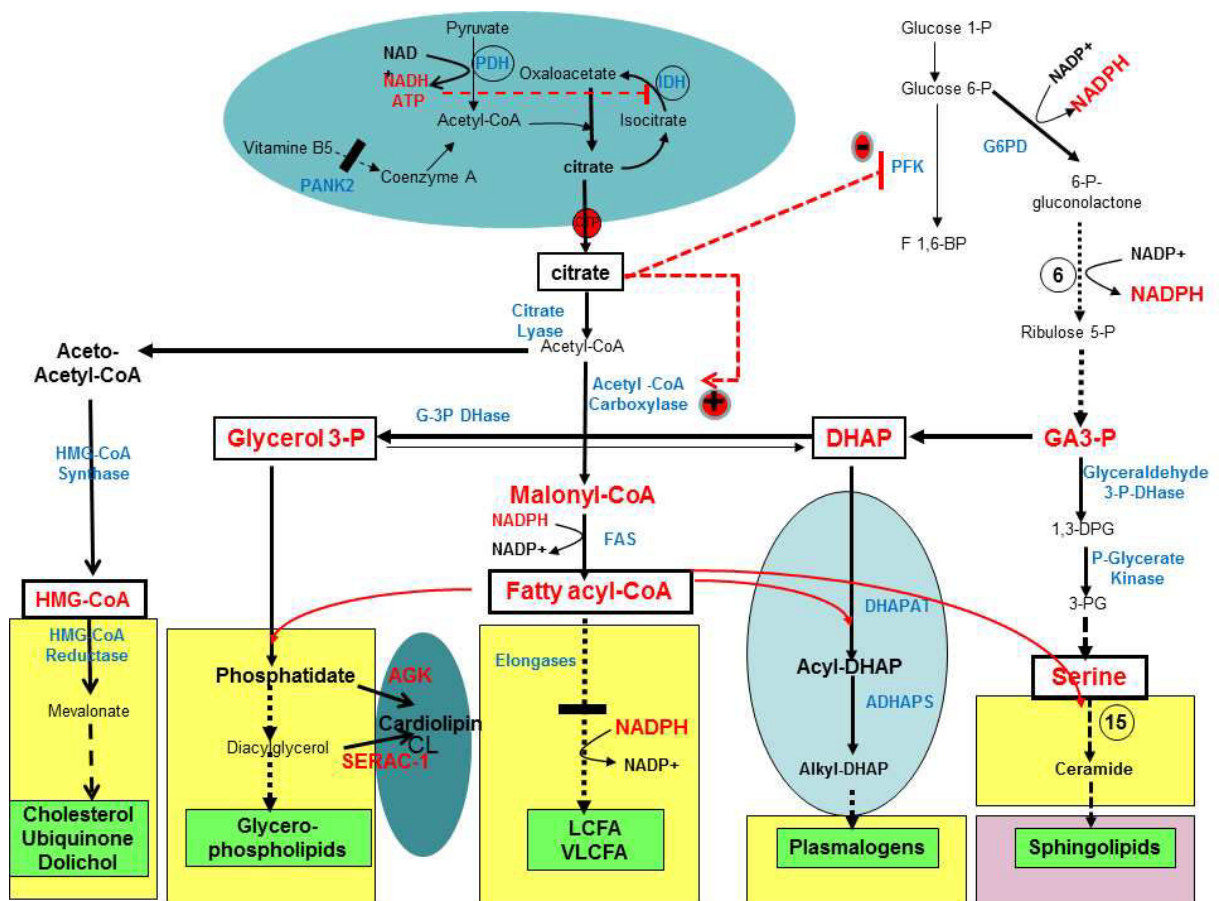


Figure 29. Voies métaboliques et biosynthèse des lipides complexes (adaptée depuis (Lamari et al; 2013))

RESULTATS

A. DEVELOPPEMENT D'APPROCHES GLOBALES DE PROFILAGE METABOLIQUE PAR CHROMATOGRAPHIE LIQUIDE COUPLEE A LA SPECTROMETRIE DE MASSE A HAUTE RESOLUTION

CHAPITRE I. METHODE GLOBALE D'ANALYSE METABOLOMIQUE

1. Introduction

Depuis que la métabolomique a émergé à la fin des années 90, de nombreuses méthodes analytiques ont été utilisées comme plate-forme pour l'acquisition des empreintes : spectroscopie infrarouge (Oliver et al; 1998), la NMR (Roberts, 2000) et la MS (Smedsgaard & Frisvad, 1996) couplée ou pas à la chromatographie en phase gazeuse (GC-MS) (Horning & Horning, 1971), la chromatographie en phase liquide (LC-MS) (Idborg-Bjorkman et al; 2003), ou l'électrophorèse capillaire (CE-MS) (Soga et al; 2002). Outre la NMR, qui était l'une des premières méthodes utilisées en métabolomique (Iles et al; 1984), les améliorations continues de la MS notamment avec (i) l'apparition de sources d'ionisation à pression atmosphérique, (ii) d'analyseurs à haute résolution et (iii) d'outils bioinformatiques d'exploration de données ont promu l'utilisation d'approches globale par MS (Boudah, Paris, & Junot, 2013, Dettmer, Aronov, & Hammock, 2007).

Les plates-formes basées sur l'ionisation à pression atmosphérique (API-MS), impliquant principalement des processus de désorption/ionisation par électrospray (ESI), sont aujourd'hui parmi les plus utilisées en métabolomique du fait de la polyvalence de détection de ces sources (détection des petites à grandes molécules, polaires à apolaires allant des protéines aux lipides). De plus, le développement continu des analyseurs de haute résolution a sensiblement amélioré la gamme dynamique, la vitesse d'acquisition ainsi que la sensibilité de détection. Il a de ce fait été possible de distinguer les métabolites isobares présents dans les liquides biologiques et de détecter les espèces minoritaires élargissant de ce fait l'information contenue dans les profils métabolomiques acquis. En outre, obtenir la composition élémentaire des métabolites détectés à partir de la mesure de masse précise a permis l'utilisation de bases de données chimiques, biochimiques et métabolomiques (HMDB (Wishart et al; 2007), Metlin (Smith et al; 2005), Massbank (Horai et al; 2010), des LipidMaps (Sud et al; 2007), lipid blast (Kind et al; 2013)) et également le développement d'outils d'annotation de données de métabolomique (Brown et al; 2009, Courant et al; 2012, Roux et al; 2012) à des fins d'identification de métabolites.

En dépit de ces améliorations technologiques, l'identification des métabolites depuis des données acquises en LC-HRMS reste difficile. En effet, même avec une précision de masse inférieure à 1 ppm, de nombreuses formules empiriques peuvent correspondre à la masse mesurée (Junot et al; 2010, Kind & Fiehn, 2006); et les requêtes dans des bases de données biochimiques et spectrales renvoient un grand nombre de métabolites putatifs. D'autres contraintes telles que la prise en compte de l'abondance isotopique, le recours à certaines règles chimiques telles que les règles de Lewis, l'échange H/D en solution ou en phase gazeuse et/ou la fragmentation de métabolites sont donc requises afin de sélectionner la composition élémentaire la plus pertinente (Kind & Fiehn, 2007). De plus, l'ESI-MS fournit des informations structurales limitées car elle ne produit principalement que les espèces moléculaires, rendant ainsi nécessaire le recours aux expériences de spectrométrie de masse type MS/MS et/ou MSⁿ. Dans ce contexte, les spectres de masse CID obtenus sur des analyseurs FTMS lors d'expériences séquentielles MSⁿ peuvent être visualisés sous la forme d'arbres de fragmentation, qui facilitent l'identification des métabolites en proposant des voies de filiation des ions produits (Kasper et al; 2012, Rasche et al; 2012, Rojas-Cherto et al; 2012a, Rojas-Cherto et al; 2012b, Schwarzenberg et al; 2013). Les spectromètres de masse de type Q - TOF fournissent également des informations précieuses pour l'identification des métabolites grâce à leur haute vitesse d'acquisition. En effet, ces appareils permettent l'implantation de modes d'acquisition sans sélection préalable d'ions précurseurs afin d'assurer une identification exhaustive du protéome ou métabolome. Nous citerons le mode d'acquisition MS^E qui assure l'acquisition simultanée de données indépendantes de spectres de fragmentation à deux énergies de collision différentes (allant de 5 à 100 eV) ou encore le mode SWATH-MS (de l'anglais « sequential window acquisition of all theoretical fragment-ion spectra ») qui permet l'acquisition parallèle de spectres MS et de spectres MS/MS sur des fenêtres de masse (Gillet et al; 2012, Plumb et al; 2006, Zhu, Chen, & Subramanian, 2014).

Il s'avère également nécessaire d'exploiter les différentes possibilités de couplage de la spectrométrie de masse à haute résolution avec différentes techniques séparatives telles que la CE et la LC. En effet, l'utilisation de la LC-HRMS permet de détecter d'avantage de métabolites, comme démontré par plusieurs études de preuve de concept (Buscher et al; 2009, Ramautar et al; 2011b, Saric et al; 2012). Par exemple, Saric et coll. ont combiné le couplage de la RP-UHPLC, de l'HILIC ainsi que la CE à la spectrométrie de masse à haute résolution pour annoter le métabolome du parasite *Fasciola hepatica*. Ils ont démontré que moins de 10% des 142 métabolites identifiés étaient communs aux cinq méthodes (Saric et al; 2012) et que chaque méthode assurait la détection unique d'un certain nombre de métabolites. Cependant, la disponibilité d'approches multi-instruments est limitée à quelques laboratoires et l'utilisation des seuls couplages LC-HRMS apparaît plus simple à mettre en œuvre et peut permettre à elle seule d'élargir la couverture métabolique.

D'un point de vue historique, les analyses métabolomiques ont principalement été réalisées avec des systèmes chromatographiques basés sur l'utilisation de phases stationnaires de type RP. Cependant, d'autres types de méthodes chromatographiques telles que l'appariement d'ions, la chromatographie liquide d'interactions hydrophiles (HILIC), phase aqueuse normale (ANP) et la fonctionnalisation des supports de silice avec des groupements fluoré, cyano-, diol, ou encore amino- (Cubbon et al; 2010, Gika et al; 2012c, Havlikova et al; 2008, Nichthauser & Stepnowski, 2009) sont disponibles pour la rétention de métabolites hautement polaires.

Plusieurs études publiées ont évalué ces méthodes chromatographiques grâce à des tests réalisés sur des composés de référence et visant à rendre compte exclusivement de la qualité de la séparation chromatographique. Ces études sont d'un intérêt certain pour une meilleure compréhension des mécanismes de rétention mais ne rendent pas compte de la capacité réelle de ces colonnes chromatographiques à assurer une détection optimale des métabolites présents dans des matrices biologiques complexes (Chen et al; 2011, Jiang et al; 2013, Lawton et al; 2008, Psychogios et al; 2011, Roux et al; 2012, Zhang et al; 2012). Ainsi, Roux et coll. ont démontré la complémentarité des systèmes (RP) et des phases stationnaires à base de silices greffées avec des groupements polyfluorophénylpropyles (PFPP) dans l'annotation d'urines humaines. 38 composés détectés dans des urines humaines en utilisant un système RP correspondaient en fait à 83 métabolites distincts une fois analysés sur une colonne PFPP (Roux et al; 2012). D'autre part, Zhang et coll. ont évalué 5 conditions chromatographiques à l'aide de 176 composés de référence ainsi que dans des échantillons d'urine. Ils ont démontré que les colonnes HILIC présentent de bonnes performances chromatographiques et une meilleure détection de métabolites polaires, par rapport à la RP-LC et l'ANP. La combinaison de deux conditions chromatographiques HILIC testées a permis à l'annotation d'environ 970 signaux de métabolites dans un pool d'échantillons urinaires (Zhang et al; 2012). Toutes ces études démontrent l'importance d'effectuer des analyses multi-plateformes pour atteindre une couverture métabolique fiable et optimale.

Certaines approches similaires ont déjà été publiées dans le cadre d'études métabolomiques d'échantillons de plasma ou de sérum humains. Ainsi, Lawton et coll. ont publié en 2008 l'utilisation combinée de GC-MS et de LC-MS pour étudier l'impact des facteurs physiologiques sur les concentrations de métabolites plasmatiques (Lawton et al; 2008). En combinant 4 méthodes d'extraction différentes et en utilisant des instruments à faible résolution (piège à ions linéaire et analyseurs de filtre quadripolaire), ils ont détecté 300 composés (40 et 60% au cours d'analyse par GC- et LC-MS, respectivement). Aucune information quant à la redondance de l'information fournie par les plateformes n'a été communiquée. Dans le même esprit, mais à l'aide d'analyseurs à haute résolution, Jiang et coll. ont détecté 196 métabolites dans le sérum humain en utilisant des couplages GC-TOFMS et UPLC-Q-TOF, dans le cadre d'une étude visant à obtenir une signature

métabolique de l'arthrite (Jiang et al; 2013). Seuls 20 métabolites d'entre eux ont été détectés par les deux plateformes. Enfin, le projet multiplateforme le plus conséquent est celui du consortium canadien HMDB (de l'anglais «Human Metabolome DataBase») et qui a pour but d'entreprendre la caractérisation du sérum humain en combinant cinq méthodes analytiques, en plus d'outils d'exploration bibliographique assistée par ordinateur. Plus de 4200 métabolites sériques putativement ou formellement identifiés ont été détectés, dont plus de 3400 lipides tels que des phospholipides, des stéroïdes ou encore des glycolipides. Là encore, les cinq méthodes ont fourni des informations complémentaires, à savoir 49, 92, 96 et 139 métabolites détectés par RMN, GC-EI-MS, LC-ESI-MS/MS et introduction directe DI/ESI-MS/MS, respectivement (Psychogios et al; 2011).

Dans ce contexte, notre objectif a été d'évaluer trois systèmes LC complémentaires (RP, PFPP et HILIC) couplées à un spectromètre de masse à transformée de Fourier de type Orbitrap, afin de détecter le plus grand nombre de métabolites dans des échantillons biologiques et d'annoter des jeux de données provenant d'analyses métabolomiques. Les travaux de mise au point ont été réalisés avec des échantillons de sérum humain. L'approche a ensuite été adaptée aux analyses de globules rouges et de liquide céphalorachidiens humains. Seule l'application aux échantillons sériques et érythrocytaires sera abordée dans ce manuscrit.

2. Matériel et méthodes

2.1. Réactifs et produits chimiques

Les 657 composés de référence proviennent de chez Sigma-Aldrich (Saint Quentin Fallavier, France), sauf les composés lipidiques qui ont été achetés chez Avanti Polar Lipids (Alabaster, États-Unis d'Amérique) et les stéroïdes qui ont été achetés chez Steraloids (Newport, États-Unis d'Amérique). Les mélanges d'étalons utilisés pour la calibration en masse des spectromètres de masse (Calmix-positif, pour le mode d'ionisation positive, comportant de la caféine, de l'acétate de L- méthionyl-arginyl-phénylanyl-alanine et de l'Ultramark 1621 ; et Calmix-négatif, pour le mode d'ionisation négative, qui contient, en plus du même mélange Calmix-positif, du dodécylsulfate de sodium et du taurocholate de sodium) ont été achetés chez Thermo Fisher Scientific (Courtaboeuf, France). L'acétonitrile (ACN) provient de SDS (Peypin, France), l'acide formique de chez Merck (Briare-le-Canal, France) et le méthanol (MeOH) de chez VWR (Fontenay -sous-Bois, France). L'eau ultra pure déionisée a été filtrée à travers un système Milli-Q (Merck-Millipore, Billerica, MA, États-Unis d'Amérique).

Les solutions mères des composés de référence (1mg/mL) ont été préparées dans des solvants appropriés, des solutions de travail à (5µg/mL par composé de référence) ont été diluées

dans un mélange H₂O/ACN, 95 :5, v/v pour les analyses en RP et PFPP, et dans un mélange H₂O/ACN, 20 :80, v/v pour l'analyse en conditions HILIC.

2.2. Echantillons biologiques et préparation des échantillons

Sérum humain :

Un mélange d'échantillons de sérum a été acheté chez Bioprédic (Rennes, France). Trois aliquotes de 100 µl chacun ont été traités avec 400 µl de MeOH contenant un mélange d'étalons internes (IS) (Répliques techniques) (voir Table 9). Les échantillons obtenus ont ensuite été homogénéisés pendant 20 s à l'aide d'un appareil de type vortex, laissés sous agitation sur de la glace pendant 30 minutes pour permettre la précipitation des protéines, puis centrifugés pendant 20min à 15000g. Les surnageants de chaque réplique ont été divisés en trois aliquotes, et évaporés sous azote. Les extraits secs ont été conservés à -80°C avant leurs analyses LC-HRMS. Deux aliquotes ont été repris dans 105µl d'un mélange 95/5/5 de H₂O/ACN/étalons externes (ExS) (voir Table 9) pour les analyses en RP et PFPP. Le troisième aliquote a été repris dans 105 µl d'un mélange 40/60/5 H₂O/ACN/ExS, pour l'analyse en ZICpHILIC.

	Composé	Formule brute	Masse exacte	EI+	EI-	Concentration
ExS	Alanine ¹³C	¹³ C ₁ ¹² C ₂ H ₇ N O ₂	90.051	91.058	\	20µg/mL
ExS	Metformin	C ₄ H ₁₁ N ₅	129.101	130.109	\	1µg/mL
ExS	Acide ethylmalonique	C ₅ H ₈ O ₄	132.042	\	131.035	3µg/mL
ExS	Aspartate ¹⁵N	C ₄ H ₇ ¹⁵ N ₁ O ₄	134.035	135.042	133.027	20µg/mL
ExS	Glucose ¹³C	¹³ C ₁ ¹² C ₅ H ₁₂ O ₆	181.067	204.056 (+Na)	216.036 (+HCl)	20µg/mL
ExS	2-Aminoanthracene	C ₁₄ H ₁₁ N	193.089	194.096	\	1µg/mL
ExS	Amiloride	C ₆ H ₈ Cl N ₇ O	229.048	230.055	228.040	10µg/mL
ExS	Imipramine	C ₁₉ H ₂₄ N ₂	280.194	281.201	\	0.5µg/mL
ExS	Atropine	C ₁₇ H ₂₃ N O ₃	289.168	290.175	\	1µg/mL
ExS	Prednisone	C ₂₁ H ₂₆ O ₅	358.178	359.185	393.147 (+HCl)	1µg/mL
ExS	Colchicine	C ₂₂ H ₂₅ N O ₆	399.168	400.175	\	0.5µg/mL
ExS	Dihydrostreptomycine	C ₂₁ H ₄₁ N ₇ O ₁₂	583.281	584.288 et 292.648	618.250 (+HCl)	20µg/mL
ExS	Roxithromycine (fragment)	C ₂₉ H ₅₄ O ₁₀ N ₂	590.378	591.385	625.347 (+HCl)	20µg/mL
IS	Dimetridazole	C ₅ H ₇ N ₃ O ₂	141.054	142.061	\	5µg/mL
IS	AMPA 2-amino-3-(3-hydroxy-5-methyl-isoxazol-4-yl)propanoic acid	C ₇ H ₁₀ N ₂ O ₄	186.064	187.071	185.057	5µg/mL
IS	MCPA 2-methyl-4-chlorophenoxyacetic acid	C ₉ H ₉ ClO ₃	200.024	\	199.017	5µg/mL
IS	Dinoseb	C ₁₀ H ₁₂ N ₂ O ₅	240.075	\	239.067	5µg/mL

Table 9. Etalons internes et étalons externes utilisés en analyse métabolomique

Les neuf aliquotes (3 répétitions pour chaque condition LC) ont également été dilués au $\frac{1}{2}$, $\frac{1}{4}$ et $\frac{1}{8}$. Chaque échantillon a été injecté une seule fois par condition LC et un échantillon biologique de contrôle qualité «QC» (un échantillon sérique extrait indépendamment) a été injecté après chaque série de dilution pour évaluer la stabilité du signal.

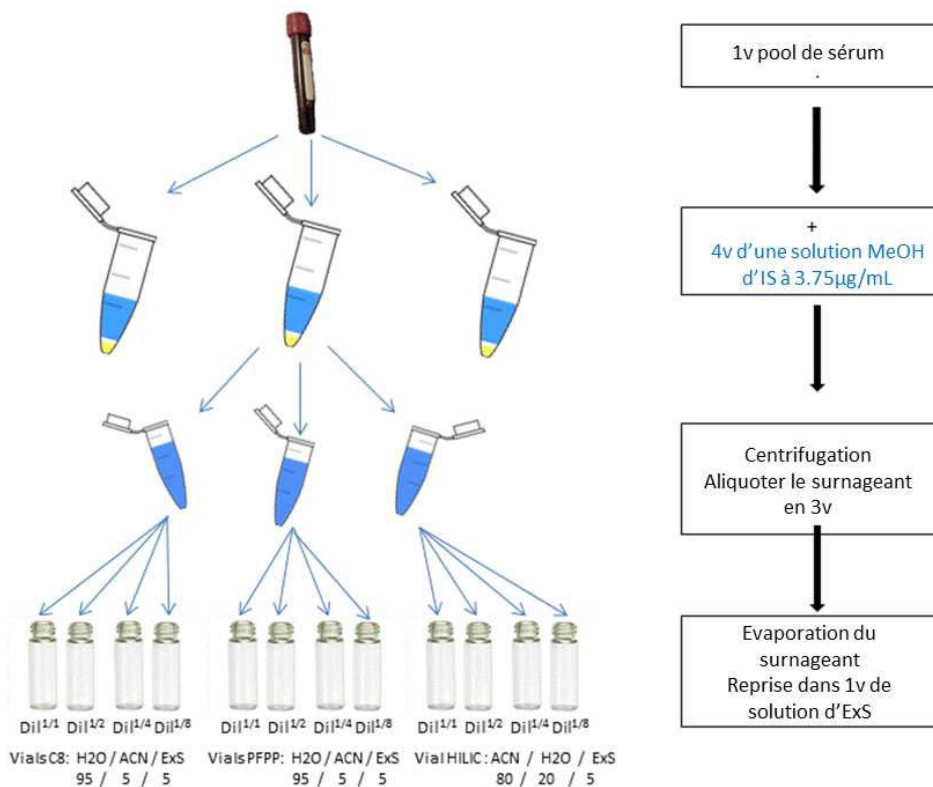


Figure 30. Démarche analytique pour l'annotation du métabolome sérique

Globules rouges humains :

Le sang total de différents patients collectés dans la cadre de différentes études menées au LEMM sont centrifugés à 150g pendant 10 minutes à 20°C. 1mL de culot (contenant globules rouges et blancs) est passé sur une colonne de 5 ml remplie de 50% de α cellulose et à 50% de Sigmacell® cellulose type 50, afin de le déleucocyter. Le culot globulaire obtenu est lavé au sérum physiologique, puis centrifugé à 1430g pendant 3 minutes à 4°C. Le culot globulaire est mélangé avec son double de volume d'eau et le mélange est incubé pendant 3 minutes à 100°C afin de provoquer une lyse cellulaire permettant de libérer les métabolites intra-érythrocytaires. Une nouvelle centrifugation à 3000g pendant 3 minutes à 4°C permet de séparer les débris membranaires du contenu cellulaire. Le surnageant est enfin centrifugé à 4000g pendant 20 minutes à 20°C sur des Amicon Ultra -4, 30 NMWL puis 10 NMWL. Le filtrat ne contient alors plus que des métabolites de masse moléculaire inférieure à 10 kDa. L'extrait total est aliquoté et conservé à -80°C jusqu'à l'analyse. Dès lors, un

mélange des extraits érythrocytaires a été constitué puis dilué au $\frac{1}{2}$, $\frac{1}{4}$ et $\frac{1}{8}$ avant analyse dans les 3 conditions LC-HRMS décrites ci-après.

2.3. Analyses par chromatographie liquide couplée à la spectrométrie de masse

Les analyses LC-HRMS ont été réalisées avec un système de chromatographie liquide de type NEXERA de la société Shimadzu (Champs-sur-Marne, France) couplé à un spectromètre de masse de type Exactive (Orbitrap) de la société Thermo Fisher Scientific (Courtabœuf, France). Les expériences LC-ESI-MS/MS ont été réalisées avec un système chromatographique de type Ultimate (Thermo Fisher Scientific, Courtabœuf, France) couplé à un spectromètre de type Q-Exactive (Thermo Fisher Scientific, Courtabœuf, France). Les spectromètres de masse étaient équipés d'une source Electrospray et les acquisitions ont été réalisées dans les deux modes d'ionisation positive et négative. Les données ont été acquises et visualisées avec le logiciel Xcalibur (version 2.1, Thermo Fisher Scientific, Courtabœuf, France).

2.3.1. Conditions chromatographiques

La chromatographie ultra-haute performance en phase liquide (UHPLC) a été réalisée avec une colonne hypersil GOLD C18 de dimension 2.1mmx150mm, 1.9 μ m (Thermo Fisher Scientific, les Ulis, France) placée dans un four thermostaté à 30°C (conditions dites «RP»). Les analyses par chromatographie liquide à haute performance (HPLC) ont été effectuées avec une colonne de type Discovery HS F5 PFPP de dimension 2.1x150mm, 5 μ m (Sigma-Aldrich, Saint Quentin Fallavier, France), thermostatée à 30°C (conditions dites «PFPP» pour pentafluorophenylpropyl), et avec une colonne de type SeQuant ZICpHILIC de dimension 2.1 x 150mm, 5 μ m, thermostatée à 15 °C (Merck, Darmstadt, Allemagne) (conditions dites «HILIC»). Tous les systèmes chromatographiques ont été équipés d'un préfiltre en ligne (Thermo Fisher Scientifics, Courtabœuf, France).

Les phases mobiles utilisées pour les séparations chromatographiques en conditions RP et PFPP étaient constituées de 100 % d'eau (phase A) et d'ACN (phase B), et contenaient toutes deux 0,1% d'acide formique. Concernant les conditions HILIC, la phase A est composée d'une solution de carbonate d'ammonium (10 mM) ajustée à un pH 10.5 avec de l'hydroxyde d'ammonium, l'ACN pur est utilisé en phase B. Les éluions chromatographiques ont été réalisées en conditions de gradient comme indiqué ci-dessous:

(i) système RP: le débit de phase mobile a été fixé à 500 μ l/min. L'élution débute avec une étape isocratique de 2 minutes à 5% de phase B. S'en suit un accroissement linéaire de 5 à 100 % de la proportion de phase B pendant les 11 minutes suivantes. Ces proportions sont maintenues constantes pendant 12,5 min avant de revenir aux conditions initiales, lesquelles sont maintenues pendant 4,5 min pour ré-équilibration du système.

(ii) système PFPP: le débit a été fixé à 250 µl/min. L'élution débute avec une étape isocratique de 2 min à 5% de la phase B. S'en suit une augmentation linéaire des proportions de phase B de 5 à 100 % en 18min. Ces proportions sont maintenues constantes pendant 4 min avant de revenir à 5 % de phase B et laisser le système s'équilibrer pendant 6 min.

(iii) système HILIC: le débit a été fixé à 200 µl/min. L'élution débute avec une étape isocratique de 2 min à 80 % de phase B. Elle se poursuit avec une décroissance linéaire des proportions de 80 à 40 % de phase B entre 2 et 12 min. Le système chromatographique est ensuite rincé pendant 5 min à 0 % de B puis rééquilibré pendant 15 min avec les proportions initiales de phases A et B.

Les facteurs de rétention ont été calculés en utilisant la formule suivante : $k = (t - t_0)/t_0$ où t désigne le temps de rétention expérimental du composé d'intérêt et t_0 est le temps mort de la colonne (t_0 a été estimé à 0.7, 1.2 et 1.7mn pour les systèmes C8, PFPP et HILIC, en nous basant sur le temps de rétention de la molécule la moins retenue des 657 molécules de référence analysées).

2.3.2. Conditions d'analyse par spectrométrie de masse

La calibration en masse du spectromètre de masse a été effectuée ou vérifiée dans les deux modes d'ionisation et avant chaque analyse en utilisant le mélange de calibration («Calmix») fourni par la société Thermo Fisher Scientifics. Le spectromètre de masse a été utilisé avec une tension du capillaire fixée à -3 kV en mode d'ionisation négative et à 5 kV à l'ionisation positive, et à une température de capillaire de 280 °C. La pression du gaz de nébulisation et la pression du gaz auxiliaire ont été respectivement fixées à 60 et 10 unités arbitraires. Le pouvoir résolutif de l'analyseur a été fixé à 50000 ($m/\Delta m$, Δm correspondant à la largeur à mi-hauteur du pic produit par un ion monochargé à m/z 200). La détection des ions a été réalisée entre 85 et 1000 u en mode d'ionisation positive, et entre 95 à 1000 u pour les analyses en conditions RP et PFPP avec détection en mode d'ionisation négative, et de m/z de 85 à 1000 pour les conditions HILIC avec détection en mode d'ionisation positive.

Les expériences de MS/MS ont été réalisées en mode d'activation non résonante de type HCD (de l'anglais «Higher energy Collision induced Dissociation») à des valeurs d'énergies de collision normalisées de 10, 12 et 23 %.

2.4. Traitement des données

Les données brutes ont été visualisées à l'aide du module Qualbrowser du logiciel Xcalibur version 2.1 (Thermo Fisher Scientific, Courtaboeuf, France). La détection automatique des pics et l'intégration des aires ont été réalisées en utilisant l'algorithme MatchedFilter du logiciel XCMS (Smith et al; 2006). Le Module « Quanbrowser » du logiciel Xcalibur version 2.1 (Thermo Fisher Scientific, Courtaboeuf, France) a été utilisé pour la détection et l'intégration des pics correspondants

aux IS et ExS en parallèle de la détection XCMS. Le logiciel CAMERA (Bioconductor version 2.8) (Kuhl et al; 2012) a été utilisé pour regrouper les ions générés au cours du processus d'ionisation à partir des ions monoisotopiques et les annoter (isotopes, adduits et pertes les plus courants). Les masses précises des ions contenues dans les matrices données issues des logiciels XCMS et CAMERA ont ensuite été annotées grâce à l'utilisation de bases de données publiques (KEGG (Kanehisa & Goto, 2000), HMDB (Wishart et al; 2007) et METLIN (Smith et al; 2005)) à l'aide d'un outil informatique développé en langage R. Les paramètres utilisés pour le traitement des données sont affichés en ANNEXE.2. Table S.1.

3. Résultats et discussion :

Le but de cette étude est de développer une base de données spectrale permettant d'annoter et d'identifier des métabolites dans des milieux biologiques. Nous avons pour cela utilisé trois systèmes LC-HRMS qui mettent en jeu des mécanismes de rétention chromatographique différents : polarité de phase inversée (RP-C18), interaction hydrophile zwitterionique (ZICp-HILIC) et une phase PentaFluoroPhenylPropyl (PFPP). Nous avons dans un premier temps mis en place une base de données spectrales à l'aide de plusieurs centaines de composés de référence, qui dans un second temps a servi à l'annotation de jeux de données de sérum humain.

3.1. Construction de la banque de données spectrales

Il a été montré qu'en moyenne 6 à 10 ions d'ions tels que des adduits, des isotopes ou des produits de fragmentation dans la source, peuvent être générés au cours du processus d'ionisation Electrospray en plus des molécules protonées ou déprotonées ((Brown et al; 2009,Werner et al; 2008a)). Cette redondance du signal est également observée dans les jeux de données de métabolomique et rend de ce fait les recherches dans les bases de données chimiques et/ou biochimiques plutôt infructueuses car les résultats peuvent être erronés (Courant et al; 2012,Evans et al; 2009,Roux et al; 2012). Il est donc nécessaire de développer des bases de données spectrales afin de mettre en évidence cette redondance du signal. Nous avons pour cela enrichi la base de données spectrales préalablement développée au laboratoire pour annoter le métabolome urinaire humain par spectrométrie de masse à haute résolution (Thèse de Doctorat de l'Université Paris 6 soutenue en 2011 par Aurélie Roux et (Roux et al; 2012)). Pour ce faire, nous avons recueilli des données spectrales (spectres de masse et spectres MS/MS) de 657 métabolites d'intérêt biochimique, ainsi que leurs temps et facteurs de rétention dans les 3 systèmes chromatographiques d'intérêt en les analysant en duplicat par groupes de 30.

Parmi les 657 composés de référence analysés, 24 n'ont pas été détectés dans les trois systèmes LC-HRMS, probablement en raison de problèmes d'instabilité en solution, d'insolubilité (exemple des

lipides) ou d'incompatibilité avec les conditions chromatographiques utilisées. 633 composés ont été détectés par au moins l'un des trois systèmes LC-HRMS, et 558 d'entre eux ont été retenus (c'est-à-dire qu'ils présentaient un facteur de rétention (k) supérieur à 1) sur au moins l'un des trois systèmes chromatographiques.

Le nombre de métabolites détectés était sensiblement le même dans les trois conditions chromatographiques (571, 552 et 579 molécules détectées en utilisant les colonnes RP, HILIC et PFPP respectivement). Comme le montre la Figure 31, la rétention des composés varie selon les conditions chromatographiques utilisées : 381 molécules ont été retenues sur la colonne PFPP, alors que 236 et 313 l'ont été sur les systèmes RP et HILIC. En outre, seulement 18 métabolites ont été spécifiquement retenus en conditions RP, contre 155 et 65 pour les systèmes HILIC et PFPP respectivement. Ceci indique que la couverture la plus globale serait atteinte par la combinaison des deux dernières conditions LC, avec la rétention de 540 métabolites sur 558.

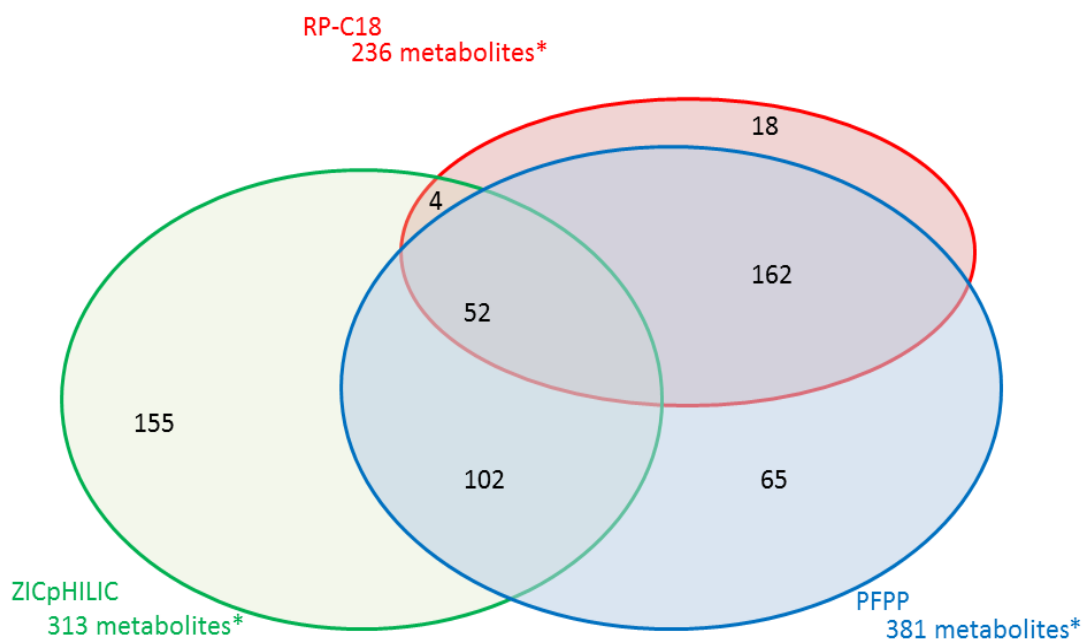


Figure 31. Répartition des 558 composés de référence retenus dans les trois systèmes LC-HRMS (Boudah et al; 2014)

* Composés de référence présentant une valeur de $k > 1$

Sans surprises, la rétention chromatographique est apparue très dépendante des propriétés physico-chimiques des métabolites, comme l'illustre la Figure 32. La colonne PFPP est particulièrement bien adaptée à la séparation d'isomères d'acides aminés. Par exemple, la Figure 32.A montre la résolution de 3 composés d'acides aminés isomères : la norleucine, l'isoleucine et la β -leucine grâce au système PFPP. En revanche, la rétention et la séparation des acides ursodésoxycholique et désoxycholique, deux acides biliaires isomères, n'a été obtenue qu'avec le

système RP (Figure 32.B), alors que la colonne HILIC s'est avérée être la plus efficace pour l'analyse des hydrates de carbone (Figure 32.C).

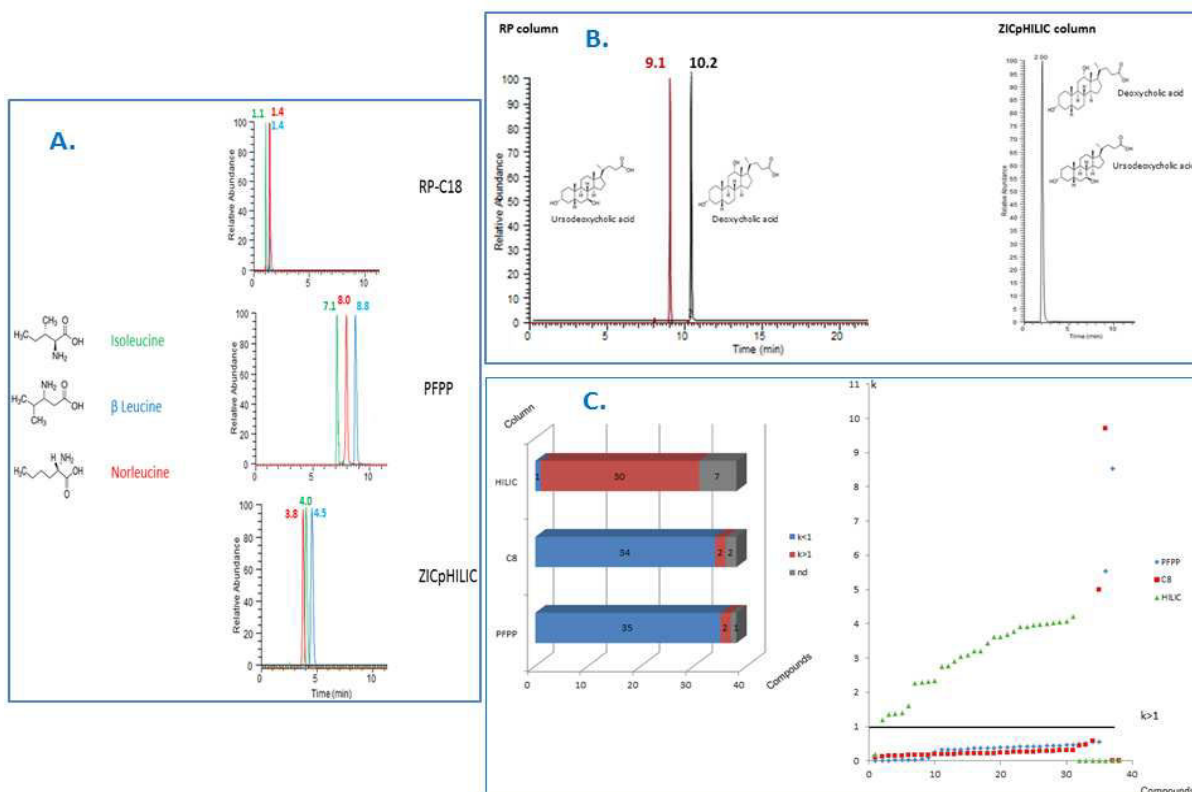


Figure 32. Complémentarité des approches LC dans la rétention et la résolution des composés de référence (Boudah et al; 2014)

En (A.) La rétention et la résolution des acides aminés isomères : isoleucine, β -leucine et norleucine par la colonne PFPP.

En (B.) La rétention et la résolution des acides biliaires isomères « acide ursodéoxycholique » et « acide désoxycholique »

sur colonne C18. En (C.) Le nombre de composés de référence de la classe des carbohydrates retenus est supérieur à celui des autres systèmes chromatographiques

L'ensemble de ces données démontrent clairement que la complémentarité des 3 systèmes chromatographiques l'emporte sur leur redondance. Les temps de rétention chromatographiques observés de chaque métabolite ont été importés dans notre base de données spectrales en vue d'annoter des jeux de données métabolomiques.

3.2. Annotation du métabolome plasmatique humain

3.2.1. Démarche d'annotation

La démarche d'annotation de jeux de données est expliquée en Figure 33. Un échantillon de sérum provenant d'individus sains a été extrait en triplicat et analysé par les 3 systèmes LC-HRMS

dans les deux modes d'ionisation Electrospray, ce qui amène à un total de 6 jeux de données. Les données brutes issues de chaque analyse LC-HRMS ont été traitées indépendamment par les logiciels XCMS (Smith et al; 2006) et CAMERA (Kuhl et al; 2012). XCMS a assuré la détection des pics, leur alignement en masse et en temps de rétention, puis a restitué les résultats sous la forme de tableaux contenant les caractéristiques de chacun des ions, ou variables (c'est-à-dire la masse précise et le temps de rétention) ainsi que les abondances (aires du pic chromatographique de l'ion extrait) dans chacun des échantillons (Smith et al; 2006). Les variables ont ensuite été traitées automatiquement par le logiciel CAMERA (Kuhl et al; 2012) qui regroupe les variables liées à un métabolite donné et générées au cours du processus de désorption. Dès lors, nous avons procédé à la vérification de la qualité des données générées.

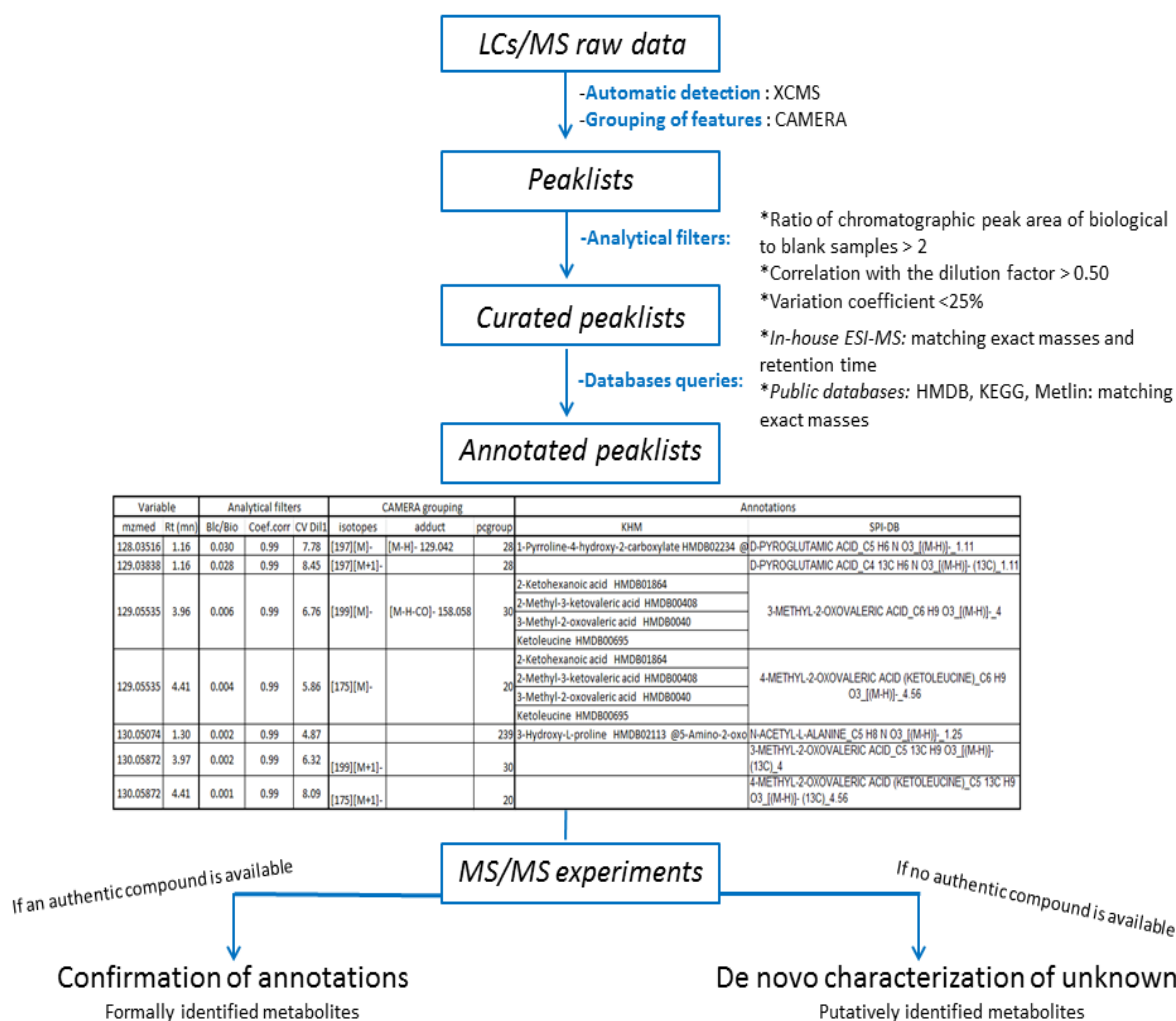


Figure 33. Démarche d'annotation et d'identification des métabolites à partir d'un jeu de données biologiques (Boudah et al; 2014)

En effet, les analyses LC-HRMS peuvent être sujettes à des dérives de temps de rétention et des écarts de précision de mesure de masse au cours d'une analyse. Afin de vérifier l'ensemble du processus d'acquisition de données, des étalons internes ont été ajoutés avant (IS) et après (ExS) l'étape de traitement des sera par le MeOH (voir Table 9). Ces étalons internes ont été choisis de

manière à couvrir les plages de masse et de temps de rétention pertinentes pour la détection des métabolites dans chaque système LC-HRMS. La bonne reproductibilité des temps de rétention et de mesure de masses précises des étalons internes a été évaluée à partir des données issues d'XCMS et également des données provenant d'intégrations manuelles réalisées avec le module QuanBrowser du logiciel XCalibur (voir ANNEXE.2. Tables S.2, S.3 et S.4). La concordance des résultats obtenus valide à la fois le bon déroulement des expériences de LC-HRMS et le processus de détection automatiques des signaux réalisé par XCMS.

Les variables générées par XCMS ont ensuite été filtrées selon les critères suivants: (i) la corrélation entre les facteurs de dilution des échantillons sériques et les aires de pics chromatographiques (les variables filtrées doivent présenter des coefficients de corrélation supérieurs à 0,5), (ii) la répétabilité (évaluée par le calcul des coefficients de variation, dont les valeurs doivent être inférieures à 25%), et (iii) les rapports des aires des pics chromatographiques dans les échantillons biologiques sur celles des blancs, dont les valeurs doivent être supérieures à 2 (voir Figure 33).

Les jeux de données ainsi filtrés ont d'abord été annotés en comparant les temps de rétention, les masses précises et si nécessaire les spectres de MS/MS obtenus dans les échantillons biologiques à ceux de la base de données spectrales du laboratoire. Cette annotation peut être considérée comme la première étape d'identification formelle des métabolites dans la mesure où elle repose sur au moins deux paramètres physico-chimiques indépendants (la masse précise, le(s) temps de rétention, et éventuellement les spectres CID) (Sumner et al; 2007).

Cette première étape d'identification a été rendue possible grâce à la robustesse des méthodes chromatographiques et aux mesures de masses précises à haute résolution. Ainsi, sur une série d'analyses de 20 échantillons, les temps de rétention des 13 étalons internes se sont avérés très stables, avec des valeurs de CV inférieures à 2% pour les 3 systèmes (ANNEXE.2. Tables S.2, S.3 et S.4). De plus, aucune différence de temps de rétention n'a été observée pour les IS qu'ils soient surchargés dans du solvant (blancs d'analyse) ou dans du sérum, suggérant ainsi que les temps de rétention obtenus à partir des composés de référence dilués dans une solution hydro-organique sont comparables à ceux mesurés dans les échantillons sériques. Ceci valide la pertinence de notre banque de données spectrale construite en solution tampon pour l'annotation de jeux de données biologiques. De plus, les exactitudes de mesures de masse en mode d'ionisation positif et négatif sont en moyenne inférieures à 3 ppm.

Ensuite, les jeux de données ont été annotés avec les masses précises théoriques de métabolites contenues dans les bases de données publiques suivantes: KEGG (Kanehisa & Goto, 2000), HMDB (Wishart et al; 2007) et METLIN (Smith et al; 2005). Cette procédure a permis de proposer des hypothèses structurales pour les variables qui ne sont pas annotées par la base de

données spectrale et pour lesquelles nous ne disposons pas du composé de référence (Figure 33). Ces propositions de structures chimiques ont ensuite été confirmées ou infirmées par des déterminations de composition élémentaire et l'interprétation de spectres de CID (Sumner et al; 2007).

3.2.2. Identification formelle des métabolites sériques humains à l'aide de la base de données spectrales

Les six matrices de données issues de XCMS (3 conditions chromatographiques et deux modes d'ionisation) ont tout d'abord été annotées par notre base de données spectrales. La tolérance d'erreur de mesure de masse a été fixée à un maximum de 10 ppm et l'intervalle de temps de rétention est exprimé selon l'équation suivante $x \pm RT^y$ (x étant un temps de rétention donné, issu d'une analyse XCMS et exprimé en minutes). Les paramètres « x » et « y » ont été respectivement fixés à 5 min et 0,8 pour les conditions HPLC, et à 2min et 0,5 pour les analyses UHPLC. Cette équation a été choisie et ses paramètres optimisés afin de modéliser le mieux possible les variations observées sur les temps de rétention, qui sont susceptibles de s'accroître au fur et à mesure du déroulement de l'élution chromatographique. Des expériences de MS/MS en mode HCD ont ensuite été réalisées afin de confirmer l'identité des métabolites annotés les plus abondants.

D'ordinaire, pour qu'un métabolite soit formellement identifié à partir d'un jeu de données biologiques, la masse précise, le(s) temps(s) de rétention et le spectre CID (si disponible) obtenus à partir d'échantillons biologiques doivent correspondre à ceux de la base de données ESI-MS. L'utilisation conjointe de trois méthodes chromatographiques a fiabilisé le processus d'annotation et a permis d'identifier 192 métabolites sériques. Les exemples décrits ci-dessous illustrent la complémentarité des 3 systèmes chromatographiques et de leur couplage à la spectrométrie de masse à haute résolution.

3.2.2.1. La séparation chromatographique des espèces isobares

Le sérum est un milieu biologique complexe, et bien que les données aient été acquises avec un spectromètre de masse à haute résolution, les ions extraits à partir des chromatogrammes peuvent inclure plusieurs espèces isobares non résolues dans le domaine spectral et qui peuvent être séparées grâce à la chromatographie. Ainsi, l'ion m/z 135,030 détecté en mode d'ionisation négative dans le sérum humain et isolé dans une fenêtre de 10 ppm peut correspondre à deux métabolites distincts: l'hypoxanthine ($C_5H_3ON_4$) ou l'acide thréonique ($C_4H_7O_5$). Deux pics élués à proximité du volume mort ont été détectés dans les conditions RP et l'identité des deux métabolites détectés fut confirmée grâce à des expériences de MS/MS (voir la Figure 34). Ces deux métabolites ont par contre été retenus et séparés en conditions HILIC, avec des temps de rétention similaires à ceux référencés

dans notre base de données spectrales pour les composés de référence de l'hypoxanthine et de l'acide thréonique. Il a donc été possible de distinguer et d'identifier sans ambiguïté ces deux molécules isobares différant seulement de 0,001 u.

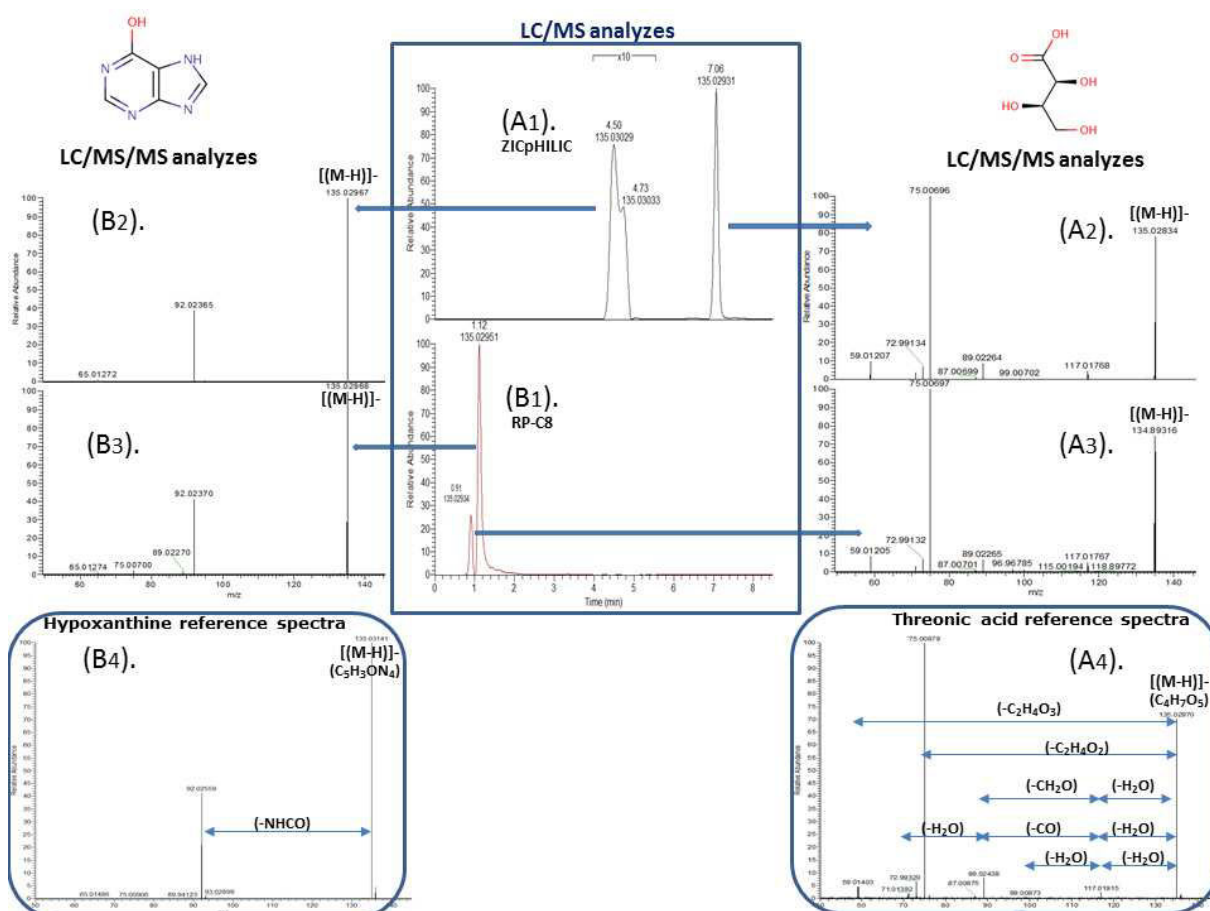


Figure 34. L'identification des métabolites isobares dans le sérum humain : discrimination entre l'acide thréonique et l'hypoxanthine (Boudah et al; 2014).

(A1) et (B1): XIC de l'ion m/z $135,029 \pm 10$ ppm récupérés à partir des profils LC-ESI(-)-Q-Exactive de sérum humain obtenus dans des conditions HILIC et RP. (A2) et (A3) sont des spectres CID non résonants (HCD) obtenus pour les pics élués à 7,1 et 0,9 min à partir de l'échantillon de sérum dans des conditions HILIC et RP, respectivement. (A4) spectre CID-non résonant (HCD) du composé de référence de l'acide thréonique. (B2) et (B3) sont des spectres CID non résonant (HCD) obtenus pour les pics élués à 4,5 et 1,1 min à partir de l'échantillon de sérum dans des conditions HILIC et RP, respectivement. (A4) spectre CID-non résonant (HCD) du composé de référence d'hypoxanthine. L'inspection visuelle des différents spectres CID non résonants permet l'identification sans équivoque de l'hypoxanthine et l'acide thréonique dans des échantillons de sérum. Notez l'inversion de l'ordre d'élué des deux métabolites entre les conditions RP et HILIC.

3.2.2.2. La séparation chromatographique des isomères

La séparation de molécules isomères représente une difficulté encore plus importante. De tels composés ont en effet des compositions élémentaires similaires et ne sont donc pas distinguables sur un spectre de masse. Des informations structurales complémentaires peuvent cependant être obtenues en utilisant (i) des expériences de MS/MS (exemple de l'acide hippurique et de la succinoylpyridine, présenté en Figure 35) (ii) plusieurs systèmes chromatographiques

complémentaires (exemple des acides dihydroxybenzoïques présenté en Figure 36) ou encore (iii) les deux à la fois (exemple de l'identification du sulfate de p-crésol décrit en Figure 37).

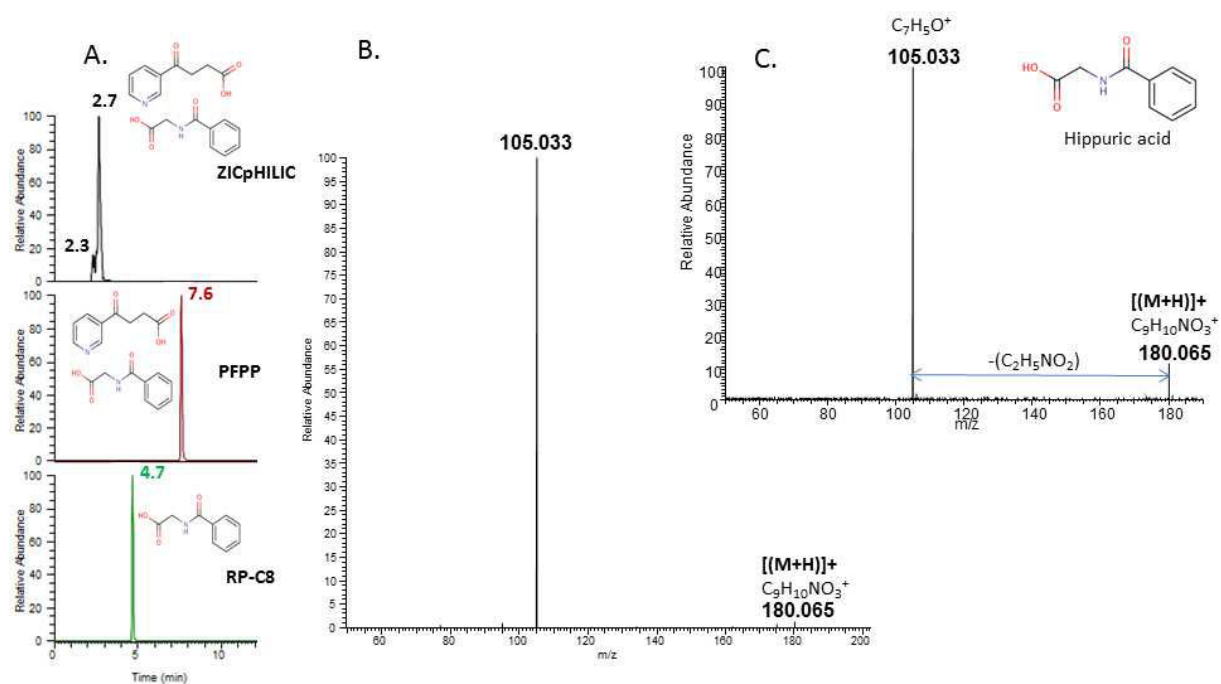


Figure 35. L'identification des métabolites isomères dans le sérum humain : discrimination entre l'acide hippurique et la succinoylpyridine (Boudah et al; 2014).

(A.) : XIC de l'ion m/z $180,065 \pm 10$ ppm récupérés à partir des profils LC-ESI(+)-Q-Exactive de sérum humain obtenus dans des conditions HILIC et PFPP et RP respectivement. La masse précise et les temps de rétention en conditions HILIC et PFPP pourraient correspondre à l'acide hippurique ($C_9H_9NO_3$) ou à la succinoylpyridine ($C_9H_9NO_3$). La discrimination a été possible grâce au temps de rétention obtenu en conditions RP pour lesquelles le composé de référence succinoylpyridine n'est pas retenu. (B.) spectre CID non résonant (HCD) obtenu à partir des échantillons sériques. (C.) spectres CID non résonant (HCD) obtenu à partir du composé de référence d'acide hippurique. Confirmation de la structure a été obtenue par des expériences MS/MS effectuées dans les trois conditions de LC : la perte des groupes $C_2H_5O_2N$ et $C_3H_5O_3N$ n'était pas compatible avec la structure de la succinoylpyridine où l'atome d'azote est engagé dans le cycle aromatique. Dans ce cas, le métabolite présente un temps de rétention spécifique et un motif de fragmentation caractéristique qui permet son identification formelle.

Valeur ajoutée de la sélectivité chromatographique

Les expériences de MS/MS sont parfois d'un intérêt limité notamment lorsque (i) les ions précurseurs sont très stables et qu'ils ne produisent pas d'ions fils « diagnostics », (ii) l'intensité de l'ion précurseur est trop faible pour permettre de réaliser des expériences MS/MS, (iii) l'ion précurseur est coélué à d'autres molécules isobares, ou enfin (iv) lorsque plusieurs molécules isomères produisent des spectres CID identiques ou proches. Ainsi, par exemple, plusieurs isomères de l'acide dihydroxybenzoïque ($C_7H_6O_4$) sont supposés être présents dans le sérum humain selon la base de données HMDB: l'acide gentisique (acide 2,5- dihydroxybenzoïque), les acides résorcyliques (acides 2,6- , 2,4- , et 3,5- dihydroxybenzoïques) et l'acide pyrocatechique (acide 2,3- dihydroxybenzoïque). Les spectres HCD des composés de référence correspondants ne présentent pas d'ions diagnostic permettant d'individualiser tous ces isomères (voir Figure 36). Les analyses

réalisées avec le système PFPP ont permis la détection sélective de trois pics dont deux correspondent aux acides 2,5- et 2,6-dihydroxybenzoïques (pics retenus et résolus). Le troisième isomère d'acide dihydroxybenzoïque détecté ne correspond à aucun des produits de référence de notre chimiothèque.

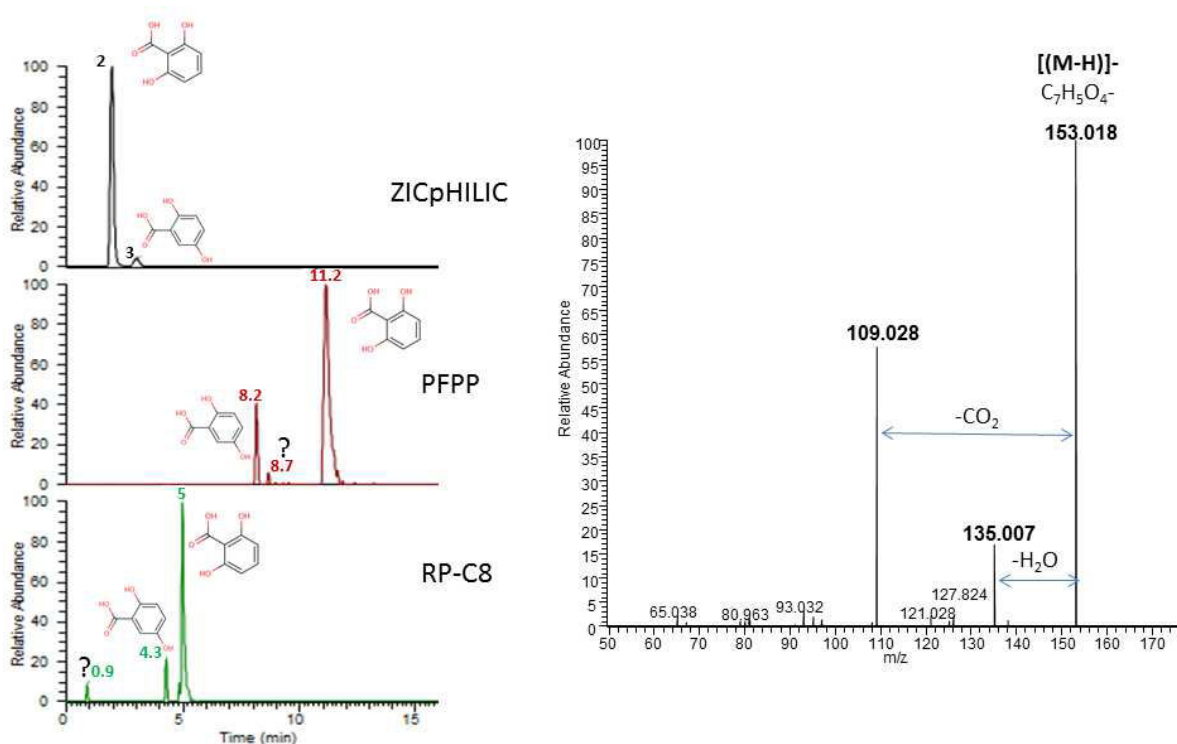


Figure 36. Discrimination entre les isomères d'acide dihydroxybenzoïque (Boudah et al; 2014).

Complémentarité des approches LC et des spectres MS/MS

Dans certaines situations extrêmes, toutes les données (masse précises, plusieurs temps de rétention et expériences de MS/MS) peuvent être requises pour l'identification d'un métabolite. Par exemple, l'ion à m/z 187,006 détecté en mode d'ionisation négative dans l'échantillon sérique peut potentiellement correspondre à deux types de structures que sont les isomères du sulfate de crésol et de l'alcool sulfobenzyle (voir Figure 37). Les expériences MS/MS ont permis d'éliminer l'hypothèse des alcools sulfobenzyles, étant donné que de telles structures ne peuvent pas produire l'anion radical à m/z 79,957 (SO_3^-), qui nécessite une liaison à un groupement phénolique. Ainsi, la présence d'isomères de sulfate de crésol semble être plus probable. La présence de sulfate de *mé*ta-crésol a été écartée car une telle position sur le noyau phényle entraverait la formation de certains ions fragments en raison de la délocalisation d'électrons. À l'inverse, les deux positions *ortho* et *para* du groupe hydroxyle sont compatibles avec le spectre MS/MS obtenu. Finalement des expériences de surcharge des composés de référence des deux isomères dans l'échantillon de sérum ont permis d'exclure l'isomère *ortho* et d'identifier le sulfate de *p*-crésol.

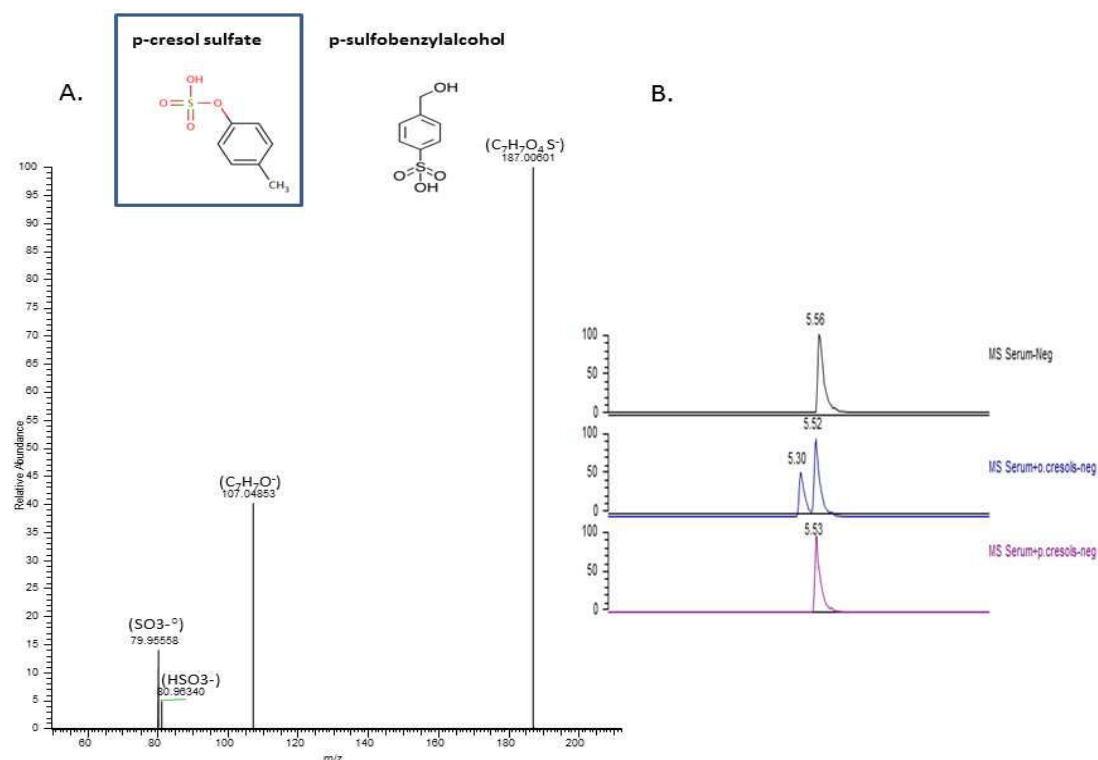


Figure 37. Identification de sulfate de p-crésol dans le sérum humain (Boudah et al; 2014).

(A): Spectre CID non résonant obtenu dans le mode d'ionisation négative de l'ion m/z 187,006 éluée à 5,5 min correspondant au sulfate de p-crésol dans des échantillons de sérum. (B): XIC de l'ion m/z 187,006 obtenu en RPLC-ESI(-)-HRMS dans des échantillons de sérum surchargés ou non avec des composés de référence de sulfate d'o-, ou de p-crésol, analysés dans des conditions RP-C18.

3.2.3. Bilan des variables annotées à l'aide de la base de données spectrales

Analyses en mode d'ionisation négatif en électrospray

Notre approche a conduit à l'annotation de 211, 63 et 154 variables présentes dans les jeux de données provenant d'acquisitions en mode d'ionisation négative en conditions HILIC, RP et PFPP. En ne prenant en compte que les ions pseudomoléculaires, ou à défaut, les adduits chlorures, 54 variables ont été annotées sur la base de mesures de masse précises et de temps de rétention. De plus, des spectres MS/MS ont également été acquis pour 86 autres variables. Ces variables correspondent à 140 métabolites formellement identifiés, selon les critères actuellement en vigueur (Sumner et al; 2007).

La méthode HILIC couplée à une détection électrospray en mode négatif est celle qui permet de détecter le plus grand nombre de métabolites. En effet, 130 métabolites sur les 140 identifiés en ionisation négative ont été détectés dans les conditions HILIC, et 103 d'entre eux ont pu être identifiés sans qu'il soit nécessaire d'avoir recours à un autre système chromatographique (Table 10). Cependant, les isomères d'acides aminés et d'acides aromatique sont mieux séparés en conditions PFPP et/ou RP. Les conditions RP apparaissent d'avantage adaptées à l'analyse des acides biliaires et

des acides gras à chaîne courte. Les acides gras à longue chaîne seront traités dans la PARTIE A.CHAPITRE II.

	Métabolites détectés	Métabolites non retenus	Métabolites identifiés	Métabolites exclusivement identifiés par la méthode LC-HRMS
C18	86	45	39	5
PFPP	91	26	53	9
HILIC	130	45	103	78

Table 10. Métabolites détectés et identifiés dans du sérum humain en mode d'ionisation négative

L'analyse en conditions RP a permis l'identification spécifique de 5 métabolites (acides biliaires et acides carboxyliques hydroxylés) ; et 9 métabolites ont été spécifiquement identifiés en conditions PFPP. Il s'agit d'amines quaternaires détectées sous forme d'adduits chlorures, de dérivés de purine et d'acides carboxyliques aromatiques.

Analyses en mode d'ionisation positif en électrospray

En ce qui concerne la détection en mode d'ionisation positive, notre travail a abouti à l'annotation de 141, 88 et 130 variables présentes dans les jeux de données issus des analyses en conditions HILIC, RP et PFPP. Cela a conduit à l'identification formelle de 104 métabolites, sur la base de leur masse précise et leur temps de rétention pour 27 d'entre eux, et grâce à l'utilisation complémentaire d'expériences de MS/MS pour 77 d'entre eux. Les analyses en conditions PFPP se sont cette fois révélées les plus informatives, en particulier pour la détection des acides aminés, des purines et de certains dérivés de carnitine. En effet, parmi les 104 molécules identifiées, 92 ont été détectés en conditions PFPP, et 71 d'entre elles ont été retenues efficacement (Table 11). Une fois de plus, les conditions RP étaient les moins informatives. Bien que 75 métabolites sur 104 aient été détectés dans ces conditions, seulement 20 d'entre eux ont été retenus efficacement par ce système chromatographique, et seulement 2 ont été spécifiquement retenus.

Enfin, en combinant les deux modes d'ionisation, 192 métabolites ont été formellement identifiés, dont 33 molécules qui sont, à notre connaissance, décrites dans le sérum humain pour la première fois (voir ANNEXE.2 Table.S.5).

	Métabolites détectés	Métabolites retenus	Métabolites identifiés	Métabolites exclusivement identifiés par la méthode LC-HRMS
C18	75	48	20	2
PFPP	92	21	65	27
HILIC	80	14	61	29

Table 11. Métabolites détectés et identifiés dans du sérum humain en mode d'ionisation positive

3.2.4. Caractérisation de métabolites inconnus

Nous nous sommes ensuite intéressés aux signaux les plus intenses qui ont été annotés par les bases de données publiques, mais pas avec notre base de données spectrales. Nous avons tenu compte des propositions de regroupement de signaux fournis par le logiciel CAMERA et nous avons vérifié que les signaux d'intérêt annotés par les bases de données publiques correspondaient bien à des ions monoisotopiques et non pas à des isotopologues naturels ou à d'autres types d'ions produits pendant le processus d'ionisation-désolvatation. Nous avons ensuite généré des compositions élémentaires à partir des mesures de masses précises en nous aidant principalement de la règle de l'azote et des signatures isotopiques pour éliminer les propositions non pertinentes. Des expériences de CID non résonantes réalisées avec un spectromètre de masse de type Q-Exactive ont ensuite été entreprises. L'interprétation de ces spectres nous a permis de valider ou d'invalider les annotations proposées par les bases de données publiques (Figure 38); ou encore de choisir la structure la plus pertinente si les bases de données publiques fournissaient plusieurs propositions (Figure 39).

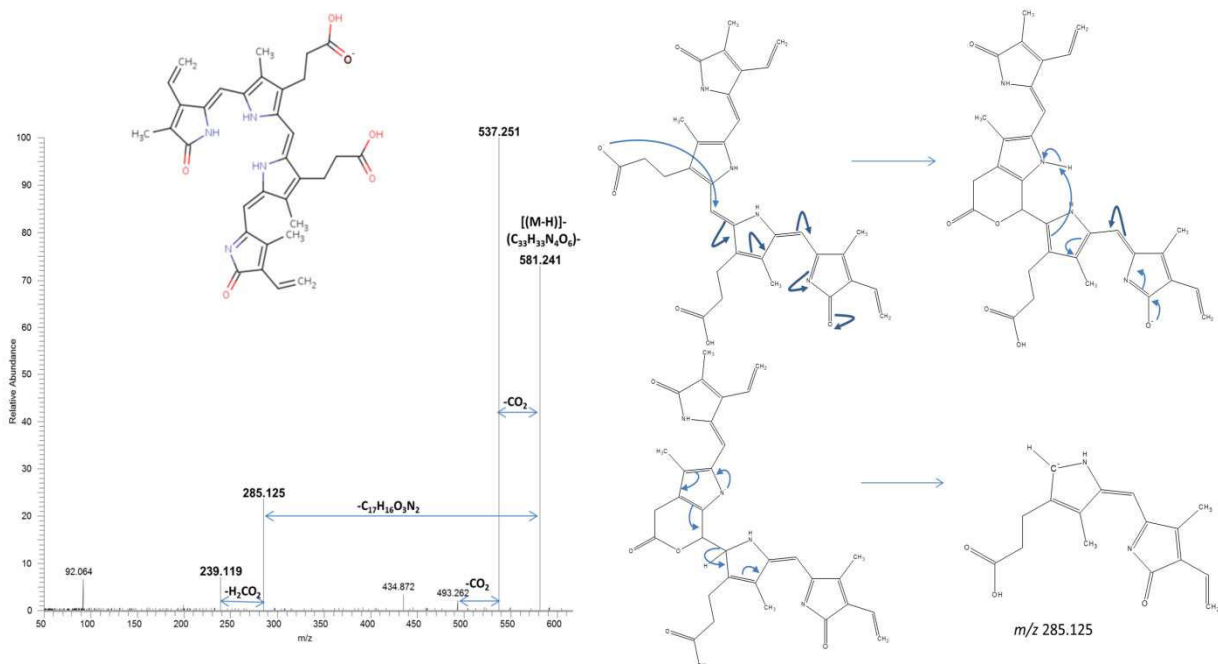


Figure 38. Identification putative de la Biliverdine dans le sérum humain (Boudah et al; 2014).

A gauche, le spectre CID non-résonant de l'ion m/z 581,239 détecté en mode d'ionisation négative à caractériser. Le pic de base (m/z 537,251) correspond à une perte de CO_2 et ne peut être considéré comme un ion diagnostique. Cela nous a conduits à nous intéresser plus particulièrement à l'ion produit m/z 285,124. À droite, le schéma de fragmentation proposé pour l'élucidation de la structure proposée pour l'ion produit à m/z 285,124 à partir de la biliverdine.

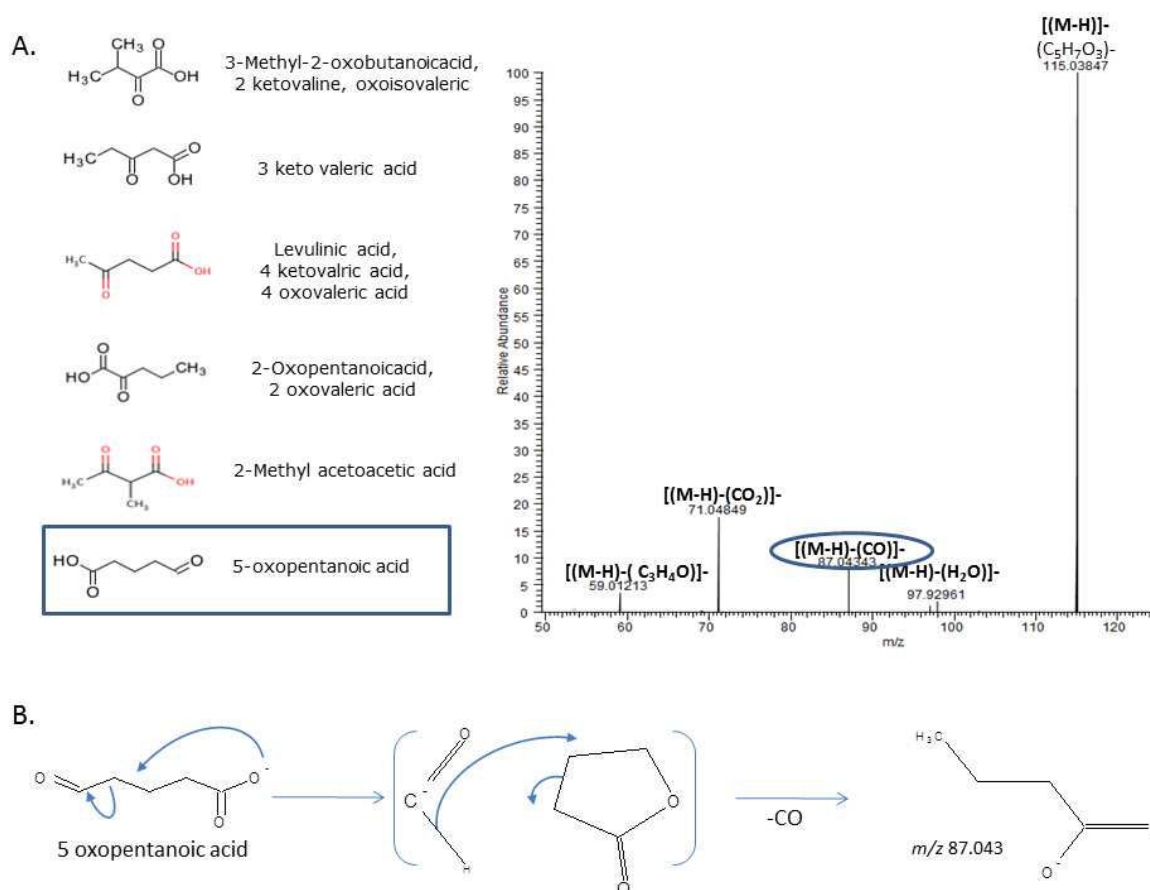


Figure 39. Identification putative de l'acide 5-oxopentanoïque dans le sérum humain (Boudah et al; 2014).

(A.) Spectre CID non résonant de l'ion m/z 115,038, acquis en mode d'ionisation négative à partir d'un échantillon de sérum. Plusieurs isomères correspondant à la composition élémentaire calculée sont proposés par les bases de données publiques. Les deux ions produits m/z 59,012 et m/z 87,043 sont particulièrement intéressants comme ils ne sont pas compatibles avec toutes les structures proposées. Le premier, correspond à l'ion $\text{CH}_3\text{-COO}^-$, ne peut être produit que par deux des structures proposées : l'acide lévulinique et l'acide 5-oxopentanoïque. La perte du groupe CO donnant l'ion m/z 87,043, a été considérée comme un ion diagnostique, car n'est compatible qu'avec la structure de l'acide 5-oxopentanoïque. Ce métabolite a donc été putativement identifié. (B.) Schéma de fragmentation proposé pour la perte de CO à partir de l'acide 5-oxo pentanoïque en mode d'ionisation négative.

Nous avons appliqué cette démarche à 38 et 40 signaux respectivement détectés en modes négatif et positif d'ionisation électrospray, et nous avons abouti à 74 identifications putatives (ANNEXE.2 Table.S.5).

3.2.5. Bilan des métabolites identifiés et annotés dans l'échantillon de sérum humain

Notre approche a conduit à l'identification de 266 métabolites dans un échantillon de sérum humain, parmi lesquels 192 ont été formellement identifiés sur la base de leur masse précise et d'au moins un temps de rétention, et 140 ont été formellement identifiés grâce à leurs masse précise et l'interprétation de spectres MS/MS (la liste exhaustive des métabolites identifiés est fournie en ANNEXE.2 Table.S.5). D'autre part, nous avons proposé une structure chimique pour 74 métabolites

pour lesquels nous ne disposons pas de composés de référence dans notre chimiothèque. Parmi ces 266 métabolites identifiés, la présence de 59 d'entre eux dans le sérum humain n'était pas mentionnée dans les bases HMDB et PUBMED au moment de cette étude (Octobre 2013). C'est par exemple le cas de la biliverdine, qui est un produit du catabolisme de l'hémoglobine qui n'est pas supposé être présent dans le sérum humain d'individus sains.

En combinant trois systèmes LC-HRMS, 209 masses précises correspondant à 266 métabolites distincts ont été annotées, ce qui indique une proportion de 27% d'isomères parmi les métabolites que nous avons identifiés dans le sérum humain. Cette proportion est similaire à celle donnée par la base de données HMDB. D'autre part, la méthode HILIC est celle qui permet de détecter le plus grand nombre de métabolites (219 métabolites sur 266), et la méthode PFPP est celle qui permet de retenir le plus de métabolites (151 métabolites sur 266) (Figure 40).

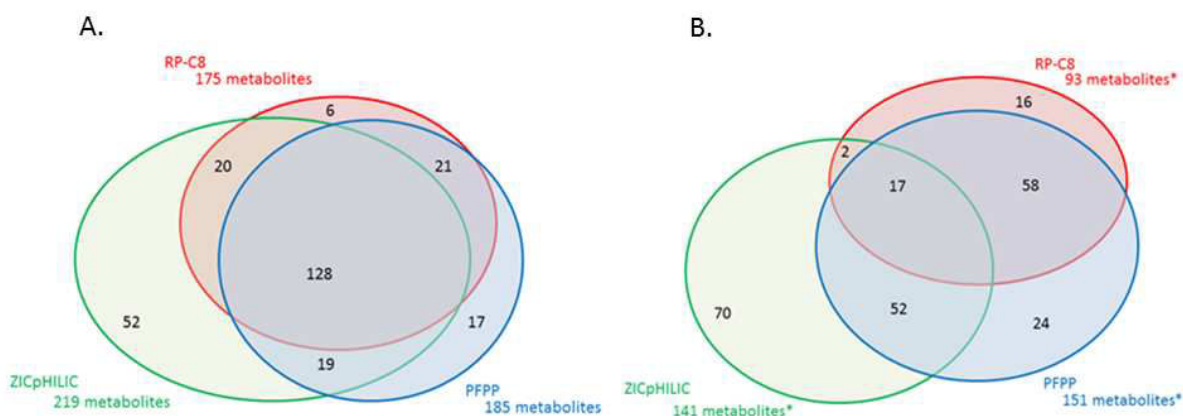


Figure 40. Contribution des plateformes LC-HRMS à l'extension de la couverture métabolique sérique (Boudah et al; 2014).

(A.) Répartition des 266 métabolites sériques humains détectés dans les trois systèmes LC-HRMS. (B.) Distribution des métabolites sériques humains retenus dans les trois systèmes LC-HRMS. * Métabolites présentant une valeur de $k > 1$

Les métabolites identifiés se répartissent sur 20 classes chimiques (Figure 41). La famille des acides aminés et de leurs dérivés est la plus représentée, avec l'identification de 72 d'entre eux. Suivent les carnitines, les sucres et les acides organiques hydroxylés. Des classes telles que les porphyrines, contenant un nombre plus limité de métabolites (bilirubine et la biliverdine) ont également été détectées. Bien que notre étude ait porté sur les espèces chimiques les plus abondantes, certains xénobiotiques tels que la théophylline, le sulfaméthoxazole, l'acétaminophène et de leurs produits de biotransformation (glucuronides ou sulfoconjugués) ont pu être détectés. Enfin, certaines classes chimiques tels que les lipides et les stéroïdes non conjugués sont peu ou pas représentées, probablement parce que notre méthode de préparation des échantillons n'est pas appropriée à leur analyse. Nous avons par exemple pu détecter des lysophosphatidylcholines en conditions RP à des temps de rétention compatibles avec leur structure chimique. Cependant, nous

n'avons pas approfondi leur caractérisation dans la mesure où nous avons développé une approche d'analyse lipidomique plus appropriée à leur analyse qui sera décrite dans la PARTIE A.CHAPITRE II.

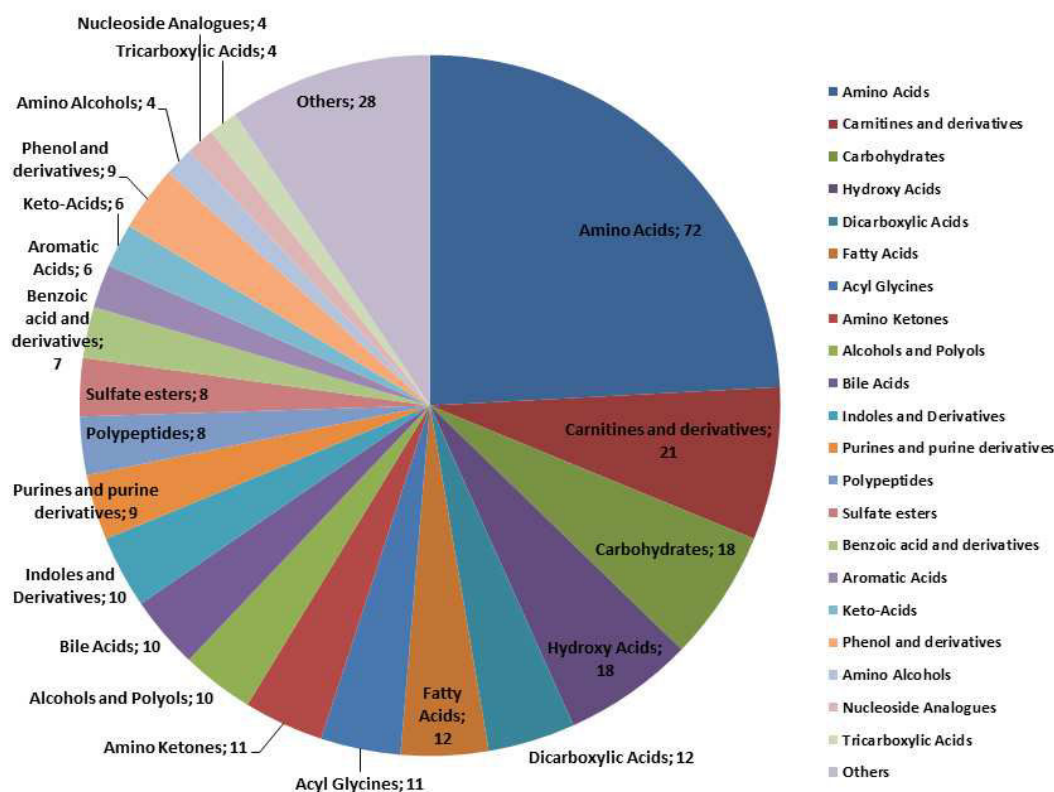


Figure 41. Métabolome sérique : Distribution des 266 métabolites sériques identifiés dans les familles chimiques (Boudah et al; 2014).

Enfin, comme le montre la Figure 42, notre approche a conduit à la détection de 30% des métabolites répertoriés dans le sang humain par la base HMDB (lipides non pris en compte, extraction de base de données réalisée le 18 Octobre 2013) ainsi que 59 métabolites décrits dans le sérum humain pour la première fois à notre connaissance. Enfin, d'autres métabolites supposés présents dans les échantillons sanguins n'ont pas été détectés en raison de problèmes de sensibilité, et aussi peut-être de la forte variabilité inter-individuelle observée dans les fluides biologiques.

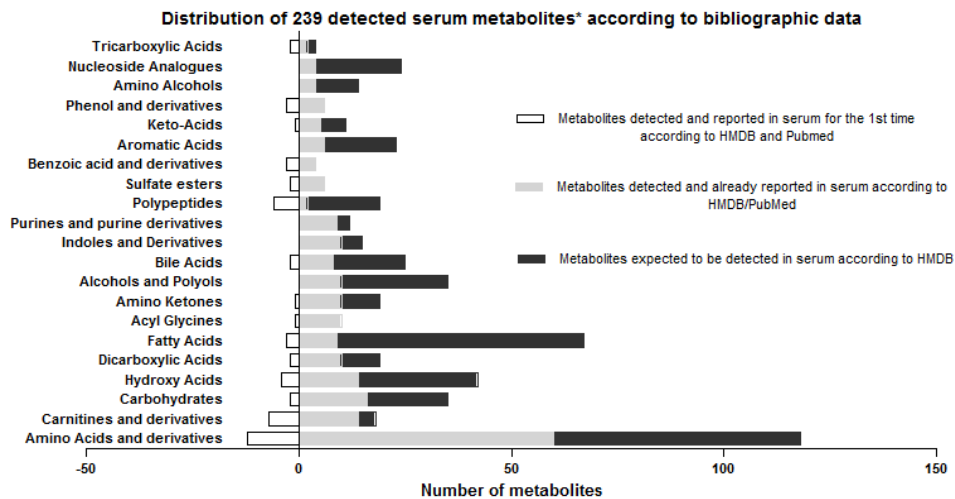


Figure 42. Métabolome sérique : Statut des 239* métabolites sériques identifiés selon les données bibliographiques (Boudah et al; 2014).

*** Métabolites appartenant aux classes chimiques les plus représentées**

Le but de ce travail était d'évaluer trois plateformes LC-HRMS pour l'annotation de signaux détectés dans un échantillon de sérum humain et l'identification de métabolites. Comme nous nous y attendions, notre étude a démontré l'intérêt de combiner plusieurs systèmes chromatographiques avec la spectrométrie de masse à haute résolution pour détecter et identifier le plus grand nombre de métabolites dans un échantillon biologique. Nous sommes également arrivés à la conclusion que le système RP, qui reste le plus fréquemment utilisé dans le domaine de l'analyse métabolomique par LC-MS est le moins efficace des 3 systèmes en matière de détection et d'identification de métabolites sériques. En effet, la plupart des métabolites détectés dans cette étude pourrait être détectés en utilisant seulement 2 conditions d'analyse sur les 6 (méthode HILIC avec une détection en mode négatif d'ionisation electrospray et méthode PFPP avec détection en mode d'ionisation positif d'ionisation).

Nous avons appliqué cette méthodologie à des extraits de GR afin d'avoir une vue exhaustive du métabolome érythrocytaire humain.

3.3. Annotation du métabolome érythrocytaire humain

Les globules rouges sont des cellules sanguines qui véhiculent l'oxygène dans tout l'organisme grâce à l'hémoglobine, pigment constitué de protéines et de fer. Il en existe 4 à 5 millions par millimètre cube de sang. Les globules rouges ou érythrocytes dérivent de la lignée érythroblastique médullaire, suite à des divisions successives puis à l'expulsion du noyau qui génère des réticulocytes qui passent dans le sang et se transforment en globules rouges matures. Ces derniers sont formés d'une membrane constituée d'une double couche lipidique à laquelle s'insèrent des protéines. Des sucres se fixent sur certaines d'entre elles, donnant lieu aux antigènes de surface qui sont à l'origine

de la détermination des groupes sanguins. D'autres protéines jouent le rôle de transporteurs et permettent ainsi de piéger et de concentrer des métabolites.

Au niveau cytosolique, l'hémoglobine constitue 98% des protéines totales, les 2% restant représentent l'arsenal enzymatique érythrocytaire qui, de ce fait, est souvent non exploité. Pendant longtemps, les seules études menées sur les protéines cytosoliques érythrocytaires se sont focalisées sur les déficits enzymatiques tels que celui de la « pyruvate kinase » ou encore de la « glucose-6 phosphate deshydrogenase ». Dès 2006, des équipes de recherche se sont intéressées au protéome des GR avec l'identification de 50 puis de 250 puis plus de 1500 autres protéines jusque là non caractérisées dans les GR, et ce essentiellement à l'aide de la déplétion en hémoglobine et techniques basées sur la MS (Kakhniashvili, Bulla, Jr., & Goodman, 2004, Pasini et al; 2006, Roux-Dalvai et al; 2008). Cette découverte soulève de nombreuses questions du point de vue de la composition du métabolome. Dans ce contexte, nous nous sommes intéressés à l'annotation du métabolome intra-érythrocytaire.

3.3.1. Bilan des métabolites annotés dans les échantillons de GR humains

Suivant la même démarche de profilage et d'annotation décrite pour le plasma humain, il nous a été possible d'annoter 205 métabolites érythrocytaires. Les conditions HILIC sont celles qui nous ont permis de détecter le plus de métabolites (voir Figure 43).

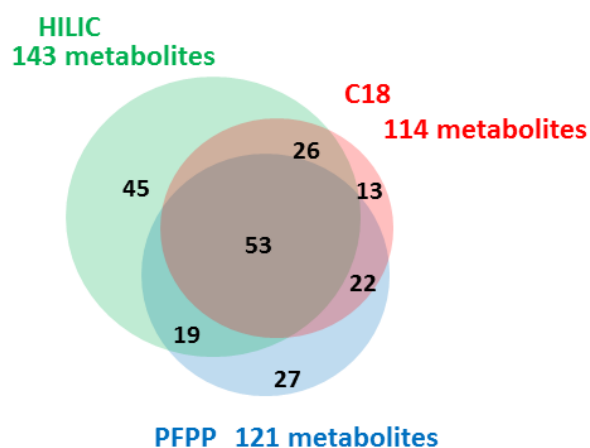


Figure 43. Contribution des plateformes LC-HRMS à l'extension de la couverture métabolique érythrocytaire

Les métabolites détectés correspondent à différentes classes chimiques (voir Figure 44). Ainsi, le métabolome du GR ne reflèterait pas uniquement les voies métaboliques érythrocytaires (qui sont essentiellement la glycolyse, ou métabolisme du glutathion), mais contiendrait en plus, d'autres métabolites tels acides aminés (classe majoritairement représentée), carnitine et dérivés,

polyamines, peptides de petite taille et vitamines (Darghouth et al; 2011a, Darghouth et al; 2011b). La présence de certains peut être due à l'arsenal enzymatique du GR, alors que d'autres ne peuvent que provenir du milieu extra-érythrocytaire (plasma et/ou autres cellules).

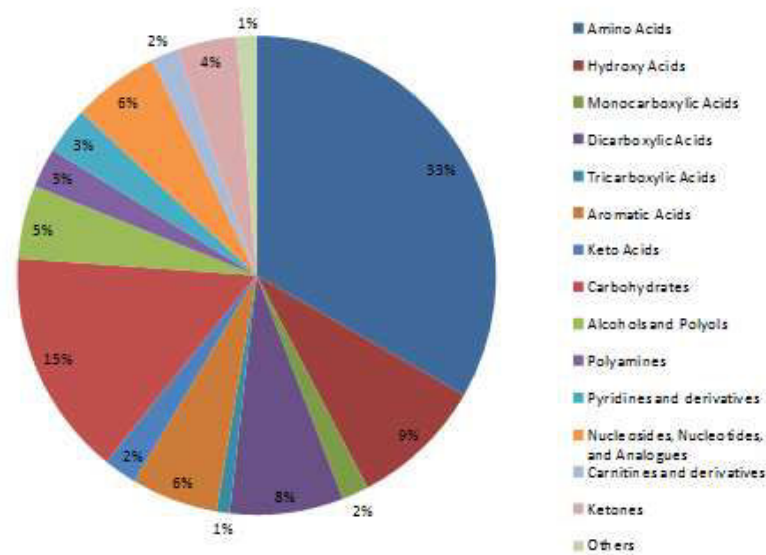


Figure 44. Métabolome érythrocytaire : Distribution des 205 métabolites érythrocytaires identifiés dans les familles chimiques

3.3.2. Du protéome expérimental au métabolome érythrocytaire théorique

Surpris de la présence de certains métabolites dans le métabolome érythrocytaire, nous avons collaboré avec l'équipe « Métabolisme des xénobiotiques » de l'INRA de Toulouse (Dr. Fabien JOURDAN). L'équipe a assuré la reconstruction du réseau métabolomique des globules rouges humains par l'intégration des données protéomiques cytosoliques existantes (Roux-Dalvai et al; 2008). Tout d'abord, il a été question de réaliser la correspondance entre les protéines identifiées et les enzymes métaboliques à l'aide du GO consortium. Ils ont ensuite ainsi généré une liste de voies métaboliques putatives à partir des identifiants EC des enzymes puis intégré le tout dans un réseau (Rbc-Cyc) à l'aide d'outils développés à l'INRA. Il s'agissait principalement de répertorier les produits mais aussi les substrats de toutes les enzymes répertoriées dans le cytosol des GR. Dans ce contexte, l'acide pantothénique par exemple (bien qu'associé au microbiote intestinal) est supposé être présent dans les GR (grâce aux protéines de transport) du fait de l'identification de l'enzyme érythrocytaire EC:2.7.1.33 qui assure sa phosphorylation. S'enchaîne alors une série de réactions biochimiques produisant le co-enzyme A.

Ainsi, il y a prédiction de 216 métabolites théoriques dans le cytosol des GR (la liste exhaustive des métabolites érythrocytaires prédits est fournie en ANNEXE.2 Table S.6). La présence de 43

d'entre eux a été confirmée dans nos empreintes LC-HRMS érythrocytaires (notamment en conditions HILIC et RP). Il s'agit essentiellement des métabolites impliqués dans la voie de la glycolyse, de la fermentation, la synthèse du glutathion ainsi que celle du NAD⁺. Ces voies assurent la production d'énergie et jouent le rôle d'un système d'oxydo-réduction. En effet, le GR doit les produire pour 2 rôles principaux : (i) maintenir l'intégrité de la membrane et ce en assurant le maintien de l'équilibre ionique par fonctionnement des pompes Na⁺, K⁺, ATPase qui nécessitent de l'ATP et (ii) maintenir l'hémoglobine sous sa forme active, c'est-à-dire réduite avec un fer divalent et ce grâce au glutathion réduit qui protège l'hémoglobine de l'oxydation (formation de méthémoglobine). Quant aux autres métabolites détectés mais pas prévus par la modélisation à savoir 162 métabolites tels que les acides aminés et carnitines, nous pouvons conclure qu'ils sont transportés depuis le milieu extérieur. De même pour les dérivés nucléosidiques, par ailleurs, ces derniers pourraient également être des vestiges des réticulocytes.

Les GR sont de ce fait une véritable mine d'informations et peuvent faire l'objet d'analyses métabolomiques intéressantes au même titre que les matrices usuelles telles que le liquide céphalorachidien (LCR), les urines ou le plasma.

4. Conclusion

Notre étude a souligné la polyvalence des approches HRMS pour la détection d'une large gamme de métabolites appartenant à différentes familles chimiques et ce dans 2 différentes matrices biologiques. Cependant, la résolution de mélanges de composés isomères reste le principal goulot d'étranglement. Ainsi, 27% des métabolites plasmatiques identifiés dans le cadre de notre étude sont des isomères. La chromatographie bidimensionnelle pourrait être un bon moyen d'aborder ce problème, comme cela a déjà été publié par GC-MS avec notamment la séparation d'isomère d'acides hydroxybutyriques (Hartonen et al; 2013). Cependant, l'utilisation de la chromatographie bidimensionnelle en phase liquide reste encore limitée par des obstacles technologiques tels que l'incompatibilité de certains types de phases mobiles lorsque l'on passe de la première à la deuxième dimension séparative, ou encore le manque d'outils de traitement informatique des données provenant de grandes séries de données produites par des approches globales. Sinon, l'électrophorèse capillaire est une approche de choix pour la séparation d'isomères polaires présentant des propriétés acido-basiques, bien qu'elle reste limitée par des problèmes de sensibilité. D'autre part, le couplage de la LC-HRMS avec la spectrométrie de mobilité ionique s'avère aussi être une approche prometteuse pour la métabolomique (Dwivedi, Schultz, & Hill, 2010), bien que la plupart des applications publiées concerne d'avantage les biomolécules de masse moléculaire supérieure à 500 Da, comme les lipides, les acides nucléiques ou les protéines (Liu et al; 2004).

Enfin, l'interprétation manuelle de ces spectres reste longue et fastidieuse. Il est nécessaire de continuer à développer des outils informatiques pour l'interprétation et la comparaison des spectres MS^2 et MS^n . Des outils tels que ceux permettant l'analyse automatisée de données MS^n et leur visualisation sous la forme d'arbres de fragmentation [22, 25], ou encore des logiciels permettant de générer des structures chimiques à partir de l'analyse automatisée de spectres de MS/MS (Wolf et al; 2010) sont maintenant disponibles. Un des principaux enjeux est maintenant de produire des données sur des composés inconnus et d'arriver à partager ce type de données au sein de la communauté scientifique.

CHAPITRE II. METHODE GLOBALE D'ANALYSE LIPIDOMIQUE

1. Introduction

Les lipides présentent une diversité structurale importante due aux différentes combinaisons possibles de têtes polaires et d'acyles gras qui complexifie leur profilage global. Dans ce contexte la technologie ESI-MS offre des perspectives intéressantes permettant de détecter un large panel de familles de lipides. Ainsi, des approches shotgun «SL» se sont popularisées. Elles assurent simultanément l'identification et la quantification de centaines d'espèces lipidiques appartenant aux différentes familles de lipides notamment à l'aide d'outils bioinformatiques dédiés, développés de façon concomitante. Néanmoins, ces méthodes souffrent d'un manque de sensibilité dû à l'effet matrice que seuls les couplages aux méthodes séparatives peut atténuer. En effet, la dimension supplémentaire apportée notamment par la chromatographie liquide permet de séparer les molécules endogènes dont les isomères, et augmente de ce fait la sensibilité de détection des espèces mineures.

Cependant, contrairement au SL, les approches intégrées couplant méthodes de profilage global et outils bioinformatiques sont peu décrites en analyses lipidomiques LC-MS. En effet, si l'on considère le plasma qui est la matrice biologique la plus communément analysée, peu d'études lipidomiques globales au sens « omiques » du terme ont été publiées (Pour des critères de recherche dans Pubmed suivants : (Plasma OR serum), AND liquid chromatography AND lipidomic*[Title], 41 références bibliographiques sont répertoriées contre 223 en métabolomique) ((Ishikawa et al; 2013,Ishikawa et al; 2014)). Nous pouvons citer, les travaux de Quehenberger et coll. (Quehenberger et al; 2010) qui ont porté sur l'annotation du lipidome plasmatique humain. Leur démarche consistait en l'application de multiples méthodes de traitement d'échantillons servant à extraire spécifiquement les différentes classes de lipides suivies d'une analyse multiplateforme compréhensive des lipides plasmatiques. Il a de ce fait été possible d'identifier plus de 500 espèces lipidiques réparties dans 21 classes et sous classes et 6 familles de lipides. Bien que cette approche analytique soit exhaustive, sa mise en œuvre est laborieuse et aucun outil de traitement de données n'a été utilisé pour l'analyse des données générées. La démarche décrite correspondait donc à une analyse de lipides à large spectre et non pas à une analyse lipidomique. Néanmoins, la richesse des informations fournies quant aux espèces lipidiques identifiées et quantifiées en font un travail de référence (dont nous discuterons les résultats en PARTIE A.CHAPITRE II.3.4). Parallèlement, d'autres travaux ont mis l'accent sur la méthodologie analytique quant au profilage global des lipides plasmatiques avec l'utilisation d'une plateforme unique. Citons l'étude de Hu et coll. (Hu et al; 2008)

qui a développé et a validé (linéarité, LOD, répétabilité et efficacité d'extraction) une méthode RPLC-IT-FTMS de profilage lipidique de plasmas de souris. Les auteurs y fournissent une liste exhaustive des 160 molécules identifiées. Une approche similaire développée sur un Q-TOF a été décrite par Sandra et coll. (Sandra et al; 2010). Bien qu'annoncée comme étude « compréhensive », l'article décrivait la méthode et répertorie 65 espèces lipidiques listées dans le cadre de la validation de la méthode analytique. Toujours sur un couplage RPLC-QTOF, Castro-Perez et coll. ont identifié 284 espèces réparties dans 12 classes de lipides (une liste exhaustive est fournie en informations supplémentaires) grâce à une acquisition des données en mode MS^E (Castro-Perez et al; 2010). De plus, les auteurs ont eu recours à MetaboLynx XS[®], un logiciel de traitement des données assurant l'alignement des spectres acquis à faible et à haute énergies de collision. La validation de la méthode à l'aide d'étalons internes appartenant à 5 classes et sous classes de lipides PE, PG, PC, LPC et TG est présentée et une application biologique a été menée en ayant recours à des statistiques multivariées dans le cadre de l'investigation de l'arthrite. Bien que très élégante, l'approche MS^E ne peut être conduite sur des appareils autres que des Q-ToF qui à l'heure actuelle, sont les seuls spectromètres de masse à assurer une vitesse d'acquisition suffisante telle que des méthodes de fragmentation DIA semblables à celles réalisées en SL peuvent être envisagées. En alternative, Yamada et coll (Yamada et al; 2013a) ont décrit une méthode RPLC-Q-Exactive[®] utilisant le mode DDA-MS² (noté dd-tMS²) avec un changement rapide de polarité. Afin d'avoir une capacité d'identification satisfaisante, la résolution de l'appareil a été fixée à 70 000 à fwhm m/z 200 (et non pas à 140000 comme le permet l'appareil) avec l'utilisation d'une colonne de diamètre 4.6mm et de granulométrie de 3 μ m. Ce paramétrage permet d'avoir une largeur de pic répondant au besoin d'assurer un nombre satisfaisant de points par pic. Cette méthode a permis ainsi l'identification de plus de 400 espèces lipidiques extraites à partir de plasma de souris après interrogation du logiciel Lipid Search[®]. Ces mêmes auteurs ont décrit une approche similaire en SFC (SFC-dd-tMS² avec interrogation de la base de données Lipid Search[®]) ayant abouti à l'identification d'autant d'espèces dans du plasma de souris (Yamada et al; 2013b). Bien que la démarche tMS² couplée aux logiciels associés soit complète (allant de l'extraction des lipides jusqu'à leur identification), elle pose le problème de l'exhaustivité car ne peut être menée sans a priori ni sur l'intégralité des espèces détectées (liste prédéfinie d'ions parents à fragmenter).

