



HAL
open science

Cartographie individuelle de la composition de méconiums humains : un reflet de l'exposome prénatal corrélé au développement d'une allergie alimentaire dans l'enfance ?

Nihel Bekhti

► To cite this version:

Nihel Bekhti. Cartographie individuelle de la composition de méconiums humains : un reflet de l'exposome prénatal corrélé au développement d'une allergie alimentaire dans l'enfance ?. Biochimie, Biologie Moléculaire. Université Paris Cité, 2021. Français. NNT : 2021UNIP5006 . tel-03882361

HAL Id: tel-03882361

<https://theses.hal.science/tel-03882361v1>

Submitted on 2 Dec 2022

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.

Université de Paris

École doctorale N°563 : Médicament Toxicologie Chimie et Imageries

Laboratoire d'Immuno-Allergie Alimentaire (LIAA) et

Laboratoire d'Etudes du Métabolisme des Médicaments (LEMM)

Cartographie individuelle de la composition de méconiums humains : un reflet de l'exposome prénatal corrélé au développement d'une allergie alimentaire dans l'enfance ?

Par **Nihel BEKHTI**

Thèse de doctorat de : Interface Chimie/Biologie

Dirigée par **Karine ADEL-PATIENT**

Présentée et soutenue publiquement le 05 mars 2021

Devant un jury composé de :

Dr. Estelle PUJOS-GUILLOT

Ingénieur de recherche HDR, INRAE, Université Clermont Ferrand

Pr. Alain PARIS

Professeur HDR, Muséum National d'Histoire Naturelle

Dr. Anne-Judith WALIGORA

Maître de conférence HDR, Université de Paris

Pr. Pierre CHAMINADE

Professeur HDR, Université Paris-Saclay

Dr. Karine ADEL-PATIENT

Directeur de recherche INRAE, HDR, UMR Université Paris-Saclay,CEA, INRAE MTS

Dr. François FENAILLE

Ingénieur chercheur CEA, HDR, UMR CEA-INRAE, MetaboHUB, Université Paris-Saclay

Rapporteur

Rapporteur

Examineur

Examineur

Directrice de thèse

Co-Encadrant de
thèse



REMERCIEMENTS

Je tiens en premier lieu à remercier l'ensemble des membres de mon jury, Estelle Pujos-Guillot et Alain Paris en tant que rapporteurs et Anne-Judith Waligora et Pierre Chaminade en tant qu'examineurs, pour avoir accepté d'évaluer mes travaux de thèse réalisés au Commissariat de l'Energie Atomique et des énergies alternatives (CEA de Saclay), au sein du Service de Pharmacologie et Immunoanalyse (SPI) et des laboratoires : LEMM (Laboratoire d'Etudes du Métabolisme des Médicaments) et LIAA (Laboratoire d'Immuno-Allergie Alimentaire).

Je tiens ensuite à remercier l'ensemble de mon équipe encadrante. Dans un premier temps, je remercie Christophe Junot, chef du SPI de m'avoir donné l'opportunité de réaliser ma thèse au sein de cette grande équipe, merci d'avoir donné suite à ma candidature spontanée. Aussi, je te remercie d'avoir toujours répondu présent aux multiples réunions de thèse, pour ta bonne humeur communicative et pour tes conseils toujours très pertinents. Je remercie dans un second temps Florence Castelli de m'avoir consacré beaucoup de son temps lors de mes débuts en métabolomique, merci pour ta disponibilité et pour tout ce que tu m'as apporté durant mes trois années de thèse. Je tiens ensuite à te remercier très chaleureusement Karine Adel-Patient de m'avoir accordé ta confiance depuis le début et d'avoir toujours répondu présente à mes questions et mes craintes, je suis heureuse et chanceuse d'avoir été accompagnée par une directrice de thèse comme toi. Merci pour ta douce personne, pour tes qualités pédagogiques et scientifiques dont j'ai beaucoup appris. Merci pour tout le temps que tu m'as consacré, pour tes réponses très tardives et pour ta réactivité. Je tiens ensuite à te remercier très sincèrement François Fenaille pour la confiance que tu m'as accordée, merci pour ta bienveillance, pour tes qualités de communication, ta pédagogie et pour tes conseils très pertinents. Merci d'avoir suivi de près l'évolution de mon travail tout en me laissant une grande indépendance qui m'a beaucoup forgée. Karine et François, Merci pour vos nombreuses relectures et scrupuleuses corrections, pour votre réactivité tard le soir, les week-ends, pendant vos vacances. Cette thèse n'aurait jamais pu se réaliser sans le soutien de chacun d'entre vous.

Je tiens ensuite à remercier l'ensemble des collaborations que j'ai pu réaliser au cours de ces trois années de thèse. D'abord un grand merci à la Clinique Sainte-Thérèse, je cite Sofiane Benachenhou qui m'a permis d'avoir un contact au sein de la maternité Ste Thérèse, et Laurent Petit et Clémence Moiron d'avoir accepté cette collaboration de collecte d'échantillons de méconium au sein de leur clinique. Je remercie ensuite Estelle Rathahao-Paris et Sandra Alves, et l'ensemble des membres de l'IPCM à Jussieu, de m'avoir ouvert grand les portes de votre laboratoire. Merci d'avoir toujours répondu présentes à notre collaboration sur le FT-ICR, merci pour ce que vous m'avez apporté, pour votre bienveillance,

votre bonne humeur, votre gentillesse et votre qualité scientifique. Je tiens ensuite à remercier Jean Armengaud, Lucia Grenga, Guylaine Miotello et l'ensemble des membres du LI2D (Karim H, Virginie J, Yannick C, Duarte G...), CEA-Marcoule au sein duquel les analyses métabolomiques ont été réalisées, sincèrement merci pour votre chaleureux accueil. Merci Jean pour tout le temps que tu m'as accordé, merci pour ta disponibilité, pour tes pertinentes explications, j'ai beaucoup apprécié discuter avec toi. Merci de m'avoir fait découvrir Avignon et pour ton chaleureux accueil. Merci Lucia pour tout le temps que tu m'as consacré, pour ta gentillesse et ta disponibilité. Je tiens ensuite à remercier Claire Cherbuy et Aurélia Bruneau et l'ensemble de l'institut Micalis de l'INRAE de Jouy en Josas pour votre chaleureux accueil. Merci Claire pour tout le temps que tu m'as consacré, et pour tout ce que tu m'as appris sur l'analyse du microbiote, merci pour ta pédagogie et pour ta bienveillance.

Au-delà de ces collaborations, j'ai passé 99 % du temps entre les murs du SPI ☺ où j'ai eu la chance d'être entourée par tant de personnes bienveillantes et chaleureuses, je suis très reconnaissante de tout ce que vous m'avez apporté durant ces trois années. Je ne sais pas dans quel ordre commencer ! je vais commencer par remercier les personnes du 2^{ème} de l'INRAE : Merci à Jihanna A, Mélanie B, Stéphane N, Stéphane H, Hervé B, Eric V, Nicolas G, Céline C, Martha G. Je tiens très spécifiquement à remercier Blanche G, Vincent M et Marine G pour tout le travail que vous avez réalisé dans l'analyse de la grosse cohorte ! Un grand Merci à toi Blanche pour ta qualité de ton travail, pour ta disponibilité et pour ta précise aide dès que j'en avais besoin. Merci aux personnes du LERI, Merci à Menel B, Victoria D, Christian M, Lucia O, Stéphanie G, Audrey R, Elise Y, Duncan D, Hervé B, Alexandra S, DOM, Cécile F, Julie D, Anne W, Karla P, Stéphanie S, Lilas P, Justine R. Merci à tous les anciens du 2^{ème} : Christelle D, Emeline R, Chloé D, Nell H, Emeric D, Rafaël C, Céline G-H, Diane G, Anita B, Vanina M, Karine M-D, Quentin B, Edyta C, Zouriatou G. J'ai adoré les petits déjeuners qu'on a partagé (quand on avait encore droit), les courses relais, les raclettes du labo, et les afterworks ensemble, je ne vous oublierai jamais. Je tiens ensuite à chaleureusement remercier les personnes du 1^{er}, le LEMM : Merci à Sophie C, Kathleen R, Thäïs H, Jessica M, Justyna S, Emeline C-V, Donatien L, Marie H, Chloé J, Julien L, Etienne T, Sandrine L, Florence V, Laurie M, François B, Céline C, Anne-Laure D, Benoit C, Jean-Claude T, Alloïse. Merci aux anciens : Daniel L, Sandrine L, Yasmina A, Hélène C, Aurélie D, Mikail B, Marie Z, Arthur A, Amir L, Insaf B-T, Patricia L, Samuel T, Karl-Joe Z, à l'équipe Medday : Jennifer H, Sylvain D, Simon B, Sandrine A-C, Chloé A, Fanny L, Loïc M, Alexandre S. Merci à tous pour nos discussions scientifiques mais pas que. Merci pour les petits déjeuners du mardi qu'on a pu partager, pour les afterworks, les congrès, séminaires et journées scientifiques où l'on a pu se retrouver... Merci pour tous ces moments mémorables. Je finis par remercier les personnes du bureau, « le bureau de l'ambiance » : Merci à Anne Sophie M (Alias le coach), Fanny R (Alias bbw), Anaïs V (Alias la positive attitude), Antoine A (Alias Antonio), Hervé V (Alias Vovo ou le chef comme vous voulez) et Benjamin M (l'animateur du bureau). Merci pour votre bonne

humeur permanente, merci pour les petits déj du bureau qu'on a pu partager et les dons culinaires de chacun qu'on a pu découvrir (j'attends toujours le Amlou Anaïs ☺), merci hervé pour les baffes, grâce ou à cause de toi je connais les paroles de Soprano par cœur maintenant ☺... La playlist du bureau, les sorties : escape game et afterworks, tout va me manquer, vous êtes des personnes en or, je ne vous oublierai jamais.

Merci à tous ces amis proches qui m'ont soutenu durant ces trois années de thèse et bien plus, Merci à Amro, Souhila, Amène, Manal, Somia, Dounya, Saliha, Fifi, Fadia, Redha, Ania, Zozo, Hind, Nabil, Sofiane, Nezy, et tous les airs andalous et les TB, vous comptez tellement pour moi et m'avez tellement apporté dans ma vie parisienne, je ne vous remercierai jamais assez.

Enfin, mes plus profonds remerciements vont à mes parents, ma sœur, mes frères, pour leur soutien tout au long de mes études, merci d'avoir toujours cru en moi et de m'avoir toujours poussé vers l'avant (Dididine, Amiya, Yaya, Samintoche, Memoun, Essoum). Merci papa de m'avoir aidé tout le long de mes études et d'avoir toujours été là pour moi. Une mention spéciale à ma maman, sans qui je ne serai jamais où j'en suis aujourd'hui, ton aide morale et ta détermination de me voir réussir sont mes plus grandes forces. Merci pour ton amour immense, pour ta patience, une vie entière ne serait pas assez pour te remercier. Merci Mima pour tes prières, merci d'exister, ta présence pour moi n'a pas de prix. Merci à mon neveu Elhadi (5 ans) qui m'appelle tous les jours pour jouer avec lui pendant ma dure période de rédaction, merci pour la force et le sourire que tu me donnes au quotidien. Merci à tous mes proches du côté Bekhti et Benmansour, je vous aime et j'ai hâte de vous revoir ! Rebbi yahfedkoum w ykhelikoum liya. Enfin merci mon Dieu.

PREAMBULE

L'incidence de l'allergie alimentaire (AA) est en constante augmentation depuis plus de 20 ans, touchant 2 à 8 % des enfants [1]. Ces allergies résultent d'une réponse immunitaire inappropriée et excessive contre des protéines alimentaires, qui sont normalement inoffensives pour l'organisme. Les AA touchent les enfants dès les premiers mois de vie, avec une symptomatologie de plus en plus sévère, suggérant des maturations incorrectes du système immunitaire et de la barrière intestinale qui débutent dès la période périnatale.

« L'exposome prénatal » auquel est exposé le fœtus regroupe les facteurs socio-économiques et de santé de la mère, son alimentation, son exposition (à des polluants environnementaux ou alimentaires), son métabolisme interne ainsi que de son microbiote (figure 1). Ces différents facteurs pourraient avoir un impact sur les différentes fonctions en développement du fœtus, notamment son système immunitaire et ses différentes barrières. Dans ce contexte, ma thèse vise à étudier cet exposome prénatal et à évaluer s'il peut être mis en relation avec la santé de l'enfant, notamment le développement d'allergies alimentaires [2].

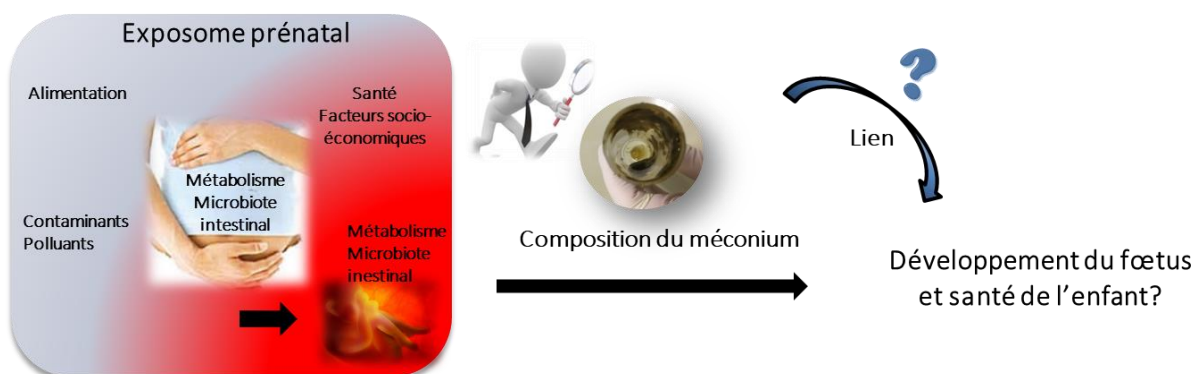


Figure 1 : Questions de mon travail de thèse : est ce que la composition du méconium peut refléter « l'exposome » prénatal, et être mise en relation avec le développement de pathologies dans l'enfance, notamment les allergies alimentaires ?

Pour appréhender cette question, nous avons cherché à obtenir un reflet au moins partiel mais objectif de cet exposome prénatal, par l'analyse de la composition des méconiums. Le méconium est une matrice inerte qui s'accumule dès le 2^{ème} mois de grossesse dans l'intestin du fœtus, et est excrétée durant les premiers jours de vie du nouveau-né [3]. Cette matrice peut donc « enregistrer » ce à quoi a été exposé le fœtus *in utero*, permettant de refléter une partie des expositions de la mère, ainsi que des premiers processus métaboliques du fœtus et de son microbiote (si toutefois celui-ci existe et est métaboliquement actif à ce stade du développement).

Pour l'analyse de cette matrice, nous avons choisi de réaliser une cartographie la plus globale possible en combinant différentes approches « omiques » : la métabolomique, la métaprotéomique et la métagénomique ciblée. Les analyses métaprotéomiques et métagénomiques ciblées ont été réalisées en collaboration avec le laboratoire LI2D du CEA de Marcoule et l'Institut Micalis de l'INRAE de Jouy en Josas, experts dans ces domaines respectifs. En parallèle, la majeure partie de mon travail de thèse a consisté à développer, optimiser et valider l'analyse métabolomique de cette matrice, jamais étudiée au laboratoire. Ceci a notamment impliqué l'optimisation de la préparation des échantillons, de l'acquisition des données métabolomiques et de leur traitement, puis l'identification des métabolites. Ce travail méthodologique a été réalisé grâce aux méconiums issus de 2 nouveau-nés disponibles au laboratoire. J'ai ensuite validé l'approche multi-omiques sur des méconiums collectés longitudinalement auprès de 11 nouveau-nés, grâce à une collaboration que j'ai établie avec la clinique Sainte Thérèse (Paris 17^{ème}). Cette analyse a permis d'apprécier l'évolution de la composition globale des méconiums au cours du temps, mais également d'en approcher la variabilité interindividuelle et d'établir une cartographie globale du métabolome, du protéome et du microbiote.

Les méthodologies mises en place ont finalement été appliquées à l'analyse de plus de 300 échantillons de méconiums, collectés au sein de la cohorte française mère-enfant EDEN. La mise en relation des informations collectées auprès des mères (lieu de vie, alimentation, expositions environnementales, données socio-économiques et de santé...) avec la composition globale des méconiums est en train d'être menée. L'analyse complète de ces données devrait permettre d'affirmer ou non, si l'analyse du méconium peut permettre d'approcher l'exposome prénatal. L'impact de la composition du méconium sur le développement d'une allergie alimentaire dans les 5 premières années de vie de l'enfant sera ensuite évalué.

TABLE DES MATIERES

REMERCIEMENTS	3
PREAMBULE	7
TABLE DES MATIERES	9
LISTE DES FIGURES	17
LISTE DES TABLEAUX	25
LISTE DES ANNEXES	27
INTRODUCTION GENERALE	31
CHAPITRE I : LES ALLERGIES ALIMENTAIRES	33
<i>I. Les allergies alimentaires : contexte général</i>	33
<i>II. Mécanismes immunologiques d'une réaction allergique médiée par des IgE</i>	34
A. La sensibilisation	35
B. Déclenchement ou réaction allergique	35
<i>III. Prévalence de l'allergie alimentaire</i>	35
<i>IV. Principaux allergènes alimentaires et doses réactogènes</i>	36
<i>V. Facteurs associés à l'augmentation de la prévalence des allergies</i>	38
A. La fenêtre prénatale	38
1. Le lieu de vie et expositions aux polluants atmosphériques	38
2. L'alimentation de la mère durant la grossesse	39
B. La fenêtre périnatale ou néonatale	40
1. Le mode de naissance	41
2. Le mode d'alimentation du nourrisson en début de vie	41
<i>VI. Vers l'analyse de l'exposome périnatal en lien avec le développement d'allergies dans l'enfance</i>	42
CHAPITRE II : ETUDE DE L'EXPOSOME FŒTAL AU TRAVERS DE L'ANALYSE DE LA COMPOSITION GLOBALE DU MECONIUM	45
<i>I. Contexte général</i>	45
<i>II. Choix de la matrice de l'étude</i>	46
A. Description et comparaison des matrices fœtales étudiées dans la littérature	46
1. L'urine	46
2. Le sang	46
3. Les cheveux	47
4. Les ongles	47

5.	La salive _____	47
6.	Le méconium _____	47
7.	Conclusion sur la matrice de l'étude _____	48
B.	La matrice méconium _____	49
1.	Formation du méconium _____	49
2.	Composition du méconium _____	50
C.	Etude de la composition du méconium par des approches omiques _____	50
1.	Les approches génomique et métagénomique - application à l'étude du microbiote intestinal _____	52
2.	L'approche protéomique _____	58
 CHAPITRE III : L'ANALYSE METABOLOMIQUE _____		63
I.	Contexte général _____	63
II.	L'approche non ciblée en métabolomique _____	65
A.	Déroulement d'une analyse métabolomique non ciblée _____	65
1.	Préparation des échantillons _____	66
2.	Acquisition des données _____	66
3.	Traitement des données acquises par LC/HRMS _____	77
4.	Analyses statistiques _____	84
5.	Outils d'aide à l'interprétation biologique des données _____	87
B.	Analyses métabolomiques réalisées sur le méconium _____	88
1.	Les analyses métabolomiques non ciblées réalisées sur le méconium _____	88
2.	Les analyses métabolomiques ciblées réalisées sur le méconium _____	89
3.	Les analyses multi-omiques réalisées sur le méconium _____	90
<i>Conclusion</i> _____		93
<i>Présentation du projet de recherche</i> _____		95
 RESULTATS ET DISCUSSION _____		97
 CHAPITRE I : MISE EN PLACE D'UNE METHODOLOGIE D'ANALYSE POUR L'EXPLORATION DU METABOLOME DU MECONIUM _____		99
I.	Introduction _____	99
A.	Préparation des échantillons _____	100
1.	Prétraitement des méconiums _____	100
2.	Extraction des métabolites : pré-extraction et choix du solvant _____	101
3.	Extraction des métabolites : homogénéisation de la matrice _____	101
4.	Séchage et reprise des métabolites extraits _____	102
B.	Séparation chromatographique en amont des acquisitions de données spectrales _____	102
C.	Conclusion _____	103
II.	Matériels et méthodes _____	104
A.	Réactifs et produits chimiques _____	104

B.	Echantillons de méconiums	104
C.	Préparation des échantillons de méconium	105
1.	Lyophilisation des méconiums	105
2.	Homogénéisation des méconiums : test comparatif de la sonde à ultrasons et de la lyse mécanique	105
3.	Reprise des échantillons de méconium pour l'analyse en LC/HRMS	107
D.	Analyse des métabolites par LC/HRMS	107
1.	Conditions chromatographiques	107
2.	Analyse par spectrométrie de masse	107
3.	Traitement des données LC/HRMS	108
E.	Analyse des métabolites par FT-ICR	108
III.	Résultats expérimentaux et discussion	109
A.	Développement et optimisation de la procédure de préparation des échantillons	109
1.	Lyophilisation du méconium	109
2.	Méthode d'homogénéisation du méconium lyophilisé	110
3.	Optimisation du volume de reprise pour l'analyse par LC/HRMS	114
4.	Etude de la stabilité des analyses de méconium par LC/HRMS pour l'analyse de grandes séries d'échantillons	118
5.	Bilan des optimisations et des développements réalisés sur le protocole de préparation de l'échantillon, vers une première description du méconium	122
6.	Analyses métabolomiques réalisées par DIMS (Direct Introduction Mass Spectrometry) sur un FT-ICR	124
B.	Cartographie métabolique du méconium, obtenue dans les conditions HILIC-ESI ⁻	131
IV.	Conclusions du chapitre I et perspectives	136

CHAPITRE II : Réalisation d'une cartographie de la composition du méconium et étude de sa variabilité interindividuelle et de son évolution durant les premiers jours de vie par des approches multi-omiques : métabolomique, métagénomique et métaprotéomique ciblée 137

I.	Introduction	137
II.	Matériels et méthodes	138
A.	Echantillons de méconiums	138
B.	Analyses métabolomiques	140
1.	« Workflow » suivi pour l'analyse métabolomique des échantillons de la cohorte clinique	140
2.	Préparation des échantillons de méconium	141
3.	Acquisition des données LC/HRMS	141
4.	Traitement des données	142
C.	Analyses métagénomiques	144
1.	« Workflow » suivi pour l'analyse métagénomique des échantillons de la cohorte clinique	144
2.	Préparation des échantillons	145
3.	Acquisition des données de métagénomique en LC-MS/MS	145
4.	Traitement des données, annotation et quantification des protéines	146
5.	Analyse des données	147

D.	Analyses métagénomiques ciblées _____	148
1.	Workflow suivi pour l'analyse métagénomique ciblée des échantillons de la cohorte clinique ____	148
2.	Préparation des échantillons _____	149
3.	Traitement des données de séquençage _____	149
III.	<i>Résultats expérimentaux et discussion</i> _____	151
A.	Analyses métabolomiques (article en préparation, Annexe 1). _____	151
1.	Traitement XCMS des données générées par les analyses HILIC-ESI ⁻ et C18-ESI ⁺ _____	151
2.	Suivi de la stabilité du système analytique dans les conditions HILIC-ESI ⁻ et C18-ESI ⁺ _____	151
3.	Annotation des variables _____	152
4.	Cartographie globale des métabolites identifiés dans le méconium _____	156
5.	Etude de l'évolution de la composition métabolique du méconium au cours des trois premiers jours de vie _____	160
B.	Analyses protéomiques _____	179
1.	Analyse des protéines d'origine humaine _____	179
2.	Analyses métaprotéomiques _____	191
C.	Analyses métagénomiques ciblées par séquençage des gènes codant pour les ARNr-16S _	197
1.	Traitement des données et attribution des OTUs (de l'anglais, Operational Taxonomic Unit) ____	197
2.	Etude de l'évolution de la composition bactérienne du méconium au cours des trois premiers jours de vie	197
IV.	<i>Discussion générale des analyses « mutli-omiques » réalisées sur le méconium</i> ____	202
V.	<i>Conclusions du chapitre II et perspectives</i> _____	208

CHAPITRE III : Analyse métabolomique des échantillons de méconium de la cohorte EDEN _____ 210

I.	<i>Introduction</i> _____	210
II.	<i>Matériels et méthodes</i> _____	212
A.	Description de la cohorte EDEN _____	212
B.	Echantillons de méconiums de la cohorte EDEN _____	213
C.	Acquisition des données _____	214
D.	Traitement des données _____	215
III.	<i>Résultats expérimentaux et discussion</i> _____	216
A.	Traitement XCMS des données générées dans les conditions analytiques HILIC-ESI ⁻ et C18-ESI ⁺ et suivi de la stabilité du système analytique _____	216
1.	Nombre de variable extraites et suivi de la stabilité du système analytique _____	216
2.	Nécessité de modifier les filtres analytiques classiquement utilisés pour répondre à notre question scientifique _____	217
B.	Annotation des variables _____	220
1.	Analyse comparative du nombre de variables annotées au sein de la cohorte clinique et de la cohorte EDEN _____	220
2.	Analyse qualitative comparative des variables annotées au sein des échantillons de la cohorte clinique et de la cohorte EDEN _____	221

C.	Analyses statistiques des données de la cohorte EDEN : effet du temps, du centre et du sexe	224
1.	Analyses non supervisées réalisées sur l'ensemble des variables extraites au sein des conditions HILIC-ESI	224
2.	Analyses non supervisées réalisées sur les métabolites annotés	226
3.	Analyses supervisées de l'effet temps et de l'effet centre sur la composition métabolique des méconiums	228
D.	Analyses préliminaires de la relation entre la composition du méconium et le devenir allergique des enfants	230
1.	Première analyse d'association sur les échantillons collectés à 24 heures (J1) au centre de Nancy	230
2.	Analyse d'association sur les échantillons collectés à 36 heures au centre de Nancy	232
IV.	<i>Conclusions du chapitre III et perspectives</i>	235
	CONCLUSIONS GENERALES ET PERSPECTIVES	237
	REFERENCES	243
	PRODUCTION SCIENTIFIQUE	265
	ANNEXES	269

LISTE DES ABBREVIATIONS

AA	Allergie Alimentaire
ACN	Acétonitrile
ACP	Analyses en composantes principales
API	Ionisation à pression atmosphérique (de l'anglais, Atmospheric Pressure Ionization)
APPI	Photoionisation à pression atmosphérique (de l'anglais, Atmospheric Pressure Photoionization)
CE	Electrophorèse capillaire (de l'anglais Capillary Electrophoresis)
CV	Coefficient de variation
EAACI	Académie Européenne d'Allergie et d'Immunologie Clinique
ESI	Ionisation electrospray
EE	Etalon externe
FT-ICR	Spectrométrie de masse à résonance cyclotronique ionique avec transformation de Fourier (de l'anglais, Fourier Transform-Ion Cyclotronic Resonance)
GC/MS	Chromatographie gazeuse couplée à la spectrométrie de masse (de l'anglais, Gas Chromatography Mass Spectrometry)
HMDB	Base de données du métabolome humain, (de l'anglais, Human Metabolome DataBase)
HMO	Oligosaccharides du lait humain (de l'anglais, Human Milk Oligosaccharides)
HRMS	Spectrométrie de masse à haute résolution (de l'anglais, High Resolution Mass Spectrometry)
Ig	Immunoglobuline
LC/MS	Chromatographie liquide couplée à la spectrométrie de masse (de l'anglais, Liquid Chromatography Mass Spectrometry)
LOESS	Régression locale, (de l'anglais, LOcally wEighted Scatterplot Smoother)
m/z	Rapport masse sur charge
MeOH	Méthanol
MS	Spectrométrie de masse (de l'anglais Mass Spectrometry)
MS/MS	Spectrométrie de masse en tandem
NP ou NPLC	Chromatographie liquide en phase normale, (de l'anglais, Normal Phase Liquid Chromatography)
OMS	Organisation Mondiale de la Santé

PLS-DA	Analyses discriminantes des moindres carrés partiels (de l'anglais, Partial Least Square Discriminant Analysis).
ppm	Partie par million
QC	Contrôle qualité (de l'anglais, Quality Control)
RMN	Résonnance magnétique nucléaire
RP ou RPLC	Chromatographie liquide en phase inverse, (de l'anglais, Reverse Phase Liquid Chromatography)
RT	Temps de rétention (de l'anglais, Retention Time)
VIP	Variable importante dans la projection, (de l'anglais, Variable with Importance in Projection)
W4M	Workflow4Metabolomics : outil collectif, www.workflow4metabolomics.usegalaxy.fr
ZIC- <i>p</i> HILIC	Chromatographie liquide à interaction hydrophile, (de l'anglais, ZwitterionIC stationary phase covalently attached to porous Polymer beads suitable for Hydrophilic interaction)

LISTE DES FIGURES

Figure 1 : Questions de mon travail de thèse : est ce que la composition du méconium peut refléter « l'exposome » prénatal, et être mise en relation avec le développement de pathologies dans l'enfance, notamment les allergies alimentaires ?	7
Figure 2 : Classification des réactions adverses aux aliments, adaptée d'une figure de Karine Adel-Patient.	33
Figure 3 : Illustration des deux phases de l'hypersensibilité allergique IgE médiée, adaptée de Prioult et Nagler-Anderson [8].	34
Figure 4 : Prévalence des allergies respiratoires et alimentaires au cours des dernières décennies (www.worldallergy.org).	36
Figure 5 : Influence de différents facteurs pré, péri et post-natals sur la modulation du microbiote intestinal du nourrisson et de l'enfant plus tard dans la vie. Figure adaptée de Milani et al [2].	40
Figure 6 : Illustration des expositions fœtales et notion d'exposome foetal.	45
Figure 7 : A. Etapes de formation du méconium, adaptée de Koren et al [73]. B. Photo du méconium.	49
Figure 8 : Arbre phylogénétique adapté de Ciccarelli et al. La section verte représente les archées, en rouge les eucaryotes et en bleu les bactéries. Dans chaque catégorie, les subdivisions les plus fréquemment utilisées sont représentées [91].	52
Figure 9 : Classification des organismes biologiques et représentation des différents liens hiérarchiques.	54
Figure 10 : A. Communautés microbiennes présentes à différentes localisations du corps humain [92]. B. Composition des communautés bactériennes dominantes le long du tube digestif [103].	55
Figure 11 : Techniques analytiques utilisées dans l'étude du protéome, adaptée de Armengaud et Bland [127].	59
Figure 12 : Représentation de la cascade des sciences « omiques », adaptée de Patti et al [147].	63
Figure 13 : Déroulement d'une analyse métabolomique globale par LC/HRMS, adapté de Junot et al [162].	65
Figure 14 : Nombre de publications en métabolomique reposant sur l'utilisation de la RMN et de la spectrométrie de masse (MS) au cours des 15 dernières années [169].	67
Figure 15 : Comparaison de la couverture métabolique du sérum humain obtenue à travers trois techniques chromatographiques, incluant en rouge la RP-C18, en vert la ZIC-pHILIC, et en bleu la PFPP [197].	70
Figure 16 : Schéma d'un spectromètre de masse [198].	71
Figure 17 : Description des deux matrices de données générées par le traitement automatique XCMS « variable metadata » en vert et « datamatrix » en bleu.	78
Figure 18 : Fréquence relative des différents signaux non biologiques et biologiques extraits lors d'une analyse métabolomique non ciblée [221].	79
Figure 19 : Description des différents signaux MS pouvant être produits pour un seul métabolite, exemple du propionyl-CoA, adaptée de Damont et al [230].	81
Figure 20 : Représentation de la somme des intensités avant (A) et après (B) normalisation. Les QC sont représentés en noir ou en gris dans les deux figures. En vert est illustrée une normalisation réalisée au sein d'un même batch. Tandis qu'en rouge et en orange, la normalisation est réalisée sur deux différents batchs.	82

Figure 21 : Processus d'annotation et d'identification des métabolites par l'utilisation d'une base de donnée interne, et confirmation par analyse MS/MS	83
Figure 22 : A. analyses univariées en fonction d'une modalité qualitative (sexe), avec representation en « boîtes à moustaches » de l'intensité de la variable analysée (acide malique) pour les deux groupes d'échantillons. B. Analyse de la corrélation entre l'intensité de la variable mesurée (caféine) et l'âge des individus (variable continue) [231].	85
Figure 23 : Représentation graphique d'une ACP. En A, le « <i>Score plot</i> » représentant une carte factorielle des individus. En B, le « <i>Loading plot</i> » représentant la carte factorielle des variables dans le jeu de données, adaptée de Roux et al [246].	87
Figure 24 : Corrélation (heatmap) entre les concentrations des 33 métabolites et l'abondance des OTUs du phylum des Actinobactéries au sein d'échantillons de fèces de nourissons. Les codes couleurs permettent de distinguer les corrélations positives (rouge) des négatives (bleu) au sein des métabolites et des OTUs [115].	92
Figure 25 : Estimation du rapport de quantités de l'ADN humain par rapport à l'ADN bactérien détectés dans les échantillons de méconium (analyse par qPCR) [264].	93
Figure 26 : Méthodologie suivie dans l'analyse métabolomique non ciblée du méconium.	103
Figure 27 : A. Pourcentage de la masse en méconium brut perdue au cours du temps de lyophilisation. B. Conditions de lyophilisation retenues pour cette étude.....	110
Figure 28 : Résultats de quantification des protéines (mg) réalisée sur 5 réplicats de méconiums homogénéisés à l'aide d'une sonde à ultrasons (bleu) ou d'une lyse mécanique (rouge). Les moyennes des masses en protéines quantifiées sont représentées par des traits horizontaux pour les deux conditions.	111
Figure 29 : Séparation des protéines hydrosolubles extraites de la matrice méconium par un gel SDS-PAGE. Le premier dépôt de gauche correspond à notre gamme de référence couvrant les masses allant de 3,5 à 160 kDa. Au sein de chacune des deux extractions, deux réplicats expérimentaux sont déposés et sont notés en rouge pour la sonde à ultrasons et en bleu pour la lyse mécanique. Le rectangle bleu montre la gamme où une concentration des bandes protéiques est observée lors de la première extraction. Le rectangle rouge montre des bandes intenses observées pour les faibles masses (< 3,5 kDa) lors de la seconde extraction.	112
Figure 30 : A. Spectre de masse des peptides issus de la digestion trypsique de la protéine à 50 KDa analysée en MALDI-ToF. Les rapports m/z entourés en rouge sont en accord avec les spectres des peptides de l' α -1-antitrypsine dans la base de données. B. Attribution de la protéine à 50 KDa à l' α -1-antitrypsine par le logiciel Mascot, en interrogeant la base de données Swissprot [131].	113
Figure 31 : Evaluation du nombre de variables extraites/filtrées/annotées au sein des solutions résultant des différents volumes de reprise, variant de 800 à 6400 μ L. A. Le nombre de variables extraites après le prétraitement XCMS. B. Le nombre de variables retenues à la suite des trois filtres appliqués. C. le nombre de variables annotées avec redondance (chimique). D. Le nombre de variables annotées sans redondance.....	116
Figure 32 : Diagramme de Venn représentant l'ensemble des annotations des métabolites détectés au sein des différents volumes de reprise allant de 800 à 6400 μ L.	117
Figure 33 : Séquence analytique réalisée pour le suivi de la répétabilité du processus analytique dans les conditions HILIC-ESI.	118
Figure 34 : Evaluation de la répétabilité des analyses au cours de 120 injections d'extraits de méconium A. Suivi des intensités des étalons internes (acide éthylmalonique, amiloride, prednisone et dinoseb) dans les échantillons biologiques et des QC. B. Evaluation de la répétabilité des analyses représentée	

par des boites à moustaches. Ces boites à moustaches nous informent sur : 1. La valeur de la médiane pour chacun des composés de références représentée par le trait au milieu de chaque boite, 2. La valeur de la moyenne représentée par une croix dans la boite. 3. La valeur minimale de la distribution représentée par l'extrémité basse du quartile. 4. La valeur maximale de la distribution représentée par l'extrémité haute du quartile. 5. La valeur du premier quartile (25 %) représentée par la valeur en bas de chaque boite, ainsi que 6. La valeur du troisième quartile (75 %) représentée par la valeur en haut de chaque boite. Uniquement les points en dehors de la distribution sont représentés (cas de l'acide éthylmalonique et de la prednisone, représentés en bleu et en vert respectivement)..... 120

Figure 35 : Etude de la stabilité et de la reproductibilité du processus analytique dans les conditions HILIC-ESI⁻. « QualityMetrics » : tracé des sommes des intensités des variables au sein de tous les échantillons en fonction de l'ordre d'injection. 121

Figure 36 : Protocole de préparation de l'échantillon de méconium final..... 123

Figure 37 : Protocole de préparation des échantillons de méconium pour l'analyse en DIMS dans le mode d'ionisation négatif. STD : solution du standard de l'acide éthylmalonique à une concentration de 0,08 ng/μL. Proportions eau/MeOH (50:50, v/v). 125

Figure 38 : A. Profils du courant total ionique (TIC) obtenus pour les différentes solutions considérées pour le méconium. B. Spectres de masse correspondants aux solutions de méconium aux fd500 à fd2500 ainsi que le spectre de masse du blanc analytique. Les rectangles rouges indiquent la région variant le plus entre les différents facteurs de dilution. Les étoiles rouges indiquent la détection du composé standard de référence dans la matrice méconium au sein des différentes solutions..... 127

Figure 39 : Evaluation de la répétabilité des analyses au cours de la séquence d'injections par : A. Les profils des courants ioniques totaux (TIC) obtenus au sein des différents réplicats d'injections d'un extrait de méconiums au fd1000. B. La représentation de la variation du signal de réponse du composé standard de référence utilisé (acide éthylmalonique) par une boite à moustache. C. Les spectres de masse obtenus pour 3 analyses répétées, considérées dans l'ordre d'injection (en début, en milieu et en fin de séquence analytique)..... 128

Figure 40 : Chromatogramme de l'ion extrait de l'acide muramique annoté dans le méconium en (A) et le spectre de masse correspondant en (B). NL indique l'intensité normalisée du pic de base, (de l'anglais Normalization Level). 131

Figure 41 : Spectres MS/MS de l'acide N-acétylaspartique pur (A) et dans l'extrait de méconium (B). NL indique l'intensité normalisée du pic de base (de l'anglais : Normalization Level) et NCE indique la valeur de l'énergie de collision normalisée (de l'anglais : Normalized Collision Energy). 132

Figure 42. Origines (A) et familles (B) des métabolites identifiés dans le méconium. 134

Figure 43 : Procédure suivie pour l'analyse métabolomique globale des échantillons de méconium de la cohorte clinique. 140

Figure 44 : Séquences analytiques réalisées dans les conditions LC/MS HILIC-ESI⁻ (a) et C18-ESI⁺ (b). 141

Figure 45 : Procédure suivie pour l'analyse métoprotéomique des échantillons de méconium de la cohorte clinique..... 144

Figure 46 : Procédure suivie pour l'analyse du microbiote des échantillons de méconium de la cohorte clinique. OTU (de l'anglais, Operational Taxonomic Unit), PCR (de l'anglais, Polymerase Chain Reaction), SILVA : base de données des gènes codant pour l'ARN ribosomique [285]. 148

Figure 47 : Somme des intensités des variables détectées et retenues après filtres en fonction de l'ordre d'injection des échantillons dans la séquence. Conditions HILIC-ESI⁻ en (A) et C18-ESI⁺ en (B). Les échantillons de QC sont représentés en rouge, les échantillons biologiques en vert, les blancs

analytiques en noir et les QC dilués en bleu (QC non dilué), orange (QC ½), jaune (QC ¼) et bordeaux (QC 1/8).	152
Figure 48 : Nombre de variables retrouvées (en bleu) parmi les 26 variables annotées au sein des échantillons A et B et manquantes au sein de la cohorte clinique. Nous faisons varier A. La tolérance en temps de rétention (+/- 45, 60, 90, 120 secondes ; avec un seuil d'intensité fixé à 30000) ou B. Le paramètre « seuil » d'intensité (0, 5000, 8000, 30000, avec une tolérance en temps de rétention fixée à +/- 45 secondes).	155
Figure 49.A. Répartition des 230 métabolites identifiés selon leurs familles. Le nombre ainsi que le pourcentage des métabolites au sein de chaque classe chimique sont indiqués. B. Origine des 230 métabolites identifiés.	157
Figure 50 : Vue d'ensemble des 25 voies métaboliques les plus enrichies classées selon leur p-values au sein des 230 métabolites identifiés dans le méconium.	159
Figure 51 : Représentation des échantillons sur les deux premières composantes de l'analyse en composantes principales réalisée sur l'ensemble des variables extraites et filtrées, correspondant respectivement à 6843 et 8555 variables dans les conditions analytiques HILIC-ESI ⁻ (A) et C18-ESI ⁺ (B). Les données sont centrées-réduites et transformées (transformation log ₁₀).	161
Figure 52 : Représentation des deux premières composantes de l'ACP réalisée sur l'ensemble des variables extraites et filtrées, correspondant respectivement à 6843 et 8555 variables dans les conditions analytiques HILIC-ESI ⁻ (A, C) et C18-ESI ⁺ (B, D). A et B : coloration en fonction des temps de prélèvements. Différents codes couleurs sont utilisés, correspondants à des temps de prélèvements continus représentés en haut à gauche de chaque figure (exprimé en minutes). C et D : coloration des échantillons en fonction du nouveau-né (de NB1 à NB11).	163
Figure 53 : Classification hiérarchique non supervisée des échantillons et des variables obtenues dans les conditions HILIC-ESI ⁻ (A) et C18-ESI ⁺ (B) (6843 et 8555 respectivement). Les échantillons de méconiums sont représentés en colonnes et les variables en lignes. La coloration indique si une variable est sous-représentée (du violet au bleu) ou sur-représentée (de l'orange au rouge) dans un échantillon donné, en comparaison à la moyenne de l'intensité de cette variable dans l'ensemble des échantillons. La couleur jaune indique des variables proches de la moyenne. Les valeurs des intensités des variables ont été au préalable centrées et réduites. Des exemples de la proximité d'échantillons appartenant à des temps de prélèvements proches sont encadrés en verts, et d'échantillons appartenant au même nouveau-né sont encadrés en rouge pour le NB7 et en bleu pour le NB9. Les échantillons extrêmes précédemment décrits sont marqués par une étoile rouge.	165
Figure 54 : A-C : Représentation des deux premières composantes de l'ACP réalisée sur les 230 variables uniques dont l'annotation a été confirmée par MS/MS . B. Coloration réalisée en fonction du temps de prélèvement (considéré comme une variable continue, code couleur en bas à gauche de la figure), et C. Coloration en fonction des individus (de NB1 à NB11, code couleur en bas à gauche de la figure). ...	166
Figure 55 : Classification hiérarchique non supervisée des échantillons et des 230 métabolites annotées dans les conditions HILIC-ESI ⁻ et C18-ESI ⁺ . Les échantillons sont représentés en colonnes (clusters 1 et 2) et les variables en lignes (clusters A et B). La coloration indique si une variable est sous-représentée (du violet au bleu) ou sur-représentée (de l'orange au rouge) dans un échantillon donné, en comparaison à la moyenne de l'intensité de cette variable dans l'ensemble des échantillons. La couleur jaune indique des variables proches de la moyenne. Les valeurs des intensités des variables ont été au préalable centrées et réduites. Les étoiles indiquent les deux échantillons qui n'appartiennent pas au seuil de 1440 minutes fixés au sein de chacun des clusters (rouge, cluster temps > 1440 minutes et en	

vert, temps < 1440 minutes). Les cadres (rouge, représente l'échantillon au temps de collecte le plus tardif : 4740 minutes, et en vert les échantillons appartenant au nouveau-né NB1).	168
Figure 56 : Représentation des deux premières composantes de l'analyse en composantes principales réalisée sur les 230 variables annotées en HILIC-ESI ⁻ et en C18-ESI ⁺ colorée selon le temps de prélèvement des échantillons. Les temps précoces (< 1440 minutes, groupe 1) sont représentés en bleu et les temps tardifs (> 1440 min, groupe 2) sont représentés en rouge.....	169
Figure 57 : Représentation en boîtes à moustaches de quelques exemples de variables significativement différentes entre les groupes 1 et 2. A. Métabolites « représentatifs » des temps tardifs (gp2). B. Métabolites « représentatifs » des temps précoces (gp1).....	170
Figure 58 : Combinaison des résultats obtenus par les analyses univariées (test de Wilcoxon) et multivariées (PLS-DA) réalisées sur l'ensemble des 230 variables identifiées. Les seuils de p-value = 0,05 et de VIP = 1,5 sont représentés par deux traits noirs vertical et horizontal respectivement. Les variables présentant une différence significative par les tests univariés entre les groupes « temps précoces » (< 1440 minutes) et « temps tardifs » (> 1440 minutes) regroupent les points représentés en vert et en rouge. Les variables représentées en vert sont également identifiées comme discriminantes dans l'analyse supervisées de type PLS-DA (VIP > 1,5). Le restant des variables en gris, ne présentant pas de caractère significativement différent entre les groupes analysés.	172
Figure 59 : Représentation du nombre de protéines humaines détectées au sein des échantillons de méconium analysés. Les échantillons appartenant au même nouveau-né sont représenté par la même couleur (de NB1 à NB11). Les pointillés rouges représentent la valeur moyenne du nombre de protéines détectées, et les tirets bleus représentent la médiane.	180
Figure 60 : Représentation des 15 protéines les plus abondantes parmi les 922 protéines humaines identifiées dans le méconium. Les protéines sont représentées dans leur ordre d'abondance relative (% NSAF).	181
Figure 61 : représentation des voies fonctionnelles les plus enrichies par les 79 protéines présentes chez l'ensemble des échantillons de méconium analysés.	183
Figure 62 : Représentation des deux premières composantes de l'ACP réalisée sur les 922 protéines humaines identifiées dans le méconium. A. Sans coloration. B. Coloration des échantillons selon le jour de prélèvement. En bleu sont représentés les échantillons collectés avant 1440 minutes suivant la naissance (J1), en vert les échantillons collectés entre 1440 et 2880 minutes (J2), et en rouge au-delà de 2880 minutes (J3). C. Coloration des échantillons selon le nouveau né (de NB1 à NB11).	186
Figure 63 : Classification hiérarchique des échantillons en partant de l'ensemble des protéines identifiées (922 protéines). Représentation des échantillons en colonnes (clusters 1, 2 et 3) et des protéines humaines identifiées en lignes (toutes les protéines ne sont pas écrites (à droite de la figure)).	187
Figure 64 : Classification hiérarchique des individus et des variables, en considérant les 10 protéines montrant le plus de variabilité dans le jeu de données. La coloration indique l'abondance relative des protéines dans les différents échantillons.	188
Figure 65 : Boxplots des protéines diminuant (A) <i>versus</i> augmentant (B) au cours du temps. En bleu sont représentés les échantillons excrétés le jour 1 (nombre d'échantillons = 19, nombre de nouveau-nés = 11), le jour 2 (nombre d'échantillons = 7, nombre de nouveau-nés = 6) et le jour 3 (nombre d'échantillons = 4, nombre de nouveau-nés = 4).....	189
Figure 66 : Représentation du taux des signaux de protéines bactériennes et humaines détectées au sein de l'ensemble de nos échantillons. En bleu est représenté le signal bactérien, en orange le signal	

humain. Les traits verticaux verts permettent de grouper les échantillons appartenant au même nouveau-né, afin de mieux les visualiser.	191
Figure 67 : Abondances relatives des différents phyla bactériens détectés dans les 30 échantillons de méconiums analysés. Les profils « blancs » correspondent aux échantillons où aucun signal bactérien n'est détecté.....	193
Figure 68 : Taux des TSMs attribués aux 4 phyla détectés dans le méconium, représentés par jour de prélèvement.	194
Figure 69 : Abondances relatives des différentes familles (A) et espèces bactériennes (B) détectées dans nos échantillons de méconium. Les échantillons pour lesquels l'espèce bactérienne n'a pas été validée sont représentés en blancs.....	195
Figure 70.A. Représentation du nombre d'OTU détectées au sein de chacun des échantillons de méconiums analysés (en colonnes). B. Vue d'ensemble des OTUs identifiées et de leurs abondances au sein des différents échantillons de méconium. L'échelle d'abondance représente le nombre de séquences attribuées pour chaque OTU.....	198
Figure 71. Représentation des abondances des phyla dans les différents échantillons de méconiums analysés. A. Représentation par individus et par temps croissant (de gauche vers la droite) ; B. Représentation des phyla bactériens au sein des échantillons regroupés par période de temps de prélèvements : C1 [0-600], C2 [600-1440], C3 [1440-2000] et C4 [2000-5000] minutes.	200
Figure 72 : Représentation des familles taxonomiques les plus abondantes au sein des différents phyla observés dans chacun des intervalles de temps considérés : C1 [0-600], C2 [600-1440], C3 [1440-2000] et C4 [2000-5000] minutes.	201
Figure 73: Evolution de la composition du microbiote intestinal dès la naissance, figure adaptée de Reinhardt et al [303].	206
Figure 74 : Normalisation du signal « LOESS » réalisé sur les échantillons de QC et les échantillons biologiques (cohorte initiale), analysés dans les conditions C18-ESI ⁺	217
Figure 75 : Comparaison du nombre de variables extraites par XCMS dans les conditions analytiques HILIC-ESI ⁻ (graphe de gauche) et C18-ESI ⁺ (graphe de droite), obtenues sur les deux échantillons A/B (rouge), sur les 33 échantillons de la cohorte clinique (vert), et sur les 257 échantillons de la cohorte EDEN (jaune).....	218
Figure 76 : Nombre de variables extraites lors du prétraitement des données dans XCMS, en fonction des valeurs de « minfrac » testées	219
Figure 77 : Nombre de variables extraites, filtrées, annotées avec redondance et sans redondance au sein de la cohorte clinique (A-B) et de la sous-cohorte initiale EDEN (C-D), dans les conditions analytiques HILIC-ESI ⁻ (panels de gauche) et C18-ESI ⁺ (panels de droite).	221
Figure 78 : Description des familles (A) et des origines (B) des métabolites annotés au sein des échantillons de méconium de la cohorte initiale EDEN, dans les deux conditions analytiques HILIC-ESI ⁻ et C18-ESI ⁺ . Les métabolites qui n'ont pas pu être attribués, ne sont pas représentés sur ces figures.	223
Figure 79 : Représentation des deux premières composantes de l'ACP réalisée sur les 4788 variables obtenues dans les conditions analytiques HILIC-ESI ⁻ (A). Les échantillons sont colorés selon leur jour de prélèvement : (B) : J1, J2, J3, J4, J5 et J6 ; (C) : J1 vs. > J1. Le jour de prélèvement de certains échantillons n'est pas renseigné (« nondefini »). (D) : le centre de prélèvement (A, Nancy en bleu et Poitiers en rouge) et (E) : le sexe du nouveau-né (B, fille en bleu, et garçon en rouge).	225
Figure 80 : Représentation des deux premières composantes de l'ACP réalisée sur les 4788 variables obtenues dans les conditions analytiques HILIC-ESI ⁻ . A. Représentation des échantillons collectés à J1	

dans les centres de Nancy (bleu) et de Poitiers (rouge). B-C : Représentation des échantillons avant ou après 24h à Nancy (B) et Poitiers (C). J1 (rouge), > J1 (bleu), les échantillons dont le temps de collecte n'est pas renseigné sont représentés en vert..... 226
 Figure 81 : Représentation des deux premières composantes de l'ACP réalisée sur les 301 variables annotées au sein des échantillons de la cohorte EDEN. A. Non colorée. B. Les échantillons sont colorés selon le jour de prélèvement : J1, J2, J3, J4, J5 et J6. C. Selon un seuil de temps de prélèvement à J1 (< ou > J1). D. Selon le centre de prélèvement (Poitiers et Nancy). E. Selon le sexe du nouveau-né. Les échantillons pour lesquels le jour de prélèvement n'a pas été renseigné sont notés « nondefini ». ... 227
 Figure 82 : Analyses discriminantes (PLS-DA) réalisées sur l'ensemble des échantillons de méconium de la cohorte initiale d'EDEN et l'ensemble des 4788 variables extraites dans les conditions HILIC-ESI⁻. 229
 Figure 83 : Représentation des deux premières composantes de l'ACP réalisée sur les 4604, 7369 variables extraites et filtrées dans les conditions analytiques HILIC-ESI⁻ (A et B) et C18-ESI⁺ (C et D) et sur les 301 variables annotées dans ces deux conditions analytiques (E et F), obtenues sur les 69 échantillons sélectionnés. Les représentations en B, C et F respectives, sont colorées en fonction du facteur « AA » (cas allergiques en bleu et non allergiques en rouge)..... 232
 Figure 84 : Représentation sous forme de boxplots de 5/42 variables dont les intensités sont significativement différentes entre les cas allergiques et les cas non allergiques (analyse univariée de type ttest avec correction de tests multiples). Les échantillons analysés proviennent du centre de Nancy, et ont été collectés dans les 36h post-partum < 36h. Les noms au dessus des figures représentent le nom des variables qui n'ont pas été annotées, appelées (**Mm/z**TRT), ainsi que la valeur corrigée de la « p value »..... 234