L'analyse LC-HRMS en lipidomique souffre alors de la difficulté de développer une plateforme analytique qui permette à la fois de détecter et d'identifier les espèces lipidiques.

S'appuyant sur le développement et le succès précédent des approches lipidomiques SL à haute résolution, nous nous proposons de développer une approche globale de profilage des lipides plasmatiques basée sur (i) une méthode robuste RPLC-HRMS pour l'acquisition des données, (ii) le

développement d'une base de données interne d'espèces lipidiques de mammifères et (iii) d'outils bioinformatiques associés assurant l'identification et la validation des espèces détectées.

Dans un premier temps, nous avons développé puis validé à l'aide d'étalons internes représentatifs de 22 classes lipidiques une méthode de profilage global des lipides plasmatiques basée sur la RPLC-MS. Cette validation a parallèlement été menée sur une matrice biologique d'intérêt biologique majeur et peu concentrée en lipides à savoir des extraits lipidiques de LCR humains. Des outils bioinformatiques dédiés ont été mis en place afin d'en faire une approche intégrée, utilisable pour l'étude d'un grand nombre d'échantillons plasmatiques. Nous avons ensuite généré une banque de données interne reposant sur des mesures de masses précises et des données expérimentales obtenues sur différentes matrices biologiques. Avoir directement recours aux extraits de cerveau de rat et de plasmas humains nous a permis de rendre compte au mieux de la diversité structurale des espèces lors de la construction de la base de données palliant de ce fait au manque de composés de référence disponibles. Enfin, des logiciels de détection automatique des signaux et d'annotation ont été développés afin d'identifier automatiquement les espèces lipidiques. La démarche ainsi développée a été utilisée pour mener une étude comprehensive du lipidome plasmatique humain et a servi à identifier un très grand nombre d'espèces lipidiques présentes dans du plasma humain.

2. Matériel et méthodes

2.1. Réactifs et produits chimiques

L'eau (H₂O) et le méthanol (MeOH) de qualité LC-MS ont été achetés chez VWR. Le formate d'ammonium, le chloroforme, le 2-propanol (IPA) et l'acide formique pour les analyses LC-MS ont été obtenus auprès de Sigma-Aldrich (Sigma Chemical Co., St Louis, MO, USA). Tous les étalons internes (IS) utilisés dans cette étude (voir Table 12) ont été fournis par la société Avanti Polar Lipids, Inc (Alabaster, AL). Les mélanges d'étalons utilisés pour la calibration externe de l'instrument MS (Calmix-positif, pour le mode d'ionisation positif, constitué de caféine, de L-méthionyl - arginyl-phénylalanine acétate et d'Ultramark 1621, et un Calmix-négatif, pour le mode d'ionisation négatif, qui contenait en plus du même mélange Calmix-positif, du dodécylsulfate de sodium et le taurocholate de sodium) sont ceux préconisés par la société Thermo Fisher Scientific (Courtaboeuf, France).

Classe lipidique	Nom	Concentration de IIS dans le mélange (ng/mL)	Masse monoisotopique	Formule brute
Acides gras libres				
C14:0(OH)	3-Hydroxytetradecanoic acid	1000	244.204	C ₁₄ H ₂₈ O ₃
Sphingolipides				
d18:1/12:0 Cer	N-lauroyl-D- <i>erythro</i> -sphingosine	100	481.449	C ₃₀ H ₅₉ NO ₃
d18:1/12:0 GalCer	D-galactosyl-β-1,1' N-lauroyl-D- <i>erythro</i> -sphingosine	200	643.502	C ₃₆ H ₆₉ NO ₈
d18:1/12:0 lacCer	D-lactosyl-β-1,1' N-lauroyl-D- <i>erythro</i> -sphingosine	100	805.555	C ₄₂ H ₇₉ NO ₁₃
d18:1/12:0 Su	3-O-sulfo-D-galactosyl-β1-1'-N-lauroyl-D- <i>erythro</i> -sphingosine	100	740.486	C ₃₆ H ₇₂ N ₂ O ₁₁ S
d18:0/12:0 SM	N-lauroyl-D- <i>erythro</i> -sphinganylphosphorylcholine	200	648.521	C ₃₅ H ₇₃ N ₂ O ₆ P
18:0 CD3 GM3	N-omega-CD ₃ -Octadecanoyl monosialoganglioside GM ₃	1000	1183.767	C ₅₉ H ₁₀₅ D ₃ N ₂ O ₂₁
Glycerophospholipides				
11:0 LPC	1-undecanoyl-2-hydroxy- <i>sn</i> -glycero-3-phosphocholine	300	425.254	C ₁₉ H ₄₀ NO ₇ P
14:0 LPA	1-myristoyl-2-hydroxy- <i>sn</i> -glycero-3-phosphate	1000	404.194	C ₁₇ H ₃₅ O ₇ P
13:0 LPE	1-tridecanoyl- <i>sn</i> -glycero-3-phosphoethanolamine	100	411.239	C ₁₈ H ₃₈ NO ₇ P
13:0 LPS	1-tridecanoyl- <i>sn</i> -glycero-3-phospho-L-serine	1000	477.210	C ₁₉ H ₃₈ NO ₉ P
13:0 LPI	1-tridecanoyl-2-hydroxy- <i>sn</i> -glycero-3-phospho-(1'-myo-inositol)	100	547.276	C ₂₂ H ₄₆ NO ₁₂ P
17:1 LPI	1-(10Z-heptadecenoyl)-2-hydroxy- <i>sn</i> -glycero-3-phospho-(1'-myo-inositol)		601.323	
13:0 LPG	1-tridecanoyl- <i>sn</i> -glycero-3-phospho-(1'-rac-glycerol)	1000	464.460	C ₁₉ H ₃₉ O ₉ P
26:0 PC	1,2-ditridecanoyl- <i>sn</i> -glycero-3-phosphocholine	2000	649.468	C ₃₄ H ₆₈ NO ₈ P
24:0 PA	1,2-dilauroyl- <i>sn</i> -glycero-3-phosphate	5000	558.330	C ₂₇ H ₅₃ O ₈ P
24:0 PE	1,2-dilauroyl- <i>sn</i> -glycero-3-phosphoethanolamine	500	579.390	C ₂₉ H ₅₈ NO ₈ P
34:1 D31 PS	1-hexadecanoyl(d31)-2-(9Z-octadecenoyl)- <i>sn</i> -glycero-3-[phospho-L-serine]	100	814.697	C ₄₀ H ₄₅ D ₃₁ NO ₁₀ P
36:0 PI	1,2-distearoyl- <i>sn</i> -glycero-3-phosphoinositol	500	883.615	C ₄₅ H ₉₀ NO ₁₃ P
24:0 PG	1,2-dilauroyl- <i>sn</i> -glycero-3-phospho-(1'-rac-glycerol)	100	632.367	C ₃₀ H ₅₉ O ₁₀ P
Glycerolipides				
1,3 - 28:0 D5 DG	1,3(d5)-ditetradecanoyl-glycerol	500	517.475	C ₃₁ H ₅₅ D ₅ O ₅
60:1 D5 TG	1,3(d5)-dieicosanoyl-2-(11Z-eicosenoyl)-glycerol	1000	977.940	C ₆₃ H ₁₁₅ D ₅ O ₆
Esters de cholestérol				
22:0 Cholesteryl Ester	cholest-5-en-3β-yl docosanoate	2220	708.678	C ₄₉ H ₈₈ O ₂

Table 12. Mélange d'étalons internes utilisé : composition et concentrations correspondantes

2.2. Instrumentation LC-MS

Conditions de fractionnement sur colonne SPE de type LC(NH₂) : Cette opération s'effectue à partir d'un extrait lipidique total évaporé sous azote et repris dans son volume initial de chloroforme (Bodenec et al; 2000). Selon nos tests, le mélange à séparer doit contenir un maximum de 250 µL de plasma par colonne ou 25 mg de poids frais de cerveau de rat pour conserver une bonne qualité de séparation.

La séparation nécessite l'utilisation de colonnes amino-propyl (dites LC-(NH₂)) (Sigma-Aldrich (Supelco) Conditionnement 1 mL(Saint Quantin Fallavier, France) conditionnées avec 2mL d'hexane.

Fraction éluee	Solvants	Volume (mL)	Classes éluées
1	Diéthyléther	2	Diglycérides , triglycérides , cholestérol et cholestérol ester
2	Chloroforme/Méthanol (23/1)	1.6	Céramides libres , monoglycérides
3	Diisopropyléther / acide acétique	1.8	Acides gras libres, acides gras hydroxylés
4	Acétone / Méthanol (9/1.2)	2	Glycolipides neutres, bases sphingoïdes libres
5	Chloroforme / Méthanol (2/1)	2	Phospholipides neutres (PC, PE, LPC)
6	Méthanol contenant 0.2M acétate ammonium	2	Phospholipides acides (PI, PS), bases sphingoïdes phosphorylées, céramides phosphorylés, sulfogalactosylcéramides

Table 13. Programme d'élution du fractionnement sur colonne SPE (LC-(NH₂))

Conditions UHPLC : Les conditions de chromatographie en phase inverse utilisées ont été adaptées à partir des travaux de Hu et coll. (Hu et al; 2008). Le MeOH a été préféré à l'ACN du fait de sa plus faible polarité réduisant l'effet mémoire observé pour certaines espèces apolaires comme les TG. Les extraits lipidiques ont été séparés soit sur un système (i) ACCELA ou (ii) Dionex Ultimate 3000 UPLC (Thermo Scientific, Les Ulis, France) en utilisant une colonne Kinetex C8 150 x 2.1 mm, 2.6 µm (Phenomenex, Sydney, NSW, Australie) avec une température de four fixée à 60°C et un débit de 400 µL/min. Le volume d'injection a été fixé à 10 µL. La phase mobile A consistait en un mélange de H₂O/MeOH 60 :40 (v/v) et 0.1% d'acide formique et la phase mobile B en un mélange de IPA/MeOH 90 :10 (v/v) et de l'acide formique à 0.1 %. Du formate d'ammonium (10 mM) a été ajouté pour le mode positif. Le gradient est décrit dans la Table 14.

Temps (min)	Phase mobile A (%)	Phase mobile B (%)	Débit ($\mu\text{L}\cdot\text{mL}^{-1}$)
0.0	68	32	400
2.5	68	32	400
3.5	55	45	400
5.0	48	52	400
7.0	42	58	400
10.0	34	66	400
13.0	30	70	400
16.0	25	75	400
20.0	20	80	400
23.0	15	85	400
26.0	10	90	400
28.0	5	95	400
30.0	3	97	400
32.0	68	32	400
40.0	68	32	400

Table 14. Gradient d'éluion chromatographique en analyse RP-LCMS lipidomique

Conditions HRMS: 10 μL d'échantillon sont injectés dans la colonne chromatographique ou sont directement introduits dans la source électrospray d'un spectromètre de masse (i) LTQ-Orbitrap discovery ou (ii) Q-Exactive (Thermo Scientific, Les Ulis, France). Les analyses sont effectuées dans les deux modes d'ionisation. Les paramètres de la source ESI sont les suivants : la tension du cône a été réglée à 3,7 kV et -3,1 kV en mode d'ionisation positive et négative, respectivement. Le capillaire est chauffé à 360 °C et les débits de gaz principal et auxiliaire (azote) sont fixés à 15 et 10 (unités arbitraires), respectivement. Les spectres de masse ont été enregistrés entre les valeurs de m/z allant de 150 à 2000 à une résolution de masse de 30000 (FWHM à m/z 400) pour le LTQ-Orbitrap discovery ou de 140 000 (FWHM à m/z 200) pour le Q-Exactive. Une calibration externe de l'instrument a été réalisée avant l'analyse à l'aide du mélange préconisé par le constructeur.

2.3. Préparation des échantillons

L'extraction lipidique à partir d'échantillons plasmatiques, sériques et échantillons de LCR : L'extraction se fait dans des tubes en verre avec un mélange de chloroforme/méthanol dans les proportions 1:1 (v/v) dit CM 1 :1 (BLIGH & DYER, 1959) avec ajout d'un mélange d'EI (voir Table 12). Chaque tube est ensuite soumis aux ultrasons (3x10secondes) à l'aide d'une sonde à sonication pour obtenir un mélange homogène. Les tubes sont laissés sous agitation pendant 2 heures dans un récipient contenant de la glace. Pour obtenir une extraction optimale, les proportions en eau milliQ+fluide biologique/chloroforme/méthanol sont ajustées à 0.7:1:1. Il y a de ce fait ajout de 0.75 mL d'eau milliQ pour un 1 mL de biofluide à extraire. Les tubes sont ensuite agités à l'aide d'un vortex pendant 1 min. Après une centrifugation de 10 min à 3000 tr/min, deux phases sont obtenues

et séparées par un anneau de couleur blanche. La phase supérieure H₂O/MeOH (aqueuse) et la phase inférieure chloroforme (organique) sont récupérées, mélangées et évaporées sous azote. L'extrait lipidique total est repris dans un mélange CM 1 :1 puis conservé à -20°C jusqu'à analyse LC-HRMS ou HRMS. Avant analyse LC-MS, les extraits plasmatiques sont dilués au centième dans un solvant de reprise approprié (soit 65 :35 :5 IPA/MeOH/H₂O) alors que les extraits de LCR sont repris dans leur volume initial.

L'extraction lipidique à partir de cerveaux de rat : Elle s'effectue selon un protocole similaire mais dans les conditions décrites par Folch et coll. soit l'utilisation de chloroforme/méthanol 2 :1 (v/v) dit CM 2/1 (Folch et al; 1951). De plus, elle nécessite de peser l'échantillon à extraire et d'adapter les volumes de solvants correspondants. Pour exemple : 1 g de cerveau est extrait par 20 mL de CM 2 :1 avec également un lavage de Folch avec 4mL d'eau milliQ.

3. Résultats

3.1. Analyses par spectrométrie de masse d'extraits lipidiques

Au vu du manque de disponibilité d'étalons internes représentatifs de l'hétérogénéité des espèces lipidiques endogènes, nous avons procédé au développement de la méthode de profilage directement à partir d'extraits lipidiques totaux de différentes matrices biologiques. Ces dernières (à savoir plasma humain et cerveau de rat) rendent compte au mieux de cette diversité tant qualitative que quantitative des espèces lipidiques présentes. Cette étape de développement de méthodes d'analyses globales de lipides a été réalisée sur un système ACCELA-LTQ-Orbitrap selon le protocole décrit en Figure 45.

Il s'agira d'évaluer différentes stratégies de profilage lipidique afin d'être le plus exhaustif possible à savoir acquisition et comparaison de spectres HRMS obtenus en DIMS ou après séparation chromatographique avec ou sans fractionnement préalable soit 4 conditions différentes : DIMS-ELT, DIMS de fractions purifiées, RPLC-MS d'ELT et RPLC-MS de fractions purifiées.

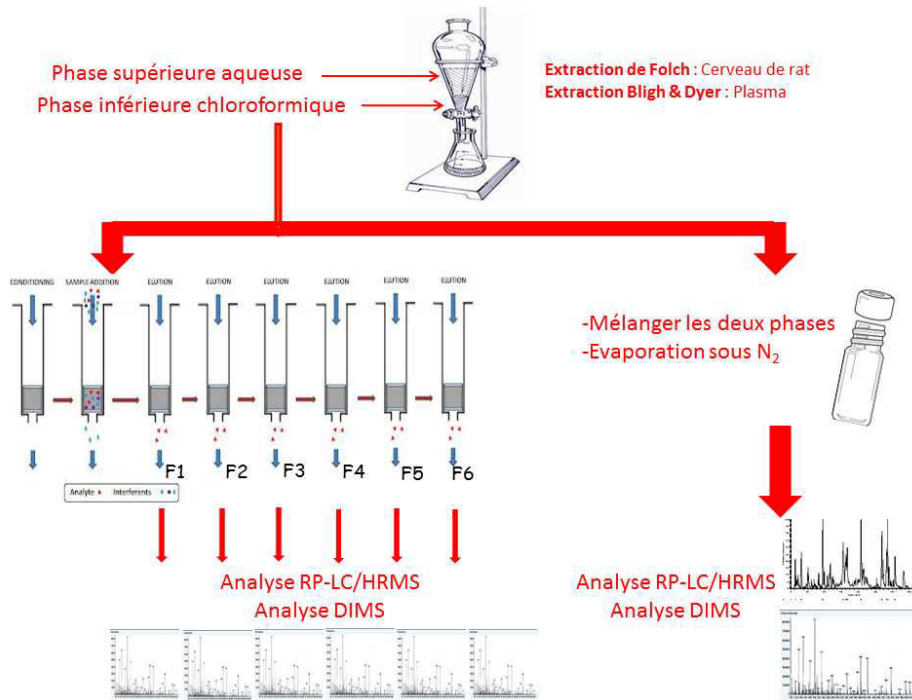


Figure 45. Les différentes approches analytiques évaluées pour le profilage des lipides

Nous avons réalisé 2 répliques d'extraction de lipides totaux depuis des échantillons de plasma humain (extraction de Bligh et Dyer) et de cerveau de rat (extraction de Folch). Un fractionnement sur colonne LC-(NH₂) a été effectué selon le protocole décrit précédemment (Bodennec et al; 2000) pour chaque ELT. Par la suite, chaque ELT et fractions correspondantes ont été analysés en DIMS et en LC-HRMS. Des analyses complémentaires de MS/MS ont été effectuées sur des molécules pré-sélectionnées.

3.1.1. Analyse des ELT par chromatographie liquide à polarité de phase inversée : RPC8-MS

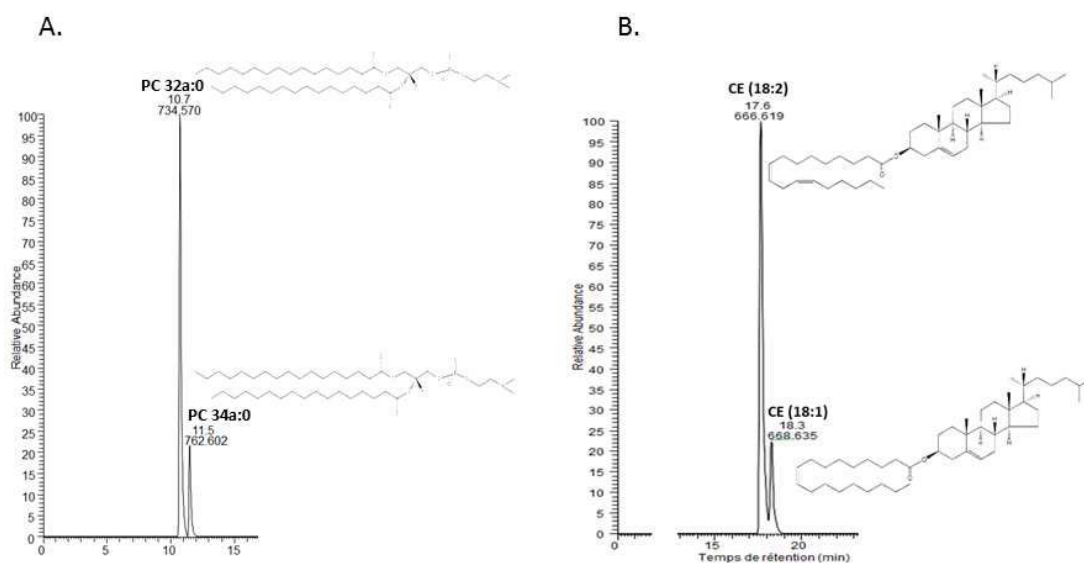


Figure 46. Détection des espèces lipidiques au sein des classes en RP(C8)-MS

Si l'on considère les espèces PC, il y a séparation avec retour à la ligne de base des espèces ne différant que par la longueur de leur chaîne carbonée de 2 atomes de carbone. Une résolution sans retour à la ligne de base est illustrée pour les CE qui diffèrent par une double liaison pour un même nombre d'atomes de carbone.

Grâce aux interactions hydrophobes qu'il assure, le greffage C8 permet de séparer des espèces lipidiques selon la longueur de chaîne et le nombre d'acyles gras, le nombre d'insaturations présentes (Figure 46.A et B), ou encore la régiosélectivité dans les chaînes d'acyles gras (Figure 68).

Cette séparation chromatographique par longueur de chaîne carbonée sous-tend une séparation des classes selon leur degré d'estérification. En effet, il y a d'abord élution des espèces comprenant une seule chaîne carbonée à savoir les formes lysoglycérophospholipides, puis celle des classes lipidiques présentant 2 chaînes carbonées qui peuvent être 2 acyles gras pour les glycérophospholipides ou une base sphingoïde et un acide gras pour les sphingolipides. La dernière zone d'élution est celle des espèces les plus hydrophobes qui contiennent 3 chaînes d'acides gras (à savoir les triglycérides). Les chromatogrammes d'extraits lipidiques totaux présentent de ce fait des zones d'élution par classes lipidiques (Figure 47) permettant de ce fait de réduire la suppression d'ionisation. Dans ce contexte, nous avons adapté la méthode RPLC-MS décrite par Hu et coll. (Hu et al; 2008) et présentée en PARTIE A.CHAPITRE II.2.2.

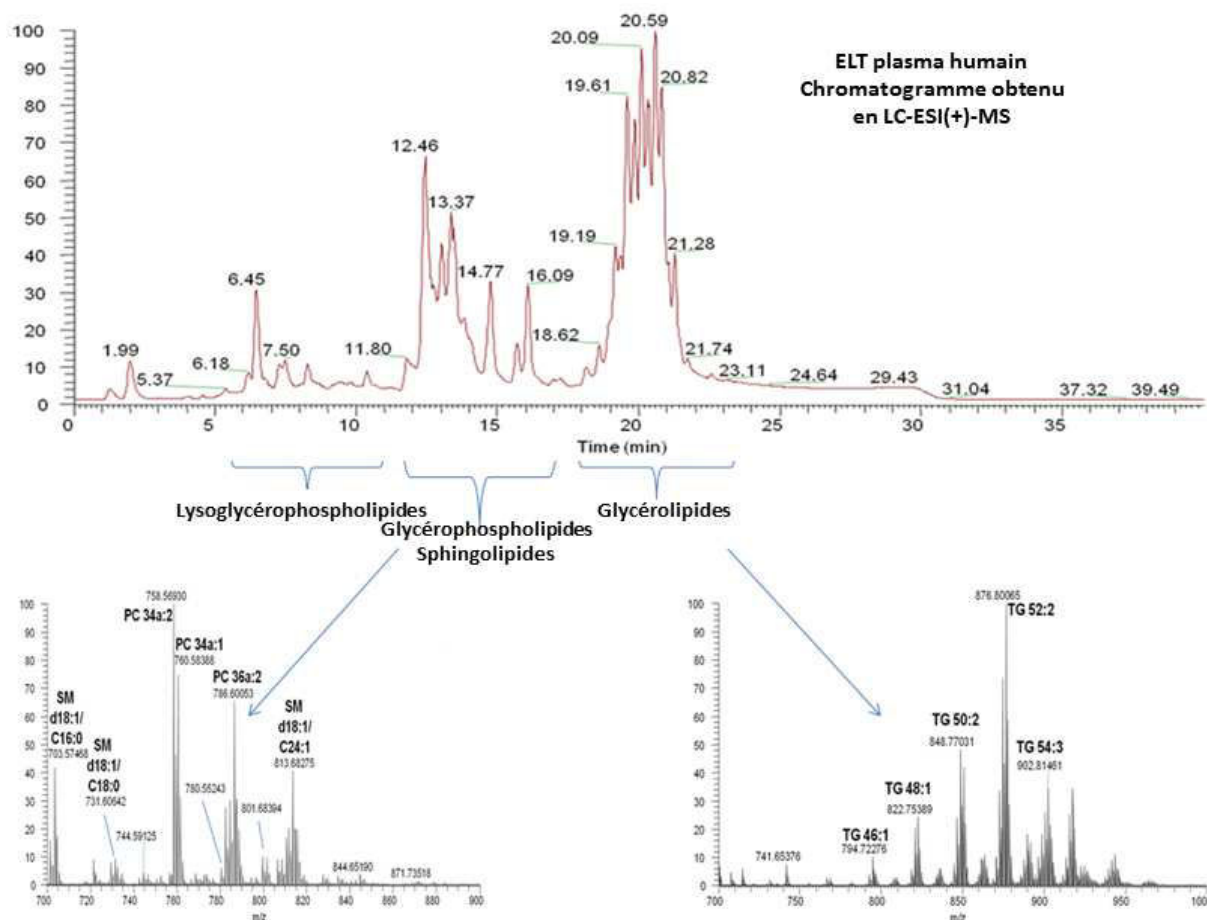


Figure 47. Chromatogramme d'un ELT de plasma détecté en mode d'ionisation positive

Intérêt de l'ajout d'acétate d'ammonium dans les phases mobiles en polarité d'ionisation positive

En présence d'ions ammoniums, une liaison hydrogène est formée entre un ion NH_4^+ et un lipide neutre. Les conditions de désolvatation choisies étant douces, tout en sachant que les esters en phase gazeuse sont moins basiques que les amines, les formes $[\text{M}+\text{NH}_4]^+$ sont majoritairement détectées. De plus, la rupture de ces ions entrainerait la formation de M (non détecté) et de l'ion NH_4^+ . Les lipides neutres (TG, DG, MG et CE) sont donc difficilement détectables sous leur forme protonée, et leur ionisation implique souvent la formation d'adduits (Bird et al; 2011, Hu et al; 2008). De plus, certaines classes de lipides telles les glycérophosphocholines forment facilement des adduits avec les métaux alcalins ((Yang et al; 2009)). Etant donné que ces ions-adduits dépendent de la matrice biologique étudiée du fait notamment des variations des concentrations et des types de sels présents (Na^+ , K^+), nous avons fait le choix d'ajouter à nos phases mobiles des sels d'ammonium (NH_4^+) dont l'affinité pour les lipides neutres a déjà été démontrée (Bird et al; 2011). Ceci nous permet de contrôler les formes moléculaires majoritaires détectées avec réduction d'un facteur 10 de l'intensité des adduits (ici Na^+) en faveur de la forme pseudomoléculaire $[(\text{M}+\text{H})]^+$ pour les espèces glycérophospholipides ou de la forme $[(\text{M}+\text{NH}_4)]^+$ des lipides neutres comme illustré en Figure 48 et Figure 49.

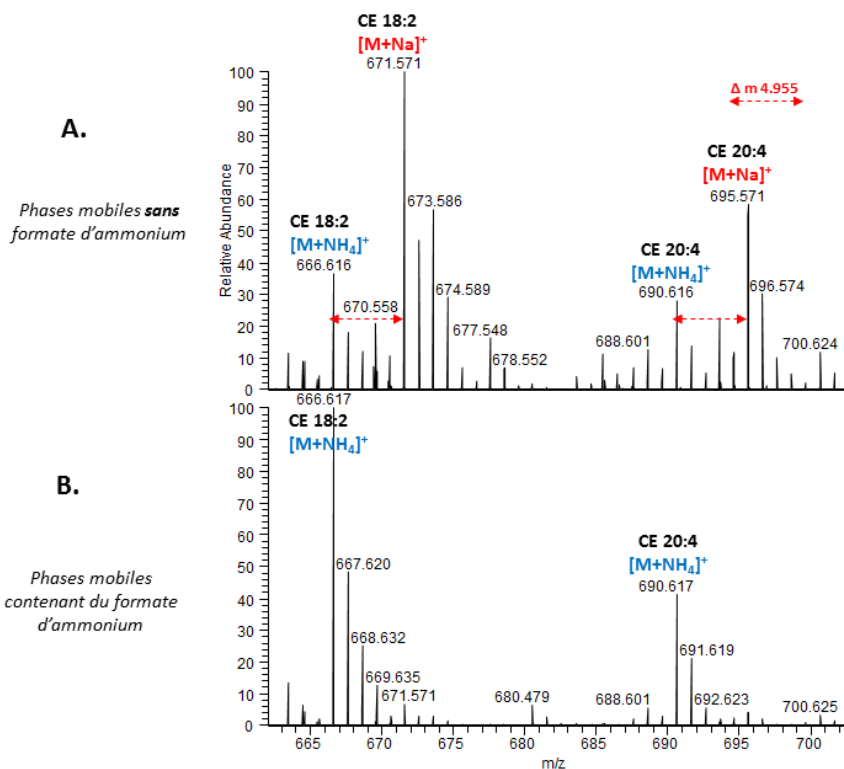


Figure 48. Analyse LC-MS dans les mêmes conditions LC avec une détection en ionisation positive avec ou sans ajout de formate d'ammonium à 10 mM : Détection des esters de cholestérol endogènes

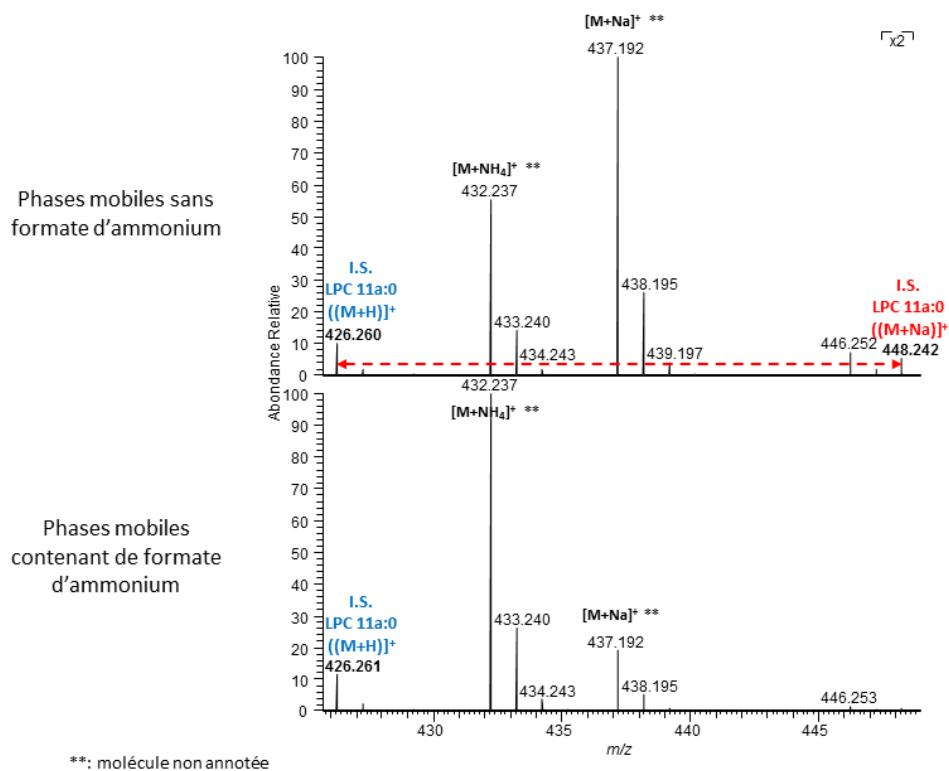


Figure 49. Analyse LC-MS d'ELT de plasma humain dans les mêmes conditions LC avec une détection en ionisation positive avec ou sans ajout de formate d'ammonium à 10 mM : Détection de l'étalon interne lysoglycérophosphocholine LPC 11a :0

La Figure 48 A a été réalisée après 30 analyses LC-MS d'un extrait plasmatique surchargé au mélange d'étalons internes (Table 12) sans ajout de formate d'ammonium au système chromatographique. On observe sur le spectre correspondant à la région des CE endogènes une multiplicité de pics correspondants à différentes formes moléculaires des mêmes espèces lipidiques, à savoir la forme adduit Na^+ et la forme adduit NH_4^+ . Cette dernière est détectée malgré le rinçage de colonne et prouve la forte affinité des ions NH_4^+ pour les lipides neutres (affinité supérieure à celle des sels naturellement présents tels Na^+ qui forment des liaisons électrostatiques). La Figure 48 B représente la même région du spectre à la même région d'élution chromatographique. Les formes Na^+ lorsqu'elles sont présentes sont minoritaires et représentent moins de 10% de l'intensité de l'espèce moléculaire adduit NH_4^+ correspondante. De même pour les espèces glycérophospholipides et lysoglycérophospholipides où il y a réduction d'un facteur 10 des adduits $[(\text{M}+\text{Na})]^+$ en faveur des formes pseudomoléculaires (Figure 49).

Grâce à la méthode LC-HRMS développée, il nous a été possible de détecter 23 classes et sous-classes de lipides appartenant à 5 familles de lipides au cours d'une même analyse RPLC-MS (voir Table 15) soit 598 espèces dans l'ELT de cerveau de rat et 700 espèces dans l'ELT de plasma humain (ceci sera présenté plus en détail en PARTIE A.CHAPITRE II.3.4).

	$[\text{M}+\text{H}]^+$	$[\text{M}+\text{NH}_4]^+$	$[\text{M}-\text{H}]^-$
GLYCEROPHOSPHOLIPIDES	LPC PC	LPA PA	LPE, LPI, LPS, LPG PE, PI, PS, PG
SPHINGOLIPIDES		SM Cer HexCer, DihexCer	Su
GLYCEROLIPIDES		MG, DG, TG	
STEROLS ET DERIVES		ChOL libre CE	
ACIDES GRAS LIBRES			FA

Table 15. Familles et classes lipidiques détectées par la méthode LC-MS développée

3.1.2. Analyse d'ELT en introduction directe HRMS

Comme décrit en introduction générale, les lipides présentent une grande diversité structurale avec notamment la présence de formes isomères dont la résolution dans le domaine spectral est impossible. De plus, des formes minoritaires isobares de molécules majoritaires peuvent être masquées en DIMS (Figure 50). Par exemple, les espèces de sphingolipides présentant une base sphingoïde de type sphingadiénine (d18:2) sont minoritaires. La détection de l'espèce Su (d18:2/C24:1) à partir d'un ELT de cerveau de rat analysé en DIMS est gênée par la détection de l'ion isobare dû à l'isotope ^{13}C de l'espèce PI 38a:4 majoritaire (m/z 886.553) en mode d'ionisation

négative. Ces deux espèces ainsi que l'espèce sulfatide majoritaire correspondante Su (d18 :1 /C24 :1) (m/z 886.607) sont parfaitement résolues en dimension RPLC. Ceci permet leurs éventuelles quantifications et expériences de caractérisation structurale.

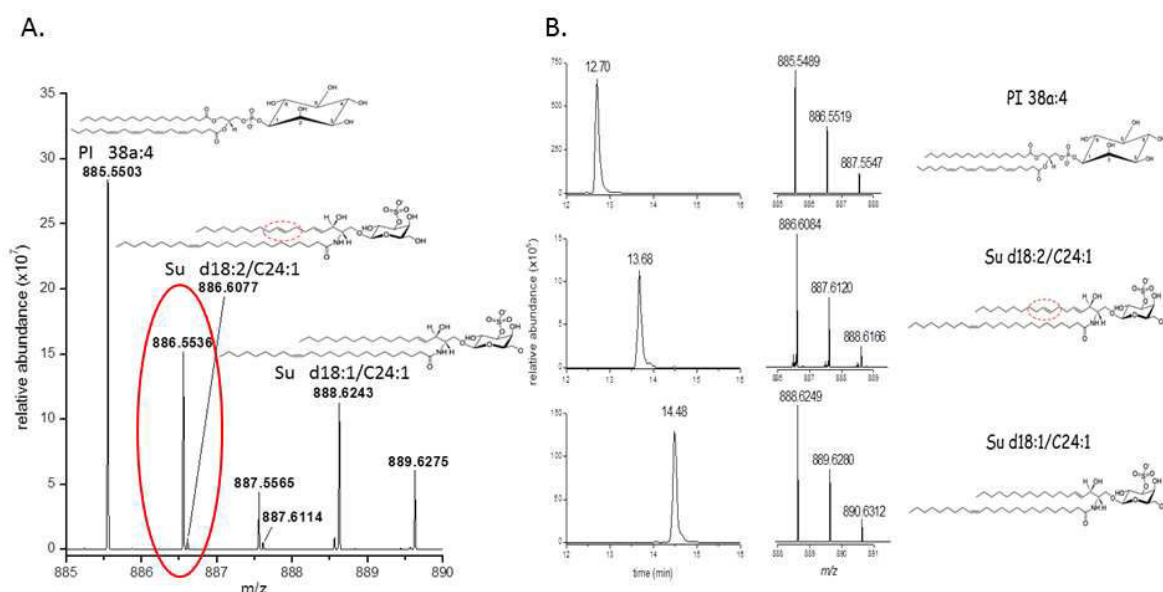


Figure 50. Intérêt de l'analyse LC-MS contrairement au DIMS dans la détection d'espèces minoritaires isobares

(A.) Le spectre DIMS montre la prédominance de l'ion m/z 886.553 (l'isotope de la forme PI plasmatique majoritaire PI 38a :4). Ce dernier entrave la détection mais surtout l'isolement pour analyse structurale de l'ion m/z 886.607 qui correspond à une forme minoritaire de Su portant une insaturation supplémentaire sur la base sphingoïde. (B.) Grâce à la RPLC, il y a séparation des trois espèces : PI 38a :4 , Su d18 :2/C24 :1 ainsi que du Su majoritaire Su d18 :1/C24 :1.

Une analyse comparative de spectres de masse obtenus en DIMS et en RPLC-MS d'un extrait lipidique total de cerveau de rat (Figure 51) nous a permis de conclure que l'analyse RP était plus concluante que le DIMS pour l'analyse d'extraits lipidiques complexes. En effet, si l'on considère les espèces détectées à partir d'un ELT de cerveau de rat, il nous a été possible d'en détecter 3 fois plus d'espèces lipidiques après séparation chromatographique qu'en analyse DIMS (selon le protocole d'annotation qui sera décrit ultérieurement et qui prend en compte notamment les valeurs de RIA).

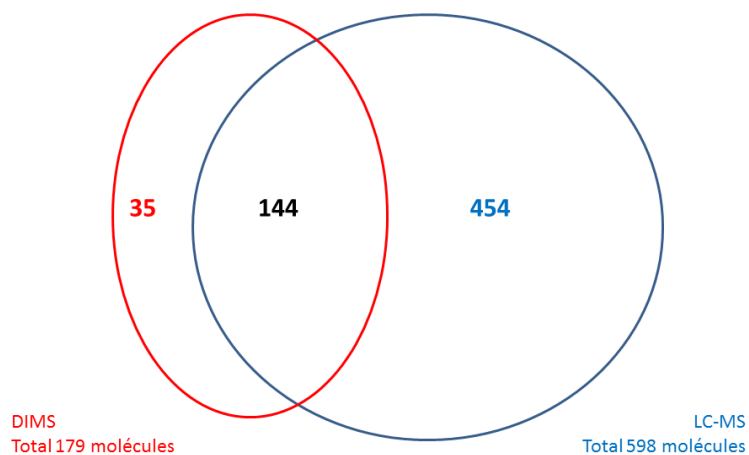


Figure 51. Comparaison du nombre de molécules détectées en analyse LC-MS et DIMS d'un même échantillon d'ELT de cerveau de rat

Cette tendance à la baisse en analyse DIMS est retrouvée dans chaque classe lipidique. Elle est plus particulièrement prononcée chez les espèces PC et TG du fait de l'existence de plusieurs isomères (Table 16).

Classes	DIMS	LC	Commun
HexCer et DiHexcer	2	59	18
PC et SM	9	250	83
TG, DG, MG et CE	15	115	26
Cer	4	8	11
LPC	3	17	6
PA	2	5	0

Table 16. Nombre d'espèces détectées par classes à partir d'un ELT de cerveau de rat en ionisation positive.

Les deux premières colonnes incluent les espèces détectées uniquement en DIMS ou en LC. La dernière colonne indique le nombre d'espèces communes détectées en DIMS et en LC.

3.1.3. Analyse de fractions (LC-NH₂) en introduction directe HRMS

A la lumière des résultats précédent, il nous a semblé évident de ne pas envisager de profilage lipidomique à partir d'un ELT en DIMS ni en shotgun tel que précédemment décrit (Ejsing et al; 2009, Ekroos et al; 2002, Han, Yang, & Gross, 2012, Schwudke et al; 2006). Nous avons de ce fait procédé à la purification des extraits biologiques avant leur analyse DIMS afin de minimiser le risque d'interférences en mesure de masse ou encore de limiter l'effet de suppression d'ionisation. Pour ce faire, nous nous sommes orientés vers une technique de SPE sur colonnes aminopropyl (LC-NH₂). Initialement développée pour la purification des sphingolipides (Bodennec et al; 2000), nous avons choisi d'exploiter les interactions dipôle-dipôle assurées par les cartouches LC(NH₂) pour réaliser un fractionnement des différentes classes lipidiques.

L'analyse par spectrométrie de masse de ces différentes fractions permet d'avoir une vue d'ensemble des espèces lipidiques présentes dans les différentes matrices biologiques et ce selon la partie polaire de leur structure.

Ci-dessous est présentée une sélection de spectres HRMS acquis après analyse DIMS des différentes fractions obtenues à partir d'un ELT de plasma humain et/ou de cerveau de rat.

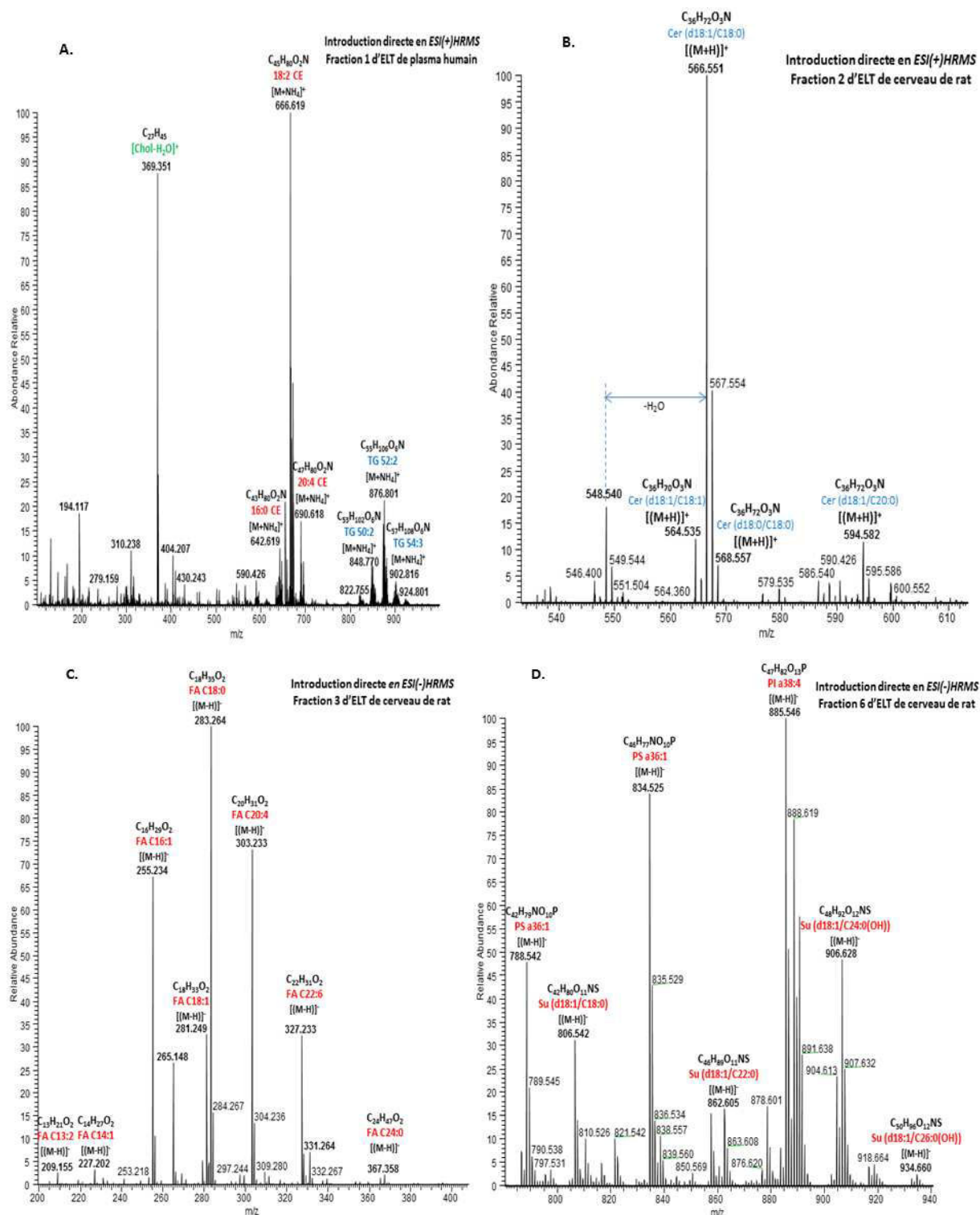


Figure 52. Spectres de masse obtenus après injection directe de différentes fractions lipidiques

(A.) Spectre de masse DIMS en mode d'ionisation positive de Fraction 1 obtenue à partir d'un ELT de plasma humain. On détecte des formes $[(M+NH_4)]^+$ des lipides neutres de classes TG et CE (B.) Spectre de masse DIMS en mode d'ionisation positive de Fraction 2 obtenue à partir d'un ELT de cerveau de rat. On détecte des espèces de céramides avec une prédominance de la sphingosine (d18 :1) comme base sphingoïde (C.) Spectre de masse DIMS en mode d'ionisation négative de Fraction 3 obtenue à partir d'un ELT de cerveau de rat. On détecte des acides gras libres (D.) Spectre de masse DIMS en mode d'ionisation négative de Fraction 6 obtenue à partir d'un ELT de cerveau de rat. On détecte des espèces de PS, PI et Su avec une prédominance de la sphingosine (d18 :1) comme base sphingoïde. Il y a existence des

formes hydroxylées. Seules les formes protonées perdent une molécule d'eau des suites du processus de la desolvatation. Les formes déprotonées ainsi que les adduits ammonium sont plus stables et on n'observe pas de perte d'une molécule d'eau.

De façon générale, le cerveau de rat a une teneur en lipides plus importante que celle du plasma humain, les fractions provenant de l'ELT de cerveau de rat sont plus abondantes en sphingolipides et glycérophospholipides notamment les espèces de céramides et de glycérophosphatidylsérines (voir figures ANNEXE.3). Elles illustrent de ce fait au mieux les résultats du fractionnement tel que démontré en Figure 52.B,C et D. Cependant, les glycérolipides neutres qui sont mieux détectés dans la fraction1 de plasma humain (présentant une concentration de 90.6mg/dl(Quehenberger et al; 2010)) du fait qu'ils soient les composants majeurs des lipoprotéines de type chylomicrons et VLDL circulants Figure 52.A.

L'analyse DIMS des différentes fractions a permis de valider les résultats décrits par Bodennec et coll. en HPTLC. En effet, les différentes classes de lipides sont éluées dans les fractions décrites (Bodennec et al; 2000) sauf pour les glycérophospholipides de classe PC et SM, censés être élués en fraction 5. Ces derniers sont également détectés en fraction 4 et leur ionisation en mode positif domine le courant ionique total (TIC) de cette fraction par rapport aux espèces HexCer attendues (voir figures ANNEXE.3). Néanmoins, les espèces de glycérophospholipides neutres (PC) et SM sont isolées en fraction 5.

Par ailleurs, la composition des fractions entre les différentes matrices biologiques diffère tant qualitativement que quantitativement tel que démontré en Figure 53. En effet, l'injection directe des fractions 5 de plasma humain et de cerveau de rat montre une répartition différente des espèces phospholipidiques majoritaires.

Ces observations soulignent l'intérêt de réaliser une étude de profilage sur différentes matrices biologiques permettant de vérifier au mieux l'exhaustivité de la méthode développée et assurer son applicabilité sans restriction dans différents contextes biologiques.

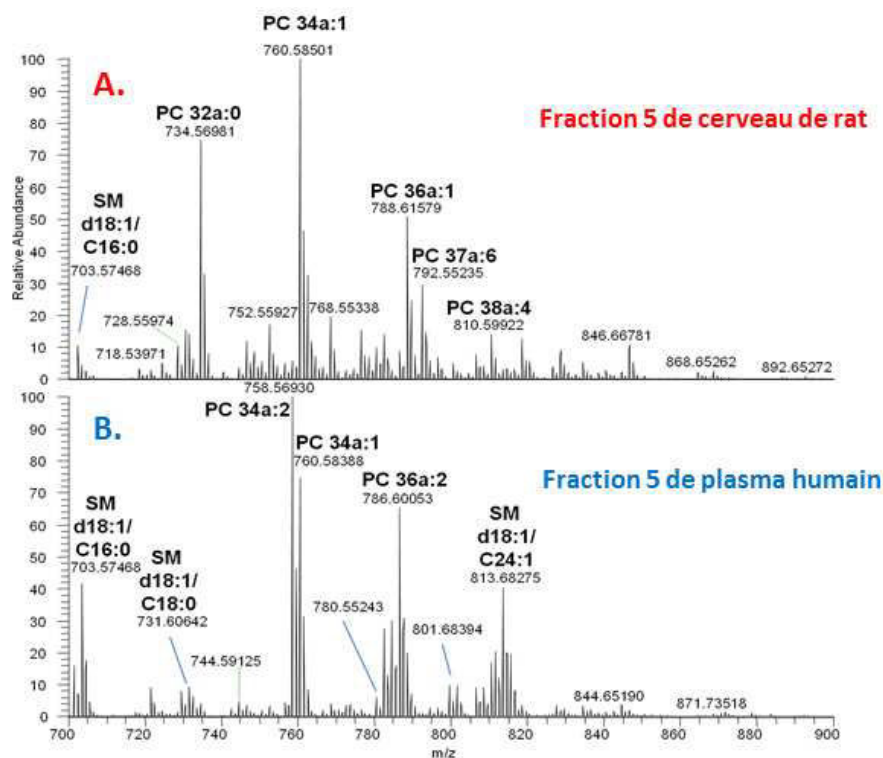


Figure 53. Diversité biologique : Matrices biologiques et abondances relatives des espèces lipidiques

Nous observons plus d'espèces SM majoritaire dans le plasma humain que dans le cerveau de rat. En outre, l'espèce PC prédominante dans le plasma est la PC 34a :2 alors que l'espèce la plus abondante dans le cerveau de rat est la PC 34a :1.

D'autre part, l'approche de fractionnement permet de réaliser une identification partielle des espèces lipidiques en distinguant les différentes classes. Par exemple, l'annotation d'un ion à m/z 826.6725 \pm 5 ppm détecté en polarité positive après analyse par introduction directe d'un extrait lipidique total de cerveau de rat (ELT) peut correspondre à deux molécules isobares : une espèce de phosphatidylcholine (PC 40p :2) ou une espèce d'hexosylcéramide hydroxylée (HexCer (d18 :1/C24 :1(OH))) dont les masses précises diffèrent par 0.008u. L'injection directe des fractions 4 et 5 obtenues à partir d'un ELT de cerveau de rat permet de conforter la structure de HexCer (d18 :1/C24 :1(OH)) et réfute celle de la PC 40p :2 et ce sans avoir recours à l'interprétation de spectres MS/MS. Cette dernière n'étant pas d'interprétation automatisée (voir Figure 54).

La procédure de fractionnement sur colonne LC-(NH₂), de mise en œuvre simple, constitue alors une aide à l'identification haut débit des espèces lipidiques notamment dans le contexte de constitution d'une base de données. Ceci sera discuté en PARTIE A.CHAPITRE II.3.2.4.

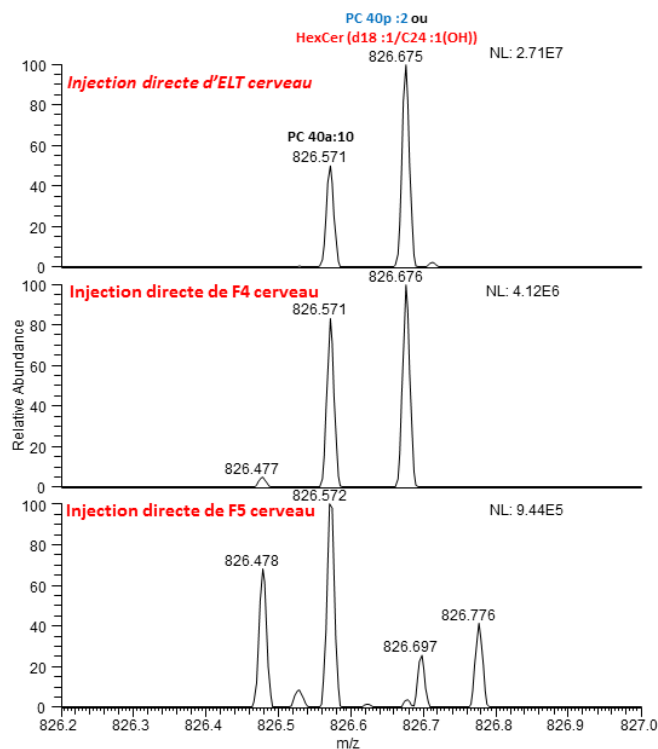


Figure 54 Fractionnement des classes lipidiques à partir d'un ELT de cerveau de rat : aide à l'identification structurale

Les spectres MS ci-dessus représentent un agrandissement sur une fenêtre de détection en masse allant de m/z 826.2 à m/z 827.0. Deux espèces lipidiques sont détectées sur le spectre DIMS d'ELT de cerveau de rat : PC 40a :10 et une espèce annotée PC ou HexCer. Cette dernière étant absente du spectre DIMS de la fraction 5 (élution des glycérophospholipides) tout en étant présente dans celui de la fraction 4 (élution des glycolipides), on déduit que l'ion m/z 826.575 correspond bien à HexCer (d18 :1/C24 :1(OH)). Notons la présence de la PC 40a :10 également dans le spectre DIMS de la fraction 4 tel que discuté précédemment, ainsi que la présence d'autres pics dans les fractions et qui ne sont pas détectés dans le spectre DIMS de l'ELT pour cause d'effet matrice.

3.1.4. Le couplage hors ligne SPE(LC-NH₂)/RPLC-MS

Les analyses NPLC et RPLC étant complémentaires, nous avons pensé les coupler afin d'obtenir le maximum de résolution dans la détection des espèces. Telle que présentée en PARTIE A.CHAPITRE II.3.1.2, la méthode SPE (LC-NH₂) permet le fractionnement grossier des différentes classes de lipides selon leurs groupements polaires (Figure 53) et donc de limiter l'effet de suppression d'ionisation rencontré en Shotgun. D'autre part, leur analyse en RPLC permet de mettre en évidence la purification des classes par une simple visualisation des chromatogrammes en comparaison aux chromatogrammes d'ELT. Le TIC du chromatogramme de l'ELT de cerveau de rat montre la forte abondance de phospholipides –qui est confirmée par le chromatogramme de la fraction 5 correspondante-(Figure 55.A et C) alors que le TIC du chromatogramme de l'ELT de plasma montre en plus une forte abondance de glycérolipides –tel que confirmée par l'analyse LC-MS de la fraction1

correspondante-(Figure 55.B et D). La Figure 56 montre la présence d'espèces PI, PS et Su dans les TIC des chromatogrammes obtenus en ionisation négative.

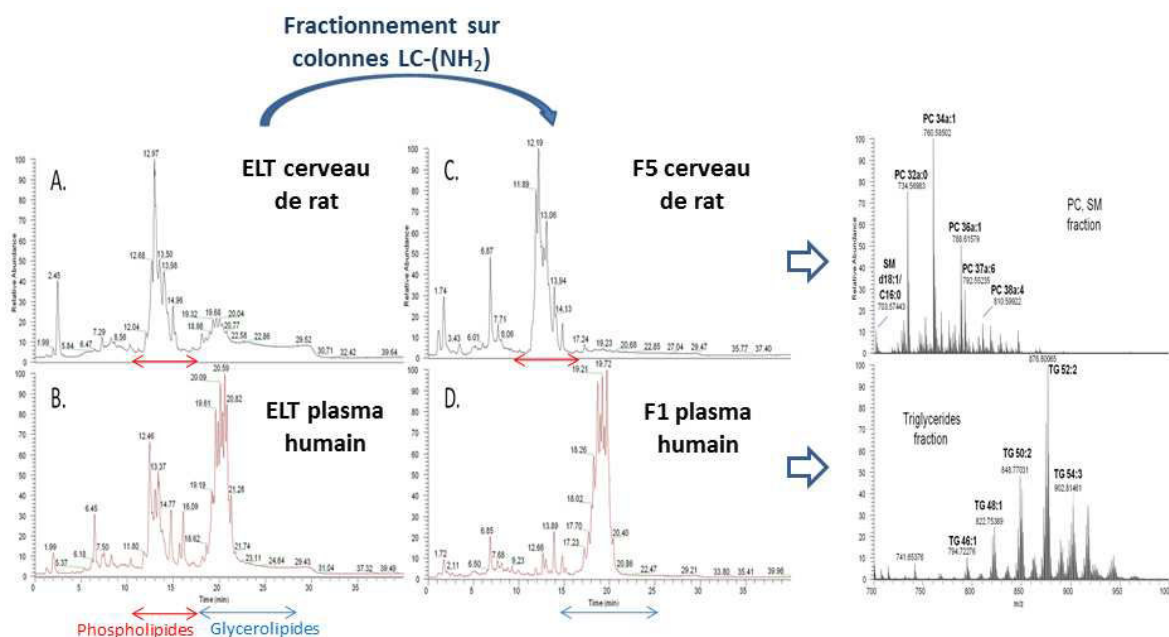


Figure 55. Chromatogrammes LC-(ESI+)MS d'ELT de cerveau de rat et de plasma avec les chromatogrammes et spectres de masse correspondants des fractions 5 et 1 respectivement .

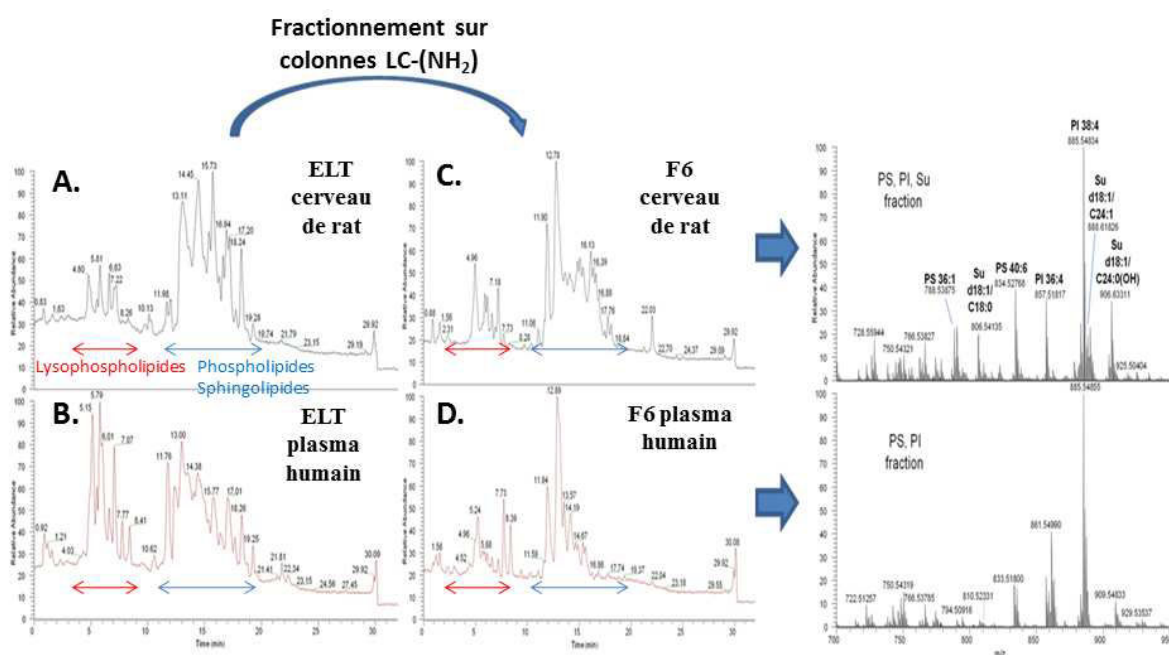


Figure 56. Chromatogrammes LC-(ESI-)MS d'ELT de cerveau de rat et de plasma avec les chromatogrammes et spectres de masse correspondants aux fractions 6 .

Nous avons également conforté par RPLC-MS les résultats obtenus en HPTLC (Bodennec et al; 2000) et en DIMS (PARTIE A.CHAPITRE II.3.1.2). En effet, la répartition des classes lipidiques respectait le profil d'éluion décrit (voir Figure 57, Figure 58 et ANNEXE.3).

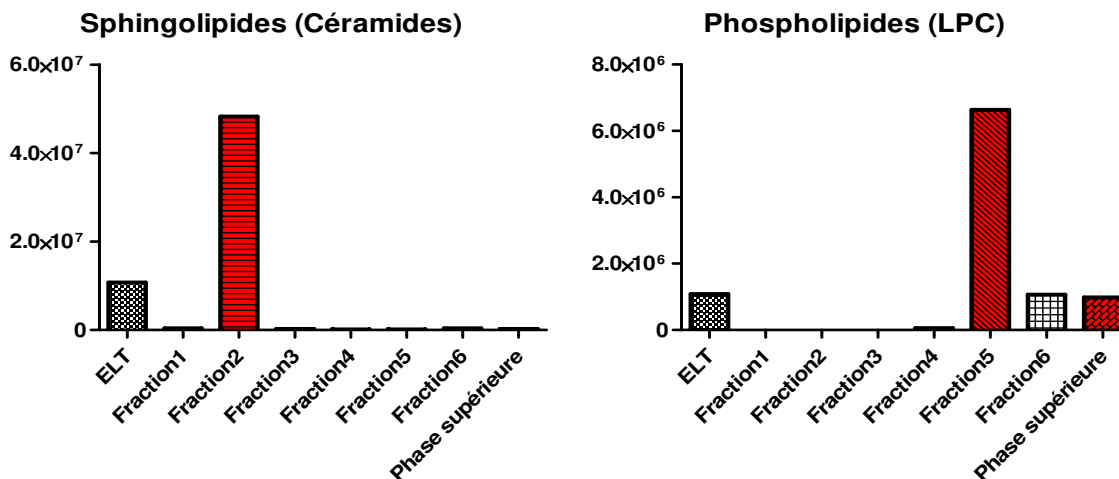


Figure 57. Fractionnement LC-(NH₂) des classes lipidiques depuis un ELT de cerveau de rat

Les barres représentent la somme totale des aires des espèces lipidiques d'une classe donnée (céramides et lysosphosphatidylcholines) détectées dans chaque fraction en analyse SPE-RPLC-MS réalisées à partir d'un ELT de cerveau de rat. En rouge la fraction d'éluion décrite selon le protocole de Bodennec et coll.

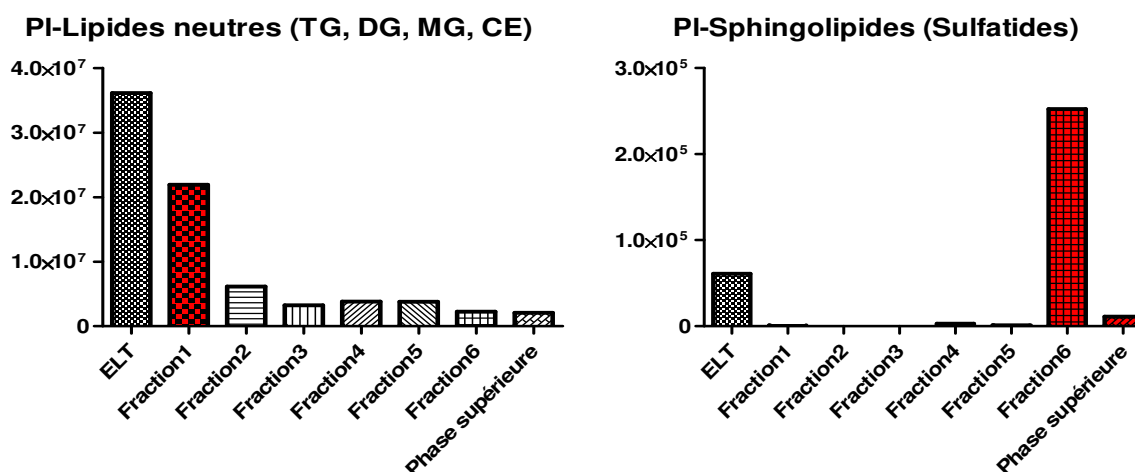


Figure 58. Fractionnement LC-(NH₂) des classes lipidiques depuis un ELT de plasma humain

Les barres représentent la somme totale des aires des espèces lipidiques (TG, DG, MG, CE et sulfatides) détectées dans chaque fraction en analyse SPE-RPLC-MS réalisées à partir d'un ELT de plasma humain. En rouge la fraction d'éluion décrite selon le protocole de Bodennec et coll.

Ces graphiques soulignent que les aires détectées en ELT sont inférieures à celles obtenues dans les fractions correspondantes, prouvant l'intérêt du fractionnement sur cartouches SPE à réduire l'effet matrice. Cependant, toutes les espèces détectées dans les fractions, le sont également dans les ELT. En plus de leur éluion dans les fractions attendues, certaines espèces sont également éluées dans des fractions adjacentes mais de façon moindre sauf pour les glycérophospholipides neutres et SM qui sont élués de façon équivalente dans les fractions 4 et 5 et les lipides neutres qui

sont retrouvés dans les chromatogrammes des différentes fractions. Concernant les PC et SM (voir ANNEXE.3), cette observation a été retrouvée en analyse DIMS des fractions 4 et 5, et montre une limitation de la méthode LC-NH₂ qui rappelons-le avait été développée dans le but de purifier les sphingolipides. Par contre, l'analyse DIMS des différentes fractions ne montre pas de surcharge de lipides neutres dans des fractions hors la fraction 1. La purification des glycérolipides neutres en fraction 1 est donc jugée bonne, et c'est plutôt leur analyse en RPLC-MS qui est mise en cause. Cette dernière entrainerait un effet mémoire des molécules hydrophobes notamment les TG (<1/10^{ème}) et qui serait à l'origine de leur détection résiduelle dans l'ensemble des fractions.

Notons que les cartouches SPE utilisées peuvent être couplées à d'autres types de cartouches telles les SPE-silice (Fauland et al; 2013) afin d'affiner le fractionnement, notamment la séparation des espèces PE et PC éluées en fraction 5.

3.1.5. Analyse structurale LC-MS/MS

Une analyse de lipides peut aussi se concevoir en mode DIMS/MS/MS ou LC-MS/MS. Elles permettent l'identification formelle des espèces lipidiques, cette dernière ne pouvant se baser que sur la mesure d'une masse précise. En effet, les lipides se fragmentent selon des règles de fragmentation reproductibles permettant de détecter ou de déduire la nature des têtes polaires et des chaînes d'acides gras suivant le mode d'ionisation utilisé. Si l'on reprend l'exemple décrit en Figure 50, l'analyse LC-MS/MS de l'ion m/z 885.548 détecté en ionisation négative nous permet d'affiner l'annotation initialement PI 38a :4 en une annotation plus développée PI (18 :0/20 :4) de par l'attribution des acides gras grâce à la détection des fragments lysophosphatidylionistol correspondants (Figure 59).

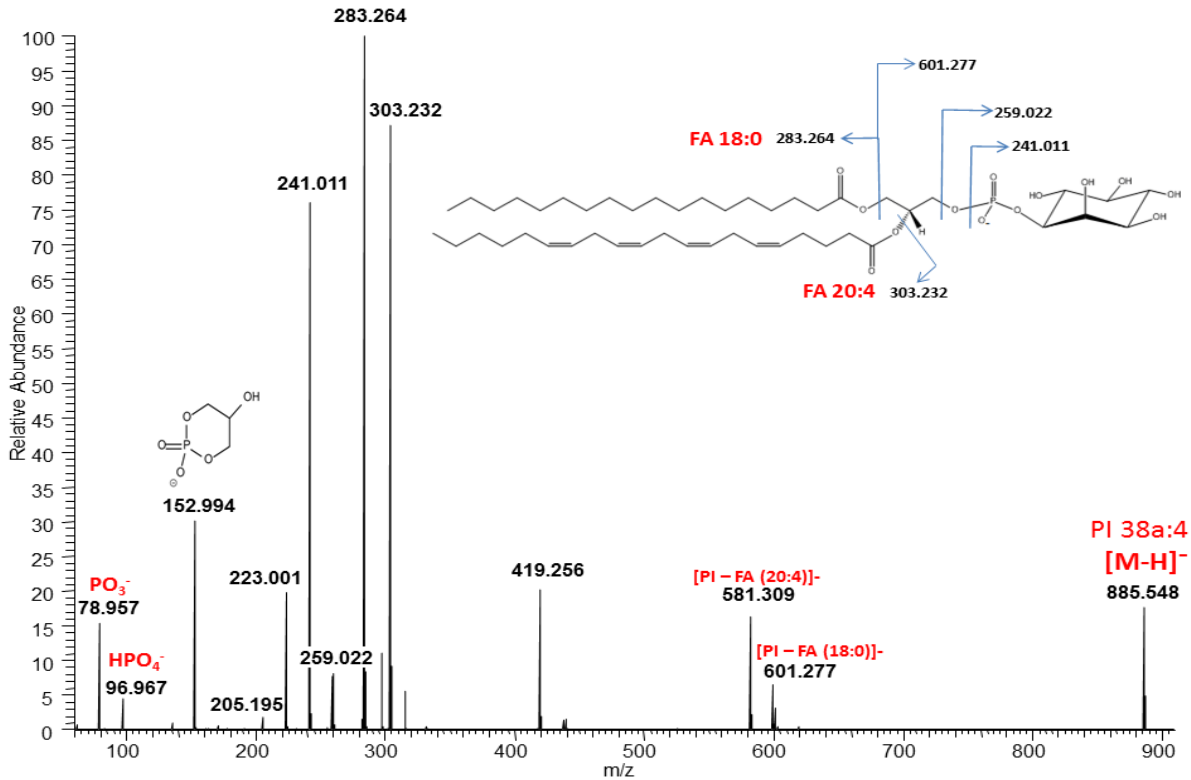


Figure 59. Spectre de fragmentation HCD de l'ion m/z 885.57 (LC-ESI(-)-MS/MS) annoté PI 38a :4 à partir d'un ELT de cerveau de rat

L'analyse LC-MS/MS permet également de distinguer les sphingolipides majoritaires présentant une base sphingoïde type sphingosine Su (d18 :1/C24 :2) de ceux minoritaires dont la base sphingoïde est la sphingadiénine Su (d18 :2/C24 :1).

Alors que la fragmentation en mode HCD permet de détecter l'ion fils m/z 96.96 correspondant à l'ion fragment HSO_4^- , mais aussi le m/z 241 qui correspond à la perte du fragment ((Galactose-Sulfate) - H_2O) aucune information sur les ions intermédiaires nous renseignant sur la nature de l'acide gras ou de la base sphingoïde n'est obtenue. Une analyse MS^2 en mode CID quant à elle ne permet pas la détection de l'ion produit HSO_4^- à cause du Low Mass Cut off « LMCO » dont souffrent les pièges ioniques. Ce dernier peut être obtenu par analyse MS^3 depuis le fragment galactose sulfate m/z 259.0. Par contre, cette fragmentation en mode CID permet de déduire la nature de l'acide gras et de la base sphingoïde à travers les pertes de neutre caractéristiques prouvant la présence d'une double insaturation au niveau de la base sphingoïde (NLS 236 pour la sphingadiénine et NLS 238 pour la sphingosine) tel que démontré en Figure 60.

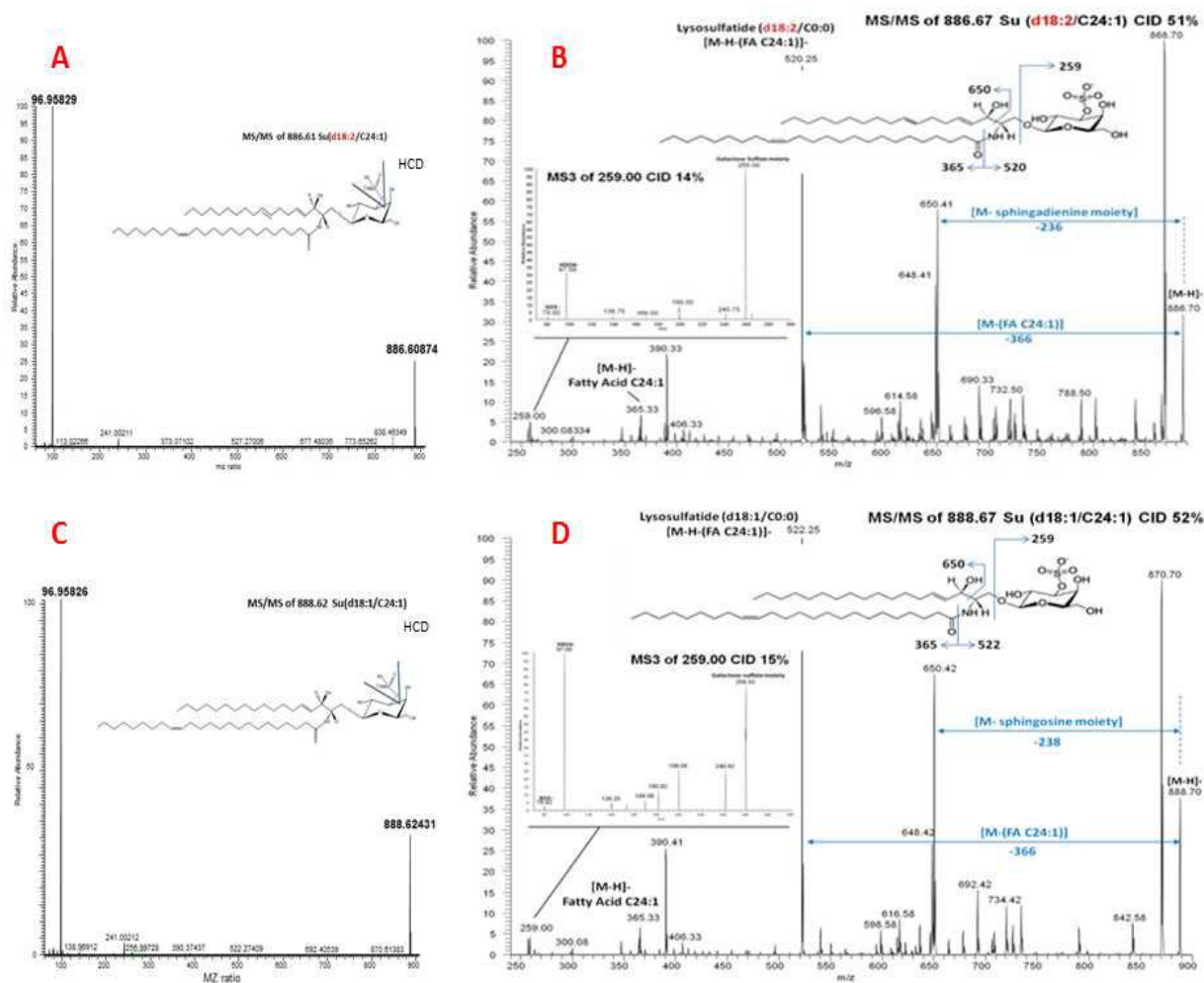


Figure 60. Discrimination par MS/MS d'espèces sulfatides à base sphingosine et à base sphingadiénine dans des ELT de cerveau de rat.

Analyse structurale de la molécule Su (d18 :2/C24 :1). En (A) Spectres MS/MS de l'ion m/z 886.608 en polarité d'ionisation négative correspondant obtenu en HCD et en (B) CID.

Analyse structurale de la molécule Su (d18 :1/C24 :1). En (C) Spectres MS/MS de l'ion m/z 888.701 en polarité d'ionisation négative correspondant obtenu en HCD et en (D) CID.

En conclusion, après évaluation des précédentes approches de profilage, l'analyse RPLC-MS d'ELT a été retenue comme méthode de profilage. En effet, l'analyse DIMS d'ELT ne permettait la détection que du tiers des espèces détectées après séparation chromatographique. Par ailleurs, bien que donnant des informations structurales importantes, le fractionnement n'a pas été retenu comme méthode globale de profilage car il n'apportait pas de plus-value en termes de détection d'espèces par rapport à la méthode RPLC-MS d'ELT. Le couplage SPE(LC-NH₂)-RPLC-MS n'a également pas été retenu comme méthode de profilage bien que des couplages SPE en NP-RPLC-MS ont déjà été décrits dans le cadre de purification et quantification ciblée de classes lipidiques (Sato et al; 2010). Ceci étant particulièrement dû à sa mise en œuvre lente (fractionnement+ 6 analyses RPLC).

Par contre, de par l'information structurale qu'il apporte, il sera utilisé comme un moyen d'assistance à l'identification haut débit dans le contexte de construction d'une base de données. Ceci sera présenté en PARTIE A.CHAPITRE II.3.2.4

3.2. Construction d'une base de données de lipides des mammifères

La méthode de profilage RPLC-MS d'ELT mise au point permet la détection de plusieurs milliers de signaux correspondant aux molécules lipidiques qui sont à identifier. L'enjeu d'assurer une identification lors d'une analyse lipidomique LC-MS reste majeur. En effet, alors que les approches SL largement documentées (Ejsing et al; 2009, Ekroos et al; 2002, Han, Yang, & Gross, 2012, Schwudke et al; 2006) sous-tendent une identification apportée par des expériences de fragmentations systématiques se faisant sans a priori, il est illusoire d'obtenir un spectre de fragmentation MS/MS pour chaque molécule détectée en LC-MS. L'identification en approche LC-MS implique de ce fait toujours une annotation sur la base d'une masse précise, parfois d'un temps de rétention pour les espèces dont le composé de référence est disponible, ou encore le cas échéant de spectres MS/MS acquis a posteriori sur une liste prédéfinie d'espèces. L'alternative à l'identification proposée est l'annotation pertinente des molécules détectées en LC-MS. Cette étape d'annotation est donc primordiale car elle est souvent la seule à assurer l'identification des espèces détectées ou encore à être à l'origine de la sélection des espèces à fragmenter.

Alors que la base de données LIPIDMAPS (Sud et al; 2007) répertorie l'ensemble des espèces lipidiques présentes dans différents milieux biologiques, l'annotation qu'elle propose, bien qu'exhaustive, reste souvent compliquée à assimiler et garde l'inconvénient de n'être basée que sur un seul critère d'identification à savoir la masse précise or ceci peut conduire à des annotations erronées (voir Figure 54). Il y a donc un réel besoin de développer une banque de données de référence assurant une annotation plus pertinente. Cependant, il y a disparité entre hétérogénéité et multiplicité des lipides endogènes et le nombre d'étalons internes commercialement disponibles. Ceci contraint souvent l'annotation à se restreindre à la correspondance entre la masse mesurée de la molécule à identifier à celle décrite dans une base de données.

3.2.1. Démarche méthodologique

Partant de ce constat, nous avons généré une base de données MS *in silico* d'espèces lipidiques théoriques qui englobe (i) des espèces répertoriées dans la base de données LIPID MAPS mais aussi (ii) des espèces calculées sur la base des motifs structuraux connus de lipides. Cette base de données renferme les informations quant aux formules brutes et masses monoisotopiques dans espèces lipidiques détectées et attribue des noms par groupe d'espèces d'isomères. La nomenclature adoptée sera développée en PARTIE A.CHAPITRE II.3.2.3.

Dans un deuxième temps, nous avons confronté la liste d'ions de référence générée aux données biologiques obtenues à partir des différentes fractions correspondant aux ELT de cerveau de rat et de plasma humain. Les fractions obtenues sur colonnes LC-(NH₂) étant utilisées en substitution des expériences de PIS ou NLS décrites en SL utilisant des TQ et servant de ce fait d'aide à l'identification. Ceci donne ainsi un degré de confiance supplémentaire aux annotations implémentées sans avoir recours aux expériences MS/MS. En outre, l'analyse RPLC-MS renseigne quant au temps de rétention de la molécule.

Enfin, afin de pallier aux interférences spectrales dues aux isotopologues, l'incrémentation des isotopes ¹³C et ²¹³C a été réalisée. La démarche est résumée en Figure 61 et chaque étape sera détaillée dans les paragraphes suivants.

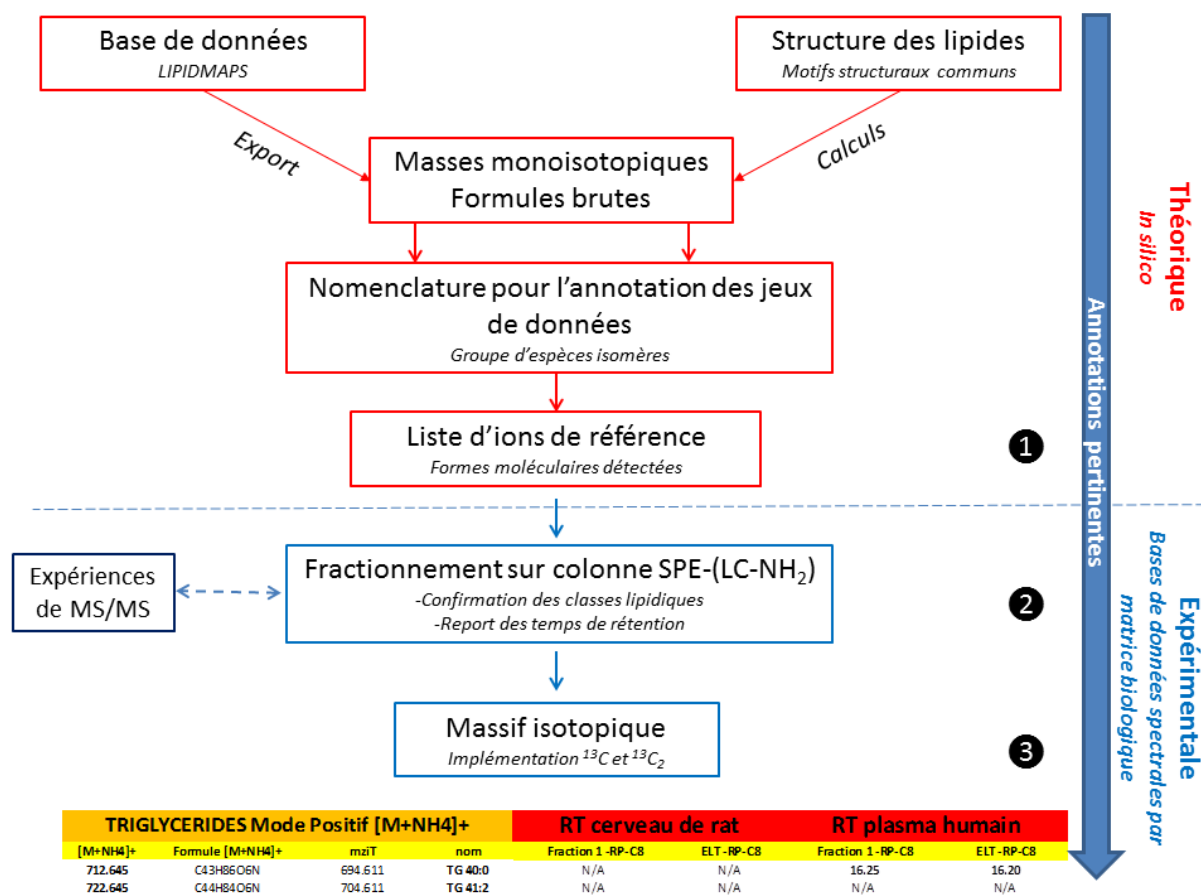


Figure 61. Démarche de construction de la base de données lipidomique

Les encadrés rouges indiquent les étapes réalisées dans une démarche *in silico* faisant appel notamment au calcul des espèces lipidiques théoriques. En bleu, les étapes réalisées à partir de données expérimentales (voir Figure 45) servant à construire les bases de données dites « biologiques ». Le degré de confiance en les annotations implémentées est augmenté avec l'ajout progressif d'informations structurales allant des formules brutes et masses monoisotopiques correspondantes aux têtes polaires associées, temps de rétention et isotopes.

3.2.2. Création d'une liste d'ions de référence depuis la base de données LIPID MAPS et calcul des masses monoisotopiques des espèces lipidiques théoriques

Nous avons choisi de nous intéresser aux lipides appartenant aux 5 familles majoritaires de lipides qui sont ubiquitaires et largement documentées dans la littérature (à savoir les glycérolipides, les glycérophospholipides, les sphingolipides, les acides gras, le cholestérol et ses esters).

Dans un premier temps, nous avons récupéré les masses exactes et les formules brutes des espèces qui leur sont affiliées et qui sont répertoriées dans la base de données de référence LIPID MAPS.

Nous avons ensuite complété cette liste à l'aide du calcul de formules brutes et de masses monoisotopiques d'espèces théoriques à l'aide des formules de calcul présentées en Table 17 pour les glycérophospholipides. Le principe de calcul repose sur la présence de motifs structuraux communs entre les espèces lipidiques. En effet, les lipides appartenant aux 5 familles peuvent être considérés comme une combinaison linéaire (i) d'une, deux ou trois chaînes aliphatiques et (ii) d'un groupe polaire rattachée à (iii) un squelette (Han, Yang, & Gross, 2012).

- (i) Les chaînes aliphatiques représentent les acyles gras estérifiés. Ces derniers se composent d'un nombre pair ou impair de carbone allant de 12 à 26. Des insaturations et des hydroxylations peuvent également être présentes. La distinction entre les conformations (*cis* et *trans*), ainsi que la régiosélectivité (*sn-1* et *sn-2*) n'ont pas été incluses dans la base de données spectrales développée. De plus, le type de liaison de la chaîne aliphatique *sn-1* et *sn-2* sur le glycérol définit également les sous-classes phosphatidyl (liaison acyl), ether (liaison alkyl) et plasmalogènes (liaison alkenyl) rencontrées pour les classes des PE et PC dans les échantillons analysés.
- (ii) Les têtes polaires considérées sont celles présentes dans les différentes classes de phospholipides à savoir le phosphate, la phosphocholine, la phosphoéthanolamine, phosphoglycérol, la phosphosérine et le phosphoinositol en plus des molécules de sucre (glucose, galactose, lactose) pour les sphingolipides.
- (iii) Les squelettes sont au nombre de trois. D'abord le « glycérol » qui est commun à toutes les classes de glycérolipides (TG, MG, AG) et glycérophospholipides (PE, LPE, PC, LPC, PA, LPA, PG, LPG, PI, LPI, PS, LPS). Les bases sphingoïdes « d18 :0, d18 :1, d18 :2, d20 :1 » présentes dans toutes les classes de sphingolipides (SM, Cer, GalCer, HexCer, SuGalCer et gangliosides). Et le cholestérol formant les esters de cholestérol.

Classe	Sous-classe	Squelette	Tête polaire	Chaines AG	Formule brute [M]	Formule détectée	
PC	diacyl PC	C ₃ H ₅ O ₃		C _m H _{2m-2n-2} O ₂	C _{m+8} H _{2m-2n+16} O ₈ PN	[(M+H)] ⁺ = [C _{m+8} H _{2m-2n+17} O ₈ PN] ⁺	
	alkenyl-acyl PC		C ₅ H ₁₃ O ₃ PN	C _m H _{2m-2n-2} O	C _{m+8} H _{2m-2n+16} O ₇ PN	[(M+H)] ⁺ = [C _{m+8} H _{2m-2n+17} O ₇ PN] ⁺	
	alkyl-acyl PC			C _m H _{2m-2n} O	C _{m+8} H _{2m-2n+18} O ₇ PN	[(M+H)] ⁺ = [C _{m+8} H _{2m-2n+19} O ₇ PN] ⁺	
PE	diacyl PE				C _m H _{2m-2n-2} O ₂	C _{m+5} H _{2m-2n+10} O ₈ PN	[(M-H)] ⁻ = [C _{m+5} H _{2m-2n+9} O ₈ PN] ⁻
	alkenyl-acyl PE		C ₂ H ₇ O ₃ PN		C _m H _{2m-2n-2} O	C _{m+5} H _{2m-2n+10} O ₇ PN	[(M-H)] ⁻ = [C _{m+5} H _{2m-2n+9} O ₇ PN] ⁻
	alkyl-acyl PE				C _m H _{2m-2n} O	C _{m+5} H _{2m-2n+12} O ₇ PN	[(M-H)] ⁻ = [C _{m+5} H _{2m-2n+11} O ₇ PN] ⁻
PS	diacyl PS			C ₃ H ₇ O ₅ PN	C _m H _{2m-2n-2} O ₂	C _{m+6} H _{2m-2n+10} O ₁₀ PN	[(M-H)] ⁻ = [C _{m+6} H _{2m-2n+9} O ₁₀ PN] ⁻
PG	diacyl PG			C ₃ H ₈ O ₅ P	C _m H _{2m-2n-2} O ₂	C _{m+6} H _{2m-2n+11} O ₁₀ P	[(M-H)] ⁻ = [C _{m+6} H _{2m-2n+10} O ₁₀ P] ⁻
PI	diacyl PI			C ₆ H ₁₂ O ₈ P	C _m H _{2m-2n-2} O ₂	C _{m+9} H _{2m-2n+15} O ₁₃ P	[(M-H)] ⁻ = [C _{m+9} H _{2m-2n+14} O ₁₃ P] ⁻
PA	diacyl PA		H ₂ O ₃ P	C _m H _{2m-2n-2} O ₂	C _{m+3} H _{2m-2n+5} O ₈ P	[(M+NH ₄) ⁺ = [C _{m+3} H _{2m-2n+9} O ₈ PN] ⁺	

Table 17. Calcul des formules brutes et des masses monoisotopiques des espèces glycérophospholipides (adaptée de (Han, Yang, & Gross, 2012))

Afin d'uniformiser les structures des espèces récupérées depuis la base de données LIPID MAPS de celles provenant des formules de calcul, les espèces répertoriées dans LIPIDMAPS mais ne répondant pas aux restrictions structurales imposées dans les formules de calcul n'ont pas été incluses dans la « liste d'ions de référence ». Ces critères de structure concernent le nombre d'atomes de carbone et de double liaisons contenus dans les chaînes d'acides gras, le type de liaison des acides gras aux squelettes, ainsi que la liste des têtes polaires envisagées. Pour illustration, si l'on considère la liste des espèces céramides répertoriées dans LIPID MAPS, cette dernière contient plusieurs composés ne répondant pas aux critères (i) de nombre de carbone ou de nombre d'insaturations -tel Cer(d18:1/2:0) (LMSP02010014) ou encore (ii) de nature de la base sphingoïde telle la forme acétylée « N-(acetyl)-sphing-4-enine- » ou encore (iii) des céramides présentant une composition de structure particulière tel le N-(30-(9Z,12Z-octadecadienoyloxy)-triacontanoyl)-sphing-4-enine (LMSP02040001). Ces structures sont donc retirées de la « liste d'ions de référence » depuis LIPID MAPS car leur présence ne peut être anticipée.

A l'aide de ce formalisme, une liste d'ions théoriques correspondant aux différentes espèces de CE, TG, DG, MG, PE, LPE, PC, LPC, PA, LPA, PG, LPG, PI, LPI, PS, LPS, SM, Cer, Su, HexCer, diHexcer, gangliosides et FFA putatives est créée. Conformément aux restrictions imposées, la base de données développée contient **2255** espèces théoriques monoisotopiques réparties dans les différentes classes et sous-classes de lipides (Table 18). Avec 14 espèces théoriques calculées, la classe des MG est la moins représentée. Cette faible représentativité est en accord avec celle de LIPIDMAPS qui ne répertorie également que 23 espèces de MG dont 8 espèces présentant 1 à 2 isomères. Les espèces PC et PE sont par contre les plus représentées, car en plus d'inclure des chaînes carbonées différentes, le type de liaison de ces dernières au squelette « glycérol » génère des sous-classes (*α*, *β* et *ε*).

Familles	Nombre d'espèces calculées
Acides gras libres	
FA	69
Sphingolipides	
Cer	52
HexCer	34
diHexCer	17
Su	45
SM	39
Glycérophospholipides	
LPC	95
LPA	18
LPE	74
LPS	38
LPI	39
LPG	36
PC	328
PA	47
PE	302
PS	208
PI	202
PG	208
Glycérolipides	
MG	14
DG	103
TG	257
Stérols	
Esters de cholestérol	30

Table 18. Composition de la base de données interne de lipides

3.2.3. Choix d'une dénomination des espèces lipidiques compatible avec les données LC-HRMS générées

La nomenclature LIPID MAPS renseigne quant à la nature et la position des acides gras estérifiés ainsi que les positions et les conformations des doubles liaisons proposant de ce fait plusieurs synonymes (ou identifiants) pour une même formule brute (Figure 27). Or, lors d'analyses MS (DIMS ou LC-MS), les isomères ne peuvent pas être distingués et seules les expériences complémentaires de fragmentation permettent cette discrimination. Ceci rend la mise en place de l'annotation telle que prévue par LIPID MAPS non pertinente et contraignante dans le cadre d'une analyse LC-MS tel que proposée en PARTIE A.CHAPITRE II.3. Nous avons de ce fait pris le parti de ne pas considérer la régiosélectivité pour la constitution de notre base de données LC-MS. La nomenclature adoptée permet d'affecter une tête polaire pour **chaque masse précise** et renseigne quant à **la somme** de la composition en atomes de carbone et degré total d'insaturation portés par les chaînes d'acide gras. De ce fait, un nom générique a été attribué par formule brute unique et non pas par espèce afin de regrouper les isomères sous un même identifiant. Cette démarche simplifie l'annotation et réduit le

nombre de propositions pour une même formule brute. Par exemple : les 2260 espèces de triglycérides répertoriées dans LIPIDMAPS correspondent à 257 groupes d'espèces de lipides dans notre base de données interne. Une démarche similaire a déjà été décrite pour l'analyse d'extraits lipidiques bactériens (Sartain et al; 2011) (voir Figure 62).

Pour les sphingolipides, la nomenclature précise la nature de la tête polaire (celle-ci pouvant être différente au sein de la même classe) et distingue de ce fait les carbones et les doubles liaisons portés par l'acide gras ou par la tête polaire (Figure 60). Le format est le suivant : **XX (dm':n'/Cm:n)** où :

- **XX** renseigne quant à la classe de sphingolipides considérée exemple « **Cer** » pour les céramides.
- **m'** et **m** correspondent aux nombres de carbone portés par la base sphingoïde (pouvant être 18 ou 20) et l'acide gras, respectivement.
- **n'** et **n** correspondent aux nombres d'insaturations portées par la base sphingoïde (pouvant être 0, 1 ou 2) et l'acide gras respectivement. Un groupement hydroxyle au niveau des chaînes d'acide gras peut également être présent, exemple **Cer (d18 :1/C20 :0(OH))**

Concernant les glycérolipides et glycérophospholipides, la nomenclature est la suivante : **XX m :n**. Les deux premiers digits correspondent à la classe lipidique (**PC** pour les glycérophosphatidylcholines, par exemple). Le (**m**) fait référence au nombre total de carbone portés par les chaînes d'acide gras (exclure les carbones portés par les squelettes et têtes polaires). Le (**n**) indique le nombre de doubles liaisons portées par les chaînes d'acide gras.

Une lettre supplémentaire entre le «**m**» et le «**:n**» est apposée pour définir les sous-classes de PE et PC. Cette lettre pouvant être un «a» si les deux liaisons en *sn-1* et *sn-2* sont de type acyl ; (p) si les liaisons d'acides gras sont de type alkenyl-acyl et (e) pour les liaisons de type alkyl-acyl, exemple **PC 36a :2**.

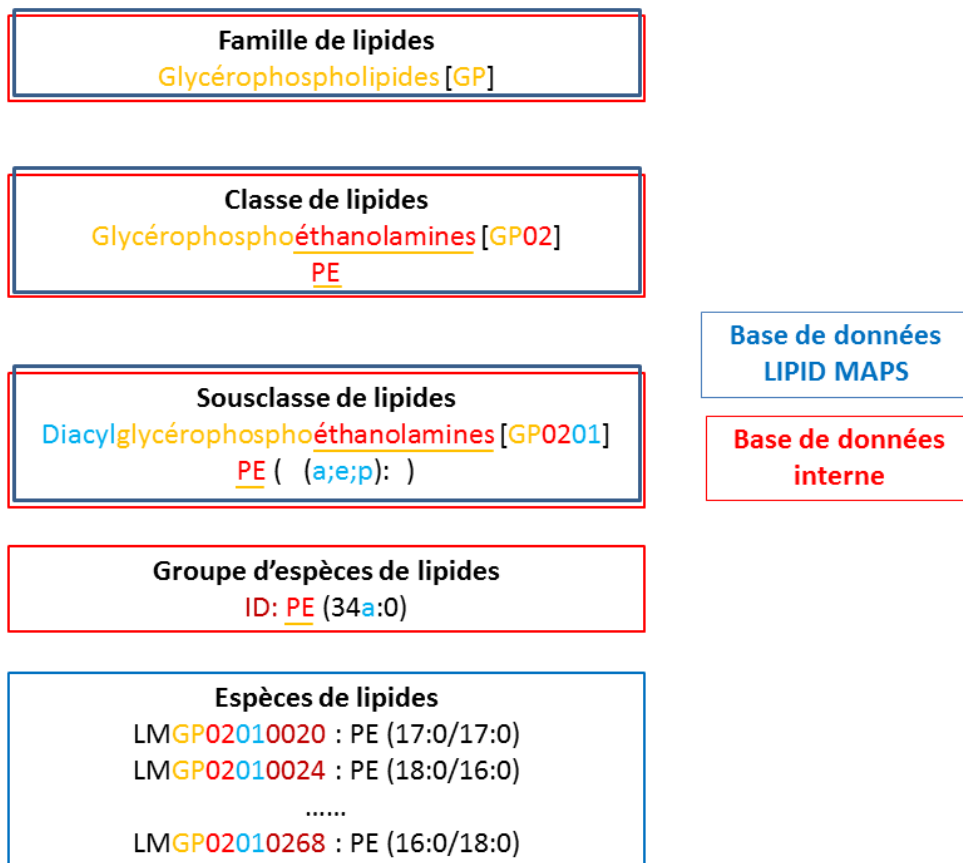


Figure 62. Organisation des nomenclatures selon LIPID MAPS et la base de données interne

Les encadrés bleus indiquent les informations répertoriées dans LIPID MAPS. Les encadrés rouges indiquent les informations répertoriées dans notre base de données interne. La nomenclature LIPID MAPS prévoit 2 digits pour chaque stade d'annotation à savoir la famille, la classe, la sous-classe et l'espèce de lipide. Une même couleur de police a été utilisée pour les digits et les éléments de l'arborescence dans la classification correspondants. 2 différences majeures existent entre la nomenclature LIPIDMAPS et celle de notre base de données interne : (i) La famille lipidique n'est pas formalisée dans la base de données interne mais plutôt déduite à partir de la classe et (ii) il y a un identifiant par groupes d'espèces de lipides isomères et non pas par isomère.

La régiosélectivité n'étant pas considérée, il n'existe qu'une seule annotation possible au sein d'une classe donnée pour chaque formule brute calculée. Alors que 36 annotations sont proposées pour une même masse précise à m/z (ESI-) 718.542, notre base de données n'en propose qu'une seule « PE 34a :0 » simplifiant de ce fait l'annotation

A ce stade-là, la base de données générée englobe la formule brute de la molécule non ionisée, la formule brute de la molécule ionisée selon le mode d'ionisation et sa forme de détection préférentielle (adduit NH_4^+ pour les lipides neutres et les PA), la masse monoisotopique de la forme détectée, le nom de la molécule ainsi qu'un numéro d'identifiant unique par formule brute.

3.2.4. Validation des annotations incluses dans la base de données interne : fractionnement des extraits lipidiques et prise en compte des temps de rétention de RPLC

Telle que conçue à ce stade, la base de données interne simplifie et élargit l'annotation telle que prévue par la base de données LIPID MAPS, gère les formes moléculaires détectées, mais reste basée sur un critère d'identification unique qui est la masse précise. Nous avons tiré parti de l'information structurale provenant des colonnes SPE-LC-(NH₂) en exploitant les données de fractionnement obtenues à partir des ELT de cerveau de rat et de plasma humain. Ces bases de données dédiées ne consignent que les annotations contrôlées par une analyse LC-MS de fractionnement effectuée sur des colonnes LC-(NH₂).

Les masses précédemment calculées pour les espèces théoriques sont donc recherchées dans les spectres SPE-LC(NH₂)/RPLC-MS de la fraction correspondant à la classe à laquelle appartient l'espèce et dans celui de l'ELT. Si le courant ionique extrait dans les 2 spectres présente la même forme de pic au même temps de rétention, l'annotation de la masse précise correspondante est maintenue dans la base de données « biologique » de la matrice. Son temps de rétention est relevé et sert alors à implémenter la base de données spectrales du milieu biologique.

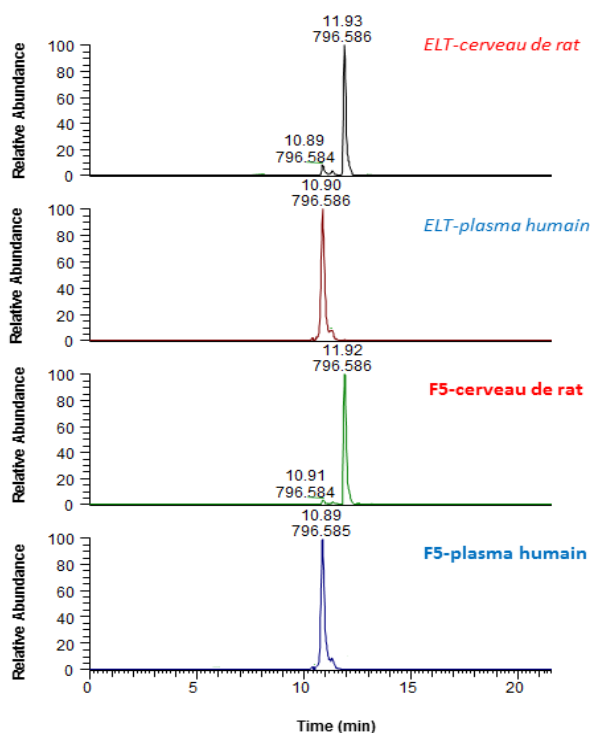


Figure 63. Implémentation de la base de données lipidomique par milieu biologique : Diversité biologique et matrices biologiques : Espèces de nature différente

L'espèce PC37a:4 (m/z 796.585 en mode d'ionisation positive) est présente dans le cerveau de rat et le plasma humain. L'annotation est validée par la masse précise et la confirmation de la tête polaire et du squelette de par la présence de l'ion dans les spectres LC-MS des fractions F5 correspondantes. Les temps de rétention observés entre les 2 matrices étant différents, on conclut à l'existence de deux espèces différentes selon la matrice biologique considérée (différences de nature des acides gras ou de leurs positions). Les différents temps de rétention sont relevés chacun dans la base de données biologique adéquate.

Une classe lipidique et un (des) temps de rétention sont alors attribués par annotation et par milieu biologique. Ces temps de rétention peuvent être différents selon le milieu biologique du fait qu'une même annotation puisse correspondre à différents isomères (Figure 63).

Le but ultime de ces banques de données «biologiques» sera d'évoluer, vers une annotation plus précise des espèces lipidiques au sein des matrices biologiques étudiées, grâce aux expériences de MS/MS qui fourniront des informations sur les différentes chaînes d'acides gras (nombre de carbones, insaturations, hydroxylation).

3.2.5. Inclusion des isotopes naturels

L'abondance naturelle de l'isotope ^{13}C du carbone est de 1.082% du ^{12}C . De ce fait, l'intensité des molécules possédant un isotope ^{13}C augmente avec le nombre d'atomes de carbone présents dans la molécule. Par exemple pour une molécule appartenant à la famille des glycérolipides tel qu'un TG à 56 carbones, l'abondance relative de ses isotopes $1[^{13}\text{C}]$ et $2[^{13}\text{C}]$ sont à environs 60.6% et 36.72% de leur isotopologue naturel ^{12}C . Cependant, la détection des isotopologues $2[^{13}\text{C}]$ peut interférer avec celle de l'isotopologue ^{12}C d'une molécule appartenant à la même classe mais présentant une insaturation en moins tel qu'illustré en Figure 64. Grâce à la RPLC, il est possible de distinguer les isotopologues $2[^{13}\text{C}]$ des ions monoisotopiques isobares présentant une insaturation en moins du fait que les insaturations font éluer la molécule plus tard au cours de l'analyse RPLC. Cette distinction étant impossible à discriminer en dimension HRMS seule. Ces deux molécules seraient donc confondues dans un spectre d'injection directe notamment pour les molécules à haute masse moléculaire.

De ce fait, les masses exactes des isotopologues $1[^{13}\text{C}]$ et $2[^{13}\text{C}]$ de chaque espèce théorique ont été intégrées à la base de données spectrales interne et ce en renseignant le temps de rétention correspondant à celui de leur ion monoisotopique. Leurs intensités serviront également à calculer des ratios isotopiques permettant de confirmer des annotations. Ceci sera présenté en PARTIE A.CHAPITRE II.3.3.2.2.

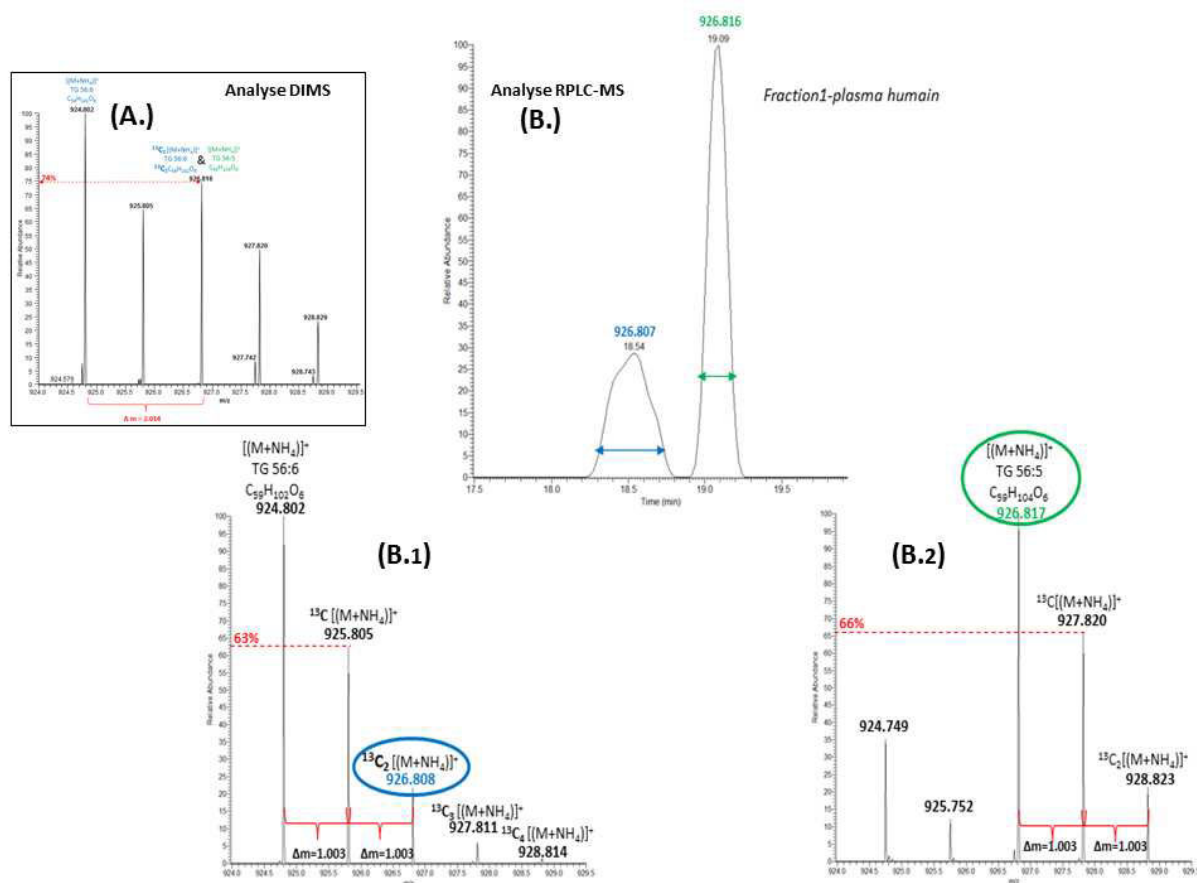


Figure 64. Interférences en détection en masse : ions monoisotopiques et isotopologues correspondants

(A) L'injection directe d'un ELT de cerveau de rat permet de détecter l'espèce TG 56 :6 sous sa forme d'adduit NH_4^+ à m/z 924.802. L'ion à m/z 926.816 peut correspondre à son isotope $^{13}\text{C}_2$ mais son intensité relative à 74% dépasse celle attendue (60.6%), on devine alors la présence d'une molécule isobare pouvant être le TG 56 :5.