LISTE DES TABLEAUX

Tableau 1 : Comparaison des matrices fœtales, adaptée de Baerer et al [54].	48
Tableau 2 : Comparaison d'études réalisées sur l'analyse de la stérilité du méconium.	57
Tableau 3 : Vue comparative d'une sélection d'analyseurs utilisés pour des analyses globales en métabolomique, adaptée de Fenaille et al [209].	74
Tableau 4 : Tableau descriptif du choix de tests univariés à utiliser selon la modalité et la distribution du jeu de données.	85
Tableau 5. Tableau récapitulatif des principales études métabolomiques réalisées sur du méconium et/ou premières selles, menées par des approches ciblées ou non ciblées.	93
Tableau 6 : Description des échantillons de méconium A et B utilisés pour les phases d'optimisations de l'analyse métabolomique.	104
Tableau 7 : Codification et analyses réalisées sur les 35 échantillons de méconium collectés.	139
Tableau 8 : Outils utilisés dans le pré-traitement XCMS et l'annotation des données obtenues dans les conditions HILIC-ESI ⁻ au sein des chapitre I et II.	153
Tableau 9 : Liste des 9 métabolites retrouvés dans le méconium, non décrits dans les fèces de la base de données « Human Fecal Metabolome ».	158
Tableau 10 : Liste des 34 variables présentant une différence significative de leur intensité ($p < 0,05$, Test Wilcoxon) et identifiées comme discriminantes ($VIP > 1,5$; PLS-DA) entre les groupes d'échantillons collectés aux temps précoces ou aux temps tardifs (seuil de 1440 minutes).	173
Tableau 11 : Liste des 163 métabolites non significativement différents entre les temps de prélèvements précoces et tardifs (seuil de temps 1440 minutes).	175
Tableau 12 : Description de quelques fonctions biologiques attribuées aux protéines détectées dans nos échantillons de méconiums [136], [137], [140], [141].	184
Tableau 13 : Liste des 11 espèces bactériennes détectées dans le méconium.	196
Tableau 14 : Description des métabolites bactériens retrouvés au sein de nos échantillons de méconium, pouvant traduire la présence de certaines bactéries les produisant.	205
Tableau 15 : Description des nouveau-nés et des échantillons de méconiums étudiés, issus de la cohorte française mère-enfant EDEN.	214
Tableau 16 : Liste des composés annotés à travers la banque de données HMDB, ayant une p-value $< 0,05$ suite aux analyses univariées (test paramétrique ttest, avec correction de tests multiples) sur les échantillons prélevés à des temps < 36 heures dans le centre de Nancy.	233

LISTE DES ANNEXES

Annexe 1 : Publication « LC/HRMS-based metabolomics profiling of human meconium ».....	271
Annexe 2 : Publication « Basal Ti level in the human placenta and meconium and evidence of a maternofoetal transfer of food-grade TiO ₂ nanoparticles in an ex vivo placental perfusion model ».	288
Annexe 3 : Liste des étalons internes utilisés en analyses métabolomiques.	304
Annexe 4 : Paramètres XCMS version 3.0.0 de W4M utilisé pour le traitement des données métabolomiques dans les conditions HILIC-ESI ⁻ (Conditions chapitre I des résultats).	305
Annexe 5 : Chromatogramme extrait de l'ATP en (A) et spectre de masse correspondant en (B) extraits du composé adénosine triphosphate analysé dans les conditions HILIC-ESI ⁻	309
Annexe 6 : Liste des 137 métabolites identifiés au niveau (a, b) et (a, b, d) dans les échantillons de méconium A et B sur la colonne SeQuant ZIC-pHILIC dans le mode d'ionisation négatif (ESI ⁻) sur un Q-Exactive.	310
Annexe 7 : « Workflow » réunissant l'ensemble des procédures réalisées lors de l'extraction automatique des signaux (l'étape du XCMS).	323
Annexe 8 : Paramètres XCMS version 3.2.1 de W4M utilisée pour le traitement des données métabolomiques dans les conditions HILIC-ESI ⁻ (conditions chapitres II et III des résultats).	325
Annexe 9 : Paramètres XCMS version 3.2.1 de W4M utilisée pour le traitement des données métabolomiques dans les conditions C18-ESI ⁺ (conditions chapitres II et III des résultats).	332
Annexe 10 : Scripts utilisés pour les analyses en composantes principales et pour les classifications hiérarchiques réalisées en protéomique.	339
Annexe 11 : Liste des 149 métabolites identifiés au niveau (a,b), (a,c), (a,b,d) et (a,c,e) sur les échantillons de méconium de la cohorte clinique à partir des colonnes SeQuant ZIC-pHILIC et Hypersil GOLD C18 en mode d'ionisation négatif et positif respectivement sur un Q-Exactive	340
Annexe 12 : Liste des 67 variables discriminantes obtenues par une analyse univariée de type wilcoxon ayant une pvalue < 0,05 entre les temps de prélèvements précoces (< 1440 minutes (groupe 1 = gp1)) et tardifs (> 1440 minutes (groupe 2 = gp2)) au sein des échantillons de méconium de la cohorte clinique.	352
Annexe 13 : Liste des 922 protéines identifiées avec au moins deux peptides spécifiques (au sein de tous les échantillons), décrites par ordre d'abondance décroissant dans le méconium (NSAF [%] décroissant).	356
Annexe 14 : Liste des 79 protéines décrivant le « core protéome » identifiées au sein de l'ensemble des échantillons de méconium étudiés, incluant les temps de prélèvements précoces et tardifs (attribuées avec au moins 2 peptides spécifiques). Les protéines sont décrites par ordre d'abondance décroissant dans le méconium (NSAF [%] décroissant).....	388
Annexe 15 : Liste des 165 OTUs détectés au sein des 24 échantillons de méconiums analysés par l'approche métagénomique ciblée de l'ARN-16S.	391
Annexe 16 : Liste des 59 métabolites supplémentaires annotés au sein de la cohorte EDEN dans les conditions HILIC-ESI ⁻ et C18-ESI ⁺	398
Annexe 17 : Graphe des pR2Y et pQ2, obtenu lors de l'analyse des moindres carrés partiels (PLS-DA) de l'effet temps de collecte (J1, >J1) au sein des échantillons de la cohorte initiale EDEN, à partir des 301 variables annotées.	400
Annexe 18 : Liste des 46 métabolites discriminants entre les prélèvements d'échantillons de méconium de la cohorte EDEN réalisés au cours des premières 24 heures (< J1) et au-delà (> J1).	401
Annexe 19 : Liste des 38 métabolites discriminants au sein des échantillons des centres de Nancy et de Poitiers de la cohorte EDEN.	403

A mes parents, j'aimerais vous dédier ma thèse.

INTRODUCTION GENERALE

CHAPITRE I : LES ALLERGIES ALIMENTAIRES

I. Les allergies alimentaires : contexte général

Les réactions adverses aux aliments ont de multiples origines et sont classées en trois types de réactions : les réactions toxiques, les réactions psychologiques (aversion alimentaire) et les réactions non-toxiques (appelées hypersensibilités alimentaires) (Figure 2).

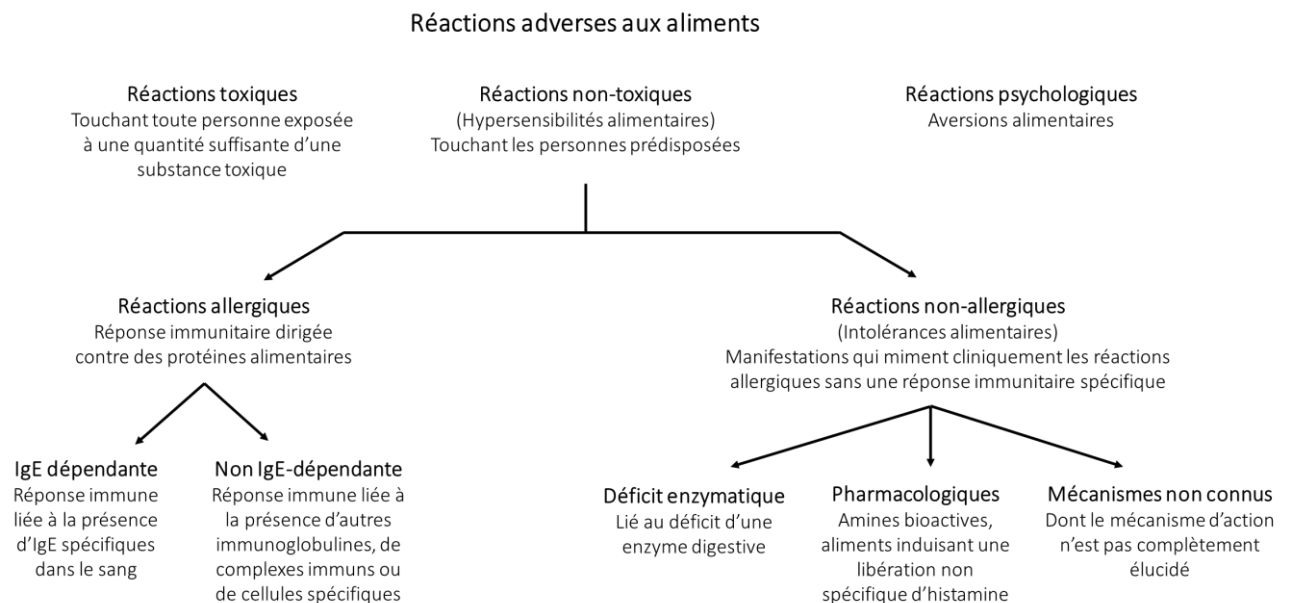


Figure 2 : Classification des réactions adverses aux aliments, adaptée d'une figure de Karine Adel-Patient.

Deux types de réactions sont décrites par l'Académie Européenne d'Allergie et d'Immunologie Clinique (EAACI) au sein des réactions non toxiques, incluant : les réactions allergiques et les réactions non-allergiques (dénommées intolérances alimentaires). Ces intolérances alimentaires ne font pas intervenir de réponses immunologiques spécifiques. Elles peuvent résulter d'un déficit enzymatique (intolérance au lactose par exemple), ou de mécanismes pharmacologiques (induits par des aliments contenant des composés histamino-libérateurs ou riches en histamine).

A l'inverse, les réactions allergiques sont un type de réaction adverse aux aliments qui met en jeu une réponse immunitaire adaptative et spécifique de l'aliment. Les allergies alimentaires (AA) correspondent ainsi à une réaction immunitaire induite contre des protéines alimentaires, qui peuvent être d'origine animale ou végétale. La réaction allergique induite se traduit par des symptômes pouvant affecter différents organes, tels que la peau, le tractus gastro-intestinal et les voies respiratoires [4]–[6].

Ce type de réaction survient chez les individus génétiquement prédisposés (atopiques), à chaque fois que ces derniers ingèrent l'aliment en cause.

Les allergies alimentaires peuvent résulter de différents mécanismes immunologiques : les allergies médiées par les immunoglobulines de type E (IgE) sont les plus fréquentes. On parle dans ce cas d'hypersensibilité allergique immédiate car ces allergies surviennent rapidement après l'ingestion de l'aliment incriminé. Il existe cependant d'autres mécanismes immunologiques conduisant à des délais variables d'apparition de symptômes. Ces réactions peuvent être complètement indépendantes des IgE, telles que le syndrome d'entérocolite induite par des protéines alimentaires (SEIPA). Elles peuvent également impliquer des réactions mixtes, telles que l'œsophagite à éosinophiles [7]. Nous nous intéresserons dans la suite de ce document aux réactions médiées par les IgE, étant les mieux connues.

II. Mécanismes immunologiques d'une réaction allergique médiée par des IgE

Les protéines alimentaires ne représentent normalement pas un danger pour l'organisme, ce sont au contraire des macronutriments nécessaires à la croissance et au maintien des fonctions de l'organisme. En conditions physiologiques normales, ces protéines alimentaires ne doivent pas induire de réponses immunitaires « défensives ». Ainsi, le système immunitaire apprend à tolérer ces protéines alimentaires, notamment par la mise en place d'une réponse immunitaire spécifique et répressive garantissant le maintien de l'homéostasie intestinale. Cependant, chez certains individus génétiquement prédisposés et dans certaines conditions d'expositions, l'introduction de ces mêmes aliments peut conduire à une AA. Le développement des AA dépendantes des IgE s'effectue en deux phases : une phase de sensibilisation asymptomatique et une phase de déclenchement de la réaction allergique (Figure 3).

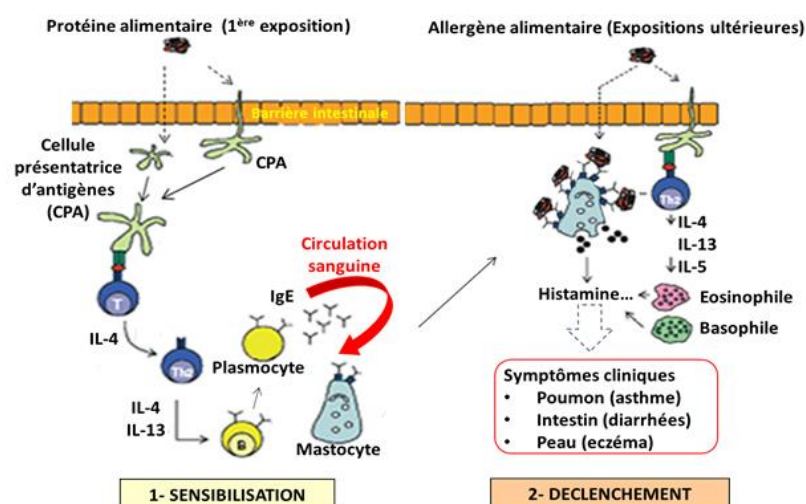


Figure 3 : Illustration des deux phases de l'hypersensibilité allergique IgE médiée, adaptée de Prioult et Nagler-Anderson [8].

A. La sensibilisation

Lors du premier contact de l'organisme avec l'antigène au niveau du système immunitaire associé à l'intestin, la réponse immunitaire spécifique s'oriente vers une voie effectrice dite de type « Th2 » au lieu de s'orienter vers la tolérance. Cette réponse adaptative Th2 se traduit par la sécrétion par les lymphocytes T activés, de différents médiateurs solubles (cytokines, par exemple IL-4, IL-5, IL-13), induisant la différenciation des lymphocytes B spécifiques en plasmocytes sécréteurs d'IgE spécifiques de la protéine alimentaire. Ces IgE, circulent dans l'organisme et se fixent sur leur récepteur de haute affinité (FcεRI) présents sur les cellules effectrices, c'est-à-dire les mastocytes muqueux et les basophiles circulants. La présence de ces IgE définit l'individu comme étant « sensibilisé » à l'allergène [8].

B. Déclenchement ou réaction allergique

Lors d'un second contact avec l'allergène, celui-ci est reconnu par les IgE présentes à la surface des cellules effectrices. Cette fixation conduit à l'activation de ces cellules et à la libération de différents médiateurs chimiques, tels que l'histamine et les leucotriènes, ainsi qu'à la production de cytokines de type Th2 amplifiant l'inflammation locale. Ces médiateurs sont responsables de l'apparition de symptômes.

L'allergie alimentaire IgE-médiée peut se traduire par différentes réactions, survenant dans les minutes jusqu'à 2 heures après l'ingestion de l'aliment. Ces réactions peuvent toucher un à plusieurs organes, de façon plus ou moins sévère. Les signes et symptômes des allergies alimentaires touchent dans 80 % des cas la peau, avec des manifestations telles que des démangeaisons, de l'eczéma ou de l'angio-œdème (au niveau du visage, des lèvres ou de la langue). Le système digestif est souvent touché, avec des manifestations de type vomissements ou diarrhées. Des manifestations au niveau du système respiratoire se traduiront par l'apparition d'asthme, de la toux ou des éternuements. Le terme anaphylaxie correspond à cette réaction allergique immédiate après le contact avec l'allergène. Elle peut être plus ou moins sévère, notamment selon le nombre d'organes touchés (grade 1 à 4). Le grade 4 correspond au choc anaphylactique (arrêt cardio-respiratoire ou asphyxie...). Ce type de réaction doit être instantanément traité par un stylo auto-injectable d'épinephrine) [1], [2].

III. Prévalence de l'allergie alimentaire

La prévalence des allergies alimentaires est en constante augmentation depuis une vingtaine d'années, touchant jusqu'à 10 % des enfants et 4 % des adultes dans les pays occidentaux [10], [11]. Cette augmentation suggère des modifications des interactions entre l'individu atopique et son environnement, notamment son alimentation (allaitement, pratiques de sevrage et de diversification,

consommation d'aliments ultra-transformés... etc), ses conditions d'hygiène (i.e. interaction avec les micro-organismes de l'environnement) et plus spécifiquement la composition et la fonction de ses microbiotes [12].

Selon l'OMS, les allergies représentent la quatrième cause de maladie dans le monde après le cancer, les maladies cardio-vasculaires et le Sida. La prévalence des allergies au sens large a doublé en cinq ans, et on estime que les allergies toucheront un français sur deux d'ici 2050. Ces allergies incluent les allergies respiratoires et médicamenteuses, les premières étant les plus fréquentes. Les études montrent que la prévalence de certaines allergies telles que la rhinite allergique ou l'asthme a tendance à se stabiliser au cours des dernières décennies. A l'inverse, les allergies alimentaires voient leur prévalence et leur sévérité augmenter depuis plus de 25 ans [9]. Ceci est notamment observé dans les sociétés industrialisées ou les zones plus urbaines des pays en voie de développement (Figure 4). En Australie, la prévalence des allergies alimentaires chez les enfants de 12 mois prouvée par des tests de provocation, est de 10 % [27]. En Europe, la prévalence de l'allergie alimentaire varie entre 1 et 5,6 % selon les pays [13].

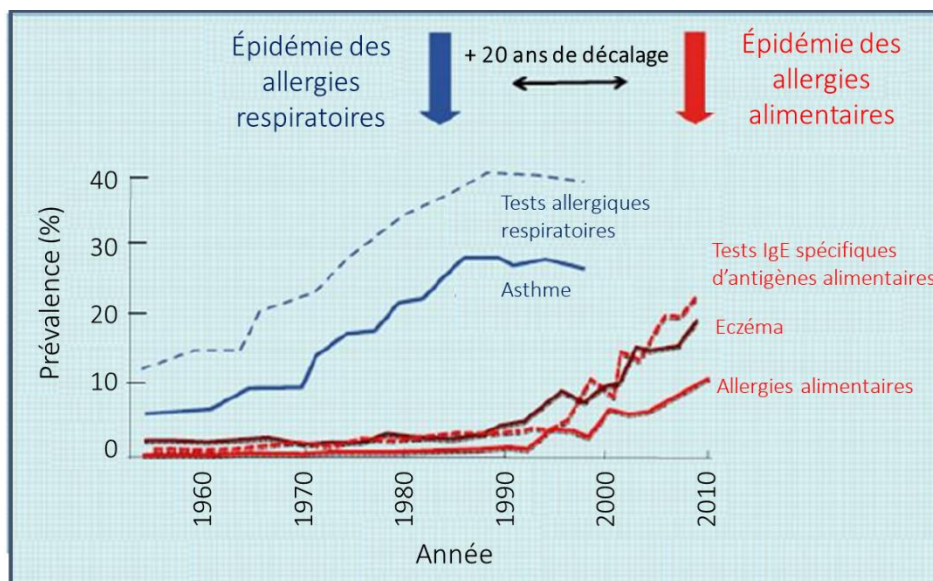


Figure 4 : Prévalence des allergies respiratoires et alimentaires au cours des dernières décennies (www.worldallergy.org).

IV. Principaux allergènes alimentaires et doses réactogènes

Les aliments incriminés dans une population donnée sont fonction des habitudes alimentaires de cette population : nous sommes allergiques aux aliments que nous consommons. Ainsi, chez le nourrisson non allaité, le lait de vache (constituant principal des préparations infantiles) est la source allergénique majeure. Chez le jeune enfant et l'adulte, une diversité protéique plus élevée est consommée, les

exposant à une plus grande variété d'allergènes alimentaires. Les principaux allergènes chez l'enfant sont le lait de vache, l'œuf, le poisson, l'arachide et les fruits à coques, tandis que les allergènes végétaux sont plus fréquemment impliqués chez l'adulte [14]. Par ailleurs, l'arachide, l'œuf, les crustacés ainsi que les fruits et légumes représentent les allergènes alimentaires les plus fréquents, responsables de près de 90 % des cas d'allergies alimentaires [15].

Chez l'adulte, les allergènes végétaux sont souvent responsables d'allergies alimentaires à symptomatologie légère, limitée à la sphère oro-pharyngée. Il s'agit dans ce cas d'allergies croisées entre des protéines végétales présentes dans les pollens et dans les fruits. L'exposition au pollen conduit à la production d'IgE spécifiques de certaines protéines du pollen (aéro-allergènes), et ces IgE peuvent ensuite reconnaître des protéines présentant des structures proches et présentes dans certains fruits. En général, les protéines en cause sont rapidement dégradées au cours de la digestion gastroduodénale, et les symptômes induits par cette reconnaissance se limitent à un prurit oro-pharyngé. Le cas classique de ces allergies « croisées » est le syndrome bouleau-pomme, avec une sensibilisation au pollen du bouleau et une réaction locale suite à la consommation d'une pomme. Cependant, de « vraies » allergies alimentaires surviennent aussi chez l'adulte, avec sensibilisation et réactions induites contre les protéines alimentaires, conduisant en général à une symptomatologie plus sévère (telles que l'arachide ou les crustacés) [16].

Un patient diagnostiqué « allergique alimentaire » devra suivre un régime d'éviction, c'est-à-dire ne plus consommer l'aliment incriminé. Cependant, des tests d'introduction reposant sur l'ingestion de quantité croissante de l'aliment incriminé (test de provocation oral), permettent de définir le seuil réactogène propre à chaque patient. Ceci permet de définir la quantité d'aliment ingéré nécessaire à l'apparition de symptômes cliniques. Ce seuil est important à connaître pour le patient, car il va dicter le niveau d'exclusion de son régime alimentaire. Ainsi, un patient pourra tolérer des traces d'allergènes, lui permettant de consommer les aliments étiquetés comme « pouvant contenir des traces de ». Il pourra également tolérer une ingestion régulière de petites doses d'allergènes, pouvant aider son système immunitaire à se réorienter vers la voie de la tolérance (Les régimes d'éviction stricte ne sont plus recommandés). Dans d'autres cas, mêmes ces petites doses ne seront pas tolérées et pourront être responsables de signes sévères.

Il n'existe pas de seuil « réactogène » absolu pour les allergènes alimentaires : la dose induisant une réaction allergique est variable en fonction de l'aliment et de l'individu. L'allergie à l'arachide représente l'une des formes les plus sévères des allergies alimentaires, où de très faibles doses (de l'ordre du microgramme) peuvent conduire au décès [17]. Pour les personnes allergiques aux crustacés, il leur suffit d'une crevette (de l'ordre de quelques grammes de protéines) pour induire une réaction allergique

[18]. Pour le lait, l'œuf ou le poisson, les doses réactogènes sont de l'ordre du milligramme de protéines [19]–[21]. D'autres seuils restent encore inconnus notamment pour la noisette et autres fruits à coque, les fruits à noyau ou encore le soja [1].