(B) Le courant ionique extrait de l'ion m/z 926.815 en mode d'ionisation positive du spectre de la Fraction1 de plasma humain correspond à l'annotation TG 56 :5. Le chromatogramme montre deux pics résolus à 18.5mn et 19.1mn suggérant la présence de 2 isomères de TG 56 :5. (B₁) Le spectre de masse [m/z 924-929.5] du pic chromatographique à 18.5mn démontre que l'ion m/z 926.808 observé est l'isotologue 2 [^{13}C] de la molécule TG 56 :6. (B₂) le spectre de masse [924-929.5] du pic chromatographique à 19.1mn prouve que c'est l'ion monoisotopique de la molécule TG 56 :5. Un seul temps de rétention pour la molécule TG 56 :5 a été noté dans la base de données à savoir 19.1mn. Le courant ionique de l'ion m/z 924.802 en mode d'ionisation positive de la Fraction1 de plasma humain et qui correspond à l'annotation TG 56 :6 a confirmé la présence d'un seul pic à 18.5mn. En effet, la différence de masse due à une insaturation de différence ($\Delta m = m[\text{XX}(m:n)] - m[\text{XX}(m:n-1)] = 2m[\text{H}] = 2 \times 1.008 = 2.016\text{Th}$) est équivalente à la différence de masse entre une molécule et son isotopologues [^{13}C]₂ ($\Delta m = m[^{13}\text{C}]_2\text{M} - [^{12}\text{C}]\text{M} = 2 \times 1.0034 = 2.007\text{Th}$), il faut être prudent lors de l'annotation des pics sur la base de leur masse précise.

Pour conclure, la base de données ainsi développée contient 2255 groupes d'espèces lipidiques uniques, réparties dans 23 classes et sous-classes de lipides. Cette dernière a été implémentée à l'aide d'extraits plasmatiques –au vu des futures applications- mais aussi d'extraits de

cerveau de rat. En effet, du fait de leur forte teneur en lipides et de la diversité des espèces lipidiques qu'ils présentent, ces extraits ont permis d'assurer une couverture plus exhaustive des espèces et groupes d'espèces lipidiques. Cette matrice biologique ne sera pas utilisée dans la suite des travaux qui vous seront présentés.

Il est ainsi possible d'envisager une annotation intelligente. La Table 19 illustre l'intérêt de l'utilisation de la base de données interne par rapport à LIPID MAPS :

- 1- Les formes $[M+NH_4]^+$ des lipides neutres sont annotées alors que l'interrogation de la base de données LIPID MAPS ne propose pas d'annotations.
- 2- Les annotations sont simplifiées. Alors que la base de données LIPID MAPS propose une multitude d'annotations pour une même masse précise qui se déclinent sous des noms souvent compliqués à assimiler à une structure chimique, notre base de données propose un nom générique qui regroupe de façon exhaustive l'ensemble des combinaisons d'acides gras possibles (même au-delà de celles décrites dans LIPID MAPS).

mz	Rt	Ech-1	Ech...	Ech-n	Base de données interne	LIPID MAPS
964.838	18.7	9.18E+05	...	9.20E+05	364_(+)_TG 59:7 (+NH4)_C62H106O6	
694.648	15.7	4.53E+07	...	2.20E+07	546_(+)_20:2 CE(+NH4)_C47H80O2	
722.808	11.1	1.13E+06	...	1.42E+06		
						1,2-di-(6Z-octadecenoyl)-sn-glycero-3-phosphocholine @ 1-(6Z-octadecenoyl)-2-(6Z-octadecenoyl)-sn-glycero-3-phosphocholine @ 1-(7E-octadecenoyl)-2-(7E-octadecenoyl)-sn-glycero-3-phosphocholine @ 1,2-di-(7Z-octadecenoyl)-sn-glycero-3-phosphocholine @ 1-(8E-octadecenoyl)-2-(8E-octadecenoyl)-sn-glycero-3-phosphocholine @ 1,2-di-(8Z-octadecenoyl)-sn-glycero-3-phosphocholine @ 1,2-di-(9E-octadecenoyl)-sn-glycero-3-phosphocholine @ 1-(9E-octadecenoyl)-2-(9E-octadecenoyl)-sn-glycero-3-phosphocholine @ 2,3-di-(9Z-octadecenoyl)-sn-glycero-3-phosphocholine @ 2,3-di-(9Z-octadecenoyl)-sn-glycero-1-phosphocholine @ 1-(9Z-octadecenoyl)-2-(9Z-octadecenoyl)-sn-glycero-3-phosphocholine @ 1-(9Z,12Z-octadecadienoyl)-2-octadecanoyl-sn-glycero-3-phosphocholine @ 1-(9Z,12Z-octadecadienoyl)-2-octadecanoyl-sn-glycero-3-phosphocholine @ 1,2-di-(4Z-octadecenoyl)-sn-glycero-3-phosphocholine.....
786.602	9.4	2.94E+09	...	3.72E+09	844_(+)_PC 36a:2_C44H84NO8P	
788.616	7.9	4.54E+05	...	5.11E+05		
788.616	9.8	4.34E+08	...	5.65E+08		
811.616	9.7	5.00E+08	...	6.54E+08		
						1-octadecanoyl-2-(8Z,14Z-eicosadienoyl)-sn-glycero-3-phosphocholine @ 1-(11Z-octadecenoyl)-2-(13Z-eicosenoyl)-sn-glycero-3-phosphocholine / @ 1-(9Z,12Z-octadecadienoyl)-2-eicosanoyl-sn-glycero-3-phosphocholine @ 1-(10Z-nonadecenoyl)-2-(10Z-nonadecenoyl)-sn-glycero-3-phosphocholine / @ 1-eicosanoyl-2-(9Z,12Z-octadecadienoyl)-sn-glycero-3-phosphocholine @ 1-eicosanoyl-2-(9Z,12Z-octadecadienoyl)-sn-glycero-3-phosphocholine
814.631	10.0	4.38E+07	...	5.81E+07	885_(+)_PC 38a:2_C46H88NO8P	

Table 19. Annotation d'un jeu de données lipidomiques à l'aide de la base de données LIPID MAPS et la base de données interne

La double annotation d'un jeu de données plasmatiques acquis en mode d'ionisation positive à l'aide de la base de données LIPID MAPS et la base de données interne s'effectue sur la base de la masse précise. L'annotation selon la base de données interne se formalise comme suit : Numéro d'identifiant_(mode d'ionisation signe + ou -)_Nom de la molécule (PARTIE A.CHAPITRE II.3.2.3)_Formule brute de la molécule non ionisée.

3.3. Développement d'une approche lipidomique

Une approche lipidomique au sens « omique » du terme est une approche multidisciplinaire, intégrée, faisant appel à la chimie analytique pour la détection et l'identification des espèces lipidiques et à des outils bioinformatiques de traitement de données en vue de leur interprétation biologique.

Après avoir sélectionné la méthode de profilage à partir d'ELT de cerveau de rat et de plasma humain (PARTIE A.CHAPITRE II.3), nous allons maintenant aborder sa validation depuis l'extraction des lipides à partir d'échantillons plasmatiques humains et de LCR humains jusqu'à leur analyse LC-HRMS. Nous présenterons ensuite la démarche bioinformatique mise au point pour l'analyse des jeux de données lipidomiques à savoir le logiciel de détection automatique des signaux et les outils d'annotations dédiés à la base de données précédemment développée en PARTIE A.CHAPITRE II.3.2.

3.3.1. Validation de la méthode de profilage lipidomique LC-MS d'extraits plasmatiques et de LCR humains

Comme présenté précédemment, la méthode de profilage lipidomique sélectionnée est l'analyse RPLC-MS d'ELTs obtenus par une extraction de Bligh & Dyer pour les biofluides biologiques. Notons que la méthode a également été appliquée aux échantillons de LCR humains car ces échantillons représentent un intérêt biologique certain et un enjeu analytique majeur du fait de leur faible teneur en lipides. La validation analytique a été réalisée sur un système LC-HRMS plus sensible (DIONEX-Q-EXACTIVE) et ce selon le protocole décrit en Figure 65. Il s'agissait d'évaluer le rendement d'extraction à savoir son efficacité et sa répétabilité, ainsi que la linéarité de la méthode sur une gamme de concentrations allant de 10 ng.mL^{-1} à 1000 ng.mL^{-1} .



Figure 65. Validation de la méthode de profilage depuis l'extraction à l'analyse LC-HRMS

3.3.1.1. Evaluation du rendement de l'extraction :

Afin d'évaluer la performance de l'extraction Bligh & Dyer modifiée choisie, des rendements d'extraction ont été calculés à partir d'échantillons de LCR et de plasma humains (n = 5) surchargés avec 22 étalons internes non endogènes (IS). Ces IS sont représentatifs des cinq grandes familles de lipides (acides gras libres, glycérophospholipides, glycérolipides, sphingolipides et esters de cholestérol (voir Table 12, voir Figure 66).

Après analyse LC-MS des extraits, les aires chromatographiques correspondant à chacun des IS ont été comparées à celles du même mélange d'IS ajouté après extraction (n=3). En utilisant la méthode décrite, les recouvrements moyens pour les cinq familles de lipides étaient toujours au-dessus de 80% pour les échantillons LCR. Ces performances étaient, de façon générale, inférieures pour le plasma (une moyenne de 57% pour les glycérophospholipides plasmatiques). Ceci doit être dû à la plus grande teneur en lipides du plasma et qui pourrait saturer leur partage (Quehenberger et al; 2010).

La moyenne de recouvrement des glycérophospholipides plasmatiques est la plus faible (voir Table 20). Les sous-classes de lysoglycérophospholipides sont celles qui sont à l'origine de cette baisse d'efficacité d'extraction moyenne (54% en moyenne contrairement à 70% pour les formes non lyso). Ceci a précédemment été démontré en analyse d'extrait lipidique de levure (Ejsing et al; 2009). Ces dernières se partagent sûrement majoritairement en phase aqueuse supérieure et ne sont de ce fait pas intégralement récupérées. D'autre part, les classes lipidiques pour lesquelles l'efficacité d'extraction est la plus faible dans les deux biofluides biologiques sont les PA et LPA pour lesquelles l'extraction Bligh & Dyer modifiée utilisée n'est pas très adaptée du fait de leur caractère polaire (Sato et al; 2010).

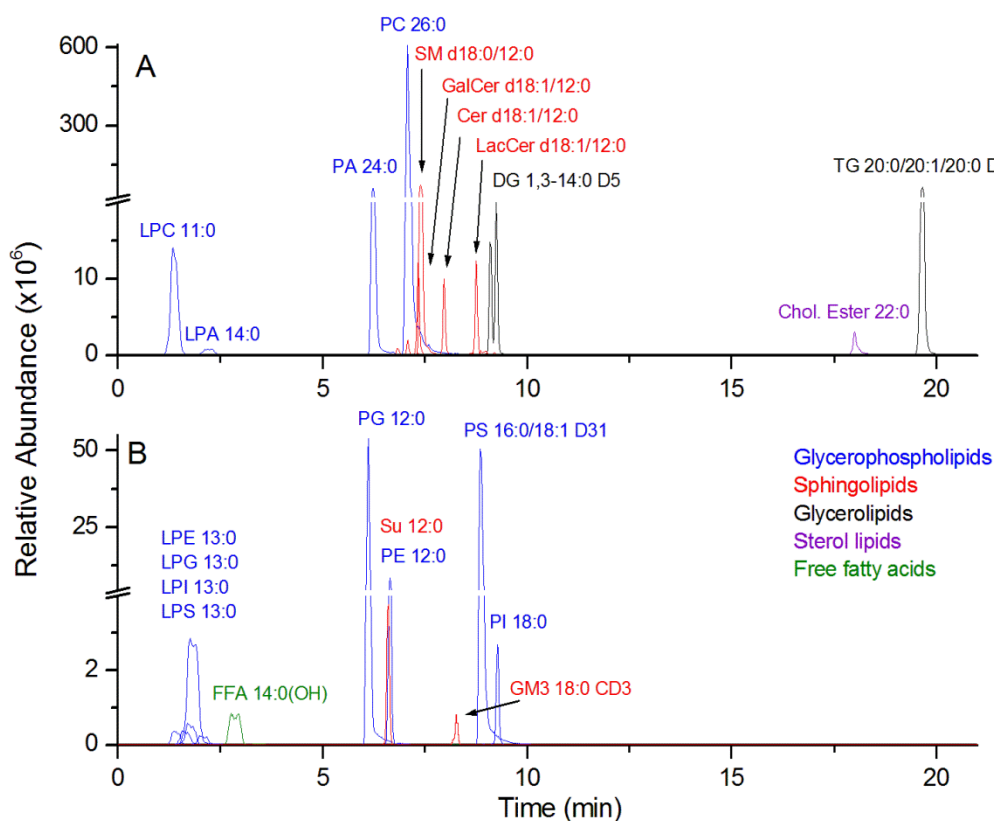


Figure 66. Chromatogrammes LC(RPC8)-ESI(+)-MS et LC(RPC8)-ESI(-)-MS du mélange des 22 étalons internes

3.3.1.2. Répétabilité de la méthode LCRP-MS

La répétabilité de la méthode depuis l'extraction des échantillons à la détection des molécules a été évaluée. Nous avons réalisé 5 extractions indépendantes d'échantillons de LCR et de plasma surchargés avec le même mélange de 22 IS à des concentrations définies (choisies en fonctions des concentrations des espèces lipidiques endogènes). Chaque extrait a été analysé en LC(RPC8)-MS. La Table 20 présente les valeurs de coefficients de variations (CV) sur les aires de pic chromatographique et sur les temps de rétention obtenus pour les deux matrices biologiques. Les valeurs de CV sur les aires sont inférieures à 8%. La valeur la plus élevée (17.4% pour les échantillons de plasma et 20.1% pour les échantillons de LCR) est rencontrée pour les glycérolipides. Ceci est dû à leur grande hydrophobicité qui provoque l'apparition d'un effet mémoire.

3.3.1.3. Linéarité de la méthode RPLC-MS

La linéarité de la méthode a été mesurée pour chaque étalon interne sur la base d'une gamme allant de 10 ng.mL^{-1} à 1000 ng.mL^{-1} (en concentrations équivalentes à celles des lipides endogènes). Le mélange de 22 IS a été ajouté aux ELT de LCR et 20 IS aux ELT de plasma (pas de FFA, ni de GM3 car pas de produits disponibles) à six concentrations différentes. La détection à 10 ng.mL^{-1} a été

possible pour la majorité des IS dans les deux matrices biologiques avec des coefficients de régression (r^2) allant de 0.926 à 0.999 (voir Table 20).

Classe lipidique	LCR humain				Plasma humain			
	Répétabilité (RSD, %)	Variation du temps de rétention (RSD, %)	Rendement d'extraction (%)	Linearité R^2	Répétabilité (RSD, %)	Variation du temps de rétention (RSD, %)	Rendement d'extraction (%)	Linearité R^2
Acides gras libres								
C14:0(OH)	5.8	1.1	88.2	0.997	2.1	1.1	109.5	nd
Sphingolipides								
d18:1/12:0 Cer	3.9	0.7	93.5	0.999	4.7	0.1	63.2	0.982
d18:1/12:0 GalCer	4.7	0.7	92.6	0.999	6.2	0.1	59.2	0.982
d18:1/12:0 lacCer	3.1	0.7	89.1	0.999	7.0	0.1	53.0	nd
d18:1/12:0 Su	3.8	0.4	85.4	0.998	1.8	0.3	66.2	0.926
d18:0/12:0 SM	5.1	0.7	93.4	0.999	6.2	0.1	60.3	0.985
GM3 CD3	2.8	0.3	89.6	0.997	2.7	0.2	84.0	nd
Glycerophospholipides								
11:0 LPC	10.6	1.7	86.9	0.992	6.2	0.5	56.0	0.986
14:0 LPA	18.5	1.9	65.5	0.979 *	7.9	0.8	24.7	nd
13:0 LPE	6.5	0.8	75.8	0.995	2.0	1.1	69.0	0.971
13:0 LPS	5.8	2	73.5	0.997	6.6	0.6	35.4	0.988
13:0 LPI	10.3	1.7	82.9	0.979				
17:1 LPI					5.6	0.7	38.1	0.965
13:0 LPG	7.6	1	82.3	0.989	0.8	0.5	71.8	0.972
26:0 PC	10.6	0.7	88.4	0.985	5.3	0.1	60.8	0.986
24:0 PA	8.9	0.7	78.5	0.998	1.8	0.1	26.7	nd
24:0 PE	6.5	0.2	78.5	0.990	1.4	0.1	73.1	0.972
34:1 D31 PS	4.9	0.2	79.1	0.996	2.6	0.2	60.2	0.968 **
36:0 PI	7.8	0.4	90.1	0.986	3.5	0.3	92.5	0.968
24:0 PG	7.8	0.4	85.2	0.994	1.6	0.2	75.5	0.958
Glycerolipides								
1,3 - 28:0 D5 DG	20.1	0.7	99.1	0.974	5.5	0.1	59.8	0.979
60:1 D5 TG	16.9	0.5	78.4	0.935 *	5.6	0.1	66.3	0.981
Esters de cholestérol								
22:0 Cholesteryl Ester	14.8	0.5	121.2	0.981 **	17.4	0.1	128.7	0.986

* (50 - 1000 ng.mL⁻¹)

** (100 - 1000 ng.mL⁻¹)

Table 20. Validation du rendement d'extraction, de la répétabilité et de la linéarité de la méthode RPLC-MS

Le calcul de linéarité pour l'IS d18 :1/12 :0 LacCer n'a pas été possible car ce dernier était naturellement présent dans les extraits plasmatiques non surchargés en IS.

Pour les espèces PA et LPA, la limite de détection était de 500ng.mL⁻¹. N'ayant eu que deux points de gamme pour la construction de la droite de régression, les valeurs de r^2 n'ont pas été reportées sur le tableau.

L'efficacité d'extraction observée pourrait être optimisée par une augmentation de la proportion en chloroforme ou de la proportion en mélange CM 1 :1 v/v. Celle-ci pourrait aider à augmenter ce rendement d'extraction tel que démontré sur des extraits lipidiques de levures (Ejsing et al; 2009). Concernant les espèces appartenant aux classes de PA et LPA dont l'efficacité d'extraction resterait faible, ce problème peut être contourné par l'addition d'acide à la phase aqueuse afin d'optimiser leur partage (Ogiso & Taguchi, 2008, Shan et al; 2008) et leur passage en

phase chloroformique. Des extractions à base de methyl tert-butyl ether (MTBE) (Gregory et al; 2013, Matyash et al; 2008, Pizarro et al; 2013) commencent également à se populariser. Ces dernières montrent des rendements d'extraction équivalents voire supérieurs à ceux rencontrés pour le Bligh & Dyer ou le Folch mais peu d'études comparatives n'ont été publiées. Par ailleurs, leur évaluation ne concerne que peu de classes lipidiques (par exemple pas de données chiffrées pour les gangliosides). Notons que les rendements d'extraction étaient, de façon générale, inférieurs pour le plasma que pour le LCR. Ceci pourrait être dû à la plus grande teneur en lipides du plasma et qui pourrait saturer leur partage.

Bien que l'efficacité de l'extraction des espèces lipidiques n'ait pas révélé de rendements optimaux, l'extraction reste répétable et permet de ce fait de comparer des échantillons traités par du CM 1 :1 v/v.

D'autre part, l'étude de la linéarité s'est effectuée sur une gamme allant de 10 ng.mL⁻¹ à 1000 ng.mL⁻¹, cette dernière pourrait aussi se préciser aux valeurs extrêmes et même au-delà des limites étudiées pour les classes dont la LOD a été provisoirement déterminée à 10 ng.mL⁻¹. Cependant, la seule étude de linéarité menée sur du plasma (Hu et al; 2008), a investigué une gamme allant de 60 ng.mL⁻¹ à 180 µg.mL⁻¹. Les valeurs de r² obtenues sur 5 étalons internes (TG, PC, PG, PE et LPC) étaient de l'ordre de (0.9994-0.9999). Notre étude peut être considérée comme plus exhaustive et donne des résultats satisfaisants.

3.3.2. Traitement des données

Une fois les méthodes d'extraction et de profilage validées, leur utilisation a permis l'acquisition de données LC-HRMS volumineuses dont la prise en charge ne peut se faire manuellement. En effet, il est impératif d'avoir recours à des outils de traitement de données assurant (i) la détection automatique des signaux et (ii) leur annotation automatique.

3.3.2.1. Détection automatique des signaux : centWave

En plus des enjeux inhérents à la détection automatique des signaux à partir de données LC-HRMS à savoir la détection simultanée de pics chromatographiques d'intensités extrêmes et de largeurs variées, les jeux de données lipidomiques générés par la méthode développée présentent certaines particularités par rapport aux données métabolomiques notamment quant à la détection d'isomères. En effet, la distinction entre les molécules isomères résolues sans retour à la ligne de base ou celles dont l'élution est proche avec une grande différence d'intensité est très délicate (Hu et al; 2008) (Barroso & Bischoff, 2005).

La solution de traitement de signal initialement envisagée utilisée était la librairie XCMS permettant la détection des pics à l'aide de l'algorithme matchedFilter. Cette dernière ne donnait

pas entière satisfaction avec notamment la multiplicité de pics correspondants à une même molécule. Une réflexion quant au choix de l'algorithme de détection automatique des signaux s'est de ce fait imposée. Dans ce contexte, nous avons collaboré avec la société Profilomic pour la mise en place de ces outils.

Après une étude bibliographique puis analyse comparative sur des algorithmes pré-sélectionnés, nous avons choisi d'utiliser l'algorithme. En effet, sa force majeure réside en le fait que contrairement à son homologue `matchedFilter`, il ne découpe pas le domaine spectral en tranche de m/z mais qu'il crée plutôt des régions d'intérêt « ROI » (de l'anglais « Region of interest »). Les spectres sont traités les uns après les autres et les ions qu'ils contiennent (i) s'ajoutent à une ROI déjà existante si ils respectent l'erreur de masse tolérée, ou bien (ii) entraînent la création d'une nouvelle ROI (Figure 67). Ces ROI doivent alors remplir des critères d'intensité et de longueur satisfaisants pour être gardés. De plus, il permet de rechercher les pics dans la dimension de temps de rétention avec une tolérance quant à la largeur de pic définie par un intervalle et non plus une valeur moyenne.

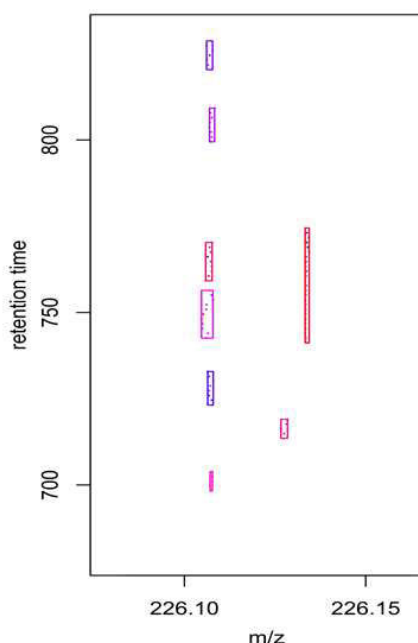


Figure 67. Création des ROI par l'algorithme centWave

Chaque point placé en dimension m/z et rt représente un scan centroïde, tandis que les rectangles autour indiquent les dimensions du ROI créé. Notons que ces dimensions sont différentes et s'adaptent aux données générées.

De plus, des filtres supplémentaires sont disponibles (intensité minimale des pics, rapport signal/bruit, modélisation gaussienne de la forme des pics chromatographiques et élimination des pics artéfactuels créés par la FT en appliquant un filtre d'intensité basé sur une distribution lorentzienne aux alentours du pic de masse) afin de s'assurer de la qualité des pics détectés tels.

Un jeu de données d'apprentissage simple qui consiste en une gamme en solvant d'étalons internes à 6 concentrations (100ng.mL^{-1} à $10\mu\text{g.mL}^{-1}$) répétée une fois par point de gamme a été acquis en RPLC-HRMS. Les contraintes imposées étaient de détecter simultanément de façon automatique à l'aide de l'algorithme mis en place des pics fins (peu intenses), des pics larges (intenses) et des molécules isomères mal résolues.

L'exemple de la détection des isomères PE 36:2 (*cis*) et (*trans*) (composés de référence) qui sont résolus sans retour à la ligne de base est présenté ci-dessous.

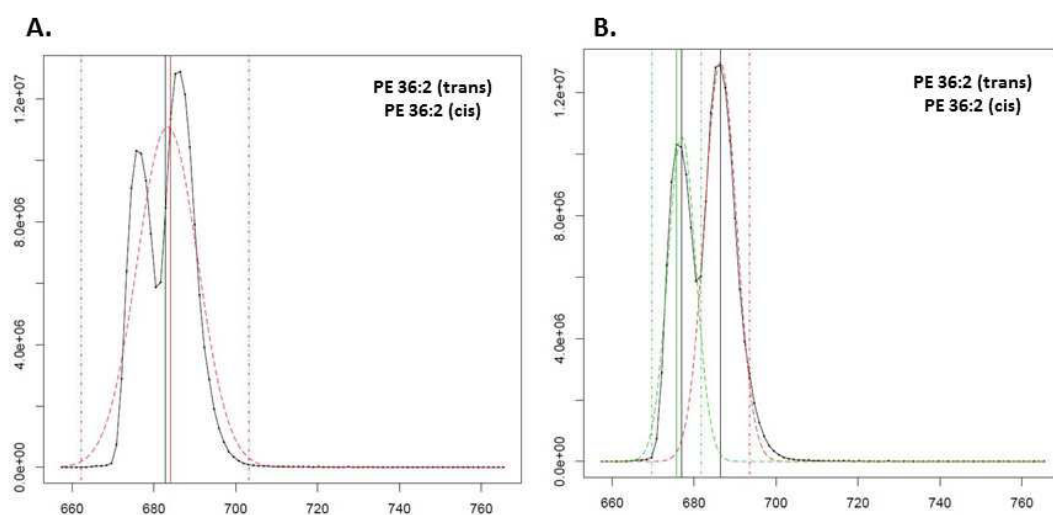


Figure 68. Détection de deux étalons internes régioisomères ($10\mu\text{g/mL}$ en tampon)

En noir, l'EIC m/z (ESI-) 742.539 correspondant aux IS PE32 :2 *cis* et *trans*. En pointillé l'intégration automatique réalisée par l'algorithme centWave. En (A), la détection automatique des pics avant optimisation des paramètres centWave, en (B) la détection automatique des espèces lipidiques après optimisation des paramètres de détection automatique des signaux.

Une fois les paramètres sélectionnés, nous avons appliqué la détection automatique des signaux au jeu de données lipidomiques ayant servi à la validation de la méthode sur des extraits lipidiques de LCR. Nous avons évalué la précision de l'algorithme de peakpicking de CentWave (logiciel XCMS) dans l'intégration des pics chromatographiques de lipides endogènes du LCR dont les aires allaient de $1 \cdot 10^5$ à $2 \cdot 10^9$. Les aires chromatographiques de 162 molécules endogènes annotées dans un ELT de LCR, et détectés par le logiciel ont été comparées aux aires mesurées par une intégration automatique Xcalibur via le module de QuanBrowser (algorithme ICIS). Les résultats sont présentés en Figure 69. La régression linéaire de ces données montrent une excellente corrélation entre les deux méthodes ($R^2 = 1.000$) et une pente de 1.000 sur ces quatre Log d'intensité considérés.

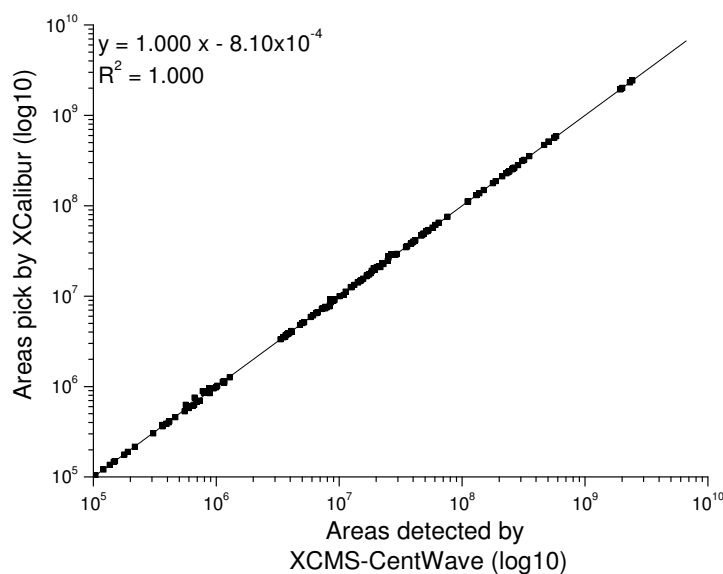


Figure 69. Comparaison de l'intégration automatique des pics chromatographiques des espèces lipidiques par XCMS (CentWave) et par le module QuanBrowser (ICIS)

Les paramètres ont de ce fait été optimisés et appliqués à la détection automatique des signaux des jeux de données lipidomiques plasmatiques.

3.3.2.2. Développement d'outils d'annotation de jeux de données lipidomiques

La détection automatique des signaux génère une liste de variables (couples temps de rétention et masse précise) qu'il reste à identifier ou à minima à annoter de la façon la plus rigoureuse possible. Pour ce faire, nous avons adopté une procédure d'annotation qui considère en plus du rapport m/z , l'intervalle de temps de rétention et leurs correspondances avec ceux décrits dans la base de données interne, et qui tient également compte du ratio isotopique. L'annotation automatique des jeux de données LC-MS développée se fait en 2 étapes :

- 1- D'abord il y a correspondance de la masse précise et du temps de rétention des ions observés et détectés par le logiciel centWave avec ceux répertoriés dans la base de données interne. Alors que la tolérance sur la mesure de masse précise est fixée à $\pm 8\text{ppm}$, l'intervalle de temps de rétention par espèce lipidique est large. Ce dernier est fixé par classe de lipides et non pas par espèce afin de ne pas se restreindre aux seuls isomères observés lors de l'analyse effectuée pour l'implémentation de la base de données (Figure 63). De plus, il y a calcul de la précision de mesure en masse exprimée en ppm pour chaque molécule ainsi annotée.

Le fichier de sortie est une liste de variables avec le cas échéant sa correspondance en annotation (ion pseudomoléculaire, isotope ^{13}C , isotope $[^{13}\text{C}_2]$ ou encore adduit NH_4^+) et sa précision de mesure en masse précise. Ces variables sont dupliquées si elles présentent plusieurs annotations cohérentes.

- 2- Dans un second temps, les annotations non pertinentes sont filtrées et ce en ne gardant que les ions monoisotopiques (ou adduit NH_4^+ pour les lipides neutres) présentant au minimum un isotope ^{13}C , et pour lesquels un calcul de l'abondance isotopique relative RIA (calculée sur l'ensemble des spectres) satisfait à $\pm 30\%$ de l'abondance isotopique théorique. Cette tolérance a été fixée à partir des performances déjà décrites sur la technologie Orbitrap (Xu et al; 2010).

Le calcul du RIA est assuré par l'équation suivante

$$\text{erreur RIA (\%)} = 100 \times \frac{\text{RIA}_{\text{exp}} - \text{RIA}_{\text{théor}}}{\text{RIA}_{\text{théor}}}$$

Après ces différents développements tant sur un plan analytique que bioinformatique, il nous a été possible d'envisager la réalisation d'une analyse lipidomique. Nous nous sommes alors penchés sur la question de l'annotation du lipidome plasmatique humain.

3.4. Annotation du lipidome plasmatique humain

A partir du jeu de données ayant servi à la validation de la méthode, nous avons d'abord généré une liste de variables détectées automatiquement à l'aide du logiciel centWave. Cette dernière est épurée à l'aide de filtres analytiques imposés à savoir : un coefficient de corrélation avec les séries de dilutions supérieur à 80%, un coefficient de variation des aires depuis l'extraction des métabolites à leur détection automatique inférieur à 20% et un rapport des aires entre les échantillons biologiques et les blancs d'analyse supérieur à 1.5. La matrice de données ainsi filtrée est ensuite soumise à la procédure d'annotation automatique précédemment décrite afin de réaliser une étude comprehensive du lipidome plasmatique humain.

Analyse des données et critères de validation

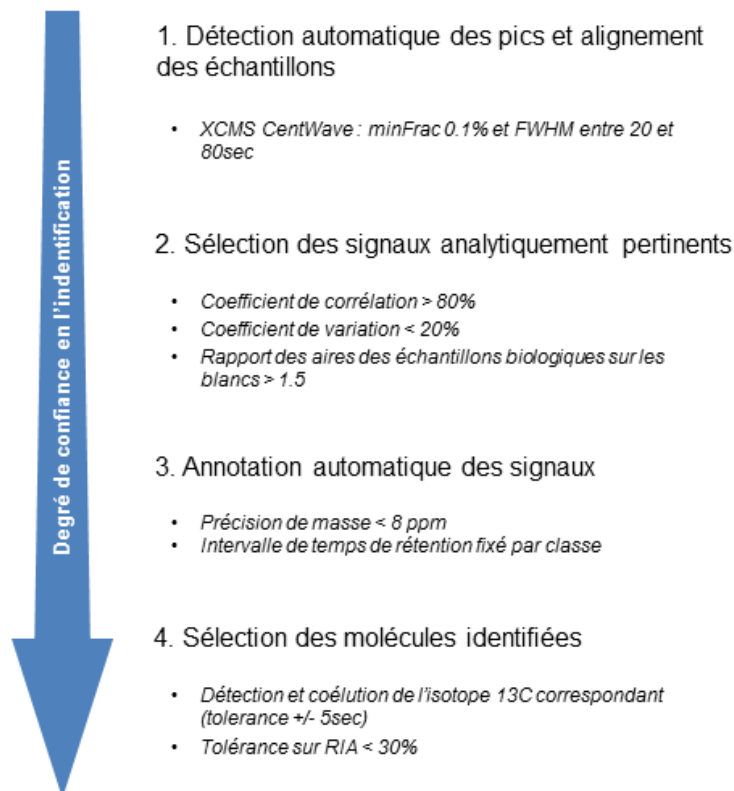


Figure 70. Démarche d'analyse globale des données lipidomiques

La démarche présentée en Figure 70 a permis l'annotation de 706 espèces lipidiques présentes dans un plasma humain et réparties dans 5 familles et 21 classes de lipides répondant aux critères de filtration analytiques imposés (**Table 21**). Le nombre d'espèces et de classes détectées dépasse celui décrit dans les études descriptives du lipidome plasmatique. En effet, Hu et coll. (Hu et al; 2008) ont identifié 160 espèces de lipides se répartissant sur 8 classes et 4 familles de lipides dans du plasma de rat en utilisant la méthode RPLC-MS qui nous a servi de référence au développement de notre méthode de profilage. Par ailleurs, Quehenberger et coll. (Quehenberger et al; 2010) en ont identifié 588 à l'aide d'une approche multiplateforme alliant GC-MS, LC-MS et LC-MS/MS sur différents instruments et après plusieurs méthodes d'extraction dédiées à chaque classe de lipides investiguée.

Familles	Nombre d'espèces uniques détectées	Forme moléculaire détectée	CV extraction (n=5) %	RIA (erreur absolue) %	Corrélation des dilutions	Précision de mesure de masse (ppm)
Acides gras libres						
FA	6	[M-H] ⁻	3.5	21	0.96	-3.5
Sphingolipides						
Cer	9	[M-H] ⁻	9.0	8.3	0.99	2.1
HexCer	6	[M-H] ⁻	11.1	9.5	0.99	1.5
LacCer	9		14.6	6.9	0.99	-3
Su	4	[M-H] ⁻	9	10.5	1.00	-3.9
SM	35	[M+H] ⁺	11.6	3.7	1.00	1.5
Gangliosides	8	[M-H] ⁻	7	4	0.96	-0.3
Glycérophospholipides						
LPC	36	[M+H] ⁺	14.3	9.1	0.99	2.8
LPA	1	[M+NH ₄] ⁺	12.3	14.4	0.98	-3.7
LPE	14	[M-H] ⁻	4.2	9.2	0.99	-2.3
LPS	4	[M-H] ⁻	4.7	10.7	1.00	-0.4
LPI	3	[M-H] ⁻	10.4	6.5	0.99	4.3
LPG	0	[M-H] ⁻	ND	ND	ND	ND
PC	173	[M+H] ⁺	12.4	6.6	0.99	1.8
PA	3	[M+NH ₄] ⁺	9.6	9.1	0.99	-0.3
PE	50	[M-H] ⁻	5.3	8.3	0.99	-1.6
PS	80	[M-H] ⁻	4.9	7.7	1.00	-0.1
PI	35	[M-H] ⁻	5.8	11.7	0.99	-1.3
PG	9	[M-H] ⁻	3.9	15	0.98	-2.7
Glycérolipides						
MG	0	[M+NH ₄] ⁺	ND	ND	ND	ND
DG	19	[M+NH ₄] ⁺	11.4	8.7	0.99	2.6
TG	170	[M+NH ₄] ⁺	11.3	6.5	0.98	1.7
Stérols						
Esters de cholestérol	32	[M+NH ₄] ⁺	12.5	6.2	0.97	2.1

Table 21. Espèces lipidiques endogènes détectées et validées dans les extraits lipidiques totaux de plasma humain

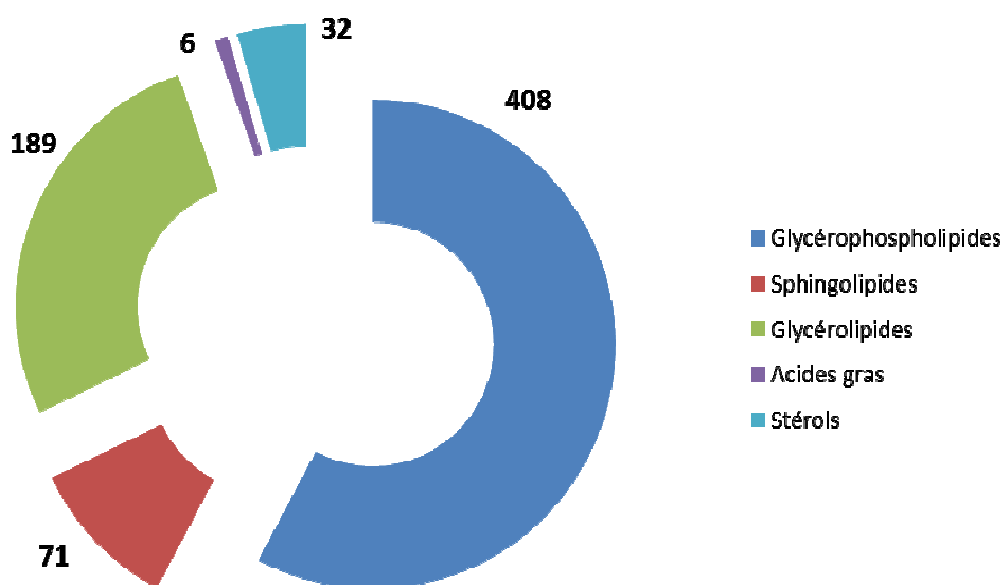


Figure 71. Répartition du nombre d'espèces lipidiques plasmatiques au sein des 5 grandes familles de lipides

Les glycérophospholipides représentent en nombre plus de 57% des espèces détectées. Les espèces PC sont les plus nombreuses dont 40% sont des formes éther ou plasmalogènes. Conformément aux résultats décrits en validation de l'extraction, peu d'espèces PA ou LPA ont été détectées. De plus, nous n'avons identifié aucune espèce LPG. La méthode développée ayant été validée à l'aide notamment d'un standard LPG (ayant assuré une efficacité d'extraction à 71.8% et 0.8% de variation) nous pouvons conclure à l'absence d'espèces LPG détectables dans le plasma humain dans les conditions d'analyse choisies.

La deuxième famille la plus représentée en termes de nombre d'espèces détectées est celle des glycérolipides (26% du nombre total d'espèces). Celles-ci appartiennent majoritairement à la classe des TG avec 170 espèces annotées (90% de TG et 10% de DG), la diversité de 3 chaînes d'acides gras aidant. De plus, ces espèces sont facilement détectables de par leur forte concentration plasmatique (l'équivalent de 90.6mg/dl pour 18 espèces détectées (Quehenberger et al; 2010)). Cette dernière étant directement dépendante de la prise alimentaire et de leur incorporation constitutive dans les lipoprotéines (chylomicrons et VLDL de l'anglais « Very Low Density Lipoprotein »). Ce nombre d'espèces TG détectées dépasse ceux précédemment décrits (18 (Quehenberger et al; 2010) et 86 (Bird et al; 2011) espèces). Leurs chaînes d'acides gras estérifiées sont considérées comme longues (le nombre total de carbone ≥ 52 dans plus de 70% des espèces détectées) et polyinsaturées (nombre total d'insaturations ≥ 3 dans plus de 66% des espèces détectées, allant jusqu'à 13 double-liaisons pour l'espèce TG 62 :13). Par ailleurs, Quehenberger et coll. ont procédé à une transformation des espèces 1,2 et 1,3-DG en difluorophénylurethanes et à leur séparation chromatographique en NPLC. Ceci a assuré la détection de 28 espèces de DG estérifiées en positions (*sn-1* et *sn-2*) et de leurs isomères (*sn-1* et *sn-3*) soit un total de 55 espèces chromatographiquement résolues. Dans notre cas, nous avons annoté 19 espèces DG (1,2 et 1,3DG confondues) avec la détection supplémentaire d'espèces à très longues chaînes et polyinsaturées non précédemment répertoriées dans le plasma (DG 44 :10, DG 44 :11 et DG 44 :12). Par ailleurs, bien que résultant du catabolisme des TG et DG aucune espèce de MG n'a été détectée. Pour l'expliquer, la méthode développée a dans un premier temps été incriminée du fait qu'aucun étalon interne de cette classe-là n'ait été utilisé lors du développement et de la validation de la méthode. Par ailleurs, ces espèces ne sont également pas décrites dans les études complètes plasmatiques notamment celles menées à l'aide d'analyses multi-plateformes (Quehenberger et al; 2010), et seules 24 espèces sont répertoriées dans la base de données LIPIDMAPS (versus 346 pour les DG et 2620 pour les TG). Nous pouvons également supposer qu'elles sont minoritaires ou bien présentes à de faibles concentrations. Ces espèces sont aussi dégradées par des lipases afin de restituer la réserve de glycérol et d'acides gras libres nécessaires à la synthèse des lipides complexes et peuvent de ce fait ne pas être détectées.

71 espèces appartenant à la famille des sphingolipides ont été identifiées dans le jeu de données plasmatiques étudié. Bien que toutes les classes soient détectées, les espèces sphingomyélines sont les plus représentées. Celles-ci présentaient des acides gras souvent saturés ou mono insaturés et parfois à nombre impair de carbone. Par contre, la nature de la base sphingoïde dans chacune des classes de sphingolipides ne peut être déterminée que par analyse MS/MS réalisée en mode CID résonant tel que nous avons pu le démontrer précédemment en fragmentation d'espèces de sulfatides sur le LTQ-Orbitrap (Figure 60). Différentes bases sphingoïdes ont été décrites dans la littérature (Quehenberger et al; 2010), et leur caractérisation au sein du laboratoire est en cours.

Les espèces d'esters de cholestérol ont également été identifiées. La moitié des espèces identifiées présentait un isomère. Les isomères de CE ne peuvent être dûs qu'à la position de la double liaison. Cette dernière n'est investiguée que par des méthodes appropriées (discutées en PARTIE A.CHAPITRE III.2.2.3 de l'introduction générale). Des espèces portant des chaînes carbonées à nombre impair de carbone ont également été détectées. D'autre part, peu d'espèces d'acide gras libres ont été détectées. Ces derniers bien que peu nombreux sont le reflet du catabolisme et anabolisme des autres familles de lipides. Le tissu adipeux en est la source majeure. Si l'on considère que les facteurs de réponse des espèces détectées sont relativement proches, les acides linoléique (FA 18 :2) et oleique (FA 18 :1) semblent être les plus abondants, suivis de l'acide arachidonique (FA 20 :4) et d'autres acides gras hydroxylés. L'acide palmitique (FA16 :0) bien qu'intense et répondant à l'intégralité des autres filtres analytiques imposés (à savoir répétabilité d'extraction et corrélation avec le facteur de dilution) n'a pas été retenu comme molécule annotée car son rapport RIA était inférieur aux limites imposées (60% d'erreur pour une tolérance à 30%). De même pour l'acide stéarique (FA18 :0) qui a été éliminé de la liste des molécules endogènes détectées quant à lui car son isotope ^{13}C n'était pas détecté. Ces deux derniers cas de figure prouvent que le système de sélection des molécules annotées est rigoureux dont les critères peuvent être trop contraignants pour les molécules présentant un faible nombre d'atomes de carbone tels les acides gras libres.

4. Conclusion et perspectives

Nous avons développé et validé une méthode RPLC-MS assurant le profilage d'extraits lipidiques totaux à partir de matrices biologiques complexes. La méthode d'extraction Bligh and Dyer choisie permet l'extraction d'une large gamme d'espèces allant des espèces polaires (exemple des gangliosides et des lysoglycérophospholipides) aux espèces hydrophobes (exemples triglycérides) avec une bonne répétabilité.

Cette méthode analytique a été couplée à des outils bioinformatiques développés (ou à minima optimisés) pour assurer le traitement haut débit des données lipidomiques générées à commencer par une base de données interne contenant plus de 2200 espèces moléculaires uniques

d'identification contrôlée. Cette dernière est en continuelle implémentation (notamment sous-classe des eicosanoides qui sont des acides gras bioactifs dérivés de l'acide arachidonique et impliqués dans les phénomènes inflammatoires ou les lipides oxydés ou encore les espèces nouvellement caractérisées). De plus, des outils de détection automatique des signaux et d'annotation dédiés sont chainés afin de réaliser une analyse intégrée.

En plus de la détection et de l'identification, une étude préliminaire visant à assurer une quantification absolue à large spectre a été menée. Celle-ci reposerait sur l'analyse DIMS de fractions (LC-NH₂) surchargées d'un étalon interne par classe. Les résultats obtenus sont présentés en ANNEXE 3.

L'approche lipidomique développée a permis d'identifier plus de 700 espèces lipidiques plasmatiques uniques appartenant à 21 classes et sous-classes de lipides. L'influence de paramètres physiopathologiques sur ces espèces a été étudiée et sera l'objet du prochain chapitre de résultats présentés (PARTIE A.CHAPITRE I.4.3PARTIE A.CHAPITRE II.5.2).

B. APPLICATIONS DES APPROCHES GLOBALES AUX COHORTES

Il est communément admis que les pathologies provoquent des modifications au niveau biochimique. Bien que le système endocrinien ou encore le système nerveux végétatif tentent de conserver le milieu intérieur à l'équilibre, des variations de concentrations adaptatives de métabolites surviennent. De plus, l'interaction du génome avec l'environnement façonne le phénotype et modifie de ce fait le métabolome. Il est donc possible de mettre en évidence un état physiopathologique ou des interactions individus-environnement à l'aide d'empreintes métabolomiques acquises sur de larges cohortes d'individus (Holmes et al; 2008, Rezzi et al; 2007). En effet, Holmes et coll. ont démontré les répercussions du régime alimentaire, de l'hygiène de vie ou encore de la flore intestinale sur les empreintes $^1\text{H-NMR}$ des urines de 24h d'une cohorte de 4630 participants (étude épidémiologique INTERMAP) (Holmes et al; 2008).

Dans ces contextes épidémiologiques où prime la taille des cohortes investigués, les approches basées sur la NMR, du fait de leur robustesse, sont les méthodes de choix. Cependant, la sensibilité des méthodes ESI-MS serait d'un intérêt certain. C'est ainsi que certaines équipes ont entrepris des analyses de larges cohortes par des méthodes ciblées assurant une quantification absolue des métabolites identifiés tel que démontré par Suhre et coll. avec la double analyse en UHPLC/MS/MS et GC-MS d'extraits sériques de 3080 individus de la cohorte KORA (Suhre et al; 2011), ou encore à l'aide de kits commercialisés par la société Biocrates (Suhre et al; 2010) dans le cadre du diagnostic du diabète. D'autres auteurs ont, quant à eux, évalué les performances analytiques des plateformes HRMS pour l'analyse de larges cohortes. Ils ont statué que la stratégie appropriée consiste à effectuer de petites séries d'analyse dans lesquelles les données acquises sont robustes et reproductibles, puis à intégrer ces sous-ensembles en un seul ensemble de données directement lié à une expérience biologique unique (Begley et al; 2009, Zelena et al; 2009). Dès lors, un protocole standardisé pour l'analyse LC- et GC-HRMS de grandes cohortes a été publié en 2011 (Dunn et al; 2011). Ce dernier traite de l'analyse métabolomique de plasma depuis l'extraction des échantillons au traitement des données. La fusion des données depuis des séries d'expériences différentes y est abordée. Cette dernière repose sur des échantillons biologiques de QC représentatifs des échantillons à analyser rendant de ce fait possible les analyses à large échelle en HRMS.

Les empreintes globales obtenues notamment en ESI-HRMS peuvent contenir des facteurs confondants. Ces derniers peuvent être d'ordre (i) analytique ou encore (ii) biologique. Par exemple, lors d'une analyse effectuée au laboratoire en RP-ESI-HRMS sur des extraits métabolomiques de globules rouges de patients atteints de stomatocytose héréditaire hyperhydratée (OHSt), de patients drépanocytaires (SCD) et de sujets sains (Figure 72), une discrimination selon la première composante principale d'une analyse PCA a été observée entre les

patients SCD vs patients OHSt, ainsi qu'une discrimination selon la seconde composante entre les patients (SCD et OHSt) et les témoins (Darghouth et coll. Données non publiées). Cependant, la population témoin se sépare également selon la première composante en fonction du jour d'analyse (D ou D+30). Ceci remet en question la discrimination obtenue entre sujets SCD et OHSt car cette dernière peut être due au jour d'analyse (biais analytique) et non pas à la physiopathologie sous-jacente.

Un deuxième biais, cette fois ci d'ordre « biologique » peut être également mis en cause. En effet, la stomatocytose étant une maladie de la perméabilité de la membrane érythrocytaire, les GR sont hyperhydratés et de ce fait, leurs extraits sont globalement dilués par rapport à ceux des témoins ou des patients atteints de SCD. Il s'agira alors de normaliser les données en fonction du volume érythrocytaire (VGM) avant la fusion des données afin de s'affranchir de ce biais inhérent à la maladie.

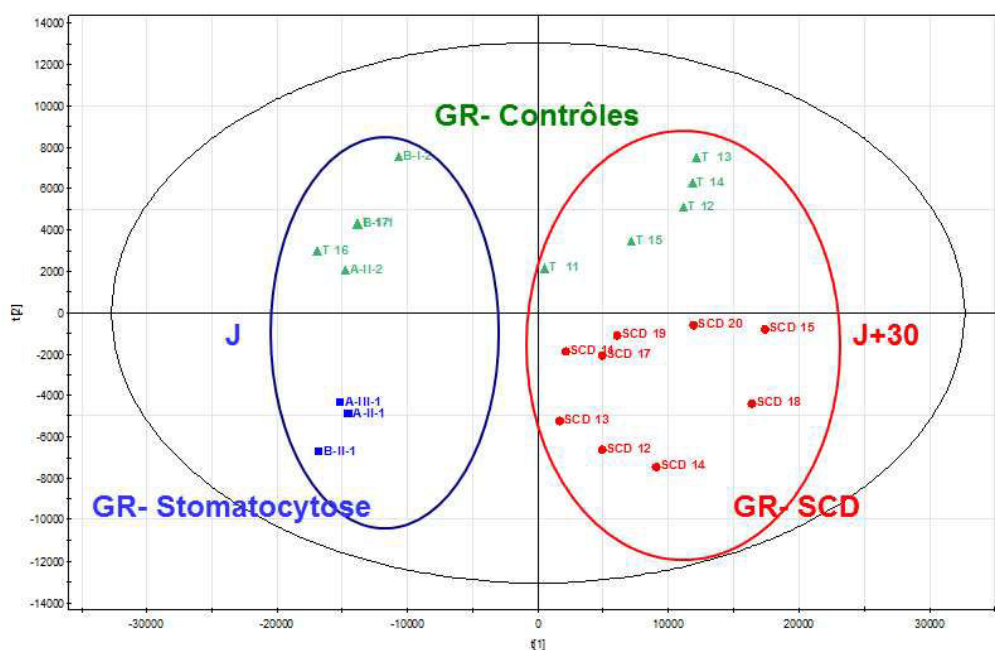


Figure 72. Carte factorielle des individus obtenue en analyse RPLC-ESI-MS d'extraits métabolomiques de globules rouges humains (Darghouth et coll. non publié)

« GR-Contrôles » sont les individus témoins. « GR-SCD » pour « sickle cell disease » sont les patients drépanocytaires. « GR-Stomatocytose » sont les patients atteints de stomatocytose (OHSt). Le cercle bleu définit une expérience menée le jour « J ». Le cercle vert englobe les échantillons analysés le jour « J+30 ».

Nous pouvons également citer les travaux épidémiologiques sur la cohorte INTERMAP et l'influence des habitudes alimentaires dans la discrimination entre population de régions géographiques différentes. Par exemple, la population japonaise présentait une augmentation de la concentration de triméthylamine-N-oxide (TMAO) dans les urines de 24 heures analysées en $^1\text{H-RMN}$

contrairement aux populations américaine et chinoise. Ce dernier est apporté par une alimentation riche en poissons et ne renseigne pas réellement quant aux origines géographiques.

Dans ce contexte, nous allons vous présenter une étude de cohorte de 305 individus, tous analysés en LC-ESI(-/+)-HRMS selon les recommandations Dunn et coll (Dunn et al; 2011). D'abord, l'analyse métabolomique et lipidomique de 227 salariés du CEA-Saclay nous a permis d'étudier l'impact du bruit de fond biologique pouvant être des potentiels facteurs confondants telles que l'influence du genre, de l'âge et du BMI (indice de masse corporelle de l'anglais « Body mass index ») sur les empreintes métabolomiques. Ces derniers pouvant être des facteurs confondants dans certains cas – comme nous allons le démontrer-. Ensuite, nous avons réalisé les analyses métabolomique et lipidomique des extraits plasmatiques de 78 sujets diabétiques et de 54 de leurs extraits érythrocytaires, afin de stratifier les patients selon l'origine de leur diabète et de comprendre au mieux la physiopathologie du diabète de type 2.

CHAPITRE I. ANALYSE DU METABOLOME ET LIPIDOME SÉRIQUES D'UNE COHORTE DE 223 SALARIES DU CEA-SACLAY

Alors que les empreintes métabolomiques sont de plus en plus analysées dans le cadre d'investigation de pathologies, peu d'intérêt a été porté aux changements dus à des modifications physiologiques. En effet, la première étude large spectre menée pour rendre compte de l'effet du genre, âge a été publiée en 2008 par la société Metabolon (Lawton et al; 2008). Cette dernière comprenait l'analyse en GC-MS et LC-MS/MS d'extraits plasmatiques de 269 individus. Des analyses statistiques univariées sur des variables catégorielles (3 groupes d'âge : 20-35, 36-50 et 51-65ans et 2 groupes de genre : femmes et hommes) ont permis de mettre en évidence des métabolites discriminants. Même si 4 extractions successives ont été menées sur les mêmes échantillons avant leur analyse, aucune ne concernait l'extraction des espèces lipidiques qui de ce fait n'ont pas été étudiées. Il a fallu attendre 2013 pour que Weir et coll. s'intéressent à l'effet du genre, âge, BMI et tabagisme sur les empreintes lipidomiques de 1000 individus. L'étude concernait 312 espèces lipidiques analysées en LC-ESI-MS/MS avec une analyse statistique de type « régression linéaire » pour l'analyse des effets des variables continues (Weir et al; 2013).

Basée sur la même méthodologie analytique que celle utilisée par Lawton et coll., une équipe japonaise a mené 6 ans plus tard, l'analyse métabolomique de 297 métabolites (Saito et al; 2014) mais aussi lipidomique de 253 espèces lipidiques d'une cohorte de 60 individus (Ishikawa et al; 2014). Ces études présentaient la force d'être menées parallèlement pour les extraits plasmatiques et sériques permettant de ce fait de renseigner quant aux différences entre les matrices disponibles à partir d'un prélèvement sanguin. Mais, une fois de plus, 4 groupes d'individus extrêmes (de 15 individus chacun) ont été définis pour les analyses statistiques uni- et multivariées (hommes ayant entre 24-34ans, hommes âgés de 55-64ans, femmes ayant entre 24-34ans, femmes âgées de 55-64ans). De plus, cette étude ne prenait pas en compte l'effet du BMI. Ce dernier paramètre, a quant à lui, fait l'objet d'une publication récente sur une cohorte de plus de 900 participants et sur 317 métabolites détectés selon le même protocole analytique. Encore une fois, des données plasmatiques et sériques ont été considérées et mêmes mises en commun lors des analyses statistiques (Moore et al; 2014).

L'ensemble de ces études récemment publiées ont mis en évidence l'importance de l'impact des facteurs physiologiques sur les empreintes métabolomiques et lipidomiques mais aucune information quant au recouvrement de ces informations n'a été fournie. De plus, ces études n'ont pas évoqué de procédure de fusion des données et seuls les résultats physiologiques ont été décrits.

Dans ce contexte, nous allons détailler ci-après les résultats de nos analyses LC-ESI-HRMS en métabolomique et lipidomique de 227 échantillons biologiques de salariés du CEA de Saclay. Cette

cohorte n'est pas considérée comme « témoin » car aucune information clinique complémentaire quant aux maladies associées n'est précisée. Elle n'est pas non plus « contrôlée » car aucun critère d'inclusion ou d'exclusion n'a été pris en compte. Cependant, elle reflète le bruit de fond biologique inhérent à chaque analyse métabolomique de cohortes et qui est souvent négligé bien que pouvant être la source de résultats erronés. Nous nous sommes focalisés sur les facteurs physiologiques principaux à savoir le genre, l'âge et le BMI sont investigués. L'analyse de cette cohorte nous a également servi de jeu de données d'apprentissage pour l'analyse et l'intégration de jeux de données (que nous appellerons dans ce qui suit fusion de données) incluant de larges quantités de données provenant de différentes analyses LC-ESI-HRMS de larges cohortes.

1. Matériels et méthodes

Les analyses métabolomiques ont été réalisées selon 3 conditions chromatographiques et dans les 2 modes d'ionisation, avec un système de chromatographie liquide de type NEXERA de la société Shimadzu (Champs-sur-Marne, France) couplé à un spectromètre de masse de type Exactive (Orbitrap) de la société Thermo Fisher Scientific (Courtabœuf, France).

Les analyses lipidomiques RPLC-MS ont été réalisées avec un système de chromatographique de type Ultimate (Thermo Fisher Scientific, Courtaboeuf, France) couplé à un spectromètre de type Q-Exactive (Thermo Fisher Scientific, Courtaboeuf, France).

Les détails des méthodes utilisées sont consignées en PARTIE A.CHAPITRE I.2.3 et PARTIE A.CHAPITRE II.2.2.

2. Echantillons biologiques

Les échantillons sériques de 227 travailleurs du CEA ont été recueillis par la médecine du travail du CEA de Saclay lors de leur examen médical annuel (courant 2008). Les volontaires ont rempli un formulaire de consentement. Les échantillons collectés sur 9 jours ont été anonymisés puis conservés à -80° C jusqu'à leur traitement et leur analyse. Les données physiologiques déclarées par les patients (âge, taille, poids, genre) ont été enregistrées pour chaque individu. Les échantillons sériques ont été recueillis auprès de 120 hommes et 107 femmes, âgés entre 18 et 81 ans et dont les indices de masse corporelle calculés vont de la maigreur à l'obésité. La description détaillée de la cohorte est présentée dans la Table 22. Aucun critère d'exclusion ou de normalisation (alimentation, tabagisme, prise médicamenteuse, exercice ...) n'a été appliqué afin d'obtenir l'échantillonnage le plus représentatif possible de la population.

Age (années)	18-34			35-54			>54			TOTAL		
Genre*	F	H	Total	F	H	Total	F	H	Total	F	H	Total
Indice de Masse Corporelle (IMC**)												
Maigreur (<18.5)	3	0	3	2	0	2	0	2	2	5	2	7
Normal (18.5-24.9)	25	30	55	45	42	87	7	1	8	77	73	150
Surpoids (25-29.9)	1	7	8	14	29	43	3	6	9	18	42	60
Obésité (>30)	0	0	0	5	2	7	2	1	3	7	3	10
TOTAL	29	37	66	66	73	139	12	10	22	107	120	227

Table 22. Description de la cohorte saclaysienne (classification des individus selon leur genre, âge et IMC)

* F : Femme; H : Homme / ** IMC = poids (kg) / taille (m²). Les classes d'âge ont été choisies selon celles investiguées par Saito et coll lors de l'analyse d'une cohorte de 60 individus répartis dans 4 classes de genre et âge (Ishikawa et al; 2014, Saito et al; 2014). Les classes d'IMC ont été choisies selon les recommandations de l'OMS (organisation mondiale de la santé)

Des échantillons biologiques de contrôle qualité «QC intraséries» consistant en un mélange des échantillons de la série à analyser ont été préparés. Un échantillon biologique de contrôle qualité «QC intraséries» consistant en un mélange d'échantillons de sérum acheté chez Bioprédic (Rennes, France) a été extrait indépendamment.

3. Traitement des données

La détection automatique des signaux a été réalisée avec le logiciel XCMS utilisant l'algorithme matchedFilter pour les données métabolomiques (les paramètres XCMS-CAMERA utilisés sont consignés en Table S.1 Annexe2). Les données ont ensuite été filtrées puis annotées selon la procédure d'annotation décrite en PARTIE A.CHAPITRE I (soit une annotation avec les bases de données publiques HMDB, KEGG et Metlin puis à l'aide de la base de données spectrales). Les aires des métabolites précédemment extraits par XCMS et identifiés ont été confirmées par des intégrations manuelles effectuées sur le module QuanBrowser (XCalibur, Thermo Scientific). La détection automatique des signaux et l'annotation des jeux de données lipidomiques a été réalisée selon la procédure décrite en PARTIE A.CHAPITRE II.

La correction des signaux au cours des analyses LC-ESI-MS s'est effectuée à l'aide de l'algorithme LOESS (de l'anglais « Low Order non linear locally Estimated Smoothing function »). La correction de la dérive analytique et les analyses statistiques univariées et multivariées en fonction des facteurs physiologiques ont été réalisées par Etienne Thévenot (DRT/LIST, CEA Saclay) sous l'environnement logiciel R (R Development Core Team, 2013).

4. Résultats

4.1. Démarche d'analyse

Cette étude a été menée sur l'intégralité des individus de la cohorte (227) en analyse métabolomique dans 2 modes d'ionisation ESI et en utilisant 3 conditions chromatographiques. Par ailleurs, l'analyse lipidomique a été faite sur 160 échantillons (parmi les 227 précédents) analysés avec une seule méthode chromatographique avec une détection dans les 2 modes d'ionisation. Ceci a abouti à plus de 1600 échantillons biologiques à analyser. Pour ce faire, nous avons adopté une démarche analytique et statistique pouvant gérer au mieux cette quantité de données (Figure 73).

Analyse lipidomique

160 individus – Réplicats techniques
Blancs d'analyse
QC intrasérie et QC interséries
Chloroforme/Méthanol 2:1

RP-C8 ESI (-)
 ESI (+) 3 séries

XCMS centWave
639 chromatogrammes

Fusion intermédiaire
1193 espèces identifiées

Correction du signal
Mise à niveau des QCinterbatch

-Test de Mann-Whitney Wilcoxon
-Corrélation de Spearman
-Régressions OPLS et OPLS-DA

1. Extraction des métabolites/lipides

2. Analyses LC-HRMS

**3. Détection automatique des signaux
& Fusion de bas niveau**

4. Filtration / Annotations des matrices

**5. Normalisation des données
Intra-séries d'analyse
Inter-séries d'analyse**

**6. Analyses statistiques uni- et
multivariées**

Analyse métabolomique

227 individus – Réplicats techniques
Blancs d'analyse
QC intrasérie et QC interséries
Méthanol

RP-C18 ESI (-)
PFPP ESI (+) 3 séries
ZICpHILIC

XCMS matchedFilter
& QuanBrowser
2070 chromatogrammes

223 métabolites identifiés

Correction du signal selon
l'algorithme « LOESS »
Mise à niveau des QCinterbatch

-Test de Mann-Whitney Wilcoxon
-Corrélation de Spearman
-Régressions OPLS et OPLS-DA

Figure 73. Démarche analytique et bioinformatique pour l'analyse métabolomique et lipidomique de la cohorte saclaysienne

Le premier enjeu a été de réaliser l'extraction et l'analyse LC-MS de l'intégralité des échantillons dans des conditions optimales sans introduire de biais. Pour ce faire, les échantillons biologiques ont été extraits puis analysés de manière aléatoire et ce dans les différents systèmes LC-MS décrits en CHAPITRE I. Pour cela, la randomisation s'est effectuée grâce au logiciel Excel. De plus, des répliques techniques ont également été réalisés (Figure 73). Un à deux échantillons pris au hasard dans chaque

lot d'analyse ont été extraits 2 fois par série d'expérience (au début puis à la fin de la série d'extraction) puis analysés de façon aléatoire au cours de l'analyse LC-MS.

Cependant, au-delà des échantillons biologiques à analyser, d'autres types d'échantillons servant à assurer et à corriger la qualité du signal sont introduits au cours des séries d'analyse (Figure 74). La séquence d'analyse commence toujours par une série des échantillons de solvant et des échantillons de contrôle qualité. Ceux-ci servent à stabiliser le système d'analyse et à vérifier la répétabilité des aires et temps de rétention des molécules détectées notamment celles des étalons internes (Table 9). Ces chromatogrammes ne seront pas considérés dans la suite du traitement de données. La structure d'une séquence d'analyse est décrite en Figure 74.

Après ce conditionnement, trois séries de dilutions en série de ces mêmes échantillons QC sont analysées. Ces séries (dilution au 1/8^{ème}, au 1/4^{ème} et au 1/2^{ème}) servent à sélectionner les ions dont les niveaux de concentration sont corrélés avec le facteur de dilution (Gika et al; 2007) (Figure 33).

Les échantillons biologiques sont ensuite analysés par lot de 10 (dans un ordre aléatoire). Une analyse LC-MS d'un blanc d'analyse et d'un échantillon de QC entre chaque série de 10 échantillons est effectuée. Les blancs d'analyse permettent d'évaluer l'effet mémoire ; les aires des différentes variables détectées servent alors à filtrer les données (cf Figure 33). En outre, le suivi de l'intensité du signal ESI dans les échantillons de QC au cours de l'analyse permet de modéliser et de corriger au mieux l'effet de la baisse du signal due à l'encrassement de la source.

En plus de ces QC propres à chaque séquence d'analyse dits « intra-série », nous analysons un autre type d'échantillon de QC dit « inter-séries ». Un aliquot de QC inter-séries est extrait en même temps que les autres échantillons puis analysé au minimum 5 fois par série d'analyse. Cet échantillon QC est donc représentatif des conditions d'analyse du lot et sert à normaliser et fusionner des données provenant de lots d'analyses différents (Dunn et al; 2011). En plus des QC inter-séries extraits autant de fois que de lot, 3 échantillons choisis au hasard ont été extraits et analysés une fois par série d'analyse. Ces derniers servent à valider la fusion réalisée grâce aux QC inter-séries.

Afin de préserver une stabilité du signal ESI-MS au cours de l'analyse, nous avons limité les séries d'analyse à 120 chromatogrammes. Ceci correspond à une moyenne de 75 échantillons biologiques par série d'analyse. A cela s'ajoute les répliques techniques, les échantillons de QC ainsi que des blancs d'analyse (Figure 74). De ce fait, l'analyse de la cohorte saclaysienne s'est donc divisée en 3 lots d'analyse qui sont fusionnés, filtrés puis soumis aux analyses statistiques.

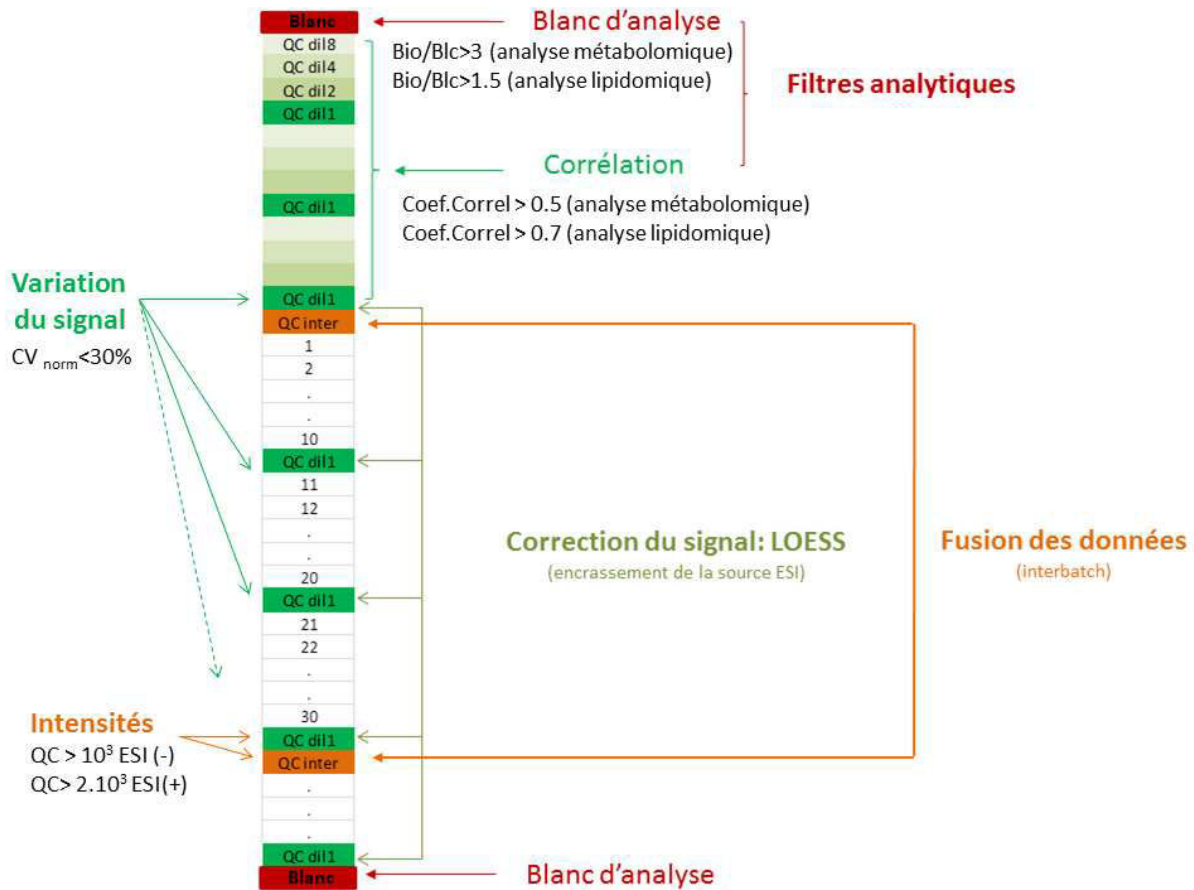


Figure 74. Structure d'une séquence d'analyse LC-MS

Le choix du type de test statistique à réaliser (à partir des tests disponibles, voir Table 23) dépend du caractère qualitatif ou quantitatif du facteur étudié, ainsi que du caractère "normal" ou non de la distribution de chaque variable. Concernant ce dernier point, et malgré la transformation logarithmique, l'inspection des histogrammes montre que l'hypothèse gaussienne n'est pas satisfaisante (ce que confirme les p-values d'un test de normalité tel que celui de Shapiro). Nous avons donc toujours utilisé des tests non-paramétriques (*i.e.* sur les rangs) :

- La variable «genre» étant catégorielle (ou qualitative) à deux classes, le test effectué sera un test de comparaison des médianes de Mann-Whitney-Wilcoxon. Ce dernier permet, pour chaque métabolite, de rejeter ou non l'hypothèse selon laquelle les populations masculine et féminine ont la même valeur médiane.
- Les variables «âge» et «IMC» étant des variables quantitatives, les tests effectués seront des tests de corrélation de Spearman pour déterminer si le coefficient de corrélation linéaire sur les rangs est ou non significativement différent de 0.

	VARIABLES QUALITATIVES						VARIABLES QUANTITATIVES	
	k = 2 classes			k > 2 classes				
Nom du Test <small>(^P) = paramétrique (^{NP}) = non paramétrique</small>	Student (^P)	Student avec correction de Welch (^P)	Mann-Whitney-Wilcoxon (^{NP})	ANOVA (^P)	ANOVA avec correction de Welch (^P)	Kruskal-Wallis (^{NP})	Pearson (^P)	Spearman (^{NP})
Détails du test	Test de comparaison de 2 moyennes	Test de comparaison de 2 moyennes	Test de comparaison de 2 médianes	Test de comparaison de plus de 2 moyennes	Test de comparaison de plus de 2 moyennes	Test de comparaison de plus de 2 médianes	Test de corrélation linéaire	Test de rang
Choix des individus dans la population	Aléatoire	Aléatoire	Aléatoire	Aléatoire	Aléatoire	Aléatoire	Aléatoire	Aléatoire
Valeurs des variables	Classes exclusives*	Classes exclusives*	Classes exclusives*	Classes exclusives*	Classes exclusives*	Classes exclusives*	Pas de valeurs manquantes**	Pas de valeurs manquantes**
Relation entre les variables							linéaire	monotone
Distribution des variables dans la population d'origine	Normale	Normale	NA	Normale	Normale	NA	Normale	NA
Variations des variables	identiques entre elles et à la population	NA	NA	identiques entre elles et à la population	NA	NA		

Table 23. Tests statistiques disponibles et critères d'utilisation

Pour chacun des tests, une correction des tests multiples a été appliquée afin de s'affranchir de l'augmentation du nombre de faux positifs engendrée par la multiplication du nombre de tests. La correction classique choisie (FDR pour "*False Discovery Rate*") garantit que, parmi les variables pour lesquelles l'hypothèse nulle a été rejetée, le nombre de celles qui le sont par erreur (faux positifs) est contrôlé (Benjamini and Hochberg, 1995).

Pour les analyses statistiques multivariées, une transformation logarithmique (base 10) a été effectuée pour chaque matrice de données. Les données fusionnées ont été ensuite centrées et réduites. Des PCA ont été réalisées sur la matrice finale après normalisation afin de vérifier la structure spontanée des jeux de données fusionnées (notamment l'absence de biais analytiques). Des analyses supervisées type OPLS et OPLS-DA ont été ensuite réalisées pour l'étude des effets physiologiques.

Ces méthodes statistiques multivariées nous permettent i) de réduire la dimensionnalité (nombre de métabolites détectés) et d'avoir ainsi une vue d'ensemble de l'ensemble des individus analysés (larges cohortes d'échantillons) et ii) de détecter les éventuelles synergies entre métabolites (contrairement aux approches univariées qui testent les variables indépendamment les unes des autres). Ainsi, ces analyses multivariées fournissent des variables latentes qui sont des combinaisons linéaires des variables d'origine mettant en avant des interactions entre métabolites et donc les phénomènes d'émergence. Cependant, il reste difficile de définir un seuil de significativité des variables à partir de leur poids dans la construction du modèle matérialisé par une valeur de VIP (de l'anglais « Variable Importance in Projection »). C'est pourquoi, des tests statistiques univariés avec

correction des tests multiples ont été réalisés en parallèle. Pour ces derniers, nous avons choisi le seuil de 5% pour le FDR.

Dans ce qui suit, nous présenterons la démarche et les résultats obtenus en analyse métabolomique puis en analyse lipidomique. L'analyse de certaines conditions (mode d'ionisation, colonne chromatographie, série d'analyse) sera détaillée dans le manuscrit. L'annexe 4 contiendra l'analyse des données non présentées dans le corps du manuscrit.

4.2. Analyse métabolomique

Nous aborderons dans cette partie l'analyse des extraits méthanoliques depuis des échantillons de 227 travailleurs du CEA-Saclay en RP, PFPP et ZICpHILIC-ESI(+/-)-MS. Après avoir présenté l'analyse statistique descriptive des 3 séries d'analyse, nous aborderons la démarche de normalisation ayant servi à corriger les signaux intra-série d'analyse puis à la fusion inter-analyses ainsi que son évaluation. Nous décrivons par la suite les résultats obtenus en analyses statistiques pour chacun des trois facteurs physiologiques investigués, puis nous discuterons enfin de leur pertinence biologique au regard d'autres études métabolomiques publiées (Lawton et al; 2008, Saito et al; 2014).

4.2.1. Constitution des séries d'analyse

Tel que précisé précédemment, afin d'assurer une bonne qualité du signal ESI-MS, l'analyse spectrale de la cohorte s'est effectuée en trois séries d'analyse. Les échantillons ont été attribués de façon aléatoire aux trois séries. La distribution des facteurs (la date de prélèvement au centre de SST, l'âge, l'IMC et le genre) a été vérifiée à l'aide de statistiques descriptives. Leurs fréquences entre les séries d'analyse sont équivalentes (Figure 75). Cette répartition démontre que chacune constitue un échantillonnage représentatif de la cohorte initiale (227 individus) ce qui nous prémunit d'introduire un biais lors de la fusion des jeux de données à partir de chacune des séries d'analyse.

A. Séries de prélèvement

B. Age

C. IMC

D. Genre

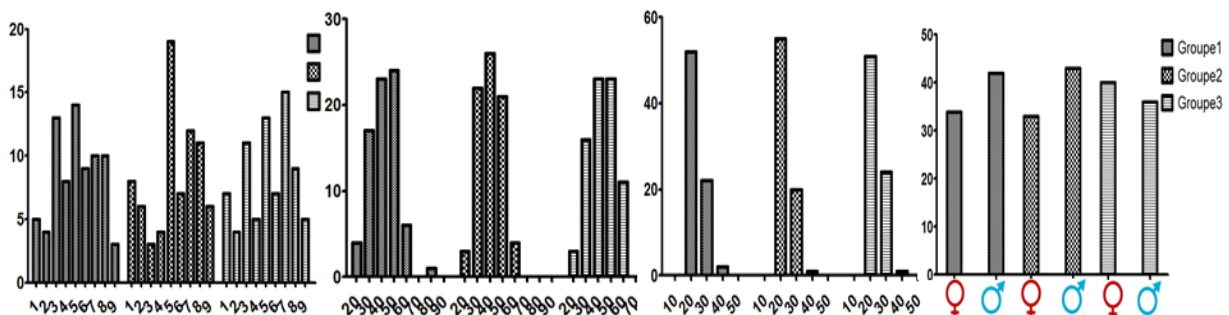


Figure 75. Histogrammes de distribution des facteurs dans les séries d'analyse métabolomique

La fusion des données en analyse métabolomique a été réalisée à bas niveau, soit détection automatique ou intégration des signaux sur l'ensemble des données brutes.

4.2.2. La normalisation des données

L'analyse LC-ESI/MS de grandes séries d'échantillons est un enjeu majeur du fait de la perte de signal liée à l'encrassement de la source ESI mais aussi de la mise en commun de données provenant de séries d'analyse différentes.

4.2.2.1. Correction de l'effet de l'encrassement de la source ESI

Bien que la randomisation des échantillons biologiques au sein d'une séquence permette de répartir aléatoirement le biais analytique dû à l'encrassement de la source, si ce dernier est trop important, il peut potentiellement masquer ou augmenter de manière fictive la variabilité biologique que l'on cherche à étudier. Cette variabilité analytique est mise en évidence en visualisant l'évolution des intensités des différents métabolites et étalons internes notamment dans les échantillons de QC au fur et à mesure de l'ordre de passage.

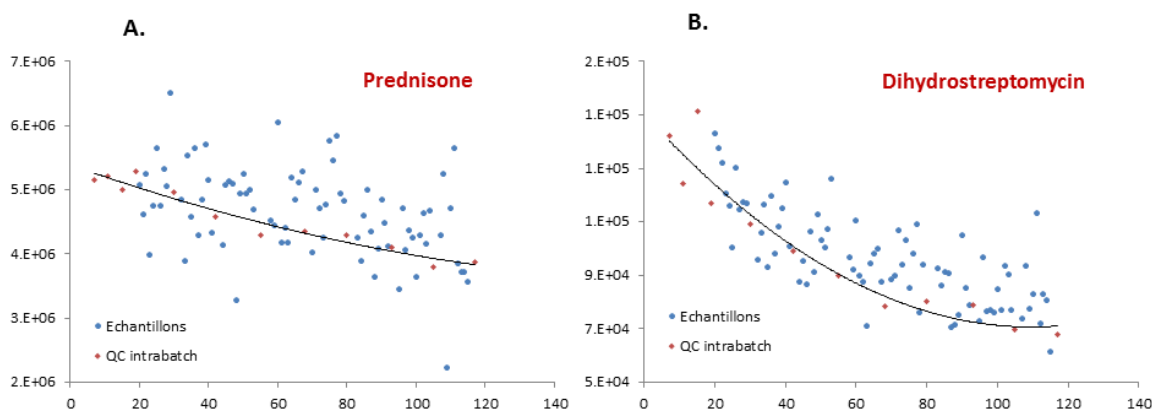


Figure 76. Evolution des intensités des étalons internes « prednisone » et « dihydrostreptomycine » selon l'ordre de passage au cours de l'analyse métabolomique de la première série RPLC-(ESI)-MS

Chaque point représente l'aire sous la courbe du chromatogramme de l'ion extrait (XIC) de l'ion calculé par le module QuanBrowser de XCalibur. La courbe représente l'approximation locale polynomiale modélisant les intensités détectées pour les échantillons de QC inter-séries.

Comme illustré en Figure 76, la dérive analytique des étalons internes suivis (assimilable à la pente de la droite) est nettement observable, elle est monotone (tendance à la baisse) mais n'est pas strictement linéaire. Il en est de même pour les métabolites endogènes. De plus, le biais analytique dépend de la molécule étudiée. Par exemple, la baisse du signal de la dihydrostreptomycine au cours de la séquence d'analyse (CV 29.4%, baisse maximale de 57% du signal) est 2 fois plus importante que celle observée pour la prednisone (CV 11.6%, baisse de 28% du signal). Il convient donc

d'effectuer une correction des données individuellement pour chaque métabolite. Pour ce faire, différentes méthodes de normalisation peuvent être utilisées.

Pour les jeux de données métabolomiques, nous avons réalisé une régression locale de type (LOWESS ou *locally weighted scatterplot smoothing*). Celle-ci s'appuie sur une modélisation polynomiale traçant la courbe qui ajuste au mieux les différentes intensités des QC. Pour chaque métabolite, une régression polynomiale locale de degré 1 a été réalisée à partir des intensités des QC. Les abondances réelles détectées pour les échantillons sont alors divisées par la valeur correspondante sur la courbe de régression afin d'obtenir les abondances normalisées. La droite de régression du nuage de points obtenu après normalisation est alors parallèle à l'axe des abscisses, signifiant que la dérive analytique (assimilable à la pente de la droite) n'impacte plus les abondances des métabolites. Cette technique de normalisation par régression locale est très utilisée pour normaliser les données d'analyses transcriptomiques (Alston et al; 2010, Sirbu, Ruskin, & Crane, 2010).

4.2.2.2. Mise à niveau des réponses inter séries d'analyse

La principale autre source de variation de la réponse de la source ESI est son manque de reproductibilité entre séries d'analyses. En effet, un même échantillon analysé en ESI-MS à des temps différents est détecté à des intensités différentes d'une série d'analyse à une autre.

La fusion de données impliquerait alors de niveler les intensités de chaque variable. Pour ce faire, nous avons exprimé l'intensité de chaque variable dans chaque échantillon à celle de la médiane de la même variable calculée sur les QC inter-séries analysés. Nous avons ensuite ramené cette valeur médiane des QC inter-séries une même valeur dans chaque lot d'analyse à savoir la médiane des médianes (Dunn et al; 2011).

Le processus de normalisation des données (correction du signal intra-analyse et la mise à l'échelle du signal en inter-analyse) est illustré pour l'analyse RPLC-ESI(-)-MS en Figure 77.

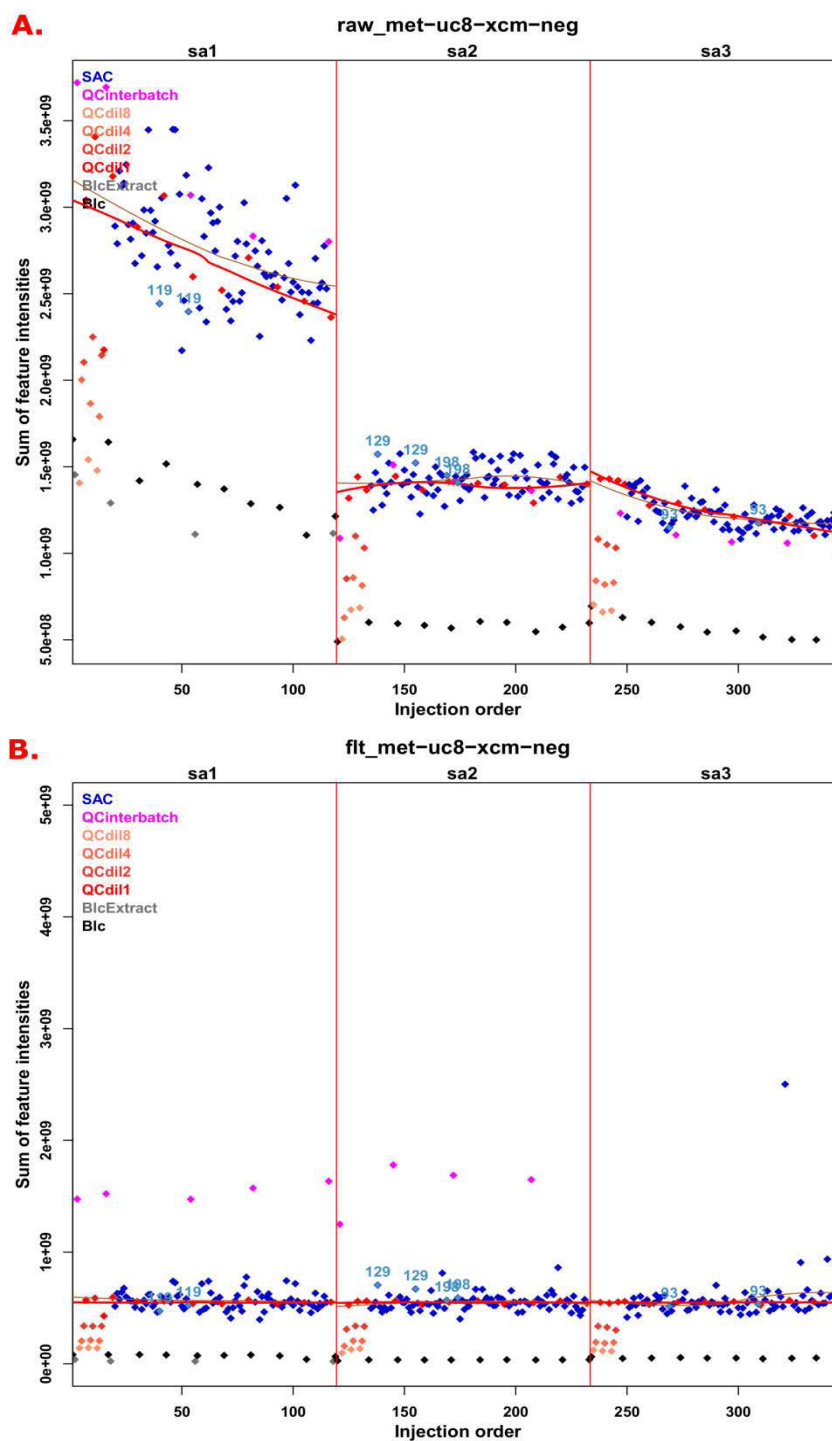


Figure 77. Correction du signal intra série d'analyse métabolomique et fusion des 3 séries d'analyses en RPLC-ESI(-)-MS

Chaque échantillon (quel que soit sa nature) est représenté par un point correspondant à la somme des intensités des variables détectées par XCMS. (A.) intensités brutes. Nous remarquons que l'intensité de départ de la première série d'analyse (sst1) est plus élevée que celles des 2 autres séries. De plus, l'effet de l'encrassement y est plus important (au vu de la pente de la courbe polynomiale). Après correction du signal et fusion des données, la somme des intensités des QC intrasérie et des individus sont au même niveau, de plus on n'observe plus de dérive analytique en fonction de l'ordre d'injection (B.).

4.2.3. Evaluation de la procédure de normalisation des données

4.2.3.1. Correction du signal au cours d'une série d'analyse

La correction du signal pour chaque série d'analyse peut être évaluée par le calcul du coefficient de variation ou encore de la baisse du signal au cours de l'analyse.

Pour exemple, si l'on reprenait les aires des étalons internes considérés en Figure 76, la correction LOESS améliore l'évolution du signal de la dihydrostreptomycine (Figure 78). En effet, les variations sur les QC intra-série sont de 8.3% (contre 29.4% avant normalisation) avec une baisse maximale du signal inférieure à 25% (contre 57% avant normalisation). En revanche, l'algorithme LOESS dégrade légèrement la correction du signal de la prednisone. Le CV augmente à 12.7% au lieu de 11.6% avec une baisse maximale du signal atteignant les 35% au lieu de 28%. Cette baisse de la qualité bien que faible est expliquée par l'aire du 3^{ème} QC intra-série qui semble particulièrement élevée. Le retrait de ce point ramène le CV à 4.6% avec une baisse maximale à 14%.

Ceci démontre que l'algorithme LOESS doit être utilisé avec précaution et éventuellement n'être appliqué qu'aux molécules dont la baisse du signal est avérée.

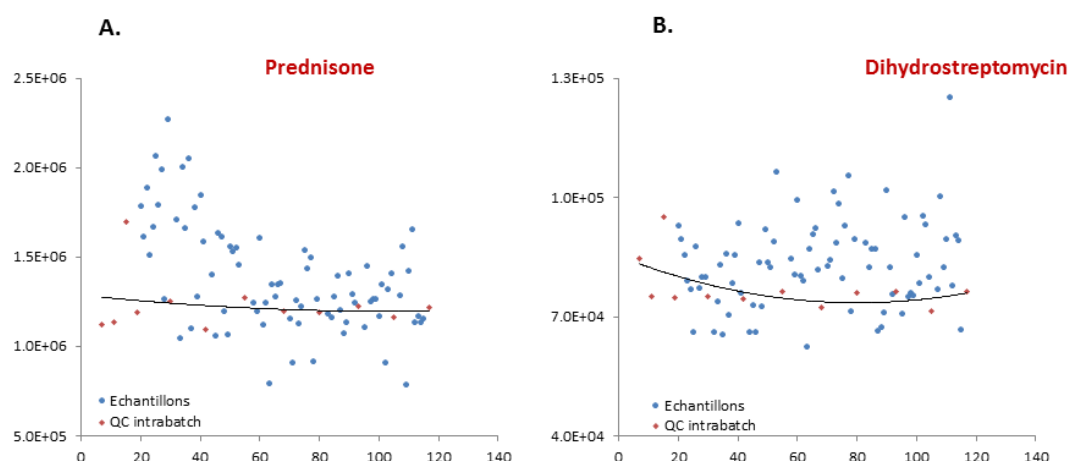


Figure 78. Evolution des intensités corrigées des étalons internes « prednisone » et « dihydrostreptomycine » selon l'ordre de passage au cours de l'analyse métabolomique de la première série RPLC-(ESI-)-MS après normalisation

4.2.3.2. Mise à niveau des séries d'analyse

La fusion des données depuis des analyses différentes repose sur la présence des métabolites au sein des échantillons de QCinter-séries. Du fait que le QCinter-séries choisi soit de la même nature biologique que les échantillons étudiés, cette condition est remplie dans 96 à 99% des variables détectées dans les QCintra-série (Figure 79).

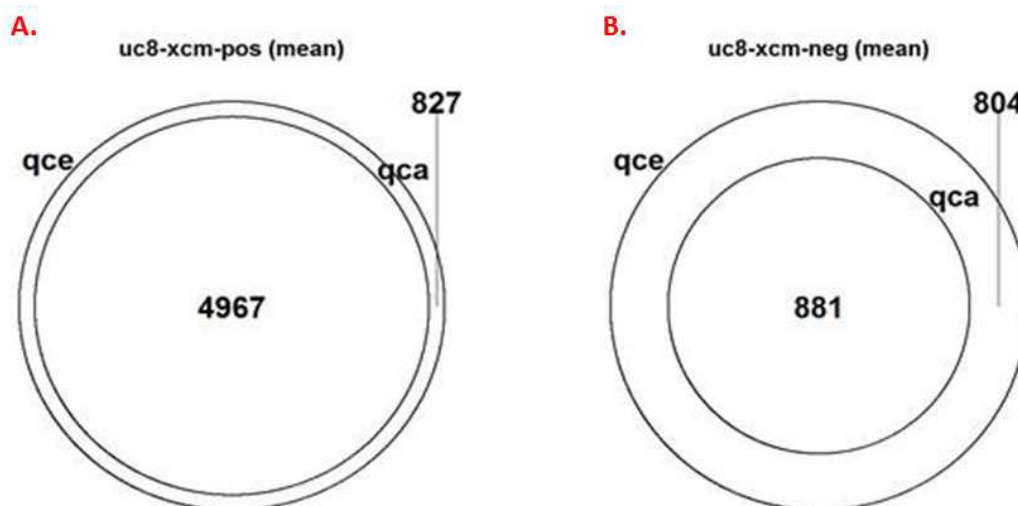


Figure 79. Représentativité des métabolites détectés par les QCinter-séries

qce correspond aux QCinter-séries. qca correspond aux QCintra-série. Les chiffres affichés correspondent au nombre de variables détectées en RPLC-ESI(+/-)-MS dans les échantillons biologiques mais absentes des qce et/ou qca. A noter : le nombre de variables totales détectées dans les échantillons biologiques est de 47912 variables en RPLC-ESI(+)-MS et 19700 variables détectées en RPLC-ESI(+/-)-MS

L'évaluation de la fusion des données a été effectuée dans un premier temps par des approches chimiométriques. En effet, comme démontré en Figure 80.A., il y a une nette séparation entre les échantillons analysés au cours de séries d'analyse différentes. Les séries d'analyses sont alors la première source de variabilité, observable en analyse statistique non supervisée. Après application de la procédure de normalisation et de fusion des données (Figure 80.B), il n'y a plus de distinction entre les séries d'analyse. La carte factorielle des individus montre un regroupement des échantillons analysés dans les différentes séries avec la mise en exergue de certains échantillons dits « aberrants » qui sortent de l'ellipse de Hotelling (correspondant à la région où se situeraient 95% des individus appartenant à une population multinormale de même covariance que celle de l'échantillon) et que nous n'avons pas considéré pour la suite de l'étude statistique (19 individus). La fusion en RPLC-ESI(-)-MS est celle pour laquelle le plus d'échantillons dits « aberrants » ont été retirés du jeu de données fusionné.

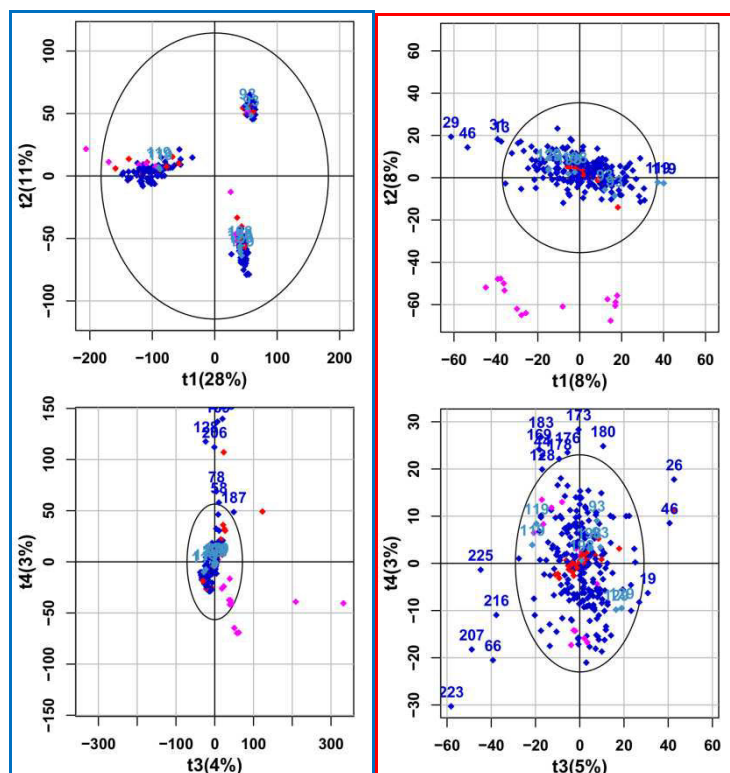


Figure 80. Carte factorielle des individus analysés en RPLC-ESI(-)-MS au cours de 3 séries d'analyse métabolomique différentes sur les axes PCA

En **bleu**, Carte factorielle des individus avant normalisation des données selon la 1^{ère} et 2^{ème} composantes (en haut) puis selon la 3^{ème} et 4^{ème} composante (en bas). On distingue trois groupes d'échantillons séparés selon les deux premières composantes principales. En **rouge**, Carte factorielle des individus après normalisation des données selon la 1^{ère} et 2^{ème} composantes (en haut) puis selon la 3^{ème} et 4^{ème} composante (en bas). Les échantillons se regroupent et ce sans distinction par rapport aux séries d'analyse et quelques individus « aberrants » sortent de l'ellipse de confiance.

Une vue détaillée de la carte factorielle en PCA nous permet de visualiser les répliques inter-analyse que nous avons analysés indépendamment au cours de chaque série d'analyse.

La Figure 81 représente la carte factorielle des individus analysés en RPLC-ESI(-)-MS obtenue en PCA après fusion des séries d'analyse. Les individus répliqués y sont consignés de la même couleur, les étiquettes correspondent au numéro de la série d'analyse. Ces répliques servent à vérifier de façon chimiométrique si après fusion, des données semblables analysées dans des lots différents sont superposables. Les distances observées entre les mêmes échantillons analysés à des temps différents (individus **58**, **134** et **149**) sont jugées satisfaisantes.

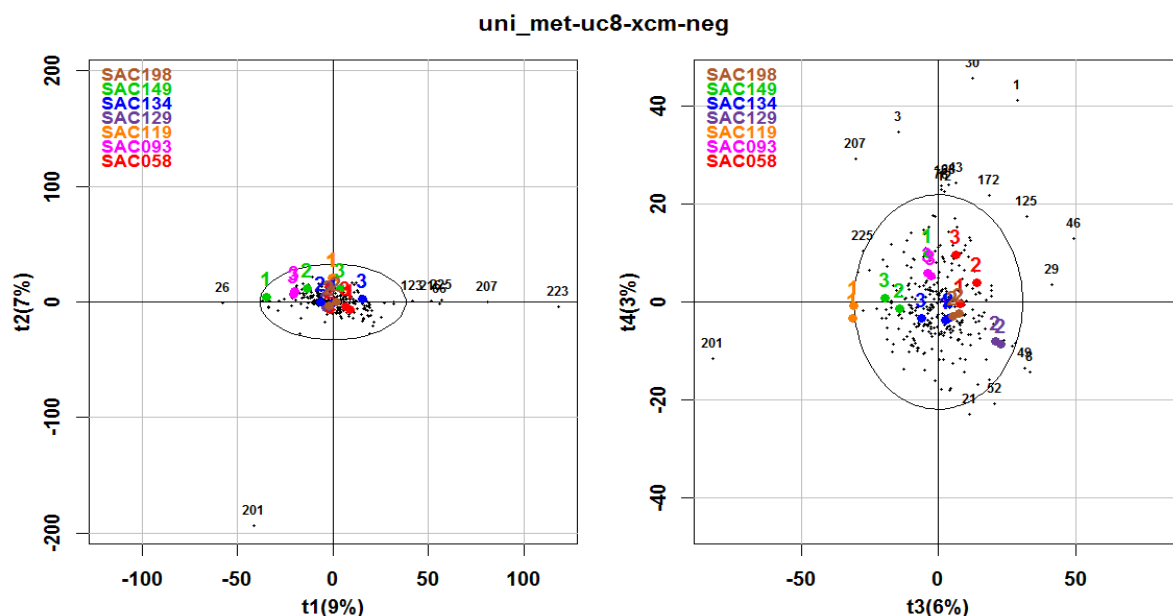


Figure 81. Répliques inter-analyses : validation chimométrique de la fusion des données métabolomique en RPLC-ESI(-)-MS

A gauche, la carte factorielle obtenue en RPLC-ESI(-)-MS selon les composantes PC1 vs PC2 et PC3 vs PC4 à droite. Les couleurs représentent des échantillons différents. Les numéros correspondent à celui de la série d'analyse

La Figure 81 nous permet également d'évaluer de façon chimométrique la répétabilité de la méthode d'extraction des métabolites et de détection des signaux. La visualisation des distances sur la carte factorielle des individus en PCA entre les répliques techniques démontre que ces derniers (individus 119, 198, 129 et 93) sont superposables et valide de ce fait notre démarche analytique et de traitement de données.

Les analyses chimométriques des autres modes d'ionisation et autres colonnes chromatographiques sont présentées en ANNEXE4.

4.2.4. Choix des matrices de données d'intérêt

Fort de nos connaissances quant à la couverture métabolique sérique (voir PARTIE A.CHAPITRE I), nous avons pensé dans un premier temps nous focaliser sur 4 matrices de données parmi les 6 générées à savoir : ZICpHILIC en détection ESI(-), PFPP en détection ESI(+) ainsi que la RP en détection ESI(+/-).

Ne seront alors conservés que les métabolites pour lesquels la variation des aires après normalisation et fusion est comprise dans l'intervalle de +/-30%. En effet si après cette étape de normalisation la variation de la réponse dans les échantillons de QC est encore trop importante, cela indique qu'il existe une source de variabilité autre que celle liée à la dérive analytique ou à la fusion des données et qui risque de biaiser les résultats. Les métabolites dans cette situation ne sont pas gardés. De plus, le rapport des aires des échantillons biologiques et des blancs d'analyse doit être

supérieur à 3. La corrélation de Spearman dans les séries de dilution doit être supérieure à 0.5 et la moyenne des aires dans les échantillons de QC inter-séries et des QC intra-série doit être supérieure à 10^3 en mode d'ionisation négatif et 2×10^3 en mode d'ionisation positif (Table 24) et ce après correction des signaux.

Les échantillons sans intérêt biologique tels les blancs d'analyse et les QC sont retirés de la matrice de données après fusion et application des filtres analytiques. Les réplicas (techniques intra-série d'analyse et inter-séries d'analyse) ont été remplacés au sein du jeu de données fusionné par leur moyenne. De plus, les individus sortant de l'ellipse de Hotelling lors de la projection sur les axes PC et/ou présentant une distribution des intensités de variables atypique (Figure 82) sont également retirés du jeu de données avant de procéder aux analyses statistiques (Table 24). On sait en effet que les techniques multivariées linéaires (PCA et PLS) sont sensibles à la présence de valeurs extrêmes.

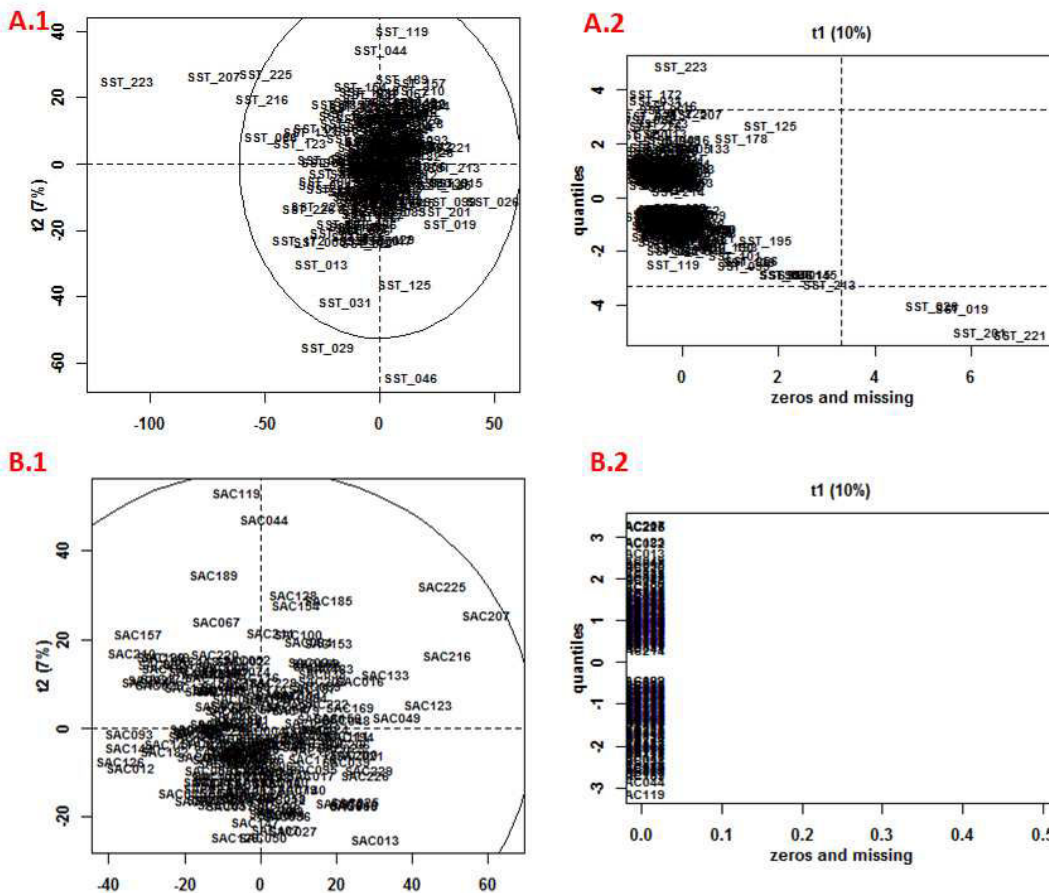


Figure 82. Echantillons atypiques non considérés

A.1 Carte factorielle des individus en PCA après normalisation des données. **A.2** graphique présentant la distribution des intensités des variables. Il y a retrait des échantillons atypiques notamment ceux projetés hors l'ellipse de Hotelling (**B.1**) tel le SST_223 ou ceux présentant une proportion anormalement élevée de valeurs manquantes (e.g. pour lesquels la normalisation de type "loess" est inadaptée; **B.2**) tel le SST_19.

	C8-ESI(-)MS		C8-ESI(+)MS		PFPP-ESI(+)MS		ZICpHILIC-ESI(-)MS	
	Observations	Variables	Observations	Variables	Observations	Variables	Observations	Variables
Données initiales	339	19700	339	47912	340	30840	340	18507
Données filtrées	227	2575	227	1802	227	6984	227	4676
Table finale (statistiques)	204	2575	212	1802	214	6984	205	4676

Table 24. Description des matrices de données fusionnées considérées

Au terme du traitement de données, des milliers de variables analytiquement pertinentes subsistent encore. Au vu de leur nombre et dans un but de trouver des variables discriminantes identifiées, nous avons fait le choix à nouveau de nous focaliser sur les seuls métabolites précédemment identifiés dans le sérum (voir PARTIE A.CHAPITRE I).

De ce fait, seuls 223 métabolites identifiés précédemment (PARTIE A.CHAPITRE I) et remplissant les critères analytiques énumérés ci-dessus ont été considérés pour la suite des analyses statistiques (Figure 83). Leurs aires dans les différents échantillons ont été intégrées à l'aide du module QuanBrowser et ont été consignées dans un tableau fusionné. Une transformation Log10 est effectuée pour se rapprocher d'une distribution gaussienne des données (la variabilité des intensités mesurées par les appareils de spectrométrie de masse étant proportionnelle à la moyenne de ces intensités, la transformation logarithmique est souvent utilisée). De plus, les aires sont centrées et réduites avant concaténation des tableaux afin d'équilibrer les contributions des différentes variables.

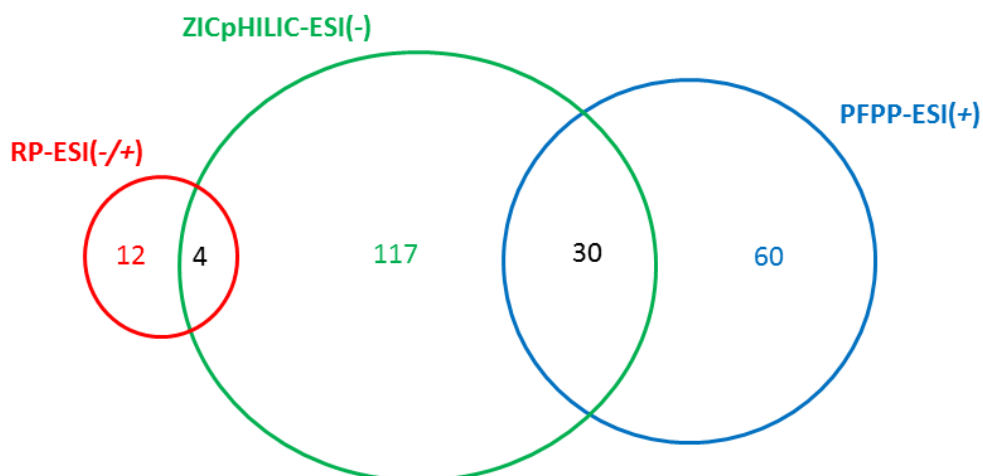


Figure 83. Distribution des métabolites considérés pour les analyses statistiques

Un total de 223 métabolites précédemment identifiés dans le sérum humain ont été concaténés dont 90 métabolites détectés en PFPP-ESI(+)-MS, 151 en ZICpHILIC-ESI(-)-MS et 16 en RP-ESI(-)-MS. Les métabolites détectés dans plusieurs conditions chromatographiques servent à vérifier la cohérence de l'agrégation à travers la similitude des résultats statistiques.

4.2.5. Analyses statistiques

Nous allons dans ce qui suit détailler l'interprétation biologique et statistique univariée des variations observées selon chaque facteur.

4.2.5.1. Variations liées à l'effet du genre

Intéressons-nous à l'effet du genre étudié dans chacune des conditions chromatographiques considérées.

Tout d'abord, 39 des 90 métabolites détectés en PFPP-ESI(+)-MS ressortent discriminants en test de Wilcoxon (dont plus de 60% le sont également en OPLS-DA, voir Figure 84). Les mêmes proportions sont retrouvées pour les 151 métabolites détectés en ZICpHILIC-ESI(-)-MS où 60 d'entre eux diffèrent selon le genre en analyses statistiques univariées après correction des tests multiples (dont plus de 60% le sont également en OPLS-DA, voir Figure 84). Par ailleurs, des 16 métabolites détectées en RP-MS, 11 sont discriminants. De ce fait, chaque colonne chromatographique apporte une information quant à l'effet du genre, ceci s'explique par la diversité des familles chimiques et voies métaboliques sont impactées.

De plus, des 30 métabolites détectés dans les deux systèmes chromatographiques (PFPP et ZICpHILIC), 4 ont des tendances discordantes selon la colonne chromatographique considérée : les isomères de méthyl-histidine dont la résolution chromatographique n'est pas possible en HILIC ainsi que l'hypoxanthine et l'acétyl-alanine qui ne sont pas retenues en PFPP (Figure 84). De même pour l'acide hydroxydécanoïque dont la non-significativité en conditions RP (où il est retenu) est discordante avec le résultat obtenu en HILIC. Dans ces cas, seul le résultat statistique obtenu dans la condition chromatographique optimale est retenu.

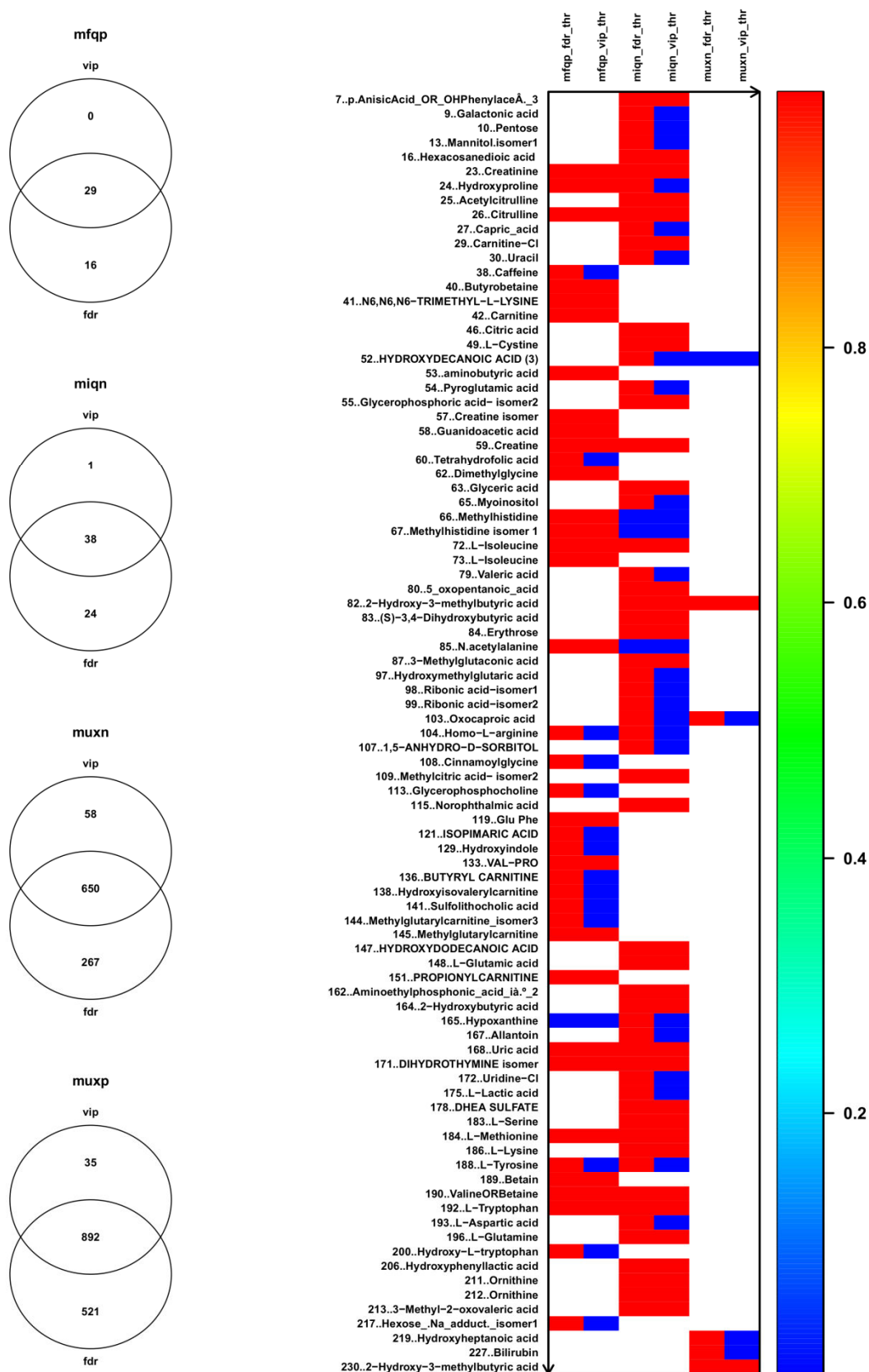


Figure 84. Métabolites discriminants détectés en ZICpHILIC et PFPP selon le genre dans les analyses uni- et multivariées

Dans ce qui suit, nous nous concentrerons de ce fait, sur la significativité des métabolites tels que détectés dans leur condition chromatographique optimale. Au total, 100 métabolites sont impactés par l'effet du genre. De façon générale, le catabolisme est augmenté chez les hommes tel que démontré en Figure 85 pour les carnitines et le glycérol phosphate (qui reflètent le catabolisme des lipides), les acides aminés et polypeptides (catabolisme des protéines), dérivés nucléosidiques (catabolisme des ARNt) ou encore métabolisme énergétique (acide lactique, acide pyruvique). De plus, l'hormone stéroïdienne DHEA-sulfatée est également augmentée chez l'homme.

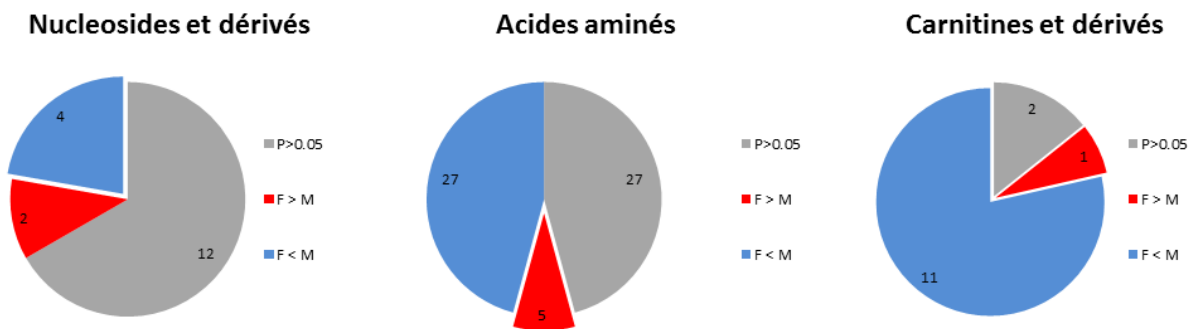


Figure 85. Classes chimiques et effet du genre

En outre, seul 18% des métabolites discriminants selon le genre sont augmentés chez les femmes (Figure 86). Notons la sérine, l'acide aspartique et l'hydroxytryptophane comme acides aminés (dont les variations selon le genre ne sont pas répertoriées dans HMDB) ou encore la caféine et l'acide anisique comme « xénométabolites » (Figure 86).

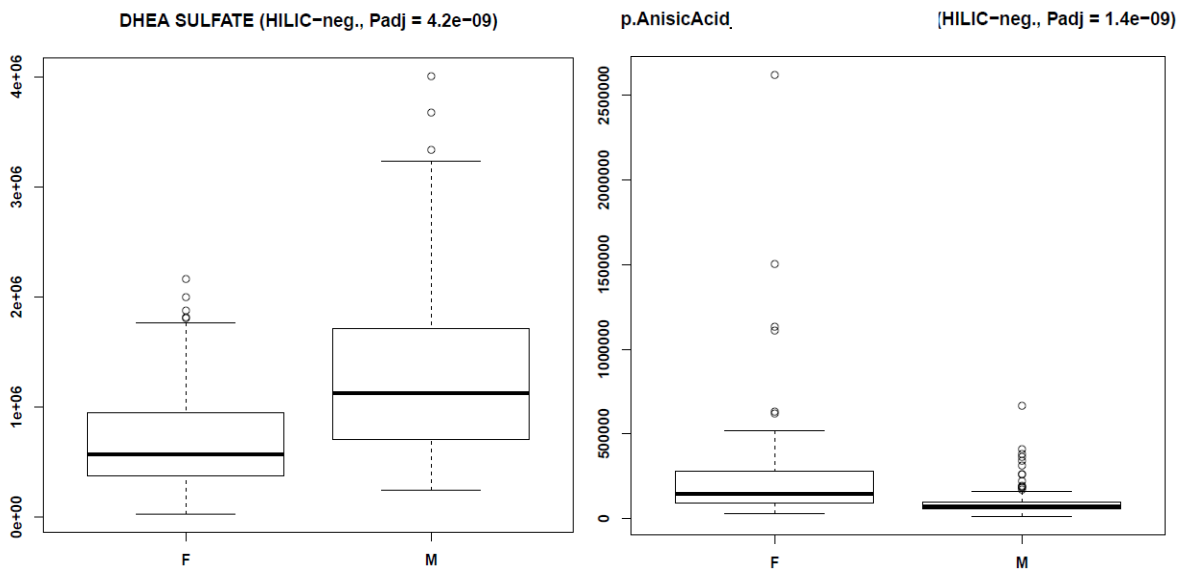


Figure 86. Représentation de l'effet du genre sur les aires des métabolites

Les boîtes à moustaches (ou BoxPlot) sont composées des valeurs des 3 quartiles (25, 50 et 75%). L'espace inter-quartiles IQR représente l'épaisseur de la boîte soit l'espace entre le 1^{er} et le 3^{ème} quartiles. Les individus au-delà de 1.5 IQR

(moustache) au-dessus ou en dessous de la boîte sont représentés par des cercles (et suggèrent des valeurs extrêmes). Si les individus extrêmes ne dépassent pas le 1.5 IQR, alors la moustache est placée à l'individu le plus éloigné. La boîte à moustache est une représentation non-paramétrique utile pour avoir une indication visuelle sur le type de distribution (à un échantillon issu d'une population gaussienne correspondra ainsi une boîte symétrique avec 1,4% des individus au-delà des moustaches) ainsi que sur les individus ayant des valeurs extrêmes devant faire l'objet d'un intérêt particulier. Les concentrations de la DHEA sulfate (hormone stéroïdienne) sont augmentées chez l'homme alors que celles de l'acide p-anisique (xénométabolite des graines d'anis) sont augmentées chez les femmes.

Si l'on s'intéresse plus particulièrement à la voie de l'acide urique, nous détectons une augmentation d'hypoxanthine chez les femmes, pas de variation significative « hommes vs femmes » pour la xanthine puis une concentration plus faible d'acide urique et d'allantoïne chez les femmes. Ceci pourrait se traduire par un plus faible taux d'oxydation (via l'enzyme xanthine oxydase) au cours du métabolisme chez les femmes, ce qui est en accord avec la prévalence de la goutte (hyperuricémie chronique morbide) chez les sujets de sexe masculin (voir Figure 87).

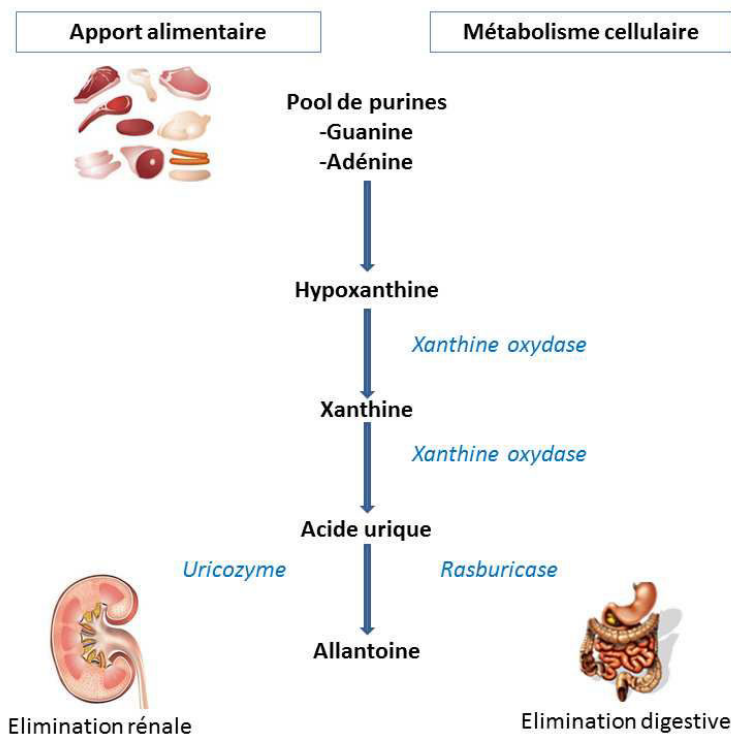


Figure 87. Voie métabolique de l'acide urique

En **bleu** : les enzymes impliquées dans la réaction biochimique

4.2.5.2. Variations liées à l'effet de l'âge

Intéressons-nous maintenant à l'effet de l'âge investigué dans chacune des conditions chromatographiques considérées. 15 des 90 métabolites détectés en PFPP-ESI(+)-MS présentent une corrélation de Spearman significative alors que 42 le sont en ZICpHILIC-ESI(-)-MS. Ce dernier

couplage apporte plus d'information biologique et son modèle de statistiques multivariées (OPLS) démontre une meilleure qualité d'explication et de prédiction (R2Y 0.763 et Q2Y 0.581) que celle du jeu de données PFPP ou du jeu de données fusionné.

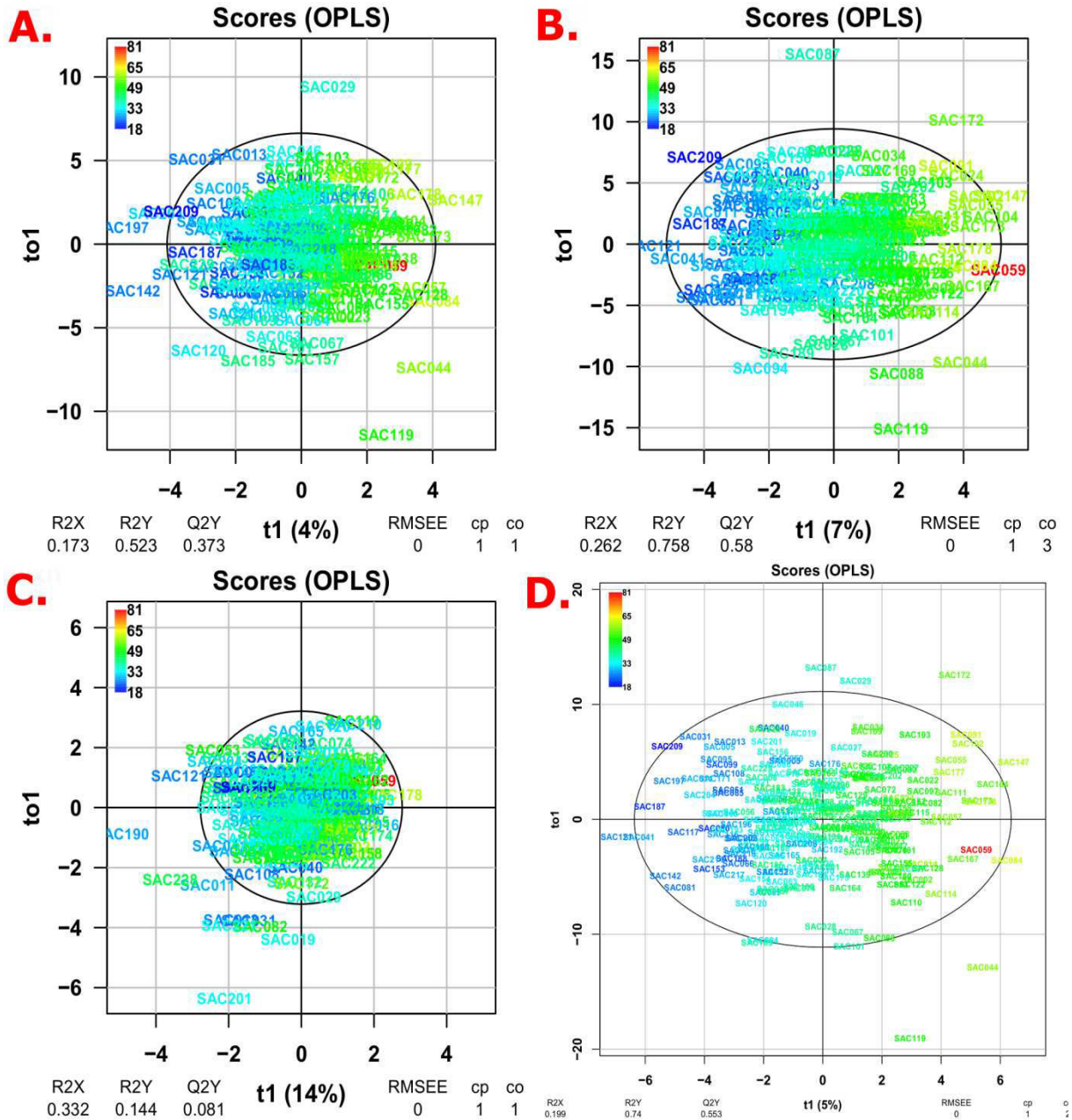


Figure 88. Carte factorielle des individus selon les conditions chromatographiques PFPP-ESI(+)-MS, ZICpHILIC-ESI(-)-MS et leur fusion dans le modèle OPLS construit en fonction de l'âge

A : jeu de données en PFPP-ESI(+) après intégration manuelle des aires sur le module Quanbrowser. B : jeu de données en ZICpHILIC-ESI(-) après intégration manuelle des aires sur le module Quanbrowser. C : jeu de données en RP-ESI(-) sélectionné depuis les données XCMS. D : Agrégation des jeux données PFPP-ESI(+)ZICpHILIC-ESI(-) et RP-ESI(-).

Si l'on se concentre sur la significativité des métabolites tels que détectés dans leur condition chromatographique optimale, 56 métabolites au total sont impactés par l'âge.

La détoxification de l'ammoniaque au cours du cycle de l'urée semble être exacerbée chez les personnes âgées (augmentation des métabolites produits dans la mitochondrie à savoir la citrulline, et l'ornithine mais aussi des métabolites annexes tels que l'acétylcitrulline et l'acide acétylglutamique qui active l'enzyme carbamoylphosphate synthase). Par contre, on ne détecte pas de changement pour les acides aminés impliqués dans la partie cytosolique du cycle à savoir l'acide aspartique et l'arginine. Le métabolisme de la caféine est également l'une des voies les plus impactées par l'âge (augmentation de la caféine, l'acide 1,7-diméthylurique, la théophylline) avec des augmentations de taux pouvant atteindre jusqu'à 3 fois leur valeur à un plus jeune âge. Ceci indique une augmentation importante de la demi-vie de la caféine, qui peut être expliquée par une diminution de l'activité des cytochromes P450 (notamment CYP1A2) avec l'âge (Chung et al; 2000, Tang, 1995). Des modifications de la demi-vie de la caféine ont déjà été décrites pour des conditions physiologiques particulières (femmes enceintes (Ortweiler et al; 1985)) ou encore pour des raisons iatrogènes (prises de contraceptifs (Meyer et al; 1991)) et prise médicamenteuse (Bolton & Null, 1981). Ces résultats ont également été démontrés dans les urines humaines de cette même cohorte de salariés du CEA-Saclay. En effet, une nette augmentation selon l'âge de la caféine et de ses métabolites (paraxanthine, 1-méthylxanthine, l'acide 1-méthylurique et l'acide 1,7-diméthylurique) a été observée dans les urines humaines. De façon générale, les concentrations des xénométabolites sont augmentées avec l'âge telles que celles de l'acide 2,5-dihydroxybenzoïque (métabolite de l'acide salicylique), l'acide hippurique (produits phénoliques), l'acide gluconique (contenu dans les fruits, les miels et les vins et souvent utilisé comme additif alimentaire), l'acide quinique (café et tabac) et la trigonelline (métabolisme de la vitamine B3, café). Nous avons également observé l'augmentation des concentrations des produits de la flore intestinale telle que la phenylacétylglutamine (catabolite de la phénylalanine).

Seuls 20% des métabolites impactés diminuent avec l'âge. On citera la DHEA-sulfate, l'acide glycérophosphorique et quelques acides gras à courte chaîne (l'acide hydroxydodecanoïque, l'acide hexacosanedioïque, et l'acide hexadecanedioïque) ou encore le fucose ou 1,5-anhydrosorbitol qui est augmenté chez les sujets les plus jeunes de la cohorte. Ce dernier est apporté par l'alimentation et subit une réabsorption rénale compétitive avec le glucose et peut de ce fait être indicateur de l'équilibre glycémique à court terme. Son augmentation sérique chez les individus jeunes impliquerait une meilleure régulation de la glycémie.

4.2.5.3. Variations liées à l'effet du BMI

Cinquante deux concentrations de métabolites sont impactées par l'effet du BMI dont les acides aminés, polypeptides mais aussi les carnitines. Ceci démontre une augmentation du catabolisme chez les individus en surpoids ou obèses (voir Figure 89). En effet, les associations entre BMI et concentrations en acides aminés peuvent être dues à une insulino-résistance entraînant la libération d'acides aminés dans le sang afin de réguler l'appétit ; ces acides aminés étant des précurseurs de neurotransmetteurs (Moore et al; 2014). Dans ce sens, leurs concentrations cérébrales doivent être mesurées. De plus, cette insulino-résistance entraîne également une augmentation de la bêta-oxydation avec élévation de l'acide hydroxybutyrique et des carnitines. A l'inverse, 6 métabolites voient leurs concentrations diminuer avec l'augmentation du BMI (l'acide indolepropionique, l'acétylglycine, l'acide caprique, l'acide hexadécanedioïque, l'acide hydroxydecanoïque et l'asparagine) (Figure 89).

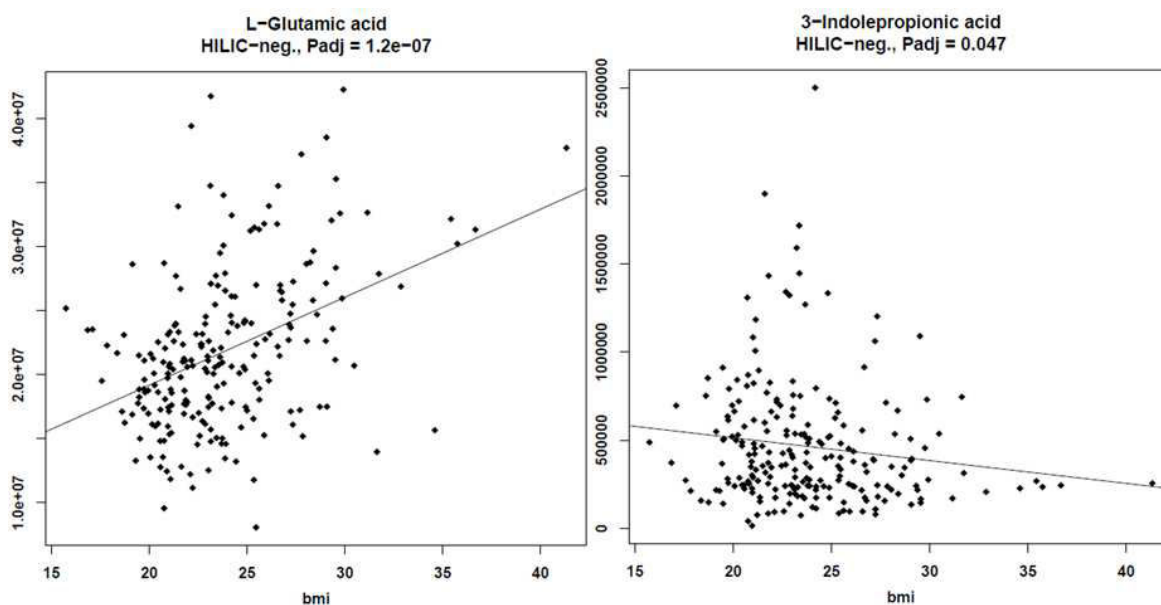


Figure 89. Variations des métabolites selon le BMI

La valeur de p value est celle du test de corrélation de Spearman

4.2.6. Discussion

Les empreintes métaboliques sériques sont impactées par les effets du genre, de l'âge et du BMI. En effet, bien que les variations qui leur sont dues soient souvent négligées, les concentrations de 64% (143 des 223) des métabolites identifiés sont modifiées par au minimum un des facteurs physiologiques étudiés. L'effet du genre apporte la plus grande variabilité avec 100 métabolites différenciant entre les hommes et les femmes (Figure 90). La deuxième source de variabilité est l'âge (56 métabolites discriminants) puis le BMI (52 métabolites discriminants). Notons que seules les variations de concentrations de 11 métabolites le sont exclusivement liées au BMI.

Cette répartition diffère de celles précédemment publiées dans le cadre d'analyse de cohortes. En effet, Lawton et coll. décrivent l'âge comme étant le facteur le plus influant dans le plasma. En utilisant les mêmes plateformes analytiques (combinaison LC-MS et GC-MS ; (Evans et al; 2009)), Saito et coll. ont récemment confirmée cette tendance dans le plasma mais aussi dans le sérum (Saito et al; 2014). Ceci démontre que les méthodes analytiques utilisées peuvent orienter les résultats avec la détection de métabolites différents ou de résolutions d'isomères ouvrant des permettant l'analyse d'autres voies métaboliques.

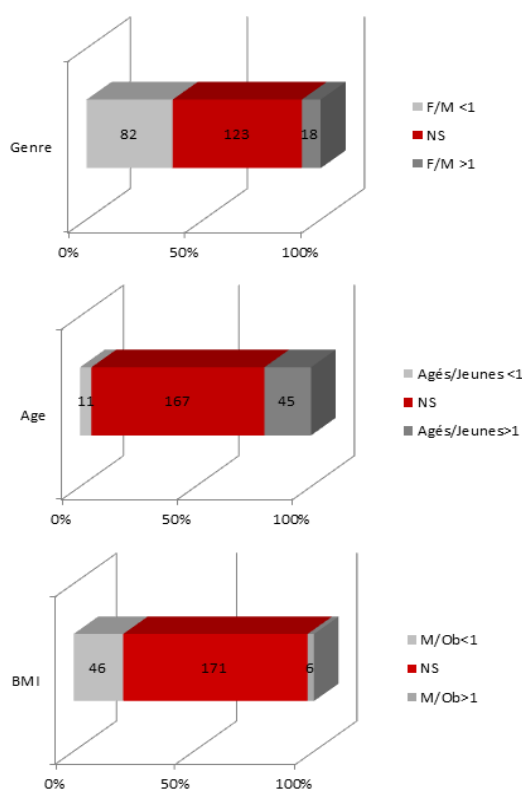


Figure 90. Métabolites discriminants et leur tendance par facteur physiologique étudié

De façon plus précise, nous nous sommes focalisés sur les 89 métabolites identifiés communs à la cohorte saclaysienne et la cohorte sérique étudiée par Saito et coll et à leur significativité selon l'âge et le genre. Pour ce faire, nous avons subdivisé la cohorte saclaysienne en 4 sous-groupes de fond biologique commun à savoir : les jeunes femmes, les jeunes hommes (<35ans), les femmes âgées et les hommes âgés (>55ans) et réalisé des tests de comparaison de moyenne, tout comme décrit par Saito et coll.. Les résultats obtenus sont illustrés dans la Figure 91.

Nous retrouvons plus de 80% de concordance dans nos résultats sauf pour l'effet de l'âge chez les femmes et l'effet du genre chez les jeunes pour lesquels la discordance peut atteindre les 34% et ce dans différentes familles chimiques. Par ailleurs, Saito et coll. ont justement observé une grande variation du métabolisme chez les jeunes femmes par rapport aux groupes homologues (jeunes

hommes ou femmes âgées). Ceci peut suggérer que la variabilité extrême observée par Saito et coll. pour ce groupe d'individus est dépendant de la cohorte étudiée et n'est peut-être pas spécifique du facteur physiologique investigué.

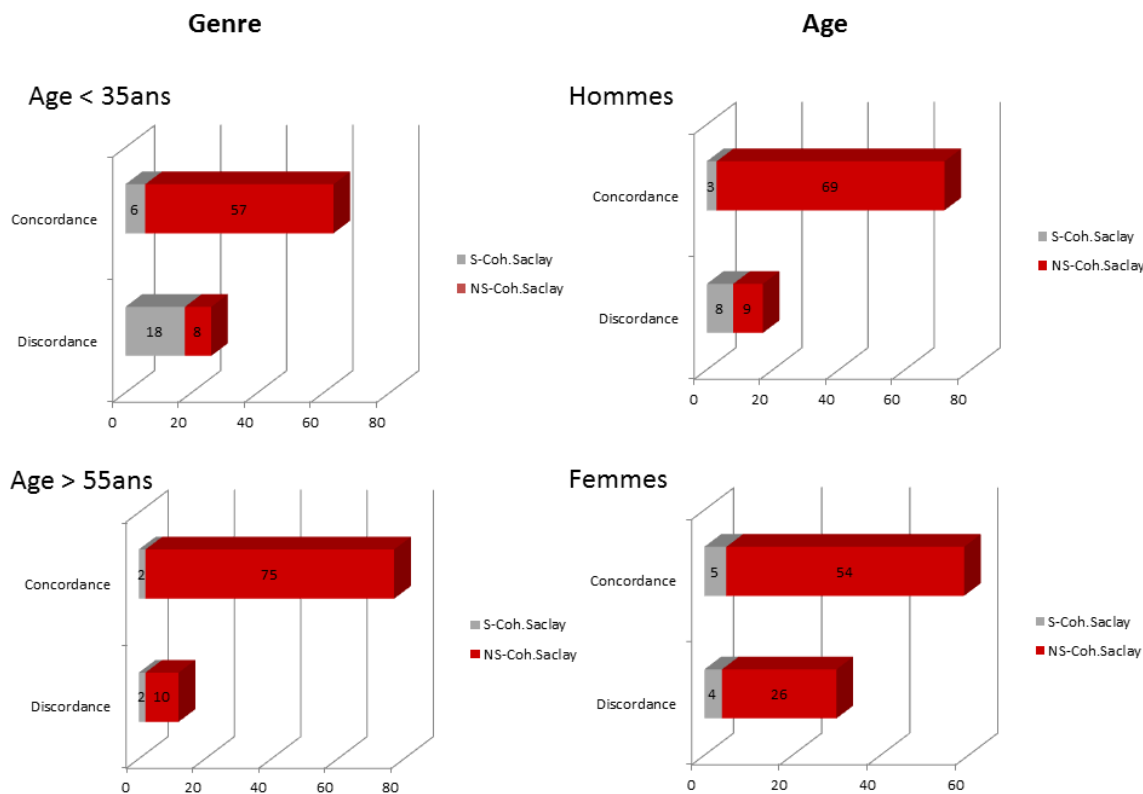


Figure 91. Comparaison des résultats statistiques obtenus pour la cohorte saclaysienne versus les résultats de Saito et coll.

Les disparités dans les résultats rendus peuvent être la conséquence de démarches expérimentales différentes (méthodes d'analyse, méthodes statistiques), de tailles de cohorte différentes, de matrices biologiques différentes (urines, plasma, sérum) ou encore de biais de recrutement des individus (notamment maladies sous-jacentes, hygiène de vie, tabagisme...). Néanmoins, nos résultats dans les sera sont cohérents avec ceux obtenus dans les urines des mêmes individus analysées dans les conditions RP (thèse Aurélie ROUX 2008-2011). En effet, les mêmes tendances sont retrouvées entre les urines et sera des mêmes individus notamment la prédominance de l'influence du genre. Ce dernier étant marqué par les hormones stéroïdiennes, leurs produits conjugués ainsi que le métabolisme énergétique (catabolisme des ARNt, des lipides ou des protéines). De plus, l'effet de l'âge dans les matrices biologiques reflète une baisse globale du métabolisme notamment celui des xénométabolites. En revanche, alors que seuls 4 métabolites urinaires ont été décrits comme impactés par le BMI, cet effet semble plus marqué dans le sérum avec un impact sur 52 métabolites notamment ceux retrouvés dans les urines.

Cette cohérence entre les résultats sériques et urinaires depuis les échantillons de mêmes individus ainsi que les disparités rencontrées entre des cohortes différentes suggère une grande influence de la variabilité interindividuelle dans les résultats obtenus.

4.3. Analyse lipidomique

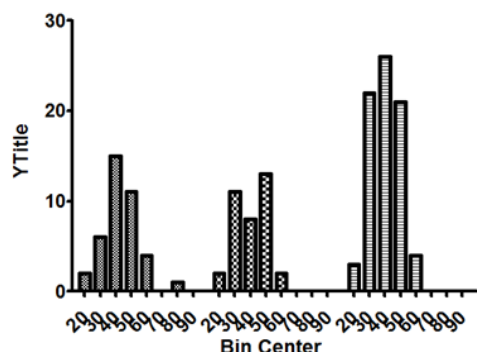
Nous aborderons dans cette partie l'analyse des extraits lipidiques totaux (extraction de Bligh & Dyer) depuis des échantillons de 160 salariés du CEA-Saclay en RP-ESI(+/-)-MS.

4.3.1. Constitution des séries d'analyse

L'analyse LC-HRMS en lipidomique s'est réalisée en 3 séries d'analyse. Les deux premières ont été jumelées à une analyse de patients diabétiques –qui feront l'objet de la PARTIE A.CHAPITRE II-. De ce fait, 40 et 40 individus ont été analysés dans les première et deuxième séries d'analyse respectivement. La troisième série d'analyse correspond à celle décrite pour l'analyse métabolomique.

A nouveau, la Figure 92 démontre que chaque série d'analyse constitue un échantillonnage représentatif de la cohorte globale.

A. Age



B. BMI

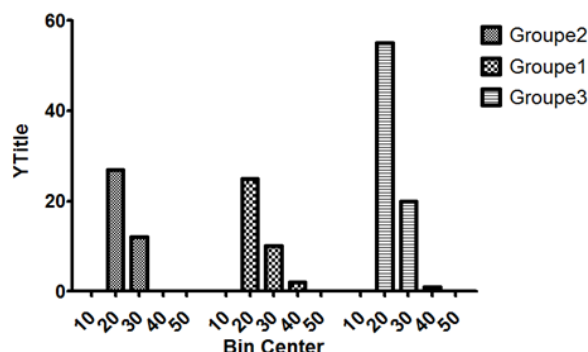


Figure 92. Histogrammes de distribution des facteurs dans les séries d'analyse lipidomique

4.3.2. Fusion intermédiaire des matrices de données en analyse lipidomique

Une source de variabilité analytique a été fortuitement introduite dans le jeu de données. En effet, La 1^{ère} série d'analyse a été réalisée sur une colonne chromatographique de type C8 provenant d'un lot de fabrication différent que celle ayant servi à l'analyse de 2^{nde} et la 3^{ème} séries d'analyse. Nous avons opté pour une fusion des données en deux temps : d'abord il y a fusion de « bas niveau » des fichiers bruts acquis lors des 2^{ème} et 3^{ème} séries d'analyse. La matrice XCMS générée est annotée, et seules les variables annotées sont ensuite agrégées à celles détectées et annotées dans la première série d'analyse « fusion intermédiaire » (voir Figure 94). Les molécules communes annotées dans les 2 matrices de données présentaient un décalage de temps de rétention variable entre -10 et

+25 secondes (Figure 93). Les décalages de temps de rétention dépassant ces limites ont été notés mais sont également associés à des précisions en mesure de masse jugées non satisfaisantes (Figure 93.B). Ces molécules ont donc été retirées de la matrice fusionnée.

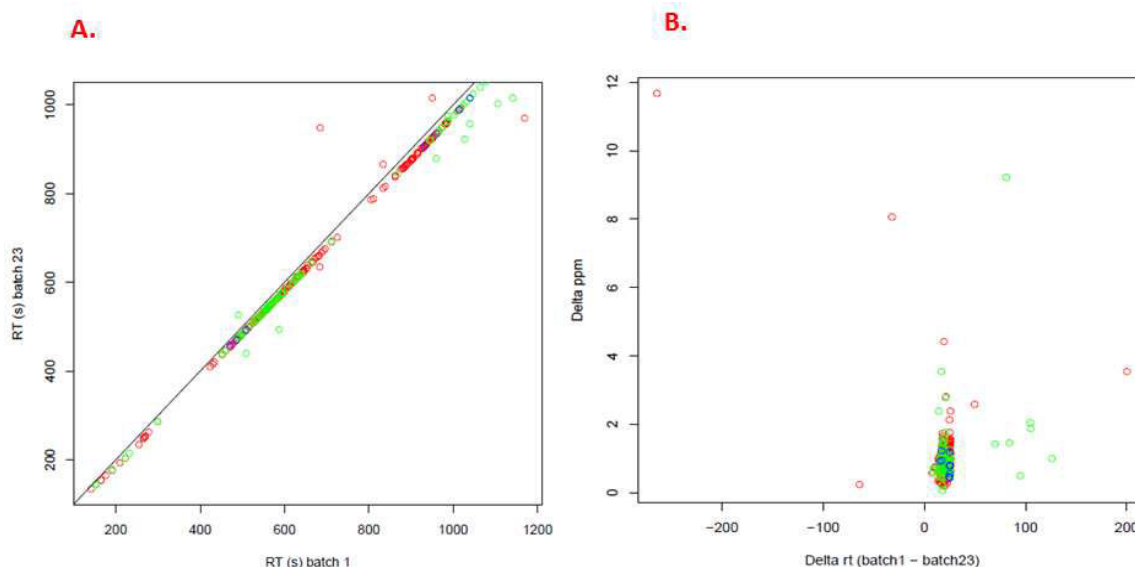


Figure 93. Evaluation du décalage de temps de rétention observé en analyse lipidomique entre la 1^{ère} série d'analyse et les 2 dernières séries d'analyse (mode d'ionisation négatif)

Les cercles représentent des molécules formellement annotées dans les 2 matrices de données. En rouge, celles pour lesquelles un seul temps de rétention est observé. En vert, les annotations présentant 2 temps de rétention. En bleu, les annotations pour lesquelles 3 isomères sont détectés. En (A.), le graphique illustre la relation entre le temps de rétention observé pour les 2^{ème} et 3^{ème} séries d'analyse contre celui observé pour la 1^{ère} série d'analyse sur la base d'une annotation commune. La relation semble être linéaire. En (B.), la différence de temps de rétention observée est évaluée selon la précision de masse calculée lors de l'annotation de la molécule dans la 1^{ère} série d'analyse.

Le décalage de temps de rétention, bien que réduit, peut entraver l'alignement des pics, notamment pour ceux correspondant à des isomères et qui peuvent ainsi se voir regroupés par erreur sous une même variable (tel qu'illustré en Figure 94).

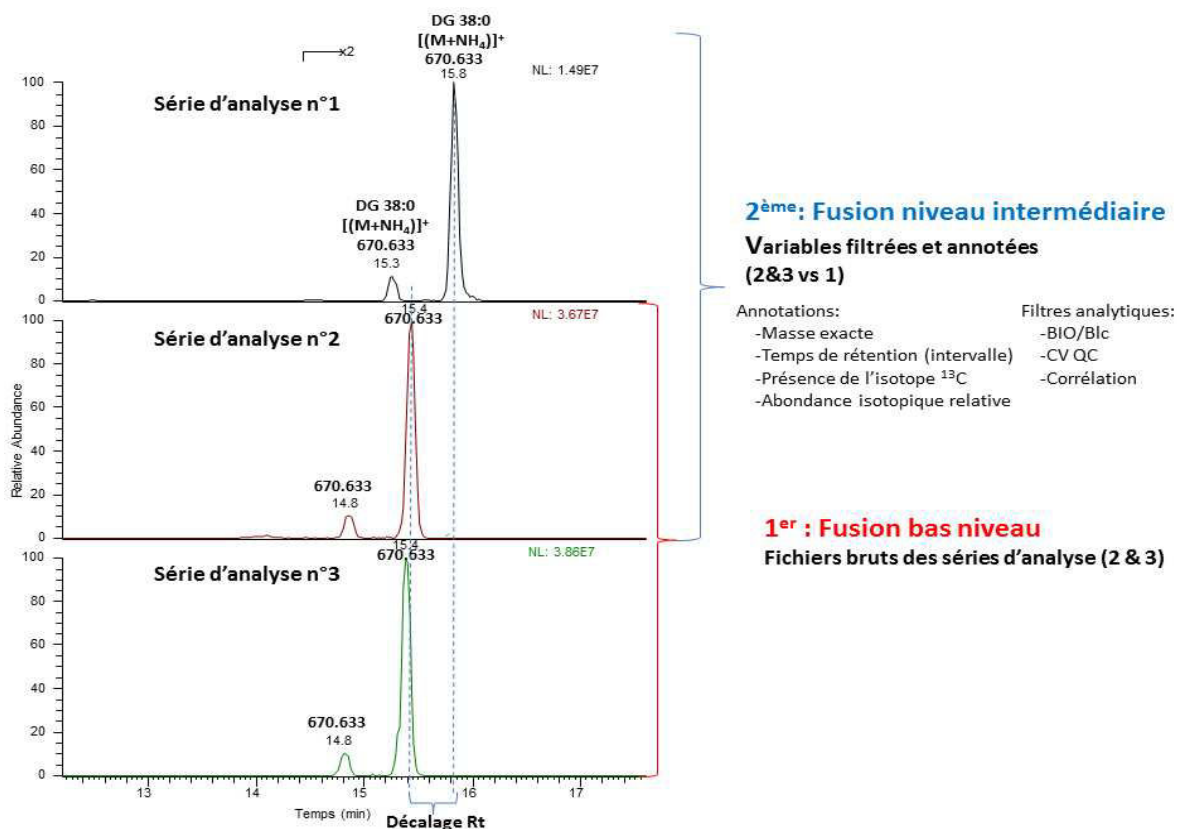


Figure 94. Fusion des données en analyse lipidomique : Exemple du DG 38 :0

La figure montre les XIC de l'ion m/z 670.633 obtenus sur un échantillon de QCinterséries lors de l'analyse LC-MS des 3 séries d'analyse. 2 isomères sont détectés à 15.3 et 15.8mn dans la 1^{ère} série puis à 14.8 et 15.4mn dans les séries d'analyse 2 et 3. Ce décalage en temps de rétention (dû à un changement de lot de colonne chromatographique) entrave l'alignement des variables en temps de rétention après détection automatique des signaux : l'intervalle de temps de rétention dépasse les limites fixées ou encore, l'alignement de l'isomère 1 de la 1^{ère} série d'analyse avec l'isomère 2 des séries d'analyse 2 et 3 peut survenir fortuitement.

Pour y pallier, nous avons modélisé le décalage de temps de rétention au cours de l'analyse RPLC-MS sur les molécules ne présentant pas d'isomère puis appliqué le facteur correctif à l'ensemble des molécules annotées. Ainsi les temps de rétention observés pour la matrice 2&3 ont été corrigés avant fusion avec la matrice 1.

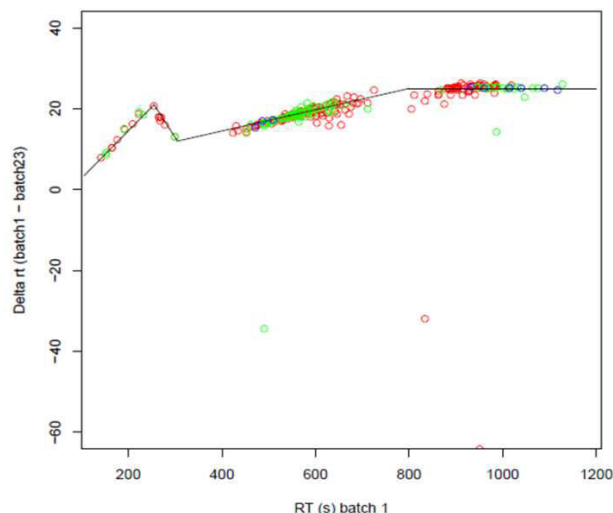


Figure 95. Modélisation du décalage de temps de rétention au cours de l'analyse lipidomique RPLC-ESI(-)-MS

En rouge, celles pour lesquelles un seul temps de rétention est observé. En vert, les annotations présentant 2 temps de rétention. En bleu, les annotations pour lesquelles 3 isomères sont détectés. 4 pentes de droite ont été considérées pour la modélisation du décalage de temps de rétention et ce selon le temps d'éluion.

En analyse lipidomique, la fusion s'est donc faite en 2 temps. Une fusion de bas niveau entre la 2^{ème} et la 3^{ème} séries d'analyse ayant été analysées à la suite après nettoyage de la source. La matrice jointe 2&3 ainsi que la matrice de données de la 1^{ère} série d'analyse sont filtrées et annotées. Puis une fusion intermédiaire des matrices annotées est effectuée sur la base des annotations

4.3.3. La normalisation des données

Les données ont été normalisées en intra puis en inter-séries d'analyse (Voir PARTIE A.CHAPITRE I.4.2.2)

Une vue détaillée nous permet également de visualiser les répliques inter-analyse que nous avons analysés indépendamment au cours de chaque série d'analyse.

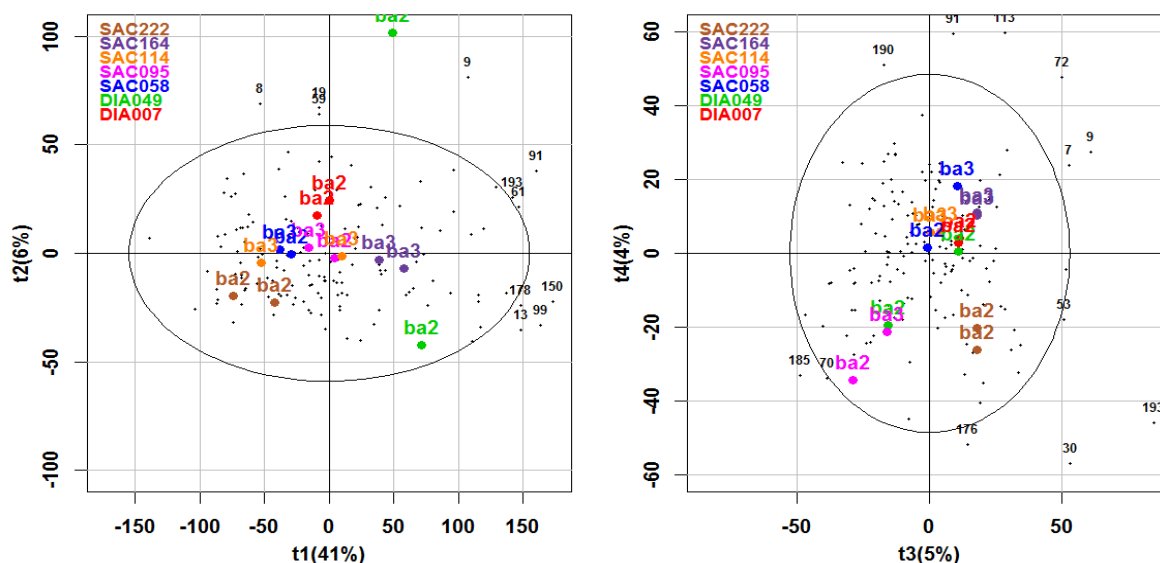


Figure 96. Répliques inter-analyses : validation chimiométrique de la fusion des données lipidomiques

A gauche, la carte factorielle des individus obtenue en RPLC-ESI(-)-MS selon les composantes PC1 vs PC2 et PC3 vs PC4 à droite après fusion de la 2ème et de la 3ème séries d'analyse. Les couleurs représentent des échantillons différents. Les numéros correspondent à celui de la série d'analyse

Les distances observées entre les mêmes échantillons analysés à des temps différents (individus 58 et 95) sont jugées satisfaisantes.

La Figure 96 nous permet également d'évaluer de façon chimiométrique la répétabilité de la méthode d'extraction des métabolites et de détection des signaux. La visualisation des distances sur la carte factorielle de PCA entre les répliques techniques démontre que ces derniers (individus 114, 222 et 164) sont superposables.

4.3.4. Analyses statistiques

Les profils lipidomiques étudiés comprennent 915 espèces lipidiques appartenant à 20 classes et sous-classes différentes (Table 25). L'interprétation des résultats se fait au niveau des classes et sous-classes et non pas par espèce lipidique. En effet, les voies biochimiques de synthèse suggèrent globalement l'implication des mêmes enzymes (pas toujours les mêmes isoformes) pour l'anabolisme ou le catabolisme d'espèces présentant la même tête polaire. Bien qu'il y ait une disparité des proportions d'espèces discriminantes selon l'approche statistique considérée (uni- ou multivariée), les valeurs de FDR sont confirmées par les valeurs de VIP.

Nous allons dans ce qui suit détailler l'interprétation biologique et statistique (tests univariés) des variations observées selon chaque facteur physiologique étudié.

Classes et sous-classes	Total	Age			BMI			Genre		
		q val %	q val	VIP >1	q val %	q val	VIP >1	q val %	q val	VIP >1
Cer	8	37.5	3	6	62.5	5	5	37.5	3	4
HexCer	5	0.0	0	0	0.0	0	0	0.0	0	2
diHexCer	13	0.0	0	3	15.4	2	3	15.4	2	5
Su	13	15.4	2	4	23.1	3	6	15.4	2	2
SGL	12	25.0	3	6	0.0	0	4	8.3	1	6
SM	37	0.0	0	25	13.5	5	12	5.4	2	8
LPC	36	0.0	0	5	0.0	0	1	30.6	11	21
PC	208	0.0	0	56	8.2	17	40	6.7	14	61
LPE	13	0.0	0	3	0.0	0	1	38.5	5	8
PE	87	9.2	8	37	4.6	4	23	12.6	11	25
LPG	2	0.0	0	1	0.0	0	1	50.0	1	2
PG	16	37.5	6	10	0.0	0	8	0.0	0	1
LPS	10	0.0	0	1	0.0	0	2	10.0	1	7
PS	126	5.6	7	36	5.6	7	45	7.9	10	42
LPI	3	0.0	0	0	0.0	0	3	0.0	0	0
PI	59	13.6	8	34	3.4	2	24	6.8	4	16
FFA	13	0.0	0	5	0.0	0	2	7.7	1	6
DG	22	0.0	0	13	27.3	6	9	9.1	2	8
TG	198	0.0	0	88	41.9	83	115	20.2	40	90
CE	34	0.0	0	17	11.8	4	12	8.8	3	13

Table 25. Espèces lipidiques annotées dans les jeux de données fusionnés de la cohorte saclaysienne et impact des facteurs physiologiques

4.3.4.1. Variations liées à l'effet du genre

Comme présenté dans la Table 25, l'effet du genre est le plus marqué des 3 facteurs physiologiques étudiés en termes de couverture métabolique (et non pas en nombre d'espèces). En effet, 18 classes et sous-classes de lipides des 20 observées présentent une différence de concentration de leurs espèces selon le sexe de l'individu. Les espèces PC, diHexCer, sulfatides, PI, les formes acyles des PE et majoritairement les PS sont augmentées chez les femmes alors que les formes lysoglycérophospholipides, TG, DG, CE, Cer et les formes « ether » des PE sont quant à elles, augmentées chez les hommes.

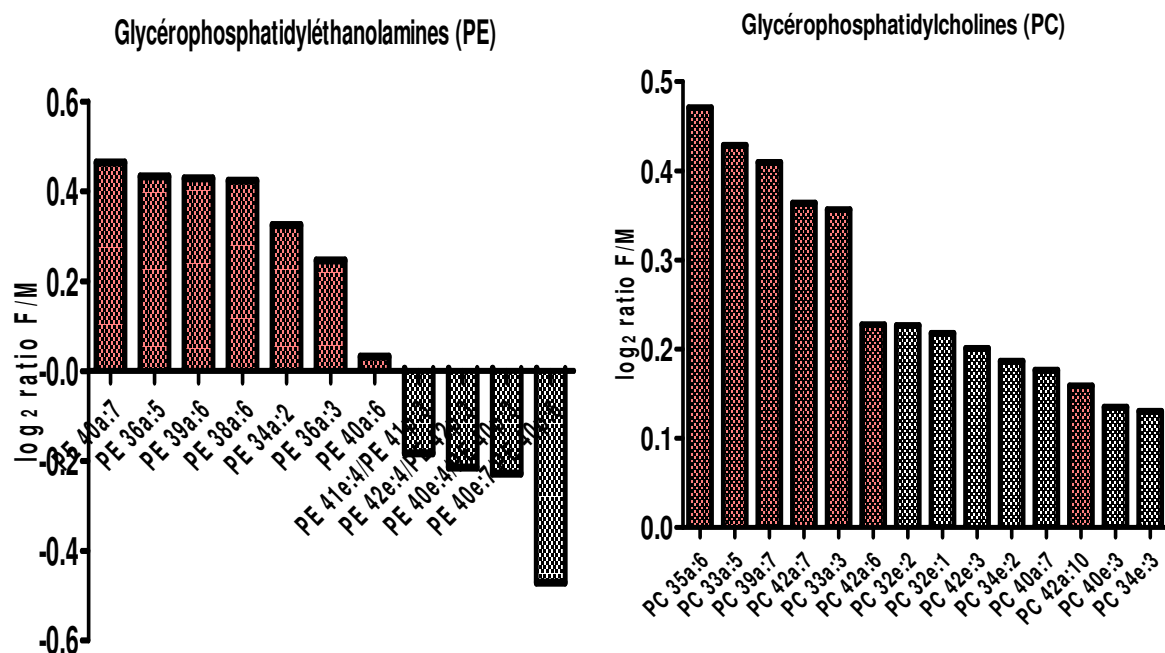


Figure 97. Espèces glycérophosphatidyléthanolamines (PE) et glycérophosphatidylcholines (PC) discriminantes selon le genre

Ceci suggère un catabolisme plus important chez les hommes (formes lysoglycérophospholipides et céramides augmentées chez les hommes versus glycérophospholipides et sphingolipides augmentés chez les femmes) ainsi qu'une activité accrue du complexe enzymatique « 1'-alkyl desaturase » microsomal (Vance & Vance, 2002). Ce dernier a pour seul substrat 3-phosphoéthanolamine (et non 3-phosphocholine) et assure l'oxydation des formes dialkylPE et alkyl-acylPE. Les formes e/pPE sont à l'origine des formes e/pPC par l'action couplée de la phospholipase C et de la cholinephosphotransférase. Les formes e/pPC, bien qu'augmentées chez les femmes par rapport aux hommes, présentent des ratios inférieurs à ceux des formes acyles (Table 26) rejoignant les tendances des PE (Figure 97).

Espèces PC annotées	Moyenne F	Moyenne H	sd F	sd H	Ratio F/M	sex_fdr	sex_vip
812_(+)_PC 35a:6_C43H74NO8P	2.54E+07	1.58E+07	1.76E+07	1.01E+07	1.60	0.022	2.026
778_(+)_PC 33a:5_C41H72NO8P	8.97E+05	5.84E+05	5.93E+05	5.01E+05	1.54	0.018	1.530
890_(+)_PC 39a:7_C47H80NO8P	1.74E+06	1.15E+06	1.14E+06	6.27E+05	1.51	0.048	1.652
945_(+)_PC 42a:7_C50H86NO8P	5.41E+06	3.76E+06	2.87E+06	1.89E+06	1.44	0.015	1.907
782_(+)_PC 33a:3_C41H76NO8P	7.72E+06	5.41E+06	4.78E+06	3.70E+06	1.43	0.030	1.787
948_(+)_PC 42a:6_C50H88NO8P	6.69E+06	5.33E+06	2.77E+06	2.09E+06	1.26	0.031	1.542
750_(+)_PC 32e:2, PC 32p:1_C40H78NO7P	1.07E+07	8.52E+06	4.46E+06	2.37E+06	1.25	0.016	1.874
752_(+)_PC 32e:1, PC 32p:0_C40H80NO7P	2.42E+07	1.95E+07	8.36E+06	5.22E+06	1.24	0.022	1.847
938_(+)_PC 42e:3, PC 42p:2_C50H96NO7P	5.81E+06	4.76E+06	2.00E+06	2.11E+06	1.22	0.030	1.644
785_(+)_PC 34e:2, PC 34p:1_C42H82NO7P	7.58E+07	6.29E+07	2.67E+07	1.82E+07	1.20	0.020	1.740
911_(+)_PC 40a:7_C48H82NO8P	1.35E+08	1.13E+08	4.37E+07	3.39E+07	1.19	0.046	1.486
936_(+)_PC 42a:10_C50H80NO8P	5.81E+06	4.95E+06	2.31E+06	1.82E+06	1.17	0.047	1.457
904_(+)_PC 40e:3, PC 40p:2_C48H92NO7P	5.72E+06	4.99E+06	1.87E+06	2.05E+06	1.14	0.046	1.473
783_(+)_PC 34e:3, PC 34p:2_C42H80NO7P	1.34E+08	1.17E+08	3.88E+07	3.20E+07	1.14	0.042	1.524

Table 26. Espèces glycérophosphatidylcholines impactées par l'effet du genre

4.3.4.2. Variations liées à l'effet de l'âge

37 espèces lipidiques sont augmentées avec l'âge. Notons les espèces Cer, PG, PI et les formes éther/plasmalogène PE. Les espèces céramides impactées présentent de longues chaînes d'acides gras (C22, C23 et C24). Ceci peut impliquer une expression évolutive de l'enzyme céramide synthase 2 qui a pour substrats préférentiels les FA-CoA à longue chaîne. Les concentrations d'espèces plasmalogènes et/ou éther corrént avec l'âge, ceci peut être dû à la plus grande prévalence du stress oxydatif chez les personnes âgées.

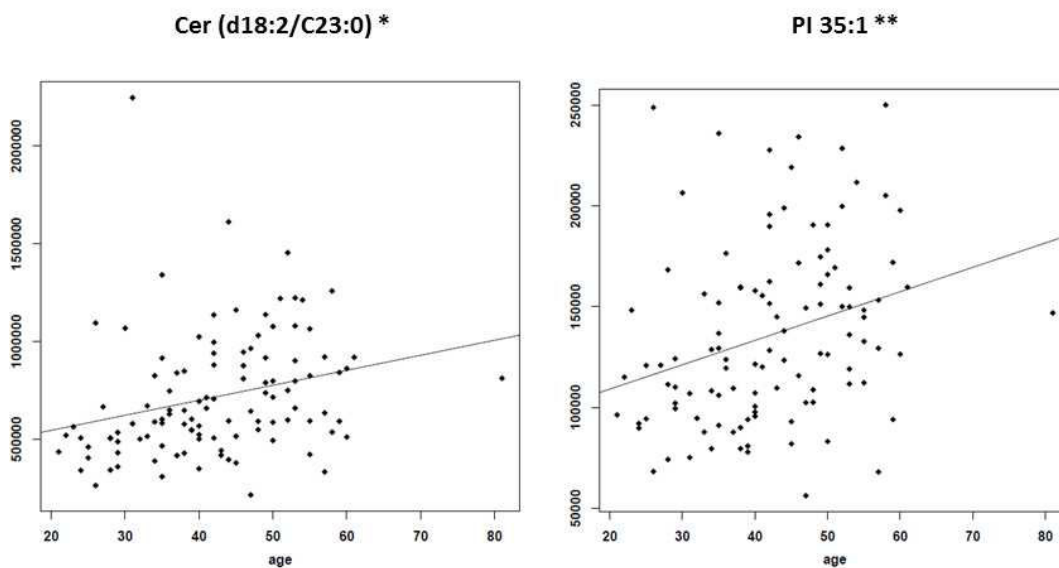


Figure 98. Variations des métabolites selon l'âge

Les valeurs de q-value indiquées sont celles du test de corrélation de Spearman

4.3.4.3. Variations liées à l'effet du BMI

Les triglycérides sont la classe la plus marquée par l'effet du BMI avec plus de 41% des espèces augmentées chez les sujets en surpoids ou obèses. Cette hypertriglycéridémie est la résultante d'une augmentation de production de lipoprotéines riches en TG et une baisse de leur dégradation. Ce processus serait donc corrélé à une baisse des HDL et privilégie l'enrichissement des LDL en TG. Les diglycérides, céramides et sphingomyélines voient leurs concentrations également augmentées avec le BMI. Il en va de même pour les formes plasmalogènes/éther PE sont également augmentées chez les individus en surpoids ou obèses pour qui la prévalence de stress oxydatif est plus importante. A l'inverse, les concentrations des espèces PC (formes acyles et formes plasmalogènes) sont diminuées avec le BMI.

Malgré le nombre important d'espèces impactées par l'effet du BMI, la discrimination des individus selon leur BMI reste délicate en analyse multivariée tel que démontré par la carte factorielle des individus obtenue en OPLS.

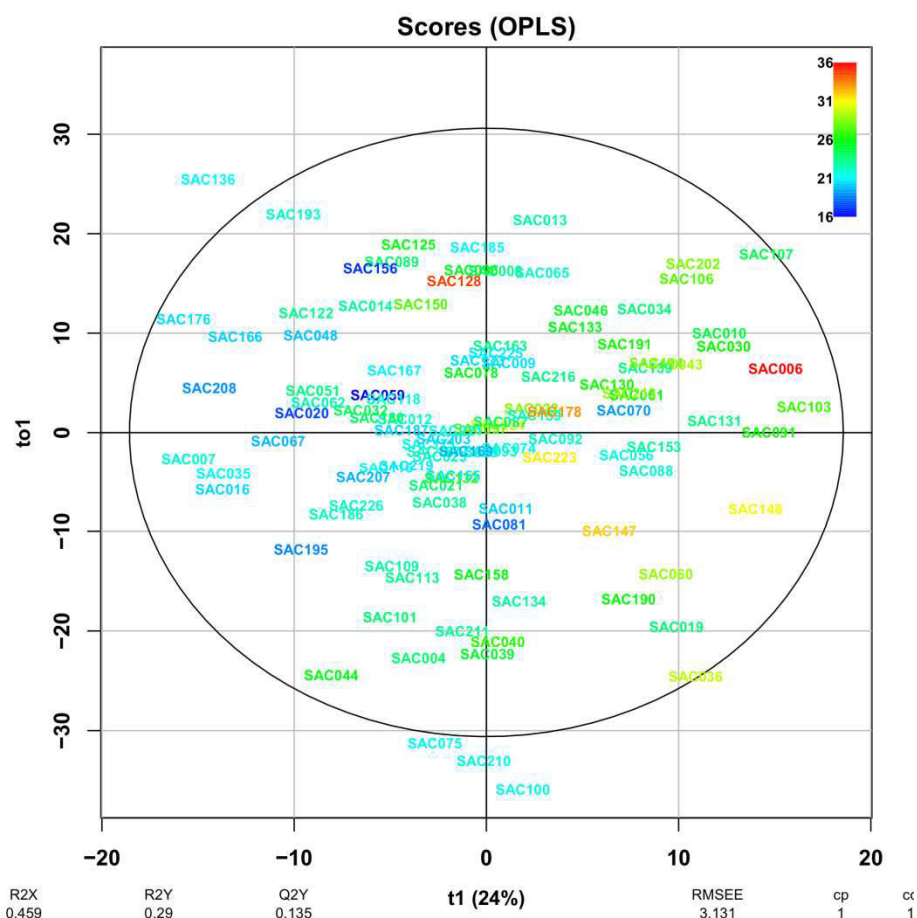


Figure 99. Carte factorielle des individus obtenue en analyse OPLS après analyse RP-ESI(+)-MS des empreintes lipidomiques annotées

4.3.5. Discussion

Les empreintes lipidomiques sériques sont impactées par l'effet du genre, de l'âge et du BMI. En effet, 25% des espèces lipidiques identifiées (232 des 919 espèces détectées) sont impactées par au minimum l'un des facteurs physiologiques étudiés. Cet impact est moindre que celui observé sur ces mêmes individus sur la base de leurs empreintes métabolomiques. Ceci pourrait être dû notamment (i) au nombre d'espèces considérées plus élevé en lipidomique, qui pourrait "diluer" le signal des variables significatives, (ii) au nombre d'individus, un tiers plus élevé en analyse métabolomique, rendant les tests statistiques plus puissants ou encore (iii) à la physiologie sous-jacente. Néanmoins, cette tendance est également retrouvée dans les travaux d'un laboratoire japonais (Ishikawa et al; 2014). Leurs analyses de 253 espèces lipidiques et 297 métabolites a démontré un impact plus important du genre et de l'âge sur les empreintes métabolomiques que lipidomiques sur une même cohorte de 60 individus.

De plus, alors que l'impact du genre apporte la plus grande variabilité et le BMI la plus faible en analyse métabolomique, l'effet de ce dernier est le plus prédominant en analyse lipidomique avec

138 espèces lipidiques discriminantes ($q \text{ val} < 0.05$). Notons que le BMI ne rend pas compte de la distribution des graisses dans le corps mais toutefois corrèle avec l'obésité et la prévalence de syndromes métaboliques. Aussi, les valeurs de BMI peuvent être imprécises car elles ont été calculées à partir des mesures de poids et de taille déclarées par les individus et non pas à partir de mesures mesurées par le personnel de la médecine du travail. L'effet du genre est le second facteur physiologique influant. Ce dernier est visible au niveau des différentes classes de lipides et reflète, au même titre qu'en analyse métabolomique, un catabolisme plus prononcé chez les hommes (augmentation des formes lysoglycérophospholipides et des carnitines). Enfin, l'effet de l'âge est le moins marqué en analyse lipidomique. Ce dernier rend compte des effets du stress oxydatif avec l'augmentation des espèces éther/plasmalogènes dans les extraits lipidiques sériques et d'une métabolisation ralentie avec augmentation de la concentration sérique des xénométabolites en analyse métabolomique.

Nos résultats obtenus en analyse lipidomique vont dans le même sens que ceux décrits par Weir et coll. en LC-MS/MS notamment avec l'augmentation des espèces CE chez les hommes ainsi que les TG et leur évolution avec l'âge (Weir et al; 2013). Par ailleurs, l'augmentation de la concentration des espèces SM chez les femmes décrite par Ishikawa et coll. n'a pas été confirmée (Ishikawa et al; 2014). Les auteurs avaient imputé cette augmentation à la sécrétion d'œstrogènes (hormones sexuelles) (Nikkila et al; 2008) (Merrill, Jr. et al; 1985) or la concentration des SM n'était pas impactée par l'âge notamment chez les femmes ménopausées remettant en question l'hypothèse émise. Cette augmentation peut être artéfactuelle est due à la taille réduite de la cohorte (30 femmes vs 30 hommes).

De plus, les analyses métabolomiques et lipidomiques combinées ont confirmé des interrelations biochimiques entre métabolites et lipides. Par exemple, l'augmentation concomitante de l'acide glycérophosphorique chez les hommes observée en analyse métabolomique ainsi que celle des lysoglycérophospholipides observée en analyse lipidomique rendent toutes deux compte d'une catabolisation des glycérophospholipides chez les hommes.

Cette exhaustivité de la couverture métabolique a également permis d'émettre des hypothèses quant aux voies métaboliques impactées. En effet, tel que décrit en PARTIE A.CHAPITRE I.4.3.4, les espèces céramides sont augmentées avec le BMI mais aussi chez les hommes. Différentes voies peuvent alors être à l'origine de cette augmentation (voir Figure 100). Les espèces sphingomyélines (précurseurs de céramides, voie I) sont augmentées avec le BMI mais ne sont pas impactées par le genre, tandis que l'acide aminé sérine (précurseurs de céramides, voie III synthèse « de novo ») est quant à lui augmenté chez les hommes et reste non influencé par le BMI. Ceci peut suggérer que seule la synthèse de novo des céramides est impactée par le genre avec une activation de l'enzyme sérine palmitoyl transférase alors que la voie de synthèse depuis les sphingomyélines est celle

impactée par le BMI. Le catabolisme des sphingolipides complexes (HexCer, diHexCer, Su et SGL) est quant à lui impacté de façon quasi-similaire par le BMI et le genre (Table 25).

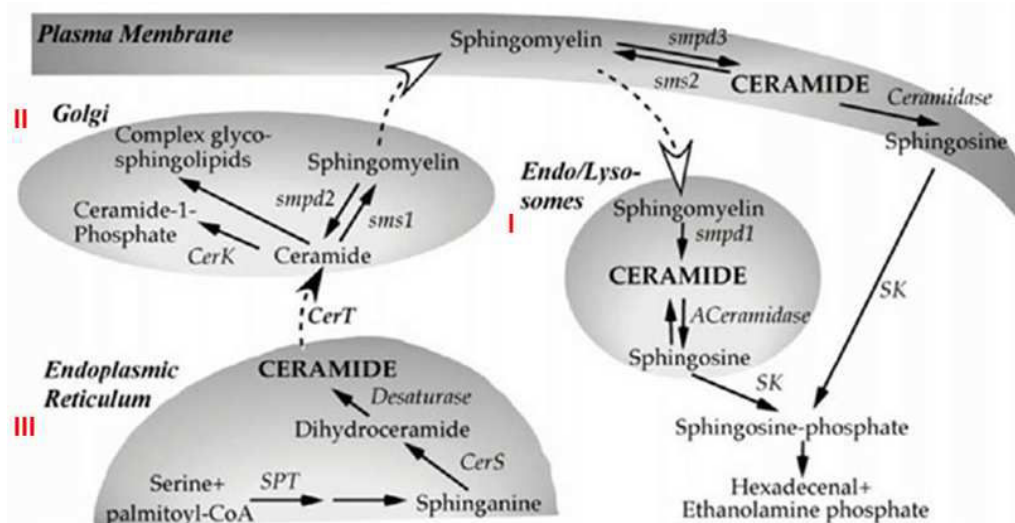


Figure 100. Voies métaboliques responsables de la synthèse de céramide et de la dégradation.

Les espèces de céramide impliquées dans la réponse cellulaire au stress sont représentées en majuscules. Les noms des enzymes sont en italique. Les noms des organites intracellulaires sont indiqués en italique gras. Des flèches pleines sont utilisées pour décrire les conversions métaboliques. Des flèches en pointillés indiquent des transferts assistés par des protéines. Les flèches blanches en pointillés indiquent le transport vésiculaire. Abréviations : SPT: Serine Palmitoyltransferase; Cers: synthases céramides; CERT: la protéine de transfert du céramide. SMS1 et 2: la sphingomyéline synthase 1 et 2; smpd1: Acide Sphingomyélinase, smpd2: sphingomyelinase1 neutre, smpd3: sphingomyelinase Neutre 2 CerK1, céramide Kinase1; SK: sphingosine kinase. ACeramidase: Acide céramidase. 3 voies métaboliques peuvent engendrer des céramides : (I) la synthèse depuis les sphingomyélines correspondantes, II la dégradation des sphingolipides complexes et (III) la synthèse de novo depuis la palmitoyl-CoA et la sérine. (source : <http://what-when-how.com/sphingolipids-as-signaling-and-regulatory-molecules/>)

Nous nous sommes posé la question de la stabilité des échantillons sériques analysés. En effet leur collecte s'est effectuée courant 2008, ces derniers ont ensuite été aliquotés puis conservés à -80°C jusqu'à 2013 où ils ont été décongelés, extraits puis analysés en LC-MS. De récentes études ont jugé l'impact de la conservation des échantillons à -80°C pour une durée de plus de 2ans et demi comme négligeable sur les spectres RMN, ce qui permettrait de ce fait le recours aux études rétrospectives (Pinto et al; 2014).

La force de notre étude a été de suivre plusieurs centaines de métabolites et de lipides (plus de 1100) sur une large cohorte d'individus (227 individus) grâce à des approches analytiques complémentaires. Ceci nous a permis de confirmer des tendances déjà décrites et d'en infirmer certaines. Si l'approche "multi-colonne" a montré son intérêt en termes de couverture du métabolome, elle nécessite le recours à des techniques chimiométriques complexes de fusion de données. L'approche présentée ici consiste à concaténer les tableaux, après avoir standardisé les variables pour équilibrer leur contribution (fusion dite "de bas niveau"). D'autres méthodes apparues récemment agrègent les variables latentes, voire les classifieurs eux-mêmes obtenus sur chacun des jeux de données (fusions dites de "moyen" et "haut-niveau", respectivement). Ces travaux ouvrent

un nouveau champ de recherche en analyse de données pour la DRT/LIST dans le cadre de notre collaboration au sein du CEA. Dans tous les cas, une attention particulière devra être portée aux échantillons contrôles (QC) et aux variables communes aux différents tableaux afin de s'assurer que le jeu de données fusionné reflète fidèlement les différentes contributions expérimentales.

CHAPITRE II. ANALYSE DES METABOLOME ET LIPIDOME SÉRIQUES D'UNE COHORTE DE 78 PATIENTS DIABÉTIQUES

Le diabète est défini par l'élévation de la concentration de glucose dans le sang (hyperglycémie) et regroupe dans un syndrome, plusieurs maladies de pathogénies différentes (trouble de la sécrétion et/ou de l'action de l'insuline).

Selon les critères actuels, le diabète sucré est défini par une glycémie à jeun supérieure à 1.26 g/L ou supérieure à 2 g/L quelque soit l'heure du prélèvement. Ce diagnostic peut également être posé devant une valeur de glycémie de 2 g/L à la 120^{ème} minute d'un examen d'HGPO (hyperglycémie provoquée par voie orale). Une glycémie à jeun modérément augmentée supérieure à 1.1 g/L mais inférieure à 1.26 g/L indiquerait par contre un trouble de l'homéostasie glucidique, ce qui équivaut à la classique intolérance au glucose (Rapport du comité d'expert américain dans le diagnostic et la classification du diabète sucré).

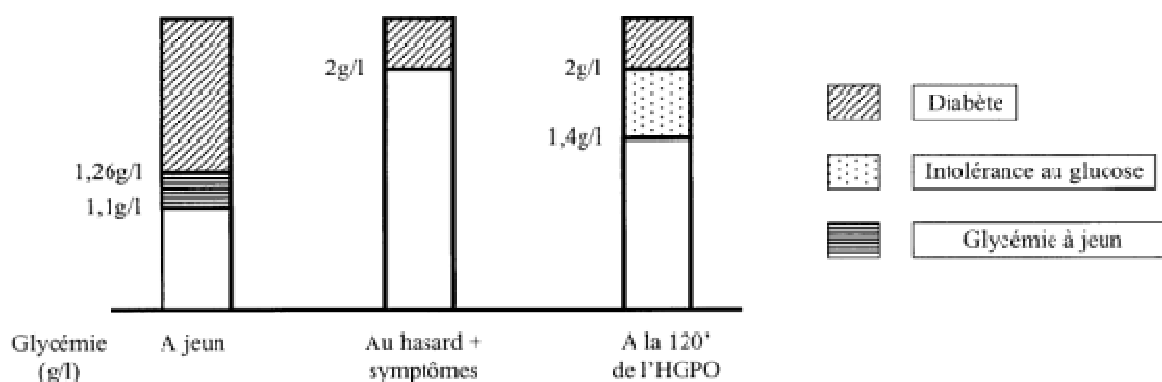


Figure 101. Diagnostic biologique du diabète sucré

Selon les différentes physiopathologies à l'origine de l'hyperglycémie chronique, on distingue 2 grandes classes de diabète (voir Table 27):

- Le diabète de type 1 (T1D) qui se manifeste suite à une carence absolue en insuline, secondaire à la destruction auto-immune des cellules β des îlots de Langerhans pancréatiques. Notons que dans de rares cas, ce diabète apparaît de façon idiopathique.
- Le diabète de type 2 (T2D) quant à lui correspond à une carence relative en insuline. L'hyperglycémie est donc liée à l'association, à divers degrés, d'une insulino-résistance et d'une insulinopénie.

A ces deux types majeurs de diabète, s'ajoutent :

- Les diabètes dits spécifiques qui sont secondaires à une maladie pancréatique, à une endocrinopathie, ou encore ceux d'origine iatrogène ou ceux liés à des anomalies génétiques.

- Le diabète gestationnel correspond à un trouble de la tolérance glucidique apparaissant entre la 24^{ème} et 28^{ème} semaine de gestation et disparaissant à l'accouchement.
- Enfin, l'altération de l'homéostasie glucidique correspond à des anomalies minimales de la régulation glycémique qui traduisent une augmentation du risque de diabète et de maladies cardiovasculaires.

Caractéristiques	Diabète de type I "T1D"	Diabète de type II "T2D"
Proportions	10-15%	85-90%
Antécédents familiaux	+	+++
Age au début	avant 30ans	après 40ans
Mode de début	brutal	progressif
Surpoids	absent	présent
Insulinosécrétion	néant	persistante
Cétose	fréquente	absente
Maladies auto-immunes associées	oui	non
Auto anticorps	présents	absents
Groupe HLA	oui	non
Traitement	insuline	insuline si insulino-résistance, régime alimentaire, exercices physiques

Table 27. Caractéristiques biologiques du diabète de type 1 et diabète de type 2

De plus, l'hyperglycémie chronique est à l'origine de l'apparition de complications à long terme qui affectent la rétine, le rein, les nerfs périphériques et le système cardio-vasculaire (micro et macroangiopathies). Ces dernières sont les causes principales de décès des patients mais peuvent être évitées ou tout du moins retardées par des traitements médicamenteux.

Le tableau clinique d'un diabète sucré est très variable. Le plus souvent le sujet est asymptomatique et le diabète est découvert à l'occasion d'un bilan de santé. Parfois, le signe d'appel sera une complication infectieuse ou une complication chronique. Ce cas de figure est l'apanage du T2D. Parfois au contraire, la clinique est au premier plan, le malade se plaignant alors d'une polyurie, polyphagie, polydipsie et d'un amaigrissement d'apparition brutale. Ce tableau correspond pratiquement toujours à la survenue d'un diabète de type 1.

Devant une telle diversité de symptômes et avec l'absence de définition stricte du T2D –qui est assimilé à toute hyperglycémie non liée à une auto-immunité avérée- le manque de biomarqueurs est un frein regrettable qui justifie l'application des analyses métabolomiques.

Dans ce cadre, des équipes de recherches se sont intéressées à l'hyperglycémie chronique. Zheng M. et coll. et Zhao et coll. ont confirmé la perturbation des voies biochimiques directement impliquées dans le syndrome métabolique (telle que la glycolyse, la lipolyse) à travers la quantification de

certaines métabolites urinaires et plasmatiques chez des sujets intolérants au glucose (Zeng et al; 2009, Zhao et al; 2010). D'autres voies métaboliques ont été explorées chez les diabétiques de type 1 traités ou privés d'insuline, telle la synthèse protéique, la protéolyse, le métabolisme énergétique, le stress oxydant ainsi que la néoglucogenèse (Lanza et al; 2010). La métabolomique s'est également avérée intéressante dans le diagnostic précoce du diabète où lors d'une étude rétrospective, Wang et coll. ont démontré des associations entre de fortes concentrations de certains acides aminés (aromatiques et à chaîne ramifiée) et l'incidence d'un T2D (Wang et al; 2011). Une récente étude menée en GC-Q-TOF visait à définir une signature métabolomique du T1D fulminant. 15, 7, 4 et 12 métabolites sériques discriminant les patients atteints de T1D fulminant aux sujets « contrôle », T2D, T1D et ceux présentant une acidocétose respectivement ont été identifiés. Trois de ces métabolites (à savoir 5-oxoproline, acide glutamique et homocystéine) permettaient la distinction des différents groupes étudiés (Lu et al; 2012). A notre connaissance, cette étude est la seule à utiliser l'analyse métabolomique pour stratifier des patients atteints de différents types de diabètes.

1. Enjeux et objectifs de l'étude

Dans la classification des diabètes, le T2D autrefois appelé diabète non-insulino-dépendant ou diabète de la maturité, qui rend compte de 90 à 95% des cas de diabète, n'a pourtant toujours pas de réelle définition. Le diagnostic reste donc un diagnostic négatif, en l'absence des marqueurs d'auto-immunité qui caractérisent le T1D et en l'absence d'une autre cause de diabète. Il est admis que plusieurs maladies coexistent sous le qualificatif de T2D. Leurs caractéristiques communes sont la résistance à l'action de l'insuline, le déficit relatif de sécrétion d'insuline, l'absence de dépendance de l'insuline et l'obésité. Les individus T2D qui ne sont pas obèses présentent un pourcentage élevé de graisse abdominale (Alberti and Zimmet, 1998, American Diabetes, 2011).

Cette définition manquant de rigueur, nous sommes à la recherche de marqueurs positifs pour le diagnostic du T2D, tout en reconnaissant la nécessité de séparer ce groupe en plusieurs entités. L'intérêt de la stratification du T2D en plusieurs entités définies positivement devrait permettre non seulement une meilleure compréhension de la physiopathologie de cette maladie, mais aussi une meilleure sélection des thérapeutiques pour le contrôle des glycémies.

En pratique clinique quotidienne, la mesure de la sécrétion d'insuline et de la sensibilité à cette hormone ne sont pas réalisables, compte tenu de la lourdeur des techniques à mettre en œuvre (clamp euglycémique et repas test). Les cliniciens sont ainsi amenés à utiliser des indices de résistance à l'insuline, telle la diminution du cholestérol HDL. Pour ce qui est de la sécrétion de l'insuline, la perte de poids est un témoin de la carence en insuline, et ainsi l'absence de perte de poids peut caractériser une carence en insuline relative, par opposition à une carence absolue attestée par la perte de poids.

Une analyse non publiée de 305 patients hospitalisés en 2009 dans une structure d'éducation du service de diabétologie de l'Hôtel Dieu de Paris indique que 26% des patients ont vu leur T2D diagnostiqué avant 40 ans (Données du Pr. Etienne Larger, Service de Diabétologie groupe hospitalier Cochin-Port-Royal, AP-HP). Seuls 43% étaient obèses ($IMC \geq 30 \text{ kg/m}^2$), mais 80 % d'entre eux étaient en surpoids ($IMC \geq 25 \text{ kg/m}^2$), 70% avaient un tour de taille excessif ($\geq 102\text{cm}$ chez les ♂ et 88cm chez les ♀), 50% présentaient une concentration de HDL normale ($\geq 1,04 \text{ mmol/l}$ chez les ♂ et $1,29\text{mmol/l}$ chez les ♀); la concentration basse de HDL étant un marqueur de résistance. De plus, 19% avaient une perte de poids de plus de 5 kg depuis le diagnostic, indiquant une carence en insuline absolue. Sur cette cohorte de patients, si l'on caractérise le diabète de type 2 par 5 critères usuels à savoir : âge au diagnostic ≥ 40 ans, surpoids, tour de taille excessif, HDL bas et perte de poids inférieure à 5 kg depuis le diagnostic, seuls 41% des patients répondent à tous ces critères.

La physiopathologie du diabète de type 2 reste mal connue, et au vue de l'hétérogénéité clinique et phénotypique des patients, les cliniciens s'interrogent quant à la pertinence de la classification actuelle des patients diabétiques. Il apparait ainsi nécessaire d'identifier des groupes de patients homogènes et de trouver des marqueurs positifs de chacune de ces catégories. Dans ce contexte, l'analyse métabolomique offre des perspectives intéressantes.

Par ailleurs, la majorité des études métabolomiques réalisées dans le domaine de la médecine sont basées sur des analyses plasmatiques ou urinaires. Bien que non ou faiblement invasifs, ces biofluides ne permettent pas toujours d'obtenir des résultats biologiquement exploitables, en raison de leur complexité et surtout des faibles concentrations circulantes de métabolites marqueurs de situations pathologiques. Partant de ce constat, nous nous sommes orientés vers l'analyse du métabolome des globules rouges (GR). Les approches globales métabolomiques des globules rouges ont fait l'objet jusque-là d'applications cantonnées à établir une signature métabolique de pathologies érythrocytaires telles que la stomatocytose (Darghouth et al; 2011a) la drépanocytose (Darghouth et al; 2011b) ou encore récemment à un état physiologique érythrocytaire propre telle que l'investigation de la durée de vie des GR (D'Alessandro, Gevi, & Zolla, 2013). Les GR sont un modèle d'étude simple puisqu'ils sont dépourvus de noyau, circulent dans tous les tissus et ont, chez l'homme, une durée de vie de 120 jours. Grâce au grand nombre de transporteurs de métabolites présents sur leur membrane, elles peuvent piéger et concentrer des métabolites qui pourraient rendre compte de l'état métabolique d'un patient sur la durée de vie du globule rouge. Cette propriété a déjà été utilisée pour suivre la stabilité de la glycémie chez les patients diabétiques par dosage de l'hémoglobine glyquée (Hb1Ac).

Nous nous proposons à travers une analyse métabolomique de mieux comprendre la physiopathologie des deux types de diabète les plus fréquemment rencontrés (T1D et T2D). Le but

ultime étant de vérifier si au-delà de l'hétérogénéité phénotypique des patients atteints de T2D, il existerait des bases biochimiques communes qui justifieraient de conserver la terminologie de diabète de type 2, ou dans le cas contraire, d'établir une stratification plus poussée sur la base de profils métaboliques homogènes. Pour cause d'inadéquation de la cohorte de diabétique étudiée, nous nous focaliserons dans ce qui suit sur la différenciation T1D versus T2D et ce à partir d'extraits plasmatiques mais aussi érythrocytaires de patients diabétiques.

2. Matériel et méthodes

Les analyses métabolomiques plasmatiques ont été réalisées selon 3 conditions chromatographiques et dans les 2 modes d'ionisation (Pas d'analyses en PFPP pour les extraits GR), avec un système de chromatographie liquide de type NEXERA de la société Shimadzu (Champs-sur-Marne, France) couplé à un spectromètre de masse de type Exactive (Orbitrap) de la société Thermo Fisher Scientific (Courtabœuf, France).

Les analyses lipidomiques plasmatiques RPLC-MS ont été réalisées avec un système de chromatographique de type Ultimate (Thermo Fisher Scientific, Courtaboeuf, France) couplé à un spectromètre de type Q-Exactive (Thermo Fisher Scientific, Courtaboeuf, France).

Les détails des méthodes LC-HRMS utilisées sont consignés en PARTIE A.CHAPITRE I.2.3 et PARTIE A.CHAPITRE II.2.2 pour les analyses métabolomiques et lipidomiques respectivement.

3. Echantillons biologiques

Un prélèvement de sang veineux sur tube-EDTA a été effectué sur 78 patients diabétiques accueillis au service de diabétologie de l'hôpital Hôtel-Dieu. Cinq campagnes de prélèvements ont été réalisées en Juin 2010 et l'acheminement depuis le service de diabétologie de l'hôpital Hôtel-Dieu a été assuré à 4°C. Des informations biochimiques complémentaires ont été relevées : HDL, LDL, TG, clairance à la créatinine, glycémie, glycémie des 12h, Hb1Ac, ASAT, ALAT et ferritine. De plus, l'âge, le BMI, le genre, l'âge au diagnostic, l'ancienneté du diagnostic, la présence de complications (type HTA, neuropathie, néphropathie) ainsi que la prise médicamenteuse de metformine ou d'insuline ont été spécifiés. On nommera cette cohorte « Db-Juin2010 ».

Le sang total a été centrifugé à 150g pendant 10 minutes à 20°C. Le surnageant (contenant le plasma) a été récupéré dans un tube Eppendorf. Il a été traité comme décrit en PARTIE A.CHAPITRE I.2.2 et en PARTIE A.CHAPITRE II.2.3 pour les analyses métabolomique et lipidomique respectivement.

Concernant la préparation des extraits de GR, 1mL de culot (contenant globules rouges et blancs) est passé sur une colonne de 5 ml remplie de 50% de α cellulose et à 50% de Sigmacell® cellulose type 50, afin de le déleucocyter. Le culot globulaire obtenu est lavé au sérum physiologique,

puis centrifugé à 1430g pendant 3 minutes à 4°C. Le culot globulaire est mélangé avec son double de volume d'eau et le mélange est incubé pendant 3 minutes à 100°C afin de provoquer une lyse cellulaire permettant de libérer les métabolites intra-érythrocytaires. Une nouvelle centrifugation à 3000g pendant 3 minutes à 4°C permet de séparer les débris membranaires du contenu cellulaire. Le surnageant est enfin centrifugé à 4000g pendant 20 minutes à 20°C sur des colonnes Amicon (Ville, Pays), de types Ultra -4, 30 NMWL puis 10 NMWL. Le filtrat ne contient alors plus que des métabolites de masse moléculaire inférieure à 10 kDa. L'extrait total est aliquoté et conservé à -80°C jusqu'à l'analyse. La procédure d'extraction des métabolites érythrocytaires étant longue, elle n'a été effectuée que sur 54 des 78 patients recrutés.

4. Traitement des données

La constitution des séquences d'analyse, le traitement informatique du signal, l'annotation des jeux de données, la normalisation intra-série d'analyse ainsi que les analyses statistiques des jeux de données plasmatiques et érythrocytaires ont été réalisées selon le protocole décrit précédemment (Voir PARTIE A.CHAPITRE I.3).

Les analyses statistiques des données métabolomiques des extraits érythrocytaires ont été réalisées à l'aide du logiciel SIMCA-P11 (Umetrics, Suède) et de Prism (GraphPad, La Jolla, USA). Pour ces données, nous n'avons pas réalisé de correction de tests multiples (FDR).

5. Résultats et discussion

5.1. La démarche

En se basant sur l'étude menée en 2009 au service de diabétologie de l'Hôtel-Dieu décrite ci-dessus (PARTIE A.CHAPITRE II.1), les cliniciens avec lesquels nous collaborons ont défini 5 critères cliniques d'atypicité des patients T2D : (i) âge au diagnostic < 40ans (ii) BMI au diagnostic < 25kg/cm² (iii) HDL < 0.4 chez les hommes (< 0.5 chez les femmes) (iv) l'absence d'hypertension artérielle (HTA) et (v) un délai avant médication à l'insuline < 5 ans (traduisant l'apparition d'une insulino-pénie). Un score allant de 0 à 5 a ainsi été attribué à chaque patient de la cohorte Db-Juin2010, ce score correspond au nombre de critères d'atypicité observés.

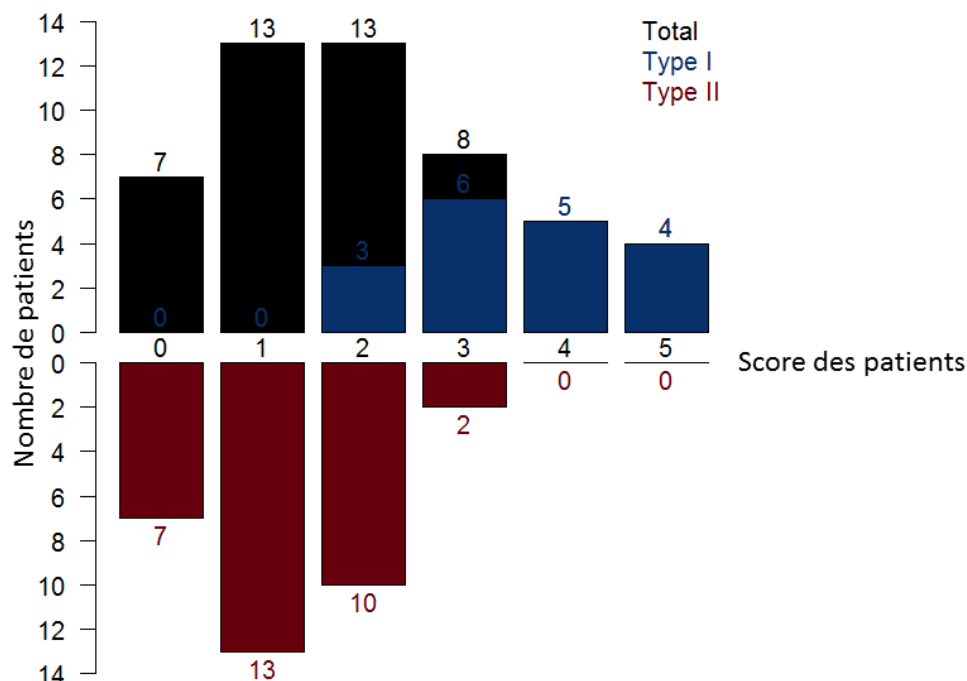


Figure 102. Répartition des patients diabétiques selon les critères d'atypicité

On observe sur la Figure 102 que moins de 20% de l'échantillonnage T2D ne possède aucun critère d'atypicité et que plus de 80% des patients T2D présentent 1 à 3 critères d'atypicité. Aucun patient T2D fortement atypique (soit possédant un score de 4 ou 5) n'a été recruté. Avec cette répartition des scores des T2D, la cohorte ne convenait pas à assurer la stratification des T2D. Par ailleurs, les patients T1D recrutés peuvent être considérés comme phénotypiquement « T2D atypiques » avec des scores allant de 2 à 5. Il nous a alors semblé pertinent de comparer la population de T1D à celle des T2D et de ne pas considérer de population « témoin » d'individus sains. Cette démarche assure la spécificité des biomarqueurs relatifs au type de diabète du fait qu'ils ne soient pas directement liés à une hyperglycémie mais plutôt à la physiopathologie sous-jacente spécifique au type de diabète.

5.2. Analyses statistiques des données plasmatiques

Nous aborderons dans cette partie l'analyse des extraits méthanoliques depuis des échantillons de 78 patients diabétiques suivis au service de diabétologie de l'Hôtel-Dieu en RP, PFPP et ZICpHILIC-ESI(+/-)-MS ainsi que les extraits lipidiques analysés en RPLC-MS. Nous décrirons par la suite les résultats obtenus en analyses statistiques pour la discrimination des patients selon le type de diabète diagnostiqué ainsi que de la mise en évidence de facteurs confondants. Enfin, nous discuterons de la prise en charge statistique des biais biologiques observés.

5.2.1. Analyses statistiques des empreintes plasmatiques indépendamment des facteurs physiologiques

La combinaison des 6 méthodes LC-ESI/HRMS développées a permis de détecter 185 métabolites précédemment identifiés (voir PARTIE A.CHAPITRE I). Tel que présenté en PARTIE A.CHAPITRE I.4.2.4, les métabolites ont été considérés dans leur condition LC-ESI/HRMS optimale afin de ne garder qu'une seule matrice de données. Cette matrice a été dans un premier temps analysée à l'aide d'une PCA afin de visualiser la répartition spontanée des échantillons dans un espace à deux dimensions. Puis, dans un second temps, une analyse statistique supervisée de type régression OPLS a été appliquée au jeu de données parallèlement aux tests statistiques univariés (type Mann-Whitney Wilcoxon et FDR), et ce afin de déterminer quelles variables participent le plus à la différenciation entre les types de diabète à travers leurs contributions factorielles « VIP ».

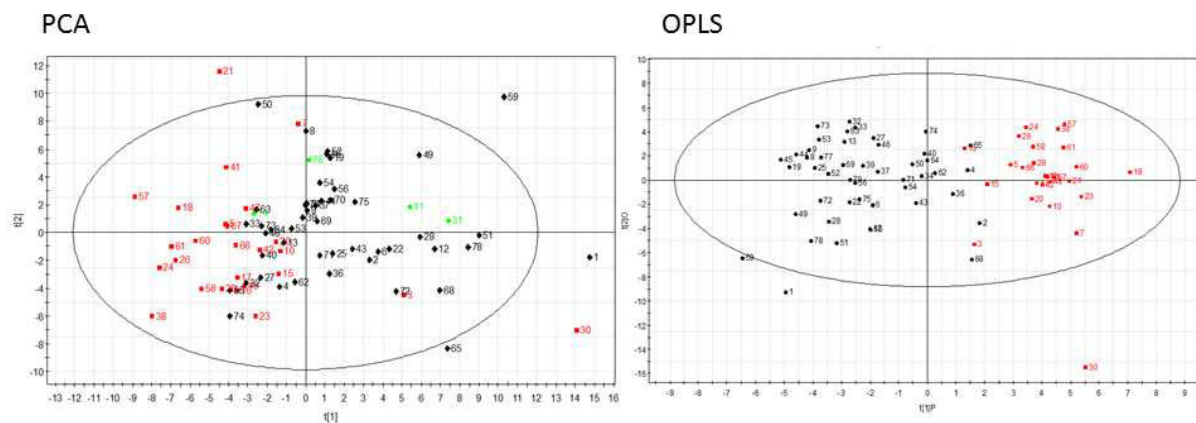


Figure 103. Carte factorielle des individus obtenue en PCA et OPLS selon le type de diabète à partir des empreintes plasmatiques métabolomiques

En rouge : T1D. **En noir :** T2D. **En vert :** LADA

Il y a séparation des patients selon le type de diabète en PCA. Néanmoins, la projection de certains patients T1D les rapprochent des T2D (Patients 15, 16 et 3). Ces derniers présentent un âge élevé notamment un âge au diagnostic très élevé pour les deux premiers.

Pour le modèle OPLS : $R^2Y = 0.774$ et $Q^2_{cum} = 0.384$.

L'application d'une PCA sur les 185 métabolites identifiés permet déjà de séparer les patients diabétiques T1D des T2D selon la première composante principale (Figure 103.A). Ceci démontre qu'il semble possible de stratifier les patients atteints de T1D et T2D sur la base de leurs empreintes métaboliques, et ce indépendamment de l'hyperglycémie. Notons que cette discrimination entre T1D et T2D est la source dominante de variation du jeu de données du fait qu'elle soit visible dès la 1^{ère} composante principale.

Cette discrimination T1D vs T2D est plus marquée en analyse supervisée (OPLS). Des valeurs de VIP ont été obtenues à partir de ce modèle ainsi que des «q-values» du test de Mann-Whitney

Wilcoxon. Ainsi, 46 métabolites présentent une valeur de «q-value» significative (Table 29). Ces métabolites appartiennent à différentes voies métaboliques dont certaines sont particulièrement impactées par le type de diabète (voir Figure 104), telles que la β oxydation des lipides (l'acide hydroxybutyrique, la carnitine et certains dérivés) ou encore la voie de la glycolyse (acide lactique et acide pyruvique). Nous avons également remarqué une plus forte concentration des acides biliaires conjugués à une glycine chez les patients T2D avec une concentration plus faible de ceux conjugués à une taurine par rapport aux patients T1D. De plus, les acides aminés à chaîne ramifiée sont augmentés chez les patients T2D. Ce constat a déjà été décrit pour l'intolérance au glucose (Wang et al; 2011).

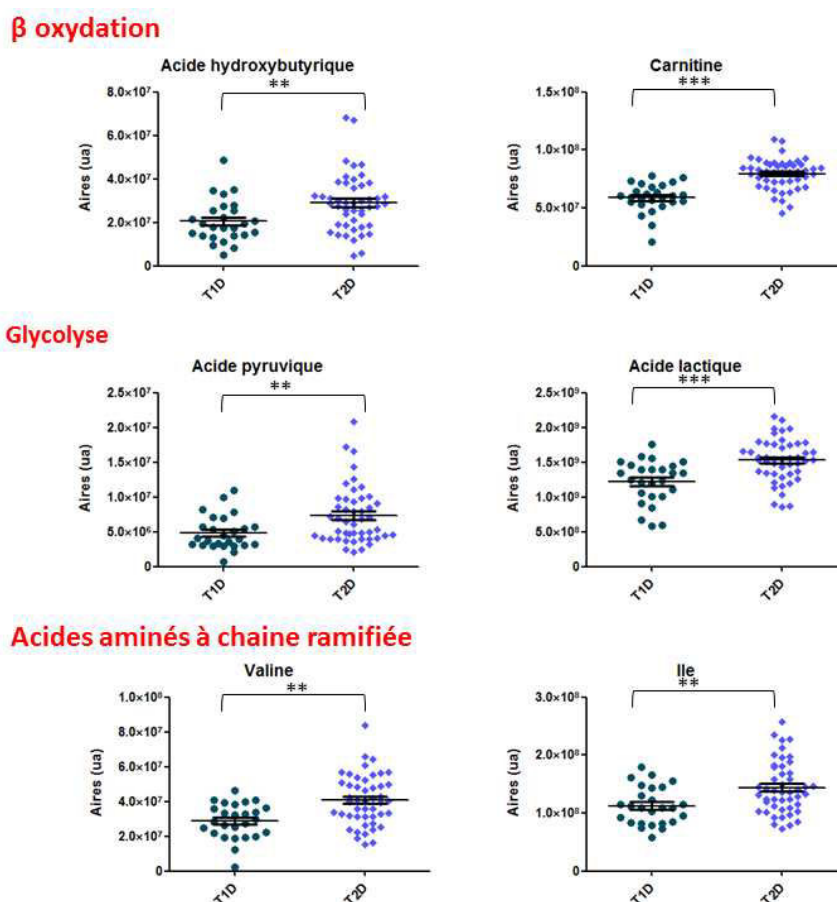


Figure 104. Voies métaboliques majeures impactées par le type de diabète à partir des empreintes plasmatiques.

Le nombre d'étoiles renseigne quant à la significativité du test statistique : *** si $q \text{ value} < 0.001$; ** si $0.001 < q \text{ value} < 0.01$ et * $0.01 < q \text{ value} < 0.05$.

Concernant l'analyse lipidomique, 827 espèces lipidiques (505 détectées en ESI(+)) et 322 en ESI(-)) ont été analysées à l'aide de tests statistiques uni- et multivariés. 21.4% de ces espèces révèlent une «q-value» significative au test de comparaison des rangs entre T2D et T1D. Un quart des espèces discriminantes présente une concentration relative plus importante chez les patients T1D notamment pour les classes de PC et PS tel qu'illustré en Figure 105 (voir Table 28). Les espèces TG

sont celles qui sont le plus impactées par le type de diabète, ces dernières présentent une concentration relative plus élevée chez les T2D, ceci pouvant être lié à la graisse abdominale. Nous n'avons établi aucune tendance quant à la longueur des chaînes carbonées. De ce fait, l'interprétation biologique des résultats n'a pu se faire qu'au niveau des espèces, et non pas au niveau des familles, comme espéré.

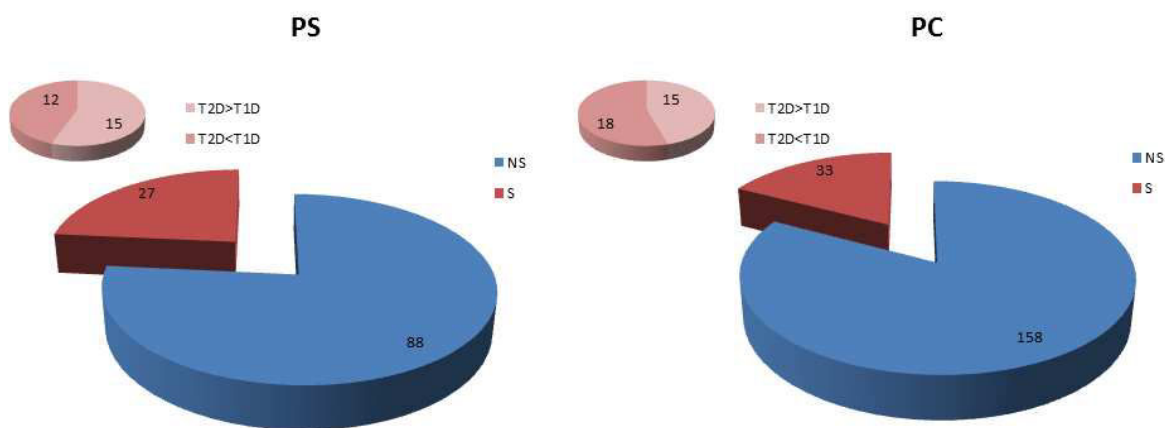


Figure 105. Espèces glycérophosphosérines et glycérophosphocholines impactées par le type de diabète.

NS : métabolites non discriminants. S : métabolites discriminants (Tests de Mann Whitney-Wilcoxon avec FDR).