V. Facteurs associés à l'augmentation de la prévalence des allergies

Les allergies résultent d'interactions complexes entre l'hôte et son environnement, incluant notamment son alimentation. La mondialisation et la commercialisation de nouveaux aliments exposent le consommateur à des produits exotiques, mais aussi des modifications dans les modes de production des aliments (procédés technologiques, produits ultratransformés... etc), peuvent participer à l'augmentation des AA [12]. L'environnement de vie a un impact important sur la survenue des AA, notamment la pollution atmosphérique, mais aussi l'hygiène et l'environnement bactérien au sens large (environnement fermier *versus* urbain, hygiène poussée, antibiothérapie... etc). L'allergie alimentaire peut se développer dès les premiers mois de vie, suggérant des modifications précoces et « néfastes » des interactions de l'individu avec son environnement.

Les fenêtres précoces, incluant la période fœtale, néonatale et postnatale sont des périodes cruciales pour le développement du système immunitaire et des différentes barrières physiologiques (intestinale, cutanée et respiratoire). Une maturation incorrecte de ces systèmes peut prédisposer à des pathologies liées à des désordres immunitaires tels que les allergies alimentaires. Nous allons décrire dans la suite, l'ensemble des paramètres et facteurs pouvant impacter et influencer la maturation du système immunitaire durant ces trois fenêtres cruciales.

A. La fenêtre prénatale

La maturation du système immunitaire *in utero* peut être influencée par l'environnement auquel le fœtus est exposé, incluant différents facteurs liés à l'environnement de la mère. Ces facteurs comprennent des expositions extérieures (associées à l'alimentation ou au lieu de vie...), mais également des expositions internes de la mère (résultant de l'activité de ses microbiotes ou de son métabolisme endogène).

1. Le lieu de vie et expositions aux polluants atmosphériques

L'impact de la situation géographique durant la grossesse sur le développement du système immunitaire du nouveau-né a beaucoup été étudié. Baiz et al ont notamment réalisé une étude sur

l'impact des expositions à des polluants atmosphériques sur la santé de l'enfant. Deux polluants majeurs des milieux urbains ont été étudiés : le dioxyde d'azote (NO₂) et les particules en suspension de taille inférieure à 10 micromètres (PM10). La mesure de ces expositions a été réalisée à partir de stations de surveillance de la qualité de l'air proche du domicile. L'impact de ces expositions a été étudié sur 370 nouveau-nés, par la mesure des populations de lymphocytes jouant un rôle important dans la régulation des réponses immunitaires (lymphocytes « régulateurs » : CD4⁺ et CD25⁺), dans le sang du cordon ombilical. Les auteurs ont retrouvé des taux de ces lymphocytes plus faibles chez les mères les plus exposées aux NO₂ et PM10. Ces résultats montrent que les expositions atmosphériques durant la grossesse pourraient altérer le dynamisme immunitaire du nouveau-né, augmentant un risque pour l'enfant de développer des maladies liées à des dérèglements immunitaires telles que les allergies ou l'asthme [22]. Zhou et al ont étudié l'impact des expositions aux polluants atmosphériques et de la consommation de tabac durant la grossesse sur le développement d'asthme au cours de la première année de vie de l'enfant. L'étude des relations entre ces expositions et l'apparition de l'asthme a été réalisée à l'aide d'outils de modélisation, démontrant des associations entre le tabagisme maternel excessif et l'exposition aux polluants atmosphériques, avec le développement d'asthme chez les enfants à 1 an. La pollution atmosphérique liée au trafic a notamment montré un impact plus important sur le développement de l'asthme chez l'enfant [23].

2. L'alimentation de la mère durant la grossesse

Dans leur étude, Willers et al ont évalué l'impact de l'alimentation de la mère durant la grossesse sur le développement d'asthme ou de maladies atopiques. Cette étude a été réalisée sur 1253 enfants suivis jusqu'à leur 5 ans. Les auteurs n'ont pas observé de lien significatif entre la consommation des aliments analysés (fruits, légumes, produits laitiers, produits à base de céréales complètes...), et l'apparition de problèmes respiratoires ou d'allergies, à l'exception de la pomme et du poisson gras, qui montraient un effet protecteur contre le développement d'asthme ou de maladies allergiques chez l'enfant [24].

En raison d'absence de stimulants antigéniques *in utero*, le système immunitaire n'est pas totalement mature à la naissance [2]. Durant les premiers jours de vie, le nouveau-né va être exposé à un certain nombre d'antigènes nouveaux, tant alimentaires que microbiens. La capacité du nouveau-né à mettre en place des réponses adaptées à l'antigène rencontré va dépendre de la maturation de sa barrière intestinale, notamment sa perméabilité et du système immunitaire qui lui est associé, qui eux-même sont directement influencés par les antigènes rencontrés.

B. La fenêtre périnatale ou néonatale

Le microbiote intestinal de par sa composition et sa fonction, est un élément majeur pour la maturation du système immunitaire et de la barrière épithéliale intestinale du nouveau-né, tant dans son architecture que dans sa fonction. A la naissance, l'intestin est considéré comme stérile, la barrière intestinale est plus perméable, et les réponses immunitaires sont biaisées vers des réponses de type Th2 du fait de l'environnement placentaire. La colonisation du tractus intestinal par le microbiote va orchestrer le développement des systèmes immunitaires et adaptatifs. Cette colonisation est progressive, avec des bactéries aérobies puis anaérobies facultatives et finalement anaérobies strictes. La diversité bactérienne augmente au cours du temps du fait des expositions environnementales et de l'alimentation. Cette colonisation, sa cinétique et sa fonction sont impactées par des facteurs tels que le mode de naissance, des traitements antibiotiques en début de vie, et le type d'alimentation (allaitement *versus* formules infantiles) (Figure 5). Différentes études associant l'impact de ces différents facteurs sur le développement de pathologies telles que les allergies, mettent en évidence le rôle de la composition du microbiote intestinal sur l'homéostasie intestinale en début de vie. A titre d'exemple, plusieurs études ont démontré une moindre diversité microbienne chez des enfants ayant développé une allergie alimentaire *versus* des enfants sains, ainsi qu'une moindre richesse en *Bifidobacterium* et en *Enterococcus* comparativement aux enfants qui ne deviendront pas allergiques [40].

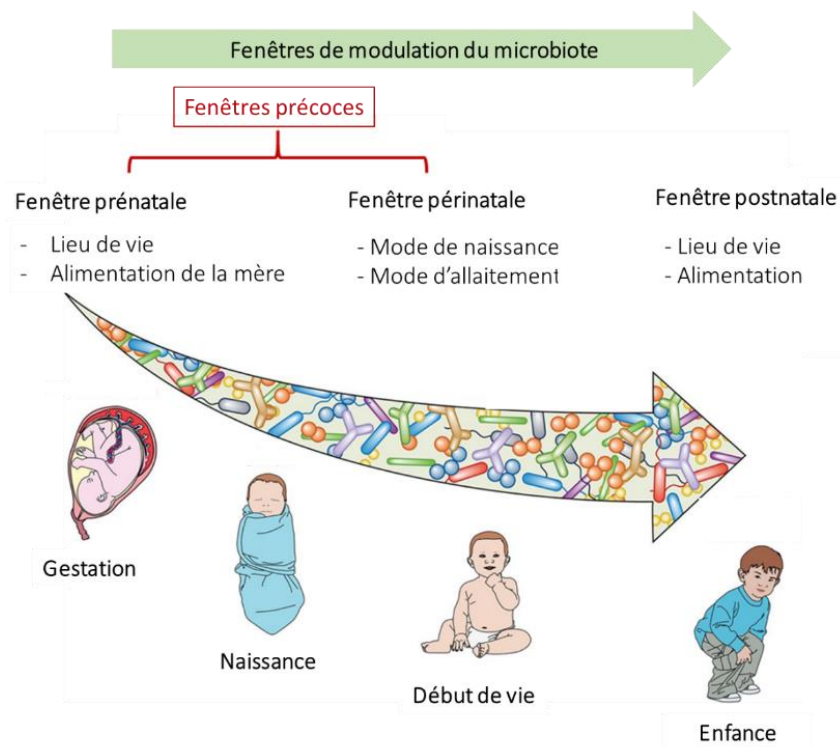


Figure 5 : Influence de différents facteurs pré, péri et post-natals sur la modulation du microbiote intestinal du nourrisson et de l'enfant plus tard dans la vie. Figure adaptée de Milani et al [2].

1. Le mode de naissance

La naissance par césarienne peut retarder voire modifier la colonisation microbienne de l'intestin du nouveau-né, un paramètre souvent relié à l'apparition de risques allergiques chez les enfants. Cette perturbation du microbiome (dysbiose) résulte d'une primocolonisation modifiée (absence de colonisation par le microbiote vaginal et fécal de la mère, et colonisation par le microbiote cutané [25]). L'utilisation périnatale d'antibiotiques, va également induire une perturbation du développement du système immunitaire du nourrisson. En effet, Hirsch et al ont indiqué qu'une consommation plus élevée d'antibiotiques chez les enfants est associée à de plus grands risques de développer une allergie [26]. Dans leur étude, Eggesbo et al ont évalué le mode de naissance par césarienne sur l'apparition d'allergies alimentaires chez les enfants. Cette étude, réalisée sur 2803 enfants, a permis de relier la naissance par césarienne au développement d'allergies alimentaires, principalement chez les enfants génétiquement prédisposés. Chez les enfants non-prédisposés, l'association a été beaucoup moins évidente [27]. Des résultats similaires ont également été publiés par Papathoma et al, vis-à-vis de l'augmentation du risque d'allergie chez les personnes atopiques nées par césarienne [28].

Stokholm et al ont observé que la composition du microbiote intestinal des naissances par césarienne est différent des naissances par voie basse. En effet, durant la première semaine et jusqu'au premier mois après la naissance, les enfants nés par césarienne ont de plus faibles abondances en Bacteroidetes et en Actinobactéries, et de plus fortes abondances en Firmicutes et en Protéobactéries, comparés aux enfants nés par voie basse. A l'âge d'un an, seules des différences mineures persistaient entre les microbiotes intestinaux au sein des deux modes de naissances. Les auteurs ont indiqué une probabilité accrue de développer de l'asthme durant les premières années de vie chez les naissances par césarienne, lorsque le microbiote intestinal de l'enfant à l'âge d'un an conservait encore la signature microbienne d'une césarienne. Ces résultats suggèrent qu'une maturation « saine et correcte » du microbiote intestinal lors de la première année de vie pouvait diminuer le risque de développer de l'asthme chez les enfants nés par césarienne [29].

2. Le mode d'alimentation du nourrisson en début de vie

Le lait maternel est l'aliment le plus adapté au nouveau-né, lui apportant tous les nutriments nécessaires pour sa croissance et son développement. Au-delà de cet aspect nutritif, l'allaitement a un effet protecteur bien établi contre les infections, mais également contre le développement de l'asthme et de certaines maladies allergiques chez des enfants génétiquement atopiques [30]–[32] [33], [34]. Ces effets protecteurs sont d'autant plus importants que la durée de l'allaitement est grande (jusqu'à au moins 4 mois) [32]. Les effets sur les allergies alimentaires sont moins clairement établis, probablement du fait

de fortes variabilités interindividuelles dans la composition du lait maternel, influencée par différents facteurs de la mère [35].

Une différence de composition du microbiote intestinal est observée entre des enfants allaités ou nourris par des formules infantiles [2], [36]. Chez les enfants allaités, une colonisation par du *Bifidobacterium* (notamment *Bifidobacterium infantis*), *Lactobacillus* et *Bacteroides* est favorisée. A l'inverse, chez les enfants nourris aux formules infantiles, le microbiote est dominé par des *Enterococcus* et *Clostridia* [37]. Ces effets reposent notamment sur la présence de très nombreux prébiotiques dans le lait maternel, qui n'ont pas d'équivalents dans le lait de vache (les oligosaccharides du lait maternel, ou HMO, de l'anglais Human Milk Oligosaccharides), ainsi que la présence d'un microbiote spécifique enrichissant celui du nouveau-né [38]–[40].

Au delà de son effet sur la composition du microbiote, le lait maternel (et principalement le colostrum : i.e. lait excrété pendant les 2 à 3 premiers jours après l'accouchement), contiennent différentes molécules bioactives (anticorps, cellules immunitaires, cytokines et nutriments immuno-modulateurs). Le colostrum participe d'une manière directe à la maturation du système immunitaire [35]. Des études précédentes menées au laboratoire ont montré que les concentrations de certains de ces composants sont variables selon des facteurs comme l'alimentation des mères pendant la grossesse ou la parité, et que la concentration de certaines de ces molécules bioactives est associée au développement de l'allergie alimentaire dans les premières années de vie [41].

Lorsque l'allaitement n'est pas possible, insuffisant ou non désiré, des formules infantiles peuvent être proposées. Ces formules sont composées pour la plupart de protéines de lait de vache, natives ou partiellement hydrolysées. Ces formules hypoallergéniques ont été recommandées par l'EAACI pour les enfants à risque atopique, dont au moins un parent ou un des membres de la fratrie a un historique allergique (allergie alimentaire, eczéma, dermatite atopique, asthme ou rhinite allergique). Cependant, l'efficacité préventive de ces préparations est remise en question par une méta-analyse récente [42]. En effet, l'utilisation de préparations infantiles partiellement hydrolysées à deux mois, a été récemment associée à un risque plus élevé de développer une allergie alimentaire durant les deux premières années de vie [43].

VI. Vers l'analyse de l'exposome périnatal en lien avec le développement d'allergies dans l'enfance

Les femmes enceintes sont sujettes à différentes expositions impactant le développement et la maturation du fœtus *in utero*, mais également son état de santé plus tard dans la vie [37]. Différents facteurs prénatals peuvent amplifier les effets de ces expositions utérines, et l'ensemble des fenêtres pré, péri et postnatales sont impliquées dans le développement et la santé future de l'enfant [1]. Parmi les facteurs présentés, l'alimentation et le lieu de vie de la mère, le mode de naissance ou encore le mode d'allaitement, peuvent être des facteurs de risque de développement d'allergies chez l'enfant.

Au sein du laboratoire, un précédent projet de thèse s'est intéressé aux différents composants du lait maternel. Cette étude a permis : i) d'apprécier la variabilité interindividuelle de la composition globale du lait maternel précoce (≤ 6 jours), ii) d'évaluer l'impact des facteurs environnementaux et maternels sur cette composition, puis iii) d'évaluer l'association entre cette composition et le développement d'allergies alimentaires au cours des 5 premières années de vie. Cette thèse, soutenue par Mikhaïl Berdi, a été réalisée sur des échantillons collectés au sein de la même cohorte de couples mères-enfants EDEN (**E**tude des **D**éterminants pré et post natals du développement et de la santé de l'**E**nfant), que la cohorte qui sera étudiée dans mon projet, cette fois-ci avec une emphase sur la fenêtre prénatale. L'objectif final de ces différents projets sera de combiner les données obtenues par les deux approches afin de caractériser l'exposome périnatal et son association avec le développement d'allergies alimentaires chez l'enfant.

Afin de réaliser cette étude, il est nécessaire de choisir la matrice biologique apportant le maximum d'information quant aux expositions transmises *in utero*, ceci durant la plus large période de la gestation possible. Dans le chapitre suivant, nous présentons et comparons les différentes matrices fœtales, ainsi que le choix réalisé en fonction des matrices disponibles dans le cadre de ce travail de thèse.

CHAPITRE II : ETUDE DE L'EXPOSOME FŒTAL AU TRAVERS DE L'ANALYSE DE LA COMPOSITION GLOBALE DU MECONIUM

I. Contexte général

« L'exposome » fœtal inclue différents facteurs d'expositions auxquels le fœtus est soumis durant la période gestationnelle [44], [45]. Ces expositions impliquent le lieu de vie et l'alimentation de la mère, mais également ses expositions professionnelles, médicamenteuses, ou encore sa consommation de produits tels que du tabac ou de l'alcool (Figure 6) [46], [47]. L'ensemble de ces expositions peut engendrer la transmission de substances externes appelées « xénobiotiques » au fœtus, suite à leur passage au travers de la barrière placentaire [37]. D'autres substances internes sont transmises au fœtus, résultant du métabolisme endogène de la mère, ainsi que du métabolisme de ses microbiotes. L'ensemble de ces expositions forment une partie de « l'exposome » fœtal. Afin d'analyser les substances transmises de la mère au fœtus durant la période gestationnelle, nous avons recherché l'ensemble des matrices fœtales préalablement étudiées dans la littérature. Les approches expérimentales utilisées pour approcher l'exposome prénatal seront ensuite présentées.

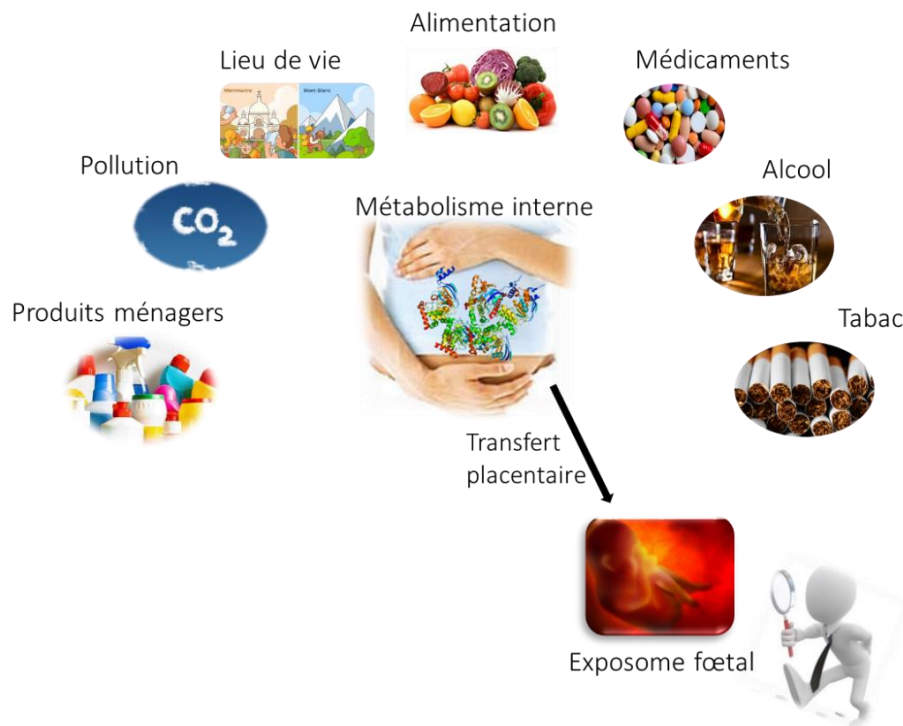


Figure 6 : Illustration des expositions fœtales et notion d'exposome foetal.

II. Choix de la matrice de l'étude

A. Description et comparaison des matrices fœtales étudiées dans la littérature

Afin de mener à bien notre étude, la matrice fœtale à considérer doit refléter la plus large fenêtre d'exposition possible. Elle doit préférentiellement être collectée de manière non invasive ainsi que démontrer une quantité suffisante permettant son analyse. Pour étudier les différents composés ayant traversé la barrière placentaire, des analyses des urines, des cheveux, des ongles ou encore des méconiums du nouveau-né ont été réalisées [48]–[50]. Différentes études se sont intéressées à l'analyse des expositions fœtales aux pesticides, aux xénobiotiques, aux drogues, au tabac ou à l'alcool, en se basant sur différentes matrices fœtales collectées après la naissance [47], [51]–[53]. Dans une revue, Bearer a notamment comparé l'ensemble de ces matrices, incluant l'urine, le sang, les cheveux, les ongles et le méconium du nouveau-né [54].

1. L'urine

L'urine fœtale peut être collectée à l'aide d'une lingette ou par une sonde. La première méthode est la plus utilisée car non invasive. Elle peut cependant poser des soucis de contamination notamment par les molécules provenant du support cellulosique. La technique de sonde est une méthode invasive, et est principalement utilisée en cas d'infections urinaires. D'un point de vue analytique, l'urine est une matrice présentant une forte richesse en métabolites grâce à l'effet de concentration des reins. La préparation des échantillons d'urines est relativement peu complexe : elle est soit réalisée par une microextraction des composés d'intérêt sur phase liquide ou solide [55], [56], ou par une simple dilution. L'urine du fœtus a déjà été étudiée pour quantifier l'exposition fœtale aux stupéfiants [57], [58]. Cette matrice ne rend cependant compte que d'une exposition récente de 2 à 3 jours, la rendant moins pertinente dans le cas de notre étude s'intéressant à la plus large fenêtre d'exposition fœtale.

2. Le sang

L'analyse du sang néonatal peut être réalisée sur : le sang du nouveau-né ou celui du cordon ombilical. Le prélèvement sanguin chez le nouveau-né est une technique invasive, avec des quantités très limitées de sang prélevées dû au faible volume sanguin du nouveau-né (250 mL). De plus les veines du nouveau-né sont très fines, rendant cette pratique très peu acceptée par les parents. Le prélèvement sanguin du cordon ombilical est quant à lui non-invasif mais nécessite d'être collecté lors de l'accouchement, ce qui complique son recueil. Comme l'urine, le sang ne reflète que les derniers jours d'exposition du fœtus, représentant une courte fenêtre.

3. Les cheveux

Les cheveux représentent une matrice accumulative, commençant à se produire dès le 6^{ème} mois de grossesse. Ainsi, cette matrice permettrait d'enregistrer une large fenêtre d'exposition fœtale contrairement à l'urine et au sang présentés précédemment. Cependant de nombreux nouveau-nés naissent sans cheveux, ce qui nécessite d'attendre que les cheveux poussent après la naissance avant de pouvoir les collecter, pouvant intégrer des éléments post-nataux. D'un point de vue analytique, la préparation des échantillons de cheveux est relativement simple et rapide. Celle-ci est réalisée par une digestion de la kératine puis une extraction sur phase solide ou liquide des composés d'intérêts [59]. Le cheveu a beaucoup été étudié notamment dans l'étude d'expositions fœtales à des contaminants [60], ou des stupéfiants [61].

4. Les ongles

Comme pour les cheveux, les ongles commencent à se former durant le dernier trimestre de gestation, enregistrant ainsi les expositions durant les 3 derniers mois de grossesse. Ils sont collectés quelques temps après la naissance, leur collecte est non invasive et leur préparation est simple mais peut être relativement longue (longue digestion de la kératine) [59]. Les ongles infantiles ont déjà été étudiés, notamment pour détecter des expositions fœtales aux stupéfiants [62], [63].

5. La salive

La salive est une matrice non invasive, facile à collecter, composée principalement d'eau et présentant une étape de préparation de l'échantillon simple et rapide. La salive présente néanmoins des quantités limitées, et plusieurs facteurs tels que le stress ou la prise de médicaments peuvent influencer sur son écoulement et ses caractéristiques physiologiques. La salive a souvent été étudiée dans la littérature en complément à d'autres matrices, telle que l'urine [64], [65].

6. Le méconium

Le méconium correspond aux premières matières fécales excrétées par le nouveau-né durant les 72 premières heures suivant la naissance [51], [66]. Il reflète la plus large fenêtre d'exposition fœtale en s'accumulant dans l'intestin du fœtus dès le 2^{ème} [67] ou 3^{ème} [68] mois de grossesse selon les références. L'échantillonnage du méconium est non invasif car collecté directement à partir de la couche. Celui-ci peut cependant présenter des contaminations par l'urine contenant de plus grandes

teneurs en métabolites, pouvant fausser les résultats des analyses fœtales. Le méconium est une matrice accumulative et disponible en quantités suffisantes pour réaliser plusieurs analyses. Il présente néanmoins une propriété très dense et visqueuse qui est complexe à analyser, et nécessite de longues étapes de préparation de l'échantillon, incluant la normalisation par un séchage, l'homogénéisation par des techniques telles que de la sonde à ultrasons... etc [46], [69].

7. Conclusion sur la matrice de l'étude

Baerer a cité dans sa revue les différents avantages et inconvénients des matrices fœtales décrites ci-dessus, exposant notamment la fenêtre d'exposition qu'elles permettent d'étudier, ainsi que la facilité ou complexité de collecte de chacune d'entre-elles [54]. Dans leur étude, Ostrea et al ont comparé la teneur en pesticides environnementaux, enregistrée dans le cheveu, dans le sang du cordon ombilical et dans le méconium de 598 couples mères/enfants. Cette étude indique de plus fortes teneurs détectées dans le méconium, du fait de son effet d'accumulation [48].

Nous décrivons et comparons l'ensemble des matrices fœtales dans le Tableau 1.

Tableau 1 : Comparaison des matrices fœtales, adaptée de Baerer et al [54].

Matrice fœtale	Avantages	Inconvénients
Urine	Forte concentration en métabolites Echantillon non invasif	Fenêtre d'exposition restreinte
Sang	Forte concentration en métabolites	Fenêtre d'exposition restreinte
Cheveux	Enregistre une large fenêtre d'exposition	Souvent absent à la naissance Contamination passive
Ongles	Facile à prélever Echantillon non invasif Enregistre une large fenêtre d'exposition	Contamination passive Préparation longue
Salive	Facile à prélever Echantillon non invasif Enregistre une large fenêtre d'exposition	Quantité limitée
Méconium	Facile à prélever Echantillon non invasif Enregistre une large fenêtre d'exposition	Analyse difficile

L'objectif de notre étude est d'apporter un maximum d'information sur « l'exposome » prénatal. La matrice enregistrant la plus large fenêtre d'exposition sera ainsi la plus pertinente pour nous, en l'occurrence le méconium. En effet, celui-ci apporte les informations les plus exhaustives sur la période prénatale [57]. Le paramètre d'accumulation ainsi que sa faible activité métabolique sont également intéressants, car beaucoup de substances sont présentes à de très faibles quantités et sont potentiellement rapidement dégradées. Il est de plus facile à collecter, tant du point de vue chronologique que du point de vue manipulation. Pour toutes ces raisons, nous avons choisi d'analyser cette matrice pour répondre à nos questions.