Classes et sous-classes	q value		VIP		Nbre d'espèces détectées
	Nbre d'espèces discriminantes	Ratio des intensités	Nbre d'espèces discriminantes	Ratio des intensités	
PS	27	15 (T2D>T1D) 12 (T2D<T1D)	57	33 (T2D>T1D) 24 (T2D<T1D)	115
LPS	1	/ 1 (T2D<T1D)	2	1 (T2D>T1D) 1 (T2D<T1D)	9
PI	2	2 (T2D>T1D) /	11	9 (T2D>T1D) 2 (T2D<T1D)	50
LPI	0	/	1	1 (T2D>T1D)	3
PE	21	21 (T2D>T1D) /	39	36 (T2D>T1D) 3 (T2D<T1D)	81
LPE	0	/ /	3	2 (T2D>T1D) 1 (T2D<T1D)	13
PG	0	/	1	1 (T2D>T1D)	15
LPG	0	/	1	1 (T2D>T1D)	2
FFA	2	2 (T2D>T1D)	5	5 (T2D>T1D)	11
Su	0	/ /	2	1 (T2D>T1D) 1 (T2D<T1D)	11
SGL	0	/	3	3 (T2D>T1D)	12
PC	33	15 (T2D>T1D) 18 (T2D<T1D)	49	19 (T2D>T1D) 30 (T2D<T1D)	191
LPC	9	9 (T2D<T1D)	12	12 (T2D<T1D)	34
SM	1	1 (T2D>T1D)	2	2 (T2D>T1D)	37
Cer	0	/	1	1 (T2D>T1D)	6
HexCer	0	/	0	/	4
diHexCer	1	1 (T2D<T1D)	3	3 (T2D<T1D)	10
DG	12	12 (T2D>T1D)	12	12 (T2D>T1D)	18
TG	103	103 (T2D>T1D)	138	138 (T2D>T1D)	171
CE	4	4 (T2D>T1D)	3	3 (T2D>T1D)	34

Table 28. Espèces lipidiques annotées dans les jeux de données lipidomiques de la cohorte Db-Juin2010 et impact du type de diabète sur leurs concentrations plasmatiques

Q value < 0.05 et VIP > 1

5.2.2. Facteurs confondants et leur prise en charge statistique

A la lumière des conclusions obtenues depuis les analyses effectuées sur la cohorte saclaysienne, nous avons pensé vérifier la distribution du genre, de l'âge et du BMI dans la cohorte Db-Juin2010. Ces deux derniers peuvent être considérés comme facteurs physiologiques confondants car leur distribution -inhérente à la physiopathologie même du diabète- est inégale au sein des deux populations (Figure 106).

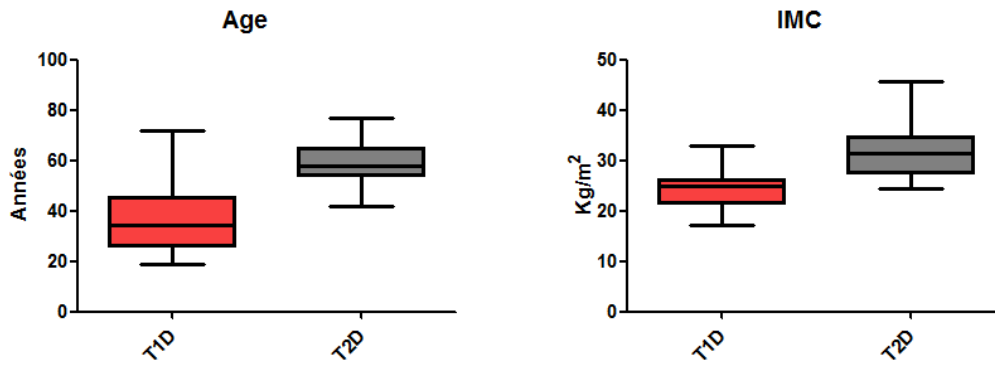


Figure 106. Répartition et effet de l'âge et du BMI sur la cohorte des diabétiques

En effet, si on reprend la carte factorielle des individus en OPLS présentée en Figure 103.B en précisant les étiquettes (âge) et (BMI) des patients, il semble qu'une discrimination T1D vs T2D sous-tend une discrimination selon l'âge ou encore selon le BMI.

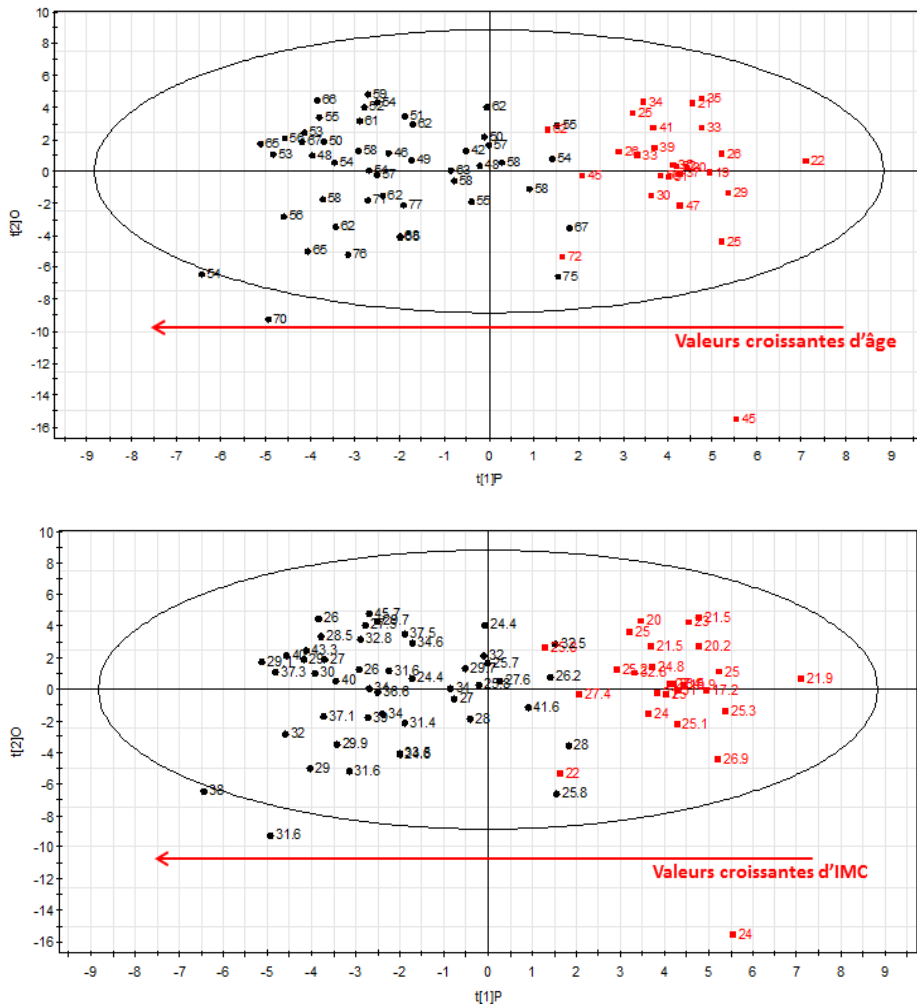
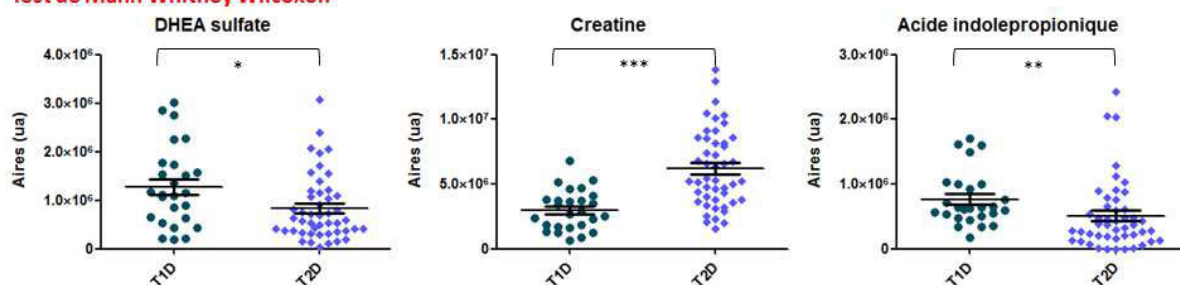


Figure 107. Carte factorielle des individus obtenue en OPLS-DA.

En rouge : T1D. En noir : T2D. Les étiquettes représentent l'âge ou le BMI du patient. Les individus se répartissent selon les valeurs d'âge et de BMI croissants au même titre qu'ils clusterisent selon leur type de diabète.

Dans ce sens, les métabolites discriminants selon le type de diabète peuvent l'être de façon artéfactuelle du fait de la superposition de facteurs confondants. Par exemple, nous avons constaté que la concentration de l'hormone stéroïdienne DHEA-sulfate était plus basse chez les patients T2D que ceux atteints de T1D. Ceci serait dû à l'âge avancé des patients T2D (comme présenté lors de l'étude de la cohorte saclaysienne PARTIE A.CHAPITRE I.4.2.5.1) et non pas à la physiopathologie du T2D. De même pour la créatine qui voit sa concentration augmenter avec l'âge ou encore l'acide indolepropionique dont la concentration baisse chez les individus en surpoids (Figure 108).

Cohorte Db-Juin2010 :
Test de Mann Whitney Wilcoxon



Cohorte Saclay-2008
Corrélation de Spearman

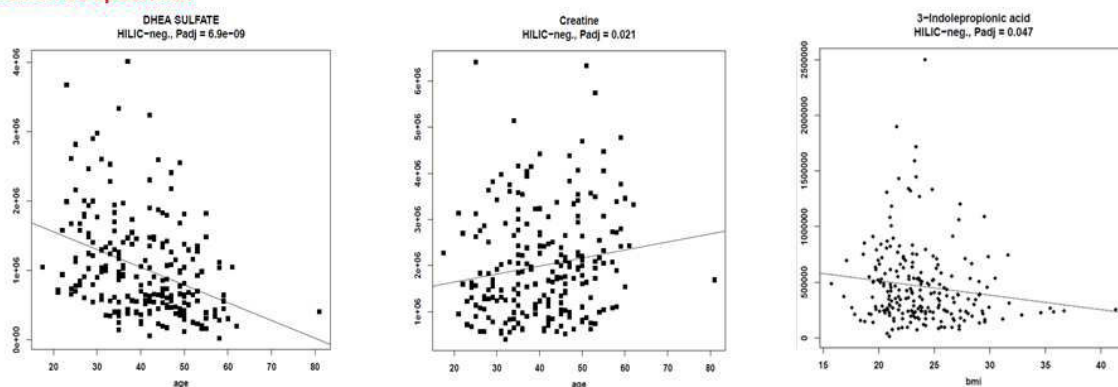


Figure 108. Influence des facteurs confondants dans la signature métabolomique du T2D.

En haut, une représentation graphique correspondant au test statistique Mann-Whitney Wilcoxon (T1D vs T2D) effectué sur les empreintes métabolomiques plasmatiques de la cohorte Db-Juin2010. En bas, les corrélations de Spearman et droites de régression obtenues sur les données de la cohorte saclaysienne pour ces mêmes métabolites.

Le nombre d'étoiles renseigne quant à la significativité du test statistique: *** si $q \text{ value} < 0.001$; ** si $0.001 < q \text{ value} < 0.01$ et * $0.01 < q \text{ value} < 0.05$.

Nous avons pensé dans un premier temps, retirer de la signature métabolomique du T2D tous les métabolites pour lesquels l'effet de l'âge et/ou du BMI a été démontré pour la cohorte saclaysienne (Table 29). Des 46 métabolites initialement recensés, seuls 17 métabolites restent exclusivement discriminants quant au type de diabète (Figure 109).

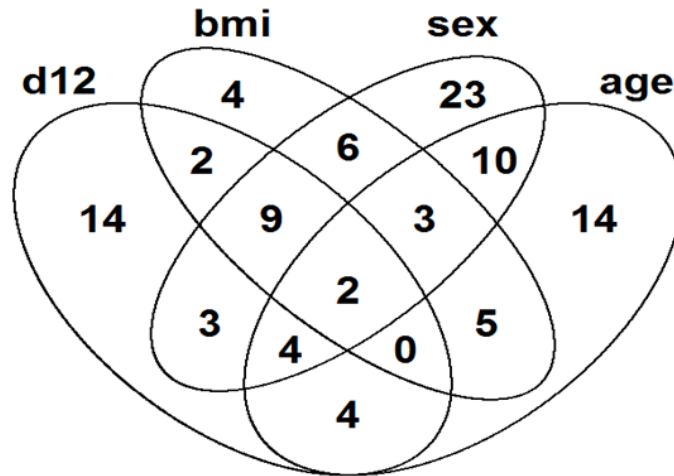


Figure 109. Recouvrement des effets physiologiques et pathologiques sur les métabolites identifiés en conditions ZICpHILIC/ESI(-)HRMS.

Les perturbations de la voie de la glycolyse ainsi que la perturbation de la conjugaison des acides biliaires observées entre les T2D et T1D restent discriminants. De plus, nous avons observé que les métabolites provenant du microbiote intestinal étaient plus concentrés chez les T2D que les T1D tels que l'acide pantothenique, le phenylsulfate, l'indoxyl sulfate ou encore l'acide dihydroxybenzoïque (Ellestad-Sayed et al; 1976, Qin et al; 2012) (voir Figure 110). Ceci va dans le sens des récentes conclusions des travaux dirigés par le Pr Karine Clément [Unité Inserm 872, UMPC et Centre de Recherche des Cordeliers et de l'institut hospitalo-universitaire ICAN (Institute of cardiology metabolism and nutrition), à Paris] et qui suggèrent que le tissu adipeux des patients T2D entretiendrait une inflammation chronique systémique contrairement au T1D qui serait dû à une inflammation pancréatique locale (Dalmas et al; 2014). Des données complémentaires sont présentées en ANNEXE.5.

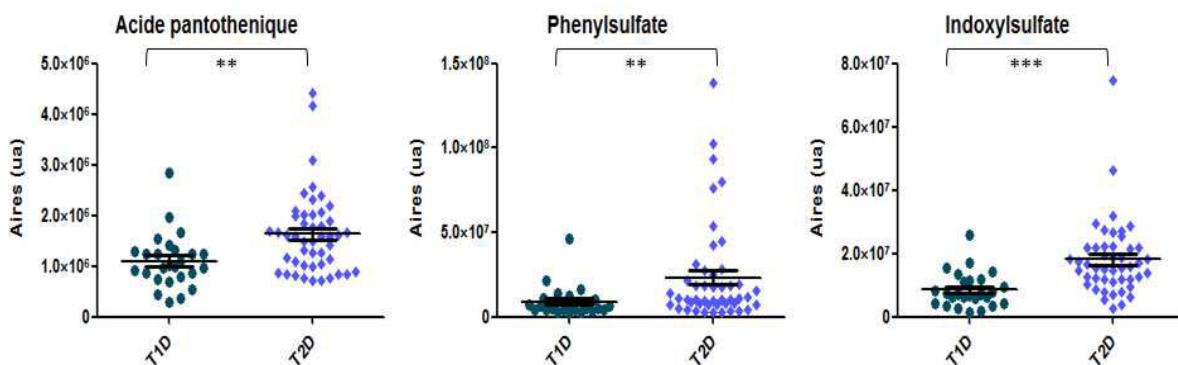


Figure 110. Microbiote intestinal et la signature métabolomique épurée du T2D.

Le symbolisme des étoiles renseigne quant à la significativité du test statistiques : *** si $q\ value < 0.001$; ** si $0.001 < q\ value < 0.01$ et * $0.01 < q\ value < 0.05$.

Cette démarche d'écrémage est stringente car les métabolites peuvent être impactés de façon multifactorielle par les facteurs physiologiques mais aussi pathologiques (en l'occurrence le

type de diabète). Par exemple, les concentrations de l'acide guanidoacétique a été jugé comme augmenté avec l'âge et avec le BMI mais on le voit diminuer chez les diabétiques de type 2 (effet antagoniste). Cette évolution contraire prouve la forte influence de sa concentration par le type de diabète, un impact qui serait tel qu'il annulerait l'effet des facteurs physiologiques.

Nous avons alors pensé corriger les effets de l'âge et du BMI en les modélisant à partir des données obtenues de la cohorte saclaysienne. Nous avons appliqué cette stratégie aux données ZICpHILIC/ESI(-)HRMS (voir Table 29). Il s'agit de considérer les 227 salariés du CEA-Saclay ainsi que les 78 patients diabétiques dans un même jeu de données (voir fusion décrite dans le CHAPITRE I.4.2.2), de réaliser une régression sur la base de la cohorte saclaysienne selon l'âge ou le BMI puis de normaliser les aires des patients diabétiques sur cette droite. L'effet de l'âge ou du BMI est ainsi mathématiquement effacé et seul restera l'effet du type de diabète (voir Figure 111).

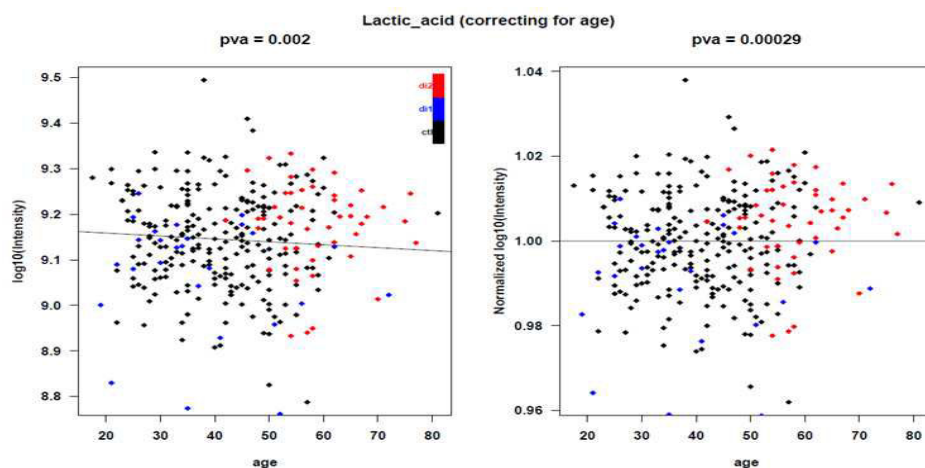


Figure 111. Correction de l'effet de l'âge pour l'acide lactique.

A gauche, les aires de l'acide lactique intégrées dans les échantillons appartenant à la cohorte saclaysienne et à la cohorte Db-Juin2010. A droite, il y a correction de l'effet de l'âge sur les aires de l'acide lactique intégrées depuis les patients diabétiques à partir des données de la cohorte saclaysienne. Ceci a permis d'améliorer sa q-value

Cette démarche nous a permis de préserver une signature métabolomique plus informative. Néanmoins, l'application de cette démarche statistique soulève l'enjeu de s'assurer de la reproductibilité des effets physiologiques entre des cohortes différentes (dans notre cas, cohorte Saclay-2008 et Db-Juin2010). Cependant, tel que nous l'avons démontré dans le chapitre précédent (PARTIE A.CHAPITRE I.4.2.6), il existe peu de recouvrement entre les différentes études menées quant à l'influence des facteurs physiologiques (Lawton et al; 2008,Saito et al; 2014). D'autres solutions statistiques peuvent être envisagées telles que les tests ANOVA à plusieurs facteurs ou des régressions PLS à multiples Y réalisés sur une même cohorte mais se posera toujours la question de la pertinence biologique des résultats statistiques obtenus car ces derniers nous éloignent des données brutes.

Nom	Classe chimique	Cohorte-Db-Juin2010			Cohorte-Saclay-2008		Cohorte-Db-Juin2010	
		q-value	T2D/T1D	VIP	q-value Age	q-value BMI	q-value Correction Age	q-value Correction BMI
Propionylcarnitine	Carnitines and derivatives	1.22E-05	1.70	2.38	4.24E-02	2.51E-04	\	\
Carnitine-[[M+Cl]]-	Quaternary Amines	5.86E-05	1.44	2.08	1.52E-01	2.99E-02	5.24E-04	1.09E-02
Creatine	Amino Acids	5.86E-05	2.08	2.28	2.10E-02	9.37E-01	4.16E-03	5.44E-05
Indoxyl sulfate	Indoles and Indole Derivatives	5.86E-05	2.14	2.14	8.93E-02	8.62E-01	1.57E-03	5.43E-05
Gluconic acid	Hydroxy Acids	6.16E-05	1.44	1.96	5.61E-05	2.12E-01	6.28E-03	1.91E-04
N-Acetyl-glutamic acid	Amino Acids	1.29E-04	1.46	2.01	7.32E-03	2.11E-01	4.16E-03	6.03E-02
Cystine	Amino Acids	1.38E-03	1.28	1.64	2.44E-06	1.08E-04	5.43E-01	5.41E-01
Pimelic acid	Fatty acids	1.38E-03	1.68	1.67	7.10E-01	2.94E-01	1.75E-03	1.09E-02
Glu Phe	Polypeptides	1.59E-03	1.28	1.79	3.01E-02	9.77E-05	\	\
Hydroxydecanoic acid	Hydroxy Acids	1.60E-03	2.27	2.00	3.59E-01	3.53E-01	\	\
Aminoethylphosphonic acid	Organic Phosphonic Acids and Derivatives	1.78E-03	1.36	1.94	9.79E-01	1.70E-05	1.75E-03	1.10E-01
N-Acetylneuraminic acid	Carbohydrates	1.99E-03	0.95	1.54	NA	NA	NA	NA
Lactic acid	Hydroxy Acids	1.99E-03	1.25	1.76	4.71E-01	3.76E-01	2.88E-04	2.94E-02
Aspartic acid	Amino Acids	3.03E-03	0.86	1.59	6.82E-01	3.28E-01	3.42E-03	5.44E-05
Tyrosine	Amino Acids	3.24E-03	1.22	1.68	9.20E-02	3.64E-03	3.78E-02	2.28E-01
Guanidoacetic acid	Amino Acids	3.33E-03	0.73	1.61	4.15E-02	4.29E-02	\	\
Uric acid	Purines and Purine Derivatives	3.84E-03	1.24	1.82	8.66E-01	3.42E-06	4.16E-03	1.70E-01
Succinyladenosine	Nucleoside Analogues	3.85E-03	0.96	1.19	6.89E-01	6.12E-01	\	\
Cotinine	Amino Ketones	4.41E-03	0.59	0.98	1.50E-02	6.61E-02	\	\
Mannitol	Alcohols and Polyols	5.48E-03	1.22	1.65	8.58E-02	3.56E-03	4.49E-02	2.32E-01
Valine	Amino Acids	5.56E-03	1.42	1.57	6.78E-01	6.82E-03	\	\
Uracil	Pyrimidine and derivative	6.46E-03	1.29	1.69	4.34E-01	4.48E-01	5.96E-03	3.90E-02
Indolepropionic acid	Indoles and Indole Derivatives	7.12E-03	0.67	1.16	1.54E-01	4.66E-02	4.49E-02	9.58E-02
D-Glucuronic acid	Carbohydrates	7.12E-03	1.03	1.38	5.74E-07	1.89E-01	4.86E-01	7.06E-02
Indolepropionic_isomer	Indoles and Indole Derivatives	7.73E-03	0.67	1.15	1.20E-01	6.34E-02	6.87E-02	9.90E-02
Pantothenic acid	Amino Acids	9.11E-03	1.49	1.61	7.04E-01	8.06E-02	1.31E-02	9.00E-02
Acetylcitrulline	Amino acids and derivatives	1.12E-02	2.08	1.61	2.69E-03	4.94E-02	2.82E-01	1.09E-01
Glutamic acid	Amino Acids	1.38E-02	1.29	1.70	6.17E-02	1.21E-07	6.54E-02	6.84E-01
N-Acetyl-hexosamine-[[M+Cl]]-	Carbohydrates	1.50E-02	1.18	1.45	1.95E-01	1.89E-01	1.13E-01	1.02E-01
Dihydroxybenzoic acid	Aromatic Acids	1.52E-02	0.93	1.43	NA	NA	NA	NA
Hydroxybutyric acid	Hydroxy Acids	1.52E-02	1.42	1.26	6.32E-01	1.84E-03	4.23E-02	3.09E-01
Phenyl sulfate	Phenol and derivatives	1.52E-02	2.60	1.38	2.29E-01	9.68E-01	1.04E-01	4.60E-02
Pyruvic acid_bis	Keto-Acids	1.81E-02	1.52	1.41	9.14E-01	3.28E-01	1.25E-02	7.06E-02
Ribonic acid-isomer	Carbohydrates	1.82E-02	1.21	1.20	5.74E-07	8.87E-01	6.16E-01	2.17E-02
4-Acetamidobutanoic acid	Amino Acids	1.98E-02	1.23	1.18	5.61E-05	7.21E-01	6.12E-01	1.03E-01
Isoleucine	Amino Acids	2.01E-02	1.28	1.51	7.62E-01	7.33E-04	\	\
Proline	Amino Acids	2.08E-02	1.29	1.39	NA	NA	5.96E-03	6.58E-02
Hexacosanedioic acid	Dicarboxylic acids	3.03E-02	0.66	0.96	6.17E-13	6.78E-01	8.53E-01	7.06E-02
Tauroursodeoxycholic acid	Bile Acids	3.92E-02	0.79	0.65	8.75E-01	6.57E-01	\	\
Tetrahydrofolic acid	Pterins	3.92E-02	0.82	0.98	9.22E-01	9.63E-01	\	\
Glycoursodeoxycholic acid	Bile Acids	3.92E-02	1.97	0.98	8.28E-01	7.87E-01	\	\
Glycochenodeoxycholic acid	Bile Acids	3.92E-02	2.97	1.25	8.75E-01	4.37E-01	\	\
N-acetylaniline	Amino Acids	4.05E-02	1.11	1.14	5.17E-01	4.21E-01	2.69E-02	2.79E-02
DHEA sulfate	Steroids	4.53E-02	0.65	1.15	6.89E-09	8.46E-01	6.39E-01	2.32E-01

Table 29. Signature métabolomique plasmatique du T2D et analyses statistiques correspondantes

Une alternative serait l'utilisation d'une matrice biologique non impactée par les effets physiologiques ou contextuels. Nous avons alors pensé à l'analyse métabolomique des GR.

5.3. Analyse du métabolome érythrocytaire de 54 patients diabétiques

Alors que le plasma circule dans tout le corps et assure le transport systémique des nutriments aux tissus ainsi que de leurs déchets aux reins, l'information métabolomique qu'il contient s'en voit

diluée et directement impactée par des événements ponctuels telle une prise alimentaire. Les GR quant à eux, de par leurs transporteurs peuvent piéger et concentrer l'information collectée localement durant leur durée de vie. Ils peuvent alors constituer un reflet fidèle d'un état métabolique donné, et de ce fait, l'influence d'événements ponctuels et de facteurs environnementaux est supposée y être moins prononcée.

Dans ce contexte, nous voulons dans ce qui suit démontrer la robustesse des empreintes métabolomiques érythrocytaires aux événements ponctuels et aux facteurs confondants physiologiques et contextuels. Pour ce faire, 3 mesures biochimiques classiques relevant de l'équilibre glycémique ont été considérées à savoir la glycémie à l'instant t, glycémie des 12h et Hb1Ac. Dans un second temps, nous avons étudié l'effet des facteurs physiologiques puis pathologiques sur les empreintes érythrocytaires. Enfin, nous avons comparé l'information biologique apportée par l'empreinte des GR à celle des plasmas correspondant.

5.3.1. Empreinte métabolomique érythrocytaire et contextualité

Chez les personnes non diabétiques, la glycémie (concentration sanguine du glucose) est finement régulée par des hormones hypo- et hyperglycémiantes à savoir l'insuline et le glucagon et ce essentiellement selon le rythme de leur sécrétion. Sa valeur varierait donc selon l'état nutritionnel. Cependant, chez les sujets diabétiques, cette régulation par l'insuline est perturbée et les glycémies à jeun ont des valeurs élevées. Afin de mieux rendre compte de l'état glycémique des patients, les cliniciens du service de diabétologie de l'Hôtel-Dieu considèrent en plus de la glycémie à jeun, une moyenne de la glycémie des 12h (à jeun, mesurée en hospitalisation de nuit) ainsi que l'hémoglobine glyquée qui est le meilleur indice de surveillance du diabète.

A partir des 158 métabolites érythrocytaires précédemment identifiés et détectés chez les 54 patients diabétiques considérés, nous avons généré des modèles PLS (où $Y = \text{Glycémie}$, Glycémie des 12h et Hb1Ac). Le logiciel SIMCA ne propose aucun modèle pour la variable « Glycémie » alors que des modèles PLS avec des performances croissantes ont été générés pour la glycémie des 12h et l'Hb1Ac respectivement (Figure 112). L'empreinte métabolomique GR modélise plus favorablement les effets de longue durée (Hb1Ac) que ceux à moyen terme (Glycémie 12h). Cette empreinte n'est pas impactée par des effets immédiats, ponctuels (Glycémie). Notons que 2 métabolites érythrocytaires participent fortement à la construction du modèle PLS $y = \text{Hb1Ac}$: l'anhydrosorbitol et les hexoses. Le premier anti corrèle avec les valeurs d'Hb1Ac, car il est en compétition avec le glucose pour son excrétion urinaire. Son augmentation dans le sang traduirait donc une diminution de la glycémie en faveur d'une glycosurie. Les hexoses quant à eux corréleraient avec les valeurs d'Hb1Ac car ces derniers sont à l'origine même de la fixation des oses sur les chaînes d'hémoglobine.

Parallèlement, la même démarche a été entreprise à partir des empreintes plasmatiques des 78 patients diabétiques. Il a ainsi été possible de générer des modèles PLS pour la glycémie, la glycémie des 12h ainsi que l'Hb1Ac avec des performances décroissantes (valeurs consignées dans la Table 30). Afin de comparer les données plasmatiques aux données érythrocytaires, nous avons généré une sous matrice de données à partir du jeu de données initial plasmatique ne contenant que les 54 échantillons dont les GR ont également été analysés. A nouveau, des modèles PLS ont été générés à partir de cette sous-matrice pour la glycémie et la glycémie des 12h avec des performances statistiques décroissantes. Cependant, avec une taille de cohorte réduite à 54 patients, aucun modèle PLS pour l'Hb1Ac n'a pu être généré sur les jeux de données plasmatiques alors que ce dernier est généré et validé pour les GR. Ceci prouve les limites d'une analyse métabolomique plasmatique à conserver une empreinte métabolomique robuste au temps et démontre la complémentarité de l'analyse métabolomique érythrocytaire.

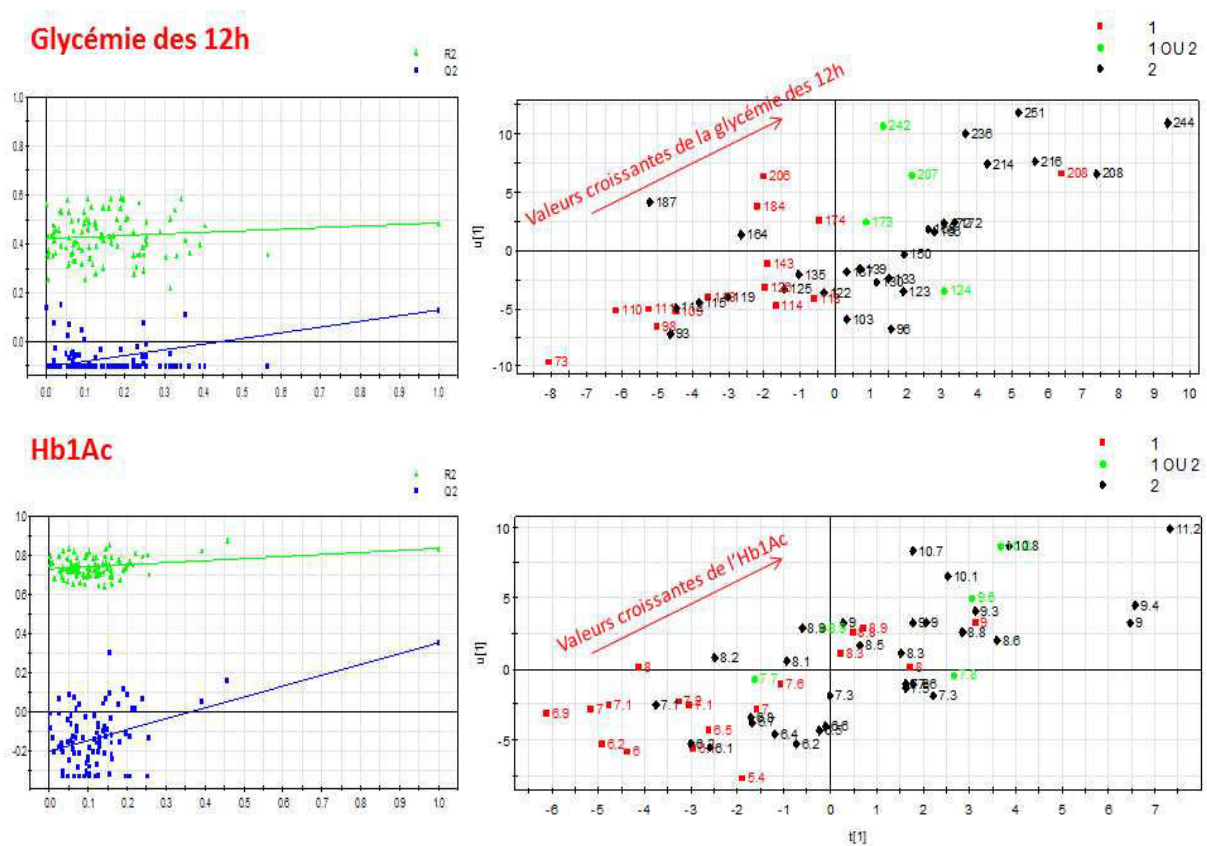


Figure 112. Empreintes métabolomique érythrocytaire et équilibre glycémique.

Cartes factorielles des individus obtenues en PLS (Y= glycémie des 12h et Hb1Ac) et leur validation par permutations croisées (100). Le modèle PLS pour l'Hb1Ac présente de meilleures performances explicatives et prédictives que celui de la glycémie des 12h. De plus, sa validation est plus concluante. Nous pouvons conclure à un plus fort impact de l'Hb1Ac sur les empreintes érythrocytaires que de la glycémie des 12h.

	GR			Plasma		
	R2X	R2Y	Q2cum	R2X	R2Y	Q2cum
Glycémie	\	\	\	0.259	0.939	0.75
Glycémie 12h	0.116	0.484	0.131	0.083	0.627	0.317
Hb1Ac	0.289	0.839	0.354	0.071	0.446	0.155

Table 30. Performances des modèles statistiques PLS générés à partir des empreintes érythrocytaires et plasmatiques en fonction des différentes mesures de l'équilibre glycémique. Cohorte 54 patients diabétiques pour les GR et de 78 patients diabétiques pour le plasma.

Bien que les GR soient en suspension dans le plasma et en continuelle interaction avec lui via les nombreux transporteurs portés par la membrane plasmique érythrocytaire, les informations métabolomiques véhiculées par ces deux matrices sanguines sont différentes et renseignent quant à des évènements métaboliques différents selon leur chronologie. Alors que le plasma modélise au mieux la glycémie ponctuelle, c'est-à-dire un évènement ponctuel, les GR rendent mieux compte de l'Hb1Ac, qui est un paramètre qui documente l'équilibre glycémique sur une période de 3 mois. Cet effet «mémoire» des GR suggère une accumulation progressive des métabolites dans le milieu intracellulaire tout au long de la durée de vie du GR (soit 120 jours), comme en témoigne par exemple les perturbations métaboliques observées au niveau érythrocytaire sur la voie de l'urée (Figure 113). En effet, il y a diminution des concentrations d'un métabolite impliqué dans cette voie (le N-carbamoyl aspartate) chez les patients présentant une forte glycémie des 12h, alors que 4 métabolites voient leur concentration décroître chez les patients non équilibrés (Hb1Ac élevée) (Citrulline, ornithine, arginine et acide guanidobutanoïque). Bien que ces métabolites soient détectés dans le plasma, la voie ne semble pas être impactée au niveau des empreintes plasmatiques.

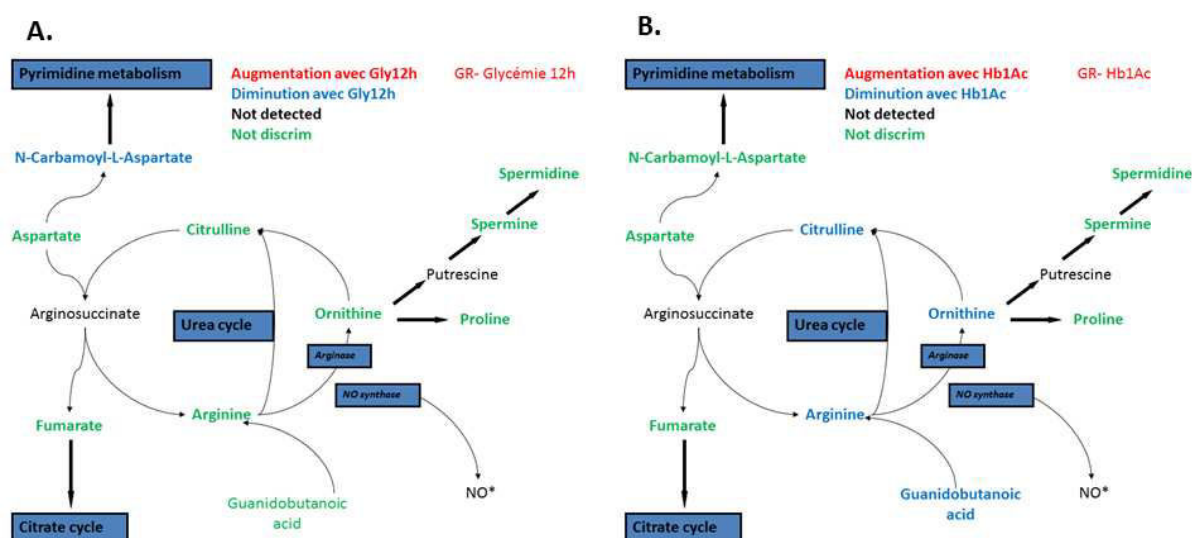


Figure 113. Effets de l'équilibre glycémique sur la voie métabolique de l'urée
Données métabolomiques érythrocytaires. On remarque une évolution au cours du temps de l'impact de la glycémie sur les empreintes érythrocytaires.

En (A.) Les perturbations de la voie métabolique de l'urée selon les valeurs de glycémie des 12h. L'augmentation de la glycémie des 12h semble affecter le « N-Carbamoyl-aspartate » qui voit sa concentration diminuer.

En (B.), on remarque une perturbation marquée du cycle de l'urée en cas de mauvais équilibre glycémique (marqué par une élévation de l'Hb1Ac). L'ornithine, la citrulline, l'arginine ainsi que l'acide guanidobutanoïque voient leurs concentrations diminuer.

Contrairement au plasma, le GR semble mieux s'affranchir du contexte biologique d'un individu. Qu'en est-il des facteurs physiologiques confondants dans le cas de la stratification des patients selon leur type de diabète ? Pour le vérifier, nous nous sommes intéressés à la signature métabolomique par type de diabète.

5.3.2. Empreinte métabolomique érythrocytaire du T2D

Nous avons procédé selon la même démarche pour les extraits érythrocytaires. L'application d'une OPLS-DA sur les 158 métabolites identifiés permet de séparer les patients diabétiques T1D des T2D (voir Figure 114). Ceci démontre que des perturbations métaboliques dues au type de diabète peuvent être également observées dans les empreintes érythrocytaires. Des valeurs de VIP ont été récupérées de ce modèle ainsi que des p-value du test de Student (confirmées par un test de Mann Whitney-Wilcoxon).

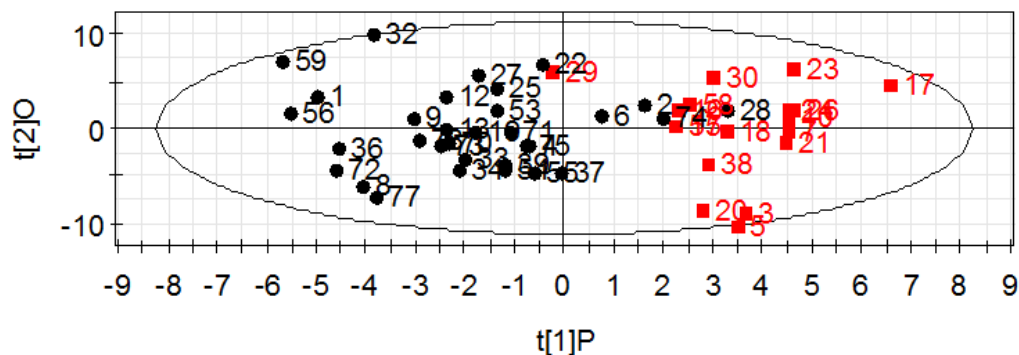


Figure 114. Carte factorielle des individus obtenue en OPLS selon le type de diabète à partir des empreintes érythrocytaires métabolomiques

En rouge : T1D. En noir : T2D. Les étiquettes renvoient au numéro de l'échantillon.

Ainsi, 22 métabolites ont une valeur de p-value significative (<0.05) dont 7 présentent des concentrations plus élevées chez les patients T2D (voir Table 31 et ANNEXE.5)

Nom	Classe chimique	p-value	T2D/T1D
Cys-gly	Peptides	0.04	0.85
Citrulline	Amino Acids	0.01	0.73
Glycine	Amino Acids	0.02	0.82
GSH	Amino Acids	0.03	0.88
Threonine	Amino Acids	0.04	0.81
Guanidinobutanoic acid	Amino Acids	0.03	0.76
Serine	Amino Acids	0.00	0.74
N6-N6-N6-Trimethyl-lysine	Amino Acids	0.05	0.81
Nicotinamide	Pyrdines and derivatives	0.01	0.82
Pentose	Carbohydrates	0.02	1.31
Trimethylamine oxide (TMAO)	Amino Acids	0.01	2.08
Glutamine	Amino Acids	0.04	0.88
Deoxyribose	Carbohydrates	0.00	1.12
Pyroglutamic acid	Amino Acids	0.03	0.86
Hydroxyproline	Amino Acids	0.05	0.87
GSSG	Amino Acids	0.04	0.81
Ketoglutaric acid	Dicarboxylic Acids	0.04	0.76
N-acetyl-L-aspartic acid	Amino Acids	0.01	0.81
Guanidoacetic acid	Amino Acids	0.00	0.71
Uric acid	Purines and Purine Derivatives	0.01	1.23
Sarcosine	Amino Acids	0.00	1.35
Pantothenic acid	Amino Acids	0.02	1.56

Table 31. Signature métabolomique érythrocytaire du T2D

p-value des tests de Student en bilatéral , effectué sur les 158 métabolites érythrocytaires identifiés.

T2D/T1D correspond au rapport de la moyenne des aires des patients T2D sur celle des patients T1D.

A nouveau, le métabolisme du microbiote intestinal semble être affecté par le type de diabète avec une concentration plus élevée de TMAO ou encore de l'acide pantothénique chez les T2D (Ellestad-Sayed et al; 1976, Qin et al; 2012). L'accumulation de TMAO a déjà été décrite chez les patients dialysés (Bain et al; 2006). Selon le métabolome érythrocytaire théorique, ces deux métabolites sont récupérés depuis le plasma vers le milieu cytosolique érythrocytaire. De plus, une voie métabolomique cellulaire semble particulièrement être impactée par le type de diabète à savoir le métabolisme du glutathion (Figure 115). En effet, le glutathion présente une activité anti-oxydante et détoxifiante importante dans les GR. Il participe aussi à la synthèse des leukotriènes, lesquels sont des médiateurs de l'inflammation ce qui expliquerait la variation des concentrations correspondantes selon le type de diabète.

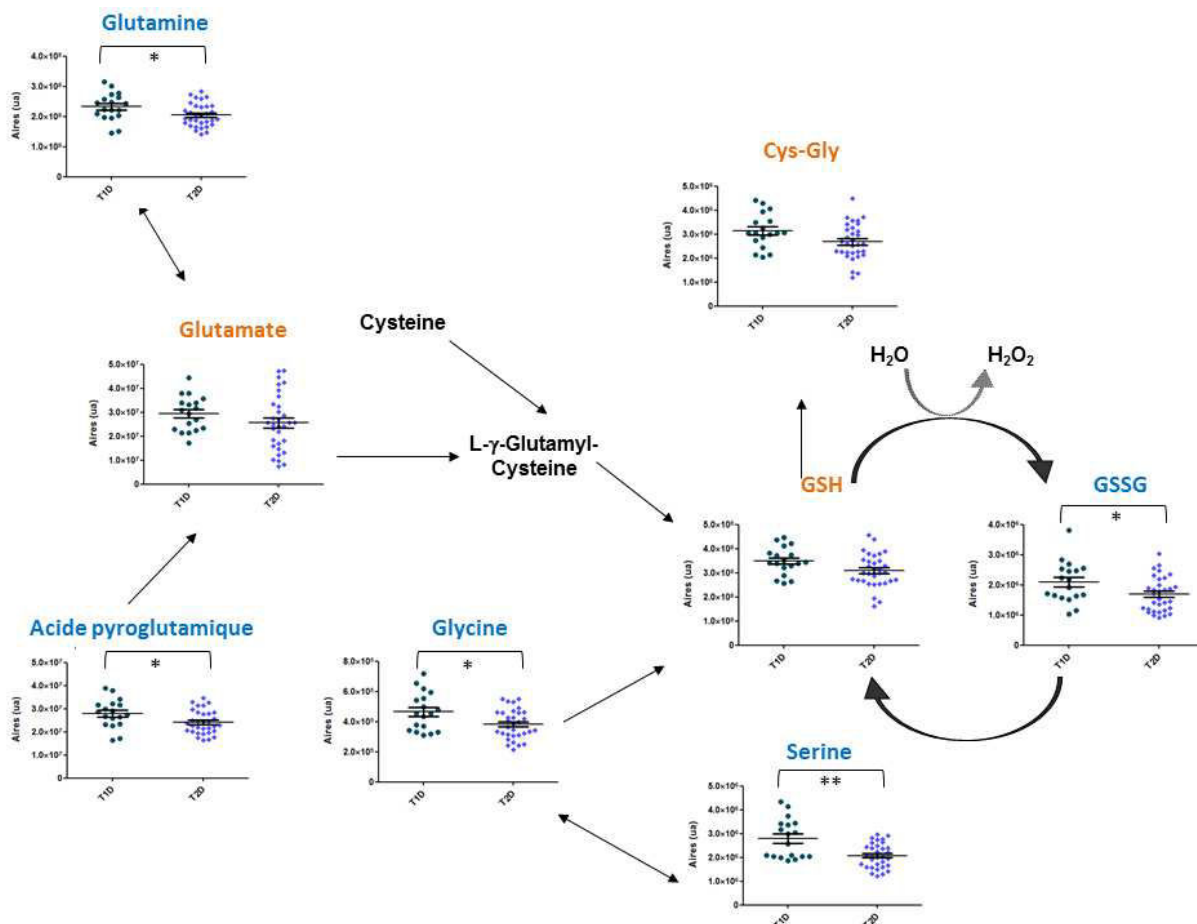


Figure 115. Influence du type de diabète sur le métabolisme du glutathion.

En orange : Métabolite non discriminant. **En bleu :** Métabolite discriminant au test de Student et au test de Mann-Whitney Wilcoxon et présentant une concentration chez les T1D > T2D. En noir : métabolite non détecté. Notons que les 2 métabolites Cys-Gly et GSH présentent des «p-values significatives» (*) au test de Student sans correction des tests multiples mais sont non discriminantes au test de Mann-Whitney Wilcoxon.

Nous avons ensuite tenté de générer, à partir des données métabolomiques érythrocytaires, un modèle PLS avec l'âge et le BMI. Aucun modèle PLS n'a pu être généré avec l'âge, par contre un modèle PLS avec le BMI a été possible (voir Figure 116).

A partir de ce modèle, 3 métabolites érythrocytaires initialement discriminants selon le type de diabète pourraient être imputés au BMI. En effet, ce dernier impacte les concentrations de sarcosine, d'acide urique ainsi que des pentoses (qui dont les concentrations augmentent chez les T2D et chez les individus en surpoids).

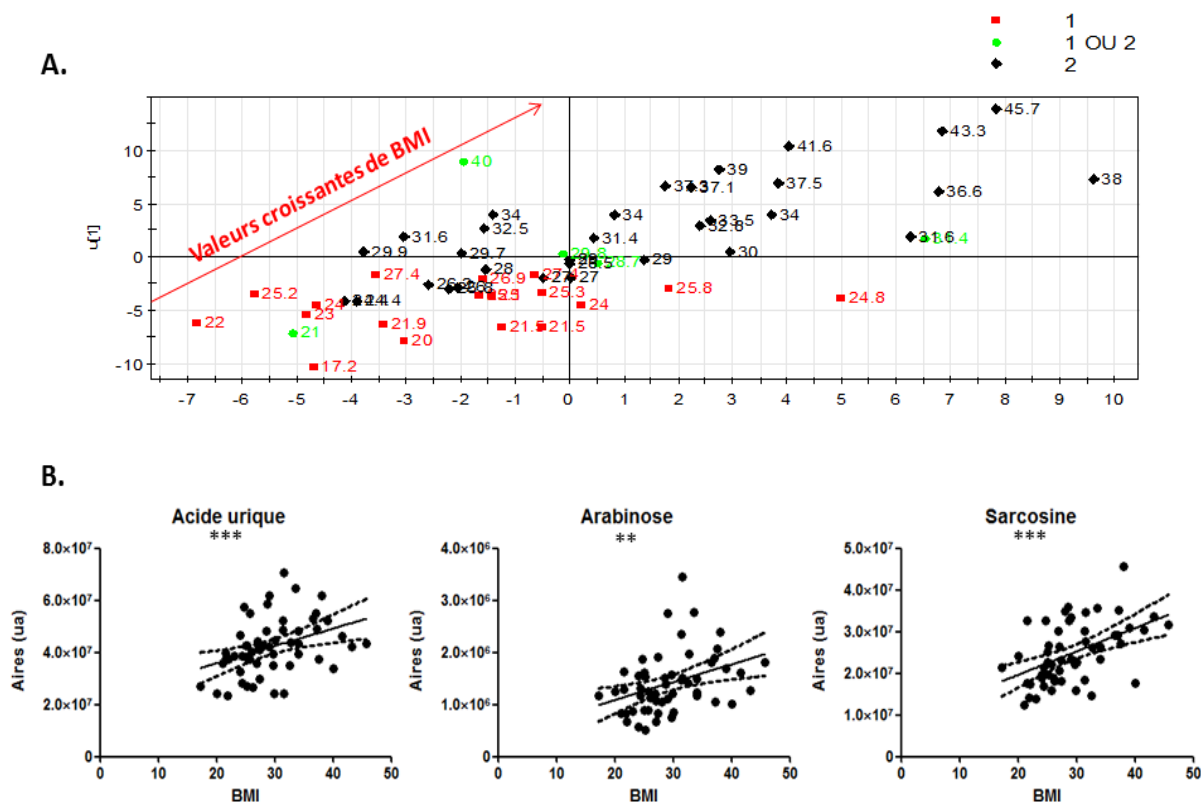


Figure 116. Influence du BMI sur les empreintes métabolomiques érythrocytaires.

(A.) Carte factorielle des individus en PLS (Y=BMI). Les étiquettes correspondent à la valeur du BMI de l'individu. Les couleurs correspondent au type de diabète diagnostiqué. (B.) Corrélations de Spearman des métabolites faisant partie de la signature métabolomique érythrocytaire du type de diabète et qui sont impactés par le BMI.

Le nombre d'étoiles renseigne quant à la significativité du test statistiques : *** si $q \text{ value} < 0.001$; ** si $0.001 < q \text{ value} < 0.01$ et * $0.01 < q \text{ value} < 0.05$.

Des 22 métabolites érythrocytaires initialement influencés par le type de diabète, 19 le sont exclusivement et les 3 autres (à savoir sarcosine, l'acide urique ainsi que les pentoses) sont parallèlement impactés par le BMI.

Bien que le GR offre une signature métabolomique initialement moins riche en métabolites que le plasma, son empreinte métabolomique est moins impactée par des effets confondants et de ce fait sa signature métabolomique épurée est autant voire plus informative (19 métabolites érythrocytaires vs 17 métabolites plasmatiques). De plus, étant une cellule présentant des transporteurs mais aussi un matériel enzymatique propre, elle permet des raisonnements biologiques concernant du biomarqueur de transporteurs ou encore de voies métaboliques contrairement au plasma qui dilue l'information.

5.3.3. Comparatif des empreintes métabolomiques de plasma et de GR

Name	Chemical class	Plasma			GR
		di1Vdi2	age	bmi	di1Vdi2
Guanidoacetic acid	Amino Acids	3.33E-03	4.15E-02	4.29E-02	0.00E+00
Serine	Amino Acids	1.27E-01	5.57E-01	8.46E-01	0.00E+00
Uric acid	Purines and Purine Derivatives	3.84E-03	8.66E-01	3.42E-06	1.00E-02
Citrulline	Amino Acids	1.27E-01	4.82E-05	3.28E-01	1.00E-02
Pantothenic acid	Amino Acids	9.11E-03	7.04E-01	8.06E-02	2.00E-02
Pentose	Carbohydrates	6.84E-01	8.30E-01	3.79E-02	2.00E-02
Pyroglutamic acid	Amino Acids	6.27E-01	9.79E-01	9.68E-01	3.00E-02
Ketoglutaric acid	Dicarboxylic Acids	5.94E-02	2.17E-02	2.95E-06	4.00E-02
Threonine	Amino Acids	3.30E-01	5.10E-01	8.46E-01	4.00E-02
Glutamine	Amino Acids	8.38E-01	7.93E-01	4.56E-01	4.00E-02
Hydroxyproline	Amino Acids	6.40E-01	6.32E-01	1.89E-01	5.00E-02
Alanine	Amino Acids	NA	1.15E-01	1.97E-03	8.00E-02
Lactic acid	Hydroxy Acids	1.99E-03	4.71E-01	3.76E-01	1.50E-01
Hydroxybutyric acid	Hydroxy Acids	1.52E-02	6.32E-01	1.84E-03	1.50E-01
Glutamic acid	Amino Acids	1.38E-02	6.17E-02	1.21E-07	1.60E-01
Tyrosine	Amino Acids	3.24E-03	9.20E-02	3.64E-03	3.50E-01
Propionylcarnitine	Carnitines and derivatives	1.22E-05	4.24E-02	2.51E-04	3.60E-01
Gluconic acid	Hydroxy Acids	6.16E-05	5.61E-05	2.12E-01	4.50E-01
Aspartic acid	Amino Acids	3.03E-03	6.82E-01	3.28E-01	4.70E-01
N.acetylalanine	Amino Acids	4.05E-02	5.17E-01	4.21E-01	4.70E-01
Carnitine	Carnitines and derivatives	3.94E-07	1.41E-01	8.42E-05	5.70E-01
Creatine	Amino Acids	5.86E-05	2.10E-02	9.37E-01	5.80E-01
Proline	Amino Acids	2.08E-02	NA	NA	8.40E-01
N-Acetyl-hexosamine	Carbohydrates	1.50E-02	1.95E-01	1.89E-01	9.00E-01

Table 32. Comparatif des empreintes métabolomiques plasmatiques et érythrocytaires.

Ce tableau répertorie l'ensemble des métabolites détectés dans le plasma et les GR et qui sont discriminants dans l'une, l'autre ou les 2 matrices biologiques selon le type de diabète.

61 métabolites détectés sont communs au plasma et globules rouges des mêmes patients diabétiques. Parmi lesquels, seuls 3 métabolites (acide guanidoacétique, acide pantothénique et acide urique) sont discriminants dans les deux empreintes (voir Table 32). Ces 3 métabolites peuvent être captés par les GR depuis le plasma vers le milieu intracellulaire car ils présentent les mêmes profils de concentrations relatives à savoir les mêmes individus extrêmes en plus d'une bonne corrélation entre les aires plasmatiques et érythrocytaires des mêmes 54 individus (voir Figure 117).

Les concentrations d'acide urique ont déjà été décrites comme augmentées dans le sang en cas de diabète (Costa et al; 2002). Le microbiote intestinal étant perturbé au cours d'un diabète (Qin et al; 2012), la concentration de l'acide pantothénique s'en voit augmentée (Ellestad-Sayed et al; 1976). L'acide guanidoacétique est quant à lui engagé dans plusieurs voies métaboliques (acides aminés, créatine, cycle de l'urée), à lui seul, il ne peut être considéré comme biomarqueur d'une voie métabolique spécifique.

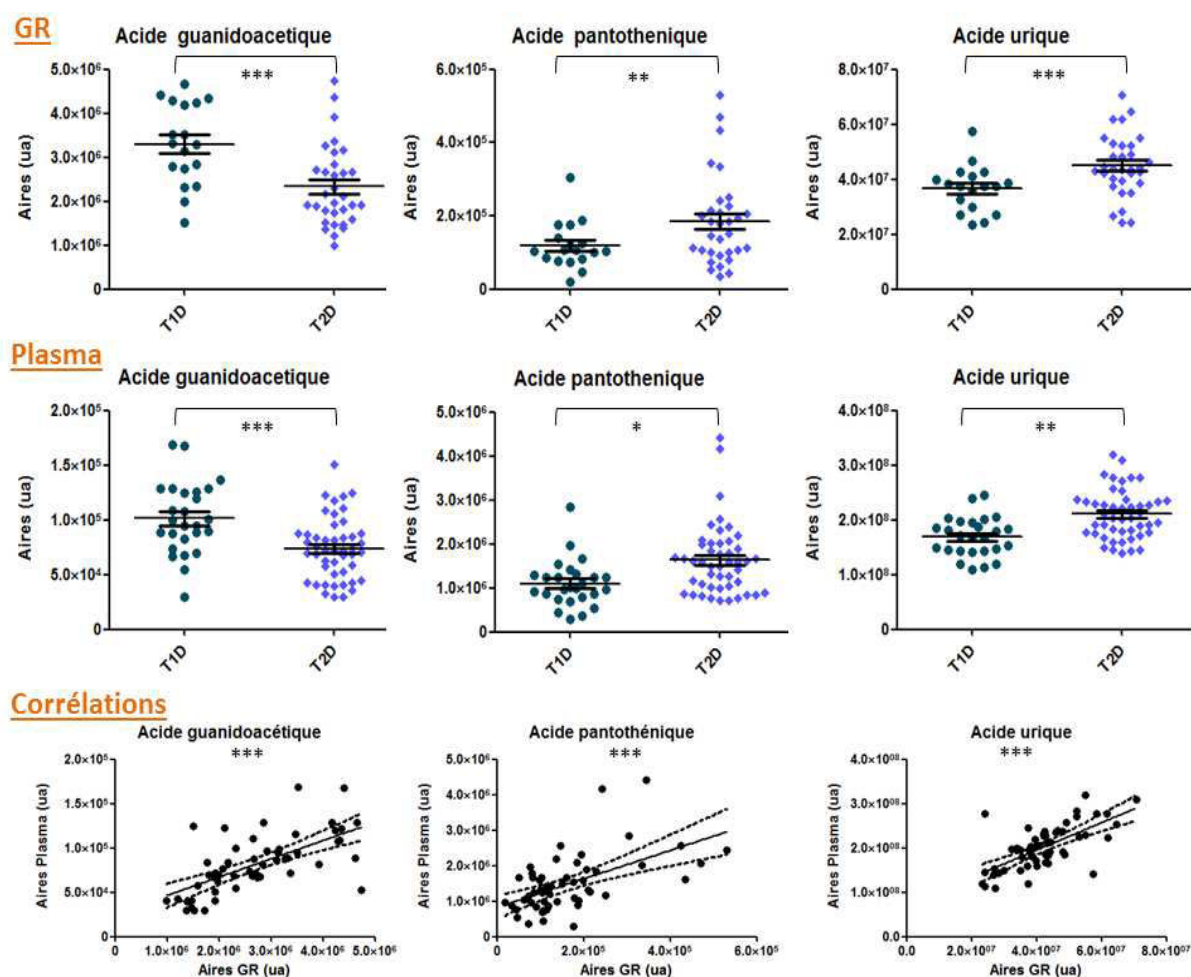


Figure 117. Métabolites plasmatiques et érythrocytaires communs discriminants selon le type de diabète.

Le nombre d'étoiles renseigne quant à la significativité du test statistiques : *** si $q \text{ value} < 0.001$; ** si $0.001 < q \text{ value} < 0.01$ et * $0.01 < q \text{ value} < 0.05$.

De plus, tel que démontré dans la Table 32, 12 métabolites sont impactés par le type de diabète dans le plasma mais ne le sont pas dans les GRs bien qu'ils y soient détectés. Nous pouvons conclure à un défaut de transport depuis le milieu extracellulaire (plasma), notamment pour les carnitines et les acides aminés ou bien à des sensibilités de perturbation métabolique différentes (exemple de l'acide lactique pouvant être produit dans les GRs mais aussi au niveau d'autres tissus et qui pourraient de ce fait présenter des réponses différentes à l'effet du diabète). Cependant, 8 de ces 12 métabolites plasmatiques discriminants ont été décrits comme impactés par l'âge et/ou le BMI dans la cohorte saclaysienne. De ce fait, nous pouvons supposer que l'empreinte érythrocytaire pourrait être spécifique du type de diabète et ne pas être affectée par les facteurs confondants expliquant leur non significativité selon le type de diabète dans les empreintes érythrocytaires.

Dans ce sens, les GR représentent une matrice biologique séduisante pouvant aider à obtenir des connaissances biologiques pertinentes. Leur analyse pourrait se faire parallèlement à celle des plasmas car elle fournit des informations complémentaires. De plus, un fractionnement des GR selon leur âge avant analyse métabolomique de la fraction la plus âgée apporterait une plus value considérable. Dans ce contexte, une collaboration avec l'équipe du Pr. Douay (équipe INSERM), les équipes de Paul-Henri Roméo (DSV/iRCM, CEA-Fontenay-aux-Roses) du GREX (laboratoire d'excellence sur le globule rouge) pour le fractionnement des GR et avec des rhumatologues (Service de Rhumatologie, Hôpital Avicenne, AP-HP) est en cours au sein du laboratoire afin de démontrer la pertinence de l'analyse métabolomique des GR âgés pour le biomarquage de maladies auto-immunes.

CONCLUSION ET PERSPECTIVES

L'analyse métabolomique est aujourd'hui un outil puissant pour l'exploration des systèmes biologiques avec des applications prometteuses dans divers domaines de la biologie. Au cours de la dernière décennie, la spectrométrie de masse a acquis une place de choix parmi les méthodes d'analyse métabolomique en raison de sa polyvalence et de sa sensibilité. Cependant, l'incapacité d'assurer une couverture complète du métabolome avec une technique d'analyse unique ainsi que la difficulté de partager des informations à partir de techniques d'analyse différentes entrave l'identification des métabolites et rend l'accès aux réseaux métaboliques difficile.

Dans ce contexte, la première partie de mes travaux de thèse a servi à étendre la couverture métabolique de différentes matrices biologiques (sérum, GR et LCR humain) en tirant profit de la polyvalence et de la sensibilité des couplages LC-ESI/HRMS. Il a ainsi été possible, grâce à des combinaisons LC-HRMS, d'identifier plus de 200 métabolites polaires dans les GR mais aussi dans les séras humains, ainsi que plus de 700 espèces lipidiques sériques différentes. Au total, plus de 25% des métabolites sériques identifiés n'étaient pas répertoriés dans le sérum humain par la base de données HMDB en Avril 2014. Il en va de même, pour le métabolome érythrocytaire dont 80% des métabolites détectés ne sont pas prédictibles par le modèle métabolique du globule rouge élaboré par Fabien Jourdan (INRA-Toulouse) à partir de données d'analyse protéomique. Nous avons également démontré que combinées aux mesures de masses précises à haute résolution, la complémentarité des sélectivités offertes par la chromatographie liquide nous a permis de détecter 27% d'espèces sériques isomères.

Par la même occasion, nous avons développé des procédures efficaces d'annotation de jeux de données biologiques, à commencer par l'implémentation de bases de données spectrales dédiées par matrice biologique. Reste l'identification systématique et formelle des métabolites et espèces lipidiques. Nous avons réalisé, au cours de cette thèse, des expériences de fragmentation sur des listes prédéfinies d'ions, puis procédé à l'interprétation manuelle des spectres CID générés. Cette démarche est longue et fastidieuse. L'utilisation d'outils informatiques permettant l'interprétation automatique des spectres CID et leur comparaison aux spectres de référence permettrait d'automatiser le procédé. Cependant, une bonne connaissance de la chimie en phase gazeuse est souvent requise et demande une expertise qu'une formule mathématique ne peut modéliser.

De plus, des procédures de standardisation, notamment des énergies déposées dans les analyseurs, doivent être développées afin de faciliter la comparaison de spectres de MS/MS provenant de laboratoires et d'instruments différents.

Une autre préoccupation majeure est d'utiliser ces outils de profilage non ciblés pour les études de grandes cohortes. En effet, le manque de reproductibilité de l'ESI-MS complique grandement son application aux études épidémiologiques ainsi que les initiatives de validation analytique, de partage

des données et leur exploitation par des biologistes. La métabolomique quantitative devrait aider à résoudre ce point, mais sa mise en œuvre est souvent compliquée et coûteuse. Grâce à une démarche analytique rigoureuse (mettant en jeu des échantillons de QC), ainsi qu'au développement d'outils informatiques pour la normalisation (Dunn et al; 2011), nous avons démontré à travers notre étude pilote, la possibilité de scinder une expérience en plusieurs séries d'analyse dans lesquelles les performances analytiques sont conservées.

Dans un premier temps, il nous a ainsi été possible de suivre plus de 200 métabolites et 900 lipides dans 227 échantillons sériques de salariés du CEA, tout en minimisant le biais analytique. Ensuite, à l'aide de tests statistiques appropriés, nous avons pu mettre en évidence l'impact des facteurs physiologiques sur les empreintes métabolomiques et lipidomiques. L'effet du genre est le plus prédominant en analyse métabolomique avec 100 métabolites impactés décrivant essentiellement un catabolisme important chez les hommes. Par ailleurs, le BMI est le facteur physiologique qui impacte le plus les empreintes lipidomiques avec notamment la corrélation de plus de 40% des concentrations de TG avec ce facteur.

Bien que nos résultats sériques restent cohérents avec ceux obtenus dans les urines des individus de la même cohorte (Aurélié Roux, Doctorat de l'Université Paris 6, 2011), avec leur 34% de discordance maximale pour certaines catégories, ils ne sont pas totalement comparables avec les données déjà publiées (Ishikawa et al; 2014, Lawton et al; 2008, Saito et al; 2014). Ceci peut être dû à des stratégies analytiques ou encore des analyses statistiques différentes pouvant mener à des conclusions différentes. Mais surtout, une variabilité inter-individuelle importante due à une population non contrôlée contribue fortement à ces disparités. Ceci peut être notamment dû à l'hygiène de vie. Par exemple, l'analyse de la cohorte américaine (Lawton et al; 2008) a mis en évidence une augmentation des concentrations de nicotine chez les individus âgés. A l'inverse, nos données indiquent une baisse de la concentration de celle-ci avec l'âge. Cela pourrait s'expliquer par des tendances différentes quant au tabagisme dans les deux pays. En effet, avant le rapport des autorités sanitaires américaines de Janvier 1964 liant le tabagisme au cancer du poumon, 42% des américains fumaient contre moins de 18% aujourd'hui, en raison d'une campagne fructueuse contre le tabagisme (<http://www.leparisien.fr/laparisienne/sante/etats-unis-apres-50-ans-de-guerre-contre-le-tabac-les-fumeurs-se-font-plus-rares-11-01-2014-3484019.php>). Au contraire, en France, on compte aujourd'hui plus de 40% de fumeurs réguliers entre 18-25ans (<http://www.cnct.fr/preserver-les-jeunes-14/le-tabagisme-chez-les-jeunes-en-france-1-21.html>). Les habitudes alimentaires peuvent également avoir un impact sur les concentrations de métabolites (Dumas et al; 2006). Citons par exemple la présence dans des spectres de RMN de signaux liés à la consommation de poisson lors d'analyse d'échantillons provenant de la population japonaise.

Notons également qu'au-delà des divergences de résultats, donner du sens biologique aux différences observées notamment dans des biofluides tels que le plasma ou le sérum est délicat. En effet, les concentrations mesurées sont la résultante d'un certain nombre de réponses adaptatives relevant de l'anabolisme et du catabolisme et qui prennent en compte l'apport nutritionnel et l'impact de l'environnemental, en plus des prédispositions génétiques de chacun. Dans ce contexte, une augmentation du taux d'acides aminés peut par exemple être une réponse à un régime alimentaire riche en protéines ou encore à une protéolyse augmentée.

Dans un second temps, nous avons appliqué la méthodologie développée (analytique et bioinformatique) à une cohorte de 78 patients diabétiques, dans le but de stratifier ces patients en fonction du type de diabète diagnostiqué. Il nous a ainsi été possible de mettre en avant une signature métabolomique plasmatique cohérente avec l'état de l'art des connaissances physiopathologiques du diabète, avec notamment l'augmentation des acides aminés à chaînes ramifiées chez les T2D ou encore une perturbation du microbiote intestinal. Cependant, la contribution de facteurs physiologiques confondants (âge et BMI) à cette stratification nous a posé le problème de la spécificité des biomarqueurs identifiés. Nous avons, de ce fait, proposé 2 stratégies de gestion de la contribution des facteurs confondants: (i) l'élimination des métabolites impactés par ces facteurs confondants de la signature métabolique du diabète, sachant que ceci revient à ignorer le caractère multifactoriel de l'homéostasie, ou encore (ii) la correction des effets confondants à l'aide des données provenant d'autres cohortes (dans notre cas, la cohorte saclaysienne). Cette dernière stratégie implique cependant de supposer que les résultats obtenus d'une cohorte à l'autre sont parfaitement reproductibles, ce qui n'est pas le cas. Les données générées ont permis de mieux cerner les contraintes liées à ce type de recherche clinique et la validation des biomarqueurs à travers l'analyse d'autres cohortes ou bien un suivi longitudinal des mêmes individus permettrait de fiabiliser les empreintes détectées et les tendances observées.

Nous nous sommes ensuite tournés vers l'analyse métabolomique des extraits de globules rouges obtenus avec les mêmes patients. Nous avons montré que les concentrations de métabolites présents dans les érythrocytes sont moins impactés par le contexte biologique et les paramètres physiologiques (âge ou glycémie). De plus, les GR possèdent leur propre système enzymatique en plus des transporteurs qui eux peuvent piéger l'information contenue dans le plasma pendant 120 jours ; et l'analyse de leur contenu métabolique permet de détecter des événements complémentaires à ceux observés dans le plasma, tout en facilitant les raisonnements à partir des voies métaboliques. Ainsi, en plus des modifications de concentrations de métabolites retrouvées dans le plasma et confirmées par l'analyse métabolomique des GR, il nous a été possible de différencier les patients atteints de T1D et T2D sur la base de concentrations de métabolites impliqués dans les voies de biosynthèse et de dégradation du glutathion.

A terme, l'objectif de ce type de travaux serait de construire une banque de données relationnelles de métabolites dans différents biofluides regroupant données analytiques (spectres MS, spectres MS/MS, temps de rétention) et biologiques (statut selon l'état physiopathologique dans un biofluide donné) directement utilisables par les scientifiques de sensibilités différentes (chimistes, biochimistes, biologistes et cliniciens).

De ce fait, les enjeux majeurs de la métabolomique dans les années à venir seront, (i) de renforcer la polyvalence de la spectrométrie de masse à travers l'amélioration des performances des analyseurs existants, voire le couplage avec des approches complémentaires telle que la spectrométrie de mobilité ionique, pouvant assurer à la fois, la détection sensible de métabolites mais aussi leur identification et quantification simultanée, comme cela est déjà décrit en analyse protéomique ((Gallien et al; 2012, Gillet et al; 2012, Hopfgartner, Tonoli, & Varesio, 2012)); (ii) de développer des outils bioinformatiques permettant une exploitation plus efficace des données, notamment en termes d'analyse structurale ou encore d'inférence de réseaux métaboliques, ainsi que (iii) de renforcer les échanges, et ce en assurant la standardisation des expériences, le développement de formats de partage des données brutes et données traitées et le développement de bases de données. Dans ce sens, l'Institut Européen de Bioinformatique a développé «MetaboLights», une base de données pour le partage des expériences de métabolomique et des métadonnées associées (Salek, Haug, & Steinbeck, 2013, Steinbeck et al; 2012).

Enfin, des sociétés savantes telle que la «Metabolomic society», qui regroupe aujourd'hui plus de 500 membres établis dans plus de 20 pays ou, en France, le Réseau Français de Métabolomique et de Fluxomique ont été créées. Des initiatives de développement de plateformes nationales ont également vu le jour, notamment en Europe (Hollande : Netherlands Metabolomics Centre, et France : MetaboHUB, par exemple) et aux Etats Unis d'Amérique («NIH Common Fund's Metabolomics program»). Toutes ces actions devraient permettre à la métabolomique de mieux se structurer autour du partage et de l'échange des données. .

LISTE DES FIGURES, TABLES ET GRAPHES

Figures

Figure 1. Apport de l'utilisation des biomarqueurs au processus de recherche et développement pharmaceutique	9
Figure 2. Stratégie de découverte de biomarqueurs protéiques (adaptée de (Rifai, Gillette, & Carr, 2006))	11
Figure 3. Métabolome et analyse métabolomique (Junot et al; 2013).....	13
Figure 4. Déroulement d'une analyse métabolomique (Boudah, Paris, & Junot, 2013)	16
Figure 5. Enjeux technologiques en analyse métabolomique (adapté de Sumner LW et coll., 2003 (Sumner, Mendes, & Dixon, 2003)).....	25
Figure 6. Schéma global d'un spectromètre de masse (adapté de (Boudah, Roux, & Junot, 2012)) ...	27
Figure 7. Sources d'ionisation et molécules détectées correspondantes (Kind & Fiehn, 2010).....	28
Figure 8. Différents modes d'acquisition en MS/MS sur des spectromètres de masse équipés d'un analyseur triple quadripolaire ((Boudah, Paris, & Junot, 2013)).	29
Figure 9. Exemple de métabolite formellement identifié selon les recommandations du CAWG	37
Figure 10. Redondance du signal en ESI/MS (Boudah, Roux, & Junot, 2012).....	38
Figure 11. Constitution d'une banque de données spectrales	41
Figure 12. Démarche de quantification absolue de métabolites à l'aide de calibrants et étalons internes	44
Figure 13. Stratégie de traitement des données mise en œuvre par XCMS.....	48
Figure 14. Application de tests statistiques univariés multiples (Darghouth et al; 2011b).	53
Figure 15. Analyses statistiques multivariées (adaptée de (Roux et al; 2011))	54
Figure 16. Voies métaboliques des acides tricarboxyliques selon KEGG (Cycle de Krebs)	56
Figure 17. Calendrier des reconstructions des voies métaboliques à l'échelle du génome (Kim et al; 2012)	57
Figure 18. Etapes de reconstruction initiale du métabolome à l'échelle du génome	58
Figure 19. Lipides et fonctions biologiques (Wenk, 2005)	65
Figure 20. Structures de lipides neutres (Han & Gross, 2005)	66
Figure 21. Structures des différentes classes de la famille des glycérophospholipides (Han & Gross, 2005)	66

Figure 22. Structures des sous-classes de glycérophospholipides (Han & Gross, 2005).....	67
Figure 23. Structure des sphingolipides (Han & Gross, 2005).....	68
Figure 24. Analyse de lipides par spectrométrie de masse (adaptée de (Lam & Shui, 2013)).....	71
Figure 25. Principe des approches MDMS (Han, Yang, & Gross, 2012)	75
Figure 26. Création et implémentation de la base de données LMSD (Subramaniam et al; 2011)	80
Figure 27. Vue détaillée d'une structure implémentée dans la base de données LMSD.....	81
Figure 28. Sphingolipidoses : Enzymes, substrats et produits associés (Kolter & Sandhoff, 2006)	87
Figure 29. Voies métaboliques et biosynthèse des lipides complexes (adaptée depuis (Lamari et al; 2013))	89
Figure 30. Démarche analytique pour l'annotation du métabolome sérique	98
Figure 31. Répartition des 558 composés de référence retenus dans les trois systèmes LC-HRMS (Boudah et al; 2014)	102
Figure 32. Complémentarité des approches LC dans la rétention et la résolution des composés de référence (Boudah et al; 2014).....	103
Figure 33. Démarche d'annotation et d'identification des métabolites à partir d'un jeu de données biologiques (Boudah et al; 2014).....	104
Figure 34. L'identification des métabolites isobares dans le sérum humain : discrimination entre l'acide thréonique et l'hypoxanthine (Boudah et al; 2014).	107
Figure 35. L'identification des métabolites isomères dans le sérum humain : discrimination entre l'acide hippurique et la succinoylpyridine (Boudah et al; 2014).	108
Figure 36. Discrimination entre les isomères d'acide dihydroxybenzoïque (Boudah et al; 2014).....	109
Figure 37. Identification de sulfate de p-crésol dans le sérum humain (Boudah et al; 2014).	110
Figure 38. Identification putative de la Biliverdine dans le sérum humain (Boudah et al; 2014).....	112
Figure 39. Identification putative de l'acide 5-oxopentanoïque dans le sérum humain (Boudah et al; 2014).....	113
Figure 40. Contribution des plateformes LC-HRMS à l'extension de la couverture métabolique sérique (Boudah et al; 2014).	114
Figure 41. Métabolome sérique : Distribution des 266 métabolites sériques identifiés dans les familles chimiques (Boudah et al; 2014).	115
Figure 42. Métabolome sérique : Statut des 239* métabolites sériques identifiés selon les données bibliographiques (Boudah et al; 2014).	116
Figure 43. Contribution des plateformes LC-HRMS à l'extension de la couverture métabolique érythrocytaire.....	117
Figure 44. Métabolome érythrocytaire : Distribution des 205 métabolites érythrocytaires identifiés dans les familles chimiques	118

Figure 45. Les différentes approches analytiques évaluées pour le profilage des lipides	128
Figure 46. Détection des espèces lipidiques au sein des classes en RP(C8)-MS	128
Figure 47. Chromatogramme d'un ELT de plasma détecté en mode d'ionisation positive.....	130
Figure 48. Analyse LC-MS dans les mêmes conditions LC avec une détection en ionisation positive avec ou sans ajout de formate d'ammonium à 10 mM : Détection des esters de cholestérol endogènes	131
Figure 49. Analyse LC-MS d'ELT de plasma humain dans les mêmes conditions LC avec une détection en ionisation positive avec ou sans ajout de formate d'ammonium à 10 mM : Détection de l'étalon interne lysoglycérophosphocholine LPC 11a :0	131
Figure 50. Intérêt de l'analyse LC-MS contrairement au DIMS dans la détection d'espèces minoritaires isobares	133
Figure 51. Comparaison du nombre de molécules détectées en analyse LC-MS et DIMS d'un même échantillon d'ELT de cerveau de rat	134
Figure 52. Spectres de masse obtenus après injection directe de différentes fractions lipidiques ...	135
Figure 53. Diversité biologique : Matrices biologiques et abondances relatives des espèces lipidiques	137
Figure 54 Fractionnement des classes lipidiques à partir d'un ELT de cerveau de rat : aide à l'identification structurale.....	138
Figure 55. Chromatogrammes LC-(ESI+)MS d'ELT de cerveau de rat et de plasma avec les chromatogrammes et spectres de masse correspondants des fractions 5 et 1 respectivement	139
Figure 56. Chromatogrammes LC-(ESI-)MS d'ELT de cerveau de rat et de plasma avec les chromatogrammes et spectres de masse correspondants aux fractions 6	139
Figure 57. Fractionnement LC-(NH ₂) des classes lipidiques depuis un ELT de cerveau de rat.....	140
Figure 58. Fractionnement LC-(NH ₂) des classes lipidiques depuis un ELT de plasma humain	140
Figure 59. Spectre de fragmentation HCD de l'ion <i>m/z</i> 885.57 (LC-ESI(-)-MS/MS) annoté PI 38a :4 à partir d'un ELT de cerveau de rat.....	142
Figure 60. Discrimination par MS/MS d'espèces sulfatides à base sphingosine et à base sphingadiénine dans des ELT de cerveau de rat.	143
Figure 61. Démarche de construction de la base de données lipidomique.....	145
Figure 62. Organisation des nomenclatures selon LIPID MAPS et la base de données interne	150
Figure 63. Implémentation de la base de données lipidomique par milieu biologique : Diversité biologique et matrices biologiques : Espèces de nature différente	151
Figure 64. Interférences en détection en masse : ions monoisotopiques et isotopologues correspondants	153
Figure 65. Validation de la méthode de profilage depuis l'extraction à l'analyse LC-HRMS	155

Figure 66. Chromatogrammes LC(RPC8)-ESI(+)-MS et LC(RPC8)-ESI(-)-MS du mélange des 22 étalons internes.....	157
Figure 67. Création des ROI par l’algorithme centWave	160
Figure 68. Détection de deux étalons internes régioisomères (10µg/mL en tampon)	161
Figure 69. Comparaison de l’intégration automatique des pics chromatographiques des espèces lipidiques par XCMS (CentWave) et par le module QuanBrowser (ICIS).....	162
Figure 70. Démarche d’analyse globale des données lipidomiques	164
Figure 71. Répartition du nombre d’espèces lipidiques plasmatiques au sein des 5 grandes familles de lipides	165
Figure 72. Carte factorielle des individus obtenue en analyse RPLC-ESI-MS d’extraits métabolomiques de globules rouges humains (Darghouth et coll. non publié)	170
Figure 73. Démarche analytique et bioinformatique pour l’analyse métabolomique et lipidomique de la cohorte saclaysienne	176
Figure 74. Structure d’une séquence d’analyse LC-MS	178
Figure 75. Histogrammes de distribution des facteurs dans les séries d’analyse métabolomique	180
Figure 76. Evolution des intensités des étalons internes « prednisone » et « dihydrostreptomycine » selon l’ordre de passage au cours de l’analyse métabolomique de la première série RPLC-(ESI-)-MS	181
Figure 77. Correction du signal intra série d’analyse métabolomique et fusion des 3 séries d’analyses en RPLC-ESI(-)-MS	183
Figure 78. Evolution des intensités corrigées des étalons internes « prednisone » et « dihydrostreptomycine » selon l’ordre de passage au cours de l’analyse métabolomique de la première série RPLC-(ESI-)-MS après normalisation	184
Figure 79. Représentativité des métabolites détectés par les QCinter-séries	185
Figure 80. Carte factorielle des individus analysés en RPLC-ESI(-)-MS au cours de 3 séries d’analyse métabolomique différentes sur les axes PCA.....	186
Figure 81. Réplicas inter-analyses : validation chimométrique de la fusion des données métabolomique en RPLC-ESI(-)-MS	187
Figure 82. Echantillons atypiques non considérés	188
Figure 83. Distribution des métabolites considérés pour les analyses statistiques.....	189
Figure 84. Métabolites discriminants détectés en ZICpHILIC et PFPP selon le genre dans les analyses uni- et multivariées.....	191
Figure 85. Classes chimiques et effet du genre	192
Figure 86. Représentation de l’effet du genre sur les aires des métabolites.....	192
Figure 87. Voie métabolique de l’acide urique	193

Figure 88. Carte factorielle des individus selon les conditions chromatographiques PFPP-ESI(+)-MS, ZICpHILIC-ESI(-)-MS et leur fusion dans le modèle OPLS construit en fonction de l'âge.....	194
Figure 89. Variations des métabolites selon le BMI.....	196
Figure 90. Métabolites discriminants et leur tendance par facteur physiologique étudié.....	197
Figure 91. Comparaison des résultats statistiques obtenus pour la cohorte saclaysienne versus les résultats de Saito et coll.....	198
Figure 92. Histogrammes de distribution des facteurs dans les séries d'analyse lipidomique	199
Figure 93. Evaluation du décalage de temps de rétention observé en analyse lipidomique entre la 1 ^{ère} série d'analyse et les 2 dernières séries d'analyse (mode d'ionisation négatif).....	200
Figure 94. Fusion des données en analyse lipidomique : Exemple du DG 38 :0.....	201
Figure 95. Modélisation du décalage de temps de rétention au cours de l'analyse lipidomique RPLC-ESI(-)-MS.....	202
Figure 96. Réplicas inter-analyses : validation chimiométrique de la fusion des données lipidomiques	203
Figure 97. Espèces glycérophosphatidyléthanolamines (PE) et glycérophosphatidylcholines (PC) discriminantes selon le genre	205
Figure 98. Variations des métabolites selon l'âge	206
Figure 99. Carte factorielle des individus obtenue en analyse OPLS après analyse RP-ESI(+)-MS des empreintes lipidomiques annotées	207
Figure 100. Voies métaboliques responsables de la synthèse de céramide et de la dégradation. Les espèces de céramide impliquées dans la réponse cellulaire au stress sont représentées en majuscules. Les noms des enzymes sont en italique. Les noms des organites intracellulaires sont indiqués en italique gras. Des flèches pleines sont utilisées pour décrire les conversions métaboliques. Des flèches en pointillés indiquent des transferts assistés par des protéines. Les flèches blanches en pointillés indiquent le transport vésiculaire. Abréviations : SPT: Serine Palmitoyltransferase; Cers: synthases céramides; CERT: la protéine de transfert du céramide. SMS1 et 2: la sphingomyéline synthase 1 et 2; smpd1: Acide Sphingomyélinase, smpd2: sphingomyelinase1 neutre, smpd3: sphingomyelinase Neutre 2 CerK1, céramide Kinase1; SK: sphingosine kinase. ACeramidase: Acide céramidase. 3 voies métaboliques peuvent engendrer des céramides : (I) la synthèse depuis les sphingomyélines correspondantes, II la dégradation des sphingolipides complexes et (III) la synthèse de novo depuis la palmitoyl-CoA et la sérine. (source : http://what-when-how.com/sphingolipids-as-signaling-and-regulatory-molecules/).....	209
Figure 101. Diagnostic biologique du diabète sucré	211
Figure 102. Répartition des patients diabétiques selon les critères d'atypicité	217

Figure 103. Carte factorielle des individus obtenue en PCA et OPLS selon le type de diabète à partir des empreintes plasmatiques métabolomiques	218
Figure 104. Voies métaboliques majeures impactées par le type de diabète à partir des empreintes plasmatiques.	219
Figure 105. Espèces glycérophosphosérines et glycérophosphocholines impactées par le type de diabète.....	220
Figure 106. Répartition et effet de l'âge et du BMI sur la cohorte des diabétiques.....	222
Figure 107. Carte factorielle des individus obtenue en OPLS-DA.	222
Figure 108. Influence des facteurs confondants dans la signature métabolomique du T2D.....	223
Figure 109. Recouvrement des effets physiologiques et pathologiques sur les métabolites identifiés en conditions ZICpHILIC/ESI(-)HRMS.	224
Figure 110. Microbiote intestinal et la signature métabolomique épurée du T2D.....	224
Figure 111. Correction de l'effet de l'âge pour l'acide lactique.	225
Figure 112. Empreintes métabolomique érythrocytaire et équilibre glycémique.....	228
Figure 113. Effets de l'équilibre glycémique sur la voie métabolique de l'urée	229
Figure 114. Carte factorielle des individus obtenue en OPLS selon le type de diabète à partir des empreintes érythrocytaires métabolomiques.....	230
Figure 115. Influence du type de diabète sur le métabolisme du glutathion.	232
Figure 116. Influence du BMI sur les empreintes métabolomiques érythrocytaires.....	233
Figure 117. Métabolites plasmatiques et érythrocytaires communs discriminants selon le type de diabète.....	235
Figure 118. Aires chromatographiques du composé de référence LPS 13:0 surchargé à 5 concentrations différentes dans un ELT de cerveau de rat, un ELT de plasma et du solvant.....	271
Figure 119. Gammes d'étalonnage de 5 espèces PS et LPS surchargées dans (A) une solution de reprise, (B) de l'ELT de cerveau de rat et (C) de l'ELT de plasma humain	272

Tables

Table 1. Principaux logiciels utilisés pour le traitement des données métabolomiques (Boudah, Roux, & Junot, 2012).....	49
Table 2. validation des résultats d’alignement de différents logiciels de détection automatique des signaux (a : données de Lange et coll(Lange et al; 2008)) (Tengstrand, Lindberg, & Aberg, 2014).....	51
Table 3. Bases de données utilisées en métabolomique (adapté de Werner et coll.(Werner et al; 2008b))	60
Table 4. Approches d’analyse de lipides (Wenk, 2005)	69
Table 5. Acquisitions en PIS et NLS pour l’identification des classes et sous-classes de lipides : adaptée (Han, Yang, & Gross, 2012)	74
Table 6. Ressources et bases de données biochimiques en ligne pour la classification des lipides	83
Table 7. Principales propriétés des logiciels bioinformatiques en analyse lipidomique	86
Table 8. Applications des analyses lipidomiques	88
Table 9. Etalons internes et étalons externes utilisés en analyse métabolomique	98
Table 10. Métabolites détectés et identifiés dans du sérum humain en mode d’ionisation négative	111
Table 11. Métabolites détectés et identifiés dans du sérum humain en mode d’ionisation positive	111
Table 12. Mélange d’étalons internes utilisé : composition et concentrations correspondantes	124
Table 13. Programme d’élution du fractionnement sur colonne SPE (LC-(NH ₂)).....	125
Table 14. Gradient d’élution chromatographique en analyse RP-LCMS lipidomique.....	126
Table 15. Familles et classes lipidiques détectées par la méthode LC-MS développée	132
Table 16. Nombre d’espèces détectées par classes à partir d’un ELT de cerveau de rat en ionisation positive.....	134
Table 17. Calcul des formules brutes et des masses monoisotopiques des espèces glycérophospholipides (adaptée de (Han, Yang, & Gross, 2012)).....	147
Table 18. Composition de la base de données interne de lipides.....	148
Table 19. Annotation d’un jeu de données lipidomiques à l’aide de la base de données LIPID MAPS et la base de données interne	154
Table 20. Validation du rendement d’extraction, de la répétabilité et de la linéarité de la méthode RPLC-MS	158
Table 21. Espèces lipidiques endogènes détectées et validées dans les extraits lipidiques totaux de plasma humain	165
Table 22. Description de la cohorte saclaysienne (classification des individus selon leur genre, âge et IMC).....	175

Table 23. Tests statistiques disponibles et critères d'utilisation.....	179
Table 24. Description des matrices de données fusionnées considérées.....	189
Table 25. Espèces lipidiques annotées dans les jeux de données fusionnés de la cohorte saclaysienne et impact des facteurs physiologiques	204
Table 26. Espèces glycérophosphatidylcholines impactées par l'effet du genre.....	205
Table 27. Caractéristiques biologiques du diabète de type 1 et diabète de type 2.....	212
Table 28. Espèces lipidiques annotées dans les jeux de données lipidomiques de la cohorte Db-Juin2010 et impact du type de diabète sur leurs concentrations plasmatiques	221
Table 29. Signature métabolomique plasmatique du T2D et analyses statistiques correspondantes	226
Table 30. Performances des modèles statistiques PLS générés à partir des empreintes érythrocytaires et plasmatiques en fonction des différentes mesures de l'équilibre glycémique.	229
Table 31. Signature métabolomique érythrocytaire du T2D.....	231
Table 32. Comparatif des empreintes métabolomiques plasmatiques et érythrocytaires.....	234

ANNEXE 1

RAPPELS THEORIQUES : LOGICIELS DE DETECTION AUTOMATIQUE DES SIGNAUX

Données d'entrée

Comme les données brutes LC-MS sont enregistrées dans des formats propriétaires, la plupart des logiciels libres et gratuits ne peuvent les lire directement. Une conversion –à l'aide des outils de conversion proposés par les fournisseurs de spectromètres de masse- dans un format universel comme «NetCDF» mzXML est obligatoire (Katajamaa & Oresic, 2005). A titre d'exemple le logiciel «MZmine 2» accepte différents formats de données en entrée tel que mzML, mzXML (Lin et al; 2005, Pedrioli et al; 2004), mzData , NetCDF et RAW des instruments Thermo Fisher Scientific.

Soustraction du bruit de fond

Tout d'abord, le filtrage spectral assure la correction de la ligne de base et la réduction du bruit de fond aléatoire en utilisant des filtres de lissage (Griffiths et al; 2010) tels que la transformée en ondelettes ou le filtre gaussien. Ce dernier supprime les "pics d'épaulement" qui correspondent aux pics artéfactuels générés par le calcul de la transformée de Fourier pour les instruments de type FTICR et Orbitrap. Cette étape est facultative et peut être omise si les données sont acquises en mode centroïde.

Détection des pics

Cette étape permet de mesurer les hauteurs et aires sous de chaque pic chromatographique dans MZmine et uniquement ces dernières dans XCMS (Smith et al; 2006).

Différents algorithmes sont disponibles. MZmine utilise une détection vectorisée des pics c'est à dire qu'il effectue la détection indépendamment dans les deux domaines « m/z » et «temps de rétention». Une contrainte supplémentaire quant à la forme des pics autorisée est assurée par une «méthode de maximum local" ou "méthode de seuil récursive". XCMS opère selon un mode opératoire différent. Il exploite d'abord le domaine spectral puis le domaine chromatographique. «Matched Filter», -le premier algorithme XCMS développé, coupe les données LC-MS en tranches spectrales (la tranche par défaut «*step* » est de 0.1 m/z). La largeur à mi-hauteur des pics chromatographiques est fixée par le paramètre «*fwhm*» et le calcul de la dérivée seconde de la fonction gaussienne (utilisée pour modéliser la forme des pics chromatographique) permet de trouver le point d'inflexion du pic chromatographique. Néanmoins, définir une taille optimale de la

tranche m/z est difficile. En effet, si le $\Delta m/z$ est trop petit, les pics chromatographiques peuvent se partager entre 2 tranches adjacentes et perdent donc leur forme. Par contre, si la distance est surestimée, les pics peuvent se superposer et les pics les moins intenses ne seront plus détectables du fait de l'augmentation du bruit de fond. Un algorithme alternatif appelé "centWave" a été développé par la suite. Celui-ci permet de ne pas fixer de tranche m/z mais plutôt (Tautenhahn, Bottcher, & Neumann, 2008) de détecter des régions d'intérêt (ROI) dans le domaine spectral de façon à adapter le découpage aux données.

Alignement des pics

L'alignement consiste à faire coïncider les pics correspondants à une même variable à travers tous les échantillons si leurs masses et temps de rétention diffèrent peu (écarts dans la limite d'une tolérance spécifiée) de sorte à ne définir qu'une seule variable. Ce processus aboutit à une matrice de données dans laquelle chaque échantillon est représenté. Cette étape est cruciale car il peut y avoir des décalages de temps de rétention et une dérive de la précision en masse au cours d'une analyse LC-MS. Les seuils de tolérance doivent donc être bien réfléchis. Ils dépendent de la méthode chromatographique et de la précision de mesure en masse du spectromètre utilisé, et ils sont spécifiés par l'utilisateur.

"Join aligner" est l'algorithme utilisé dans MZmine. Il crée d'abord une liste princeps de pics à partir d'un spectre puis aligne les listes de pics générées des autres échantillons avec. «Random Sample Consensus Aligner», mis en œuvre dans «MZmine2», peut gérer de forts écarts de temps de rétention, mais est du coup susceptible d'introduire des valeurs aberrantes (Pluskal et al; 2010).

La méthode dite «group nearest» est implémentée dans le logiciel XCMS. Elle aligne les données provenant de différents échantillons selon des fenêtres de m/z et temps de rétention définies. Contrairement à «Join aligner», l'alignement se fait simultanément dans tous les échantillons simultanément au lieu d'un appariement par paires. Ceci évite ainsi les problèmes liés à la sélection d'un échantillon de référence. L'algorithme «group density» mis en œuvre par défaut dans XCMS quant à lui assure un alignement d'abord dans le domaine spectrale en prenant en compte la précision de mesure en masse du spectromètre de masse utilisé. Il y a découpage arbitraire des intervalles m/z selon le paramètre «*mzwid*» défini avec un chevauchement de la moitié des intervalles afin d'éviter de scinder les pics en deux. Cet algorithme calcule également lors la répartition globale des pics chromatographiques et identifie de manière dynamique les régions où de nombreux pics présentent des temps de rétention similaires. Ces derniers seront lissés et regroupés (fonction «*bw*»).

Des contraintes quant à la présence de la variable dans le jeu de données peuvent être imposées à ce stade grâce aux valeurs de «*minfrac*» et «*minsamp*». Enfin, après l'alignement, les pics n'ayant pas de correspondance dans certains échantillons sont traités par des algorithmes tels

que « Gap filler » de MZmine ou « fillPeaks » de XCMS. Ils permettent d'éviter ces valeurs manquantes, en estimant hauteur et/ou surface des pics manqués.

Normalisation

Certains logiciels effectuent une étape de «normalisation» pour réduire l'erreur systématique non désirée entre des échantillons en ajustant les intensités. MZmine proposent deux approches différentes : "normalisation linéaire" par le même facteur qui pourrait être : (i) la hauteur moyenne des pics, (ii) la moyenne des hauteurs maximales au carré, (iii) la hauteur maximum du pic ou du signal brut total. Une autre alternative consiste en l'utilisation d'un ou plusieurs étalons internes ajoutés dans chaque échantillon pour calculer un facteur de correction optimal nommée «NOMIS» de l'anglais NOrmalization Multiple Internal Standards (Katajamaa & Oresic, 2007, Sysi-Aho et al; 2007). Cette dernière approche a été publiée pour traiter des données d'analyses lipidomiques.