B. La matrice méconium

1. Formation du méconium

Le méconium est accumulé dans l'intestin du fœtus dès sa formation et jusqu'à la naissance [70]. Les différentes substances pouvant être retrouvées dans le méconium correspondent aux substances produites par le fœtus ainsi que toutes les substances transmises par la mère, traversant la barrière placentaire. Ce transfert s'effectue principalement par diffusion passive (transfert des molécules du milieu le plus concentré vers le milieu le moins concentré). Celui-ci dépend de plusieurs paramètres physico-chimiques des molécules considérées tels que la taille, le pKa, l'hydrophobicité, ou encore le pH sanguin [71], [72]. La formation du méconium inclue cinq paramètres représentés en Figure 7.A, et décrits ci-dessous :

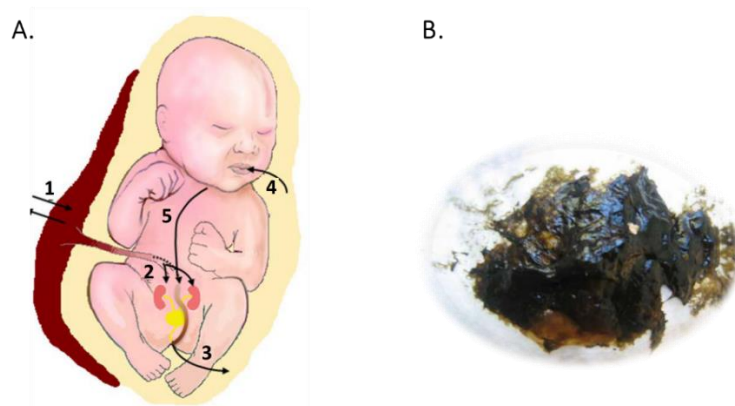


Figure 7 : A. Etapes de formation du méconium, adaptée de Koren et al [73]. B. Photo du méconium.

1. Le transfert des molécules de la mère vers le fœtus à travers le placenta *via* le cordon ombilical.
2. Les molécules transmises de la mère sont alors directement métabolisées par le foie puis accumulées dans le méconium, ou filtrées au niveau des reins et se retrouvent dans la vessie du fœtus.

3. Les molécules transférées vers la vessie vont être excrétées dans le liquide amniotique par l'urine.
4. Ces molécules excrétées *via* l'urine sont présentes dans le liquide amniotique, qui va être ingéré par le fœtus.
5. Ces molécules sont ensuite métabolisées par le foie et stockées à leur tour dans le méconium.

2. Composition du méconium

Le méconium présente une texture visqueuse, de couleur verdâtre ou brunâtre et est inodore (Figure 7.B). Selon les études, celui-ci est totalement expulsé après 24 à 48h [46], [47], 72h [51], [66] ou encore jusqu'au jour 5 suivant la naissance [74]. Ahanya et al ont rapporté que l'excrétion du méconium chez des nouveau-nés sains se fait généralement au cours des 24 à 48 heures après la naissance [75]. Le méconium est parfois excrété *in utero* en cas de stress important ou d'infections fœtales. L'évènement de naissance est également considéré comme un processus stressant pour le fœtus pouvant également impacter l'excrétion du méconium [75]. La composition de cette matrice a été décrite par Unsworth et Vause [67], incluant :

- 70 à 80 % d'eau
- Des protéines
- Des métabolites
- Des lipides
- Des acides biliaires
- Des minéraux
- Des sécrétions intestinales
- Des cheveux néonataux
- Des résidus de liquide amniotique

C. Etude de la composition du méconium par des approches omiques

De plus en plus d'études sont réalisées sur le méconium, avec environ 1000 publications en 2019 *versus* moins de 800 publications en 2010 (mots clés recherchés « meconium » et « newborn faeces », ScienceDirect, Août 2020). Les analyses réalisées incluent aussi bien l'étude d'exposition à des métaux [76], des nanoparticules [77], des pesticides [69], [78], des composés organiques toxiques (par exemple les « BTEX » : Benzène, Toluène, Éthylbenzène et Xylènes) [47], des médicaments [68], du tabac [53], [79], [80], ou encore des drogues ou de l'alcool [74], [81].

Dans le cadre de ce travail de thèse, nous avons pour objectif d'étudier de façon non ciblée, et la plus exhaustive possible la composition globale du méconium. Ceci en intégrant l'analyse de toutes ses petites molécules organiques (métabolites) et de ses protéines respectivement par des approches dites de métabolomique et de protéomique. Des analyses des espèces bactériennes présentes dans le méconium seront également réalisées après extraction de l'ADN et séquençage du gène de l'ARN ribosomal 16S (ARNr16S), spécifique des bactéries. Nous présentons dans ce chapitre, les différentes approches « omiques », suivies par un état de l'art des études réalisées sur la composition du microbiote et des protéines présentes dans le méconium. Enfin, l'approche métabolomique sera présentée individuellement dans le dernier chapitre de l'introduction (partie III), correspondant à l'approche sur laquelle j'ai le plus travaillé durant ce travail de thèse.

Les approches « omiques » permettent de caractériser voire de quantifier l'ensemble des composés biologiques impliqués dans la structure et le fonctionnement d'un organisme donné. L'ensemble des gènes a été appelé « génome » pour la première fois par Hans Winkler en 1920 [82]. Le séquençage de l'ADN et l'identification des gènes a été appelée « génomique » par Mckusick et Ruddle en 1987 [83]. Ce concept a été étendu aux analyses des autres biomolécules donnant lieu à la « transcriptomique » s'intéressant à l'étude des ARNm, à la « protéomique » étudiant les protéines et à la « métabolomique » étudiant les petites molécules de masses moléculaires inférieures à 1500 Dalton. Ces petites molécules appelées métabolites peuvent être issus du métabolisme interne d'un individu, mais peuvent également intégrer les interactions de l'individu avec son environnement (métabolites issus des microbiotes vivant en symbiose avec l'individu, ou encore des métabolites exogènes liés à l'alimentation ou à des expositions à des xénobiotiques).

Les analyses omiques offrent de nouveaux moyens d'étude de l'évolution adaptative à l'échelle des populations [84]. Elles permettent de mieux comprendre le processus vis-à-vis d'une pathologie ou d'un contexte physiologique ou environnemental donné. Les sciences « omiques » se sont développées grâce à l'essor des techniques de biologie et de chimie à haut débit [85], [86], incluant des techniques de séquençage ou encore d'analyse par spectrométrie de masse haute résolution [87], [88]. Ces approches peuvent être réalisées soit de façon globale, non ciblée et sans *a priori*, soit à moindre échelle, dans le but d'acquérir une vision ciblée sur une famille de gènes, de protéines ou de métabolites d'intérêts. Dans ce cas de figure, ces composés doivent être préalablement identifiés par des approches non ciblées ou par des hypothèses mécanistiques.

1. Les approches génomique et métagénomique - application à l'étude du microbiote intestinal

a) Analyse génomique : du procaryote à l'eucaryote

L'analyse génomique étudie l'ensemble des gènes d'un organisme ou d'un tissu [89]. La classification des organismes est caractérisée en 3 sections : les eucaryotes, les archées et les bactéries, et a été introduite pour la première fois en 1977 par Woese et Fox (Figure 8) [90]. Ces organismes appartiennent à deux familles : les eucaryotes, et les procaryotes (incluant les bactéries et archées).



Figure 8 : Arbre phylogénétique adapté de Ciccarelli et al. La section verte représente les archées, en rouge les eucaryotes et en bleu les bactéries. Dans chaque catégorie, les subdivisions les plus fréquemment utilisées sont représentées [91].

La famille des eucaryotes possède des cellules avec un noyau, qui contient l'information génétique (ADN). Cette famille compte les animaux, les plantes, les micro-algues, les champignons ainsi que les levures. De l'autre côté, la famille des procaryotes est caractérisée par l'absence de noyau et donc la présence de l'ADN dans le cytoplasme de la cellule. Cette famille est majoritairement représentée par les bactéries.

b) Techniques d'analyses génomiques et métagénomiques

L'analyse génomique concerne l'analyse d'un génome (par exemple, analyse du génome humain). Lorsque les génomes de plusieurs individus d'espèces différentes (par exemple, un ensemble de

communauté bactérienne) sont considérés, l'approche utilisée est appelée « métagénomique ». C'est cette approche qui est notamment utilisée pour caractériser les quelques 10^{14} bactéries et autres procaryotes qui peuplent notre intestin. Les analyses métagénomiques reposent sur une étape de purification puis d'amplification de l'ADN, suivie par une technique de séquençage exécutée pour la première fois par Sanger et al en 1977 [122]. Le premier séquençage complet réalisé est celui de la bactérie *Haemophilus influenzae*, par Fleischmann et al en 1995. Les bactéries représentent les organismes avec le plus grand nombre de génomes séquencés. A l'inverse, les organismes eucaryotes présentent des séquences non codantes, c'est-à-dire des séquences d'ADN non traduites et ont par conséquent été moins séquencés (séquences beaucoup plus longues) [111]. La technique « expressed sequence tag » (EST) a été initialement utilisée pour le décryptage du séquençage chez les eucaryotes, représentant uniquement la partie codante de l'ADN. A la suite des avancées en termes de techniques de séquençage, une nouvelle approche dite « Next Generation Sequencing » (NGS) a été utilisée, offrant une meilleure qualité et un plus haut débit de séquençage [105], [123], [124].

Il existe au sein du génome des régions variables à hypervariables et d'autres constantes, l'ensemble de cette information génétique est extraite et amplifiée pour être séquencée. Le séquençage est réalisé sur des fragments d'ADN de courtes longueurs, qui sont ensuite comparés à des génomes de référence [105]. Cette étape présente la plus grosse difficulté en génomique car peu d'espèces ont leur génome séquencé et donc connu à ce jour.

Deux régions sont d'intérêt pour l'analyse des gènes procaryotes et eucaryotes, les gènes codant pour l'ARNr16S et l'ARNr18S respectivement et spécifiques à chacun de ces 2 organismes [2]. Ainsi, la région 16S représente une région taxonomique, c'est-à-dire un référentiel permettant de remonter à la famille, au genre ou encore à la souche de la bactérie étudiée (Figure 9). Ceci est réalisé par comparaison des séquences à des bases de données disponibles, telles que Greengenes (GG) [127], SILVA [128], RDP (de l'anglais, Ribosomal Database Project) [129], [130]. Le séquençage de l'ensemble des gènes codants pour les ARNr16S, associé à des outils bioinformatiques, permet ainsi d'étudier la composition complexe du microbiote.

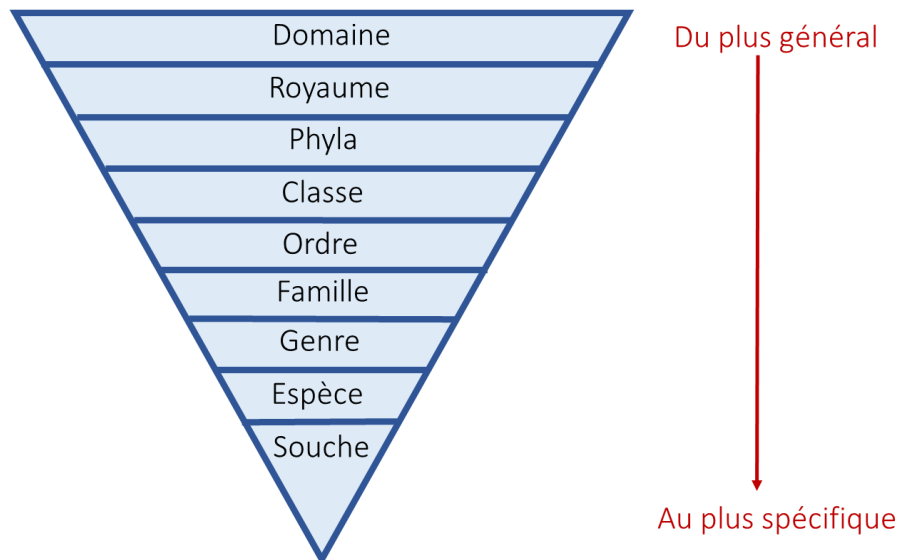


Figure 9 : Classification des organismes biologiques et représentation des différents liens hiérarchiques.

c) Analyse métagénomique des microbiotes humains

c.1) Les microbiotes humains

Le corps humain contient une grande diversité microbienne associée aux différentes muqueuses et surfaces de l'organisme, avec qui il vit en symbiose (Figure 10.A) [92]. Les cellules bactériennes sont 10 à 100 fois supérieures en nombre par rapport à l'ensemble des cellules constituant le corps humain [93]. En effet, le génome humain est composé d'environ 23000 gènes alors que le nombre de gènes des microorganismes retrouvés dans l'intestin se compte en millions [90], [94]. Cette composition en microorganismes assure de nombreuses fonctions essentielles au maintien de notre santé [95]–[97]. En effet, l'absence de diversité microbienne est souvent associée à l'apparition de maladies chroniques, telles que l'obésité [98], le diabète [96] ou encore des allergies telles que l'asthme [99], [100].

Parmi les neuf régions les plus colonisées en bactéries dans le corps adulte, on retrouve le microbiote intestinal. Celui-ci contient la plus grande densité microbienne avec plus de 10^{11} bactéries par gramme de matière fécale [93], représentant environ 1 kg de notre masse corporelle totale [94]. Leur densité suit un gradient oro-anal représenté en Figure 10.B. Bien que sa composition reste spécifique à chaque individu, les Firmicutes et les Bacteroidetes représentent à elles seules environ 90 % des populations bactériennes chez tous les individus [101]. On retrouve ensuite les Actinobactéries, représentant environ 10 % des populations bactériennes, et moins de 2 % du microbiote est composé de *Lactobacillus*, de *Streptococcus* ou d'*Enterobacter* [93], [102]. Un microorganisme est considéré comme dominant lorsque sa densité est supérieure à 10^7 bactéries/g de contenu. La colonisation de l'intestin

en microorganismes dominants reste globalement stable dans le temps pour un individu donné et retrouve naturellement son état initial après une perturbation donnée (phénomène de résilience, suite à des traitements antibiotiques par exemple) [93], [94]. Les populations sous-dominantes ayant des densités inférieures à 10^7 bactéries/g de contenu sont généralement en transit et leur durée de présence dans l'intestin peut varier d'une souche à une autre, comme pour certaines souches bactériennes probiotiques.

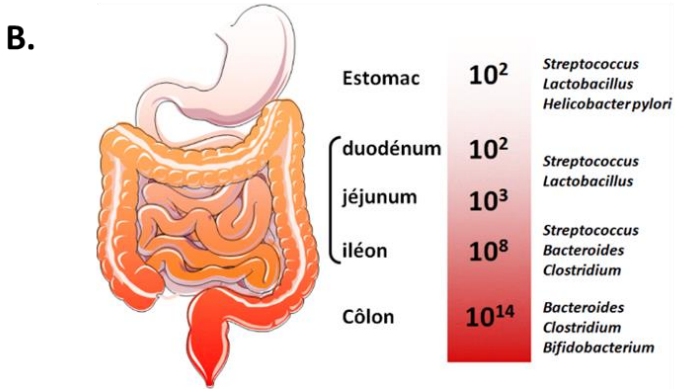
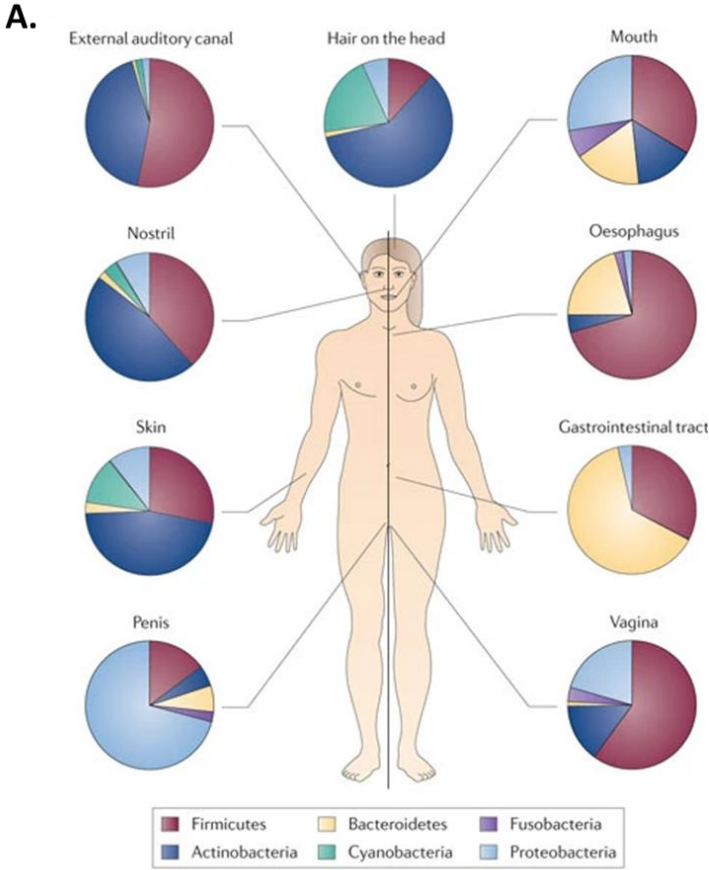


Figure 10 : A. Communautés microbiennes présentes à différentes localisations du corps humain [92]. B. Composition des communautés bactériennes dominantes le long du tube digestif [103].

c.2) Analyse du microbiote du méconium

L'environnement du fœtus humain a durant des décennies été considéré comme étant stérile [104], [105]. L'absence de bactéries vivantes est appuyée par l'utilisation de méthodes traditionnelles de culture ou de microscopie, offrant quelques limites notamment vis-à-vis de l'incapacité de détecter des bactéries non cultivables [2]. Des études récentes utilisant de nouvelles techniques de pointe ont remis en question cette stérilité, discutant un début de développement du microbiote humain *in utero* [104], [106], même si les démonstrations fournies restent très controversées.

En effet, différents auteurs ont démontré la présence de bactéries telles que des *Staphylococcus* ou des *Enterobacteriaceae* dans de nombreuses matrices fœtales telles que le placenta, le cordon ombilical, le liquide amniotique ou encore le méconium [107]–[110]. La présence de microorganismes dans ces matrices fœtales est considérée comme une preuve soutenant l'hypothèse de la colonisation microbienne *in utero* [111], [112]. Dans leur étude, Hansen et al ont réalisé l'analyse par la méthode FISH (de l'anglais, Fluorescent In-Situ Hybridisation) de 31 échantillons de méconium excrétés dans les 24 heures suivant la naissance, démontrant une présence bactérienne dans 67 % des échantillons [113]. Les familles bactériennes les plus fréquentes étaient les *Bifidobacteriaceae*, les *Enterobacteriaceae*, et les *Enterococcaceae* [113]. Dans une autre étude réalisée par Hu et al sur 23 échantillons de méconium de nouveau-nés sains issus de naissances par voie basse ou par césarienne, les auteurs ont observé une présence bactérienne au sein de tous les échantillons de méconium, incluant des Bacteroidetes et des Protéobactéries [114]. Plusieurs études réalisées sur des échantillons de méconium issus de nouveau-nés sains ont également démontré la présence d'ADN issus de différentes bactéries [114]–[118], suggérant une transmission bactérienne de la mère vers le fœtus et un début de colonisation microbienne intestinale avant la naissance [2]. Dans une autre étude de méconiums collectés chez des naissances prématurées, les auteurs ont décrit une plus forte colonisation en *Lactobacillus* chez les nouveau-nés très prématurés (< 32 semaines de gestation), comparés aux naissances à terme [39]. Ardissonne et al ont observé de plus fortes abondances des genres *Enterobacter*, *Enterococcus* et *Lactobacillus* dans le méconium de nouveau-nés prématurés, qui pourraient être responsables de réactions inflammatoires au niveau de l'intestin [110].

A l'inverse, d'autres études suggèrent une contamination des matières fécales au moment de la naissance. Makino et al ont considéré des échantillons de méconium et de fèces chez 17 couples mères-enfants en bonne santé et issus de naissances par voie basse *versus* des naissances par césarienne. Cette étude a démontré la présence de souches appartenant au genre *Bifidobacterium* chez 11 des 12 méconiums des naissances par voie basse, mais pas dans les accouchements par césarienne. Ces

résultats suggèrent ainsi une contamination microbienne liée au mode d'accouchement, et démontrent la transmission de souches uniques et spécifiques du microbiote de la mère vers l'enfant, lors de la naissance par voie basse [117]. Quelques exemples d'études réalisées sur la stérilité ou non du méconium sont regroupés dans le Tableau 2 ci-dessous.

Tableau 2 : Comparaison d'études réalisées sur l'analyse de la stérilité du méconium.

Nombre d'échantillons et temps de prélèvement	Méthode d'analyse	Résultats obtenus	Auteurs
52 échantillons (< 48 heures)	Séquençage ARNr 16S	65 % des méconiums sont non stériles	Ardissonne et al [110]
23 échantillons (< 48 heures)	Séquençage ARNr 16S	Méconium non stérile Grande abondance en Proteobactéries Faible abondance en Bacteroidetes	Hu et al [114]
31 échantillons (< 24 heures)	Séquençage ARNr 16S	66 % des méconiums sont non stériles	Hansen et al [113]
21 échantillons (< 2 heures)	Culture et séquençage ARNr 16S	Méconium non stérile Espèces prédominantes : <i>Staphylococcus</i> et <i>E.Coli</i>	Jiménez et al [104]

De nombreuses études dans la littérature supportent l'hypothèse de colonisation microbienne *in utero*, à travers l'analyse du méconium ou d'autres matrices fœtales, telles que le cordon ombilical ou le placenta [2].

L'analyse métagénomique ciblée (ADN 16S) réalisée dans ce travail de thèse vise à identifier les organismes bactériens présents dans le méconium. Ceux-ci pouvant impacter dès la vie *in utero* la maturation de l'épithélium intestinal et de son système immunitaire associé, qu'ils soient vivants et métaboliquement actifs ou non. L'approche métagénomique ne permet en effet pas de garantir la présence de bactéries vivantes dans les échantillons.

Cette approche est réalisée en collaboration avec l'institut Micalis de l'INRAE de Jouy en Josas, avec Claire Cherbuy et Aurélia Bruneau, spécialisées dans le domaine de l'étude du microbiote.

2. L'approche protéomique

Les protéines présentent des propriétés physico-chimiques diverses et jouent un rôle primordial dans le fonctionnement des cellules et de l'organisme. Elles contribuent à des fonctions structurales (par exemple, les protéines du cytosquelette), au transport de substances dans les différents fluides (les hémoglobines dans le sang), à la motricité (les protéines responsables de la contraction musculaire), ou encore au sein du système immunitaire (les anticorps ou les cytokines). Les protéines font ainsi un lien direct entre l'information génétique et la (patho)physiologie des organismes [119]. L'étude de l'ensemble de ces constituants forme le domaine de la « protéomique ».

La protéomique est apparue au début des années 1990 contribuant pour beaucoup dans l'analyse de milieux complexes [120], [121]. Elle tire son émergence des avancées en techniques séparatives des mélanges protéiques et peptidiques complexes (telles que l'électrophorèse ou la chromatographie liquide), en analyse par spectrométrie de masse, en séquençage et en construction de bases de données de séquences nucléiques et protéiques [1]. L'objectif des approches protéomiques fut d'abord d'identifier les différentes protéines présentes dans un milieu biologique, mais elles ont ensuite connues de nombreuses applications telles que la caractérisation structurale des protéines, l'étude de leurs interactions dans l'organisme mais également de leur localisation et leur quantification dans des cellules ou des tissus [122], [123].

a) Les différentes approches utilisées en analyse protéomique

L'analyse protéomique est réalisée selon deux approches méthodologiques principales : « bottom-up » ou « top-down » respectivement, approches ascendante ou descendante.

L'approche « bottom-up », aussi appelée « shotgun proteomics » est l'approche la plus utilisée. Elle consiste à digérer les protéines (en mélange ou préalablement fractionnées) en fragments peptidiques avant de les analyser par spectrométrie de masse [124]. La digestion des protéines se fait à l'aide d'enzymes, souvent la trypsine dont la spécificité pour les résidus lysine et arginine est bien établie. Les fragments peptidiques sont ensuite analysés généralement par LC-(ESI)MS/MS, et l'identification des protéines est réalisée en utilisant des bases de données permettant d'attribuer les séquences peptidiques aux protéines. Un taux de recouvrement est obtenu à cette étape, permettant d'apprécier la pertinence des identifications. L'approche « bottom-up » offre beaucoup d'avantages, avec notamment une couverture plus complète des protéomes [122]. A l'inverse l'approche « top-down » consiste à analyser par spectrométrie de masse un mélange protéique ou une fraction protéique sans recourir à la protéolyse [125]. L'avantage majeur de cette technique est qu'elle permet de caractériser les protéoformes d'une protéine sans ambiguïté, à l'inverse de l'approche « bottom-up » pour laquelle

la présence de peptides modifiés au niveau post-traductionnel peut rendre l'identification des protéines difficile. Cette approche n'est cependant pas la plus utilisée du fait du manque de répétabilité de la séparation des protéines en amont de l'analyse [126].

b) Méthodologie suivie en analyse protéomique

b.1) Techniques analytiques utilisées en protéomique

Afin de décomplexifier le protéome des systèmes biologiques étudiés, une séparation est généralement réalisée au préalable, soit par migration sur un gel SDS-PAGE (électrophorèse capillaire en gel de polyacrylamide contenant du laurylsulfate de sodium), mono ou bidimensionnel, soit par une chromatographie en phase liquide (Figure 11). La séparation monodimensionnelle des protéines est réalisée selon leur poids moléculaire. La séparation bidimensionnelle sur gel implique une première séparation basée sur une focalisation isoélectrique et une seconde séparation basée sur le poids moléculaire. Dans les deux cas, les gels sont colorés et les spots protéiques d'intérêts excisés. La digestion des protéines est ensuite réalisée dans le gel, suivie par l'analyse des peptides générés par spectrométrie de masse (avec au préalable une éventuelle séparation par chromatographie liquide). La séparation sur gel est intéressante dans le cas de l'analyse de matrices biologiquement complexes, mais n'est pas systématiquement réalisée. Une digestion des protéines peut être directement réalisée à partir de l'échantillon primaire, suivie par une séparation des peptides par chromatographie liquide et une analyse par spectrométrie de masse [127].

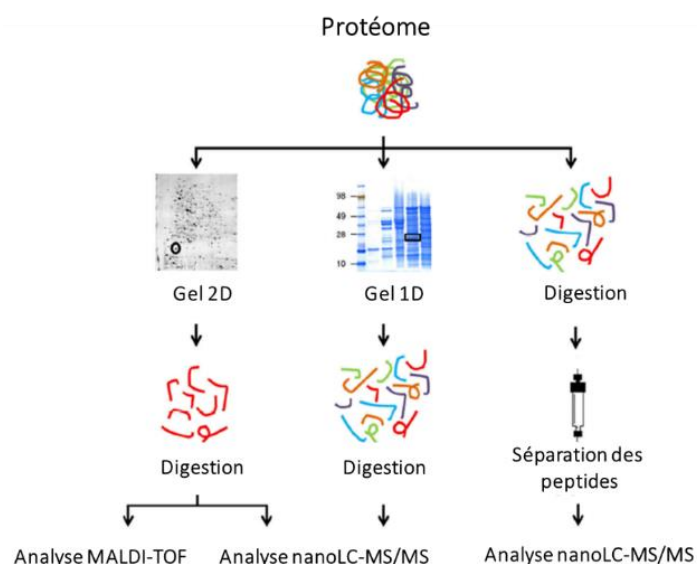


Figure 11 : Techniques analytiques utilisées dans l'étude du protéome, adaptée de Armengaud et Bland [127].

b.2) Analyse des données de protéomique et identification des protéines

Face au volume des données produites par les analyses protéomiques, il existe aujourd'hui des logiciels et outils automatisés permettant différents prétraitements (correction des intensités, réduction du bruit, filtre des données...) tels que Mascot, Xalign [128], Sequest [129] ou encore Maxquant [130]. L'attribution des séquences peptidiques est réalisée par homologie à des bases de données de séquences protéiques telles que Swissprot [131] ou PRIDE [132]. Lors de cette étape, il est conseillé d'utiliser des bases de données spécifiques aux protéines d'intérêts (par exemple, des protéines humaines), permettant de diminuer le taux de faux positifs [133]. L'annotation est réalisée en prenant en considération différents paramètres tels que l'enzyme utilisée, la tolérance en modification de certains acides aminés... etc. A chaque spectre MS/MS obtenu (explicité plus tard dans le chapitre III de l'introduction) pour un peptide candidat, est attribué un score permettant de considérer son taux d'homologie avec la base de donnée utilisée [134]. Plus le taux de couverture de la protéine est élevé, plus l'identification est de confiance [135]. Les peptides identifiés peuvent être attribués à plus d'une protéine simultanément (cas des protéines homologues). Ainsi une protéine est considérée comme identifiée lorsqu'elle est représentée par un ou plusieurs peptides spécifiques à elle.

c) Analyses métaprotéomiques du méconium

Différentes études ont analysé la composition protéique du méconium et les premières selles. Lisowska-Myjak et al ont réalisé une analyse protéomique sur 50 échantillons de méconiums issus de 10 nouveau-nés. Quatre à cinq prélèvements sont collectés par nouveau-né dès la naissance et jusqu'à 52 heures post-partum [136]. A l'aide d'une approche « bottom-up » par LC-MS/MS, les auteurs ont détecté 946 protéines, dont 430 sont représentées par au moins deux peptides spécifiques. Parmi les protéines majoritaires, les auteurs ont décrit les enzymes α -1-antitrypsine (A1AT) et aminopeptidase-N ou encore la neprilysine. De nombreuses protéines parmi les 946 identifiées ont par la suite été définies comme étant impliquées dans la fonction immunitaire « innée », se développant *in utero*, telles que les immunoglobulines (Ig *kappa*, *lambda*...) et des protéines dérivant de neutrophiles (lactotransferrines et les protéines S100-A8/A9). D'autres types de protéines ont été identifiées, incluant l'uromoduline, impliquée dans la défense contre les infections urinaires, ou encore l'attractine, impliquée dans les réactions immunitaires agissant durant l'inflammation. Des analyses supplémentaires permettront d'améliorer la compréhension du rôle de chacune de ces protéines dans les processus normaux et pathologiques et dans le développement fœtal [137].

Dans la continuité de leurs travaux, Lisowska-Myjak et al ont également réalisé une analyse quantitative d'une des protéines majoritaires du méconium, la α -antitrypsine. L'objectif de l'étude est de comparer l'activité et la concentration de cette protéine pendant la vie intra et extra utérine, par une analyse longitudinale de méconiums. Cette étude a été réalisée sur 118 échantillons de méconiums excrétés entre les jours 1 et 4 suivant la naissance chez 10 nouveau-nés sains. Les auteurs décrivent une diminution systématique de l' α -antitrypsine (antiprotéase active déposée dans l'intestin du fœtus). Le prélèvement en série de fèces pendant 2 à 4 jours peut fournir un reflet de la dynamique de l'augmentation postnatale des fonctions exocrines du pancréas du nouveau-né [138].

Au sein d'une autre étude, les auteurs ont quantifié quatre protéines définissant des biomarqueurs de l'environnement prénatal, permettant d'évaluer l'inflammation/l'homéostasie intestinale, soit : la calprotectine, la lactoferrine, la myéloperoxydase et l'élastase. L'analyse quantitative de ces protéines a été réalisée sur 81 échantillons de méconiums issus de 20 nouveau-nés sains. Les résultats ont démontré des concentrations protéiques bien supérieures dans les méconiums que dans les fèces présentant une inflammation intestinale (exemple, le syndrome du côlon irritable [139]). La calprotectine a notamment été retrouvée à des concentrations 19 fois plus élevées dans le méconium (286,5 $\mu\text{g/g}$ *versus* 15 $\mu\text{g/g}$ dans les fèces). Les auteurs ont suggéré que ces protéines dérivées notamment de neutrophiles, pourraient être utilisées comme marqueurs pour évaluer l'état de maturation et d'activité du système immunitaire inné à la fin de la période fœtale [140]. De même, Laforgia et al ont défini des taux plus élevés en calprotectine chez des enfants prématurés par rapport à des naissances à terme, suggérant une immaturité de la muqueuse intestinale lors des naissances prématurées [141].

Dans le cadre de ce travail de thèse, nous nous intéresserons à la composition en protéines du méconium et à l'impact des expositions fœtales sur cette composition. L'approche protéomique a été effectuée en collaboration avec le LI2D du CEA de Marcoule (Laboratoire Innovation pour la Détection et le Diagnostic), impliquant Jean Armengaud, Lucia Grenga et Guylaine Miotello, spécialisés dans le domaine.

CHAPITRE III : L'ANALYSE METABOLOMIQUE

I. Contexte général

Le métabolome qualifie l'ensemble des composés organiques de faibles poids moléculaires, inférieurs à 1500 Dalton. La métabolomique est une science relativement jeune, introduite dans les années 1990-2000 par Fiehn et Nicholson [142], [143]. Le terme métabolomique a été introduit par Oliver [144] et le terme métabonomique par Nicholson [143], à partir de la fin des années 1990. Ces termes ont respectivement été utilisés dans la réalisation du profilage global des métabolites présents dans une matrice biologique donnée, et dans la quantification d'une réponse métabolique suite à une perturbation connue. Aujourd'hui le terme métabolomique est celui qui fait référence. La métabolomique s'est initialement intéressée à étudier le contexte pharmacologique et toxicologique du métabolome des espèces vivantes, dans le but d'évaluer le plus finement possible les changements phénotypiques. La métabolomique connaît une grande expansion depuis une dizaine d'années suite à l'évolution des techniques analytiques haut débit mais également au développement d'outils informatiques performants [88], [145], [146].

L'approche génomique permet d'identifier et d'expliquer un certain nombre de prédispositions génétiques. De l'autre côté, la métabolomique décrit le reflet phénotypique le plus détaillé de la réponse de l'organisme dans un environnement donné ou suite à une perturbation (état pathophysiologique, exposition à des xénobiotiques, changement d'environnement...). Cette approche symbolise la dernière étape de la cascade des sciences « omiques » (Figure 12), et sera détaillée dans ce chapitre.

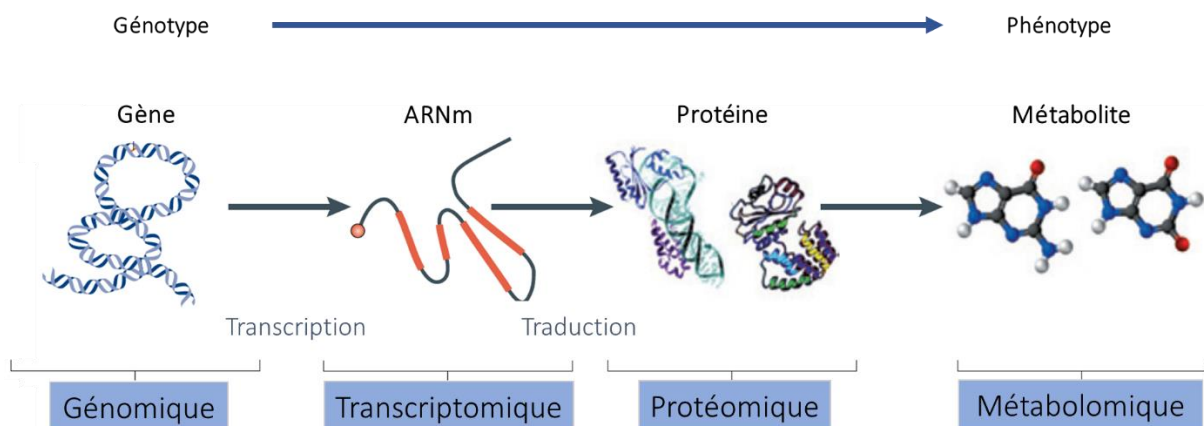


Figure 12 : Représentation de la cascade des sciences « omiques », adaptée de Patti et al [147].

Le métabolome inclut les substances naturellement produites par le métabolisme de l'hôte, appelées « métabolites endogènes », incluant les aminoacides, les lipides, les nucléotides, les carbohydrates... [85]. Le métabolome inclut également des substances exogènes apportées par l'environnement au sens large : alimentation, contaminants, médicaments... [148]. L'étude de l'ensemble de ces métabolites aux propriétés physico-chimiques très variées, est réalisée soit par une approche quantitative et ciblée d'un sous-ensemble spécifique de métabolites [149], ou par une approche globale, non ciblée, dans le but de suivre un maximum de métabolites dans une matrice biologique donnée [150], [151]. L'analyse métabolomique non ciblée vise à obtenir une vue métabolique globale et sans « *a priori* » d'un système biologique donné. Elle est un outil prometteur pour le diagnostic de multiples pathologies en combinant diverses disciplines : la chimie analytique, la biologie et la bioinformatique. Cette approche permet d'accéder au plus grand nombre possible de métabolites permettant de connecter différentes voies métaboliques et ainsi d'expliquer des perturbations et des désordres pathologiques [152]–[154].

La limite de l'analyse par profilage métabolomique est liée au volume des données générées nécessitant des outils informatiques performants, robustes et automatisés pour leur traitement et leur interprétation [155]–[157]. L'approche métabolomique non ciblée ne permet pas de quantifier les métabolites de manière absolue mais uniquement de façon relative. Cependant, une approche ciblée quantitative peut être envisagée par la suite pour un sous-ensemble de métabolites d'intérêts [158]. La quantification est dite « relative » lorsqu'elle est réalisée en comparant l'intensité des signaux entre eux en l'absence de composé(s) de référence. Celle-ci est alors dite « absolue » lorsqu'une concentration est définie à l'aide de standards ou d'isotopes marqués. L'objectif de cette approche est souvent de confirmer la présence d'une perturbation métabolique, ou de quantifier des métabolites biomarqueurs relatifs à des pathologies [159]–[161]. De plus, l'approche ciblée représente une méthode plus sensible comparativement à l'approche non ciblée, grâce à l'utilisation de méthodes d'extractions spécifiques et des méthodes analytiques adaptées aux métabolites recherchés.

Dans ce projet de thèse, une approche métabolomique non ciblée a été utilisée, permettant de cartographier de façon la plus exhaustive possible les composants du méconium, une matrice très peu décrite dans la littérature par cette approche.

II. L'approche non ciblée en métabolomique

A. Déroulement d'une analyse métabolomique non ciblée

L'analyse métabolomique se compose de différentes étapes allant de la préparation de l'échantillon à l'identification des métabolites et leur interprétation biologique au sein du système étudié. Le déroulement de l'analyse non ciblée par LC/HRMS, telle que réalisée au laboratoire est décrite en Figure 13 [162]. Celle-ci inclut six étapes essentielles et capitales, tant pour la quantité que pour la qualité des données générées. En effet, chacune de ces étapes peut induire un biais pouvant directement impacter les résultats et fausser l'interprétation biologique finale. La première étape consiste en la préparation de l'échantillon, permettant l'extraction des métabolites aux propriétés physico-chimiques variées, suivie d'une étape d'acquisition des données spectrales par LC/HRMS. Le prétraitement des données jusqu'à l'annotation des variables (étape C) est réalisé grâce à des outils automatisés. Des outils de visualisation (permettant par exemple d'identifier des individus ou des variables abberants) sont utilisés, puis des analyses statistiques sont conduites permettant d'identifier les variables différenciant les groupes étudiés. Des analyses MS/MS sont réalisées sur les composés d'intérêts permettant de valider les annotations putatives obtenues à l'étape C. Enfin une interprétation biologique des données est réalisée en dernier lieu, par exemple par la construction de réseaux métaboliques.

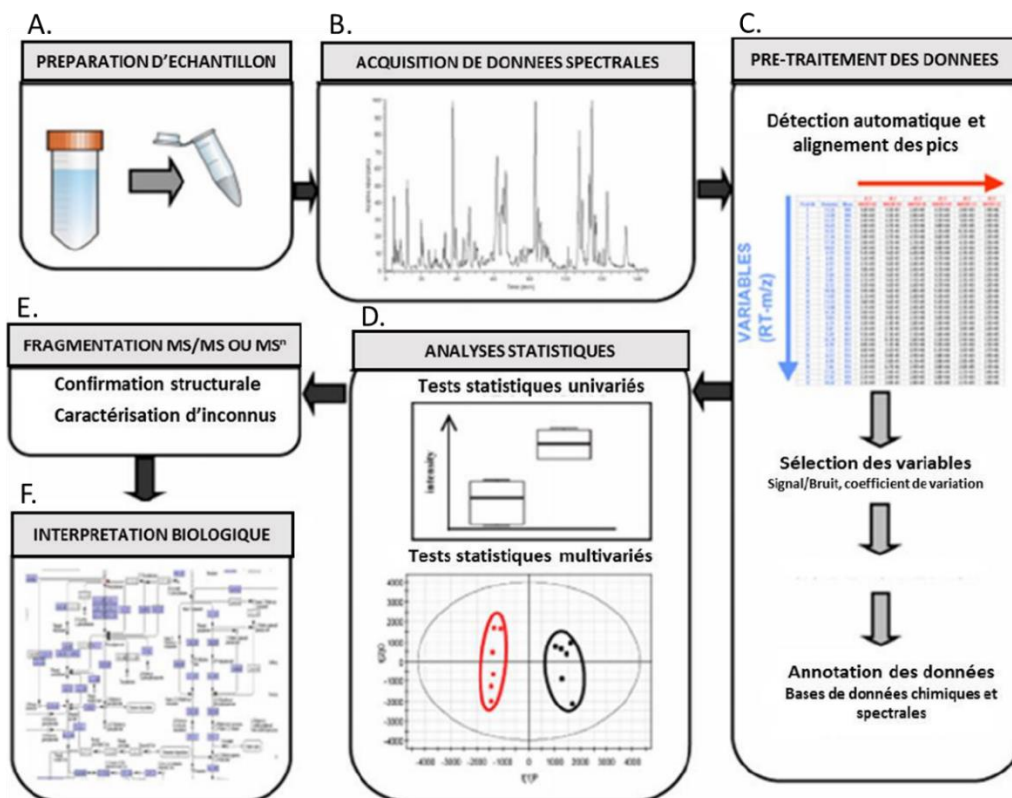


Figure 13 : Déroulement d'une analyse métabolomique globale par LC/HRMS, adapté de Junot et al [162].

1. Préparation des échantillons

La première étape du « workflow » de l'analyse métabolomique consiste à extraire l'ensemble du métabolome en un minimum d'étapes. Celle-ci représente une étape primordiale, conditionnant la couverture des métabolites extraits. Elle doit répondre à différents paramètres de normalisation (entre les échantillons) et de reproductibilité. Cette étape doit également assurer une extraction non-sélective, permettant de retenir un maximum d'information [163]. La méthode de préparation de l'échantillon est très dépendante de la matrice étudiée. Dans des fluides biologiques comme le sérum ou le plasma, l'extraction métabolique est principalement réalisée par précipitation des protéines à l'aide d'un solvant organique, souvent le méthanol, puis récupération du surnageant contenant les métabolites. L'extraction des métabolites au méthanol permet d'avoir une couverture métabolique complète et est très répétable [164]–[166]. Il existe néanmoins d'autres types d'échantillons plus complexes, visqueux ou solides tels que des fèces, des tissus ou d'organes biologiques nécessitant une homogénéisation préalable à l'aide d'ultrasons ou de lyse mécanique aidant à l'extraction métabolique de la matrice biologique.

2. Acquisition des données

Il existe non seulement une grande diversité physico-chimique des métabolites (poids moléculaire, hydrophobicité, valeurs de pKa...), mais également une grande variabilité de leur gamme de concentration dans les matrices biologiques. En effet, au sein d'une même matrice les concentrations en métabolites peuvent varier du nano au millimolaire, rendant l'analyse simultanée des métabolites très difficile.

Différentes techniques analytiques peuvent être utilisées en métabolomique : la chromatographie bidimensionnelle sur couche mince (CCM-2D) [167], la spectroscopie infrarouge [144], la résonance magnétique nucléaire (RMN) [168], ou encore la spectrométrie de masse [162]. Les techniques analytiques permettant l'élucidation structurale des métabolites font parties des techniques les plus utilisées actuellement, incluant la (RMN) et la spectrométrie de masse (MS). A travers ces deux techniques respectives, plus de 700 et 2500 publications ont été réalisées au cours de l'année 2019. La spectrométrie de masse devenant prépondérante à partir de 2010 (Figure 14).

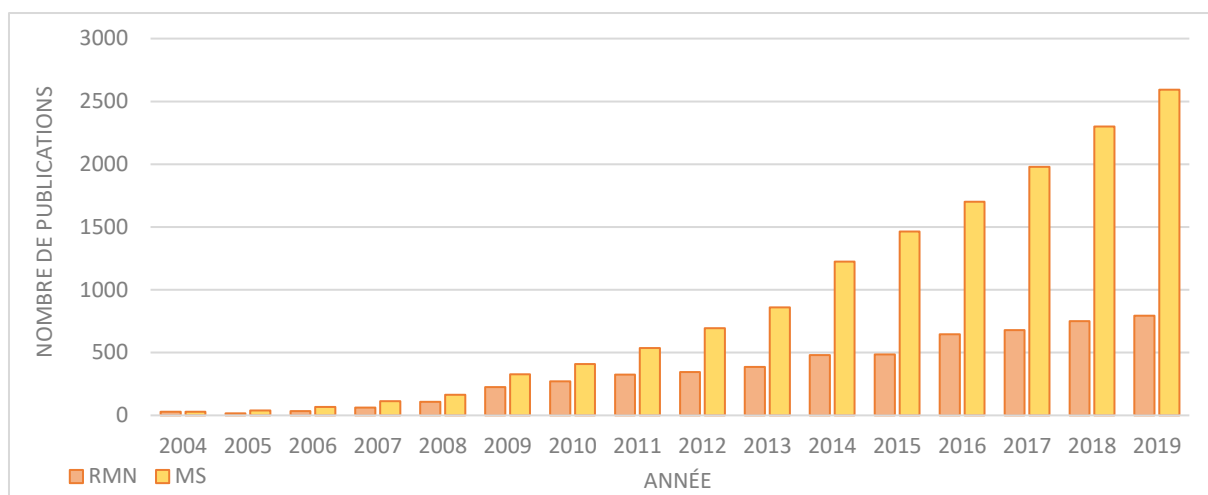


Figure 14 : Nombre de publications en métabolomique reposant sur l'utilisation de la RMN et de la spectrométrie de masse (MS) au cours des 15 dernières années [169].

a) L'analyse métabolomique par RMN

La RMN est une technique exploitant les propriétés magnétiques de certains noyaux atomiques possédant un spin nucléaire non nul (exemple : ^1H , ^{13}C , ^{31}P ...). L'application d'un champ magnétique à ces noyaux va aligner leurs spins de façon parallèle ou antiparallèle au champ appliqué. Le retour du système à l'équilibre par un phénomène de relaxation va alors émettre un champ de radiofréquence qui sera enregistré. L'utilisation de la transformée de Fourier va permettre de convertir ces fréquences en spectres qui pourront être interprétés. Ces spectres vont définir, pour le proton pris comme exemple, des valeurs de déplacements chimiques de tous les protons de la molécule étudiée. Le couplage de chaque proton avec son voisinage va définir un déplacement chimique qui permettra à l'aide d'une bibliothèque, d'identifier son voisinage et ainsi de remonter à la structure moléculaire. L'analyse des spectres de différents noyaux atomiques (^1H , ^{13}C , ^{15}N , ^{31}P ...) est complémentaire et est souvent utilisée pour l'élucidation structurale de molécules complexes telles que des protéines, des sucres... etc.

La RMN a été l'une des premières technologies utilisées pour le profilage métabolomique [168]. En présence de matrices complexes, la RMN à deux dimensions (RMN de corrélation homo ou hétéro-nucléaire) peut également être envisagée, permettant de remédier aux problèmes de chevauchements des signaux. Le couplage bidimensionnel apporte une résolution supplémentaire, facilitant l'identification des métabolites [170]. La RMN est une technique nécessitant peu d'étapes de préparation de l'échantillon, et est non destructive. Sa grande reproductibilité est intéressante en particulier dans l'analyse de larges cohortes, rendant les résultats facilement comparables entre « batchs » analytiques ou entre laboratoires. La RMN est de plus une technique quantitative, où l'intensité des pics est directement proportionnelle à la concentration relative des métabolites. Une

quantification absolue peut être envisagée en injectant dans la matrice étudiée une quantité connue de composés de référence. La limite majeure de la technique RMN demeure dans sa faible sensibilité.

b) L'analyse métabolomique par LC-MS et par GC-MS

La technique de spectrométrie de masse à haute résolution est devenue la technique de référence pour les analyses métabolomiques, du fait de sa grande sensibilité, sélectivité et spécificité. Elle se base sur la mesure du ratio masse/charge (m/z) des ions moléculaires et de leurs fragments permettant de déterminer leur identité. Différentes méthodes d'introduction de l'échantillon dans le spectromètre de masse sont possibles. Une introduction dite directe DIMS (de l'anglais, Direct Introduction Mass Spectrometry) [171], permettant de réaliser des analyses rapides (< 5 minutes) et est facile à mettre en place [172]. Un autre type d'introduction faisant intervenir une séparation en amont de l'analyse par spectromètre de masse, inclut soit une séparation par électrophorèse capillaire [173], par chromatographie en phase gazeuse (GC) [174], [175] ou par chromatographie en phase liquide (LC) [176]

b.1) Le couplage avec les méthodes séparatives

b.1.a) La CE/MS

L'électrophorèse capillaire est une technique séparative utilisée en couplage à la spectrométrie de masse dans l'analyse globale des métabolites [173]. Elle est basée sur la séparation des composés polaires et chargés suivant leur rapport taille sur charge [177], [178]. Elle présente un réel atout résolutif pour certains métabolites difficiles à distinguer par des systèmes chromatographiques. L'électrophorèse capillaire est cependant peu reproductible et ne permet que de faibles volumes d'analyses, représentant les limites majeures de cette technique [179].

b.1.b) La GC/MS

Le couplage de la chromatographie en phase gazeuse (GC) à la spectrométrie de masse a depuis longtemps été utilisé pour le profilage métabolomique [180]. La GC (de l'anglais, Gas Chromatography) a été utilisée dans de nombreuses études telles que l'analyse des métabolites dans les plantes ou dans l'urine [142], [181], [182]. Cette technique utilise l'ionisation par impact électronique, lui conférant une grande reproductibilité spectrale et facilitant l'identification d'un grand nombre de métabolites par l'utilisation de banques de données universelles [183], [184].