ANNEXE 2

APPROCHE GLOBALE DE PROFILAGE METABOLOMIQUE

LC conditions	C8	PFPP	HILIC
fwhm	4	10	10
step	0.005	0.005	0.005
steps	2	2	2
mzdiff	0.005	0.005	0.005
sleep	0	0	0
max	50	50	50
snthresh	5	5	5
index	FALSE	FALSE	FALSE
profmeth	bin	bin	bin
minfrac_GA	0.75	0.75	0.75
minsamp_GA	1	1	1
bw_GA	5	5	5
mzwid_GA	0.01	0.01	0.01
max_GA	50	50	50
sleep_GA	0	0	0
missing	1	1	1
extra	1	1	1
method	loess	loess	loess
span	0.2	0.2	0.2
family	gaussian	gaussian	gaussian
plotype	deviation	deviation	deviation
minfrac_GB	0.75	0.75	0.75
minsamp_GB	1	1	1
bw_GB	5	5	5
mzwid_GB	0.01	0.01	0.01
max_GB	50	50	50
sleep_GB	0	0	0
class1	BLC	BLC	BLC
class2	BIO	BIO	BIO
filebase	table	table	table
eicmax	50	50	50
eicwidth	200	200	200
sortpval	TRUE	TRUE	TRUE
metlin	0	0	0
sigma	6	6	6
perfwhm	0.6	1	1
cor_eic_th	0.6	0.6	0.6
maxcharge	2	2	2
maxiso	2	2	2
ppm	5	5	5
polarity	negative/positive	negative/positive	negative/positive

Table S.1 : Paramètres XCMS de traitement du signal LC-HRMS

UPLC-C18												
		Pos						Neg				
Compound	Exact Mass	Retention time			Mass accuracy (ppm)	Retention time			Mass accuracy (ppm)	Retention time		
		Rt (min)	Maximum shift (mn)	CV ² (%)		Rt (min)	Maximum shift (mn)	CV ² (%)		Rt (min)	Maximum shift (mn)	CV ² (%)
ExS Alanine ¹³ C	90.05103	0.88	0.18	4.7	-0.1	\	\	\	\	\	\	
ExS Metformin	129.10145	0.83	\	\	-1.25	\	\	\	\	\	\	
ExS Ethylmalonic acid	132.04226	\	\	\	\	2.82	0.16	1.71	-2.23	0.07	2.23	
ExS Aspartate ¹⁵ N	134.03454	0.81	0.26	7.46	-1.78	0.92	0.07	2.23	-2.82	0.1	3.24	
ExS Glucose ¹³ C	181.06674	0.85	0.25	6.66	nd	0.93	0.1	3.24	-2.93	\	\	
ExS 2-Aminoanthracene	193.08915	8.05	0.24	0.82	-0.9	\	\	\	\	\	\	
ExS Amiloride	229.04789	3.1	0.5	4.56	-1.2	3.22	0.21	1.87	-1.22	\	\	
ExS Imipramine	280.19395	7.5	0.17	0.63	-1.86	\	\	\	\	\	\	
ExS Atropine	289.16779	5.2	0.18	0.99	-1.2	\	\	\	\	\	\	
ExS Prednisone	358.17802	6.75 and 7.00	0.17	0.67	-0.86	6.75 and 7.00	0.07	0.3	-0.2	\	\	
ExS Coldicine	389.16819	6.44	0.15	0.71	-0.93	\	\	\	\	\	\	
ExS Dihydrostreptomycin	583.28132	0.72	0.13	4.7	-1.5	0.73	0.04	1.88	-0.8	\	\	
ExS Roxithromycin (fragment)	590.37784	5.87 and 6.08	0.14	0.66	-2.8 and -5	5.91 and 6.11	0.06	0.32	-0.6	\	\	
IS Dimetridazole	141.05383	3.3	0.74	5.45	-0.37	\	\	\	\	\	\	
IS AIMP 2-amino-3-(3-hydroxy-5-methylisoxazol-4-yl)propanoic acid	186.06406	1.1	0.19	4.54	-2.2	1.1	0.06	1.67	-2.5	\	\	
IS IMCPA 2-methyl-4-chlorophenoxyacetic acid	200.02402	\	\	\	\	8.05	0.1	0.35	-3.25	\	\	
IS Dinoseb	240.07462	\	\	\	\	10.15	0.11	0.3	-1.6	\	\	

Table S.2 : Performances de la méthode RPLC-HRMS

		HSF5									
		Pos					Neg				
Compound	Exact Mass	Retention time			Mass measurement Mass accuracy (ppm)	Retention time			Mass measurement Mass accuracy (ppm)		
		Rt (min)	Maximum shift (mn)	CV (%)		Rt (min)	Maximum shift (mn)	CV (%)			
ExS Alanine ¹³ C	90.051	2.35	0.12	1.48	-0.2	\	\	\	\	\	
ExS Metformin	129.101	nd	\	\	1	\	\	\	\	\	
ExS Ethylmalonic acid	132.042	\	\	\	\	5.75	0.35	1.39	0.7	0.7	
ExS Aspartate ¹⁵ N	134.035	2.06	0.14	1.5	-1.6	2.06	0.06	0.76	0.7	0.7	
ExS Glucose ¹³ C	181.067	1.7	\	\	-2	1.67	0.05	0.82	0.4	0.4	
ExS 2-Aminoanthracene	193.089	18.87	0.19	0.28	-3.3	\	\	\	\	\	
ExS Amiloride	229.048	11.5	0.46	1.37	-7.2	12.48	0.51	0.9	0.8	0.8	
ExS Imipramine	280.194	nd	\	\	nd	\	\	\	\	\	
ExS Atropine	289.168	9.1	0.37	1.34	-3.7	\	\	\	\	\	
ExS Prednisone	358.178	10.5 and 10.80	0.07	0.24	-3.6	10.50 and 10.80	0.08	0.21	0.8	0.8	
ExS Colchicine	399.168	9.7	0.08	0.25	-3.3	\	\	\	\	\	
ExS Dihydrostreptomycin	583.281	nd	\	\	nd	nd	\	\	nd	nd	
ExS Roxithromycin (fragment)	590.378	nd	\	\	nd	nd	\	\	nd	nd	
IS Dimetridazole	141.054	8.3	0.15	0.67	-3.6	\	\	\	\	\	
IS AMPA 2-amino-3-(3-hydroxy-5-methyl-isoxazol-4-yl)propanoic acid	186.064	3.55	0.14	1.03	-2.5	3.55	0.15	1.22	-2.6	-2.6	
IS MCPA 2-methyl-4-chlorophenoxyacetic acid	200.024	\	\	\	\	13.02	0.06	0.13	-4	-4	
IS Dinoseb	240.075	\	\	\	\	16.1	0.07	0.1	-1.3	-1.3	

Table S.3 : Performances de la méthode PFPF-HRMS

		HSF5					
		Pos			Neg		
Compound	Exact Mass	Retention time		Mass measurement Mass accuracy (ppm)	Retention time		Mass measurement Mass accuracy (ppm)
		Rt (min)	Maximum shift (mn)		CV (%)	Rt (min)	
ExS Alanine ¹³ C	90.051	2.35	0.12	1.48	\	\	\
ExS Metformin	129.101	nd	\	\	\	\	\
ExS Ethylmalonic acid	132.042	\	\	\	5.75	0.35	1.39
ExS Aspartate ¹⁵ N	134.035	2.06	0.14	1.5	2.06	0.06	0.76
ExS Glucose ¹³ C	181.067	1.7	\	\	1.67	0.05	0.82
ExS 2-Aminoanthracene	193.089	18.87	0.19	0.28	\	\	\
ExS Amiloride	229.048	11.5	0.46	1.37	12.48	0.51	0.9
ExS Imipramine	280.194	nd	\	\	\	\	\
ExS Atropine	289.168	9.1	0.37	1.34	\	\	\
ExS Prednisone	358.178	10.5 and 10.80	0.07	0.24	10.50 and 10.80	0.08	0.21
ExS Colchicine	399.168	9.7	0.08	0.25	\	\	\
ExS Dihydrostreptomycin	583.281	nd	\	\	nd	\	\
ExS Roxithromycin (fragment)	590.378	nd	\	\	nd	\	\
IS Dimetridazole	141.054	8.3	0.15	0.67	\	\	\
IS AMPA 2-amino-3-(3-hydroxy-5-methylisoxazol-4-yl)propanoic acid	186.064	3.55	0.14	1.03	3.55	0.15	1.22
IS MCPA 2-methyl-4-chlorophenoxyacetic acid	200.024	\	\	\	13.02	0.06	0.13
IS Dinoseb	240.075	\	\	\	16.1	0.07	0.1

Table S.3 : Performances de la méthode PFP-PP-HRMS

		ZICpHILIC										
		Pos					Neg					
Compound	Exact Mass	Retention time			Mass measurement (ppm)	Retention time			Mass measurement (ppm)	Retention time		
		Rt (min)	Maximum shift (mn)	CV (%)		Rt (min)	Maximum shift (mn)	CV (%)		Rt (min)	Maximum shift (mn)	CV (%)
ExS Alanine ¹³ C	90.05103	7.08	0.13	0.48	-2.3	\	\	\	-2.3	\	\	\
ExS Metformin	129.10145	8.12 / 9.35 / 7	0.14	0.29	-2.7	\	\	\	-2.7	\	\	\
ExS Ethylmalonic acid	132.04226	\	\	\	\	7.65	0.11	0.34	\	8.08	0.1	0.31
ExS Aspartate ¹⁵ N	134.03454	8.06	0.08	0.32	-0.43	7.2	0.32	0.96	-0.43	7.2	0.32	0.96
ExS Glucose ¹³ C	181.06674	nd	\	\	\	\	\	\	\	\	\	\
ExS 2-Aminoanthracene	193.08915	2.09	0.07	1.12	-5.9	\	\	\	-5.9	\	\	\
ExS Amiloride	229.04789	6.5 / 7.03	0.65	4.31	-8	6.7	0.46	2.9	-8	6.7	0.46	2.9
ExS Imipramine	280.19395	1.9	0.16	2.64	-5	\	\	\	-5	\	\	\
ExS Atropine	289.16779	2 à 3	0.52	5.73	-3.7	\	\	\	-3.7	\	\	\
ExS Prednisone	358.17802	2	0.1	1.43	-5.7	1.95	0.07	1.19	-5.7	1.95	0.07	1.19
ExS Colchicine	399.16819	1.9	0.12	1.83	-4.5	\	\	\	-4.5	\	\	\
ExS Dihydrostreptomycin	583.28132	nd	\	\	nd	nd	\	\	nd	nd	\	rd
ExS Roxithromycin (fragment)	590.37784	1.86	0.08	1.08	-3	1.8	0.04	0.71	-3	1.8	0.04	0.71
IS Dimetridazole	141.05383	2.1	0.07	1.15	-3.8	\	\	\	-3.8	\	\	\
IS AMPA 2-amino-3-(3-hydroxy-5-methyl-isoxazol-4-yl)propanoic acid	186.06406	7.1	0.25	1.25	-5.8	7.1	0.21	1.16	-5.8	7.1	0.21	1.16
IS MCPA 2-methyl-4-chlorophenoxyacetic acid	200.02402	\	\	\	\	1.9	0.04	1.07	\	1.9	0.04	1.07
IS Dinoseb	240.07462	\	\	\	\	1.7	0.03	1.58	\	1.7	0.03	1.58

Table S.4 : Performances de la méthode ZICpHILIC-HRMS

MI	Detection in the positive ionization mode	Detection in the negative ionization mode	Chemical class	Identified metabolite	Already identified in human serum samples	Status of identification	Ret HILIC	Ret PPPP	Ret OS
88.017	nd	87.008	Keto-Acids	Pyruvic acid	yes	a,b	4.14		
88.017	nd	87.008	Keto-Acids	Pyruvic acid	yes	a,b	3.28		
89.047	90.055	88.040	Amino Acids	Alanine	yes	a,b,c	7.10	1.94	0.91
89.047	90.055	nd	Amino Acids	Sarcosine	yes	a,b		2.34	
90.032	nd	89.024	Hydroxy Acids	Lactic acid	yes	a,b,c	4.11		
101.032	nd	101.024	Keto-Acids	Ketobutyric acid	yes	a,b,c	7.33		
101.069	nd	101.060	Fatty Acids	Valeric acid	yes	a,b	2.63		5.15
103.062	104.070	nd	Amino Acids	aminobutyric acid	yes	a,b,c	5.34	2.65	0.94
103.062	104.071	nd	Amino Acids	N,N-DIMETHYLGLYCINE	yes	a,b,c		2.10	
104.048	nd	103.039	Hydroxy Acids: Fatty Acids	Hydroxybutyric acid isomer1	yes	a,b,c	4.04	3.35	1.60
104.048	nd	103.040	Hydroxy Acids: Fatty Acids	Hydroxybutyric acid isomer2	yes	a,b,c	3.33		1.33
105.041	106.049	104.035	Amino Acids	Serine	yes	a,b,c	7.37	2.13	0.62
106.027	nd	105.019	Hydroxy Acids	Glycnic acid	yes	a,b,c	5.77	1.89	0.93
110.037	nd	109.029	Phenol and derivatives	Hydroquinone	yes	a,b	2.34		3.35
112.028	nd	111.019	Pyrimidine and derivatives	Uracil	yes	a,b,c	4.01		
113.057	114.066	112.051	Amino Ketones	Creatinine	yes	a,b,c	3.60	3.68	0.93
113.058	114.066	nd	Amino Ketones	Creatinine isom	No	a,b		4.32	
115.062	116.070	nd	Amino Acids	Proline	yes	a,b,c	6.13	2.32	0.94
116.084	nd	115.076	Fatty Acids	Caproic acid	yes	a,b	2.04		
117.043	118.050	116.035	Acyl Glycines	N-acetyl glycine	yes	a,b	4.48	2.29	1.04
117.053	118.061	nd	Amino Acids	GUANIDOACETIC ACID	yes	a,b,c	7.92	2.34*	1.04
117.077	118.086	116.071	Amino Acids	Valine	yes	a,b,c	5.14	3.76	1.02
117.078	118.086	152.048 (Cl-)	Amino Acids	Betaine	yes	a,b,c		2.10	
118.027	nd	117.019	Dicarboxylic Acids	Succinic acid	yes	a,b,c	6.12	3.00	1.32
118.064	nd	117.055	Hydroxy Acids	2 hydroxy methylbutyric acid	yes	a,b,c	2.73	5.82	3.30
118.064	nd	117.055	Hydroxy Acids	3 hydroxy methylbutyric acid	yes	a,b,c	2.70	5.84	1.96
119.057	120.065	118.050	Amino Acids	THREONINE/Homoserine	yes	a,b,c	7.03	2.21	0.92
120.043	nd	119.034	Alcohols and polyols	Erythrose	yes	a,b	4.56	2.02	0.90
122.047	123.055	nd	Nucleotides	NICOTINAMIDE	yes	a,b,c	2.69		
125.015	126.022	124.007	Amino Acids	Taurine	yes	a,b,c	7.11	1.65	0.92
125.024	nd	124.016	NA	Aminomethylphosphonic acid isomer	No	a,b	7.03	2.40	1.00
125.057	126.065	nd	Pyrimidines and Derivatives	5-METHYLCTOSINE	No	a,b,c	3.87*		3.300
126.008	nd	125.000	Alcohols and Polyols	Glyceraldehyde-Cl	yes	a,b	1.68	2.39	1.11
128.057	129.066	nd	Pyrimidines and Derivatives	DIHYDROTHYMINE isomer	yes	a,b	7.51	2.80	0.93*
129.043	130.050	128.035	Amino Acids: Amino Ketones	Pyroglutamic acid	yes	a,b,c	4.63	2.60	1.16
129.078	130.086	128.071	Amino Acids	Pipecolic acid	yes	a,b	4.89	2.29	

M	Detection in the positive ionization mode	Detection in the negative ionization mode	Chemical class	Identified metabolite	Already identified in human serum samples	Status of identification	Rt HILIC	Rt PFPF	Rt C8
129.078	130.086	nd	Amino Acids	Pipecolic acid isomer2	No	a,b,c	3.41	1.04	
129.078	130.086	nd	Amino Acids	Pipecolic acid isomer3	No	a,b	3.14	3.96	
130.064	nd	129.055	Keto-Acids	3-METHYL-2-OXOVALERIC ACID	yes	a,b,c	1.97	7.62	
130.064	nd	129.055	Keto-Acids	4-METHYL-2-OXOVALERIC ACID	yes	a,b,c	8.30	4.41	
131.052	132.061	130.051	Amino Acids	N-ACETYL-L-ALANINE	No	a,b	3.77*	6.52*	
131.059	132.065	130.050	Amino Acids	Hydroxyproline isomer	yes	a,b	7.43	2.00	
131.068	132.076	130.062	Amino Acids	Creatine	yes	a,b,c	7.45	2.84	
131.069	132.077	nd	Amino Acids	Creatine isom	No	a,b,c	3.16	3.16	
131.083	132.101	130.086	Amino Acids	Leu/ile/Norleu	yes	a,b,c	3.97	6.99	
131.084	132.102	130.087	Amino Acids	Leu/ile/Norleu	yes	a,b,c	4.28	6.52	
132.054	133.060	131.046	Amino Acids	Asparagine	yes	a,b,c	7.23	2.11	
132.089	133.097	131.082	Amino Acids	Ornithine	yes	a,b,c	8.70	3.39	
133.036	134.045	132.030	Amino Acids	Aspartic acid	yes	a,b,c	8.07	2.05	
133.052	134.060	nd	Indoles and Derivatives	Hydroxyindole	yes	a,b	2.61	6.42*	
134.022	nd	133.014	Dicarboxylic Acids	Malic acid	yes	a,b,c	8.45	2.07	
135.034	136.042	nd	Amino Acids	HOMOCYSTEINE	yes	a,b	2.53	2.53	
136.038	137.045	135.029	Hydroxy Acids	Threonic acid	yes	a,b,c	7.01	2.40	
136.039	137.045	135.030	Purines and Purine Derivatives	Hypoxanthine	yes	a,b,c	4.44	2.58	
137.046	138.055	nd	Pyridines and Derivatives	Trigonelline	yes	a,b,c	5.46	2.38	
137.048	nd	136.040	Benzoic acid and derivatives	SAUCYLAMIDE	yes	a,b	1.99	6.34	
138.032	nd	137.024	Benzoic acid and derivatives	Salicylic acid	yes	a,b,c	2.08	11.33	
139.028	nd	138.019	Phenol and derivatives	4-NITROPHENOL	yes	a,b,c	2.15	11.40	
140.057	141.065	nd	Imidazoles	METHYLIMIDAZOLEACETIC ACID	yes	a,b	3.92*	6.43	
143.093	144.101	nd	Amino Acids	STACHYDRINE	yes	a,b,c	4.65	2.55	
145.072	146.081	nd	Amino Ketones; Amino Acids	4-ACETAMIDOBUTANOIC ACID isomer	yes	a,b,c	5.51*	1.17	
145.074	nd	144.066	Acyl glycines; Amino Acids	i-butrylylglycine	yes	a,b,c	2.86	1.50	
145.074	146.080	144.066	Amino Ketones; Amino Acids	4-ACETAMIDOBUTANOIC ACID	yes	a,b,c	3.87	2.66	
145.084	146.092	nd	Amino Acids	4-GUANIDINOBTANOIC ACID	yes	a,b	7.66	1.40	
146.022	nd	145.014	Dicarboxylic Acids	KETOGLUTARIC ACID	yes	a,b	8.21	1.98	
146.056	147.064	nd	Dicarboxylic Acids	Methylglutaric acid/ Dimethylsuccinic acid	yes	a,b	7.24	2.17	
146.057	147.076	145.061	Amino Acids	GLUTAMINE	yes	a,b,c	7.24	0.85	
146.068	147.076	145.062	Amino Acids	GLUTAMINE isomer	yes	a,b	2.54	6.27	
146.085	nd	145.086	Hydroxy acids	Hydroxyheptanoic acid	No	a,b,c	9.07	3.72	
146.104	147.113	145.097	Amino Acids	Lysine	yes	a,b,c	5.02	0.79	
147.053	nd	146.045	Amino Acids	O-acetylserine	yes	a,b,c	8.02	2.19	
147.053	148.060	146.045	Amino Acids	Glutamic acid	yes	a,b,c	8.02	0.91	

MI	Detection In the positive ionization mode	Detection in the negative ionization mode	Chemical class	Identified metabolite	Already identified in human serum samples	Status of identification	Rt HILIC	Rt PFPF	Rt CB
149.050	150.058	148.043	Amino Acids	Methionine	yes	a,b,c	4.05	4.26	1.13
150.053	nd	149.045	Carbohydrates	Pentose	yes	a,b,c	7.00	3.35	
150.053	nd	149.045	Carbohydrates	Pentose	yes	a,b,c	7.36	1.89	0.90
151.061	152.070	nd	Amino Acids and derivatives	Phenylglycine/Acetoxyarylamine	yes	a,b,c	6.48 ⁿ	3.73	1.59
151.062	152.070	nd	Phenol and derivatives	Acetaminophen	yes	a,b,c	2.23 ⁿ	6.21	2.89
152.034	153.040	151.025	Purines and Purine Derivatives	Xanthine-isomer1	yes	a,b,c	4.59	2.62	1.16
152.034	nd	151.025	Purines and Purine Derivatives	Xanthine-isomer2	No	a,b,c	5.04	2.82	
152.048	nd	151.039	Benzoic acid and derivatives	p-anisic acid	No	a,b,c	2.03	8.53	
154.027	nd	153.019	Aromatic Acids	2,6-DIHYDROXYBENZOIC ACID	yes	a,b,c	1.96	11.13	5.00
154.027	nd	153.019	Aromatic Acids	2,5-DIHYDROXYBENZOIC ACID	yes	a,b,c	3.03	8.17	4.30
155.068	156.076	154.062	Amino Acids	Histidine	yes	a,b,c	7.06	2.66	0.80
155.068	156.076	nd	Amino Acids	Histidine isomer	No	a,b,c		3.34	
157.074	nd	156.066	Amino Acids	N-Tigloylglycine/Methylcrotonylglycine	yes	a,b	1.92		
158.044	nd	157.036	Amino Ketones	Allantoin	yes	a,b,c	7.01	1.85	0.90
159.090	nd	158.082	Acyl Glycines	ISOVALERYLGLYCINE	yes	a,b	2.26		
159.090	nd	158.082	Amino acids	N-Acetyl-valine/N-Acetyl-norvaline	No	a,b,c	2.65	7.90	5.09
160.110	nd	159.102	Fatty acids	methyladipic acid/ dimethylglutaric acid/Plimelic acid	yes	a,b,c	1.94	7.35	4.73
161.059	nd	160.061	Amino Acids	O-Acetyl-L-homoserine	No	a,b,c	4.17	2.55	
161.103	162.112	nd	Carnitines and derivatives	Carnitine	yes	a,b,c	6.90	3.58	0.88
161.104	162.112	nd	Carnitines and derivatives	Carnitine isomer	No	a,b,c	7.39	4.23	
162.053	nd	161.045	Hydroxy acids	Hydroxymethylglutaric acid	yes	a,b,c			
164.048	nd	163.040	Aromatic Acids	Phenylpyruvic acid	yes	a,b	1.96		
165.077	166.085	164.071	Amino Acids	Phenylalanine	yes	a,b,c	3.39	8.05	2.17
165.078	166.086	164.071	Amino Acids	Phenylalanine isomer	No	a,b,c	2.95	2.55	
166.064	nd	165.055	Phenol and derivatives	3-Methoxyphenylacetic acid / 2-Phenoxypropionic acid / 3-(4-Hydroxyphenyl)propionic acid / 4-(Hydroxymethyl)phenylacetic acid	yes	a,b	2.85	11.45	7.10
166.064	nd	165.055	Phenol and derivatives	4-(Hydroxymethyl)phenylacetic acid / 3-Phenoxypropionic acid	yes	a,b	1.97	8.81	5.30
166.100	nd	165.092	Aromatic Acids	Perillic acid	yes	a,b,c	1.88		0.90
167.022	nd	166.014	Dicarboxylic Acids	QUINOLINIC ACID	yes	a,b,c	8.08		
168.029	nd	167.020	Purines and Purine Derivatives	Uric acid	No	a,b,c	6.27		
168.029	169.035	167.020	Purines and Purine Derivatives	Uric acid	yes	a,b,c	6.95	2.23	1.00
168.043	nd	167.034	Benzoic acids and derivatives	Methoxysalicylic acid	yes	a,b		10.21	5.85
169.072	170.080	nd	Pyridoxals and Derivatives	Pyridoxine	yes	a,b	3.34		
169.084	170.092	168.077	Amino Acids	Methylhistidine	yes	a,b,c	5.57	3.34	0.83 ⁿ
169.084	170.092	nd	Amino Acids	Methylhistidine isomer	yes	a,b		2.28	

M	Detection in the positive ionization mode	Detection in the negative ionization mode	Chemical class	Identified metabolite	Already identified in human serum samples	Status of identification	Rt HILIC	Rt EPPF	Rt CS
172.014	nd	171.006	Acyl Phosphates	Glycerophosphoric acid- isomer1	yes	a,b,c	7.69		
172.014	nd	171.006	Acyl Phosphates	Glycerophosphoric acid- isomer2	yes	a,b,c	8.11		
172.083	173.091	nd	Polypeptides/NA	PRO-GLY	yes	a,b	7.24		1.29
173.104	174.112	nd	Amino acids; Acyl glycines	Hexamoyl glycine	yes	a,b	3.34*		1.12
173.999	nd	172.991	Phenol and derivatives	Phenyl sulfate	yes	a,b,c	2.08	6.01	4.00
174.101	nd	173.092	Amino Acids	Acetylornithine	yes	a,b,c	6.77	8.05	1.38
174.110	175.119	173.104	Amino Acids	Arginine	yes	a,b,c	9.53	4.09*	0.81
175.048	nd	174.040	Amino Acids	N-ACETYL-L-ASPARTIC ACID	yes	a,b,c	7.91	2.61	1.14
175.062	176.070	174.056	Indoles and Derivatives	Indoleacetic acid	yes	a,b,c	3.8 and 8.1	10.25	6.32
175.094	176.102	174.088	Amino Acids	CITRULLINE	yes	a,b,c	7.59	2.28	0.93
176.032	nd	175.024	Hydroxy Acids	Ascorbic acid	yes	a,b,c	8.06	2.60	
176.094	177.102	nd	Amino Ketones	COTININE	yes	a,b,c	2.09	6.38	15 and 1
179.059	180.065	178.051	Acyl Glycines; Amino Alcohols	Hippuric acid- isomer1	yes	a,b,c	2.28	7.68	4.64
179.059	nd	178.051	Acyl Glycines; Amino Alcohols	Hippuric acid- isomer2	No	a,b,c	2.65		
180.063	181.071	179.057	Purines and Derivatives	THEOBROMINE	yes	a,b,c	1.66 and 3.	6.37	3.71
180.063	181.072	179.057	Purines and Derivatives	Theophylline	yes	a,b,c	3.6 and 3.3.	5.64	2.60
180.064	181.072	179.056	Alcohols and Polyols	Hexose isomer1	yes	a,b,c	7.06	1.66	0.82
180.064	nd	179.055	Alcohols and Polyols	Hexose isomer2	yes	a,b,c	7.35	2.44	1.12
180.064	nd	179.056	Alcohols and Polyols	Hexose isomer3	yes	a,b	8.74		
181.072	182.080	180.066	Amino Acids	Tyrosine	yes	a,b,c	5.32	6.03	1.23
182.058	nd	181.050	Aromatic Acids	P-HYDROXYPHENYLACTIC ACID	yes	a,b,c	3.56	6.78	3.89
182.079	nd	181.071	Alcohols and Polyols	Mannitol	yes	a,b,c	7.05		
183.052	184.060	182.045	Pyridoxals and Derivatives	4-PYRIDOXIC ACID	yes	a,b,c	2.11	5.25	1.40
183.124	184.133	nd	Phenylacetic Acid Derivatives	Acetyltropine	No	a,b	8.01	8.07	1.11
185.009	nd	184.001	Amino Acid Phosphates	PHOSPHOSERINE	yes	a,b	6.19		0.92
186.030	nd	185.021	Carbohydrates	Fructose-Cl- isomer3	yes	a,b	1.91	7.47	5.35
188.015	nd	187.007	Carboxylic acids ; Sulfate esters	p-Cresol sulfate	yes	a,b,c	6.21		
188.045	nd	187.037	Carbohydrates	L-ARABITOL / ADONITOL-Cl	yes	a,b	9.87	4.33*	0.91
188.125	189.134	nd	Amino Acids	Homocysteine	yes	a,b,c	1.84	12.75	8.64
188.141	nd	187.133	Hydroxy acids	HYDROXYDECANOIC ACID (3)	No	a,b,c	8.69	2.96	0.81
188.141	nd	187.133	Hydroxy acids	HYDROXYDECANOIC ACID- isomer1	No	a,b,c	7.70		
188.151	189.159	nd	Amino Acids	NE, NG- TRIMETHYL- L- LYSINE	yes	a,b,c	2.68	6.60	
189.064	nd	188.056	Amino Acids	N-ACETYL- L- GLUTAMIC ACID	No	a,b,c	9.40	2.44	1.12
191.062	nd	190.053	Amino Acids	N-ACETYL- L- METHIONINE	No	a,b,c	7.02	1.85	0.90
192.028	5.0161385 (Na add.)	191.019	Tricarboxylic acids	Citric acid	yes	a,b,c			
192.064	nd	191.055	Hydroxy Acids	QUINIC ACID	yes	a,b,c			

M	Detection in the positive ionization mode	Detection in the negative ionization mode	Chemical class	Identified metabolite	Already identified in human serum samples	Status of identification	Rt HILIC	Rt PFPF	Rt C8
194.043	nd	193.034	Carbohydrates	GALACTURONIC ACID isomer1	yes	a,b,c	8.41	1.65	
194.043	nd	193.035	Carbohydrates	GALACTURONIC ACID isomer2	yes	a,b,c	7.62		
194.079	195.087	nd	Purines and Derivatives	CAFFEINE	yes	a,b,c	2.12	7.25	4.51
196.058	nd	195.050	Hydroxy Acids	Gluconic acid	yes	a,b,c	7.26	6.21*	
196.058	nd	195.050	Hydroxy Acids	Galactonic acid	yes	a,b,c	7.50		
196.058	nd	195.050	Purines and purine derivatives	1,7-Dimethyluric acid	yes	a,b	3.36	6.21*	
197.082	nd	196.074	Quaternary Amines	Carnitine-C1	yes	a,b	6.86		
200.045	nd	199.037	Carbohydrates	1,5-ANHYDRO-D-SORBITOL/fucose-C1	yes	a,b	5.46	1.86	0.92
202.121	nd	201.113	Dicarboxylic acid	Sebacic acid	yes	a,b,c	6.24	10.21	6.70
203.113	204.121	nd	Carnitines and derivatives	ACETYL-L-CARNITINE	yes	a,b,c	5.14	7.77	1.08
204.088	205.096	203.081	Amino Acids; Indoles and Derivatives	TRYPTOPHAN	yes	a,b,c	4.16	10.73	3.98
205.072	206.080	nd	Indoles and Derivatives	Indole/lactic acid	yes	a,b,c		9.60	
205.072	206.080	nd	Indoles and Derivatives	Cinnamoylglycine	yes	a,b,c		9.50	
205.074	nd	204.066	Indoles and Derivatives	5-METHOXYINDOLEACETIC ACID	yes	a,b	2	9.37	5.80
207.089	nd	206.081	Amino acids	N-ACETYL-L-PHENYLALANINE	No	a,b,c	1.97		
208.083	209.091	207.077	Amino Acids	KYNURENE	yes	a,b,c	3.67	8.63	2.17
214.130	215.139	nd	Polypeptides	VAL-PRO	No	a,b	3.56		1.15
216.173	nd	215.165	Hydroxy Acids	HYDROXYDODECANOIC ACID	yes	a,b,c	1.81	13.63	9.56
217.129	218.138	nd	Carnitines and derivatives	PROPIONYL-CARNITINE	yes	a,b,c	4.04	10.72	1.58
217.129	218.138	nd	Carnitines and derivatives	PROPIONYL-CARNITINE isomer	yes	a,b,c			
218.056	nd	217.048	Alcohols and Polyols	Mannitol-C1	No	a,b,c			
219.109	220.117	nd	Amino Alcohols; Amino Acids	Pantothenic acid	yes	a,b	7.10	1.66	
231.145	232.153	nd	Carnitines and derivatives	BUTYRYL CARNITINE	yes	a,b,c	3.64	4.61	2.62
231.146	232.154	nd	Carnitines and derivatives	BUTYRYL CARNITINE isomer	yes	a,b,c	3.33	15.53	3.11
239.093	nd	238.085	Quaternary Amines	ACETYL-L-CARNITINE-C1	No	a,b,c	16.43	3.44	
257.067	nd	256.059	Carbohydrates	N-ACETYL-D-HEXOSAMINE-C1	yes	a,b		8.25	1.09
259.176	260.184	nd	Carnitines and derivatives	HEXANOYL-CARNITINE	yes	a,b,c	6.01		
259.176	260.184	nd	Carnitines and derivatives	HEXANOYL-CARNITINE isomer	No	a,b,c	2.23	9.64	5.73
261.154	262.163	nd	Carnitines and derivatives	Hydroxyvaleryl-carnitine	yes	a,b,c	2.67	16.43	
264.110	265.118	263.103	Amino Acids	PHENYLACETYL-L-GLUTAMINE	yes	a,b,c	4.97	9.11	1.61
269.080	269.089	nd	Nucleoside Analogues	INGOSINE	yes	a,b,c	68 and 2.1	7.38	4.76
280.046	245.076	19.038017 (Cl adduc	Nucleoside Analogues	Uridine-C1	yes	a,b,c	5 and 5.4	2.74	1.24
281.109	282.118	nd	Nucleoside Analogues	METHYLADENOSINE	yes	a,b	4.2	2.60	1.13
286.214	nd	285.206	Dicarboxylic acids	HEXADECANEDIOIC ACID	yes	a,b	4.01	.5 and 11.1	1.00
287.207	288.215	nd	Carnitines and derivatives	Octanoylcarnitine	No	a,b	1.96	14.03	9.94
289.150	290.159	nd	Carnitines and derivatives	Methylglutaryl-carnitine	yes	a,b,c	2.02		7.07
289.150	290.159	nd	Carnitines and derivatives	Methylglutaryl-carnitine isomer2	yes	a,b,c	7.09	10.94	
289.150	290.159	nd	Carnitines and derivatives	Methylglutaryl-carnitine isomer3	No	a,b,c		10.25	

M	Detection in the positive ionization mode	Detection in the negative ionization mode	Chemical class	Identified metabolite	Already Identified in human serum samples	Status of Identification	RT HILIC	RT PFPF	RT CB
294.119	295.127	nd	Polypeptides	Glu Phe	No	a, b, c	9.24	9.24	2.60
294.120	295.128	nd	Polypeptides	Glu Phe isomer	No	a, b, c	9.99	9.95	4.52
309.106	nd	308.098	Carbohydrates	N-Acetylneuraminic acid	Yes	a, b, c	7.32	13.50	8.70
327.095	328.103	326.087	Glucuronides	ACETAMINOPHENE GLUCURONIDE	Yes	a, b, c	5.6	3.86	1.59
342.116	nd	341.108	Carbohydrates	(1S)-MALTULOSE/Lactulose/ Sucrose-2	Yes	a, b, c	8.27		
342.116	nd	341.108	Carbohydrates	(1S)-MALTULOSE/Lactulose/ Sucrose-1	Yes	a, b, c	7.79		
368.165	nd	367.157	Sulfate esters : Steroids	DHEA SULFATE	Yes	a, b, c	1.72	11.04	7.37
392.294	nd	391.286	Bile Acids	URSODEOXYCHOLIC ACID	Yes	a, b	1.95	13.38	9.74
449.315	450.319	448.306	Bile Acids: Acyl Glycines	GLYCOURSODEOXYCHOLIC ACID	Yes	a, b, e	1.95	12.01	8.51
465.310	466.316	464.302	Bile Acids: Acyl Glycines	GLYCOCHOLIC ACID	Yes	a, b, c	1.98	10.86	7.62
499.295	500.303	498.288	Bile Acids	Taurochenodesoxycholic acid	Yes	a, b	9.63		7.72
515.293	nd	514.285	Bile Acids	Taurocholic Acid	Yes	a, b	8.77	7.04	7.04
529.271	nd	528.263	Acyl Glycines: Bile Acids : sulfate esters	Glycochenodeoxycholate-sulfate	Yes	a, b, c	1.96	9.40	7.47
594.333	595.341	nd	Tetrapyrroles and Derivatives	Stercobilin	No	a, b, c	1.93	8.15*	5.96**
85.052	86.060	nd	Amino Ketones	Pyrollidinone	Yes	a, d			1.38
95.036	96.044	nd	Pyridines and Derivatives	Hydroxypyridine	No	a, d	2.78	2.59*	1.3**
111.983	nd	110.975	Sulfate esters	MeSulfate	No	a, d	3.25		
116.048	nd	115.040	Keto-Acids	5-oxo-pentanoic acid (preferer MOD C8)	No	a, d	2.02		1.91
120.043	nd	119.035	Hydroxy Acids	Dihydroxybutyric acid	Yes	a, d	7.37		
122.037	nd	121.029	Aldehydes	Hydroxybenzaldehyde	Yes	a, d	2.61		5.01**
144.043	nd	143.034	Dicarboxylic Acids	Methylglutaconic	Yes	a, d	7.34		
145.073	146.081	nd	Keto acids and derivatives	2-Keto-6-aminocaproic acid	No	a, d	7.15*		
145.109	146.117	nd	Amino Fatty acids	Butyrobetaine	Yes	a, d	6.95	5.37	0.93
145.109	146.118	nd	Amino Fatty acids	Butyrobetaine isomer	No	a, d		4.95	
148.037	nd	147.029	Carbohydrates	Ribonolactone-isomer1	No	a, d	8.16	2.43	1.29
148.037	nd	147.029	Carbohydrates	Ribonolactone-isomer2	No	a, d	7.92	2.82	1.12
150.068	nd	149.060	Aromatic Acids	Hydrocinamic acid	Yes	a, d	2.02		
158.067	159.076	nd	Amino acids derivatives	methyleneglutamine	No	a, d	7.59	2.28	
158.094	nd	157.086	Carboxylic acids	Oxovalproic acid	Yes	a, d	1.84	10.4*	6.64
160.120	161.128	nd	Amino Acids	N(6)-Methyllysine	Yes	a, d	8.76	4.14*	0.80
165.046	nd	164.038	Amino Acids	Methionine sulfoxide	Yes	a, d	6.88		
166.048	nd	165.040	Carbohydrates	Ribonic acid-isomer1	Yes	a, d	7.03		0.91
166.048	nd	165.040	Carbohydrates	Ribonic acid-isomer2	No	a, d	7.32		1.63
170.131	nd	169.123	Fatty Acids	Decenolic acid	Yes	a, d	1.85		

M	Detection in the positive ionization mode	Detection in the negative ionization mode	Chemical class	Identified metabolite	Already identified in human serum samples	Status of identification	RT: HILIC	RE: PFP	RT: CB
172.146	nd	171.138	Fatty Acids	Decanoic acid	yes	a,d	1.83		
174.099	175.107	nd	Polypeptides	Gly-Val	No	a,d	6.85	7.64	1.38
178.099	nd	177.091	Fatty acids	Eugenolmethylether	No	a,d	1.87		
186.125	nd	185.117	Fatty acids	Oxocaproic acid / decenoic acid	No	a,d	1.86	12.07	8.06
189.079	nd	188.071	Indoles and Indole Derivatives	3-Indolepropionic acid	yes	a,d	2.68		
189.994	nd	188.986	Phenol and derivatives : Sulfate esters	Pyrocatechol sulfate / Quinol sulfate	No	a,d	2.33		
190.093	191.102	nd	Amino acids and derivatives	Diaminopimelic acid	yes	a,d	9.09		
206.043	nd	205.035	Tricarboxylic acids	Methylcitric acid- isomer2	yes	a,d	9.09		
206.043	nd	205.035	Tricarboxylic acids	Methylcitric acid- isomer3	No	a,d	9.33		
206.043	nd	205.035	Tricarboxylic acids	Methylcitric acid- isomer1	No	a,d	7.22	6.89	
211.085	212.093	nd	Amino acids and derivatives	3-Methoxytyrosine	yes	a,d		6.46	4.67
213.010	nd	212.002	Indoles and Indole Derivatives	Indoxyl sulfate	yes	a,d	2.82	6.46	4.67
216.121	217.129	nd	Amino Acids	Acetyl-L-arginine	yes	a,d	7.09	4.63	1.03
217.105	218.113	nd	Amino acids and derivatives	Acetyldrulline	yes	a,d			1.14
217.989	nd	216.980	Benzoic acid and derivatives : sulfate esters	Sulfosalicylic acid	yes	a,d	8.88	6.30	3.28
218.089	219.097	nd	Polypeptides	Glu-Ala	No	a,d	7.93	2.75	
220.084	221.092	nd	Amino Acids; Indoles and Indole Derivatives	Hydroxy-L-tryptophan	yes	a,d		6.03	4.66
222.088	223.096	nd		Dodecatetraenedioic acid	yes	a,d	11.67	11.67	7.78
222.098	223.107	nd		Gly Phe	No	a,d	12.13	3.99	
230.152	nd	229.144	Polypeptides	Dodecane dioic acid-Isomer1	yes	a,d	11.72	11.72	7.82
230.152	nd	229.144	Dicarboxylic Acids	Dodecane dioic acid-Isomer2	No	a,d			7.68
238.082	239.090	nd	Benzoic acids and derivatives	Trimethoxydinamic acid Isom	No	a,d		8.59	5.75
240.024	nd	239.016	Amino Acids	Cystine	yes	a,d	8.14		
240.100	nd	239.092	Hydroxy acids	2-Hydroxy-3-(4-methoxyethylphenoxy)-propanoic acid	No	a,d	2.80	11.68	7.76
245.161	246.169	nd	Carnitines and derivatives	isovalerylcarnitine Isomer	No	a,d	2.96		4.77
245.161	246.169	nd	Carnitines and derivatives	isovalerylcarnitine	yes	a,d			
247.103	248.111	nd	Carnitines and derivatives	Malonylcarnitine	yes	a,d	7.33	5.91	
253.050	254.058	nd	Aminobenzenesulfonamide s	Sulfamethoxazole	yes	a,d	2.5	10.34	6.08
253.989	nd	254.981	Alcohols and polyols : sulfate esters	Ascorbate-sulfate-Isomer1	yes	a,d	9.28		
253.989	nd	254.981	Alcohols and polyols : sulfate esters	Ascorbate-sulfate-Isomer2	No	a,d	10.97		
257.101	258.109	nd	Glycerophospholipid s	Glycerophosphocholine	yes	a,d	7.72	1.64	0.85
258.184	nd	257.176	Dicarboxylic acids	Tetradecanedioic acid-Isomer1	yes	a,d		12.82	8.69
258.184	nd	257.176	Dicarboxylic acids	Tetradecanedioic acid-Isomer2	No	a,d		12.94	8.90

Detection in the positive ionization mode	Detection in the negative ionization mode	Chemical class	Identified metabolite	Already identified in human serum samples	Status of identification	Rt HILIC	Rt PPPP	Rt CB
nd	271.226	Fatty acids	2-hydroxyhexadecanoic acid	No	a,d	1.72	16.58	12.39
276.119	274.104	Amino acids	Noropthalamic acid	yes	a,d	8.24	2.20	
282.279	nd	Fatty amides	Oleamide	yes	a,d		16.95	12.92
296.068	nd	Aminobenzenesulfonamides	Acetylsulfamethoxazole	yes	a,d	2.05	10.37	6.27
300.289	nd	Amino Alcohols	Sphingosine	yes	a,d		15.99	12.21*
303.231	301.216	Eicosanoids	ISOPIMARIC ACID	yes	a,d	1.71	15.26	10.91
344.277	nd	Carnitines and derivatives	Lauroylcarnitine	yes	a,d	1.95		9.22
361.203	nd	Steroids and Steroid Derivatives	Aldosterone/Cortisone	yes	a,d		10.53	
382.271	nd	Sphingolipids	Sphinganine 3-phosphate	yes	a,d		14.33	9.89
384.114	nd	Nucleoside Analogues	Succinyladenosine	yes	a,d	7.84	5.86	2.85
400.339	nd	Carnitines and derivatives	Palmitoylcarnitine	yes	a,d	1.86		11.06
424.340	nd	Carnitines and derivatives	Linoleyl carnitine	yes	a,d	1.85		10.67
nd	425.364	Dicarboxylic acids	Hexacosanedioic acid	No	a,d			13.89
428.373	nd	Carnitines and derivatives	Stearoylcarnitine	yes	a,d		1.85	11.81
463.230	nd	Steroids and Steroid Derivatives	Hydrocortisone succinate	yes	a,d		12.63	8.87
nd	512.269	Acyl Glycines; Bile Acids/NA	Sulfolithocholylglycine	yes	a,d		12.45	8.21
583.256	581.239	Porphyrins	Biliverdin	No	a,d	1.93		8.12
585.271	583.255	Porphyrins	Bilirubin	yes	a,d	1.94	19.51	7.72
626.355	nd	Bile Acids; Glucuronides	Glycochenodeoxycholic acid 3-glucuronide	yes	a,d		10.93	7.62

Table S.5 : Métabolites formellement et putativement identifiés dans le sérum humain

if peak shape
 eed on accurate measured mass
 eed on retention time similarity with that of an authentic standard
 eed on the MS² spectral similarity with that of an authentic standard
 eed on the MS² spectral interpretation

Métabolite théorique	Détection	Métabolite théorique	Détection
uridine		indole-3-acetate	
D-mannose-6-phosphate		glutathione	ok
orotate		D-sedoheptulose-7-phosphate	
porphobilinogen		CDP	
adenylo-succinate		D-ribulose-5-phosphate	
phosphoribosyl-formamido-carboxamide		shikimate-3-phosphate	
RX		oxygen	
phosphoenolpyruvate	ok	inosine-5'-phosphate	
3,4-dihydroxyphenylacetaldehyde		a 2(R)-hydroperoxy fatty acid	
H ₂ O		formaldehyde	
4-aminobutyrate	ok	5'-phosphoribosyl-N-formylglycineamide	
L-serine	ok	carbamoyl-phosphate	
S-adenosyl-L-methionine		hydroxymethylbilane	
D-sorbitol	ok	betanidin quinone	
xylitol		GDP-α-D-mannose	
D-myo-inositol (1,4,5)-trisphosphate		H ⁺	
nicotinamide mononucleotide		D-erythrose-4-phosphate	
uroporphyrinogen-III		glycogen	
S-adenosyl-L-homocysteine		acetyl-CoA	
shikimate		acetaldehyde	
uridine-5'-diphosphate		dimethylthiophosphate	
palmitoyl-CoA		nicotinate adenine dinucleotide	
CTP		4-fumaryl-acetoacetate	
L-cysteine		betanidin	
4-nitrophenol	ok	a 1,2-diacylglycerol	
a glutathione-toxin conjugate		formate	
pyridoxal 5'-phosphate		5-phosphoribosyl-N-formylglycineamidine	
oxaloacetate		2-phosphoglycerate	ok
D-4'-phosphopantothenate		quinolinate	
indole-3-acetonitrile		2-amino-3-carboxymuconate semialdehyde	
thiocyanate		4'-phosphopantetheine	
fructose-1,6-bisphosphate	ok	NADP ⁺	ok
α-D-glucose 1-phosphate	ok	pyridoxamine	
H ₂ O ₂		5'-phosphoribosyl-4-(N-succinocarboxamide)-5-aminoimidazole	
D-ribose-5-phosphate	ok	5-oxoproline	ok
palmitate		pantothenate	ok
a limit dextrin		acetoacetate	
2-methylacetoacetyl-CoA		CMP	
glutathione disulfide		L-homocysteine	
S-lactoyl-glutathione		α-D-glucose 6-phosphate	ok
ribose-1-arsenate		myo-inositol	
methyl parathion		L-ornithine	ok
arsenate		L-arabinose	ok
3-phospho-serine		dihydroorotate	
nicotinamide riboside		D-glyceraldehyde-3-phosphate	
ethanol		4-carboxy aminoimidazole ribonucleotide	
dUTP		dihydrouracil	
a debranched limit dextrin		D-lactate	ok
pyruvate	ok	O ₂ -	
homovanillate	ok	5-phosphoribosyl 1-pyrophosphate	
R-4'-phosphopantothenoyl-L-cysteine		aminoimidazole carboxamide ribonucleotide	
UDP-D-glucose		coenzyme A	
S-methyl-5'-thioadenosine		3-ureidopropionate	ok
spermidine	ok	dCTP	

Métabolite théorique	Détection	Métabolite théorique	Détection
dTTP		spermine	ok
NAD+	ok	UTP	
β-D-glucose-6-phosphate	ok	10-formyl-tetrahydrofolate	
GDP-D-rhamnose		dihydroxyacetone phosphate	ok
dTDP		fumarate	ok
a limit dextrin with short branches		S-adenosyl-L-methioninamine	
N-carbamoyl-L-aspartate	ok	D-xylulose-5-phosphate	
putrescine		L-arginine	ok
L-aspartate	ok	guanosine	
deoxyribose-5-phosphate		hypusine	
5,10-methylene-THF		methylglyoxal	
3-sulfinyl-pyruvate		L-arabitol	ok
cytidine		thiosulfate	
3-phosphoglycerate	ok	a 2,3,4-saturated fatty acyl CoA	
L-fucose	ok	a phosphatidylcholine	
L-glutamate	ok	a fatty aldehyde	
3-mercaptopyruvate		dephospho-CoA	
tetrahydrofolate		uridine-5'-phosphate	
stearoyl-CoA		N-acetyl-4-aminobutyrate	
L-methionine	ok	3-methoxytyramine	
nicotinamide	ok	propionyl-CoA	
D-myo-inositol (3)-monophosphate		S-formylglutathione	
a 1-acyl-2-lyso-glycerophosphocholine		stearate	
D-glucosamine-6-phosphate		dTMP	
L-arginino-succinate		a glucosylated glycogenin	
D-fructose		3,4-dihydroxyphenylacetate	
β-L-fucose 1-phosphate		pyridoxine-5'-phosphate	
tetrahydrobiopterin		N-acetylneuraminic acid	
GDP-L-fucose		pyridoxal	
L-glutamine	ok	L-γ-glutamylcysteine	
dCDP		5-aminoimidazole ribonucleotide	
L-xylulose		1,3-diphosphateglycerate	
dopamine		6-phospho-D-gluconate	
linoleate		dihydrobiopterin	
β-D-glucose	ok	bicarbonate	
betaxanthin		elf5A deoxyhypusine	ok
N-acetyl-4-aminobutanal		3-phospho-hydroxypyruvate	
an L-phosphatidate		acetoacetyl-CoA	
5,10-methenyltetrahydrofolate		N-acetyl-D-mannosamine	
pyridoxamine 5'-phosphate		acetate	
glycine	ok	(5-L-glutamyl)-L-amino acid	
citrate	ok	biliverdin-IX-α	
α-D-mannose 1-phosphate		a fatty acyl CoA	
nicotinate mononucleotide		GDP-4-dehydro-6-deoxy-D-mannose	
D-fructose-6-phosphate		1-phosphatidyl-1D-myo-inositol 4,5-bisphosphate	
linoleoyl-CoA		xanthosine-5-phosphate	
sulfite		3-sulfinoalanine	
adenosine-5'-phosphate	ok	S-hydroxymethylglutathione	
bilirubin		D-glucono-δ-lactone-6-phosphate	
paraoxon		malate	ok
a fatty acid		L-lactate	ok
orotidine-5'-phosphate		coproporphyrinogen III	
dUDP		diethylphosphate	
guanine		guanosine-5'-phosphate	

TableS.6: Métabolome érythrocytaire théorique et recouvrement avec les empreintes détectées

ANNEXE 3

ANALYSE LC-NH₂ DES EXTRAITS LIPIDIQUES

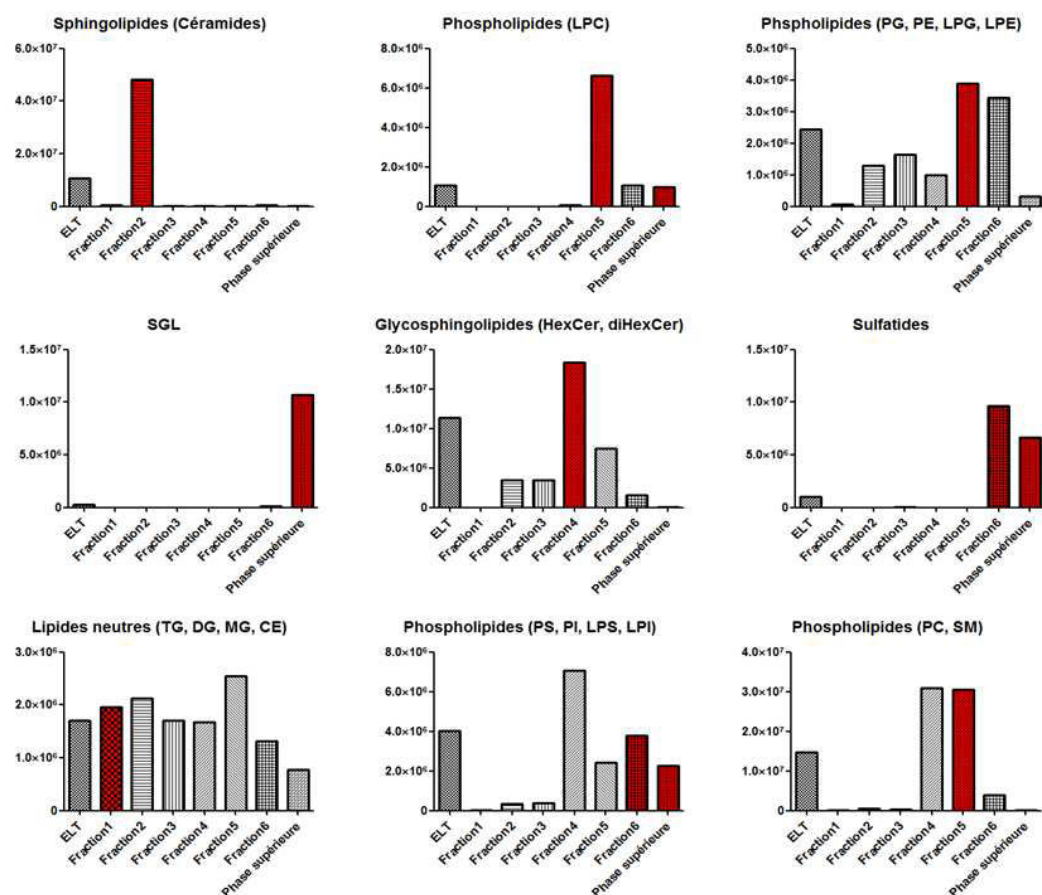


Figure.S.1 : Analyses RPLC-HRMS des fractions LC-NH₂ des extraits lipidiques de cerveau de rat

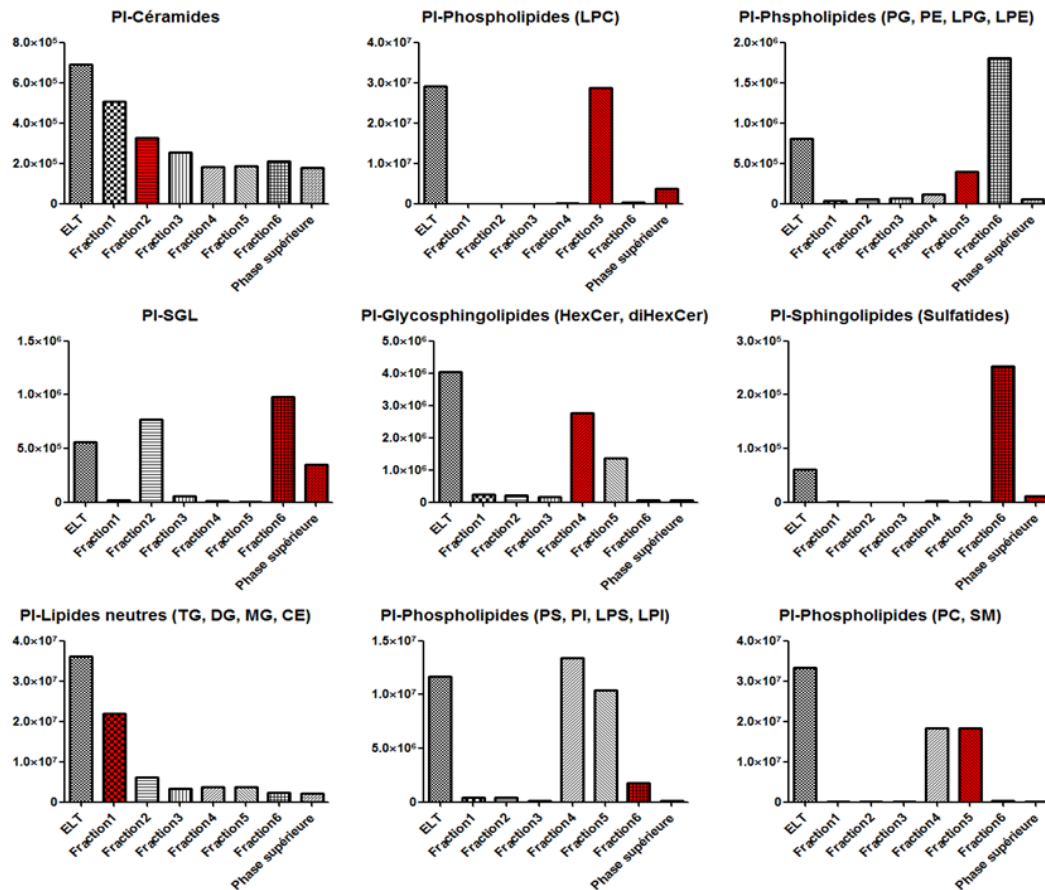


Figure.S.2 : Analyses RPLC-HRMS des fractions LC-NH₂ des extraits lipidiques de plasma humain

QUANTIFICATION DES LIPIDES : ETUDE PRELIMINAIRE

La finalité de chaque analyse métabolomique (dont l'analyse lipidomique) est d'aboutir à une liste de métabolites (en l'occurrence lipides) identifiés et quantifiés de façon absolue dans l'ensemble des échantillons témoins et malades. La quantification absolue peut être envisagée dans le cas où un étalon interne pour chaque espèce à quantifier est surchargé dans le mélange CM 1 :1 utilisé pour l'extraction. Appliquée au jeu de données plasmatiques précédent, plus de 700 composés de référence devraient être disponibles dans une démarche ciblée à large spectre, ce qui n'est pas envisageable. Cependant dans le cas des lipides, l'ionisation est assurée par les têtes polaires lesquelles sont communes aux différentes espèces au sein d'une classe de lipides pouvant permettre l'utilisation d'un étalon interne par classe de lipides. Cette propriété est utilisée en quantification absolue en SL (Ejsing et al; 2009, Ekroos et al; 2002) ou encore en NPLC-MS (Cifkova et al; 2012) ou en RPLC-MS dans des conditions isocratiques (Hermansson et al; 2005) car les espèces appartenant à la même classe de lipides sont ionisées sous les mêmes conditions. Par contre, ceci n'est pas vérifié en

RPLC-MS où les espèces d'une même classe sont éluées dans des proportions de phases mobiles différentes rendant l'efficacité d'ionisation entre espèces d'une même classe différente.

Nous avons de ce fait réalisé des analyses préliminaires en RPLC-MS visant à vérifier la différence d'ionisation de différents composés de référence appartenant à une même classe. Le mélange d'étalons internes testé contient un minimum de 2 composés de référence pour chacune des 14 classes lipidiques considérées. Ce dernier est surchargé à des concentrations différentes (allant de 100ng.mL^{-1} à $10\mu\text{g.mL}^{-1}$) dans des ELT de cerveau de rat, de plasma humain et dans la solution de reprise (65 :35 :5 v/v/v IPA/MeOH/H₂O).

Dans un premier temps, nous avons vérifié que les aires de chaque étalon interne à chaque concentration étaient équivalentes dans les deux matrices biologiques que dans la gamme tampon.

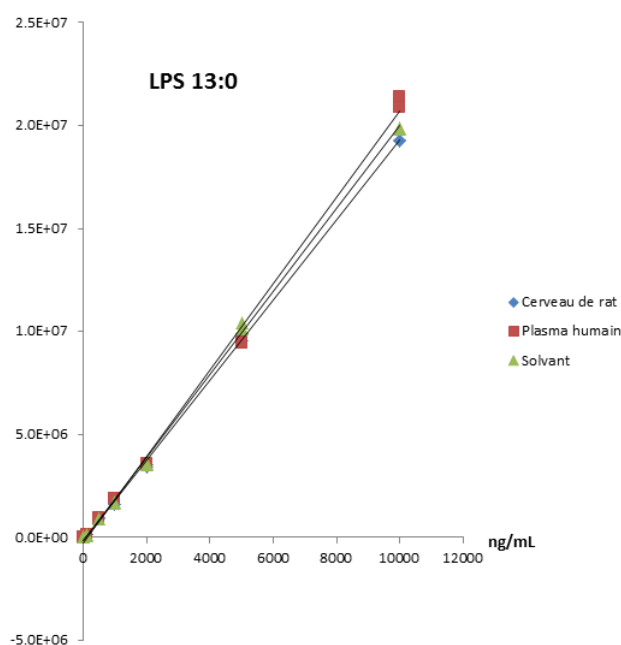


Figure 118. Aires chromatographiques du composé de référence LPS 13 :0 surchargé à 5 concentrations différentes dans un ELT de cerveau de rat, un ELT de plasma et du solvant

Les trois courbes d'étalonnage sont linéaires $r^2=[0.997-0.999]$ et superposables. Ceci prouve que la méthode est linéaire (de 100ng.mL^{-1} à $10\mu\text{g.mL}^{-1}$) et que l'effet matrice est négligeable dans ces 2 matrices biologiques analysées selon la méthode RPLC-MS développée.

Dans un second, nous avons évalué les facteurs de réponses de différents étalons internes par classe. Le but serait de pouvoir modéliser la réponse des différentes espèces au sein d'une classe selon les conditions d'éluion ou en d'autres termes, selon la longueur de la chaîne carbonée.

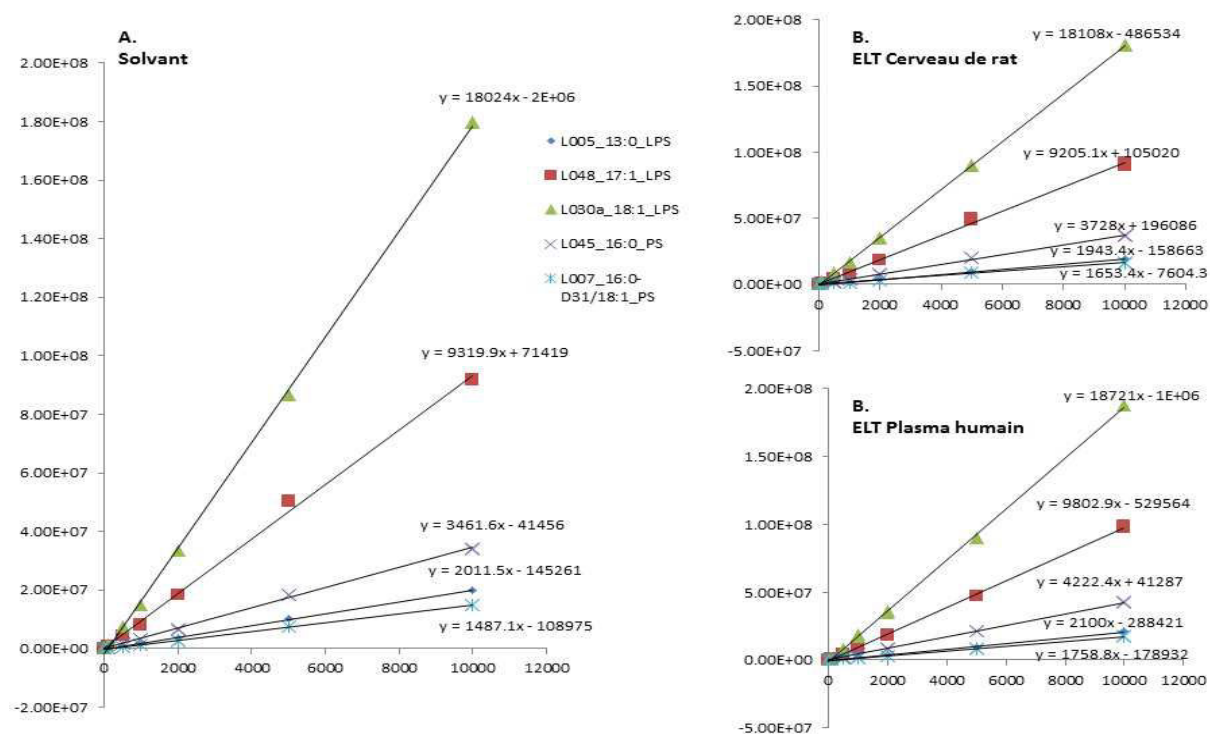


Figure 119. Gammes d'étalonnage de 5 espèces PS et LPS surchargées dans (A) une solution de reprise, (B) de l'ELT de cerveau de rat et (C) de l'ELT de plasma humain

On remarque que les composés LPS présentant le plus grand nombre d'atomes de carbone sont ceux qui répondent le mieux en ionisation électrospray. Cette tendance semble être inversée pour les espèces PS. Aussi, pour une longueur de chaîne équivalente sur une position *sn-1* ou *sn-2* donnée, les espèces lysophosphatidylsérines s'ionisent mieux que les formes phosphatidylsérines. Ceci pourrait être en rapport avec la baisse en proportion d'eau et l'augmentation de l'isopropanol favorisant l'ionisation des espèces.

Une investigation plus approfondie avec la modélisation des facteurs de réponse selon la nature de la chaîne carbonée pourrait être envisagée. Celle-ci requiert une liste plus exhaustive d'étalons internes par classe afin de pouvoir rendre compte des différentes longueurs de chaînes, du nombre d'insaturations et de la régiosélectivité qui sont autant de paramètres pouvant influencer les facteurs de réponse.

Cependant, alors que les analyses SL ne souffrent pas de ce manque de représentativité des étalons interne mais sont sujettes à un effet matrice non négligeable, l'analyse DIMS de fractions (LC-NH₂) surchargées d'un étalon interne par classe pourrait permettre d'assurer une quantification absolue à large spectre sans risque de suppression d'ionisation. Ceci est en cours de mise en œuvre au laboratoire.

ANNEXE 4

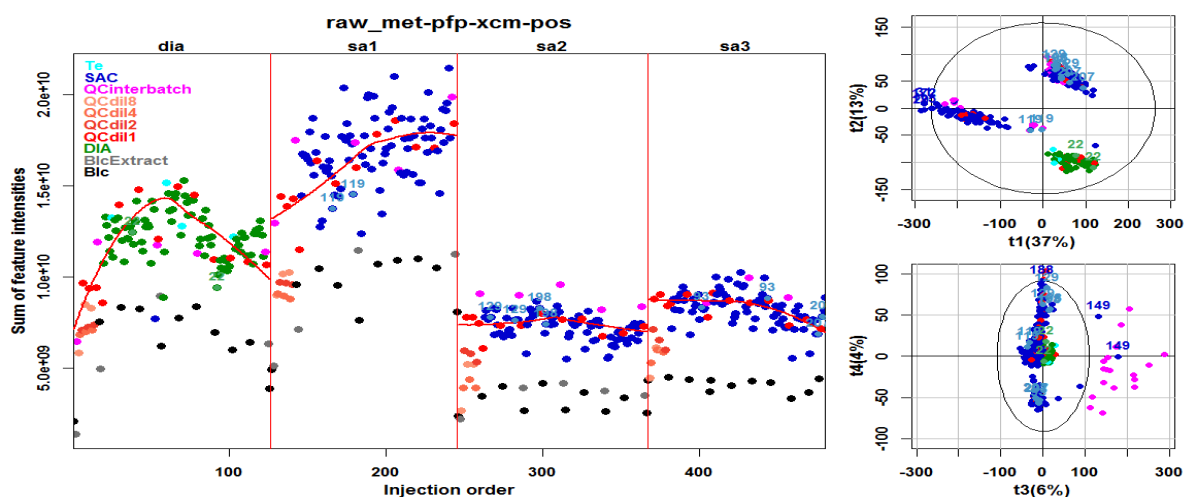
FUSION DES DONNEES METABOLOMIQUES ET LIPIDOMIQUES SERIQUES ET PLASMATIQUES DES SUJETS « TEMOINS » ET PATIENTS DIABETIQUES

1. Analyse Métabolomique

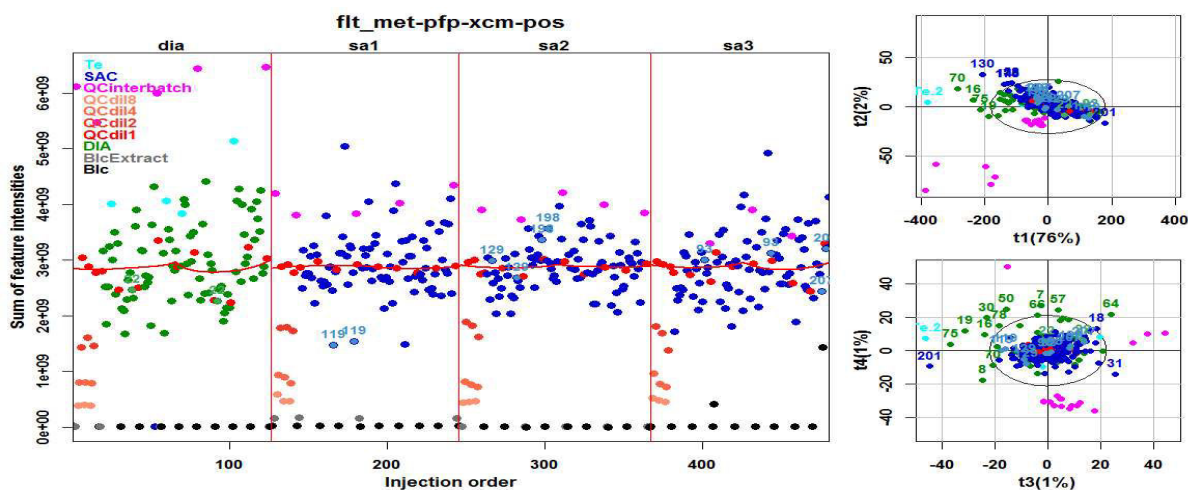
1.1. Matrices de données XCMS en analyse métabolomique

1.1.1. met-pfp-xcm-pos

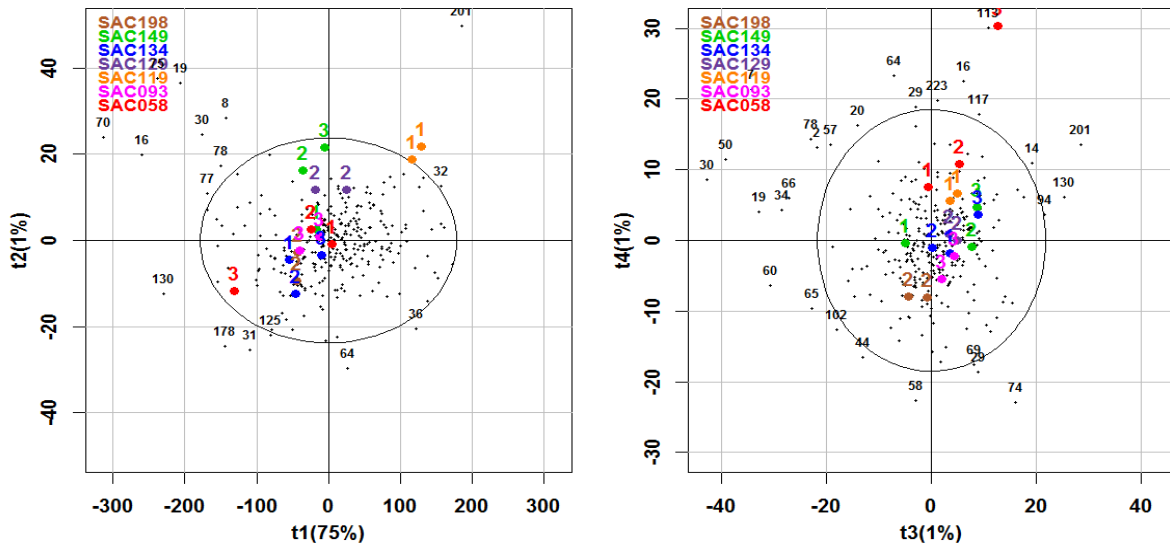
1.1.1.1. raw



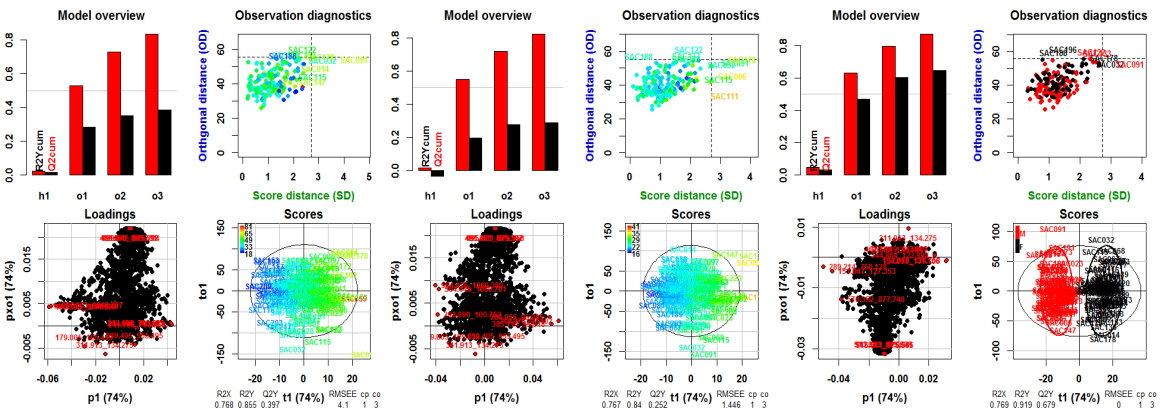
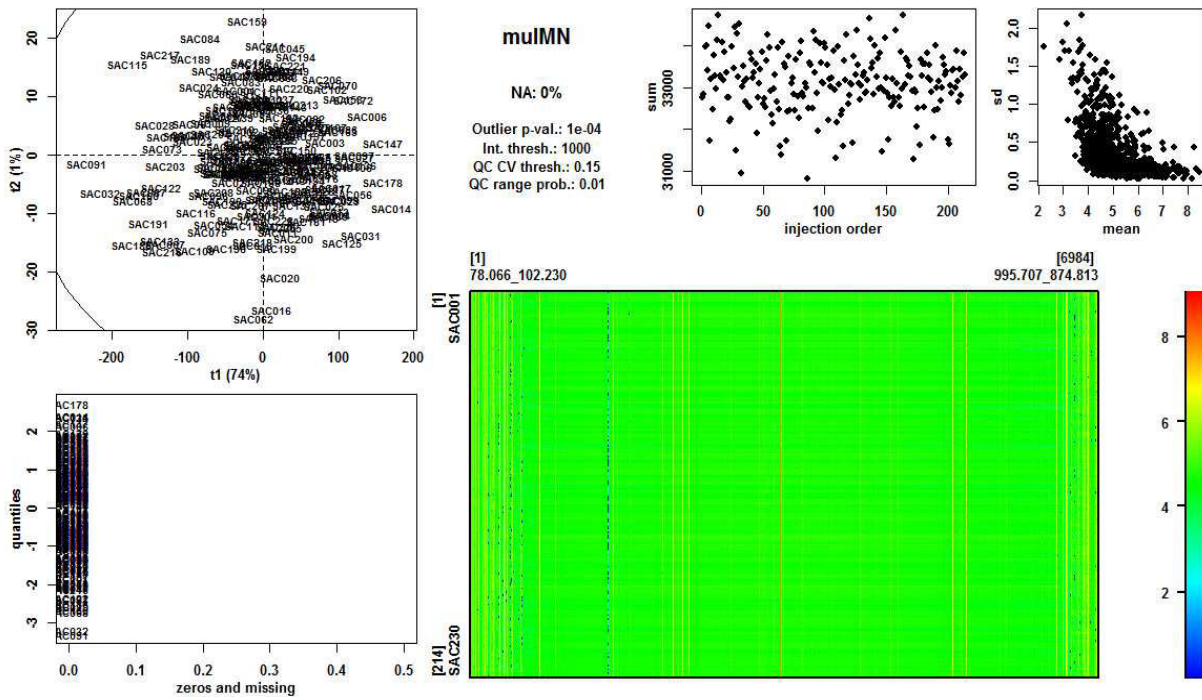
1.1.1.2. flt

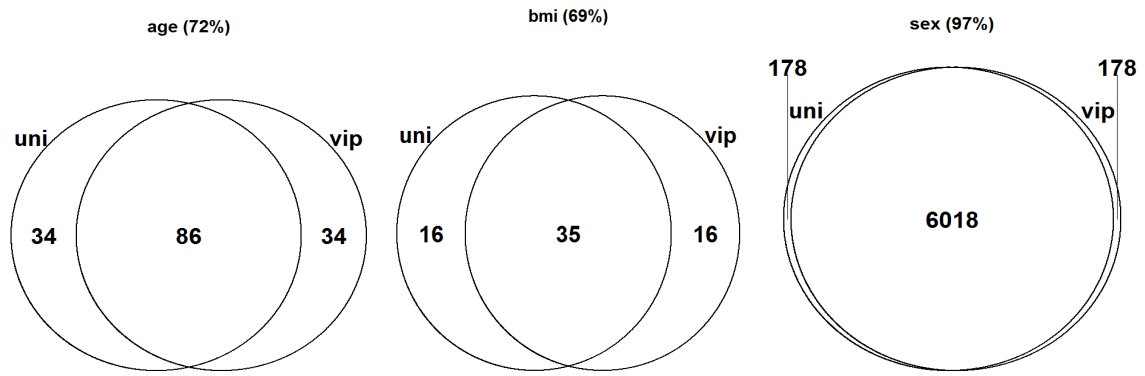


uni_met-pfp-xcm-pos



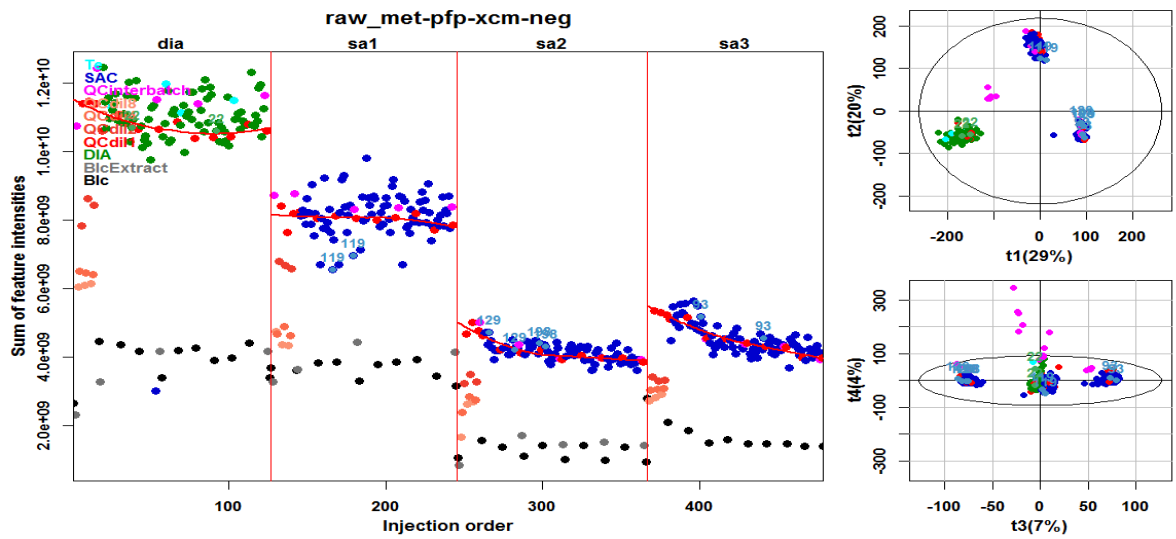
1.1.1.3. mul



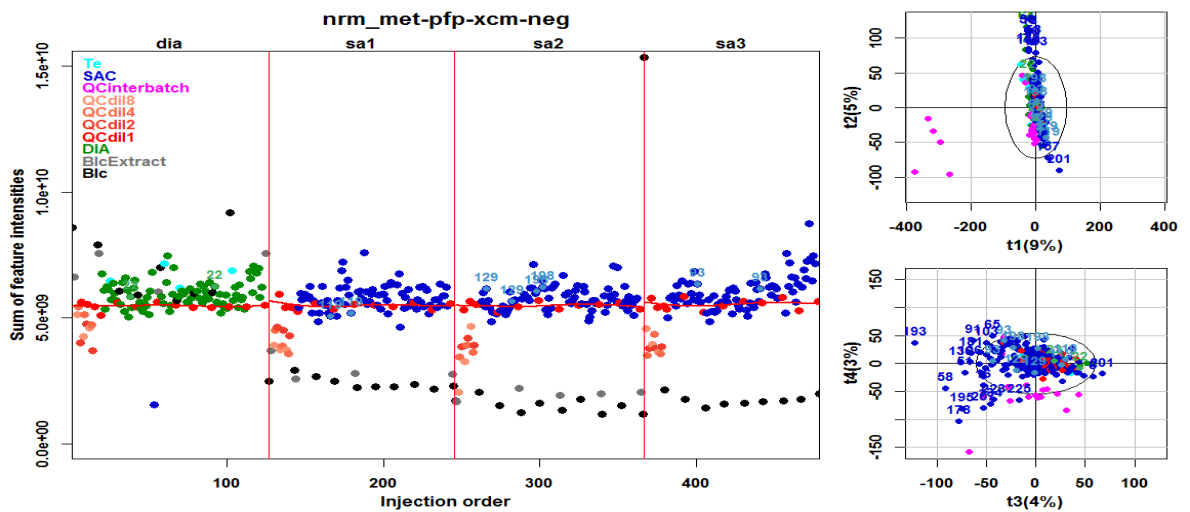


1.1.2. met-pfp-xcm-neg

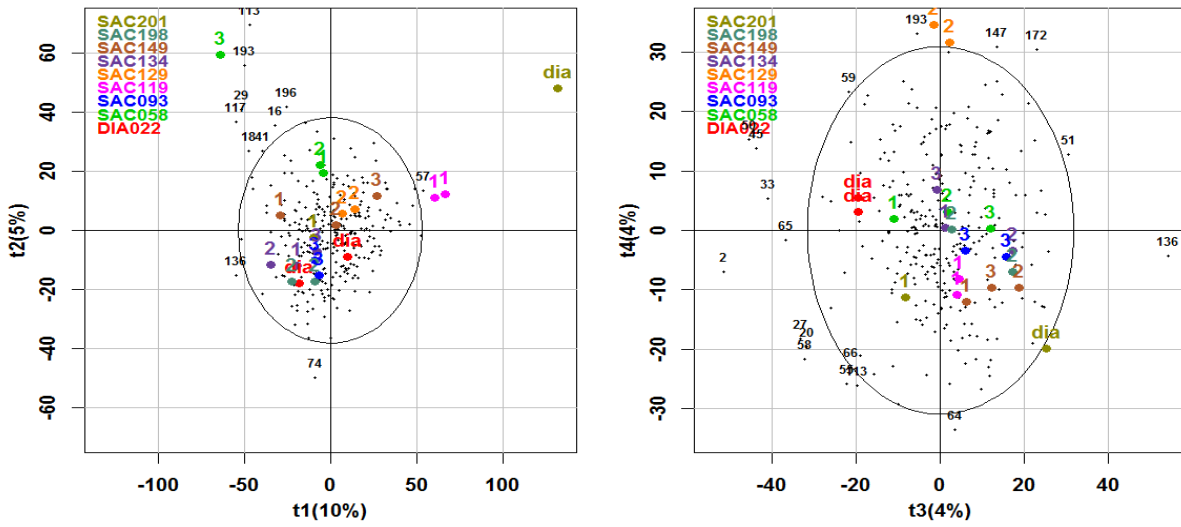
1.1.2.1. raw



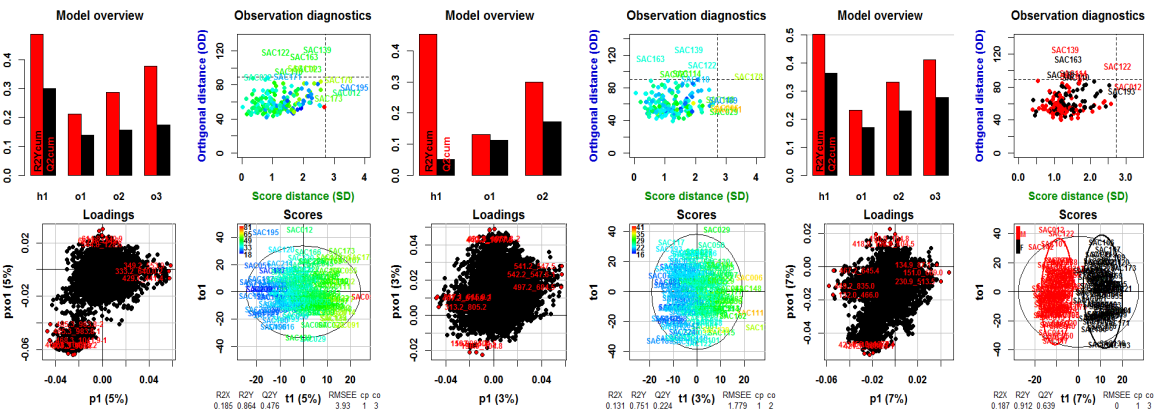
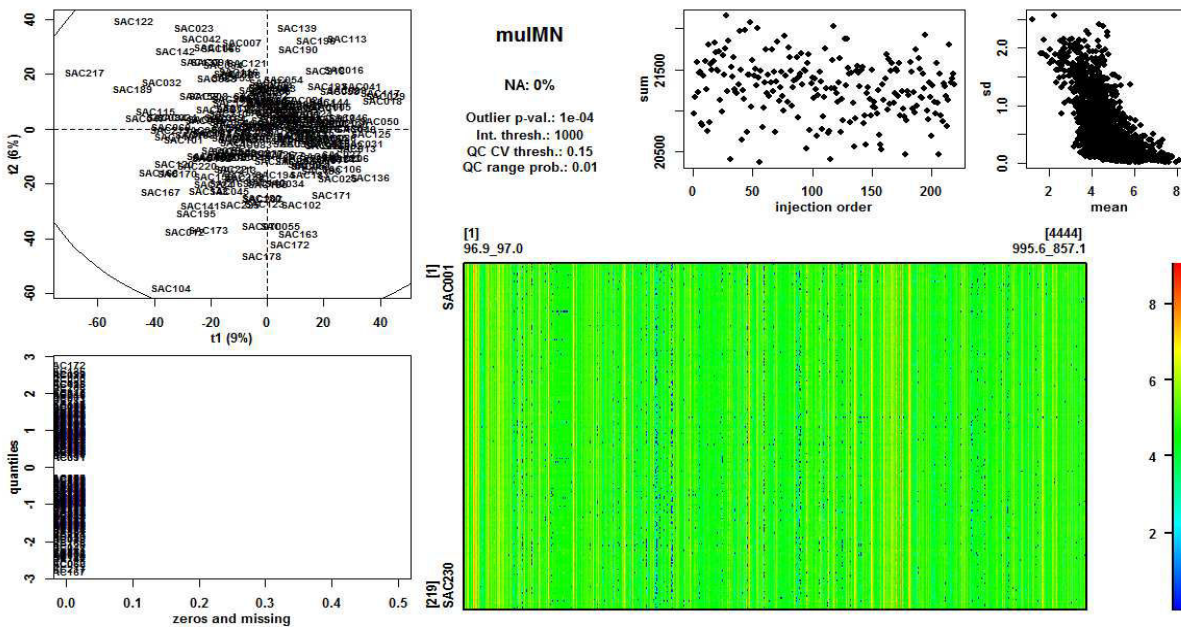
1.1.2.2. nrm

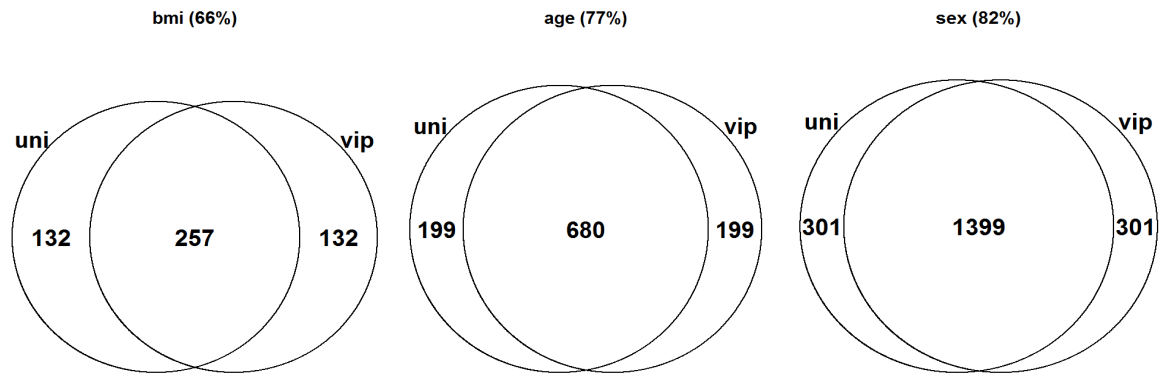


uni_met-pfp-xcm-neg



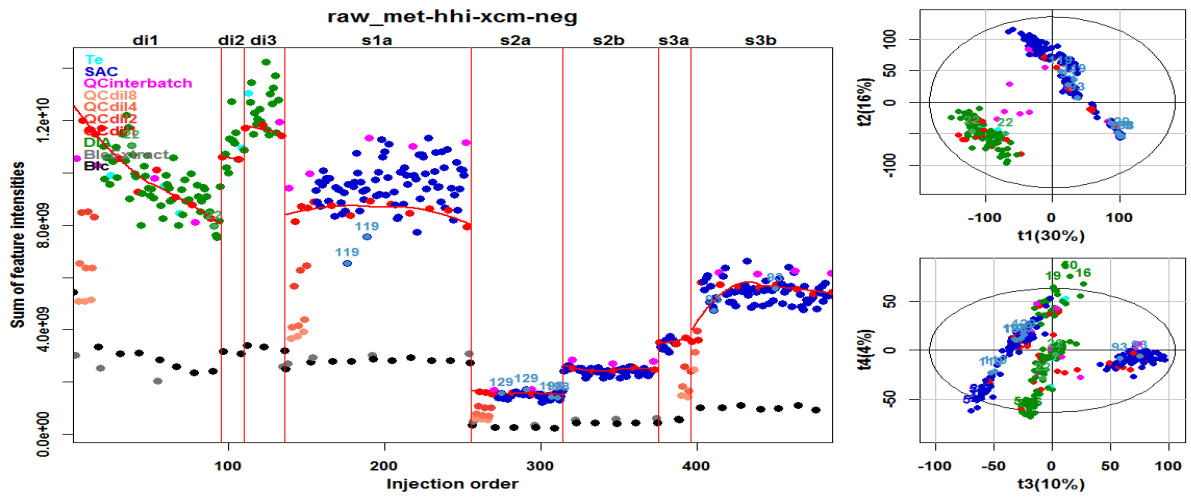
1.1.2.3. mul



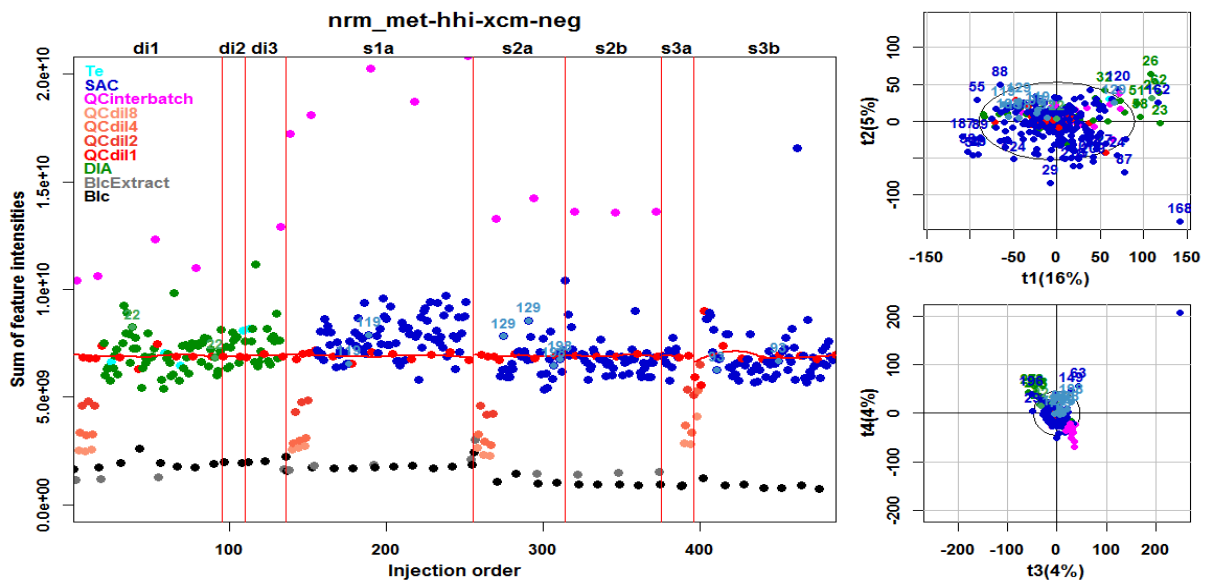


1.1.3. met-hhi-xcm-neg

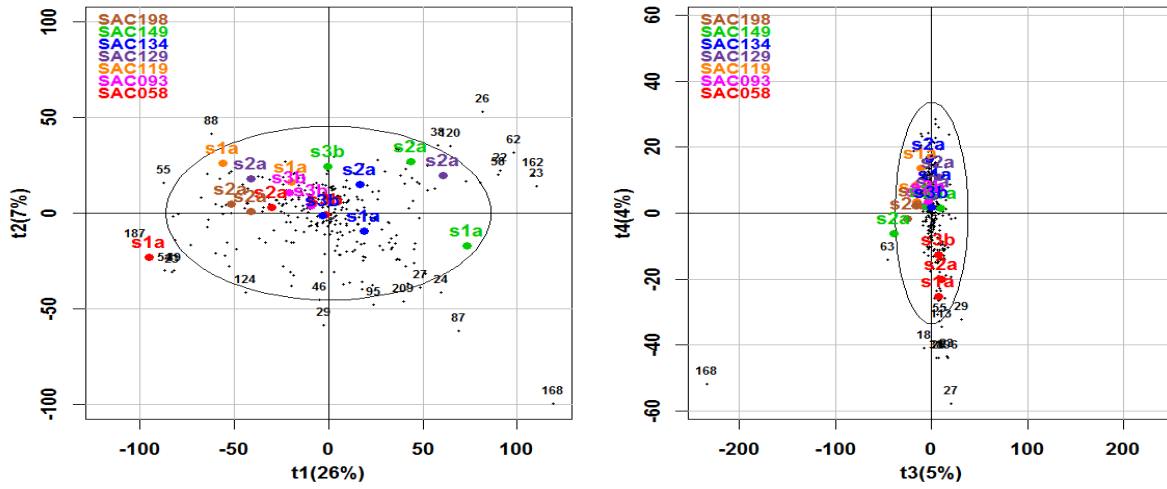
1.1.3.1. raw



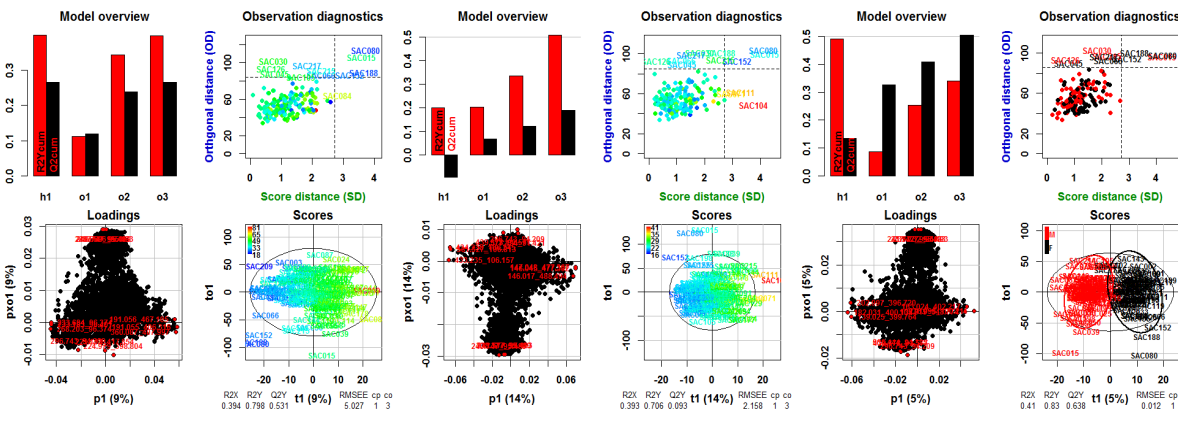
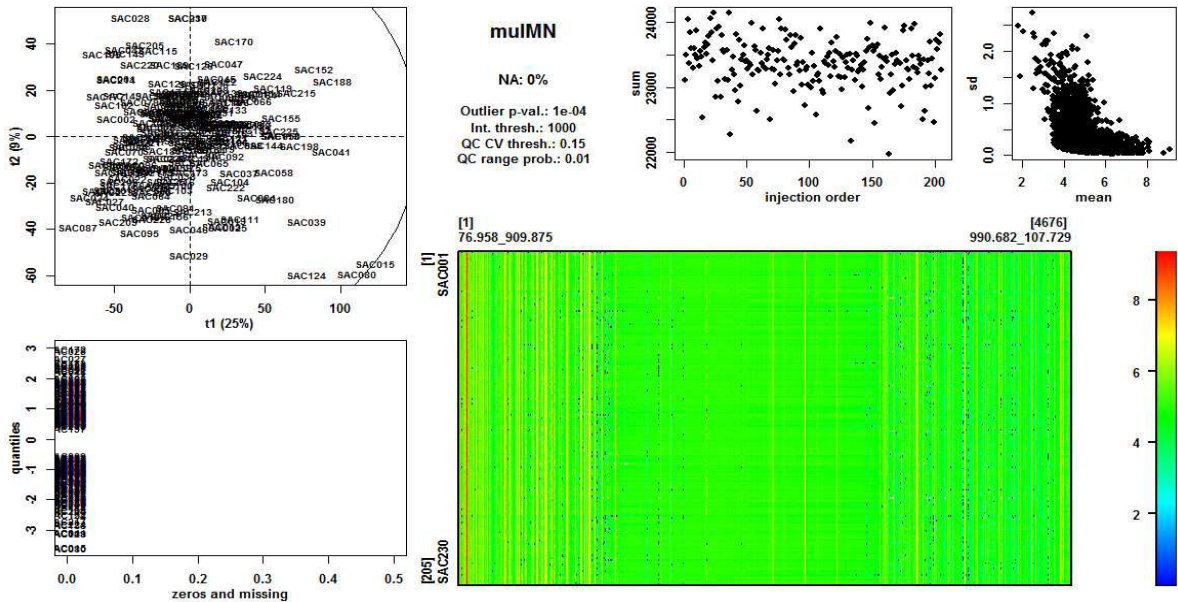
1.1.3.2. nrm

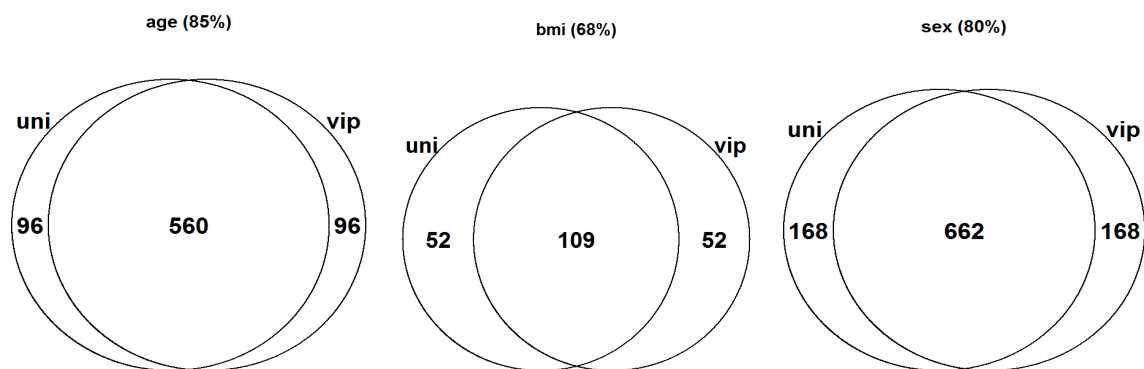


uni_met-hhi-xcm-neg



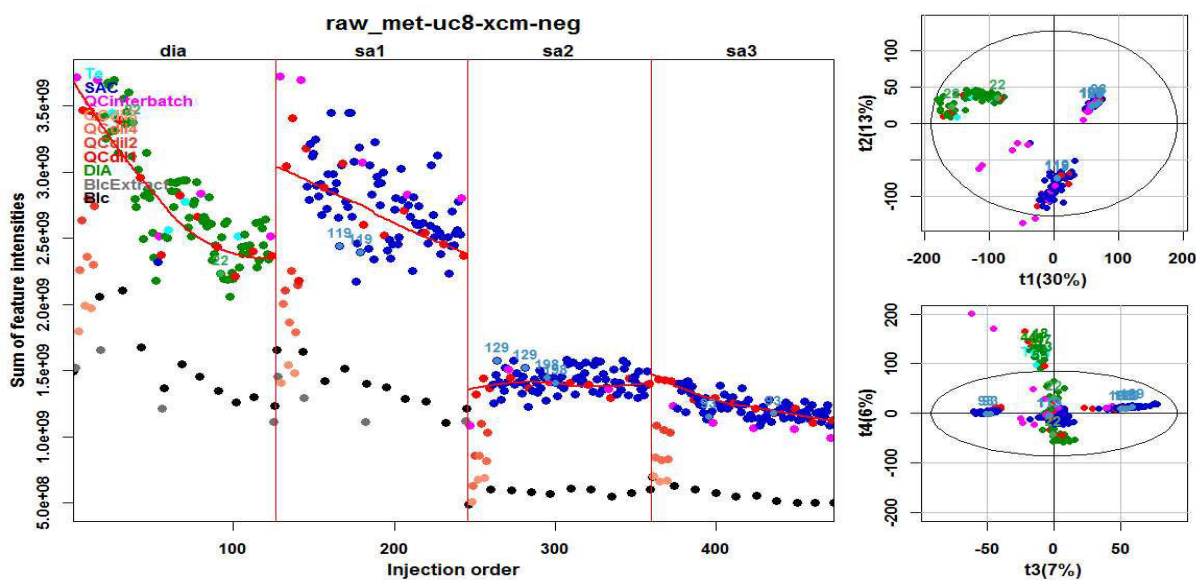
1.1.3.3. mul



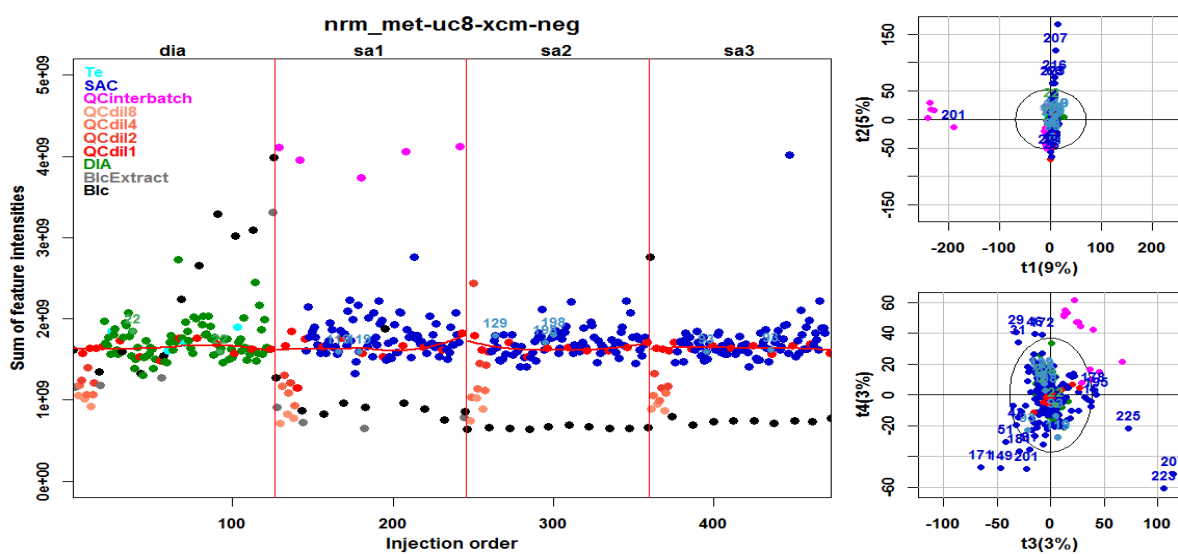


1.1.4. met-uc8-xcm-neg

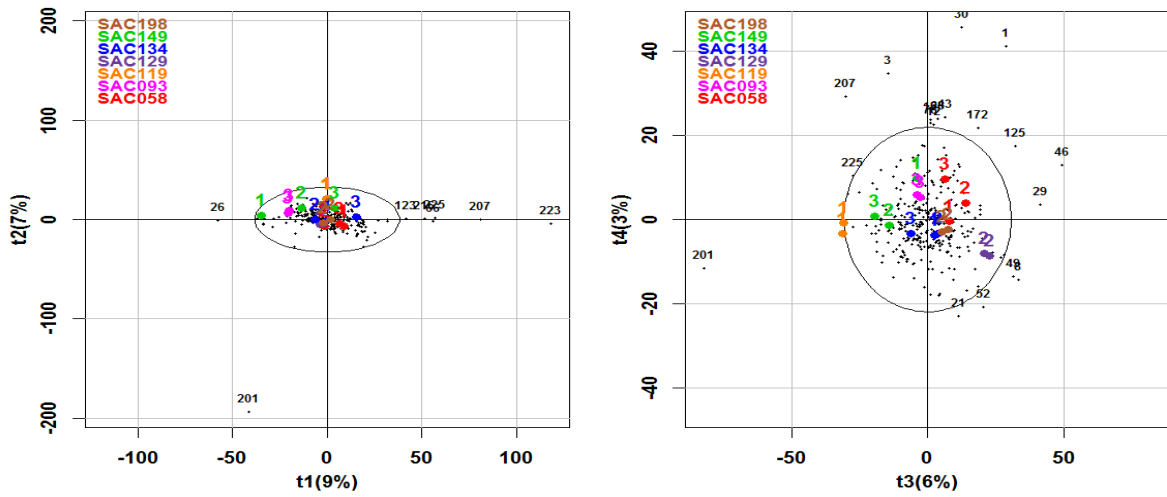
1.1.4.1. raw



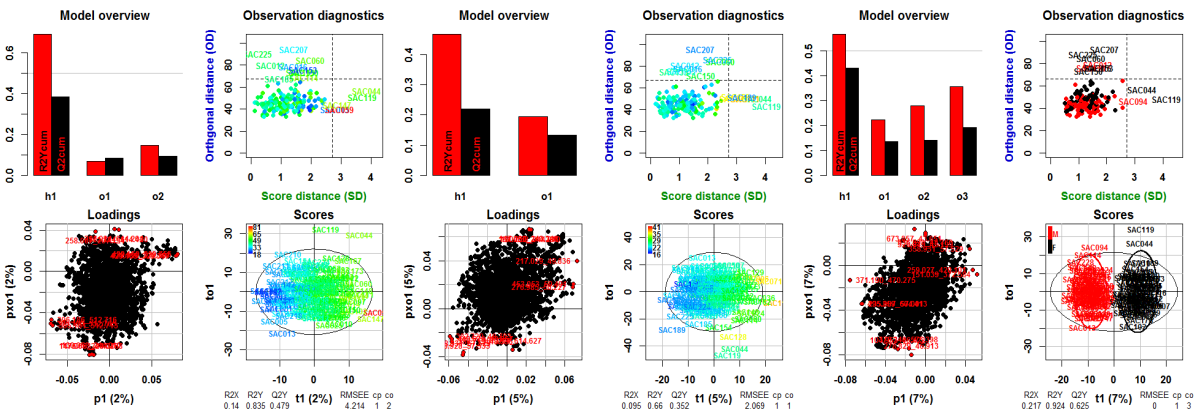
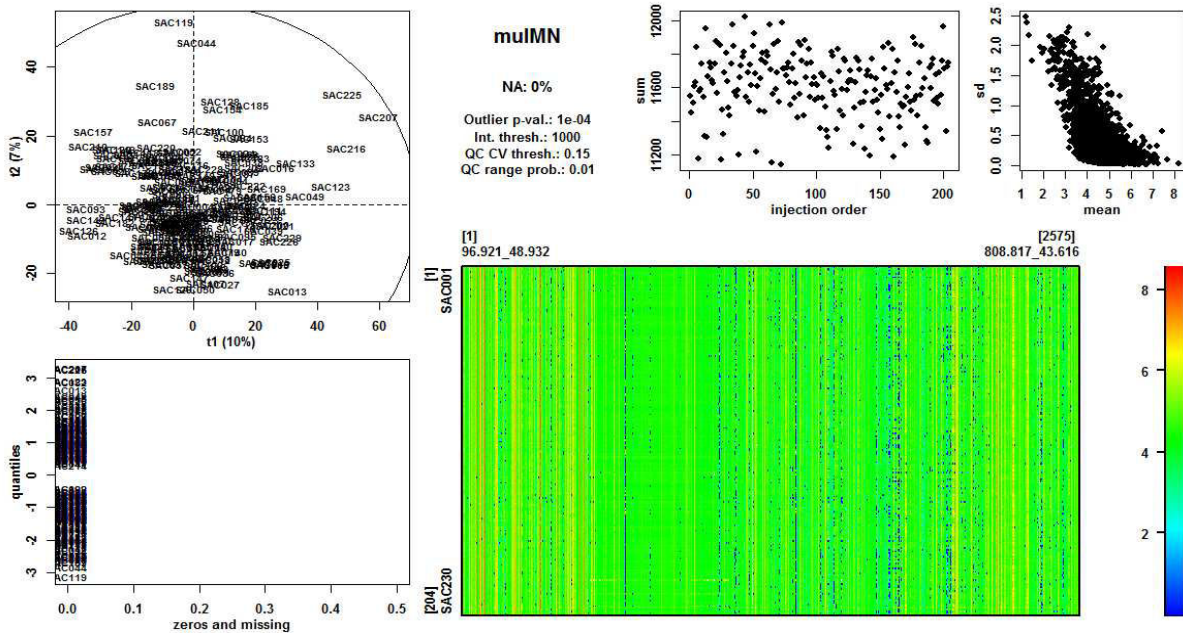
1.1.4.2. nrm

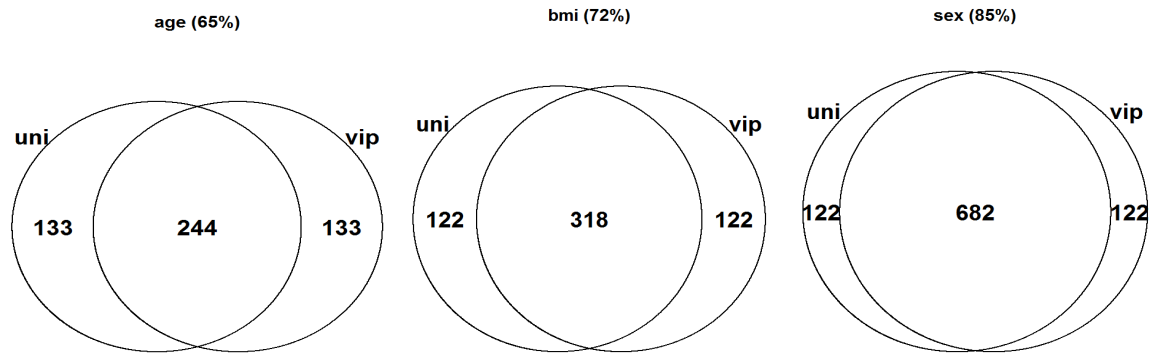


uni_met-uc8-xcm-neg



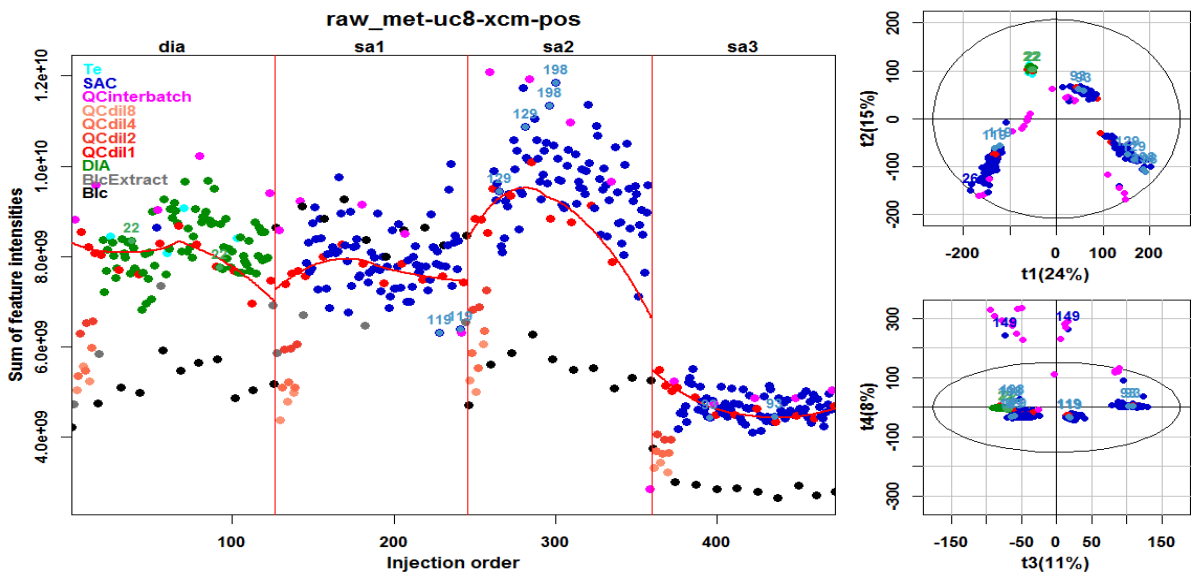
1.1.4.3. mul



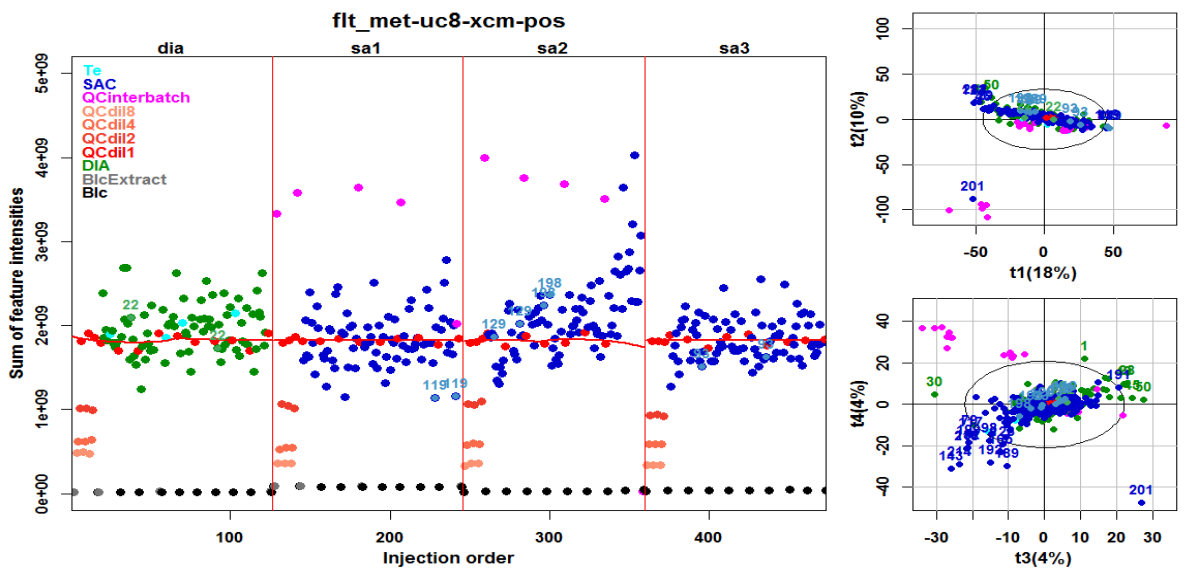


1.1.5. met-uc8-xcm-pos

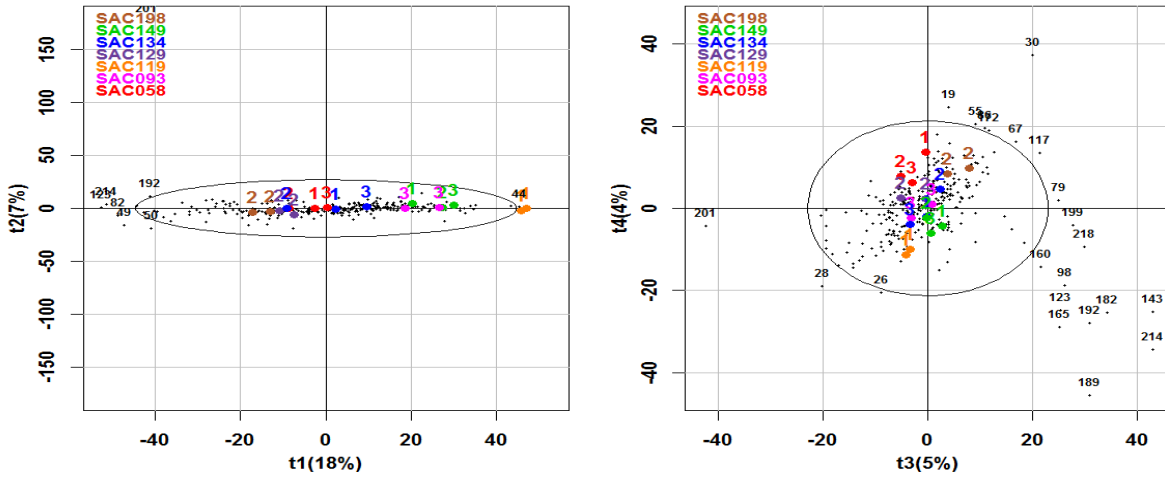
1.1.5.1. raw



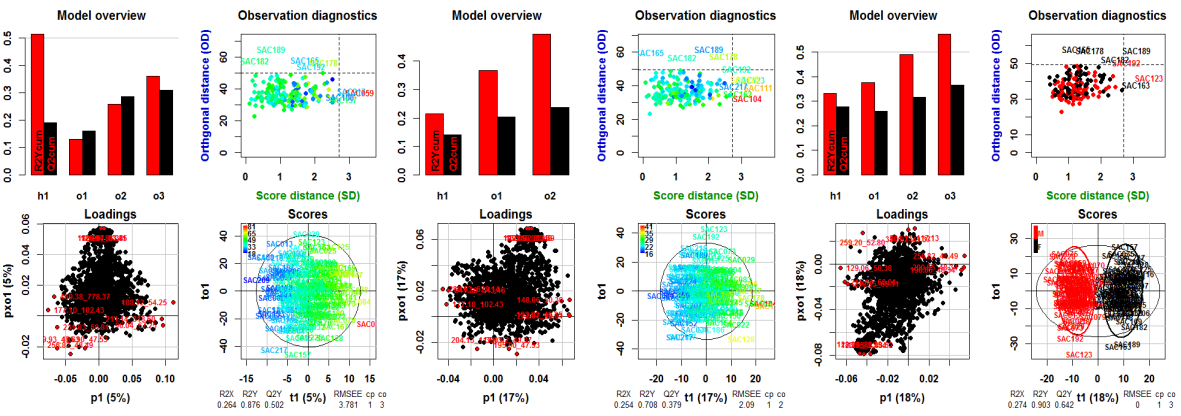
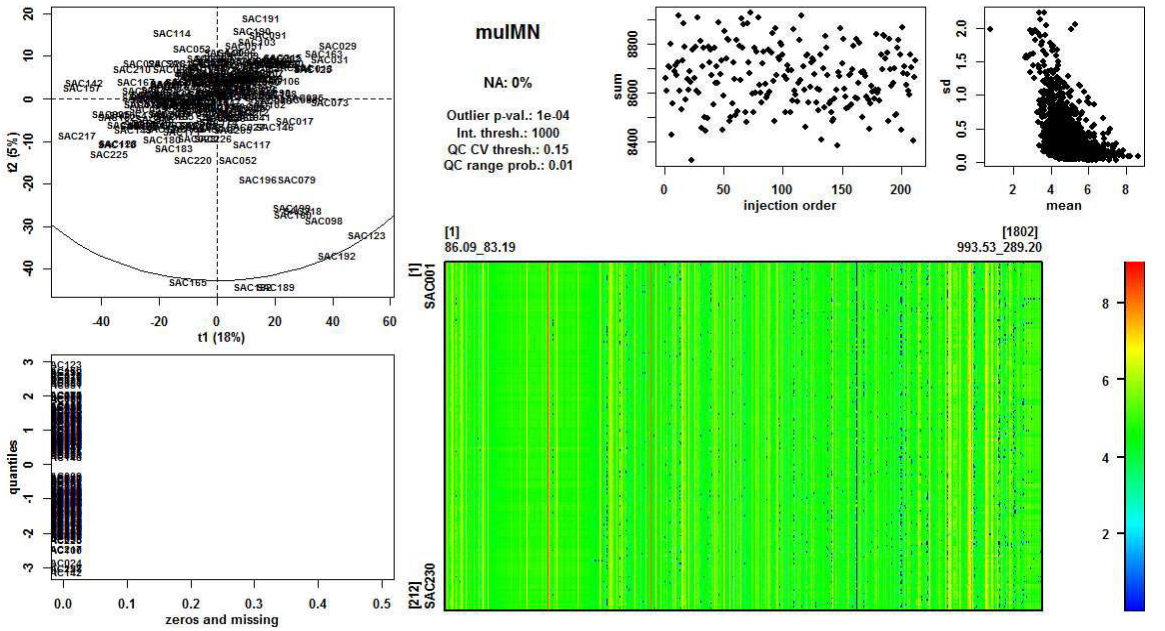
1.1.5.2. flt-Warn