La GC-2D peut être envisagée pour l'analyse de matrices biologiques complexes en couplant deux colonnes chromatographiques montées en série. Celle-ci permet de réaliser deux séparations complémentaires voire orthogonales, telles qu'une séparation des composés polaires au sein de la première dimension et une séparation des composés apolaires au sein de la seconde dimension [185], [186]. La GC/MS est cependant limitée aux composés volatiles apolaires et thermostables. Il est donc nécessaire d'envisager une étape de dérivation chimique des métabolites polaires non volatiles. La dérivation peut être chronophage et peut également entraîner la dégradation artéfactuelle de certains métabolites fragiles [187]. Cette étape pourrait également montrer un biais expérimental si la dérivation montre une mauvaise efficacité/reproductibilité.

b.1.c) La LC/MS

La chromatographie liquide (LC) couplée à la spectrométrie de masse représente la technique de référence pour l'analyse globale du métabolome. Elle permet l'analyse d'un large éventail de métabolites allant des composés très polaires aux composés très apolaires *via* l'utilisation de différentes colonnes chromatographiques [188], [189]. Deux types de systèmes chromatographiques en phase liquide peuvent être distingués, la haute pression (HPLC, de l'anglais, High Pressure Liquid Chromatography), et l'ultra haute performance (UHPLC, de l'anglais, Ultra High Performance Liquid Chromatography). La méthode UHPLC utilise des phases stationnaires avec des diamètres de particules réduits (< 2 μm), permettant de travailler à des débits et des pressions plus importants. Les temps d'analyses peuvent ainsi être réduits, générant souvent des pics plus intenses car plus fins, augmentant ainsi le ratio (signal spécifique) / (bruit de fond). La finesse des pics chromatographiques peut cependant se montrer limitante au travers de la vitesse d'acquisition du spectromètre de masse utilisé en aval.

Différents types de greffons sont utilisés sur la couche silicique de la colonne chromatographique. Des greffons alkyles apolaires (RPLC, de l'anglais, Reverse Phase Liquid Chromatography), permettent la rétention et la séparation des composés apolaires et plutôt hydrophobes. De par sa reproductibilité, la RPLC est très répandue en analyse métabolomique et lipidomique [188], [190]. Un autre type de greffons correspond à l'utilisation de groupements alcools, assurant la rétention d'analytes polaires et hydrophiles (NPLC, de l'anglais, Normal Phase Liquid Chromatography) [191], [192]. De multiples études ont montré l'intérêt d'utiliser des colonnes chromatographiques aux configurations complémentaires afin de maximiser la couverture métabolique de la matrice biologique étudiée [188], [189], [193]. Dans leur étude, Boudah et al ont montré qu'une grande partie du métabolome est décrite en combinant les colonnes PFPP (de l'anglais, PentaFluoroPhenylPropyl), et ZIC-pHILIC (de l'anglais, ZwitterionIC-

Hydrophilic liquid chromatography) (Figure 15). Cependant, les colonnes C18 offrent une meilleure reproductibilité, notamment au niveau des temps de rétention, comparativement aux colonnes PFPP, et sont souvent privilégiées en analyses métabolomiques [194], [195]. La colonne HILIC est très utilisée en métabolomique pour la séparation et l'analyse des composés polaires tels que les acides aminés, les acides nucléiques ou les carbohydrates. Cette colonne possède une fine couche d'eau semi-immobilisée à la surface de la couche silicique, permettant des interactions polaires et hydrophiles avec les analytes [192]. Contrairement à la NPLC, la colonne HILIC utilise des phases mobiles directement compatibles avec la spectrométrie de masse. Pour ces différentes raisons, les deux colonnes HILIC et C18 sont les principales colonnes chromatographiques utilisées de façon complémentaires au laboratoire, et ont été utilisées dans l'étude du métabolome de nombreuses matrices biologiques (sérum, plasma, lait maternel...) [35], [188], [196].

De par la grande couverture métabolique obtenue par LC-MS, ce couplage sera utilisé dans le cadre de mes travaux de thèse. Les deux colonnes chromatographiques HILIC et C18 seront utilisées pour couvrir l'ensemble du métabolome du méconium.

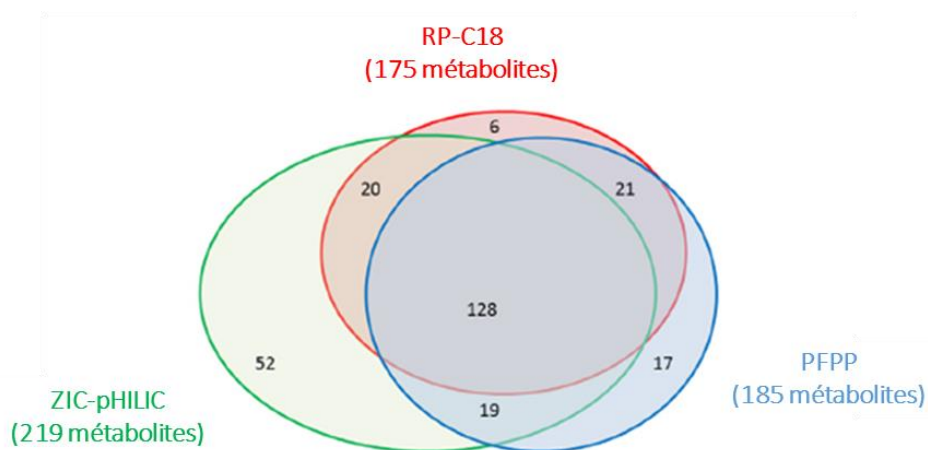


Figure 15 : Comparaison de la couverture métabolique du sérum humain obtenue à travers trois techniques chromatographiques, incluant en rouge la RP-C18, en vert la ZIC-pHILIC, et en bleu la PFPP [197].

b.2) La spectrométrie de masse

Un spectromètre de masse est composé de trois éléments principaux, incluant la source d'ionisation, l'analyseur et le détecteur (Figure 16).

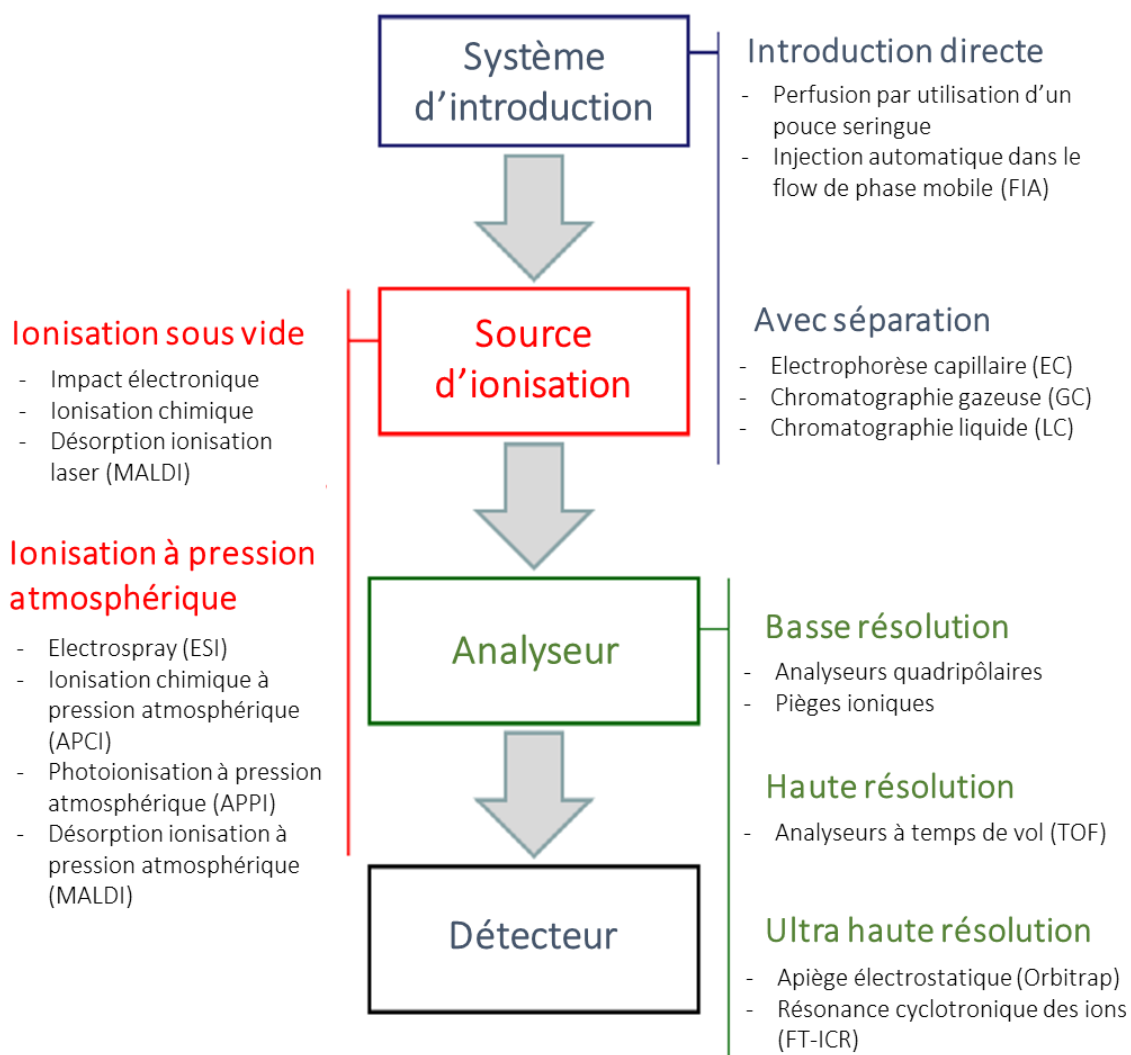


Figure 16 : Schéma d'un spectromètre de masse [198].

b.2.a) La source d'ionisation

La source d'ionisation, au sein de laquelle les ions en phase gazeuse sont produits, peut opérer dans les modes positif et négatif. L'application d'une tension positive ou négative au sein de la source va définir la charge des ions détectés. Parmi les méthodes d'ionisations les plus utilisées, nous décrivons l'ionisation par electrospray (ESI), la désorption-ionisation par laser assistée par une matrice (MALDI), ou encore l'ionisation chimique à pression atmosphérique (APCI) : conduisant notamment à la détection de formes protonnées $[M+H]^+$ ou déprotonnées $[M-H]^-$. Parmi les sources décrites ci-dessus, la source d'ionisation à pression atmosphérique electrospray est la plus utilisée en métabolomique, et sera utilisée dans le cadre de ce travail.

b.2.b) L'analyseur

Le deuxième élément essentiel d'un spectromètre de masse est l'analyseur, il permet de séparer les ions produits dans la source selon leur rapport masse sur charge (m/z). Différents types d'analyseurs peuvent être connectés à une source electrospray (ESI). Ces analyseurs possèdent des propriétés spécifiques décrites selon quatre caractéristiques principales : i) la résolution, ii) la précision de la mesure de masse, iii) la sensibilité et la gamme dynamique, et iv) la vitesse d'acquisition.

- i. **La résolution (R_s) (ou le pouvoir résolutif (R_p))** représente l'aptitude d'un analyseur à séparer deux ions de rapports m/z très proches. La résolution est le plus souvent définie à l'aide de la mesure de la largeur des pics à mi-hauteur (FWHM, de l'anglais, Full Width at Half Maximum), et peut être calculée selon la formule :

$$R_s = \frac{m}{\Delta m (\text{FWHM})}$$

- ii. **La précision de la mesure de masse** indique la justesse de la masse mesurée par rapport à la masse théorique. La précision de la mesure de masse s'exprime souvent en ppm (pour partie par million), et est calculée selon la formule :

$$\text{erreur de la mesure de masse (ppm)} = \frac{m (\text{expérimentale}) - m (\text{théorique})}{m (\text{théorique})} \times 10^6$$

- iii. **La sensibilité et la gamme dynamique :** la sensibilité est notamment caractérisée par le pouvoir de transmission de l'analyseur, et dépend directement de la proportion d'ions transmis de l'analyseur au détecteur [171]. La sensibilité d'un instrument peut être calculée selon la formule :

$$\text{Sensibilité} = \frac{\text{Intensité (signal spécifique)}}{\text{Intensité (bruit)}}$$

- iv. **La vitesse d'acquisition** correspond au temps nécessaire à un analyseur pour enregistrer un spectre de masse. Ce paramètre est important en particulier lors d'un couplage LC/MS, où le temps imparti pour l'enregistrement d'un spectre est défini par la largeur du pic chromatographique. Dans ce cas de figure, il est essentiel d'envisager un compromis entre la finesse d'un pic chromatographique et le temps imparti pour l'acquisition des spectres de masse. La vitesse d'acquisition d'un analyseur peut être exprimée soit sous forme d'un temps

(en millisecondes), définissant le moment de la formation des ions et l'enregistrement du signal, soit sous forme d'une fréquence (nombre de spectres par secondes, en Hertz).

Il existe différents types d'analyseurs de masse, de haute et d'ultra haute résolution. Les analyseurs de basse résolution, tels que les analyseurs quadripolaires, ont un pouvoir résolutif < 10000 et sont largement utilisés en analyses ciblées. Ceux-ci montés en série permettent des analyses quantitatives des métabolites. Une sélection d'analyseurs de haute à ultra haute résolution avec un pouvoir résolutif > 10 000, très répandus en profilage métabolomique, est décrite dans le

Tableau 3. Les mesures de masses à haute résolution permettent la séparation de molécules isobares, possédant la même masse nominale et ne différant que par leur défaut de masse. Ainsi, plus un analyseur sera résolutif, plus cette séparation sera significative, facilitant l'identification des métabolites à l'aide de bases de données spectrales. Ces analyseurs incluent les technologies Orbitrap [199]–[201], les analyseurs à temps de vol (ToF, de l'anglais Time of Flight) [202], [203], ou FT-ICR (de l'anglais, Fourier Transform – Ion Cyclotron Resonance) [204]–[206]. Chacun de ces analyseurs présente des avantages et des inconvénients : les analyseurs ToF sont moins résolutifs comparés aux Orbitrap et aux cellules ICR, mais présentent la plus haute vitesse d'acquisition permettant le développement de nouvelles approches d'acquisition. A l'opposé, le FT-ICR est l'analyseur le plus résolutif mais nécessite un temps plus long pour enregistrer un spectre et est relativement plus coûteux. La technologie Orbitrap présente de très bonnes caractéristiques en termes de résolution, de précision en masse et de vitesse d'acquisition.

La majorité des spectromètres de masse actuels contiennent plusieurs analyseurs montés en série (hybrides) [199], [203]. A titre d'exemple, l'Orbitrap fusion possède trois analyseurs de masse : un quadripôle, un analyseur de technologie Orbitrap et un piège à ions linéaire. Cette technologie combinant plusieurs analyseurs permet de réaliser des analyses complémentaires de fragmentation (MS/MS ou MSⁿ) au moment de l'acquisition des spectres [207], [208].

La technologie Orbitrap représente aujourd'hui la technique la plus utilisée en analyses métabolomiques. Au laboratoire, les analyses métabolomiques sont principalement réalisées sur des spectromètres de masse de type Exactive/Q-Exactive (technologie Orbitrap de Thermo Fisher Scientific). Pour ce travail de thèse, les analyses métabolomiques ont été réalisées sur un Q-Exactive.

Tableau 3 : Vue comparative d'une sélection d'analyseurs utilisés pour des analyses globales en métabolomique, adaptée de Fenaille et al [209].

	Instruments	Pouvoir résolutif/ Résolution	Précision en masse	Vitesse d'acquisition
TOF	6545 Q-ToF (Agilent)	Jusqu'à 45 000	< 1 ppm	50 Hz (à une résolution de 45 000 à m/z 2722)
	timsToF (Bruker)	Jusqu'à 50 000	< 0,8 ppm	100 Hz (à une résolution de 50 000 à m/z 1222)
Orbitrap	Exactive (Thermo Fisher Scientific)	Jusqu'à 100 000 (m/z 200)	< 1 ppm	12 Hz (à une résolution de 17 500 à m/z 200)
	Q-Exactive HF (Thermo Fisher Scientific)	Jusqu'à 240 000 (m/z 200)	< 1 ppm	18 Hz (à une résolution de 7 500 à m/z 200)
	Orbitrap Fusion Lumos (Thermo Fisher Scientific)	Jusqu'à 500 000 (m/z 200)	< 1 ppm	20 Hz (à une résolution de 15 000 à m/z 200)
FT-ICR	Solarix XR FT 7T (Bruker)	Jusqu'à 750 000 (m/z 400)	< 1 ppm	1 Hz (à une résolution de 100 000 à m/z 400)

b.2.c) Le détecteur

Enfin arrive en dernier lieu le détecteur, celui-ci permet de dénombrer les ions, suivi d'un système de traitement des données permettant l'enregistrement et la visualisation des spectres, afin de permettre leur interprétation.

b.3) L'acquisition de spectres MS/MS

L'acquisition de spectres de fragmentation est essentielle à l'identification « formelle » des métabolites « putativement » annotés. Au cours de ce travail de thèse, nous utilisons le mode ciblé, permettant d'isoler et de fragmenter les ions d'intérêts en renseignant leur valeur m/z et leur temps de rétention. L'acquisition des spectres MS/MS utilise un système composé de deux analyseurs montés en série. Le processus de fragmentation est réparti en trois étapes : une première étape de sélection des ions parents est réalisée au sein du premier analyseur généralement de basse résolution, tel qu'un quadripôle ou piège ionique. Une deuxième étape permettant l'activation des ions parents en augmentant leur énergie interne, induisant leur fragmentation en ions fils. Il faut noter que tous les ions ne se fragmentent pas à la même énergie de collision. Il est souvent nécessaire de tester différentes valeurs d'énergies de telle sorte à observer l'ensemble des ions fils ainsi qu'une intensité suffisante de l'ion parent permettant de l'identifier. Enfin, les ions arrivent au sein du deuxième analyseur pour être analysés. Celui-ci est le plus souvent de haute résolution, tel que l'Orbitrap ou le temps de vol (ToF).

Les spectres de fragmentation obtenus sont ensuite comparés manuellement ou automatiquement aux spectres des composés de références purs analysés dans les mêmes conditions et sur les mêmes instruments. L'analyse par spectrométrie de masse en tandem (MS/MS ou MS²) permet de valider, ou au moins de conforter, les propositions attribuées lors de l'annotation. Quatre niveaux d'identification ont été décrits par Sumner et al [210], permettant d'harmoniser les résultats présentés d'un laboratoire à un autre. Ces niveaux d'annotation sont définis comme suit :

- i. **Le niveau 1 correspond au plus haut niveau d'identification**, alliant au moins deux paramètres indépendants et orthogonaux. Un métabolite peut être dit identifié au niveau 1 si son temps de rétention et sa masse précise, ou sa masse précise et son spectre de fragmentation correspondent à ceux d'un composé de référence pur analysé dans les mêmes conditions expérimentales et sur le même instrument.
- ii. **Le niveau 2 correspond à l'annotation putative** mais très probable des métabolites, réalisée à partir d'une banque de donnée spectrale externe ou publique. Cette annotation n'est pas validée au niveau 1 par l'absence du composé de référence au sein du laboratoire réalisant l'étude.
- iii. **Le niveau 3 correspond à une annotation incomplète ou incertaine** quand plusieurs propositions de métabolites sont obtenues. Les informations disponibles permettent néanmoins de classer ces métabolites au sein d'une famille chimique.

- iv. **Le niveau 4 correspond aux composés inconnus**, le dernier niveau d'identification dans le cas où aucune conclusion démonstrative ne peut être tirée quant à la nature du composé.

3. Traitement des données acquises par LC/HRMS

L'analyse métabolomique non ciblée génère de gros volumes de données dont le traitement nécessite une automatisation et l'utilisation de logiciels robustes. Celui-ci comprend de nombreuses étapes qu'il est essentiel de maîtriser, afin de générer des résultats fiables et biologiquement pertinents. Ces étapes de traitement permettent de réduire l'impact du bruit de fond, de contrôler les différents biais biologiques ainsi qu'analytiques existants, et de réduire la variance entre les échantillons [211], [212]. Les principales étapes des traitements de données en métabolomique sont décrites ci-dessous.

a) La détection automatique des signaux

Cette étape consiste à transformer les signaux extraits en une matrice de données regroupant l'ensemble des informations sur les échantillons et les variables détectées. Chaque variable est définie par un couple masse/charge (m/z) x temps de rétention (RT), et l'intensité de cette variable dans chaque échantillon est reportée dans la matrice. Pour une seule condition analytique et un même échantillon, plusieurs milliers de variables sont détectées à cette étape, d'où la nécessité de l'automatiser. Différents logiciels sont disponibles qui sont soit commercialisés par les constructeurs de spectromètres de masse [213]–[215], soit disponibles en libre accès sur plusieurs plateformes telles que XCMS online [216], MetaboAnalyst [217] ou Workflow4Metabolomics (W4M, utilisé au laboratoire) [211], [218]. W4M regroupe de multiples algorithmes offrant la possibilité d'optimiser différents paramètres appliqués lors de l'extraction des signaux (XCMS), mais permet également l'annotation des variables ainsi que l'analyse statistique des données.

Le traitement XCMS génère deux tableaux (Figure 17) : un premier tableau (matrice des données ou « data matrix »), qui regroupe les variables (en lignes) et les échantillons (en colonnes). Cette matrice renseigne les aires des pics détectés pour chaque variable au sein de chaque échantillon. Le deuxième tableau généré représente les métadonnées des variables (« variable metadata »), regroupant les variables (en lignes) et les informations sur les variables (en colonnes), telles que le temps de rétention, le m/z ... L'ensemble des variables détectées est décrit selon la nomenclature MiTj où i indique le rapport masse sur charge et j le temps de rétention (en secondes). Il faut noter que plusieurs variables peuvent être issues d'un même pic et peuvent donc représenter un seul et même métabolite.

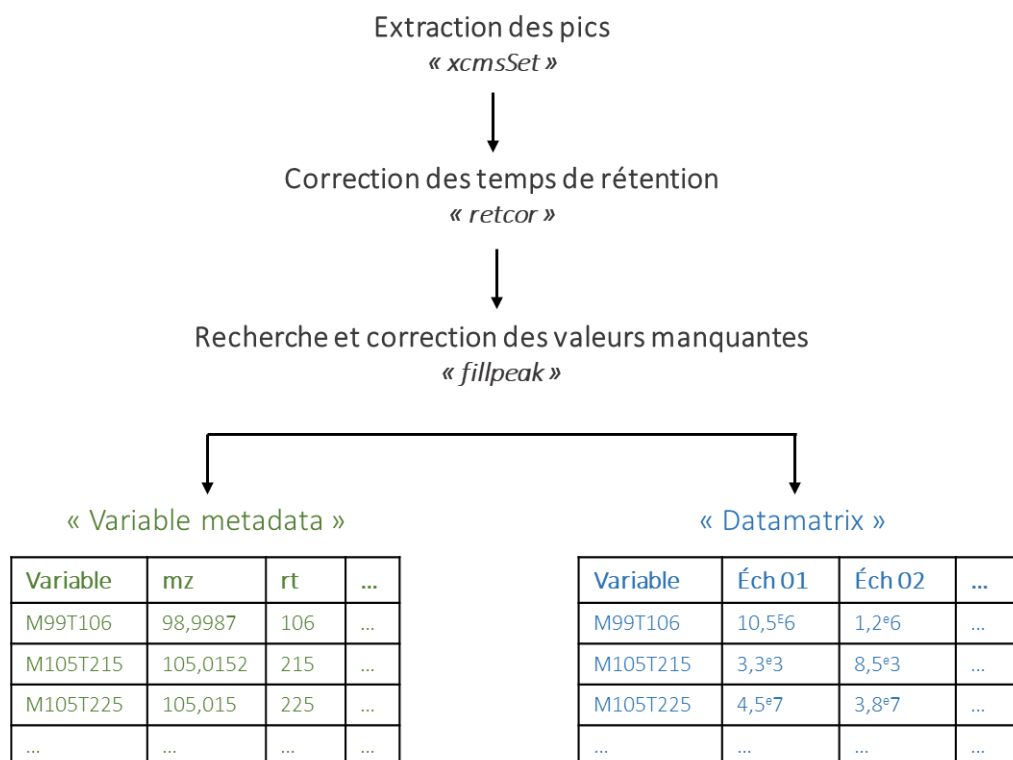


Figure 17 : Description des deux matrices de données générées par le traitement automatique XCMS « variable metadata » en vert et « datamatrix » en bleu.

b) L'alignement et intégration des pics

L'étape de l'alignement des pics va permettre de rechercher et d'aligner les pics (en m/z et RT) en cas de décalage entre les différentes analyses au sein de la même séquence. Ce décalage peut résulter par exemple d'un encrassement de la source ou de la colonne chromatographique. L'alignement inter-échantillons appliqué sur les temps de rétention et les ratios m/z permet de regrouper les variables identiques au sein des différents échantillons. A la suite de cet alignement/regroupement, une correction des temps de rétention permettant de recalibrer les temps enregistrés est souvent réalisée. Il faut noter qu'une nouvelle étape de regroupement des pics peut être réalisée à la suite de cette correction pour mettre en évidence de nouveaux groupes de signaux [211]. Lors de la détection automatique des signaux, le logiciel peut observer l'absence de certains pics en comparant les différents échantillons. Une méthode de complétion des données manquantes peut être appliquée à ce stade. Cette étape, appelée « fillpeak », permet de réintégrer le signal non détecté en raison d'un bruit de fond trop important, d'une forme de pic non gaussienne ou d'une intensité inférieure au seuil de détection [219]. Ceci permet de s'affranchir de valeurs manquantes pouvant fausser les analyses statistiques.

c) Le filtre analytique des variables

Tautenhahn et al ont montré qu'en sortie des analyses métabolomiques non ciblées, seul un signal sur quatre détectés peut être pertinent [220]. Parallèlement, Sindelar et al montrèrent la complexité de l'origine des signaux extraits en analyse métabolomique, confirmant que toute l'information apportée n'est pas pertinente (Figure 18) [221]. Ces paramètres montrent la nécessité de filtrer les signaux non significatifs pouvant provenir d'un biais analytique, tels que des contaminants issus des phases mobiles, des tubulures ou des fioles d'injections. Ces signaux peuvent également provenir d'un biais biologique, ou de « redondance spectrale » ou artéfacts, provenant de l'enregistrement de plusieurs signaux pour un même pic [222][223]. Ces signaux non pertinents rendent souvent la détermination du nombre exact des métabolites uniques difficile.