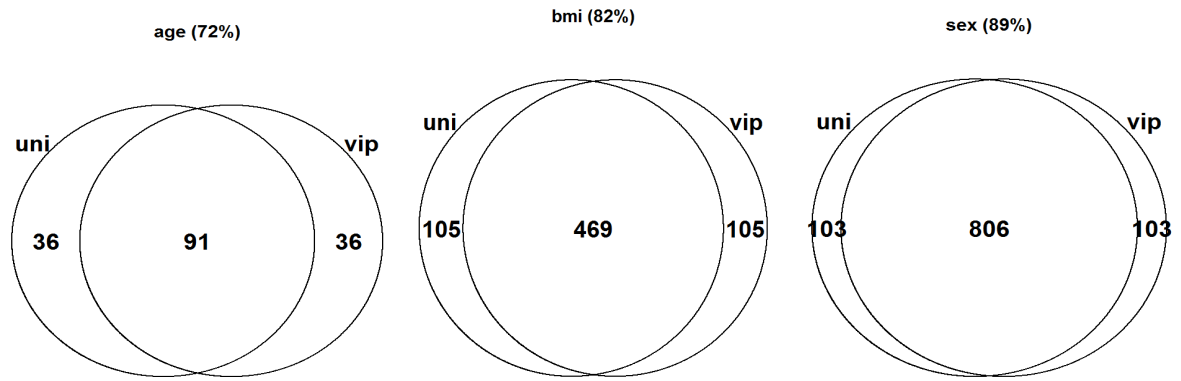


uni_met-uc8-xcm-pos



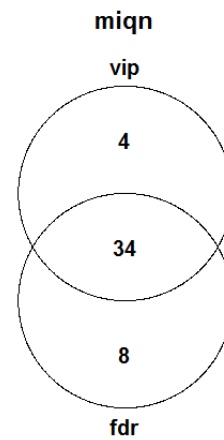
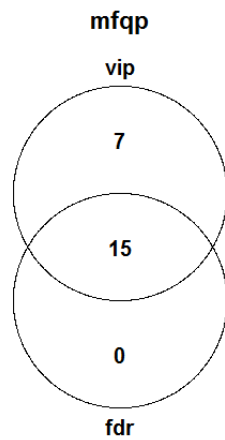
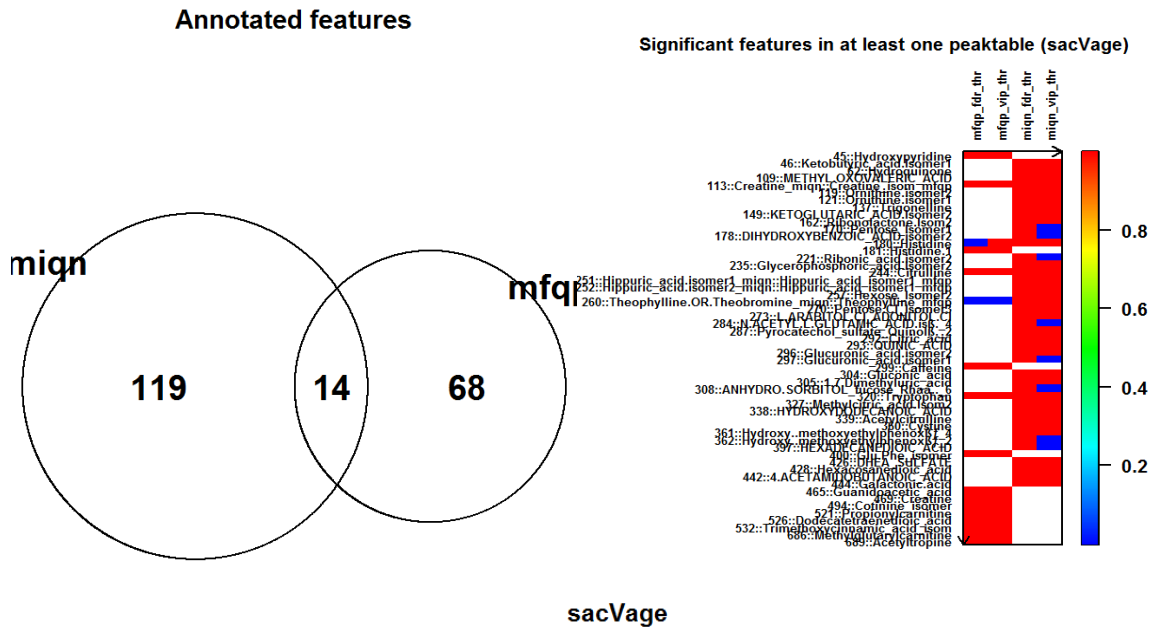
1.1.5.3. mul

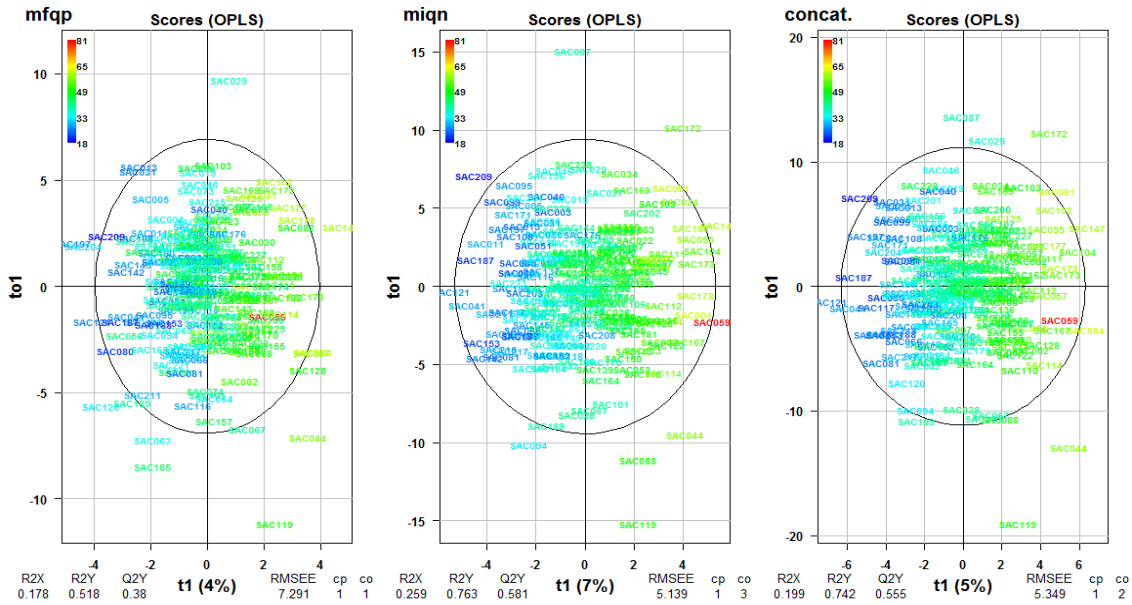




1.2. Matrices de données QuanBrowser en analyse métabolomique

1.2.1. sacVage

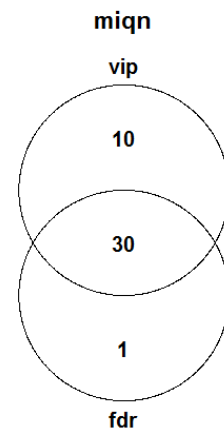
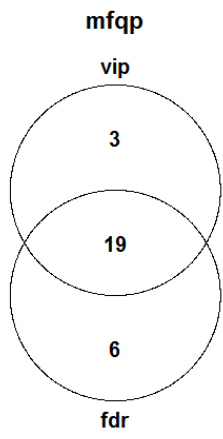
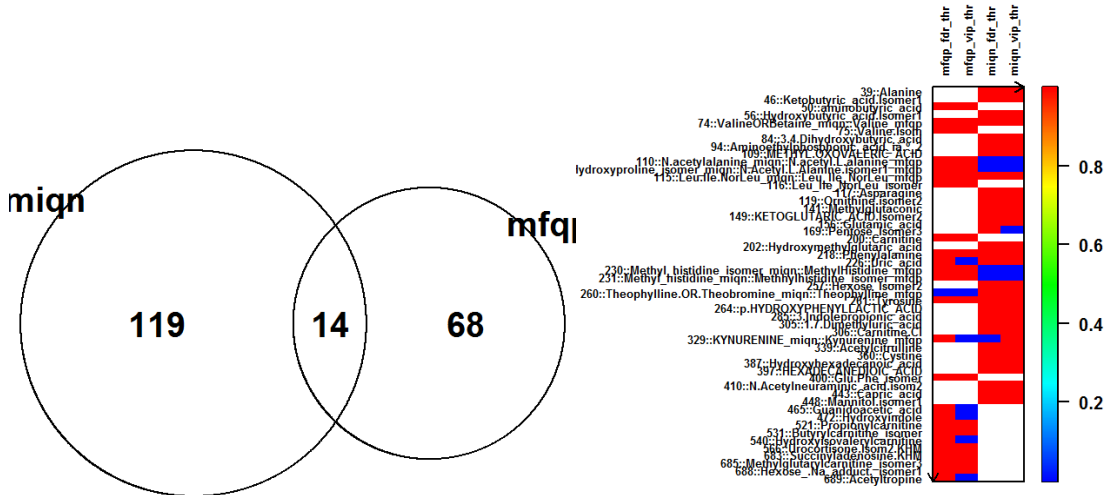


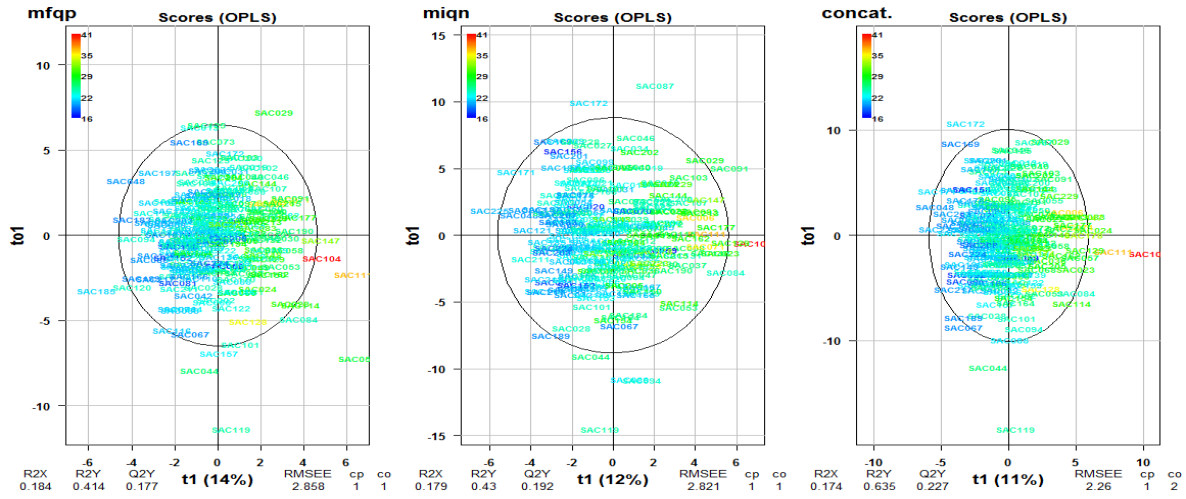


1.2.2. sacVbmi

Annotated features

Significant features in at least one peaktable (sacVbmi)

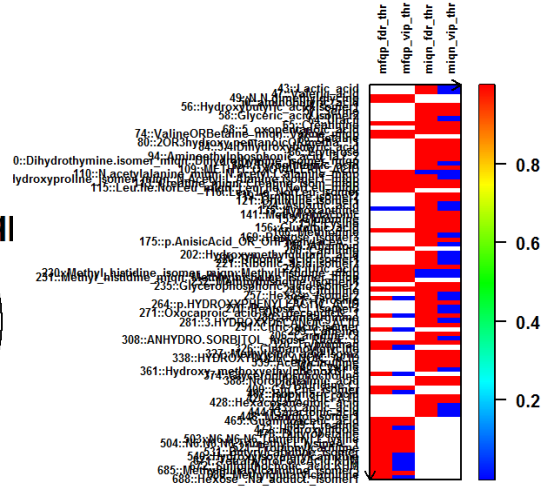
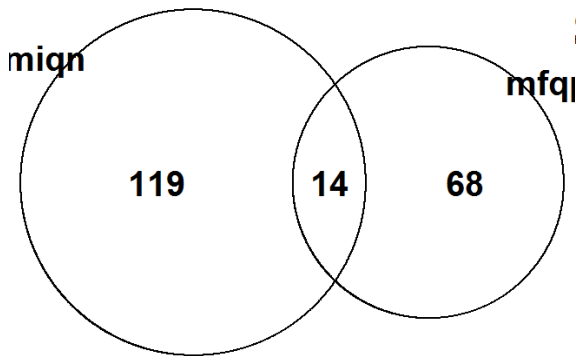




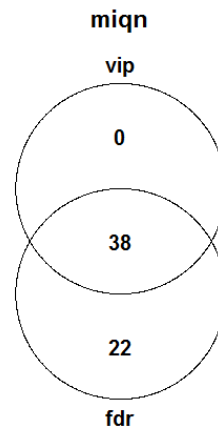
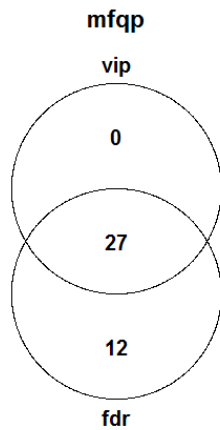
1.2.3. sacVsex

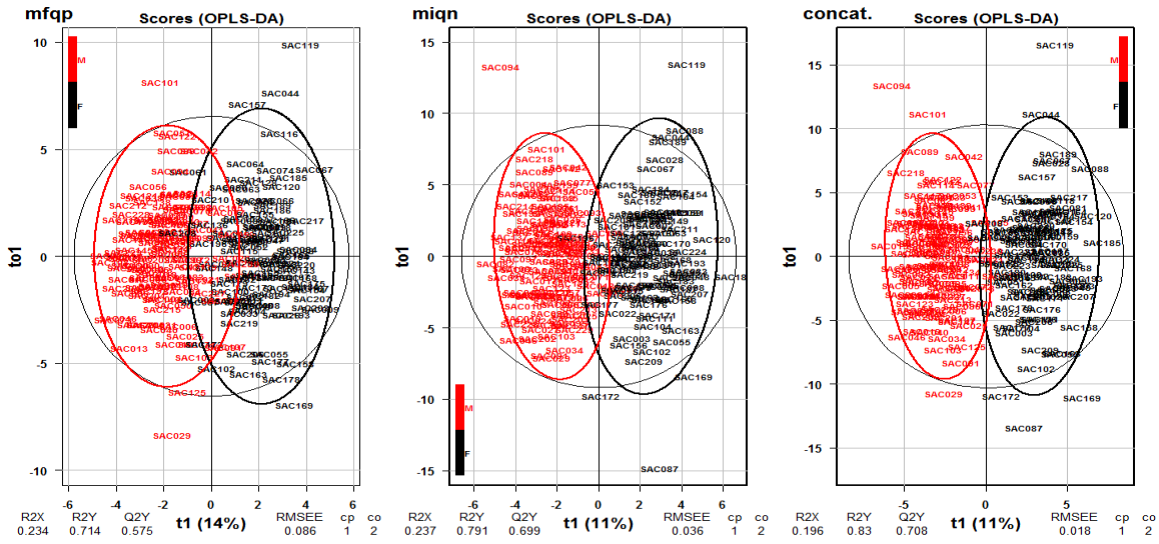
Annotated features

Significant features in at least one peaktable (sacVsex)



sacVsex





2. Analyse lipidomique

2.1. lip-xcm-pos

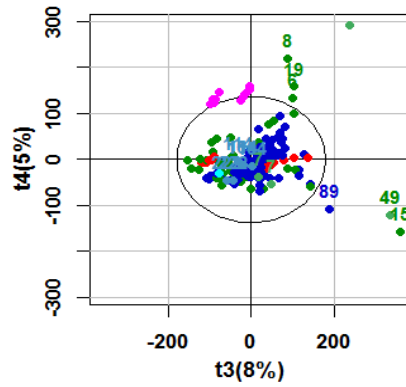


Fig S : PCA des données lipidomiques fusionnées (RPLC-ESI(+)/HRMS)

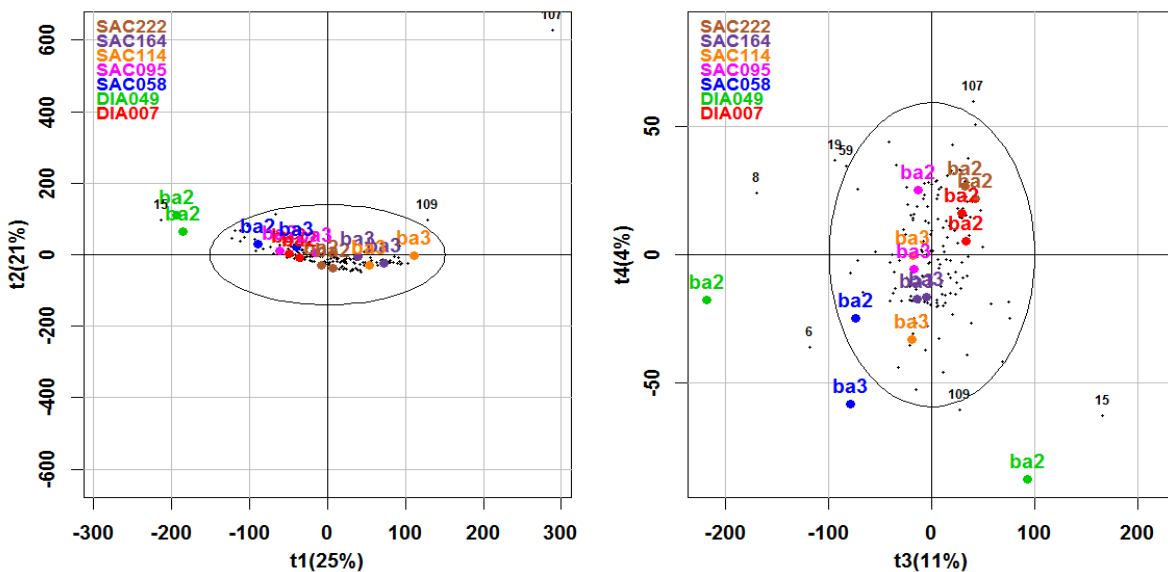


Fig S. : Visualisation des réplicats techniques après fusion des données lipidomiques (RPLC-ESI(+)/HRMS)

2.2. lip-xcm-neg

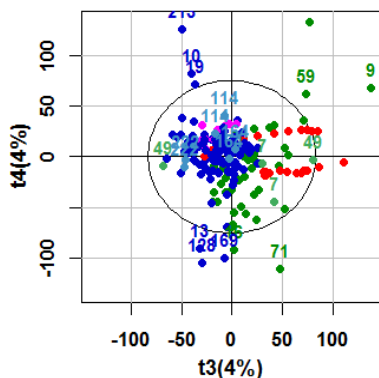


Fig S. : PCA des données lipidomiques fusionnées (RPLC-ESI(-)/HRMS)

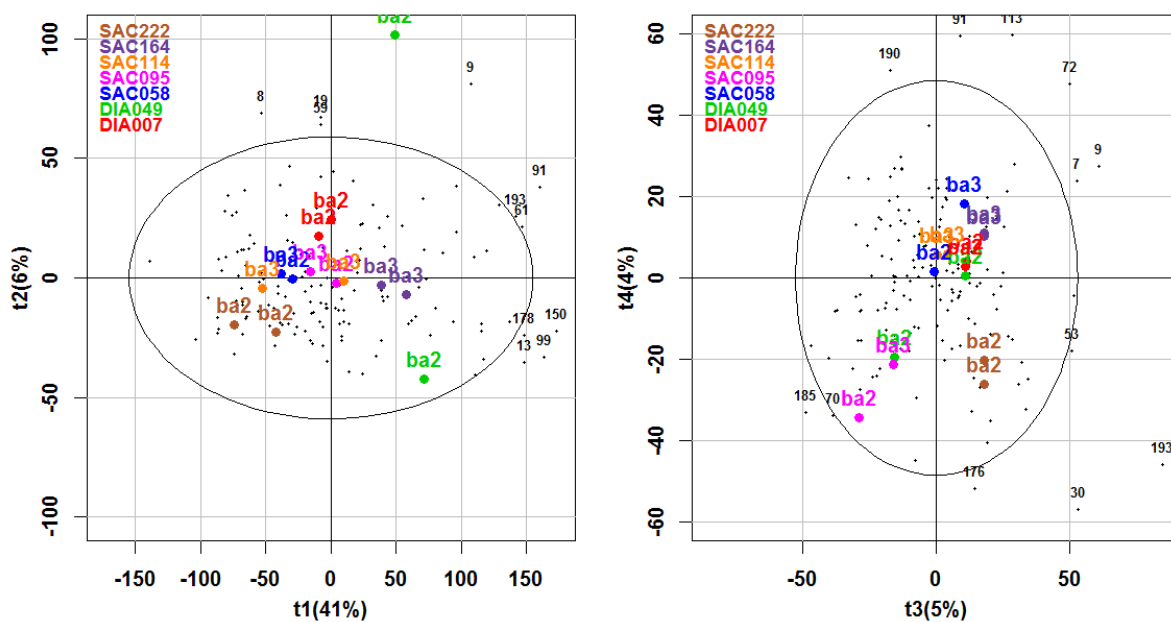
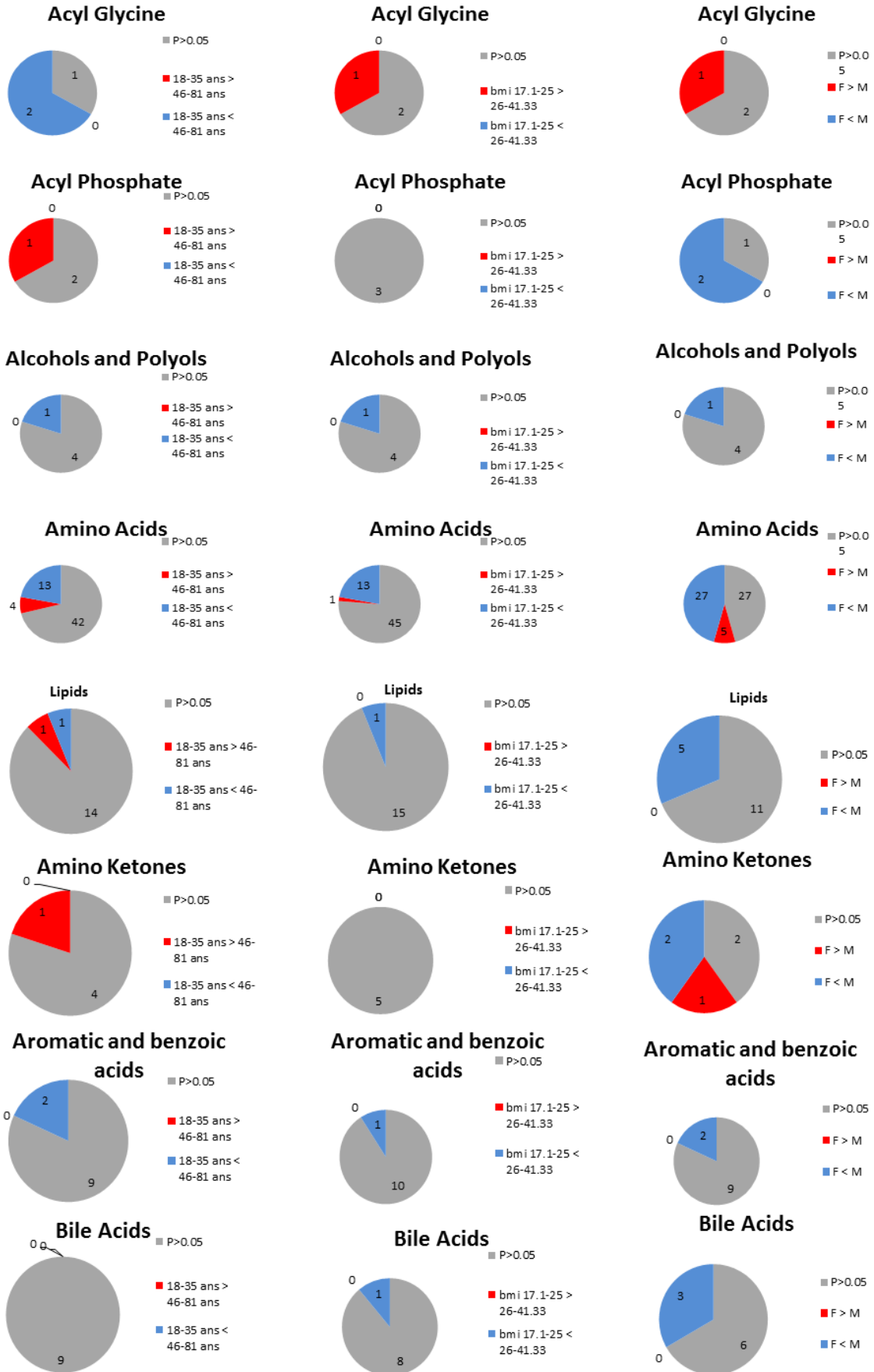
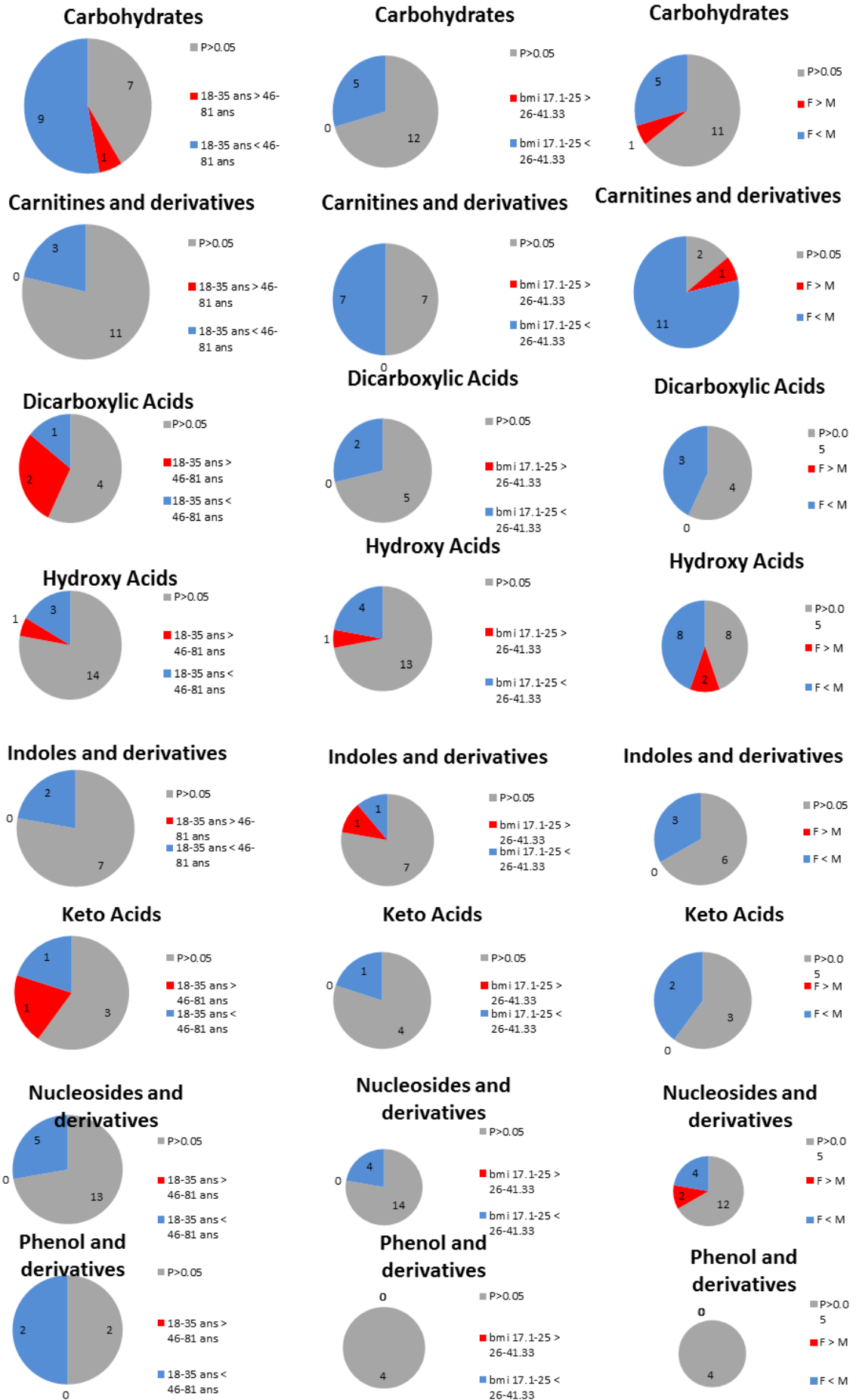
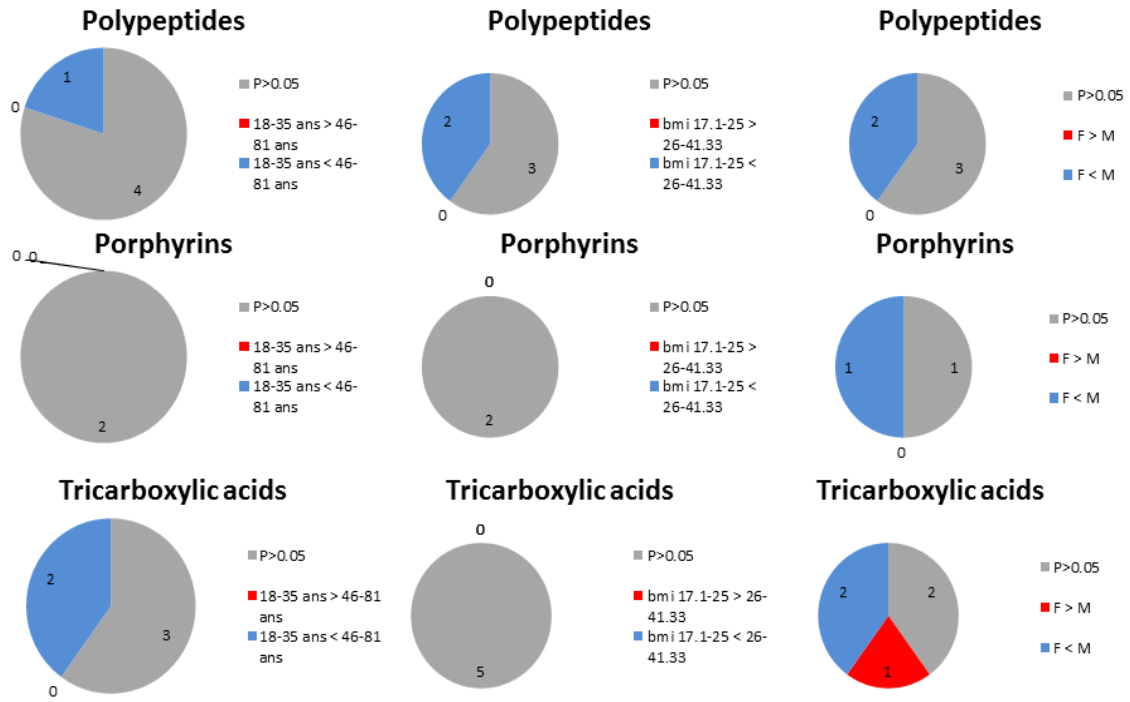


Fig S. : Visualisation des réplicats techniques après fusion des données lipidomiques (RPLC-ESI(+)/HRMS)

RESULTATS DES TESTS DE L'ANALYSE DE LA COHORTE SACLAYSIENNE





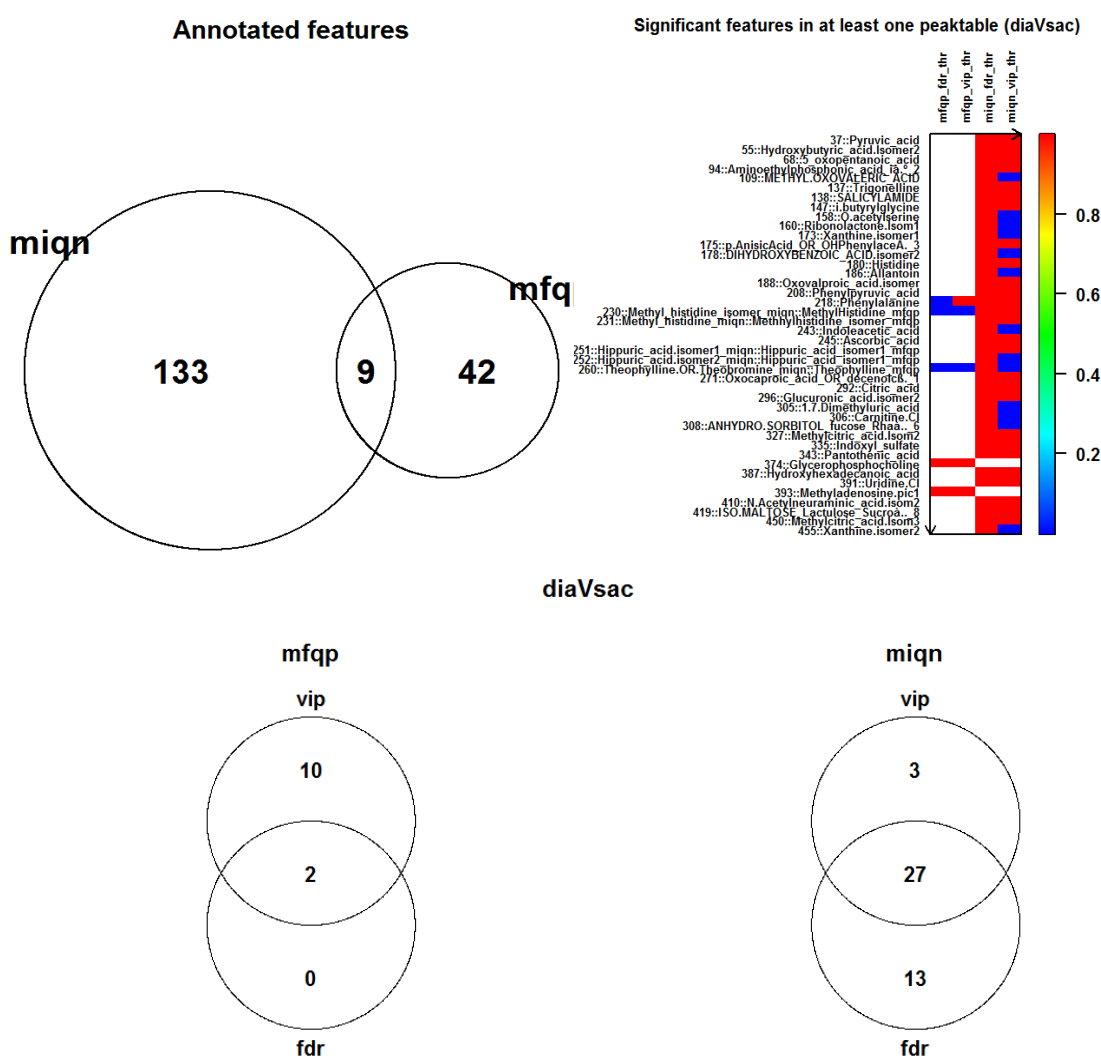


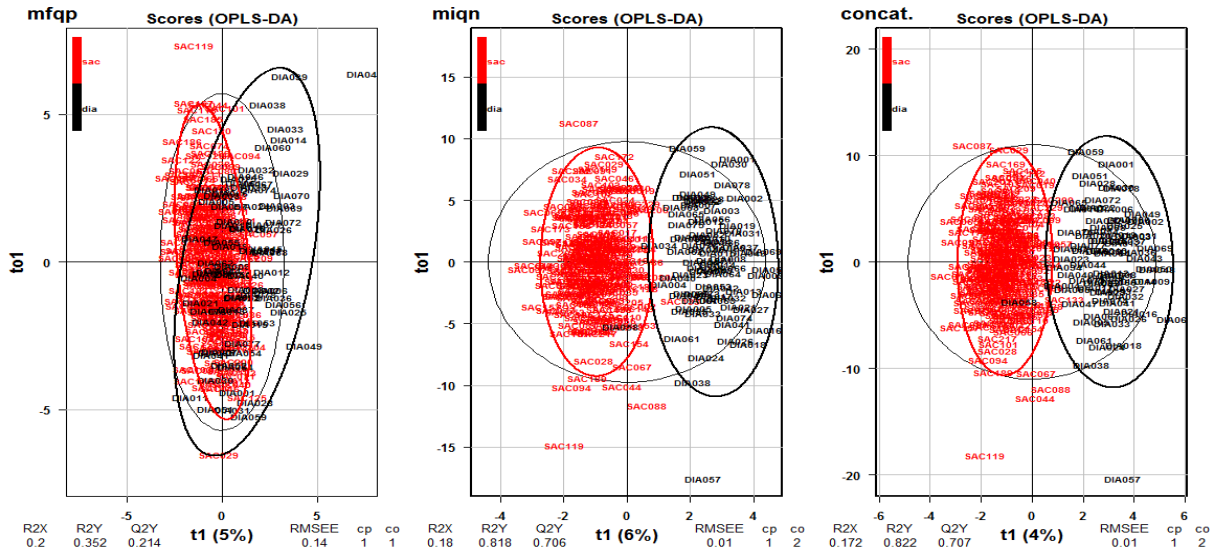
ANNEXE 5

RESULTATS DES TESTS STATISTIQUES DE L'ANALYSE DE LA COHORTE DE PATIENTS DIABETIQUES – DONNEES PLASMATIQUES

1. QuanBrowser (PFPP-pos and HILIC-neg)

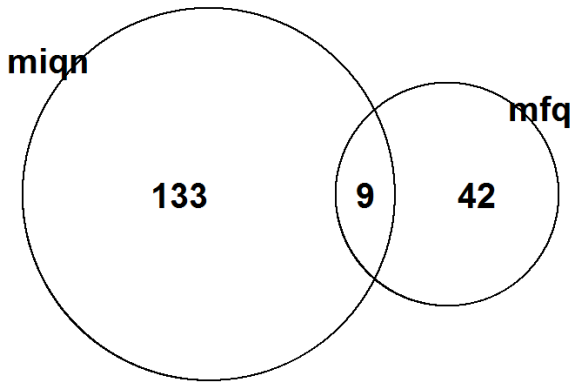
1.1. diaVsac



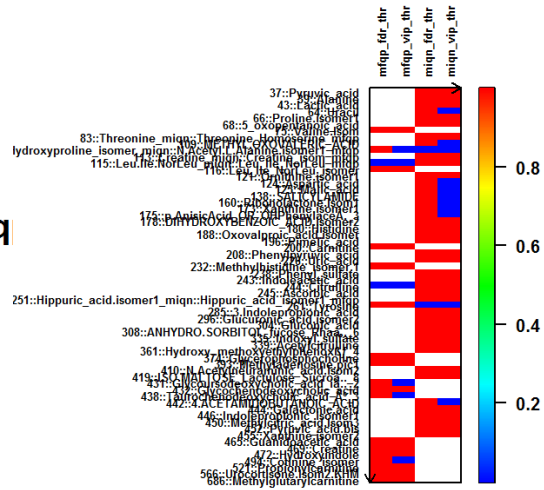


1.2. di1Vsac

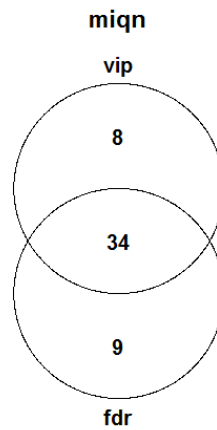
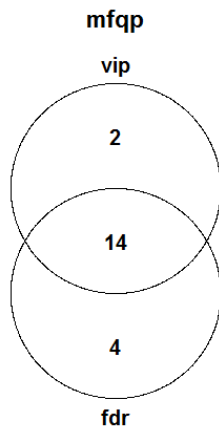
Annotated features

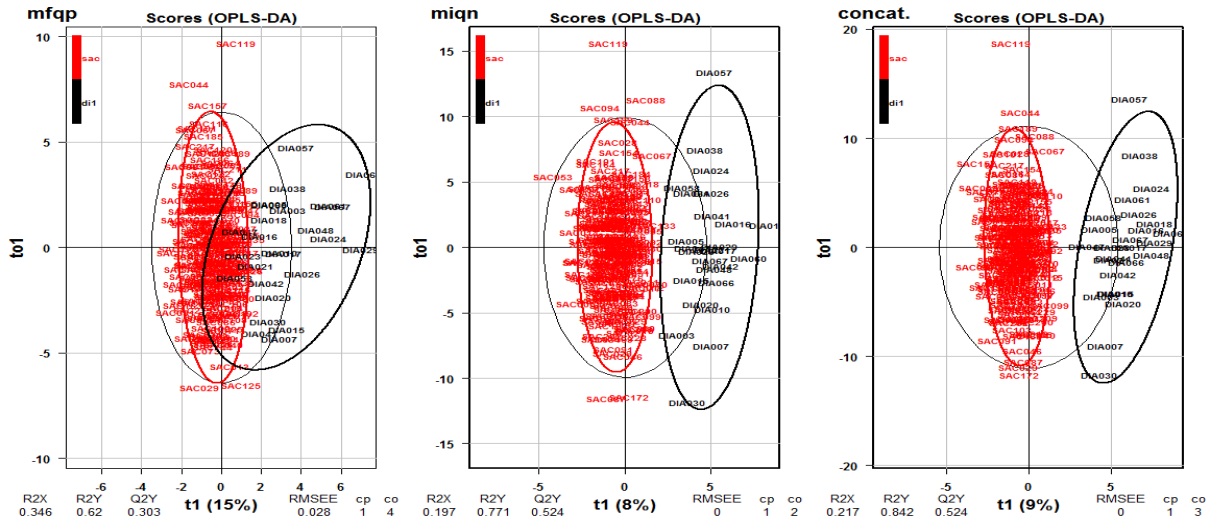


Significant features in at least one peaktable (di1Vsac)



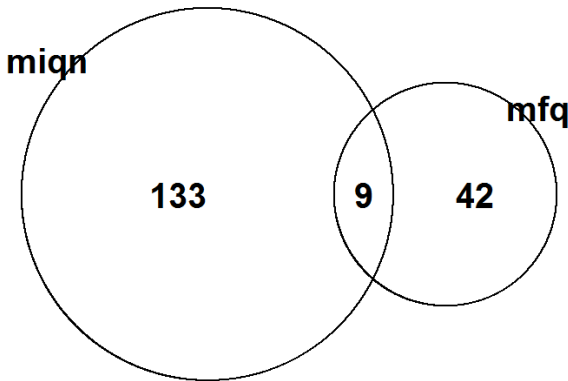
di1Vsac



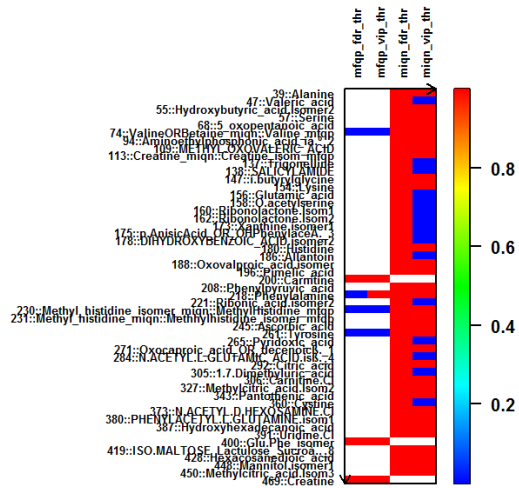


1.3. di2Vsc

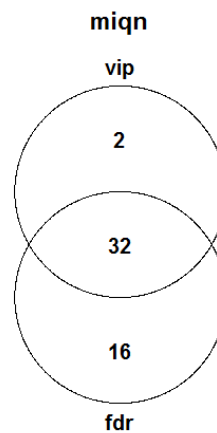
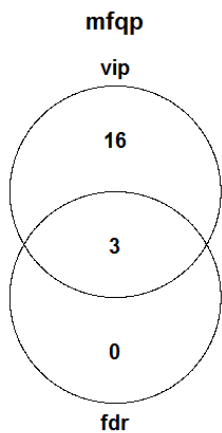
Annotated features

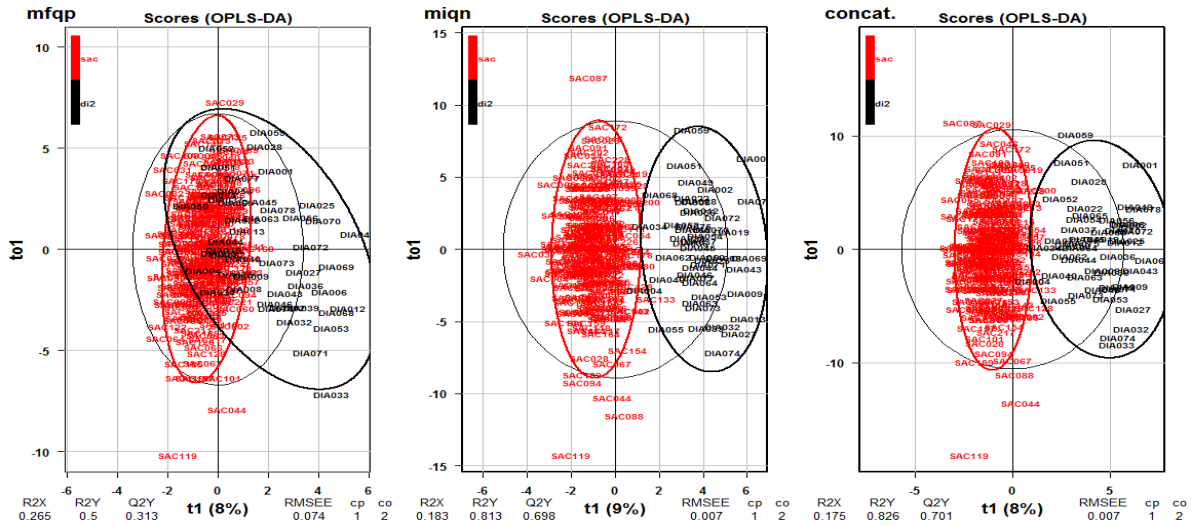


Significant features in at least one peaktable (di2Vsc)

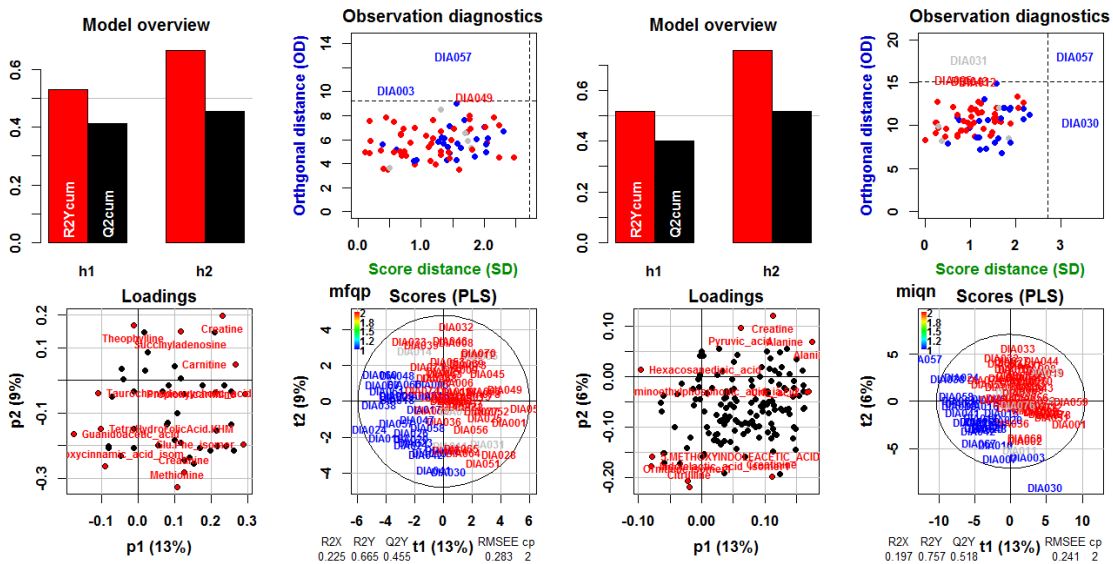


di2Vsc

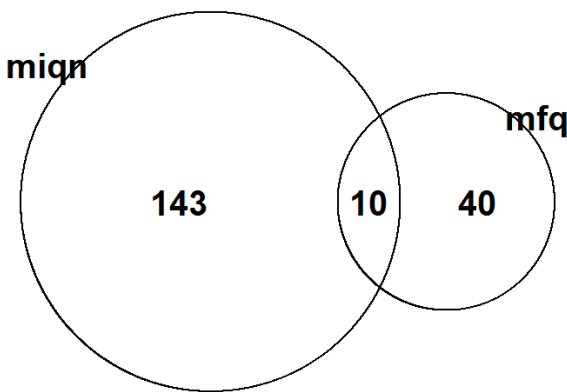




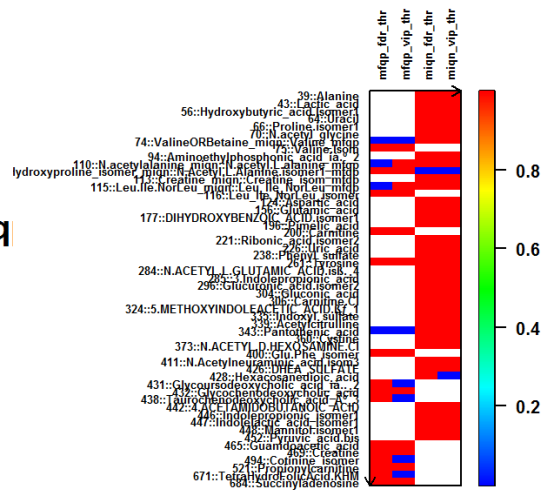
1.4. di1Vdi2



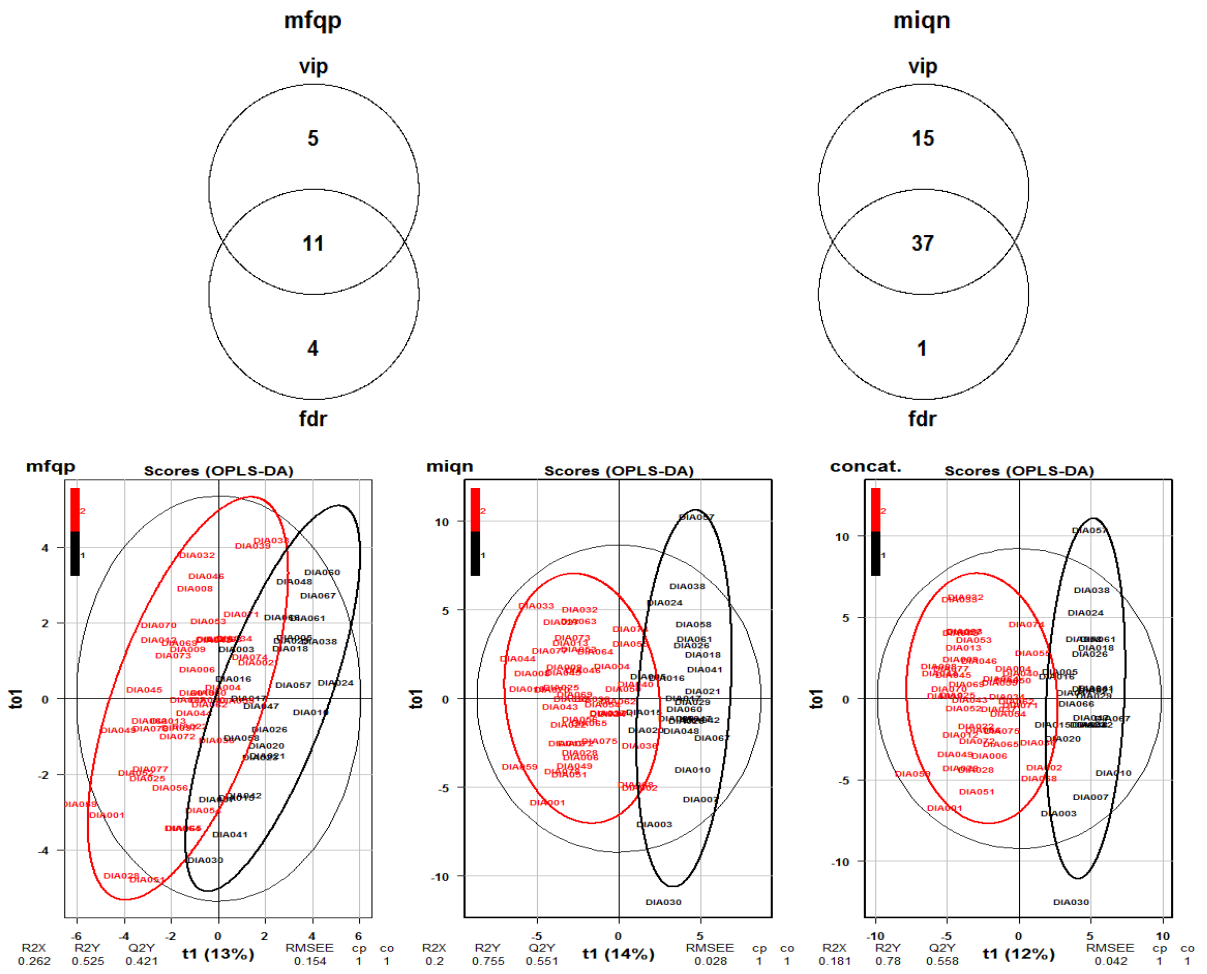
Annotated features



Significant features in at least one peaktable (di1Vdi2)

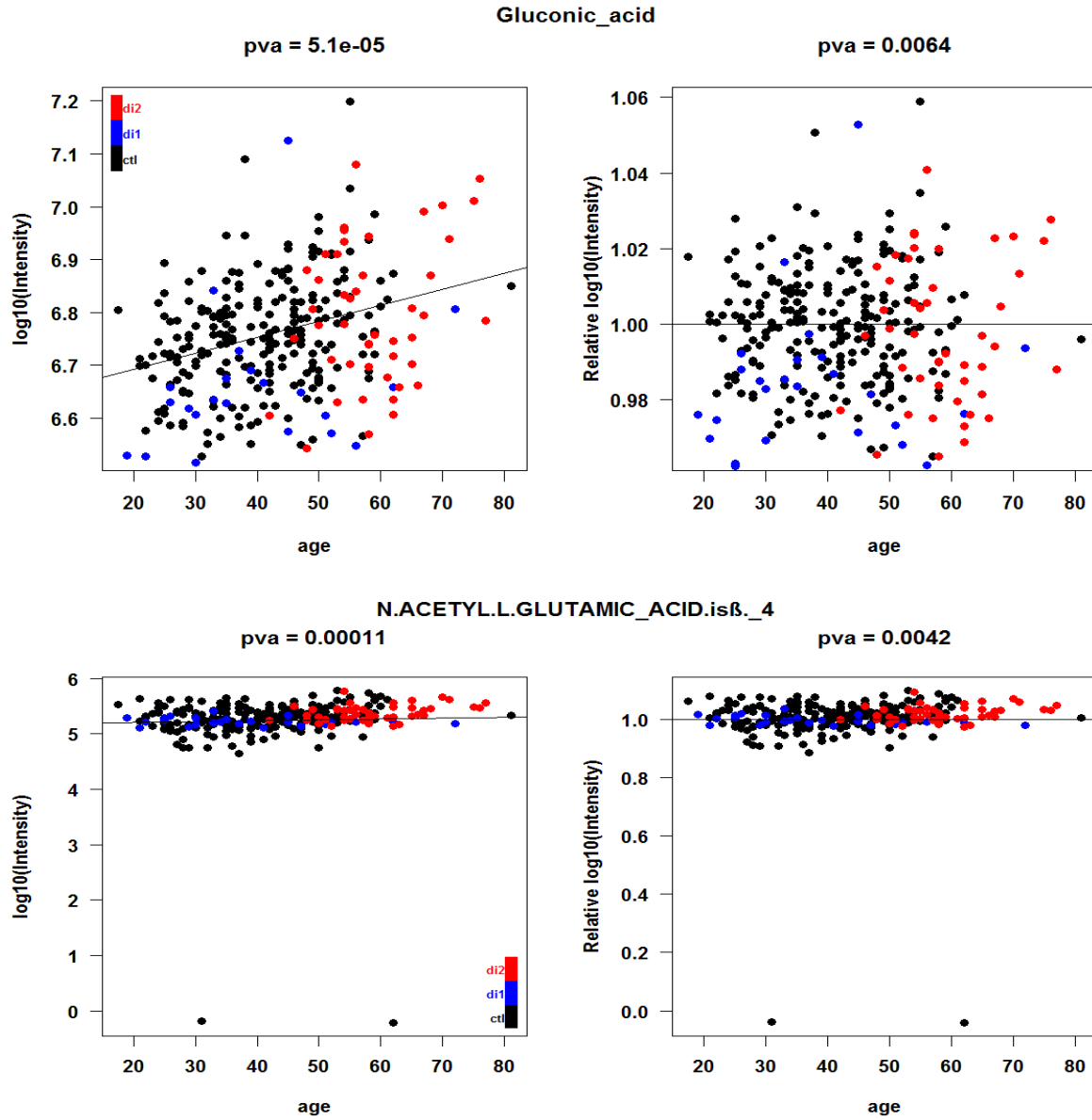


di1Vdi2

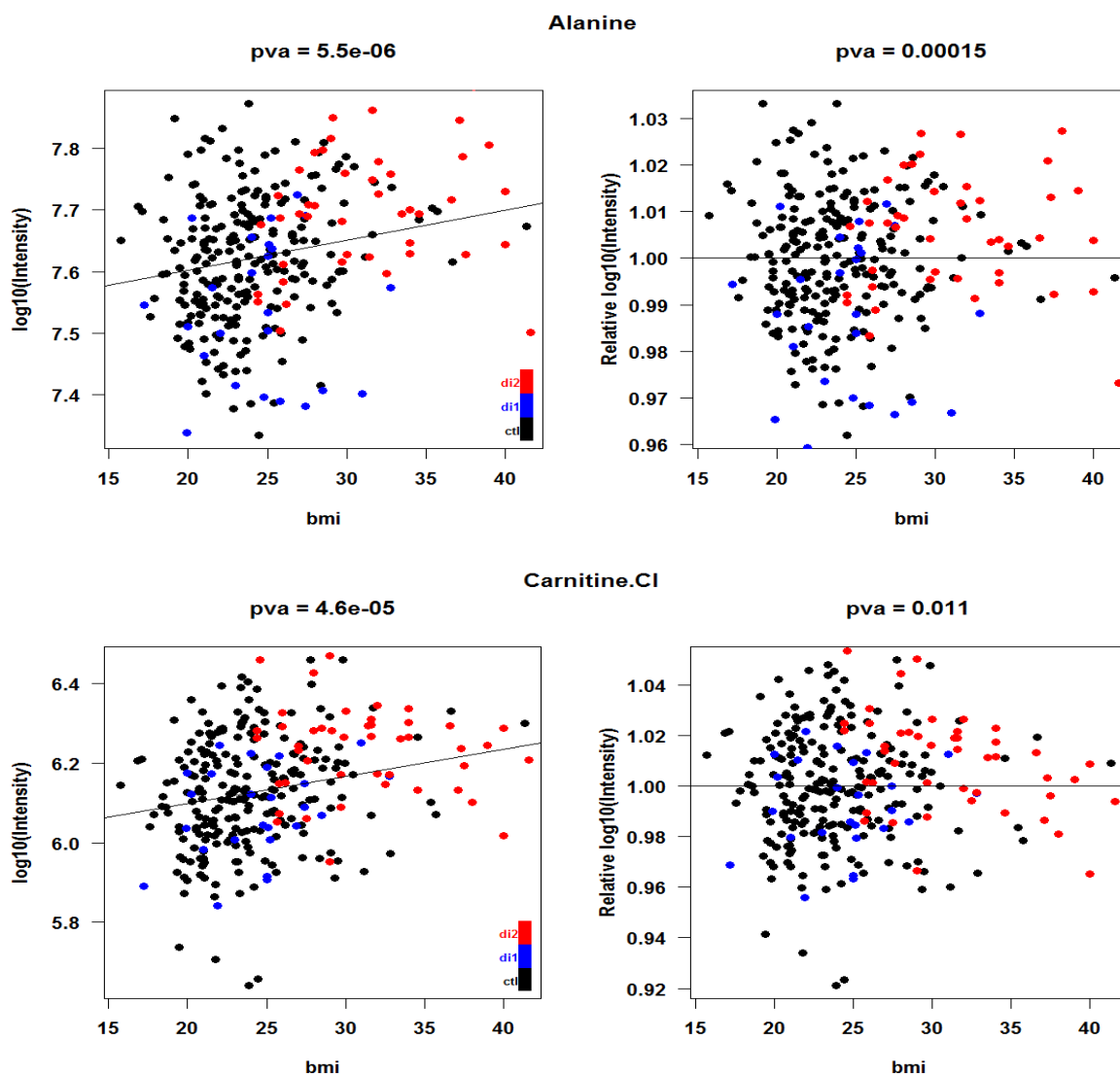


2. Correction du signal en fonction des facteurs physiologiques confondants

2.1. Age

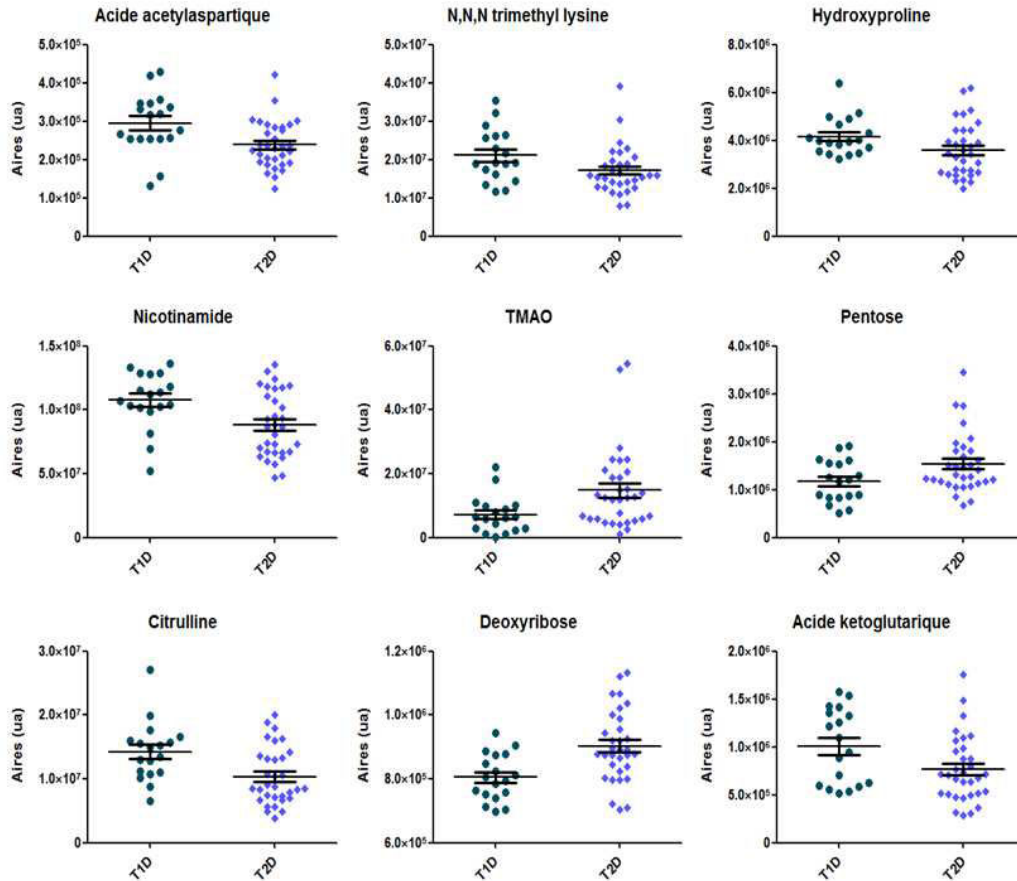


2.2. BMI



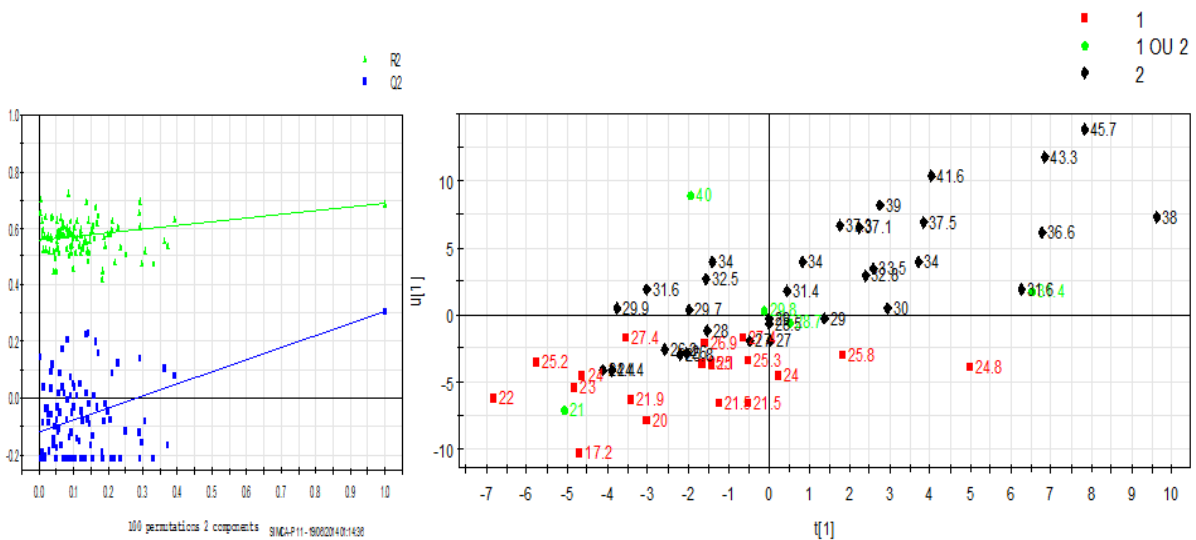
DONNEES ERYTHROCYTAIRES

1. Type de diabète



Empreinte métabolomique érythrocytaire : T1D vs T2D

2. Empreinte érythrocytaire et BMI



Modèle PLS ($y = \text{BMI}$) et sa validation par des tests de permutations croisées sur les empreintes érythrocytaires

PRODUCTIONS SCIENTIFIQUES

Communications affichées

1/ LC-HRMS and data mining tools for the combined metabolomic and lipidomic serum profiling of human cohorts.

S.Boudah, E.Thévenot, A.Seyer, S.Broudin, L.Oliveira, F.Castelli, J-C.Tabet, B.Colsch, C.Junot.

62èmes conférences de la société américaine de spectrométrie de masse (ASMS), 15-19 Juin 2014, Baltimore, USA.

2/ Characterization and localization of d18:2 sphingadiénine based sulfatides in rat cerebellum using 2D offline LC-HRMS and MALDI imaging

B.Colsch; S.Boudah; C.Junot; A.Woods.

62èmes conférences de la société américaine de spectrométrie de masse (ASMS), 9-13 Juin , 2013, Minneapolis, USA.

3/ Comprehensive Liquid Chromatography coupled to High Resolution Mass Spectrometry methods for the global metabolic profiling of human serum

S.Boudah, M-F.Olivier, S.Aros-Calt, L.Oliveira, F.Fenaille, J-C.Tabet, C.Junot.

62èmes conférences de la société américaine de spectrométrie de masse (ASMS), 9-13 Juin , 2013, Minneapolis, USA.

Communications orales

1/ LC-HRMS and data mining tools for the combined metabolomic and lipidomic serum profiling of human cohorts.

S.Boudah, E.Thévenot, A.Seyer, S.Broudin, L.Oliveira, F.Castelli, J-C.Tabet, B.Colsch, C.Junot.

8èmes journées du RFMF (Session Jeunes), 19-21 Mai 2014, Lyon, France.

2/ Characterization and localization of d18:2 sphingadiénine based sulfatides in rat cerebellum using 2D offline LC-HRMS and MALDI imaging.

B.Colsch; S.Boudah; C.Junot; A.Woods

Club Jeune de Spectrométrie de Masse. 8-12 Avril 2013, Pau, France.

3/ Liquid chromatography-Mass spectrometry based analysis of plasma and red blood cells metabolome for the stratification of type 2 diabetes.

S.Boudah, B.Colsch, E.Thévenot, E.Nicodème, E.Larger, C.Junot.

6èmes journées du RFMF, 21-23 Mai 2012, Nantes, France et DSV-PhD meetings, 15-17 Octobre 2012, Iles de Porquerolles, France.

Publications

1/ Annotation of human serum metabolome by coupling three distinct liquid chromatography conditions to high-resolution mass spectrometry.

S.Boudah, M-F.Olivier, S.Aros-Calt, L.Oliveira, F.Fenaille, J-C.Tabet, C.Junot.

J.Chrom.B 2014, Metabolomics special issue, disponible en ligne

2/ Liquid chromatography coupled to mass spectrometry based metabolomics and the concept of biomarker

S.Boudah, A.Paris, C.Junot

Advances in botanical researches, Volume 67 (2013): «Metabolomics coming on age with its technological diversity», Chapitre 4, P159-218.

3/ Advances and challenges in liquid chromatography-mass spectrometry based metabolomics (*en français*)

S.Boudah, A.Roux, C.Junot.

Spectra Analyse N°284 –2012, P52-68 et Spectra Biologie N°193, Avril 2012, P22-38.

4/ Global lipidomic approach based on ultra-performance liquid chromatography coupled to high resolution mass spectrometry and integrated data processing tools: Application to human cerebrospinal fluid.

A.Seyer§, S.Boudah§, S.Broudin, C.Junot, B.Colsch

(En préparation)

REFERENCES

- A J, Trygg, J., Gullberg, J., Johansson, A. I., Jonsson, P., Antti, H., Marklund, S. L., and Moritz, T. (2005). Extraction and GC/MS analysis of the human blood plasma metabolome. *Anal.Chem.* 77, 24, 8086-8094.
- Addona, T.A., Shi, X., Keshishian, H., Mani, D. R., Burgess, M., Gillette, M. A., Clauser, K. R., Shen, D., Lewis, G. D., Farrell, L. A., Fifer, M. A., Sabatine, M. S., Gerszten, R. E., and Carr, S. A. (2011). A pipeline that integrates the discovery and verification of plasma protein biomarkers reveals candidate markers for cardiovascular disease. *Nature Biotechnology* 29, 7, 635-643.
- Adosraku, R.K., Choi, G. T., Constantinou-Kokotos, V., Anderson, M. M., and Gibbons, W. A. (1994). NMR lipid profiles of cells, tissues, and body fluids: proton NMR analysis of human erythrocyte lipids. *J.Lipid Res.* 35, 11, 1925-1931.
- Afonso, C., Riu, A., Xu, Y., Fournier, F., and Tabet, J. C. (2005). Structural characterization of fatty acids cationized with copper by electrospray ionization mass spectrometry under low-energy collision-induced dissociation. *J.Mass Spectrom.* 40, 3, 342-349.
- Aharoni, A., Ric, d., V, Verhoeven, H. A., Maliepaard, C. A., Kruppa, G., Bino, R., and Goodenowe, D. B. (2002). Nontargeted metabolome analysis by use of Fourier Transform Ion Cyclotron Mass Spectrometry. *OMICS* 6, 3, 217-234.
- Al-Saad, K.A., Siems, W. F., Hill, H. H., Zabrouskov, V., and Knowles, N. R. (2003). Structural analysis of phosphatidylcholines by post-source decay matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *J.Am.Soc.Mass Spectrom.* 14, 4, 373-382.
- Albert, C.J., Anbukumar, D. S., Messner, M. C., and Ford, D. A. (2009). Chromatographic methods for the analyses of 2-halofatty aldehydes and chlorohydrin molecular species of lysophosphatidylcholine. *J.Chromatogr.B Analyt.Technol.Biomed.Life Sci.* 877, 26, 2768-2777.
- Alonso, A., Julia, A., Beltran, A., Vinaixa, M., Diaz, M., Ibanez, L., Correig, X., and Marsal, S. (2011). AStream: an R package for annotating LC/MS metabolomic data. *Bioinformatics.* 27, 9, 1339-1340.
- Alston, M.J., Seers, J., Hinton, J. C., and Lucchini, S. (2010). BABAR: an R package to simplify the normalisation of common reference design microarray-based transcriptomic datasets. *BMC.Bioinformatics.* 11, 73-
- Alymani, N.A., Smith, M. D., Williams, D. J., and Petty, R. D. (2010). Predictive biomarkers for personalised anti-cancer drug use: discovery to clinical implementation. *Eur.J.Cancer* 46, 5, 869-879.
- Anderson, N.L. (2010). The clinical plasma proteome: a survey of clinical assays for proteins in plasma and serum. *Clin.Chem.* 56, 2, 177-185.
- Angerer, J., Ewers, U., and Wilhelm, M. (2007). Human biomonitoring: State of the art. *International Journal of Hygiene and Environmental Health* 210, 3-4, 201-228.

- Apffel, A., Fischer, S., Goldberg, G., Goodley, P. C., and Kuhlmann, F. E. (1995). Enhanced sensitivity for peptide mapping with electrospray liquid chromatography-mass spectrometry in the presence of signal suppression due to trifluoroacetic acid-containing mobile phases. *J.Chromatogr.A* 712, 1, 177-190.
- Armstrong, J.A. (2007). Urinalysis in Western culture: a brief history. *Kidney Int.* 71, 5, 384-387.
- Bader, M., Lehnert, G., and Angerer, J. (1994). GC/MS determination of N-phenylvaline, a possible biomarker for benzene exposure in human hemoglobin by the "N-alkyl Edman method". *Int.Arch.Occup.Environ.Health* 65, 6, 411-414.
- Bain, M.A., Faull, R., Fornasini, G., Milne, R. W., and Evans, A. M. (2006). Accumulation of trimethylamine and trimethylamine-N-oxide in end-stage renal disease patients undergoing haemodialysis. *Nephrol.Dial.Transplant.* 21, 5, 1300-1304.
- Bajad, S. and Shulaev, V. (2007). Highly-parallel metabolomics approaches using LC-MS for pharmaceutical and environmental analysis. *Trends Analyt.Chem.* 26, 6, 625-636.
- Bajad, S.U., Lu, W., Kimball, E. H., Yuan, J., Peterson, C., and Rabinowitz, J. D. (2006). Separation and quantitation of water soluble cellular metabolites by hydrophilic interaction chromatography-tandem mass spectrometry. *J.Chromatogr.A* 1125, 1, 76-88.
- Baker, M. (2005). In biomarkers we trust? *Nature Biotechnology* 23, 3, 297-304.
- Bamba, T., Lee, J. W., Matsubara, A., and Fukusaki, E. (2012). Metabolic profiling of lipids by supercritical fluid chromatography/mass spectrometry. *J.Chromatogr.A* 1250, 212-219.
- Bando, K., Kawahara, R., Kunimatsu, T., Sakai, J., Kimura, J., Funabashi, H., Seki, T., Bamba, T., and Fukusaki, E. (2010). Influences of biofluid sample collection and handling procedures on GC-MS based metabolomic studies. *J.Biosci.Bioeng.* 110, 4, 491-499.
- Barber, M., Bordoli, R. S., Sedgwick, R. D., and Tyler, A. N. (1981). Fast atom bombardment of solids as an ion source in mass spectrometry. *Nature* 293, 270-275.
- Barber, M.N., Risis, S., Yang, C., Meikle, P. J., Staples, M., Febbraio, M. A., and Bruce, C. R. (2012). Plasma lysophosphatidylcholine levels are reduced in obesity and type 2 diabetes. *PLoS.One.* 7, 7, e41456-
- Barroso, B. and Bischoff, R. (2005). LC-MS analysis of phospholipids and lysophospholipids in human bronchoalveolar lavage fluid. *J.Chromatogr.B Analyt.Technol.Biomed.Life Sci.* 814, 1, 21-28.
- Begley, P., Francis-McIntyre, S., Dunn, W. B., Broadhurst, D. I., Halsall, A., Tseng, A., Knowles, J., Goodacre, R., and Kell, D. B. (2009). Development and performance of a gas chromatography-time-of-flight mass spectrometry analysis for large-scale nontargeted metabolomic studies of human serum. *Anal.Chem.* 81, 16, 7038-7046.
- Bennett, B.D., Yuan, J., Kimball, E. H., and Rabinowitz, J. D. (2008). Absolute quantitation of intracellular metabolite concentrations by an isotope ratio-based approach. *Nat.Protoc.* 3, 8, 1299-1311.
- Bennett, D.A. and Waters, M. D. (2000). Applying biomarker research. *Environ.Health Perspect.* 108, 9, 907-910.

- Benton, H.P., Wong, D. M., Trauger, S. A., and Siuzdak, G. (2008). XCMS2: processing tandem mass spectrometry data for metabolite identification and structural characterization. *Anal.Chem.* 80, 16, 6382-6389.
- Bernard, C. (1865). Introduction à l'étude de la médecine expérimentale.
- Berry, K.A. and Murphy, R. C. (2005). Analysis of cell membrane aminophospholipids as isotope-tagged derivatives. *J.Lipid Res.* 46, 5, 1038-1046.
- Biomarkers Definitions Working Group (2001). Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol.Ther.* 69, 3, 89-95.
- Bird, S.S., Marur, V. R., Sniatynski, M. J., Greenberg, H. K., and Kristal, B. S. (2011). Serum lipidomics profiling using LC-MS and high-energy collisional dissociation fragmentation: focus on triglyceride detection and characterization. *Anal.Chem.* 83, 17, 6648-6657.
- BLIGH, E.G. and DYER, W. J. (1959). A rapid method of total lipid extraction and purification. *Can.J.Biochem.Physiol* 37, 8, 911-917.
- Boccard, J. and Rutledge, D. N. (2013). A consensus orthogonal partial least squares discriminant analysis (OPLS-DA) strategy for multiblock Omics data fusion. *Anal.Chim.Acta* 769, 30-39.
- Bodennec, J., Brichon, G., Zwingelstein, G., and Portoukalian, J. (2000). Purification of sphingolipid classes by solid-phase extraction with aminopropyl and weak cation exchanger cartridges. *Methods Enzymol.* 312, 101-114.
- Bolten, C.J., Kiefer, P., Letisse, F., Portais, J. C., and Wittmann, C. (2007). Sampling for metabolome analysis of microorganisms. *Anal.Chem.* 79, 10, 3843-3849.
- Bolton, S. and Null, G. (1981). Caffeine Psychological Effects, Use and Abuse. *Journal of Orthomolecular Psychiatry* 10, 3, 202-211.
- Boswell, P.G., Schellenberg, J. R., Carr, P. W., Cohen, J. D., and Hegeman, A. D. (2011). Easy and accurate high-performance liquid chromatography retention prediction with different gradients, flow rates, and instruments by back-calculation of gradient and flow rate profiles. *J.Chromatogr.A* 1218, 38, 6742-6749.
- Bou, K.M., Hou, W., Zhou, H., Elisma, F., Swayne, L. A., Blanchard, A. P., Yao, Z., Bennett, S. A., and Figeys, D. (2010). Lipidomics era: accomplishments and challenges. *Mass Spectrom.Rev.* 29, 6, 877-929.
- Boudah, S., Olivier, M. F., ros-Calt, S., Oliveira, L., Fenaille, F., Tabet, J. C., and Junot, C. (2014). Annotation of the human serum metabolome by coupling three liquid chromatography methods to high-resolution mass spectrometry. *J.Chromatogr.B Analyt.Technol.Biomed.Life Sci.*
- Boudah, S., Roux, A., and Junot, C. (2012). Analyse métabolomique par chromatographie liquide couplée à la spectrométrie de masse à haute résolution : état des lieux et perspectives. *Spectra Analyse* 284, 52-68.
- Boudah, S., Paris, A., and Junot, C. (2013). Liquid Chromatography Coupled to Mass Spectrometry-Based Metabolomics and the Concept of Biomarker. 159-218.

- Breitling, R., Pitt, A. R., and Barrett, M. P. (2006). Precision mapping of the metabolome. *Trends Biotechnol.* 24, 12, 543-548.
- Brison, D.R., Houghton, F. D., Falconer, D., Roberts, S. A., Hawkhead, J., Humpherson, P. G., Lieberman, B. A., and Leese, H. J. (2004). Identification of viable embryos in IVF by non-invasive measurement of amino acid turnover. *Hum.Reprod.* 19, 10, 2319-2324.
- Brown, M., Dunn, W. B., Dobson, P., Patel, Y., Winder, C. L., Francis-McIntyre, S., Begley, P., Carroll, K., Broadhurst, D., Tseng, A., Swainston, N., Spasic, I., Goodacre, R., and Kell, D. B. (2009). Mass spectrometry tools and metabolite-specific databases for molecular identification in metabolomics. *Analyst* 134, 7, 1322-1332.
- Brownsill, R., Combal, J. P., Taylor, A., Bertrand, M., Luijten, W., and Walther, B. (1994). The application of electrospray and neutral-loss mass spectrometry to the identification of the metabolites of S12813 in rat liver slices. *Rapid Commun.Mass Spectrom.* 8, 5, 361-365.
- Bruce, S.J., Tavazzi, I., Parisod, V., Rezzi, S., Kochhar, S., and Guy, P. A. (2009). Investigation of human blood plasma sample preparation for performing metabolomics using ultrahigh performance liquid chromatography/mass spectrometry. *Anal.Chem.* 81, 9, 3285-3296.
- Bure, C., Ayciriex, S., Testet, E., and Schmitter, J. M. (2013). A single run LC-MS/MS method for phospholipidomics. *Anal.Bioanal.Chem.* 405, 1, 203-213.
- Buscher, J.M., Czernik, D., Ewald, J. C., Sauer, U., and Zamboni, N. (2009). Cross-platform comparison of methods for quantitative metabolomics of primary metabolism. *Anal.Chem.* 81, 6, 2135-2143.
- Byrdwell, W.C. (2001). Atmospheric pressure chemical ionization mass spectrometry for analysis of lipids. *Lipids* 36, 4, 327-346.
- Byrdwell, W.C. (2005). The bottom-up solution to the triacylglycerol lipidome using atmospheric pressure chemical ionization mass spectrometry. *Lipids* 40, 4, 383-417.
- Caffrey, M. and Hogan, J. (1992). LIPIDAT: a database of lipid phase transition temperatures and enthalpy changes. DMPC data subset analysis. *Chem.Phys.Lipids* 61, 1, 1-109.
- Cai, S.S. and Syage, J. A. (2006). Atmospheric pressure photoionization mass spectrometry for analysis of fatty acid and acylglycerol lipids. *J.Chromatogr.A* 1110, 1-2, 15-26.
- Canelas, A.B., ten, P. A., Ras, C., Seifar, R. M., van Dam, J. C., van Gulik, W. M., and Heijnen, J. J. (2009). Quantitative evaluation of intracellular metabolite extraction techniques for yeast metabolomics. *Anal.Chem.* 81, 17, 7379-7389.
- Caprioli, R.M., Farmer, T. B., and Gile, J. (1997). Molecular imaging of biological samples: localization of peptides and proteins using MALDI-TOF MS. *Anal.Chem.* 69, 23, 4751-4760.
- Carr, S.A. and Anderson, L. (2008). Protein quantitation through targeted mass spectrometry: the way out of biomarker purgatory? *Clin.Chem.* 54, 11, 1749-1752.
- Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C. A., Holland, T. A., Keseler, I. M., Kothari, A., Kubo, A., Krummenacker, M., Latendresse, M., Mueller, L. A., Ong, Q., Paley, S., Subhraveti, P., Weaver, D. S., Weerasinghe, D., Zhang, P., and Karp, P. D. (2014). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res.* 42, Database issue, D459-D471.

- Castrillo, J.I., Hayes, A., Mohammed, S., Gaskell, S. J., and Oliver, S. G. (2003). An optimized protocol for metabolome analysis in yeast using direct infusion electrospray mass spectrometry. *Phytochemistry* 62, 6, 929-937.
- Castro-Perez, J.M., Kamphorst, J., DeGroot, J., Lafeber, F., Goshawk, J., Yu, K., Shockcor, J. P., Vreeken, R. J., and Hankemeier, T. (2010). Comprehensive LC-MS E lipidomic analysis using a shotgun approach and its application to biomarker detection and identification in osteoarthritis patients. *J.Proteome.Res.* 9, 5, 2377-2389.
- Chace, D.H. (2001). Mass spectrometry in the clinical laboratory. *Chem.Rev.* 101, 2, 445-477.
- Chace, D.H., Lim, T., Hansen, C. R., Adam, B. W., and Hannon, W. H. (2009a). Quantification of malonylcarnitine in dried blood spots by use of MS/MS varies by stable isotope internal standard composition. *Clin.Chim.Acta* 402, 1-2, 14-18.
- Chace, D.H., Lim, T., Hansen, C. R., De, J., V, and Hannon, W. H. (2009b). Improved MS/MS analysis of succinylacetone extracted from dried blood spots when combined with amino acids and acylcarnitine butyl esters. *Clin.Chim.Acta* 407, 1-2, 6-9.
- Chace, D.H., Singleton, S., Diperna, J., Aiello, M., and Foley, T. (2009c). Rapid metabolic and newborn screening of thyroxine (T4) from dried blood spots by MS/MS. *Clin.Chim.Acta* 403, 1-2, 178-183.
- Chen, J., Zhao, X., Fritsche, J., Yin, P., Schmitt-Kopplin, P., Wang, W., Lu, X., Haring, H. U., Schleicher, E. D., Lehmann, R., and Xu, G. (2008). Practical approach for the identification and isomer elucidation of biomarkers detected in a metabolomic study for the discovery of individuals at risk for diabetes by integrating the chromatographic and mass spectrometric information. *Anal.Chem.* 80, 4, 1280-1289.
- Chen, L., Zhou, L., Chan, E. C., Neo, J., and Beuerman, R. W. (2011). Characterization of the human tear metabolome by LC-MS/MS. *J.Proteome.Res.* 10, 10, 4876-4882.
- Cheng, S., Rhee, E. P., Larson, M. G., Lewis, G. D., McCabe, E. L., Shen, D., Palma, M. J., Roberts, L. D., Dejam, A., Souza, A. L., Deik, A. A., Magnusson, M., Fox, C. S., O'Donnell, C. J., Vasan, R. S., Melander, O., Clish, C. B., Gerszten, R. E., and Wang, T. J. (2012). Metabolite profiling identifies pathways associated with metabolic risk in humans. *Circulation* 125, 18, 2222-2231.
- Chirita, R.I., West, C., Zubrzycki, S., Finaru, A. L., and Elfakir, C. (2011). Investigations on the chromatographic behaviour of zwitterionic stationary phases used in hydrophilic interaction chromatography. *J.Chromatogr.A* 1218, 35, 5939-5963.
- Chung, W.G., Kang, J. H., Park, C. S., Cho, M. H., and Cha, Y. N. (2000). Effect of age and smoking on in vivo CYP1A2, flavin-containing monooxygenase, and xanthine oxidase activities in Koreans: Determination by caffeine metabolism. *Clinical Pharmacology & Therapeutics* 67, 3, 258-266.
- Cifkova, E., Holcapek, M., Lisa, M., Ovcacikova, M., Lycka, A., Lynen, F., and Sandra, P. (2012). Nontargeted quantitation of lipid classes using hydrophilic interaction liquid chromatography-electrospray ionization mass spectrometry with single internal standard and response factor approach. *Anal.Chem.* 84, 22, 10064-10070.
- Cody, R.B., Laramee, J. A., and Durst, H. D. (2005). Versatile new ion source for the analysis of materials in open air under ambient conditions. *Anal.Chem.* 77, 8, 2297-2302.

- Collier, T.S., Hawkrigde, A. M., Georgianna, D. R., Payne, G. A., and Muddiman, D. C. (2008). Top-down identification and quantification of stable isotope labeled proteins from *Aspergillus flavus* using online nano-flow reversed-phase liquid chromatography coupled to a LTQ-FTICR mass spectrometer. *Anal.Chem.* 80, 13, 4994-5001.
- Colsch, B., Afonso, C., Turpin, J. C., Portoukalian, J., Tabet, J. C., and Baumann, N. (2008). Sulfogalactosylceramides in motor and psycho-cognitive adult metachromatic leukodystrophy: relations between clinical, biochemical analysis and molecular aspects. *Biochim.Biophys.Acta* 1780, 3, 434-440.
- Costa, A., Iguale, I., Bedini, J., Quinto, L., and Conget, I. (2002). Uric acid concentration in subjects at risk of type 2 diabetes mellitus: relationship to components of the metabolic syndrome. *Metabolism* 51, 3, 372-375.
- Courant, F., Royer, A. L., Chereau, S., Morvan, M. L., Monteau, F., Antignac, J. P., and Le, B. B. (2012). Implementation of a semi-automated strategy for the annotation of metabolomic fingerprints generated by liquid chromatography-high resolution mass spectrometry from biological samples. *Analyst* 137, 21, 4958-4967.
- Creek, D.J., Jankevics, A., Breitling, R., Watson, D. G., Barrett, M. P., and Burgess, K. E. (2011). Toward global metabolomics analysis with hydrophilic interaction liquid chromatography-mass spectrometry: improved metabolite identification by retention time prediction. *Anal.Chem.* 83, 22, 8703-8710.
- Croixmarie, V., Umbdenstock, T., Cloarec, O., Moreau, A., Pascussi, J. M., Boursier-Neyret, C., and Walther, B. (2009). Integrated comparison of drug-related and drug-induced ultra performance liquid chromatography/mass spectrometry metabolomic profiles using human hepatocyte cultures. *Anal.Chem.* 81, 15, 6061-6069.
- Cubbon, S., Antonio, C., Wilson, J., and Thomas-Oates, J. (2010). Metabolomic applications of HILIC-LC-MS. *Mass Spectrom.Rev.* 29, 5, 671-684.
- D'Alessandro, A., Gevi, F., and Zolla, L. (2013). Red blood cell metabolism under prolonged anaerobic storage. *Mol.Biosyst.* 9, 6, 1196-1209.
- Dai, J. and Carr, P. W. (2005). Role of ion pairing in anionic additive effects on the separation of cationic drugs in reversed-phase liquid chromatography. *J.Chromatogr.A* 1072, 2, 169-184.
- Dai, J. and Carr, P. W. (2009). Effect of mobile phase anionic additives on selectivity, efficiency, and sample loading capacity of cationic drugs in reversed-phase liquid chromatography. *J.Chromatogr.A* 1216, 39, 6695-6705.
- Dai, J., Mendonsa, S. D., Bowser, M. T., Lucy, C. A., and Carr, P. W. (2005). Effect of anionic additive type on ion pair formation constants of basic pharmaceuticals. *J.Chromatogr.A* 1069, 2, 225-234.
- Daimon, M., Soga, T., Hozawa, A., Oizumi, T., Kaino, W., Takase, K., Karasawa, S., Jimbu, Y., Wada, K., Kameda, W., Susa, S., Kayama, T., Saito, K., Tomita, M., and Kato, T. (2012). Serum glycerophosphate levels are increased in Japanese men with type 2 diabetes. *Intern.Med.* 51, 6, 545-551.
- Dalmas, E., Venteclef, N., Caer, C., Poitou, C., Cremer, I., ron-Wisnewsky, J., Lacroix-Desmazes, S., Bayry, J., Kaveri, S. V., Clement, K., Andre, S., and Guerre-Millo, M. (2014). T Cell-Derived IL-

- 22 Amplifies IL-1beta-Driven Inflammation in Human Adipose Tissue: Relevance to Obesity and Type 2 Diabetes. *Diabetes* 63, 6, 1966-1977.
- Dang, L., White, D. W., Gross, S., Bennett, B. D., Bittinger, M. A., Driggers, E. M., Fantin, V. R., Jang, H. G., Jin, S., Keenan, M. C., Marks, K. M., Prins, R. M., Ward, P. S., Yen, K. E., Liao, L. M., Rabinowitz, J. D., Cantley, L. C., Thompson, C. B., Vander Heiden, M. G., and Su, S. M. (2009). Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* 462, 7274, 739-744.
- Darghouth, D., Koehl, B., Heilier, J. F., Madalinski, G., Bovee, P., Bosman, G., Delaunay, J., Junot, C., and Romeo, P. H. (2011a). Alterations of red blood cell metabolome in overhydrated hereditary stomatocytosis. *Haematologica* 96, 12, 1861-1865.
- Darghouth, D., Koehl, B., Madalinski, G., Heilier, J. F., Bovee, P., Xu, Y., Olivier, M. F., Bartolucci, P., Benkerrou, M., Pissard, S., Colin, Y., Galacteros, F., Bosman, G., Junot, C., and Romeo, P. H. (2011b). Pathophysiology of sickle cell disease is mirrored by the red blood cell metabolome. *Blood* 117, 6, e57-e66.
- Daykin, C.A., Foxall, P. J., Connor, S. C., Lindon, J. C., and Nicholson, J. K. (2002). The comparison of plasma deproteinization methods for the detection of low-molecular-weight metabolites by (1)H nuclear magnetic resonance spectroscopy. *Anal.Biochem.* 304, 2, 220-230.
- Delobel, A., Roy, S., Touboul, D., Gaudin, K., Germain, D. P., Baillet, A., Brion, F., Prognon, P., Chaminade, P., and Laprevote, O. (2006). Atmospheric pressure photoionization coupled to porous graphitic carbon liquid chromatography for the analysis of globotriaosylceramides. Application to Fabry disease. *J.Mass Spectrom.* 41, 1, 50-58.
- Delobel, A., Touboul, D., and Laprevote, O. (2005). Structural characterization of phosphatidylcholines by atmospheric pressure photoionization mass spectrometry. *Eur.J.Mass Spectrom.(Chichester, Eng)* 11, 4, 409-417.
- DeLong, C.J., Baker, P. R., Samuel, M., Cui, Z., and Thomas, M. J. (2001). Molecular species composition of rat liver phospholipids by ESI-MS/MS: the effect of chromatography. *J.Lipid Res.* 42, 12, 1959-1968.
- DeLong, C.J., Shen, Y. J., Thomas, M. J., and Cui, Z. (1999). Molecular distinction of phosphatidylcholine synthesis between the CDP-choline pathway and phosphatidylethanolamine methylation pathway. *J.Biol.Chem.* 274, 42, 29683-29688.
- Denery, J.R., Nunes, A. A., and Dickerson, T. J. (2011). Characterization of differences between blood sample matrices in untargeted metabolomics. *Anal.Chem.* 83, 3, 1040-1047.
- Dennis, E.A. (2009). Lipidomics joins the omics evolution. *Proc.Natl.Acad.Sci.U.S.A* 106, 7, 2089-2090.
- Dettmer, K., Almstetter, M. F., Appel, I. J., Nurnberger, N., Schlamberger, G., Gronwald, W., Meyer, H. H., and Oefner, P. J. (2010). Comparison of serum versus plasma collection in gas chromatography--mass spectrometry-based metabolomics. *Electrophoresis* 31, 14, 2365-2373.
- Dettmer, K., Aronov, P. A., and Hammock, B. D. (2007). Mass spectrometry-based metabolomics. *Mass Spectrom.Rev.* 26, 1, 51-78.
- Dettmer, K. and Hammock, B. D. (2004). Metabolomics--a new exciting field within the "omics" sciences. *Environ.Health Perspect.* 112, 7, A396-A397.

- Deventer, K., Pozo, O. J., Van, E. P., and Delbeke, F. T. (2009). Qualitative detection of diuretics and acidic metabolites of other doping agents in human urine by high-performance liquid chromatography-tandem mass spectrometry: comparison between liquid-liquid extraction and direct injection. *J.Chromatogr.A* 1216, 31, 5819-5827.
- Dill, A.L., Eberlin, L. S., Zheng, C., Costa, A. B., Ifa, D. R., Cheng, L., Masterson, T. A., Koch, M. O., Vitek, O., and Cooks, R. G. (2010). Multivariate statistical differentiation of renal cell carcinomas based on lipidomic analysis by ambient ionization imaging mass spectrometry. *Anal.Bioanal.Chem.* 398, 7-8, 2969-2978.
- Dill, A.L., Ifa, D. R., Manicke, N. E., Costa, A. B., Ramos-Vara, J. A., Knapp, D. W., and Cooks, R. G. (2009). Lipid profiles of canine invasive transitional cell carcinoma of the urinary bladder and adjacent normal tissue by desorption electrospray ionization imaging mass spectrometry. *Anal.Chem.* 81, 21, 8758-8764.
- Duarte, N.C., Becker, S. A., Jamshidi, N., Thiele, I., Mo, M. L., Vo, T. D., Srivas, R., and Palsson, B. O. (2007). Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc.Natl.Acad.Sci.U.S.A* 104, 6, 1777-1782.
- Dumas, M.E., Maibaum, E. C., Teague, C., Ueshima, H., Zhou, B., Lindon, J. C., Nicholson, J. K., Stamler, J., Elliott, P., Chan, Q., and Holmes, E. (2006). Assessment of analytical reproducibility of ¹H NMR spectroscopy based metabonomics for large-scale epidemiological research: the INTERMAP Study. *Anal.Chem.* 78, 7, 2199-2208.
- Dunn, W.B., Bailey, N. J., and Johnson, H. E. (2005). Measuring the metabolome: current analytical technologies. *Analyst* 130, 5, 606-625.
- Dunn, W.B., Broadhurst, D., Begley, P., Zelena, E., Francis-McIntyre, S., Anderson, N., Brown, M., Knowles, J. D., Halsall, A., Haselden, J. N., Nicholls, A. W., Wilson, I. D., Kell, D. B., and Goodacre, R. (2011). Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat.Protoc.* 6, 7, 1060-1083.
- Dwivedi, P., Schultz, A. J., and Hill, H. H. (2010). Metabolic Profiling of Human Blood by High Resolution Ion Mobility Mass Spectrometry (IM-MS). *Int.J.Mass Spectrom.* 298, 1-3, 78-90.
- Eberlin, L.S., Norton, I., Dill, A. L., Golby, A. J., Ligon, K. L., Santagata, S., Cooks, R. G., and Agar, N. Y. (2012). Classifying human brain tumors by lipid imaging with mass spectrometry. *Cancer Res.* 72, 3, 645-654.
- Edmond de Hoffmann and Vincent Stroobant (2007). *Mass Spectrometry: Principles and Applications*, 3rd Edition. 3,
- Eikel, D., Vavrek, M., Smith, S., Bason, C., Yeh, S., Korfmacher, W. A., and Henion, J. D. (2011). Liquid extraction surface analysis mass spectrometry (LESA-MS) as a novel profiling tool for drug distribution and metabolism analysis: the terfenadine example. *Rapid Commun.Mass Spectrom.* 25, 23, 3587-3596.
- Ejsing, C.S., Duchoslav, E., Sampaio, J., Simons, K., Bonner, R., Thiele, C., Ekroos, K., and Shevchenko, A. (2006). Automated identification and quantification of glycerophospholipid molecular species by multiple precursor ion scanning. *Anal.Chem.* 78, 17, 6202-6214.

- Ejsing, C.S., Sampaio, J. L., Surendranath, V., Duchoslav, E., Ekroos, K., Klemm, R. W., Simons, K., and Shevchenko, A. (2009). Global analysis of the yeast lipidome by quantitative shotgun mass spectrometry. *Proc.Natl.Acad.Sci.U.S.A* 106, 7, 2136-2141.
- Ekroos, K., Chernushevich, I. V., Simons, K., and Shevchenko, A. (2002). Quantitative profiling of phospholipids by multiple precursor ion scanning on a hybrid quadrupole time-of-flight mass spectrometer. *Anal.Chem.* 74, 5, 941-949.
- Ellestad-Sayed, J.J., Nelson, R. A., Adson, M. A., Palmer, W. M., and Soule, E. H. (1976). Pantothenic acid, coenzyme A, and human chronic ulcerative and granulomatous colitis. *Am.J.Clin.Nutr.* 29, 12, 1333-1338.
- Euerby, M.R. and Petersson, P. (2003). Chromatographic classification and comparison of commercially available reversed-phase liquid chromatographic columns using principal component analysis. *J.Chromatogr.A* 994, 1-2, 13-36.
- Evans, A.M., DeHaven, C. D., Barrett, T., Mitchell, M., and Milgram, E. (2009). Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. *Anal.Chem.* 81, 16, 6656-6667.
- Fahy, E., Cotter, D., Byrnes, R., Sud, M., Maer, A., Li, J., Nadeau, D., Zhau, Y., and Subramaniam, S. (2007a). Bioinformatics for lipidomics. *Methods Enzymol.* 432, 247-273.
- Fahy, E., Subramaniam, S., Brown, H. A., Glass, C. K., Merrill, A. H., Jr., Murphy, R. C., Raetz, C. R., Russell, D. W., Seyama, Y., Shaw, W., Shimizu, T., Spener, F., van, M. G., VanNieuwenhze, M. S., White, S. H., Witztum, J. L., and Dennis, E. A. (2005). A comprehensive classification system for lipids. *J.Lipid Res.* 46, 5, 839-861.
- Fahy, E., Subramaniam, S., Murphy, R. C., Nishijima, M., Raetz, C. R., Shimizu, T., Spener, F., van, M. G., Wakelam, M. J., and Dennis, E. A. (2009). Update of the LIPID MAPS comprehensive classification system for lipids. *J.Lipid Res.* 50 Suppl, S9-14.
- Fahy, E., Sud, M., Cotter, D., and Subramaniam, S. (2007b). LIPID MAPS online tools for lipid research. *Nucleic Acids Res.* 35, Web Server issue, W606-W612.
- Fauland, A., Trotsmuller, M., Eberl, A., fiuni-Zadeh, S., Kofeler, H., Guo, X., and Lankmayr, E. (2013). An improved SPE method for fractionation and identification of phospholipids. *J.Sep.Sci.* 36, 4, 744-751.
- Feist, A.M., Henry, C. S., Reed, J. L., Krummenacker, M., Joyce, A. R., Karp, P. D., Broadbelt, L. J., Hatzimanikatis, V., and Palsson, B. O. (2007). A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol.Syst.Biol.* 3, 121-
- Feist, A.M., Herrgard, M. J., Thiele, I., Reed, J. L., and Palsson, B. O. (2009). Reconstruction of biochemical networks in microorganisms. *Nat.Rev.Microbiol.* 7, 2, 129-143.
- Fenn, J.B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. *Science* 246, 4926, 64-71.
- Fiehn, O. (2001). Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comp Funct.Genomics* 2, 3, 155-168.

- Fiehn, O. (2002). Metabolomics--the link between genotypes and phenotypes. *Plant Mol.Biol.* 48, 1-2, 155-171.
- Folch, J., ASCOLI, I., LEES, M., MEATH, J. A., and LeBARON, N. (1951). Preparation of lipide extracts from brain tissue. *J Biol Chem* 191, 2, 833-841.
- Fuhrer, T., Heer, D., Begemann, B., and Zamboni, N. (2011). High-throughput, accurate mass metabolome profiling of cellular extracts by flow injection-time-of-flight mass spectrometry. *Anal.Chem.* 83, 18, 7074-7080.
- Gallien, S., Duriez, E., Crone, C., Kellmann, M., Moehring, T., and Domon, B. (2012). Targeted proteomic quantification on quadrupole-orbitrap mass spectrometer. *Mol.Cell Proteomics.* 11, 12, 1709-1723.
- Gao, X., Zhang, Q., Meng, D., Isaac, G., Zhao, R., Fillmore, T. L., Chu, R. K., Zhou, J., Tang, K., Hu, Z., Moore, R. J., Smith, R. D., Katze, M. G., and Metz, T. O. (2012). A reversed-phase capillary ultra-performance liquid chromatography-mass spectrometry (UPLC-MS) method for comprehensive top-down/bottom-up lipid profiling. *Anal.Bioanal.Chem.* 402, 9, 2923-2933.
- Garcia-Villalba, R., Carrasco-Pancorbo, A., Oliveras-Ferraros, C., Vazquez-Martin, A., Menendez, J. A., Segura-Carretero, A., and Fernandez-Gutierrez, A. (2010). Characterization and quantification of phenolic compounds of extra-virgin olive oils with anticancer properties by a rapid and resolutive LC-ESI-TOF MS method. *J Pharm.Biomed.Anal.* 51, 2, 416-429.
- Gevorgyan, A., Poolman, M. G., and Fell, D. A. (2008). Detection of stoichiometric inconsistencies in biomolecular models. *Bioinformatics.* 24, 19, 2245-2251.
- Giavalisco, P., Kohl, K., Hummel, J., Seiwert, B., and Willmitzer, L. (2009). ¹³C isotope-labeled metabolomes allowing for improved compound annotation and relative quantification in liquid chromatography-mass spectrometry-based metabolomic research. *Anal.Chem.* 81, 15, 6546-6551.
- Gidden, J., Denson, J., Liyanage, R., Ivey, D. M., and Lay, J. O. (2009). Lipid Compositions in *Escherichia coli* and *Bacillus subtilis* During Growth as Determined by MALDI-TOF and TOF/TOF Mass Spectrometry. *Int.J.Mass Spectrom.* 283, 1-3, 178-184.
- Gika, H., Theodoridis, G., Mattivi, F., Vrhovsek, U., and Pappa-Louisi, A. (2012b). Hydrophilic interaction ultra performance liquid chromatography retention prediction under gradient elution. *Anal.Bioanal.Chem.* 404, 3, 701-709.
- Gika, H., Theodoridis, G., Mattivi, F., Vrhovsek, U., and Pappa-Louisi, A. (2012a). Retention prediction of a set of amino acids under gradient elution conditions in hydrophilic interaction liquid chromatography. *J.Sep.Sci.* 35, 3, 376-383.
- Gika, H.G., Theodoridis, G. A., Vrhovsek, U., and Mattivi, F. (2012c). Quantitative profiling of polar primary metabolites using hydrophilic interaction ultrahigh performance liquid chromatography-tandem mass spectrometry. *J.Chromatogr.A* 1259, 121-127.
- Gika, H.G., Theodoridis, G. A., and Wilson, I. D. (2008). Liquid chromatography and ultra-performance liquid chromatography-mass spectrometry fingerprinting of human urine: sample stability under different handling and storage conditions for metabonomics studies. *J.Chromatogr.A* 1189, 1-2, 314-322.

- Gika, H.G., Theodoridis, G. A., Wingate, J. E., and Wilson, I. D. (2007). Within-day reproducibility of an HPLC-MS-based method for metabolomic analysis: application to human urine. *J.Proteome.Res.* 6, 8, 3291-3303.
- Gilbert-Lopez, B., Garcia-Reyes, J. F., Fernandez-Alba, A. R., and Molina-Diaz, A. (2010). Evaluation of two sample treatment methodologies for large-scale pesticide residue analysis in olive oil by fast liquid chromatography-electrospray mass spectrometry. *J.Chromatogr.A* 1217, 24, 3736-3747.
- Gillet, L.C., Navarro, P., Tate, S., Rost, H., Selevsek, N., Reiter, L., Bonner, R., and Aebersold, R. (2012). Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol.Cell Proteomics.* 11, 6, O111-
- Godat, E., Madalinski, G., Muller, L., Heilier, J. F., Labarre, J., and Junot, C. (2010). Mass spectrometry-based methods for the determination of sulfur and related metabolite concentrations in cell extracts. *Methods Enzymol.* 473, 41-76.
- Gorden, D.L., Ivanova, P. T., Myers, D. S., McIntyre, J. O., VanSaun, M. N., Wright, J. K., Matrisian, L. M., and Brown, H. A. (2011). Increased diacylglycerols characterize hepatic lipid changes in progression of human nonalcoholic fatty liver disease; comparison to a murine model. *PLoS.One.* 6, 8, e22775-
- Gregory, K.E., Bird, S. S., Gross, V. S., Marur, V. R., Lazarev, A. V., Walker, W. A., and Kristal, B. S. (2013). Method development for fecal lipidomics profiling. *Anal.Chem.* 85, 2, 1114-1123.
- Griffiths, W.J., Koal, T., Wang, Y., Kohl, M., Enot, D. P., and Deigner, H. P. (2010). Targeted metabolomics for biomarker discovery. *Angew Chem Int Ed Engl* 49, 32, 5426-5445.
- Guittard, J., Hronowski, X. L., and Costello, C. E. (1999). Direct matrix-assisted laser desorption/ionization mass spectrometric analysis of glycosphingolipids on thin layer chromatographic plates and transfer membranes. *Rapid Commun.Mass Spectrom.* 13, 18, 1838-1849.
- Guo, K., Ji, C., and Li, L. (2007). Stable-isotope dimethylation labeling combined with LC-ESI MS for quantification of amine-containing metabolites in biological samples. *Anal.Chem.* 79, 22, 8631-8638.
- Guo, K. and Li, L. (2009). Differential ¹²C-/¹³C-isotope dansylation labeling and fast liquid chromatography/mass spectrometry for absolute and relative quantification of the metabolome. *Anal.Chem.* 81, 10, 3919-3932.
- Guo, K. and Li, L. (2010). High-performance isotope labeling for profiling carboxylic acid-containing metabolites in biofluids by mass spectrometry. *Anal.Chem.* 82, 21, 8789-8793.
- Gustavsson, S.A., Samskog, J., Markides, K. E., and Langstrom, B. (2001). Studies of signal suppression in liquid chromatography-electrospray ionization mass spectrometry using volatile ion-pairing reagents. *J.Chromatogr.A* 937, 1-2, 41-47.
- Gutnikov, G. (1995). Fatty acid profiles of lipid samples. *J.Chromatogr.B Biomed.Appl.* 671, 1-2, 71-89.
- Haber, M.H. (1988). Pisse prophecy: a brief history of urinalysis. *Clin.Lab Med.* 8, 3, 415-430.

- Hadadi, N., Cher, S. K., Seijo, M., Zisaki, A., Guan, X., Wenk, M. R., and Hatzimanikatis, V. (2014). A computational framework for integration of lipidomics data into metabolic pathways. *Metab Eng* 23C, 1-8.
- Haimi, P., Uphoff, A., Hermansson, M., and Somerharju, P. (2006). Software tools for analysis of mass spectrometric lipidome data. *Anal.Chem.* 78, 24, 8324-8331.
- Han, J., Datla, R., Chan, S., and Borchers, C. H. (2009). Mass spectrometry-based technologies for high-throughput metabolomics. *Bioanalysis* 1, 9, 1665-1684.
- Han, X. and Gross, R. W. (1994). Electrospray ionization mass spectroscopic analysis of human erythrocyte plasma membrane phospholipids. *Proc.Natl.Acad.Sci.U.S.A* 91, 22, 10635-10639.
- Han, X. and Gross, R. W. (2005). Shotgun lipidomics: electrospray ionization mass spectrometric analysis and quantitation of cellular lipidomes directly from crude extracts of biological samples. *Mass Spectrom.Rev.* 24, 3, 367-412.
- Han, X., Yang, K., and Gross, R. W. (2012). Multi-dimensional mass spectrometry-based shotgun lipidomics and novel strategies for lipidomic analyses. *Mass Spectrom.Rev.* 31, 1, 134-178.
- Han, X., Yang, K., Yang, J., Cheng, H., and Gross, R. W. (2006). Shotgun lipidomics of cardiolipin molecular species in lipid extracts of biological samples. *J.Lipid Res.* 47, 4, 864-879.
- Han, X.L. and Gross, R. W. (1991). Alterations in membrane dynamics elicited by amphiphilic compounds are augmented in plasmemylcholine bilayers. *Biochim.Biophys.Acta* 1069, 1, 37-45.
- Handa, S., Kushi, Y., Kambara, H., and Shizukuishi, K. (1983). Secondary ion mass spectra of neutral sphingoglycolipids. *J.Biochem.* 93, 1, 315-318.
- Harrison, K.A., Clay, K. L., and Murphy, R. C. (1999). Negative ion electrospray and tandem mass spectrometric analysis of platelet activating factor (PAF) (1-hexadecyl-2-acetyl-glycerophosphocholine). *J.Mass Spectrom.* 34, 4, 330-335.
- Hartler, J., Tharakan, R., Kofeler, H. C., Graham, D. R., and Thallinger, G. G. (2013). Bioinformatics tools and challenges in structural analysis of lipidomics MS/MS data. *Brief.Bioinform.* 14, 3, 375-390.
- Hartonen, M., Mattila, I., Ruskeepaa, A. L., Oresic, M., and Hyotylainen, T. (2013). Characterization of cerebrospinal fluid by comprehensive two-dimensional gas chromatography coupled to time-of-flight mass spectrometry. *J.Chromatogr.A* 1293, 142-149.
- Haufroid, V., Jakubowski, M., Janasik, B., Ligocka, D., Buchet, J. P., Bergamaschi, E., Manini, P., Mutti, A., Ghittori, S., Arand, M., Hangen, N., Oesch, F., Hirvonen, A., and Lison, D. (2002). Interest of genotyping and phenotyping of drug-metabolizing enzymes for the interpretation of biological monitoring of exposure to styrene. *Pharmacogenetics* 12, 9, 691-702.
- Havlikova, L., Matysova, L., Hajkova, R., Satinsky, D., and Solich, P. (2008). Advantages of pentafluorophenylpropyl stationary phase over conventional C18 stationary phase-- application to analysis of triamcinolone acetonide. *Talanta* 76, 3, 597-601.
- Hayasaka, T., Goto-Inoue, N., Sugiura, Y., Zaima, N., Nakanishi, H., Ohishi, K., Nakanishi, S., Naito, T., Taguchi, R., and Setou, M. (2008). Matrix-assisted laser desorption/ionization quadrupole ion trap time-of-flight (MALDI-QIT-TOF)-based imaging mass spectrometry reveals a layered

- distribution of phospholipid molecular species in the mouse retina. *Rapid Commun.Mass Spectrom.* 22, 21, 3415-3426.
- Henderson, S.C., Valentine, S. J., Counterman, A. E., and Clemmer, D. E. (1999). ESI/ion trap/ion mobility/time-of-flight mass spectrometry for rapid and sensitive analysis of biomolecular mixtures. *Anal.Chem.* 71, 2, 291-301.
- Hermansson, M., Uphoff, A., Kakela, R., and Somerharju, P. (2005). Automated quantitative analysis of complex lipidomes by liquid chromatography/mass spectrometry. *Anal.Chem.* 77, 7, 2166-2175.
- Hilvo, M., Denkert, C., Lehtinen, L., Muller, B., Brockmoller, S., Seppanen-Laakso, T., Budczies, J., Bucher, E., Yetukuri, L., Castillo, S., Berg, E., Nygren, H., Sysi-Aho, M., Griffin, J. L., Fiehn, O., Loibl, S., Richter-Ehrenstein, C., Radke, C., Hyotylainen, T., Kallioniemi, O., Iljin, K., and Oresic, M. (2011). Novel theranostic opportunities offered by characterization of altered membrane lipid metabolism in breast cancer progression. *Cancer Res.* 71, 9, 3236-3245.
- Holmes, E., Loo, R. L., Cloarec, O., Coen, M., Tang, H., Maibaum, E., Bruce, S., Chan, Q., Elliott, P., Stamler, J., Wilson, I. D., Lindon, J. C., and Nicholson, J. K. (2007). Detection of urinary drug metabolite (xenometabolome) signatures in molecular epidemiology studies via statistical total correlation (NMR) spectroscopy. *Anal.Chem* 79, 7, 2629-2640.
- Holmes, E., Loo, R. L., Stamler, J., Bictash, M., Yap, I. K., Chan, Q., Ebbels, T., De, I. M., Brown, I. J., Veselkov, K. A., Daviglus, M. L., Kesteloot, H., Ueshima, H., Zhao, L., Nicholson, J. K., and Elliott, P. (2008). Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature* 453, 7193, 396-400.
- Hopfgartner, G., Tonoli, D., and Varesio, E. (2012). High-resolution mass spectrometry for integrated qualitative and quantitative analysis of pharmaceuticals in biological matrices. *Anal.Bioanal.Chem.* 402, 8, 2587-2596.
- Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K., Ojima, Y., Tanaka, K., Tanaka, S., Aoshima, K., Oda, Y., Kakazu, Y., Kusano, M., Tohge, T., Matsuda, F., Sawada, Y., Hirai, M. Y., Nakanishi, H., Ikeda, K., Akimoto, N., Maoka, T., Takahashi, H., Ara, T., Sakurai, N., Suzuki, H., Shibata, D., Neumann, S., Iida, T., Tanaka, K., Funatsu, K., Matsuura, F., Soga, T., Taguchi, R., Saito, K., and Nishioka, T. (2010). MassBank: a public repository for sharing mass spectral data for life sciences. *J.Mass Spectrom.* 45, 7, 703-714.
- Horai, H., Suwa, K., Arita, M., Nihei, Y., and Nishioka, T. (2007). Mass Spectral Database for Metabolome Analysis. 55th ASMS Conference on Mass Spectrometry and Allied Topics
- Horning, E.C. and Horning, M. G. (1971). Metabolic profiles: gas-phase methods for analysis of metabolites. *Clin.Chem.* 17, 8, 802-809.
- Horning, E.C., Horning, M. G., Carroll, D. I., Dzidic, I., and Stillwell, R. N. (1973). Chemical ionization mass spectrometry. *Adv Biochem Psychopharmacol.* 7, 15-31.
- Hsu, F.F., Bohrer, A., and Turk, J. (1998). Electrospray ionization tandem mass spectrometric analysis of sulfatide. Determination of fragmentation patterns and characterization of molecular species expressed in brain and in pancreatic islets. *Biochim.Biophys.Acta* 1392, 2-3, 202-216.
- Hsu, F.F. and Turk, J. (2009). Electrospray ionization with low-energy collisionally activated dissociation tandem mass spectrometry of glycerophospholipids: mechanisms of

fragmentation and structural characterization. *J.Chromatogr.B Analyt.Technol.Biomed.Life Sci.* 877, 26, 2673-2695.

Hsu, F.F. and Turk, J. (2008a). Elucidation of the double-bond position of long-chain unsaturated fatty acids by multiple-stage linear ion-trap mass spectrometry with electrospray ionization. *J.Am.Soc.Mass Spectrom.* 19, 11, 1673-1680.

Hsu, F.F. and Turk, J. (2008b). Structural characterization of unsaturated glycerophospholipids by multiple-stage linear ion-trap mass spectrometry with electrospray ionization. *J.Am.Soc.Mass Spectrom.* 19, 11, 1681-1691.

Hsu, F.F. and Turk, J. (2000). Charge-driven fragmentation processes in diacyl glycerophosphatidic acids upon low-energy collisional activation. A mechanistic proposal. *J.Am.Soc.Mass Spectrom.* 11, 9, 797-803.

Hu, C., van, D. J., van der, H. R., Spijksma, G., Reijmers, T. H., Wang, M., Slee, E., Lu, X., Xu, G., van der, G. J., and Hankemeier, T. (2008). RPLC-ion-trap-FTMS method for lipid profiling of plasma: method validation and application to p53 mutant mouse model. *J.Proteome.Res.* 7, 11, 4982-4991.

Huang, X., Chen, Y. J., Cho, K., Nikolskiy, I., Crawford, P. A., and Patti, G. J. (2014). X13CMS: global tracking of isotopic labels in untargeted metabolomics. *Anal.Chem.* 86, 3, 1632-1639.

Hubner, G., Crone, C., and Lindner, B. (2009). lipID--a software tool for automated assignment of lipids in mass spectra. *J.Mass Spectrom.* 44, 12, 1676-1683.

Ichou F, Schwarzenberg A, esage D, Ives S, unot C, achuron-Mandard X, and abet JC (2014). Comparison of the activation time effects and the internal energy distributions for the CID, PQD and HCD excitation modes. *J Mass Spectrom*

Ichou, F., Lesage, D., huron-Mandard, X., Junot, C., Cole, R. B., and Tabet, J. C. (2013). Collision cell pressure effect on CID spectra pattern using triple quadrupole instruments: a RRKM modeling. *J.Mass Spectrom.* 48, 2, 179-186.

Idborg, H., Zamani, L., Edlund, P. O., Schuppe-Koistinen, I., and Jacobsson, S. P. (2005). Metabolic fingerprinting of rat urine by LC/MS Part 1. Analysis by hydrophilic interaction liquid chromatography-electrospray ionization mass spectrometry. *J.Chromatogr.B* 828, 1-2, 9-13.

Idborg-Bjorkman, H., Edlund, P. O., Kvalheim, O. M., Schuppe-Koistinen, I., and Jacobsson, S. P. (2003). Screening of biomarkers in rat urine using LC/electrospray ionization-MS and two-way data analysis. *Anal.Chem.* 75, 18, 4784-4792.

Ikeda, K. and Taguchi, R. (2010). Highly sensitive localization analysis of gangliosides and sulfatides including structural isomers in mouse cerebellum sections by combination of laser microdissection and hydrophilic interaction liquid chromatography/electrospray ionization mass spectrometry with theoretically expanded multiple reaction monitoring. *Rapid Commun.Mass Spectrom.* 24, 20, 2957-2965.

Iles, R.A., Snodgrass, G. J., Chalmers, R. A., and Stacey, T. E. (1984). Rapid screening of metabolic diseases by proton NMR. *Lancet* 2, 8413, 1221-1222.

Imbert, L., Gaudin, M., Libong, D., Touboul, D., Abreu, S., Loiseau, P. M., Laprevote, O., and Chaminade, P. (2012a). Comparison of electrospray ionization, atmospheric pressure

- chemical ionization and atmospheric pressure photoionization for a lipidomic analysis of *Leishmania donovani*. *J.Chromatogr.A* 1242, 75-83.
- Imbert, L., Ramos, R. G., Libong, D., Abreu, S., Loiseau, P. M., and Chaminade, P. (2012b). Identification of phospholipid species affected by miltefosine action in *Leishmania donovani* cultures using LC-ELSD, LC-ESI/MS, and multivariate data analysis. *Anal.Bioanal.Chem.* 402, 3, 1169-1182.
- Ipsen, A., Want, E. J., Lindon, J. C., and Ebbels, T. M. (2010). A statistically rigorous test for the identification of parent-fragment pairs in LC-MS datasets. *Anal.Chem.* 82, 5, 1766-1778.
- Ishikawa, M., Maekawa, K., Saito, K., Senoo, Y., Urata, M., Murayama, M., Tajima, Y., Kumagai, Y., and Saito, Y. (2014). Plasma and serum lipidomics of healthy white adults shows characteristic profiles by subjects' gender and age. *PLoS.One.* 9, 3, e91806-
- Ishikawa, M., Tajima, Y., Murayama, M., Senoo, Y., Maekawa, K., and Saito, Y. (2013). Plasma and serum from nonfasting men and women differ in their lipidomic profiles. *Biol.Pharm.Bull.* 36, 4, 682-685.
- Ito, T., Odake, T., Katoh, H., Yamaguchi, Y., and Aoki, M. (2011). High-throughput profiling of microbial extracts. *J.Nat.Prod.* 74, 5, 983-988.
- Ivanova, P.T., Cerda, B. A., Horn, D. M., Cohen, J. S., McLafferty, F. W., and Brown, H. A. (2001). Electrospray ionization mass spectrometry analysis of changes in phospholipids in RBL-2H3 mastocytoma cells during degranulation. *Proc.Natl.Acad.Sci.U.S.A* 98, 13, 7152-7157.
- Jackson, S.N., Ugarov, M., Post, J. D., Egan, T., Langlais, D., Schultz, J. A., and Woods, A. S. (2008). A study of phospholipids by ion mobility TOFMS. *J.Am.Soc.Mass Spectrom.* 19, 11, 1655-1662.
- Jackson, S.N., Wang, H. Y., Woods, A. S., Ugarov, M., Egan, T., and Schultz, J. A. (2005). Direct tissue analysis of phospholipids in rat brain using MALDI-TOFMS and MALDI-ion mobility-TOFMS. *J.Am.Soc.Mass Spectrom.* 16, 2, 133-138.
- Jaffe, J.D., Keshishian, H., Chang, B., Addona, T. A., Gillette, M. A., and Carr, S. A. (2008). Accurate inclusion mass screening: a bridge from unbiased discovery to targeted assay development for biomarker verification. *Mol.Cell Proteomics.* 7, 10, 1952-1962.
- Ji, H., Morre, J., Voinov, V. G., Deinzer, M. L., and Barofsky, D. F. (2007a). Charge-remote metastable ion decomposition of free fatty acids under FAB MS: evidence for biradical ion structures. *Anal.Chem.* 79, 7, 2822-2826.
- Ji, H., Voinov, V. G., Deinzer, M. L., and Barofsky, D. F. (2007b). Distinguishing between cis/trans isomers of monounsaturated fatty acids by FAB MS. *Anal.Chem.* 79, 4, 1519-1522.
- Jiang, M., Chen, T., Feng, H., Zhang, Y., Li, L., Zhao, A., Niu, X., Liang, F., Wang, M., Zhan, J., Lu, C., He, X., Xiao, L., Jia, W., and Lu, A. (2013). Serum metabolic signatures of four types of human arthritis. *J.Proteome.Res.* 12, 8, 3769-3779.
- Jonasdottir, H.S., Nicolardi, S., Jonker, W., Derks, R., Palmblad, M., Ioan-Facsinay, A., Toes, R., van der Burgt, Y. E., Deelder, A. M., Mayboroda, O. A., and Giera, M. (2013). Detection and structural elucidation of esterified oxylipids in human synovial fluid by electrospray ionization-fourier transform ion-cyclotron mass spectrometry and liquid chromatography-ion trap-MS(3): detection of esterified hydroxylated docosapentaenoic acid containing phospholipids. *Anal.Chem.* 85, 12, 6003-6010.

- Jourdan, F., Breitling, R., Barrett, M. P., and Gilbert, D. (2008). MetaNetter: inference and visualization of high-resolution metabolomic networks. *Bioinformatics*. 24, 1, 143-145.
- Jourdan, F., Cottret, L., Huc, L., Wildridge, D., Scheltema, R., Hillenweck, A., Barrett, M. P., Zalko, D., Watson, D. G., and Debrauwer, L. (2010). Use of reconstituted metabolic networks to assist in metabolomic data visualization and mining. *Metabolomics* 6, 2, 312-321.
- Junot, C., Fenaille, F., Colsch, B., and Becher, F. (2013). High resolution mass spectrometry based techniques at the crossroads of metabolic pathways. *Mass Spectrom.Rev.*
- Junot, C., Madalinski, G., Tabet, J. C., and Ezan, E. (2010). Fourier transform mass spectrometry for metabolome analysis. *Analyst* 135, 9, 2203-2219.
- Kakhniashvili, D.G., Bulla, L. A., Jr., and Goodman, S. R. (2004). The human erythrocyte proteome: analysis by ion trap mass spectrometry. *Mol.Cell Proteomics*. 3, 5, 501-509.
- Kanehisa, M. and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 1, 27-30.
- Karp, P.D., Riley, M., Paley, S. M., and Pellegrini-Toole, A. (2002). The MetaCyc Database. *Nucleic Acids Res.* 30, 1, 59-61.
- Karp, P.D., Riley, M., Saier, M., Paulsen, I. T., Paley, S. M., and Pellegrini-Toole, A. (2000). The EcoCyc and MetaCyc databases. *Nucleic Acids Res.* 28, 1, 56-59.
- Kasper, P.T., Rojas-Cherto, M., Mistrik, R., Reijmers, T., Hankemeier, T., and Vreeken, R. J. (2012). Fragmentation trees for the structural characterisation of metabolites. *Rapid Commun.Mass Spectrom.* 26, 19, 2275-2286.
- Katajamaa, M., Miettinen, J., and Oresic, M. (2006). MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics* 22, 5, 634-636.
- Katajamaa, M. and Oresic, M. (2005). Processing methods for differential analysis of LC/MS profile data. *BMC.Bioinformatics*. 6, 179-
- Katajamaa, M. and Oresic, M. (2007). Data processing for mass spectrometry-based metabolomics. *J Chromatogr.A* 1158, 1-2, 318-328.
- Kehoe, R.A., Thamann, F., and Cholak, J. (1933). Lead absorption and excretion in relation to the diagnosis of lead poisoning. *Journal of Industrial Hygiene and Toxicology* 15, 5, 320-340.
- Kiefer, P., Portais, J. C., and Vorholt, J. A. (2008). Quantitative metabolome analysis using liquid chromatography-high-resolution mass spectrometry. *Anal.Biochem.* 382, 2, 94-100.
- Kim, J.K., Harada, K., Bamba, T., Fukusaki, E., and Kobayashi, A. (2005). Stable isotope dilution-based accurate comparative quantification of nitrogen-containing metabolites in *Arabidopsis thaliana* T87 cells using in vivo (15)N-isotope enrichment. *Biosci.Biotechnol.Biochem.* 69, 7, 1331-1340.
- Kim, T.Y., Sohn, S. B., Kim, Y. B., Kim, W. J., and Lee, S. Y. (2012). Recent advances in reconstruction and applications of genome-scale metabolic models. *Curr.Opin.Biotechnol.* 23, 4, 617-623.
- Kind, T. and Fiehn, O. (2010). Advances in structure elucidation of small molecules using mass spectrometry. *Bioanal.Rev.* 2, 1-4, 23-60.

- Kind, T. and Fiehn, O. (2006). Metabolomic database annotations via query of elemental compositions: mass accuracy is insufficient even at less than 1 ppm. *BMC.Bioinformatics*. 7, 234-
- Kind, T. and Fiehn, O. (2007). Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC.Bioinformatics*. 8, 105-
- Kind, T., Liu, K. H., Lee, d. Y., DeFelice, B., Meissen, J. K., and Fiehn, O. (2013). LipidBlast in silico tandem mass spectrometry database for lipid identification. *Nat.Methods* 10, 8, 755-758.
- Kinsley Narainsamy, Corinne Cassier-Chauvat, Christophe Junot, and Franck Chauvat (2011). High performance analysis of the cyanobacterial metabolism via liquid chromatography coupled to a LTQ-Orbitrap mass spectrometer: evidence that glucose reprograms the whole carbon metabolism and triggers oxidative stress. *Metabolomics*
- Kirkland, J.J. (2004). Development of some stationary phases for reversed-phase high-performance liquid chromatography. *J.Chromatogr.A* 1060, 1-2, 9-21.
- Kishimoto, K., Urade, R., Ogawa, T., and Moriyama, T. (2001). Nondestructive quantification of neutral lipids by thin-layer chromatography and laser-fluorescent scanning: suitable methods for "lipidome" analysis. *Biochem.Biophys.Res.Commun.* 281, 3, 657-662.
- Kitteringham, N.R., Jenkins, R. E., Lane, C. S., Elliott, V. L., and Park, B. K. (2009). Multiple reaction monitoring for quantitative biomarker analysis in proteomics and metabolomics. *J Chromatogr.B Analyt.Technol Biomed.Life Sci* 877, 13, 1229-1239.
- Kliman, M., May, J. C., and McLean, J. A. (2011). Lipid analysis and lipidomics by structurally selective ion mobility-mass spectrometry. *Biochim.Biophys.Acta* 1811, 11, 935-945.
- Koal, T. and Deigner, H. P. (2010). Challenges in mass spectrometry based targeted metabolomics. *Curr.Mol.Med.* 10, 2, 216-226.
- Koeberle, A., Shindou, H., Koeberle, S. C., Laufer, S. A., Shimizu, T., and Werz, O. (2013). Arachidonoyl-phosphatidylcholine oscillates during the cell cycle and counteracts proliferation by suppressing Akt membrane binding. *Proc.Natl.Acad.Sci.U.S.A* 110, 7, 2546-2551.
- Kohl, M., Wiese, S., and Warscheid, B. (2011). Cytoscape: software for visualization and analysis of biological networks. *Methods Mol.Biol.* 696, 291-303.
- Koivusalo, M., Haimi, P., Heikinheimo, L., Kostianen, R., and Somerharju, P. (2001). Quantitative determination of phospholipid compositions by ESI-MS: effects of acyl chain length, unsaturation, and lipid concentration on instrument response. *J.Lipid Res.* 42, 4, 663-672.
- Kolter, T. and Sandhoff, K. (2006). Sphingolipid metabolism diseases. *Biochim.Biophys.Acta* 1758, 12, 2057-2079.
- Kopka, J., Schauer, N., Krueger, S., Birkemeyer, C., Usadel, B., Bergmuller, E., Dormann, P., Weckwerth, W., Gibon, Y., Stitt, M., Willmitzer, L., Fernie, A. R., and Steinhauser, D. (2005). GMD@CSB.DB: the Golm Metabolome Database. *Bioinformatics* 21, 8, 1635-1638.
- Krieger, C.J., Zhang, P., Mueller, L. A., Wang, A., Paley, S., Arnaud, M., Pick, J., Rhee, S. Y., and Karp, P. D. (2004). MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res.* 32, Database issue, D438-D442.

- Kuhl, C., Tautenhahn, R., Bottcher, C., Larson, T. R., and Neumann, S. (2012). CAMERA: An Integrated Strategy for Compound Spectra Extraction and Annotation of Liquid Chromatography/Mass Spectrometry Data Sets. *Anal.Chem* 84, 1, 283-289.
- Kuksis, A., Suomela, J. P., Tarvainen, M., and Kallio, H. (2009). Lipidomic analysis of glycerolipid and cholesteryl ester autooxidation products. *Mol.Biotechnol.* 42, 2, 224-268.
- Kurvinen, J.P., Rua, P., Sjoval, O., and Kallio, H. (2001). Software (MSPECTRA) for automatic interpretation of triacylglycerol molecular mass distribution spectra and collision induced dissociation product ion spectra obtained by ammonia negative ion chemical ionization mass spectrometry. *Rapid Commun.Mass Spectrom.* 15, 13, 1084-1091.
- Lafaye, A., Junot, C., Pereira, Y., Lagniel, G., Tabet, J. C., Ezan, E., and Labarre, J. (2005a). Combined proteome and metabolite-profiling analyses reveal surprising insights into yeast sulfur metabolism. *J.Biol.Chem.* 280, 26, 24723-24730.
- Lafaye, A., Junot, C., Ramounet-Le, G. B., Fritsch, P., Ezan, E., and Tabet, J. C. (2004). Profiling of sulfoconjugates in urine by using precursor ion and neutral loss scans in tandem mass spectrometry. Application to the investigation of heavy metal toxicity in rats. *J Mass Spectrom* 39, 6, 655-664.
- Lafaye, A., Labarre, J., Tabet, J. C., Ezan, E., and Junot, C. (2005b). Liquid chromatography-mass spectrometry and ¹⁵N metabolic labeling for quantitative metabolic profiling. *Anal.Chem.* 77, 7, 2026-2033.
- Lam, S.M. and Shui, G. (2013). Lipidomics as a principal tool for advancing biomedical research. *J.Genet.Genomics* 40, 8, 375-390.
- Lamari, F., Mochel, F., Sedel, F., and Saudubray, J. M. (2013). Disorders of phospholipids, sphingolipids and fatty acids biosynthesis: toward a new category of inherited metabolic diseases. *J.Inherit.Metab Dis.* 36, 3, 411-425.
- Lanekoff, I., Sjoval, P., and Ewing, A. G. (2011). Relative quantification of phospholipid accumulation in the PC12 cell plasma membrane following phospholipid incubation using TOF-SIMS imaging. *Anal.Chem.* 83, 13, 5337-5343.
- Lange, E., Tautenhahn, R., Neumann, S., and Gropl, C. (2008). Critical assessment of alignment procedures for LC-MS proteomics and metabolomics measurements. *BMC.Bioinformatics.* 9, 375-
- Lanza, I.R., Zhang, S., Ward, L. E., Karakelides, H., Raftery, D., and Nair, K. S. (2010). Quantitative metabolomics by H-NMR and LC-MS/MS confirms altered metabolic pathways in diabetes. *PLoS.One.* 5, 5, e10538-
- Larsen, A., Uran, S., Jacobsen, P. B., and Skotland, T. (2001). Collision-induced dissociation of glycerophospholipids using electrospray ion-trap mass spectrometry. *Rapid Commun.Mass Spectrom.* 15, 24, 2393-2398.
- Lawton, K.A., Berger, A., Mitchell, M., Milgram, K. E., Evans, A. M., Guo, L., Hanson, R. W., Kalhan, S. C., Ryals, J. A., and Milburn, M. V. (2008). Analysis of the adult human plasma metabolome. *Pharmacogenomics.* 9, 4, 383-397.
- Leandro, C.C., Hancock, P., Fussell, R. J., and Keely, B. J. (2006). Comparison of ultra-performance liquid chromatography and high-performance liquid chromatography for the determination

- of priority pesticides in baby foods by tandem quadrupole mass spectrometry. *J Chromatogr.A* 1103, 1, 94-101.
- Leavell, M.D. and Leary, J. A. (2006). Fatty acid analysis tool (FAAT): An FT-ICR MS lipid analysis algorithm. *Anal.Chem.* 78, 15, 5497-5503.
- Lesnefsky, E.J., Stoll, M. S., Minkler, P. E., and Hoppel, C. L. (2000). Separation and quantitation of phospholipids and lysophospholipids by high-performance liquid chromatography. *Anal.Biochem.* 285, 2, 246-254.
- Lewis, G.D., Wei, R., Liu, E., Yang, E., Shi, X., Martinovic, M., Farrell, L., Asnani, A., Cyrille, M., Ramanathan, A., Shaham, O., Berriz, G., Lowry, P. A., Palacios, I. F., Tasan, M., Roth, F. P., Min, J., Baumgartner, C., Keshishian, H., Addona, T., Mootha, V. K., Rosenzweig, A., Carr, S. A., Fifer, M. A., Sabatine, M. S., and Gerszten, R. E. (2008). Metabolite profiling of blood from individuals undergoing planned myocardial infarction reveals early markers of myocardial injury. *J Clin Invest* 118, 10, 3503-3512.
- Liebisch, G., Drobnik, W., Lieser, B., and Schmitz, G. (2002). High-throughput quantification of lysophosphatidylcholine by electrospray ionization tandem mass spectrometry. *Clin.Chem.* 48, 12, 2217-2224.
- Lievre, A., Bachet, J. B., Le, C. D., Boige, V., Landi, B., Emile, J. F., Cote, J. F., Tomasic, G., Penna, C., Ducreux, M., Rougier, P., Penault-Llorca, F., and Laurent-Puig, P. (2006). KRAS mutation status is predictive of response to cetuximab therapy in colorectal cancer. *Cancer Res.* 66, 8, 3992-3995.
- Likic, V.A., McConville, M. J., Lithgow, T., and Bacic, A. (2010). Systems biology: the next frontier for bioinformatics. *Adv.Bioinformatics.* 268925-
- Lin, S.M., Zhu, L., Winter, A. Q., Sasinowski, M., and Kibbe, W. A. (2005). What is mzXML good for? *Expert.Rev.Proteomics.* 2, 6, 839-845.
- Liu, X., Plasencia, M., Ragg, S., Valentine, S. J., and Clemmer, D. E. (2004). Development of high throughput dispersive LC-ion mobility-TOFMS techniques for analysing the human plasma proteome. *Brief.Funct.Genomic.Proteomic.* 3, 2, 177-186.
- Lommen, A. (2009). MetAlign: interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing. *Anal.Chem.* 81, 8, 3079-3086.
- Lu, J., Zhou, J., Bao, Y., Chen, T., Zhang, Y., Zhao, A., Qiu, Y., Xie, G., Wang, C., Jia, W., and Jia, W. (2012). Serum metabolic signatures of fulminant type 1 diabetes. *J.Proteome.Res.* 11, 9, 4705-4711.
- Lu, W., Bennett, B. D., and Rabinowitz, J. D. (2008). Analytical strategies for LC-MS-based targeted metabolomics. *J.Chromatogr.B Analyt.Technol.Biomed.Life Sci.* 871, 2, 236-242.
- Madalinski, G., Godat, E., Alves, S., Lesage, D., Genin, E., Levi, P., Labarre, J., Tabet, J. C., Ezan, E., and Junot, C. (2008). Direct introduction of biological samples into a LTQ-Orbitrap hybrid mass spectrometer as a tool for fast metabolome analysis. *Anal.Chem.* 80, 9, 3291-3303.
- Mahrous, E.A., Lee, R. B., and Lee, R. E. (2008). A rapid approach to lipid profiling of mycobacteria using 2D HSQC NMR maps. *J.Lipid Res.* 49, 2, 455-463.

- Mak, T.D., Laiakis, E. C., Goudarzi, M., and Fornace, A. J., Jr. (2014). MetaboLyzer: a novel statistical workflow for analyzing Postprocessed LC-MS metabolomics data. *Anal.Chem.* 86, 1, 506-513.
- Manicke, N.E., Wiseman, J. M., Ifa, D. R., and Cooks, R. G. (2008). Desorption electrospray ionization (DESI) mass spectrometry and tandem mass spectrometry (MS/MS) of phospholipids and sphingolipids: ionization, adduct formation, and fragmentation. *J.Am.Soc.Mass Spectrom.* 19, 4, 531-543.
- Marchi, I., Rudaz, S., and Veuthey, J. L. (2009). Atmospheric pressure photoionization for coupling liquid-chromatography to mass spectrometry: a review. *Talanta* 78, 1, 1-18.
- Margaria, T., Kubczak, C., and Steffen, B. (2008). Bio-jETI: a service integration, design, and provisioning platform for orchestrated bioinformatics processes. *BMC.Bioinformatics.* 9 Suppl 4, S12-
- Marhuenda-Egea, F.C., Martinez-Sabater, E., Gonsalvez-Alvarez, R., Lledo, B., Ten, J., and Bernabeu, R. (2009). A crucial step in assisted reproduction technology: human embryo selection using metabolomic evaluation. *Fertil.Steril.*
- Marto, J.A., White, F. M., Seldomridge, S., and Marshall, A. G. (1995). Structural characterization of phospholipids by matrix-assisted laser desorption/ionization Fourier transform ion cyclotron resonance mass spectrometry. *Anal.Chem.* 67, 21, 3979-3984.
- Mashego, M.R., Wu, L., van Dam, J. C., Ras, C., Vinke, J. L., van Winden, W. A., van Gulik, W. M., and Heijnen, J. J. (2004). MIRACLE: mass isotopomer ratio analysis of U-13C-labeled extracts. A new method for accurate quantification of changes in concentrations of intracellular metabolites. *Biotechnol.Bioeng.* 85, 6, 620-628.
- Matyash, V., Liebisch, G., Kurzchalia, T. V., Shevchenko, A., and Schwudke, D. (2008). Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. *J.Lipid Res.* 49, 5, 1137-1146.
- McClellan, J.E., Quarmby, S. T., and Yost, R. A. (2002). Parent and neutral loss monitoring on a quadrupole ion trap mass spectrometer: screening of acylcarnitines in complex mixtures. *Anal.Chem.* 74, 22, 5799-5806.
- McFarland, M.A., Marshall, A. G., Hendrickson, C. L., Nilsson, C. L., Fredman, P., and Mansson, J. E. (2005). Structural characterization of the GM1 ganglioside by infrared multiphoton dissociation, electron capture dissociation, and electron detachment dissociation electrospray ionization FT-ICR MS/MS. *J.Am.Soc.Mass Spectrom.* 16, 5, 752-762.
- Meng, D., Zhang, Q., Gao, X., Wu, S., and Lin, G. (2014). LipidMiner: a software for automated identification and quantification of lipids from multiple liquid chromatography/mass spectrometry data files. *Rapid Commun.Mass Spectrom.* 28, 8, 981-985.
- Merlo, M.E., Jankevics, A., Takano, E., and Breitling, R. (2011). Exploring the metabolic state of microorganisms using metabolomics. *Bioanalysis.* 3, 21, 2443-2458.
- Merrill, A.H., Jr., Sullards, M. C., Allegood, J. C., Kelly, S., and Wang, E. (2005). Sphingolipidomics: high-throughput, structure-specific, and quantitative analysis of sphingolipids by liquid chromatography tandem mass spectrometry. *Methods* 36, 2, 207-224.
- Merrill, A.H., Jr., Wang, E., Innis, W. S., and Mullins, R. (1985). Increases in serum sphingomyelin by 17 beta-estradiol. *Lipids* 20, 4, 252-254.

- Metz, T.O., Zhang, Q., Page, J. S., Shen, Y., Callister, S. J., Jacobs, J. M., and Smith, R. D. (2007). The future of liquid chromatography-mass spectrometry (LC-MS) in metabolic profiling and metabolomic studies for biomarker discovery. *Biomark.Med* 1, 1, 159-185.
- Meyer, F.P., Canzler, E., Giers, H., and Walther, H. (1991). Time Course of Impaired Elimination of Caffeine by the Oral Depot Contraceptive Deposiston. *Zentralblatt fur Gynakologie* 113, 6, 297-302.
- Meyer, H., Liebeke, M., and Lalk, M. (2010). A protocol for the investigation of the intracellular *Staphylococcus aureus* metabolome. *Anal.Biochem.* 401, 2, 250-259.
- Michael-Jubeli, R., Tfyali, A., Bleton, J., and Baillet-Guffroy, A. (2011). Chemometric approach for investigating the skin surface lipids (SSLs) composition: influence of geographical localization. *Eur.J.Dermatol.* 21 Suppl 2, 63-71.
- MICHALEC, C. (1963). PAPER CHROMATOGRAPHY OF SOME CHOLESTEROL DERIVATIVES. *J.Lipid Res.* 4, 110-112.
- Michopoulos, F., Edge, A. M., Theodoridis, G., and Wilson, I. D. (2010a). Application of turbulent flow chromatography to the metabolomic analysis of human plasma: comparison with protein precipitation. *J.Sep.Sci.* 33, 10, 1472-1479.
- Michopoulos, F., Theodoridis, G., Smith, C. J., and Wilson, I. D. (2010b). Metabolite profiles from dried biofluid spots for metabolomic studies using UPLC combined with oaToF-MS. *J.Proteome.Res.* 9, 6, 3328-3334.
- Mikesh, L.M. and Bruns, D. E. (2008). Stabilization of glucose in blood specimens: mechanism of delay in fluoride inhibition of glycolysis. *Clin.Chem.* 54, 5, 930-932.
- Millot, G. (2009). Comprendre et réaliser les tests statistiques avec R.
- Mochel, F., Sedel, F., Vanderver, A., Engelke, U. F., Barritault, J., Yang, B. Z., Kulkarni, B., Adams, D. R., Clot, F., Ding, J. H., Kaneshi, C. R., Verheijen, F. W., Smits, B. W., Seguin, F., Brice, A., Vanier, M. T., Huizing, M., Schiffmann, R., Durr, A., and Wevers, R. A. (2009). Cerebellar ataxia with elevated cerebrospinal free sialic acid (CAFSA). *Brain* 132, Pt 3, 801-809.
- Moe, M.K. and Jensen, E. (2004). Structure elucidation of unsaturated fatty acids after vicinal hydroxylation of the double bonds by negative electrospray ionisation low-energy tandem mass spectrometry. *Eur.J.Mass Spectrom.(Chichester, Eng)* 10, 1, 47-55.
- Moore, S., Matthews.C.E, Sampson, J., Solomon, R., Zheng, W., Cai, Q., Tan, Y., Chow, W. H., Ji, B. T., Liu, D. K., Xiao, Q., Boca, S., Leitzmann, M. F., Yang, G., Xiang, T. B., Sinha, R., Shu, X. O., and Cross, A. J. (2014). Human Metabolic Correlates of Body Mass Index. *Metabolomics* 10, 259-269.
- Mossoba, M.M., Yurawecz, M. P., Roach, J. A., Lin, H. S., McDonald, R. E., Flickinger, B. D., and Perkins, E. G. (1994). Rapid determination of double bond configuration and position along the hydrocarbon chain in cyclic fatty acid monomers. *Lipids* 29, 12, 893-896.
- Mueller, D.M., Duretz, B., Espourteille, F. A., and Rentsch, K. M. (2011). Development of a fully automated toxicological LC-MS(n) screening system in urine using online extraction with turbulent flow chromatography. *Anal.Bioanal.Chem.* 400, 1, 89-100.

- Neuweger, H., Albaum, S. P., Dondrup, M., Persicke, M., Watt, T., Niehaus, K., Stoye, J., and Goemann, A. (2008). MeltDB: a software platform for the analysis and integration of metabolomics experiment data. *Bioinformatics*. 24, 23, 2726-2732.
- Nicholson, J.K., Lindon, J. C., and Holmes, E. (1999). 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* 29, 11, 1181-1189.
- Nichthäuser, J. and Stepnowski, P. (2009). Retention mechanism of selected ionic liquids on a pentafluorophenylpropyl polar phase: investigation using RP-HPLC. *J.Chromatogr.Sci.* 47, 3, 247-253.
- Nie, H., Liu, R., Yang, Y., Bai, Y., Guan, Y., Qian, D., Wang, T., and Liu, H. (2010). Lipid profiling of rat peritoneal surface layers by online normal- and reversed-phase 2D LC QToF-MS. *J.Lipid Res.* 51, 9, 2833-2844.
- Nikkilä, J., Sysi-Aho, M., Ermolov, A., Seppänen-Laakso, T., Simell, O., Kaski, S., and Oresic, M. (2008). Gender-dependent progression of systemic metabolic states in early childhood. *Mol.Syst.Biol.* 4, 197-
- Nordström, A., O'Maille, G., Qin, C., and Siuzdak, G. (2006). Nonlinear data alignment for UPLC-MS and HPLC-MS based metabolomics: quantitative analysis of endogenous and exogenous metabolites in human serum. *Anal.Chem.* 78, 10, 3289-3295.
- Oberacher, H., Pavlic, M., Libiseller, K., Schubert, B., Sulyok, M., Schuhmacher, R., Csaszar, E., and Kofeler, H. C. (2009a). On the inter-instrument and inter-laboratory transferability of a tandem mass spectral reference library: 1. Results of an Austrian multicenter study. *J Mass Spectrom* 44, 4, 485-493.
- Oberacher, H., Pavlic, M., Libiseller, K., Schubert, B., Sulyok, M., Schuhmacher, R., Csaszar, E., and Kofeler, H. C. (2009b). On the inter-instrument and the inter-laboratory transferability of a tandem mass spectral reference library: 2. Optimization and characterization of the search algorithm. *J Mass Spectrom* 44, 4, 494-502.
- Ogiso, H. and Taguchi, R. (2008). Reversed-phase LC/MS method for polyphosphoinositide analyses: changes in molecular species levels during epidermal growth factor activation in A431 cells. *Anal.Chem.* 80, 23, 9226-9232.
- Oliver, S.G., Winson, M. K., Kell, D. B., and Baganz, F. (1998). Systematic functional analysis of the yeast genome. *Trends Biotechnol.* 16, 9, 373-378.
- Ortweiler, W., Simon, H. U., Splinter, F. K., Peiker, G., Siegert, C., and Traeger, A. (1985). Elimination of Caffeine and Metamizol in Pregnancy and After Delivery As An In vivo Method for Characterization of Various Cytochrome-P-450 Dependent Biotransformation Reactions. *Biomedica Biochimica Acta* 44, 7-8, 1189-1199.
- Otieno, A.C. and Mwangela, S. M. (2008). Capillary electrophoresis-based methods for the determination of lipids--a review. *Anal.Chim.Acta* 624, 2, 163-174.
- Palit, M. and Mallard, G. (2009). Fragmentation energy index for universalization of fragmentation energy in ion trap mass spectrometers for the analysis of chemical weapon convention related chemicals by atmospheric pressure ionization-tandem mass spectrometry analysis. *Anal.Chem* 81, 7, 2477-2485.

- Panchal, M., Gaudin, M., Lazar, A. N., Salvati, E., Rivals, I., Ayciriex, S., Dauphinot, L., Dargere, D., Auzeil, N., Masserini, M., Laprevote, O., and Duyckaerts, C. (2014). Ceramides and sphingomyelinases in senile plaques. *Neurobiol.Dis.* 65, 193-201.
- Parvy, P., Bardet, J., Rabier, D., Bonnefont, J. P., and Kamoun, P. (1989). A new pitfall in plasma amino acid analysis. *Clin.Chem.* 35, 1, 178-
- Pasikanti, K.K., Ho, P. C., and Chan, E. C. (2008). Gas chromatography/mass spectrometry in metabolic profiling of biological fluids. *J.Chromatogr.B Analyt.Technol.Biomed.Life Sci.* 871, 2, 202-211.
- Pasini, E.M., Kirkegaard, M., Mortensen, P., Lutz, H. U., Thomas, A. W., and Mann, M. (2006). In-depth analysis of the membrane and cytosolic proteome of red blood cells. *Blood* 108, 3, 791-801.
- Paustenbach, D.J. (2001). *The practice of exposure assessment.*
- Pearce, S.J., Schrenk, H. H., and Yant, W. P. (1936). Microcolorimetric determination of benzene in blood and urine. *3302 Chem Abstr* 30:5600,
- Pedrioli, P.G., Eng, J. K., Hubley, R., Vogelzang, M., Deutsch, E. W., Raught, B., Pratt, B., Nilsson, E., Angeletti, R. H., Apweiler, R., Cheung, K., Costello, C. E., Hermjakob, H., Huang, S., Julian, R. K., Kapp, E., McComb, M. E., Oliver, S. G., Omenn, G., Paton, N. W., Simpson, R., Smith, R., Taylor, C. F., Zhu, W., and Aebersold, R. (2004). A common open representation of mass spectrometry data and its application to proteomics research. *Nature Biotechnology* 22, 11, 1459-1466.
- Pepe, M.S., Etzioni, R., Feng, Z., Potter, J. D., Thompson, M. L., Thornquist, M., Winget, M., and Yasui, Y. (2001). Phases of biomarker development for early detection of cancer. *J.Natl.Cancer Inst.* 93, 14, 1054-1061.
- Pesek, J.J., Matyska, M. T., and Fischer, S. M. (2011). Improvement of peak shape in aqueous normal phase analysis of anionic metabolites. *J.Sep.Sci.* 34, 24, 3509-3516.
- Pesek, J.J., Matyska, M. T., Fischer, S. M., and Sana, T. R. (2008). Analysis of hydrophilic metabolites by high-performance liquid chromatography-mass spectrometry using a silica hydride-based stationary phase. *J.Chromatogr.A* 1204, 1, 48-55.
- Peters, S., van, V. E., and Janssen, H. G. (2009). Parameter selection for peak alignment in chromatographic sample profiling: objective quality indicators and use of control samples. *Anal.Bioanal.Chem.* 394, 5, 1273-1281.
- Pflueger, M., Seppanen-Laakso, T., Suortti, T., Hyotylainen, T., Achenbach, P., Bonifacio, E., Oresic, M., and Ziegler, A. G. (2011). Age- and islet autoimmunity-associated differences in amino acid and lipid metabolites in children at risk for type 1 diabetes. *Diabetes* 60, 11, 2740-2747.
- Pham, H.T., Ly, T., Trevitt, A. J., Mitchell, T. W., and Blanksby, S. J. (2012). Differentiation of complex lipid isomers by radical-directed dissociation mass spectrometry. *Anal.Chem.* 84, 17, 7525-7532.
- Pinto, J., Domingues, M. R., Galhano, E., Pita, C., Almeida, M. C., Carreira, I. M., and Gil, A. M. (2014). Human plasma stability during handling and storage: impact on NMR metabolomics. *Analyst* 139, 5, 1168-1177.

- Pizarro, C., renzana-Ramila, I., Perez-del-Notario, N., Perez-Matute, P., and Gonzalez-Saiz, J. M. (2013). Plasma lipidomic profiling method based on ultrasound extraction and liquid chromatography mass spectrometry. *Anal.Chem.* 85, 24, 12085-12092.
- Plumb, R.S., Johnson, K. A., Rainville, P., Smith, B. W., Wilson, I. D., Castro-Perez, J. M., and Nicholson, J. K. (2006). UPLC/MS(E); a new approach for generating molecular fragment information for biomarker structure elucidation. *Rapid Commun.Mass Spectrom.* 20, 13, 1989-1994.
- Plumb, R.S., Stumpf, C. L., Granger, J. H., Castro-Perez, J., Haselden, J. N., and Dear, G. J. (2003). Use of liquid chromatography/time-of-flight mass spectrometry and multivariate statistical analysis shows promise for the detection of drug metabolites in biological fluids. *Rapid Commun.Mass Spectrom.* 17, 23, 2632-2638.
- Pluskal, T., Castillo, S., Villar-Briones, A., and Oresic, M. (2010). MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC.Bioinformatics.* 11, 395-
- Preiano, M., Pasqua, L., Gallelli, L., Galasso, O., Gasparini, G., Savino, R., and Terracciano, R. (2012). Simultaneous extraction and rapid visualization of peptidomic and lipidomic body fluids fingerprints using mesoporous aluminosilicate and MALDI-TOF MS. *Proteomics.* 12, 22, 3286-3294.
- Psychogios, N., Hau, D. D., Peng, J., Guo, A. C., Mandal, R., Bouatra, S., Sinelnikov, I., Krishnamurthy, R., Eisner, R., Gautam, B., Young, N., Xia, J., Knox, C., Dong, E., Huang, P., Hollander, Z., Pedersen, T. L., Smith, S. R., Bamforth, F., Greiner, R., McManus, B., Newman, J. W., Goodfriend, T., and Wishart, D. S. (2011). The human serum metabolome. *PLoS.One.* 6, 2, e16957-
- Pulfer, M. and Murphy, R. C. (2003). Electrospray mass spectrometry of phospholipids. *Mass Spectrom.Rev.* 22, 5, 332-364.
- Qi, X., Zhang, Y., Gao, J., Chen, T., Zhao, A., Yan, Y., and Jia, W. (2011). Metabolite profiling of hemodialysate using gas chromatography time-of-flight mass spectrometry. *J.Pharm.Biomed.Anal.* 55, 5, 1142-1147.
- Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., Liang, S., Zhang, W., Guan, Y., Shen, D., Peng, Y., Zhang, D., Jie, Z., Wu, W., Qin, Y., Xue, W., Li, J., Han, L., Lu, D., Wu, P., Dai, Y., Sun, X., Li, Z., Tang, A., Zhong, S., Li, X., Chen, W., Xu, R., Wang, M., Feng, Q., Gong, M., Yu, J., Zhang, Y., Zhang, M., Hansen, T., Sanchez, G., Raes, J., Falony, G., Okuda, S., Almeida, M., LeChatelier, E., Renault, P., Pons, N., Batto, J. M., Zhang, Z., Chen, H., Yang, R., Zheng, W., Li, S., Yang, H., Wang, J., Ehrlich, S. D., Nielsen, R., Pedersen, O., Kristiansen, K., and Wang, J. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* 490, 7418, 55-60.
- Quehenberger, O., Armando, A. M., Brown, A. H., Milne, S. B., Myers, D. S., Merrill, A. H., Bandyopadhyay, S., Jones, K. N., Kelly, S., Shaner, R. L., Sullards, C. M., Wang, E., Murphy, R. C., Barkley, R. M., Leiker, T. J., Raetz, C. R., Guan, Z., Laird, G. M., Six, D. A., Russell, D. W., McDonald, J. G., Subramaniam, S., Fahy, E., and Dennis, E. A. (2010). Lipidomics reveals a remarkable diversity of lipids in human plasma. *J.Lipid Res.* 51, 11, 3299-3305.
- Quinones, M.P. and Kaddurah-Daouk, R. (2009). Metabolomics tools for identifying biomarkers for neuropsychiatric diseases. *Neurobiol.Dis.* 35, 2, 165-176.

- Ramautar, R., Busnel, J. M., Deelder, A. M., and Mayboroda, O. A. (2012). Enhancing the coverage of the urinary metabolome by sheathless capillary electrophoresis-mass spectrometry. *Anal.Chem.* 84, 2, 885-892.
- Ramautar, R., Mayboroda, O. A., Somsen, G. W., and de Jong, G. J. (2011a). CE-MS for metabolomics: Developments and applications in the period 2008-2010. *Electrophoresis* 32, 1, 52-65.
- Ramautar, R., Nevedomskaya, E., Mayboroda, O. A., Deelder, A. M., Wilson, I. D., Gika, H. G., Theodoridis, G. A., Somsen, G. W., and de Jong, G. J. (2011b). Metabolic profiling of human urine by CE-MS using a positively charged capillary coating and comparison with UPLC-MS. *Mol.Biosyst.* 7, 1, 194-199.
- Ramautar, R., Somsen, G. W., and de Jong, G. J. (2012). CE-MS for metabolomics: developments and applications in the period 2010-2012. *Electrophoresis*
- Rasche, F., Scheubert, K., Hufsky, F., Zichner, T., Kai, M., Svatos, A., and Bocker, S. (2012). Identifying the unknowns by aligning fragmentation trees. *Anal.Chem.* 84, 7, 3417-3426.
- Rezzi, S., Ramadan, Z., Martin, F. P., Fay, L. B., van, B. P., Lindon, J. C., Nicholson, J. K., and Kochhar, S. (2007). Human metabolic phenotypes link directly to specific dietary preferences in healthy individuals. *J.Proteome.Res.* 6, 11, 4469-4477.
- Rifai, N., Gillette, M. A., and Carr, S. A. (2006). Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nature Biotechnology* 24, 8, 971-983.
- Robb, D.B., Covey, T. R., and Bruins, A. P. (2000). Atmospheric pressure photoionization: an ionization method for liquid chromatography-mass spectrometry. *Anal.Chem* 72, 15, 3653-3659.
- Roberts, J.K. (2000). NMR adventures in the metabolic labyrinth within plants. *Trends Plant Sci.* 5, 1, 30-34.
- Robledo, V.R. and Smyth, W. F. (2014). Review of the CE-MS platform as a powerful alternative to conventional couplings in bio-omics and target-based applications. *Electrophoresis*
- Rojas-Cherto, M., Peironcely, J. E., Kasper, P. T., van der Hooft, J. J., de Vos, R. C., Vreeken, R., Hankemeier, T., and Reijmers, T. (2012a). Metabolite identification using automated comparison of high-resolution multistage mass spectral trees. *Anal.Chem.* 84, 13, 5524-5534.
- Rojas-Cherto, M., van, V. M., Peironcely, J. E., van, D. R., Kooyman, M., te, B. T., van Driel, M. A., Hankemeier, T., and Reijmers, T. (2012b). MetiTree: a web application to organize and process high-resolution multi-stage mass spectrometry metabolomics data. *Bioinformatics.* 28, 20, 2707-2709.
- Rolfsson, O., Palsson, B. O., and Thiele, I. (2011). The human metabolic reconstruction Recon 1 directs hypotheses of novel human metabolic functions. *BMC.Syst.Biol.* 5, 155-
- Roux, A., Lison, D., Junot, C., and Heilier, J. F. (2011). Applications of liquid chromatography coupled to mass spectrometry-based metabolomics in clinical chemistry and toxicology: A review. *Clin Biochem* 44, 1, 119-135.
- Roux, A., Xu, Y., Heilier, J. F., Olivier, M. F., Ezan, E., Tabet, J. C., and Junot, C. (2012). Annotation of the human adult urinary metabolome and metabolite identification using ultra high performance liquid chromatography coupled to a linear quadrupole ion trap-Orbitrap mass spectrometer. *Anal.Chem.* 84, 15, 6429-6437.

- Roux-Dalvai, F., Gonzalez de, P. A., Simo, C., Guerrier, L., Bouyssie, D., Zanella, A., Citterio, A., Burlet-Schiltz, O., Boschetti, E., Righetti, P. G., and Monsarrat, B. (2008). Extensive analysis of the cytoplasmic proteome of human erythrocytes using the peptide ligand library technology and advanced mass spectrometry. *Mol.Cell Proteomics*. 7, 11, 2254-2269.
- Sabatine, M.S., Liu, E., Morrow, D. A., Heller, E., McCarroll, R., Wiegand, R., Berriz, G. F., Roth, F. P., and Gerszten, R. E. (2005). Metabolomic identification of novel biomarkers of myocardial ischemia. *Circulation* 112, 25, 3868-3875.
- Saito, K., Maekawa, K., Pappan, K. L., Urata, M., Ishikawa, M., Kamagai, M., and Saito, Y. (2014). Differences in Metabolite Profiles between Blood Matrices, Ages, and Sexes among Caucasian Individuals and Their Inter-Individual Variations. *Metabolomics* 10, 402-413.
- Salek, R.M., Haug, K., and Steinbeck, C. (2013). Dissemination of metabolomics results: role of MetaboLights and COSMOS. *Gigascience*. 2, 1, 8-
- Sandra, K., Pereira, A. S., Vanhoenacker, G., David, F., and Sandra, P. (2010). Comprehensive blood plasma lipidomics by liquid chromatography/quadrupole time-of-flight mass spectrometry. *J.Chromatogr.A* 1217, 25, 4087-4099.
- Saric, J., Want, E. J., Duthaler, U., Lewis, M., Keiser, J., Shockcor, J. P., Ross, G. A., Nicholson, J. K., Holmes, E., and Tavares, M. F. (2012). Systematic evaluation of extraction methods for multiplatform-based metabotyping: application to the *Fasciola hepatica* metabolome. *Anal.Chem.* 84, 16, 6963-6972.
- Sartain, M.J., Dick, D. L., Rithner, C. D., Crick, D. C., and Belisle, J. T. (2011). Lipidomic analyses of *Mycobacterium tuberculosis* based on accurate mass measurements and the novel "Mtb LipidDB". *J.Lipid Res.* 52, 5, 861-872.
- Sato, Y., Nakamura, T., Aoshima, K., and Oda, Y. (2010). Quantitative and wide-ranging profiling of phospholipids in human plasma by two-dimensional liquid chromatography/mass spectrometry. *Anal.Chem.* 82, 23, 9858-9864.
- Savitski, M.M., Sweetman, G., Askenazi, M., Marto, J. A., Lang, M., Zinn, N., and Bantscheff, M. (2011). Delayed fragmentation and optimized isolation width settings for improvement of protein identification and accuracy of isobaric mass tag quantification on Orbitrap-type mass spectrometers. *Anal.Chem.* 83, 23, 8959-8967.
- Scheltema, R.A., Jankevics, A., Jansen, R. C., Swertz, M. A., and Breitling, R. (2011). PeakML/mzMatch: a file format, Java library, R library, and tool-chain for mass spectrometry data analysis. *Anal.Chem.* 83, 7, 2786-2793.
- Schiller, J. and Arnold, K. (2002). Application of high resolution ³¹P NMR spectroscopy to the characterization of the phospholipid composition of tissues and body fluids - a methodological review. *Med.Sci.Monit.* 8, 11, MT205-MT222.
- Schuhmann, K., Almeida, R., Baumert, M., Herzog, R., Bornstein, S. R., and Shevchenko, A. (2012). Shotgun lipidomics on a LTQ Orbitrap mass spectrometer by successive switching between acquisition polarity modes. *J.Mass Spectrom.* 47, 1, 96-104.
- Schuhmann, K., Herzog, R., Schwudke, D., Metelmann-Strupat, W., Bornstein, S. R., and Shevchenko, A. (2011). Bottom-up shotgun lipidomics by higher energy collisional dissociation on LTQ Orbitrap mass spectrometers. *Anal.Chem.* 83, 14, 5480-5487.

- Schwarzenberg, A., Ichou, F., Cole, R. B., Huron-Mandard, X., Junot, C., Lesage, D., and Tabet, J. C. (2013). Identification tree based on fragmentation rules for structure elucidation of organophosphorus esters by electrospray mass spectrometry. *J. Mass Spectrom.* 48, 5, 576-586.
- Schwudke, D., Hannich, J. T., Surendranath, V., Grimard, V., Moehring, T., Burton, L., Kurzchalia, T., and Shevchenko, A. (2007). Top-down lipidomic screens by multivariate analysis of high-resolution survey mass spectra. *Anal. Chem.* 79, 11, 4083-4093.
- Schwudke, D., Oegema, J., Burton, L., Entchev, E., Hannich, J. T., Ejsing, C. S., Kurzchalia, T., and Shevchenko, A. (2006). Lipid profiling by multiple precursor and neutral loss scanning driven by the data-dependent acquisition. *Anal. Chem.* 78, 2, 585-595.
- Seo, S. and Lewin, H. A. (2009). Reconstruction of metabolic pathways for the cattle genome. *BMC Syst. Biol.* 3, 33-
- Shah, V.P., Midha, K. K., Findlay, J. W., Hill, H. M., Hulse, J. D., McGilveray, I. J., McKay, G., Miller, K. J., Patnaik, R. N., Powell, M. L., Tonelli, A., Viswanathan, C. T., and Yacobi, A. (2000). Bioanalytical method validation--a revisit with a decade of progress. *Pharm. Res.* 17, 12, 1551-1557.
- Shan, L., Jaffe, K., Li, S., and Davis, L. (2008). Quantitative determination of lysophosphatidic acid by LC/ESI/MS/MS employing a reversed phase HPLC column. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 864, 1-2, 22-28.
- Shibue, M., Mant, C. T., and Hodges, R. S. (2005a). The perchlorate anion is more effective than the trifluoroacetate anion as an ion-pairing reagent for reversed-phase chromatography of peptides. *J. Chromatogr. A* 1080, 1, 49-57.
- Shibue, M., Mant, C. T., and Hodges, R. S. (2005b). Effect of anionic ion-pairing reagent concentration (1-60 mM) on reversed-phase liquid chromatography elution behaviour of peptides. *J. Chromatogr. A* 1080, 1, 58-67.
- Shibue, M., Mant, C. T., and Hodges, R. S. (2005c). Effect of anionic ion-pairing reagent hydrophobicity on selectivity of peptide separations by reversed-phase liquid chromatography. *J. Chromatogr. A* 1080, 1, 68-75.
- Shiva, S., Vu, H. S., Roth, M. R., Zhou, Z., Marepally, S. R., Nune, D. S., Lushington, G. H., Visvanathan, M., and Welti, R. (2013). Lipidomic analysis of plant membrane lipids by direct infusion tandem mass spectrometry. *Methods Mol. Biol.* 1009, 79-91.
- Shlomi, T., Herrgard, M., Portnoy, V., Naim, E., Palsson, B. O., Sharan, R., and Ruppin, E. (2007). Systematic condition-dependent annotation of metabolic genes. *Genome Res.* 17, 11, 1626-1633.
- Shortreed, M.R., Lamos, S. M., Frey, B. L., Phillips, M. F., Patel, M., Belshaw, P. J., and Smith, L. M. (2006). Ionizable isotopic labeling reagent for relative quantification of amine metabolites by mass spectrometry. *Anal. Chem.* 78, 18, 6398-6403.
- Sigurdsson, M.I., Jamshidi, N., Steingrimsdottir, E., Thiele, I., and Palsson, B. O. (2010). A detailed genome-wide reconstruction of mouse metabolism based on human Recon 1. *BMC Syst. Biol.* 4, 140-

- Simon-Manso, Y., Lowenthal, M. S., Kilpatrick, L. E., Sampson, M. L., Telu, K. H., Rudnick, P. A., Mallard, W. G., Bearden, D. W., Schock, T. B., Tchekhovskoi, D. V., Blonder, N., Yan, X., Liang, Y., Zheng, Y., Wallace, W. E., Neta, P., Phinney, K. W., Remaley, A. T., and Stein, S. E. (2013). Metabolite profiling of a NIST Standard Reference Material for human plasma (SRM 1950): GC-MS, LC-MS, NMR, and clinical laboratory analyses, libraries, and web-based resources. *Anal.Chem.* 85, 24, 11725-11731.
- Simonsen, K.W., Hermansson, S., Steentoft, A., and Linnet, K. (2010a). A validated method for simultaneous screening and quantification of twenty-three benzodiazepines and metabolites plus zopiclone and zaleplone in whole blood by liquid-liquid extraction and ultra-performance liquid chromatography- tandem mass spectrometry. *J.Anal.Toxicol.* 34, 6, 332-341.
- Simonsen, K.W., Steentoft, A., Buck, M., Hansen, L., and Linnet, K. (2010b). Screening and quantitative determination of twelve acidic and neutral pharmaceuticals in whole blood by liquid-liquid extraction and liquid chromatography-tandem mass spectrometry. *J.Anal.Toxicol.* 34, 7, 367-373.
- Sinclair, A.J., Viant, M. R., Ball, A. K., Burdon, M. A., Walker, E. A., Stewart, P. M., Rauz, S., and Young, S. P. (2010). NMR-based metabolomic analysis of cerebrospinal fluid and serum in neurological diseases--a diagnostic tool? *NMR Biomed.* 23, 2, 123-132.
- Singh, A. and Prasad, R. (2011). Comparative lipidomics of azole sensitive and resistant clinical isolates of *Candida albicans* reveals unexpected diversity in molecular lipid imprints. *PLoS.One.* 6, 4, e19266-
- Sirbu, A., Ruskin, H. J., and Crane, M. (2010). Cross-platform microarray data normalisation for regulatory network inference. *PLoS.One.* 5, 11, e13822-
- Smedsgaard, J. and Frisvad, J. C. (1996). Using direct electrospray mass spectrometry in taxonomy and secondary metabolite profiling of crude fungal extracts. *Journal of Microbiological Methods* 25, 1, 5-17.
- Smilde, A.K., van der Werf, M. J., Bijlsma, S., van der Werff-van der Vat BJ, and Jellema, R. H. (2005). Fusion of mass spectrometry-based metabolomics data. *Anal.Chem.* 77, 20, 6729-6736.
- Smith, C.A., O'Maille, G., Want, E. J., Qin, C., Trauger, S. A., Brandon, T. R., Custodio, D. E., Abagyan, R., and Siuzdak, G. (2005). METLIN: a metabolite mass spectral database. *Therapeutics and Drugs Monitoring* 27, 6, 747-751.
- Smith, C.A., Want, E. J., O'Maille, G., Abagyan, R., and Siuzdak, G. (2006). XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal.Chem* 78, 3, 779-787.
- Smith, L.M., Maher, A. D., Want, E. J., Elliott, P., Stamler, J., Hawkes, G. E., Holmes, E., Lindon, J. C., and Nicholson, J. K. (2009). Large-scale human metabolic phenotyping and molecular epidemiological studies via ¹H NMR spectroscopy of urine: investigation of borate preservation. *Anal.Chem.* 81, 12, 4847-4856.
- Smolinska, A., Blanchet, L., Coulier, L., Ampt, K. A., Luiders, T., Hintzen, R. Q., Wijmenga, S. S., and Buydens, L. M. (2012a). Interpretation and visualization of non-linear data fusion in kernel space: study on metabolomic characterization of progression of multiple sclerosis. *PLoS.One.* 7, 6, e38163-

- Smolinska, A., Posma, J. M., Blanchet, L., Ampt, K. A., Attali, A., Tuinstra, T., Luider, T., Doskocz, M., Michiels, P. J., Girard, F. C., Buydens, L. M., and Wijmenga, S. S. (2012b). Simultaneous analysis of plasma and CSF by NMR and hierarchical models fusion. *Anal.Bioanal.Chem.* 403, 4, 947-959.
- Soga, T., Ueno, Y., Naraoka, H., Ohashi, Y., Tomita, M., and Nishioka, T. (2002). Simultaneous determination of anionic intermediates for *Bacillus subtilis* metabolic pathways by capillary electrophoresis electrospray ionization mass spectrometry. *Anal.Chem.* 74, 10, 2233-2239.
- Sokol, E., Almeida, R., Hannibal-Bach, H. K., Kotowska, D., Vogt, J., Baumgart, J., Kristiansen, K., Nitsch, R., Knudsen, J., and Ejsing, C. S. (2013). Profiling of lipid species by normal-phase liquid chromatography, nanoelectrospray ionization, and ion trap-orbitrap mass spectrometry. *Anal.Biochem.* 443, 1, 88-96.
- Sommer, U., Herscovitz, H., Welty, F. K., and Costello, C. E. (2006). LC-MS-based method for the qualitative and quantitative analysis of complex lipid mixtures. *J.Lipid Res.* 47, 4, 804-814.
- Song, H., Hsu, F. F., Ladenson, J., and Turk, J. (2007). Algorithm for processing raw mass spectrometric data to identify and quantitate complex lipid molecular species in mixtures by data-dependent scanning and fragment ion database searching. *J.Am.Soc.Mass Spectrom.* 18, 10, 1848-1858.
- Song, H., Ladenson, J., and Turk, J. (2009). Algorithms for automatic processing of data from mass spectrometric analyses of lipids. *J.Chromatogr.B Analyt.Technol.Biomed.Life Sci.* 877, 26, 2847-2854.
- Spickett, C.M., Reis, A., and Pitt, A. R. (2011). Identification of oxidized phospholipids by electrospray ionization mass spectrometry and LC-MS using a QQLIT instrument. *Free Radic.Biol.Med.* 51, 12, 2133-2149.
- Sreekumar, A., Poisson, L. M., Rajendiran, T. M., Khan, A. P., Cao, Q., Yu, J., Laxman, B., Mehra, R., Lonigro, R. J., Li, Y., Nyati, M. K., Ahsan, A., Kalyana-Sundaram, S., Han, B., Cao, X., Byun, J., Omenn, G. S., Ghosh, D., Pennathur, S., Alexander, D. C., Berger, A., Shuster, J. R., Wei, J. T., Varambally, S., Beecher, C., and Chinnaiyan, A. M. (2009). Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature* 457, 7231, 910-914.
- Stahlman, M., Ejsing, C. S., Tarasov, K., Perman, J., Boren, J., and Ekroos, K. (2009). High-throughput shotgun lipidomics by quadrupole time-of-flight mass spectrometry. *J.Chromatogr.B Analyt.Technol.Biomed.Life Sci.* 877, 26, 2664-2672.
- Steinbeck, C., Conesa, P., Haug, K., Mahendraker, T., Williams, M., Maguire, E., Rocca-Serra, P., Sansone, S. A., Salek, R. M., and Griffin, J. L. (2012). MetaboLights: towards a new COSMOS of metabolomics data management. *Metabolomics* 8, 5, 757-760.
- Stella, C., Rudaz, S., Gauvrit, J. Y., Lanteri, P., Huteau, A., Tchaplal, A., and Veuthey, J. L. (2007). Characterization and comparison of the chromatographic performance of different types of reversed-phase stationary phases. *J.Pharm.Biomed.Anal.* 43, 1, 89-98.
- Sturm, M. and Kohlbacher, O. (2009). TOPPView: an open-source viewer for mass spectrometry data. *J Proteome.Res.* 8, 7, 3760-3763.
- Subramaniam, S., Fahy, E., Gupta, S., Sud, M., Byrnes, R. W., Cotter, D., Dinasarapu, A. R., and Maurya, M. R. (2011). Bioinformatics and systems biology of the lipidome. *Chem.Rev.* 111, 10, 6452-6490.

- Sud, M., Fahy, E., Cotter, D., Brown, A., Dennis, E. A., Glass, C. K., Merrill, A. H., Jr., Murphy, R. C., Raetz, C. R., Russell, D. W., and Subramaniam, S. (2007). LMSD: LIPID MAPS structure database. *Nucleic Acids Res.* 35, Database issue, D527-D532.
- Suhre, K., Meisinger, C., Doring, A., Altmaier, E., Belcredi, P., Gieger, C., Chang, D., Milburn, M. V., Gall, W. E., Weinberger, K. M., Mewes, H. W., Hrabe de, A. M., Wichmann, H. E., Kronenberg, F., Adamski, J., and Illig, T. (2010). Metabolic footprint of diabetes: a multiplatform metabolomics study in an epidemiological setting. *PLoS.One.* 5, 11, e13953-
- Suhre, K., Shin, S. Y., Petersen, A. K., Mohny, R. P., Meredith, D., Wagele, B., Altmaier, E., Deloukas, P., Erdmann, J., Grundberg, E., Hammond, C. J., de Angelis, M. H., Kastenmuller, G., Kottgen, A., Kronenberg, F., Mangino, M., Meisinger, C., Meitinger, T., Mewes, H. W., Milburn, M. V., Prehn, C., Raffler, J., Ried, J. S., Romisch-Margl, W., Samani, N. J., Small, K. S., Wichmann, H. E., Zhai, G., Illig, T., Spector, T. D., Adamski, J., Soranzo, N., and Gieger, C. (2011). Human metabolic individuality in biomedical and pharmaceutical research. *Nature* 477, 7362, 54-60.
- Sullards, M.C., Wang, E., Peng, Q., and Merrill, A. H., Jr. (2003). Metabolomic profiling of sphingolipids in human glioma cell lines by liquid chromatography tandem mass spectrometry. *Cell Mol.Biol.(Noisy.-le-grand)* 49, 5, 789-797.
- Sumner, L.W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., Fan, T. W. M., Fiehn, O., Goodacre, R., Griffin, J. L., Hankemeier, T., Hardy, N., Harnly, J., Higashi, R., Kopka, J., Lane, A. N., Lindon, J. C., Marriott, P., Nicholls, A. W., Reilly, M. D., Thaden, J. J., and Viant, M. R. (2007). Proposed minimum reporting standards for chemical analysis. *Metabolomics* 3, 3, 211-221.
- Sumner, L.W., Mendes, P., and Dixon, R. A. (2003). Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry* 62, 6, 817-836.
- Surinova, S., Schiess, R., Huttenhain, R., Cerciello, F., Wollscheid, B., and Aebersold, R. (2011). On the development of plasma protein biomarkers. *J.Proteome.Res.* 10, 1, 5-16.
- Syka, J.E., Coon, J. J., Schroeder, M. J., Shabanowitz, J., and Hunt, D. F. (2004). Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc.Natl.Acad.Sci.U.S.A* 101, 26, 9528-9533.
- Sysi-Aho, M., Katajamaa, M., Yetukuri, L., and Oresic, M. (2007). Normalization method for metabolomics data using optimal selection of multiple internal standards. *BMC.Bioinformatics.* 8, 93-
- Taguchi, R., Hayakawa, J., Takeuchi, Y., and Ishida, M. (2000). Two-dimensional analysis of phospholipids by capillary liquid chromatography/electrospray ionization mass spectrometry. *J.Mass Spectrom.* 35, 8, 953-966.
- Taguchi, R., Nishijima, M., and Shimizu, T. (2007). Basic analytical systems for lipidomics by mass spectrometry in Japan. *Methods Enzymol.* 432, 185-211.
- Takats, Z., Wiseman, J. M., Gologan, B., and Cooks, R. G. (2004). Mass spectrometry sampling under ambient conditions with desorption electrospray ionization. *Science* 306, 5695, 471-473.
- Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y., Yoshida, T., and Matsuo, T. (1988). Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun.Mass Spectrom.* 2, 151-153.

- Tang, B.K. (1995). Caffeine Metabolism As A Measure of Cytochrome-P450 Function. *Journal of Cellular Biochemistry* 187-187.
- Tautenhahn, R., Bottcher, C., and Neumann, S. (2008). Highly sensitive feature detection for high resolution LC/MS. *BMC.Bioinformatics*. 9, 504-
- Tautenhahn, R., Patti, G. J., Kalisiak, E., Miyamoto, T., Schmidt, M., Lo, F. Y., McBee, J., Baliga, N. S., and Siuzdak, G. (2011). metaXCMS: second-order analysis of untargeted metabolomics data. *Anal.Chem.* 83, 3, 696-700.
- Tautenhahn, R., Patti, G. J., Rinehart, D., and Siuzdak, G. (2012). XCMS Online: a web-based platform to process untargeted metabolomic data. *Anal.Chem.* 84, 11, 5035-5039.
- Teahan, O., Gamble, S., Holmes, E., Waxman, J., Nicholson, J. K., Bevan, C., and Keun, H. C. (2006). Impact of analytical bias in metabonomic studies of human blood serum and plasma. *Anal.Chem.* 78, 13, 4307-4318.
- Tengstrand, E., Lindberg, J., and Aberg, K. M. (2014). TracMass 2-A Modular Suite of Tools for Processing Chromatography-Full Scan Mass Spectrometry Data. *Anal.Chem.* 86, 7, 3435-3442.
- Thiele, I. and Palsson, B. O. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat.Protoc.* 5, 1, 93-121.
- Thomas Moyon, abien Le Marec, El Mostafa Qannari, velyne Vigneau, Aurélie Le Plain, Frédérique Courant, Jean-Philippe Antignac, Patricia Parnet, and Marie-Cécile Alexandre-Gouabau (2012). Statistical strategies for relating metabolomics and proteomics data: a real case study in nutrition research area. *Metabolomics* 8, 6, 1090-1101.
- Thomas, M.C., Mitchell, T. W., and Blanksby, S. J. (2006). Ozonolysis of phospholipid double bonds during electrospray ionization: a new tool for structure determination. *J.Am.Chem.Soc.* 128, 1, 58-59.
- Thomas, M.C., Mitchell, T. W., Harman, D. G., Deeley, J. M., Nealon, J. R., and Blanksby, S. J. (2008). Ozone-induced dissociation: elucidation of double bond position within mass-selected lipid ions. *Anal.Chem.* 80, 1, 303-311.
- Timbrell, J.A. (1998). Biomarkers in toxicology. *Toxicology* 129, 1, 1-12.
- Tiziani, S., Emwas, A. H., Lodi, A., Ludwig, C., Bunce, C. M., Viant, M. R., and Gunther, U. L. (2008). Optimized metabolite extraction from blood serum for 1H nuclear magnetic resonance spectroscopy. *Anal.Biochem.* 377, 1, 16-23.
- Tolstikov, V.V. and Fiehn, O. (2002). Analysis of highly polar compounds of plant origin: combination of hydrophilic interaction chromatography and electrospray ion trap mass spectrometry. *Anal.Biochem* 301, 2, 298-307.
- Touboul, D., Piednoel, H., Voisin, V., De La, P. S., Brunelle, A., Halgand, F., and Laprevote, O. (2004). Changes of phospholipid composition within the dystrophic muscle by matrix-assisted laser desorption/ionization mass spectrometry and mass spectrometry imaging. *Eur.J.Mass Spectrom.(Chichester, Eng)* 10, 5, 657-664.

- Touboul, D., Roy, S., Germain, D. P., Baillet, A., Brion, F., Prognon, P., Chaminade, P., and Laprevote, O. (2005). Fast fingerprinting by MALDI-TOF mass spectrometry of urinary sediment glycosphingolipids in Fabry disease. *Anal.Bioanal.Chem.* 382, 5, 1209-1216.
- Townsend, M.K., Clish, C. B., Kraft, P., Wu, C., Souza, A. L., Deik, A. A., Tworoger, S. S., and Wolpin, B. M. (2013). Reproducibility of metabolomic profiles among men and women in 2 large cohort studies. *Clin.Chem.* 59, 11, 1657-1667.
- Trusheim, M.R., Burgess, B., Hu, S. X., Long, T., Averbuch, S. D., Flynn, A. A., Lieftucht, A., Mazumder, A., Milloy, J., Shaw, P. M., Swank, D., Wang, J., Berndt, E. R., Goodsaid, F., and Palmer, M. C. (2011). Quantifying factors for the success of stratified medicine. *Nat.Rev.Drug Discov.* 10, 11, 817-833.
- Trygg, J., Holmes, E., and Lundstedt, T. (2007). Chemometrics in metabonomics. *J.Proteome.Res.* 6, 2, 469-479.
- Tweeddale, H., Notley-McRobb, L., and Ferenci, T. (1998). Effect of slow growth on metabolism of *Escherichia coli*, as revealed by global metabolite pool ("metabolome") analysis. *J.Bacteriol.* 180, 19, 5109-5116.
- van den Berg, R.A., Hoefsloot, H. C., Westerhuis, J. A., Smilde, A. K., and van der Werf, M. J. (2006). Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* 7, 142-
- van, '., V, Dai, H., van, d., V, He, Y. D., Hart, A. A., Mao, M., Peterse, H. L., van der, K. K., Marton, M. J., Witteveen, A. T., Schreiber, G. J., Kerkhoven, R. M., Roberts, C., Linsley, P. S., Bernards, R., and Friend, S. H. (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415, 6871, 530-536.
- Vance, D.E. and Vance, J. E. (2002). *Biochemistry of Lipids, Lipoproteins and Membranes*, 4th Edition. 4th, 9-Ether linked lipids and their bioactive species,
- Vaughan, A.A., Dunn, W. B., Allwood, J. W., Wedge, D. C., Blackhall, F. H., Whetton, A. D., Dive, C., and Goodacre, R. (2012). Liquid chromatography-mass spectrometry calibration transfer and metabolomics data fusion. *Anal.Chem.* 84, 22, 9848-9857.
- Vervoort, R.J., Ruyter, E., Debets, A. J., Claessens, H. A., Cramers, C. A., and de Jong, G. J. (2002). Characterisation of reversed-phase stationary phases for the liquid chromatographic analysis of basic pharmaceuticals by thermodynamic data. *J.Chromatogr.A* 964, 1-2, 67-76.
- Veselkov, K.A., Vingara, L. K., Masson, P., Robinette, S. L., Want, E., Li, J. V., Barton, R. H., Boursier-Neyret, C., Walther, B., Ebbels, T. M., Pelczar, I., Holmes, E., Lindon, J. C., and Nicholson, J. K. (2011). Optimized preprocessing of ultra-performance liquid chromatography/mass spectrometry urinary metabolic profiles for improved information recovery. *Anal.Chem.* 83, 15, 5864-5872.
- Vitic, J., Nikolic, B., and Nikolic, V. (1981). A specific type of dislipidaemia-a complex between monotypic immunoglobulins and lipids. *J.Clin.Lab Immunol.* 6, 3, 227-231.
- Vuckovic, D. (2012). Current trends and challenges in sample preparation for global metabolomics using liquid chromatography-mass spectrometry. *Anal.Bioanal.Chem.* 403, 6, 1523-1548.

- Vuckovic, D. and Pawliszyn, J. (2011). Systematic evaluation of solid-phase microextraction coatings for untargeted metabolomic profiling of biological fluids by liquid chromatography-mass spectrometry. *Anal.Chem.* 83, 6, 1944-1954.
- Walter B.Cannon (1932). *The Wisdom of the Body*.
- Wang, T.J., Larson, M. G., Vasan, R. S., Cheng, S., Rhee, E. P., McCabe, E., Lewis, G. D., Fox, C. S., Jacques, P. F., Fernandez, C., O'Donnell, C. J., Carr, S. A., Mootha, V. K., Florez, J. C., Souza, A., Melander, O., Clish, C. B., and Gerszten, R. E. (2011). Metabolite profiles and the risk of developing diabetes. *Nat.Med.* 17, 4, 448-453.
- Want, E.J., Masson, P., Michopoulos, F., Wilson, I. D., Theodoridis, G., Plumb, R. S., Shockcor, J., Loftus, N., Holmes, E., and Nicholson, J. K. (2013). Global metabolic profiling of animal and human tissues via UPLC-MS. *Nat.Protoc.* 8, 1, 17-32.
- Want, E.J., O'Maille, G., Smith, C. A., Brandon, T. R., Uritboonthai, W., Qin, C., Trauger, S. A., and Siuzdak, G. (2006). Solvent-dependent metabolite distribution, clustering, and protein extraction for serum profiling with mass spectrometry. *Anal.Chem.* 78, 3, 743-752.
- Want, E.J., Wilson, I. D., Gika, H., Theodoridis, G., Plumb, R. S., Shockcor, J., Holmes, E., and Nicholson, J. K. (2010). Global metabolic profiling procedures for urine using UPLC-MS. *Nat.Protoc.* 5, 6, 1005-1018.
- Watanabe, K., Yasugi, E., and Oshima, M. (2000). How to search the glycolipid data in "LIPIDBANK for Web", the newly developed lipid database in Japan. *Trends in Glycoscience and Glycotechnology* 12, 65, 175-184.
- Watson, A.D. (2006). Thematic review series: systems biology approaches to metabolic and cardiovascular disorders. Lipidomics: a global approach to lipid analysis in biological systems. *J.Lipid Res.* 47, 10, 2101-2111.
- Wei, J., Buriak, J. M., and Siuzdak, G. (1999). Desorption-ionization mass spectrometry on porous silicon. *Nature* 399, 6733, 243-246.
- Wei, R., Li, G., and Seymour, A. B. (2010). High-throughput and multiplexed LC/MS/MRM method for targeted metabolomics. *Anal.Chem.* 82, 13, 5527-5533.
- Weir, J.M., Wong, G., Barlow, C. K., Greeve, M. A., Kowalczyk, A., Almasy, L., Comuzzie, A. G., Mahaney, M. C., Jowett, J. B., Shaw, J., Curran, J. E., Blangero, J., and Meikle, P. J. (2013). Plasma lipid profiling in a large population-based cohort. *J.Lipid Res.* 54, 10, 2898-2908.
- Wenk, M.R. (2005). The emerging field of lipidomics. *Nat.Rev.Drug Discov.* 4, 7, 594-610.
- Wenk, M.R., Lucast, L., Di, P. G., Romanelli, A. J., Suchy, S. F., Nussbaum, R. L., Cline, G. W., Shulman, G. I., McMurray, W., and De, C. P. (2003). Phosphoinositide profiling in complex lipid mixtures using electrospray ionization mass spectrometry. *Nature Biotechnology* 21, 7, 813-817.
- Werner, E., Croixmarie, V., Umbdenstock, T., Ezan, E., Chaminade, P., Tabet, J. C., and Junot, C. (2008a). Mass spectrometry-based metabolomics: accelerating the characterization of discriminating signals by combining statistical correlations and ultrahigh resolution. *Anal.Chem.* 80, 13, 4918-4932.

- Werner, E., Heilier, J. F., Ducruix, C., Ezan, E., Junot, C., and Tabet, J. C. (2008b). Mass spectrometry for the identification of the discriminating signals from metabolomics: current status and future trends. *J Chromatogr.B Analyt.Technol Biomed.Life Sci* 871, 2, 143-163.
- Wharfe, E.S., Jarvis, R. M., Winder, C. L., Whiteley, A. S., and Goodacre, R. (2010). Fourier transform infrared spectroscopy as a metabolite fingerprinting tool for monitoring the phenotypic changes in complex bacterial communities capable of degrading phenol. *Environ.Microbiol.* 12, 12, 3253-3263.
- White, T., Bursten, S., Federighi, D., Lewis, R. A., and Nudelman, E. (1998). High-resolution separation and quantification of neutral lipid and phospholipid species in mammalian cells and sera by multi-one-dimensional thin-layer chromatography. *Anal.Biochem.* 258, 1, 109-117.
- Whiteaker, J.R., Lin, C., Kennedy, J., Hou, L., Trute, M., Sokal, I., Yan, P., Schoenherr, R. M., Zhao, L., Voytovich, U. J., Kelly-Spratt, K. S., Krasnoselsky, A., Gafken, P. R., Hogan, J. M., Jones, L. A., Wang, P., Amon, L., Chodosh, L. A., Nelson, P. S., McIntosh, M. W., Kemp, C. J., and Paulovich, A. G. (2011). A targeted proteomics-based pipeline for verification of biomarkers in plasma. *Nature Biotechnology* 29, 7, 625-634.
- Whitehead, S.N., Hou, W., Ethier, M., Smith, J. C., Bourgeois, A., Denis, R., Bennett, S. A., and Figeys, D. (2007). Identification and quantitation of changes in the platelet activating factor family of glycerophospholipids over the course of neuronal differentiation by high-performance liquid chromatography electrospray ionization tandem mass spectrometry. *Anal.Chem.* 79, 22, 8539-8548.
- Williams, R.E., Major, H., Lock, E. A., Lenz, E. M., and Wilson, I. D. (2005). D-Serine-induced nephrotoxicity: a HPLC-TOF/MS-based metabolomics approach. *Toxicology* 207, 2, 179-190.
- Wilson, I.D., Nicholson, J. K., Castro-Perez, J., Granger, J. H., Johnson, K. A., Smith, B. W., and Plumb, R. S. (2005a). High resolution "ultra performance" liquid chromatography coupled to oa-TOF mass spectrometry as a tool for differential metabolic pathway profiling in functional genomic studies. *J.Proteome.Res.* 4, 2, 591-598.
- Wilson, I.D., Plumb, R., Granger, J., Major, H., Williams, R., and Lenz, E. M. (2005b). HPLC-MS-based methods for the study of metabolomics. *J.Chromatogr.B Analyt.Technol.Biomed.Life Sci.* 817, 1, 67-76.
- Wishart, D.S. (2011). Advances in metabolite identification. *Bioanalysis.* 3, 15, 1769-1782.
- Wishart, D.S., Tzur, D., Knox, C., Eisner, R., Guo, A. C., Young, N., Cheng, D., Jewell, K., Arndt, D., Sawhney, S., Fung, C., Nikolai, L., Lewis, M., Coutouly, M. A., Forsythe, I., Tang, P., Shrivastava, S., Jeroncic, K., Stothard, P., Amegbey, G., Block, D., Hau, D. D., Wagner, J., Miniaci, J., Clements, M., Gebremedhin, M., Guo, N., Zhang, Y., Duggan, G. E., Macinnis, G. D., Weljie, A. M., Dowlatabadi, R., Bamforth, F., Clive, D., Greiner, R., Li, L., Marrie, T., Sykes, B. D., Vogel, H. J., and Querengesser, L. (2007). HMDB: the Human Metabolome Database. *Nucleic Acids Res* 35, Database issue, D521-D526.
- Wolf, S., Schmidt, S., Muller-Hannemann, M., and Neumann, S. (2010). In silico fragmentation for computer assisted identification of metabolite mass spectra. *BMC.Bioinformatics.* 11, 148-
- Wu, C.C., MacCoss, M. J., Howell, K. E., Matthews, D. E., and Yates, J. R., III (2004). Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis. *Anal.Chem.* 76, 17, 4951-4959.

- Wu, H., Volponi, J. V., Oliver, A. E., Parikh, A. N., Simmons, B. A., and Singh, S. (2011). In vivo lipidomics using single-cell Raman spectroscopy. *Proc.Natl.Acad.Sci.U.S.A* 108, 9, 3809-3814.
- Wu, H., Zhang, J., Norem, K., and El-Shourbagy, T. A. (2008). Simultaneous determination of a hydrophobic drug candidate and its metabolite in human plasma with salting-out assisted liquid/liquid extraction using a mass spectrometry friendly salt. *J.Pharm.Biomed.Anal.* 48, 4, 1243-1248.
- Xia, J., Psychogios, N., Young, N., and Wishart, D. S. (2009). MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Res.* 37, Web Server issue, W652-W660.
- Xia, J. and Wishart, D. S. (2011). Web-based inference of biological patterns, functions and pathways from metabolomic data using MetaboAnalyst. *Nat.Protoc.* 6, 6, 743-760.
- Xu, Y. and Brenna, J. T. (2007). Atmospheric pressure covalent adduct chemical ionization tandem mass spectrometry for double bond localization in monoene- and diene-containing triacylglycerols. *Anal.Chem.* 79, 6, 2525-2536.
- Xu, Y., Heilier, J. F., Madalinski, G., Genin, E., Ezan, E., Tabet, J. C., and Junot, C. (2010). Evaluation of accurate mass and relative isotopic abundance measurements in the LTQ-orbitrap mass spectrometer for further metabolomics database building. *Anal.Chem.* 82, 13, 5490-5501.
- Yalow, R.S. and Berson, S. A. (1960a). Plasma insulin concentrations in nondiabetic and early diabetic subjects. Determinations by a new sensitive immuno-assay technic. *Diabetes* 9, 254-260.
- Yalow, R.S. and Berson, S. A. (1960b). Immunoassay of endogenous plasma insulin in man. *J.Clin.Invest* 39, 1157-1175.
- Yamada, T., Uchikata, T., Sakamoto, S., Yokoi, Y., Fukusaki, E., and Bamba, T. (2013a). Development of a lipid profiling system using reverse-phase liquid chromatography coupled to high-resolution mass spectrometry with rapid polarity switching and an automated lipid identification software. *J.Chromatogr.A* 1292, 211-218.
- Yamada, T., Uchikata, T., Sakamoto, S., Yokoi, Y., Nishiumi, S., Yoshida, M., Fukusaki, E., and Bamba, T. (2013b). Supercritical fluid chromatography/Orbitrap mass spectrometry based lipidomics platform coupled with automated lipid identification software for accurate lipid profiling. *J.Chromatogr.A* 1301, 237-242.
- Yang, J., Yang, S., Gao, X., and Yuan, Y. J. (2011). Integrative investigation of lipidome and signal pathways in human endothelial cells under oxidative stress. *Mol.Biosyst.* 7, 8, 2428-2440.
- Yang, K., Cheng, H., Gross, R. W., and Han, X. (2009). Automated lipid identification and quantification by multidimensional mass spectrometry-based shotgun lipidomics. *Anal.Chem.* 81, 11, 4356-4368.
- Yasugi, E. and Watanabe, K. (2002). [LIPIDBANK for Web, the newly developed lipid database]. *Tanpakushitsu Kakusan Koso* 47, 7, 837-841.
- Yetukuri, L., Ekroos, K., Vidal-Puig, A., and Oresic, M. (2008). Informatics and computational strategies for the study of lipids. *Mol.Biosyst.* 4, 2, 121-127.

- Yetukuri, L., Katajamaa, M., Medina-Gomez, G., Seppanen-Laakso, T., Vidal-Puig, A., and Oresic, M. (2007). Bioinformatics strategies for lipidomics analysis: characterization of obesity related hepatic steatosis. *BMC.Syst.Biol.* 1, 12-
- Yin, P., Peter, A., Franken, H., Zhao, X., Neukamm, S. S., Rosenbaum, L., Lucio, M., Zell, A., Haring, H. U., Xu, G., and Lehmann, R. (2013). Preanalytical aspects and sample quality assessment in metabolomics studies of human blood. *Clin.Chem.* 59, 5, 833-845.
- Yoon, H.R., Kim, H., and Cho, S. H. (2003). Quantitative analysis of acyl-lysophosphatidic acid in plasma using negative ionization tandem mass spectrometry. *J.Chromatogr.B Analyt.Technol.Biomed.Life Sci.* 788, 1, 85-92.
- Yu, Z., Kastenmuller, G., He, Y., Belcredi, P., Moller, G., Prehn, C., Mendes, J., Wahl, S., Roemisch-Margl, W., Ceglarek, U., Polonikov, A., Dahmen, N., Prokisch, H., Xie, L., Li, Y., Wichmann, H. E., Peters, A., Kronenberg, F., Suhre, K., Adamski, J., Illig, T., and Wang-Sattler, R. (2011). Differences between human plasma and serum metabolite profiles. *PLoS.One.* 6, 7, e21230-
- Yunoki, K., Sato, M., Seki, K., Ohkubo, T., Tanaka, Y., and Ohnishi, M. (2009). Simultaneous quantification of plant glyceroglycolipids including sulfoquinovosyldiacylglycerol by HPLC-ELSD with binary gradient elution. *Lipids* 44, 1, 77-83.
- Zelena, E., Dunn, W. B., Broadhurst, D., Francis-McIntyre, S., Carroll, K. M., Begley, P., O'Hagan, S., Knowles, J. D., Halsall, A., Wilson, I. D., and Kell, D. B. (2009). Development of a robust and repeatable UPLC-MS method for the long-term metabolomic study of human serum. *Anal.Chem.* 81, 4, 1357-1364.
- Zeng, M., Xiao, Y., Liang, Y., Wang, B., Chen, X., Cao, D., Li, H., Wang, M., and Zhou, Z. (2009). Metabolic alterations of impaired fasting glucose by GC/MS based plasma metabolic profiling combined with chemometrics. *Metabolomics* 4, 2, 303-311.
- Zeng, Z., Liu, X., Dai, W., Yin, P., Zhou, L., Huang, Q., Lin, X., and Xu, G. (2014). Ion Fusion of High-Resolution LC-MS-Based Metabolomics Data to Discover More Reliable Biomarkers. *Anal.Chem.*
- Zhang, T., Creek, D. J., Barrett, M. P., Blackburn, G., and Watson, D. G. (2012). Evaluation of coupling reversed phase, aqueous normal phase, and hydrophilic interaction liquid chromatography with Orbitrap mass spectrometry for metabolomic studies of human urine. *Anal.Chem.* 84, 4, 1994-2001.
- Zhao, X., Fritsche, J., Wang, J., Chen, J., Rittig, K., Schmitt-Kopplin, P., Fritsche, A., Haring, H. U., Schleicher, E. D., Xu, G., and Lehmann, R. (2010). Metabonomic fingerprints of fasting plasma and spot urine reveal human pre-diabetic metabolic traits. *Metabolomics* 6, 3, 362-374.
- Zhou, X., Mao, J., Ai, J., Deng, Y., Roth, M. R., Pound, C., Henegar, J., Welte, R., and Bigler, S. A. (2012). Identification of plasma lipid biomarkers for prostate cancer by lipidomics and bioinformatics. *PLoS.One.* 7, 11, e48889-
- Zhu, X., Chen, Y., and Subramanian, R. (2014). Comparison of information-dependent acquisition, SWATH, and MS(All) techniques in metabolite identification study employing ultrahigh-performance liquid chromatography-quadrupole time-of-flight mass spectrometry. *Anal.Chem.* 86, 2, 1202-1209.

Le profilage global des liquides biologiques humains est un enjeu majeur en analyse métabolomique. En effet, les métabolites présentent des propriétés physico-chimiques différentes, il est donc nécessaire d'utiliser des méthodes complémentaires afin d'assurer une couverture optimale. Cela implique généralement l'utilisation combinée de techniques de RMN et de SM. Toutefois, la polyvalence des couplages LC-HRMS permet à elle seule de surmonter ce problème. Un autre obstacle est l'utilisation de ces outils de profilage non ciblés pour l'analyse de grandes cohortes. Ceci peut être surmonté par le développement d'outils bioinformatiques dédiés assurant différentes fonctions (c'est-à-dire, la détection et l'alignement des variables, les analyses statistiques, la fusion de données et l'identification des métabolites). Dans ce contexte, notre travail vise à développer et à mettre en œuvre des approches basées sur la LC-HRMS pour les analyses métabolomique et lipidomique d'échantillons de sérum humain pour l'étude de larges cohortes. Dans un premier temps, nous avons développé puis évalué la pertinence de 4 méthodes LC-HRMS dans l'annotation du métabolome et lipidome sériques humains. Des centaines de milliers de signaux ont été détectés à partir d'extraits de sérum toutefois l'identification d'une majeure partie demeurait inconnue. Dans ce contexte, une base de données spectrales a été implémentée à l'aide de composés chimiques de référence et d'extraits lipidiques purifiés. Les spectres MS, MS/MS ainsi que les temps de rétention de 800 métabolites et 2000 espèces de lipides ont été enregistrés afin d'assurer l'annotation de jeux de données LC-HRMS, gérer la redondance du signal et l'identification des métabolites. La combinaison de méthodes RP-, HILIC- et PFPP-HRMS a permis l'identification de 266 métabolites dans des échantillons de sérum répartis sur 20 classes chimiques dont 27% d'espèces isomères. Les couplage HILIC-ESI(-)HRMS assure la meilleure couverture métabolique alors que la PFPP-ESI(+)HRMS apporte la meilleure résolution. En outre, 700 espèces lipidiques uniques correspondant à 24 classes et sous-classes ont été identifiées dans le lipidome données en fonction de leur rapport masse sur charge, leur temps de rétention et le rapport isotopique. Ces outils ont été appliqués, dans un second temps, à de larges séries d'analyse afin de proposer des biomarqueurs permettant de mieux caractériser des états physiopathologiques. Ainsi, nous avons analysé des échantillons plasmatiques humains d'une cohorte de 78 patients diabétiques dans le cadre d'une étude pilote visant à les stratifier sur la base de leurs empreintes métaboliques. Outre les voies métabolique perturbés lors d'un syndrome métabolique (glycolyse, beta-oxydation), nos analyses préliminaires ont montré l'impact délétère de facteurs physiologiques confondants -tels que l'âge et l'indice de masse corporelle IMC- sur la mise en évidence de biomarqueurs de stratification. Nous avons de ce fait évalué l'influence de ces facteurs physiologiques (âge, genre et IMC) sur les métabolites et lipides sériques précédemment identifiés, et ce à l'aide d'une cohorte de 227 salariés du CEA-Saclay, prélevés dans le cadre de leur suivi de médecine du travail. Nous avons constaté que les empreintes lipidomiques étaient plus robustes à l'influence des facteurs physiologiques que celles obtenues en analyses métaboliques. Néanmoins, l'impact de l'IMC est marqué notamment pour la famille des lipides neutres. L'effet du genre démontre un catabolisme masculin plus prononcé. L'effet de l'âge se manifeste par des activités enzymatiques altérées. Ces études sont parmi les premières à combiner des analyses LC/ESI-HRMS globales métabolomiques et lipidomiques sur les mêmes échantillons dans de larges cohortes humaines. Elles constituent la première étape de construction d'une base de données relationnelle incluant données spectrales (spectres MS, spectres MS/MS) et biologiques (variations des conditions physiologiques et physio-pathologiques observées) qui pourra être utilisée pour la caractérisation de biomarqueurs dans le cas d'études cliniques.

Global profiling of human biofluids is a major bottleneck in metabolomics. Indeed, metabolites display huge differences in their physicochemical properties. It is thus required to use complementary methods to achieve an optimal coverage. This usually involves the combined use of NMR and MS technologies. However, the versatility of HRMS allows to overcome this issue. Another key concern is to use such untargeted profiling tools for large human cohort studies which is managed by dedicated bioinformatics tools (i.e., peak detection and alignment, statistics, data fusion and metabolite identification). In this context, our work aims to develop and to implement LC-HRMS-based approaches for metabolomic and lipidomic analyzes of human serum samples for studying large cohort studies using advanced data mining tools. In this study, we first evaluated the relevance of our LC-HRMS platform to perform a comprehensive study of the human serum metabolome and lipidome. Hundreds to thousands of signals were detected from serum extracts however major part of this information remained unknown. In this context, In-house spectral libraries were implemented using chemical compounds and purified lipid extracts. MS, MS/MS spectra and retention times of 800 metabolites and 2000 lipid species were recorded in order to ensure LC-HRMS dataset annotation, handle signal redundancy and assisted metabolite identification. Using these tools, the combination of RP, HILIC and PFPP coupled to HRMS resulted in the identification of 266 metabolites in serum samples distributed over 20 chemical classes including 27% of isomeric species. HILIC ensures the greatest metabolome coverage in the negative ionization mode whereas PFPP was the most effective in the positive one. In addition, more than 700 unique lipid species corresponding to 24 classes and subclasses were identified in RPLC-HRMS lipidomics dataset based on their exact mass-to-charge ratio, retention time and isotopic ratio. Then, these developed tools were applied to large samples series for better characterize pathophysiological conditions. Thus, we analyzed human plasma samples from a 78 diabetic patients cohort in order to classify patients on the basis of their metabolic fingerprints. Besides metabolic pathways disruption during a metabolic syndrome (glycolysis, beta-oxidation), our preliminary analysis showed the deleterious impact of physiological confounding factors-such as age and body mass index BMI-on setting Identification of biomarkers for classification. We therefore evaluated the influence of these physiological factors (age, gender and BMI) on the previously identified metabolites and serum lipids using a cohort of 227 CEA employees. We found that the lipidomic fingerprints were more robust to the influence of physiological factors than those obtained in metabolic analyzes. Nevertheless, the impact of BMI is particularly marked for neutral lipids. The effect of gender shows a more pronounced catabolism in male. The effect of age is characterized by altered enzyme activities. These studies are among the first to combine LC/ESI-HRMS global profiling metabolomics and lipidomics on the same samples of large human cohorts. They are the first step to build a relational database including spectral data (MS spectra, MS / MS spectra) and biological (changes in physiopathological conditions) that can be used for the characterization of biomarkers in clinical studies.