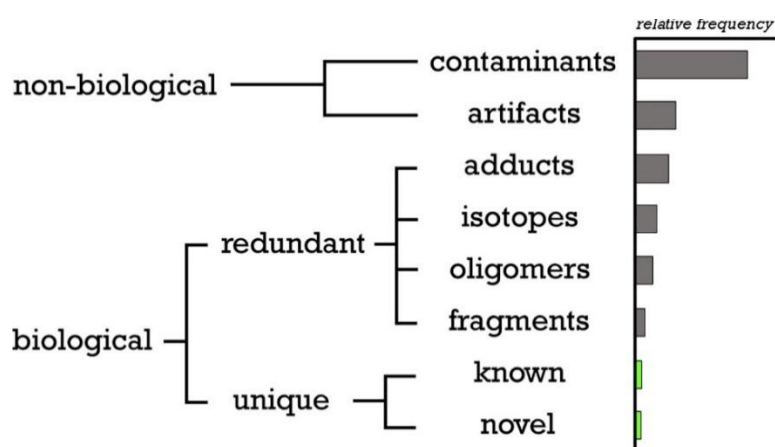


Figure 18 : Fréquence relative des différents signaux non biologiques et biologiques extraits lors d'une analyse métabolomique non ciblée [221].

Afin de déterminer les éventuelles dérives analytiques, de nombreuses études font appel à l'utilisation de contrôles qualité (QC) insérés à des intervalles réguliers tout le long de la séquence analytique [224], [225]. Dans leur étude, Begou et al préconisent l'injection de QC tous les cinq échantillons biologiques lorsque le nombre d'échantillons total est < 100, ou tous les dix échantillons biologiques lorsque leur nombre total est > 100 [225]. Plus le nombre de QC analysés dans la séquence est grand, plus le suivi de la séquence analytique sera maîtrisé. Cependant, les multiples injections de QC rallongent les temps d'analyses, nécessitant de trouver un compromis. En général, l'ensemble des QC représentent environ 10 % de la totalité des analyses [225]. L'échantillon de QC doit être représentatif des échantillons étudiés. Il est ainsi préparé en mélangeant un volume équivalent de tous les échantillons de la séquence. Pour l'analyse de cohortes, l'échantillon de QC peut être préparé à partir de l'ensemble des échantillons ou, si la préparation se montre compliquée, à partir des échantillons contenus dans une même série d'analyse [166], [225], [226].

L'utilisation de QC dilués a également été préconisée dans le but d'améliorer l'identification des signaux artéfactuels ou interférents, et de les filtrer [220], [227]. Les signaux pertinents doivent en effet diminuer proportionnellement avec la dilution réalisée [228]. Des « blancs » analytiques composés d'un mélange de phases mobiles dans les proportions analytiques peuvent être insérés en début et/ou en fin de séquence, permettant de définir le bruit de fond de la séquence analytique, et de filtrer les variables non issues des échantillons biologiques étudiés [225].

c.1) Filtres des signaux non pertinents utilisés au laboratoire

Trois différents filtres sont réalisés au laboratoire, permettant de nettoyer au mieux les données générées. Ces filtres sont appliqués pour chaque variable sur : i) Le coefficient de variation de l'intensité des variables au sein des QC, qui doit être $< 30\%$, ii) Le coefficient de corrélation de l'intensité des variables avec le facteur de dilution des QC, qui doit être $> 70\%$, iii) Le ratio des signaux obtenus au sein des échantillons biologiques par rapport aux signaux des blancs analytiques, qui doit être au moins égal à 3. Cette méthode permet de réduire le jeu de données de cinq à six fois, sans perte significative d'information pertinente [223], [229].

c.2) Filtres des signaux redondants (redondance chimique)

D'un autre côté, il y-a les redondances chimiques, résultant de la présence de plusieurs formes moléculaires détectées pour un même composé. Celles-ci incluent les isotopes, les adduits, les dimères ou les fragments, surévaluant le nombre de métabolites réellement présents (Figure 19). La présence de plusieurs ions pour un seul métabolite peut cependant aider dans l'identification du métabolite parent, par la mesure de plusieurs masses précises liées au même composé. Cette redondance peut être nettoyée lors de l'annotation des métabolites, en retenant la forme moléculaire la plus abondante par métabolite (souvent représentée par le pic monoisotopique).

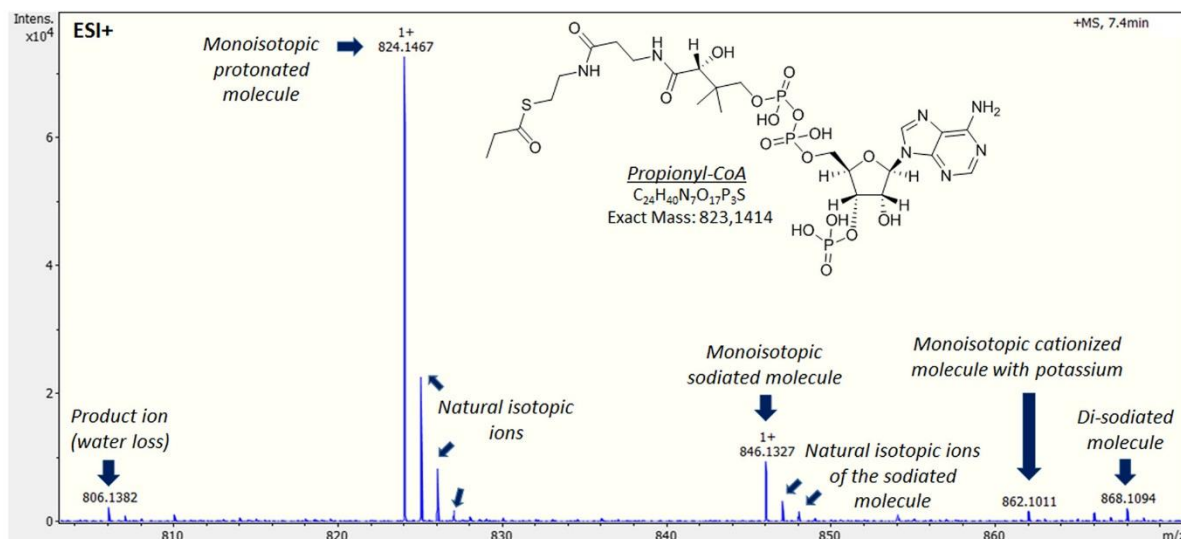


Figure 19 : Description des différents signaux MS pouvant être produits pour un seul métabolite, exemple du propionyl-CoA, adaptée de Damont et al [230].

d) La normalisation des données

Le couplage LC/MS peut être source de nombreuses dérives, telles que des décalages dans les temps de rétention ou dans la mesure des ratios m/z , observés notamment lors de l'analyse de grandes séries d'échantillons [228]. Au sein d'une même séquence d'analyse regroupant une partie des échantillons biologiques (« batch »), ces dérives peuvent être maîtrisées par l'ajout de standards internes, idéalement marqués permettant de les différencier des métabolites réellement présents dans les échantillons biologiques. L'utilisation de QC permet également de réaliser cette normalisation en les insérant à des intervalles réguliers au cours de la séquence analytique [164], [225]. Cette normalisation peut être réalisée au sein d'un même ou de plusieurs « batchs », notamment dans le cas de larges cohortes [225], [228], [231]. En cas de dérive, une correction basée sur les signaux enregistrés au sein des QC peut être réalisée à partir d'un algorithme comme le « LOESS » (de l'anglais, Locally Estimated Scatterplot Smoothing). Cet algorithme permet de réaliser une régression locale et spécifique aux intensités mesurées dans les QC, et de corriger les signaux enregistrés au sein des échantillons (Figure 20).

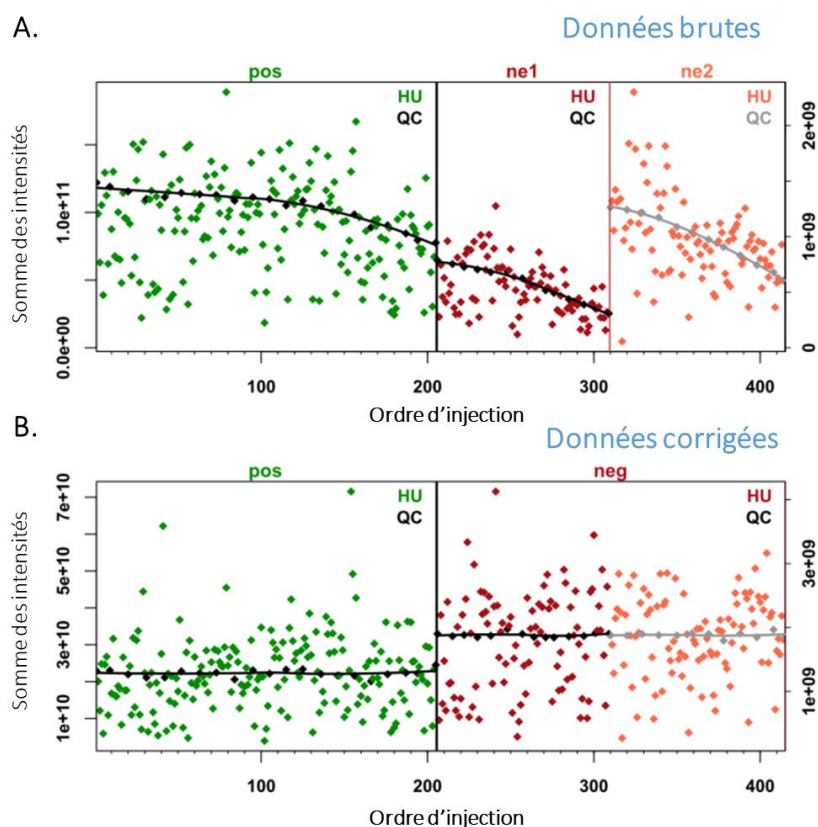


Figure 20 : Représentation de la somme des intensités avant (A) et après (B) normalisation. Les QC sont représentés en noir ou en gris dans les deux figures. En vert est illustrée une normalisation réalisée au sein d'un même batch. Tandis qu'en rouge et en orange, la normalisation est réalisée sur deux différents batches.

e) L'annotation des variables

Dans une approche d'identification globale des métabolites, l'ensemble des variables est pris en compte pour l'étape de l'annotation. A l'inverse dans une approche de recherche de biomarqueurs, uniquement les variables significativement différentes entre les groupes étudiés sont recherchées. L'étape de l'annotation représente l'une des principales limites en analyses métabolomiques, en raison du nombre restreint de métabolites connus à ce jour [232], [233]. L'annotation des métabolites est réalisée par comparaison à des bases de données spectrales publiques et/ou internes. La banque interne est généralement utilisée en premier quand celle-ci est disponible. L'interrogation de bases de données externes peut ensuite être réalisée pour annoter davantage de variables, mais avec un moindre niveau de confiance.

e.1) Utilisation de bases de données internes

La création d'une base de données interne en LC/HRMS consiste à rassembler les informations spectrales et chromatographiques à partir de composés de référence sur un spectromètre de masse et une colonne chromatographique donnés. Au sein du laboratoire, une base de données a été construite contenant environ 1200 composés, incluant majoritairement des métabolites endogènes et quelques métabolites exogènes (médicaments, cosmétiques, alimentaires...) [35]. Nous obtenons à partir de cette base de données, moins de 10 % d'annotation des signaux détectés et analytiquement pertinents (retenus après les différents filtres appliqués). Bien que les bases de données internes soient limitées en nombre de métabolites, elles permettent néanmoins d'obtenir des annotations plus fiables.

Le procédé d'annotation suivi au laboratoire, décrit en Figure 21, utilise les informations des variables incluant leurs m/z ainsi que leurs RT, qui sont à rapprocher des métabolites inclus dans la base de données interne. Lors de l'annotation, des valeurs de tolérance sur les m/z et les RT sont à fixer par l'opérateur permettant de considérer de faibles décalages présents dans leurs mesures. Ces valeurs de tolérance sont spécifiques au spectromètre de masse et à la colonne chromatographique utilisés. Une fois l'annotation réalisée, les métabolites sont alors « putativement » annotés et nécessitent d'être validés à travers des expériences MS/MS (réalisées par l'utilisation de composés standards analysés dans les mêmes conditions analytiques [232]).

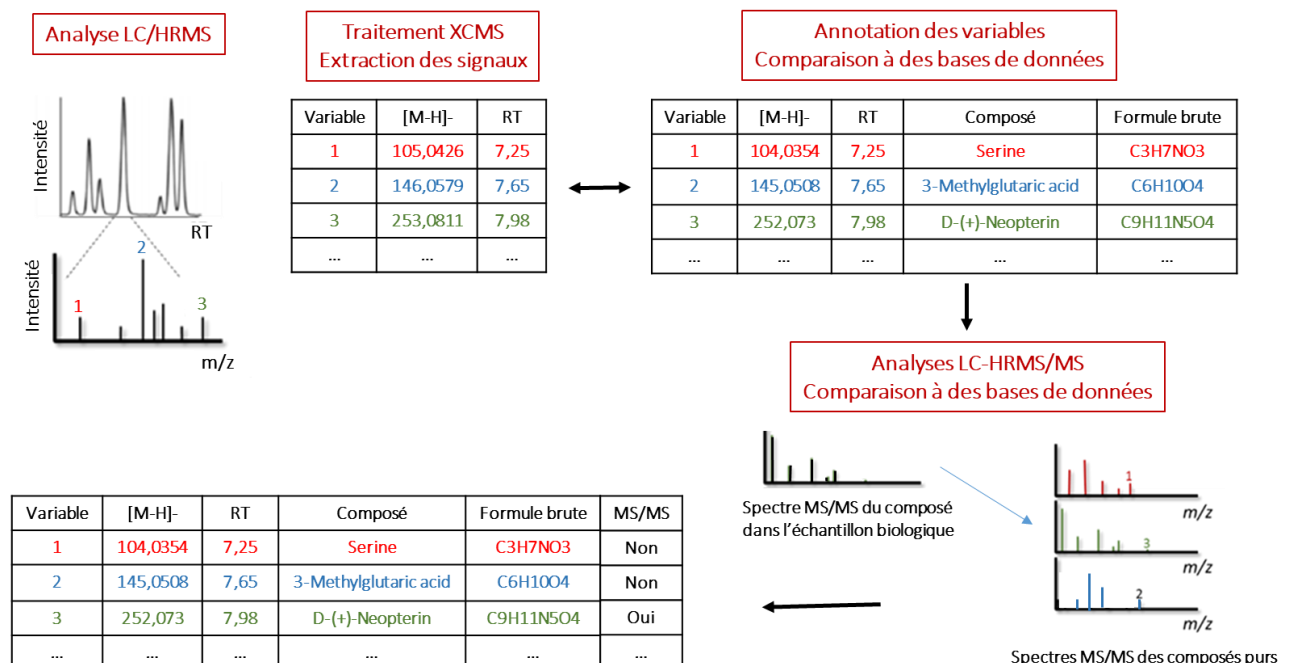


Figure 21 : Processus d'annotation et d'identification des métabolites par l'utilisation d'une base de donnée interne, et confirmation par analyse MS/MS .

e.2) Utilisation de bases de données externes

Les banques publiques peuvent être alimentées par plusieurs laboratoires simultanément. Elle répertorient des données spectrales MS et parfois MS/MS acquises sur différents spectromètres de masses et dans différentes conditions analytiques [234]. Dans ce contexte, les bases de données publiques peuvent contenir plusieurs centaines de milliers de molécules chimiques, sans lien avec le type d'échantillon analysé. Il est donc souvent intéressant de privilégier l'utilisation de bases de données plus spécialisées, amenant à l'annotation de métabolites biologiquement pertinents à l'étude [235]. Il existe plusieurs banques disponibles, telles que HMDB (de l'anglais, Human Metabolome DataBase), ou plus globales telles que KEGG (de l'anglais, Kyoto Encyclopedia of Genes and Genomes), regroupant des données chimiques, métaboliques et biochimiques [236]–[240].

L'annotation procède par l'attribution de compositions élémentaires aux données spectrales selon leur m/z en appliquant une tolérance dépendante de l'instrument utilisé. Le fait d'opérer en HRMS permet d'avoir des valeurs de m/z plus précises, réduisant notablement le nombre de candidats possibles. En effet, l'attribution de multiples annotations par pic rendent souvent l'identification formelle difficile [241]. Cette caractéristique représente à la fois une limite des bases externes où de multiples candidats isobares et/ou isomères sont proposés pour une même molécule. Afin de valider l'annotation réalisée par des bases de données externes, il est nécessaire de se procurer le composé pur et de l'analyser dans les mêmes conditions que l'échantillon biologique.

4. Analyses statistiques

Les analyses statistiques permettent de faire ressortir les variables pouvant être marqueurs d'un statut physio(patho)logique donné, en analysant par exemple des patients *versus* des contrôles. Dans ce contexte, deux grandes catégories d'analyses statistiques peuvent être réalisées en métabolomique, présentées ci-dessous.

a) Analyses statistiques univariées

Les techniques d'analyses univariées sont classiquement utilisées en biologie. Elles permettent d'analyser une variable donnée en fonction de différentes modalités (lieu de vie, mode de naissance, statut pathologique... etc) [231]. En fonction du jeu de données, de la tendance de distribution et de la taille de la population, un test paramétrique ou non paramétrique est utilisé. Un test paramétrique est appliqué lorsque le nombre d'échantillons est suffisant (> 30) et que la distribution des données suit

une loi normale. Dans ce cas, il peut être utilisé un t-test (test de Student) ou une ANOVA (de l'anglais, ANalysis Of Variance), pour analyser les différences entre respectivement 2 ou plus de 2 groupes différents. Ces groupes sont définis selon des modalités qualitatives (par exemple le sexe ou le statut malade vs sain). Un test de corrélation de Pearson peut également être utilisé pour évaluer une association (dépendance) entre variables quantitatives (ou variables continues) au sein des individus (par exemple l'âge ou l'indice de masse corporelle) (Tableau 4 et Figure 22). A l'inverse, lorsque la distribution des variables ne suit pas une loi normale ou que le nombre d'échantillons est faible (< 30), des tests non-paramétriques sont utilisés. Le test de rang de Mann-Withney-Wilcoxon et le test de Kruskal-Wallis sont ainsi des alternatives non paramétriques du t-test et de l'ANOVA respectivement. De l'autre côté le test de corrélation de Spearman permettra de tester l'association de la variable mesurée avec des modalités quantitatives. Différentes procédures de corrections peuvent être utilisées lors de ces tests, telles que les corrections de Bonferroni, de Benjamini-Hochberg ou FDR (False Discovery Rate), fréquemment utilisées en métabolomique [242].

Tableau 4 : Tableau descriptif du choix de tests univariés à utiliser selon la modalité et la distribution du jeu de données.

Réponse Y		Paramétrique	Non paramétrique
Qualitative	2 niveaux	Student	Mann-Withney-Wilcoxon
	> 2 niveaux	Anova	Kruskal Wallis
Quantitative	Corrélation	Pearson	Spearman

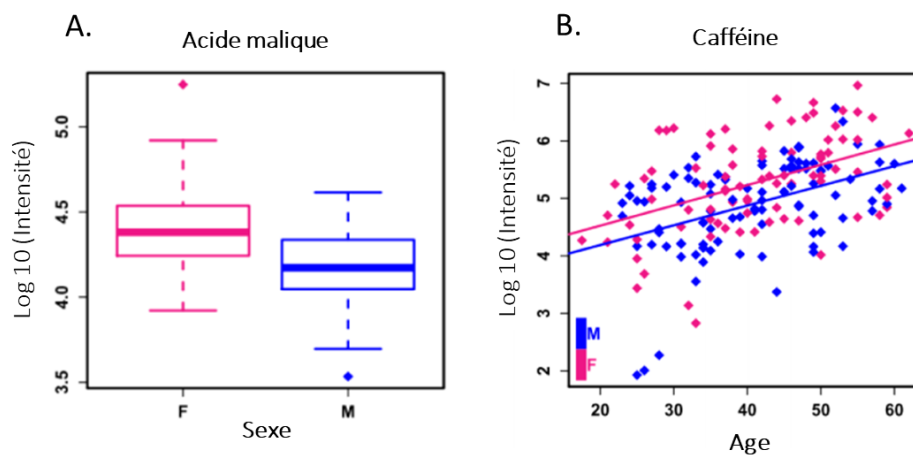


Figure 22 : A. analyses univariées en fonction d'une modalité qualitative (sexe), avec représentation en « boîtes à moustaches » de l'intensité de la variable analysée (acide malique) pour les deux groupes

d'échantillons. B. Analyse de la corrélation entre l'intensité de la variable mesurée (caféine) et l'âge des individus (variable continue) [231].

b) Analyses statistiques multivariées

En analyses métabolomiques, un grand volume de données est généré, nécessitant une représentation graphique de ces données afin de permettre leur visualisation [243], [244]. Les analyses multivariées reposent sur une réduction de la dimension des données, permettant de représenter dans un espace réduit l'ensemble de l'information. Il existe deux types d'analyses multivariées, supervisées et non-supervisées, décrites ci-dessous.

b.1) Les analyses non supervisées

Les analyses multivariées non supervisées, telles que les analyses en composantes principales (ACP), vont chercher à faire ressortir la variance existante dans le jeu de données sans *a priori* sur les échantillons. Ces analyses sont en général réalisées en premier lieu lors des traitements statistiques. Elles permettent d'avoir une représentation de la structure latente des données par la construction spontanée et globale dans un espace réduit [244], [245]. Les ACP permettent également d'identifier des échantillons ou des variables aberrants.

Lors d'une ACP, les individus sont projetés sur un premier axe, résumant la plus grande part de la variance du jeu de données. Un second axe orthogonal au premier est ensuite constitué, résumant la plus grande part de la variance résiduelle. En théorie, le nombre d'axes sera égal au nombre de variables, mais les premiers axes représentent en général un pourcentage suffisant de la variabilité totale du jeu de données. L'analyse de ces premiers axes et la contribution des individus et des variables à leur construction, permet une première interprétation des données. Deux types de graphiques sont à retenir pour l'ACP, d'abord le « *score plot* » où les échantillons/individus sont représentés sous forme de carte factorielle dans un espace réduit. Parallèlement le « *loading plot* » montre une représentation spontanée et globale des variables. En analysant ces deux graphiques, il est possible de mettre en évidence la présence de variables ou d'échantillons aberrants au sein des données, mais aussi des variables explicatives de la structuration du jeu de données et du regroupement sans « *a priori* » des individus (Figure 23).

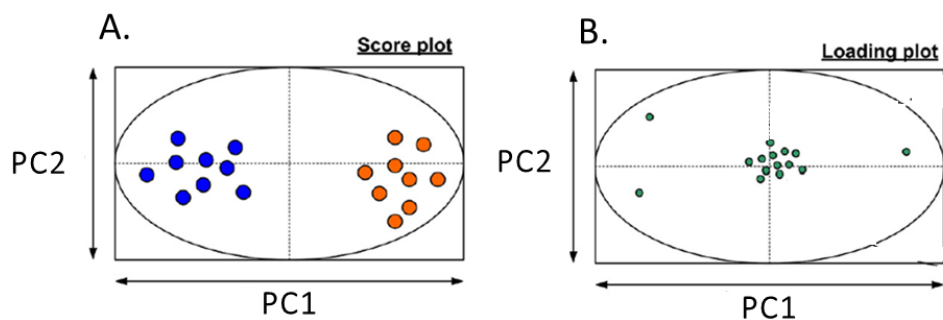


Figure 23 : Représentation graphique d'une ACP. En A, le « *Score plot* » représentant une carte factorielle des individus. En B, le « *Loading plot* » représentant la carte factorielle des variables dans le jeu de données, adaptée de Roux et al [246].

b.2) Les analyses supervisées

Les analyses supervisées dites prédictives visent à discriminer des groupes de données selon des variables explicatives par la construction d'un modèle mathématique. Ces analyses sont réalisées en renseignant le paramètre de réponse qu'on souhaite étudier. Si un modèle est construit, la contribution de chaque variable dans le modèle est calculée, en lui attribuant un score VIP (de l'anglais, Variable of Importance in Projection). Une variable dont le VIP est supérieur à 1 est considérée comme participant « significativement » à la construction du modèle et donc à la discrimination des groupes (variable discriminante) [222], [247]. Parmi les méthodes explicatives, on trouve la régression PLS-DA (de l'anglais, Partial Least Square - Discriminant Analysis) et l'OPLS-DA (de l'anglais, Orthogonal Projection to Latent Structure - Discriminant Analysis). L'analyse OPLS-DA offre une meilleure discrimination et donc séparation des groupes d'échantillons par un phénomène de permutation. Les variables considérées discriminantes peuvent à la suite être identifiées pour aider à l'interprétation biologique des données.

5. Outils d'aide à l'interprétation biologique des données

Une fois les métabolites identifiés, leur interprétation biologique est réalisée en les replaçant au sein de réseaux métaboliques connus. Ces réseaux permettent de connecter les métabolites entre eux par des réactions de biosynthèse (anabolisme) ou de biodégradation (catabolisme). Ces voies métaboliques vont permettre de mieux comprendre le rôle des métabolites identifiés dans le contexte biologique étudié et les mécanismes biochimiques sous-jacents [248]. Ces voies ont souvent été étudiées de façon indépendantes en ne considérant qu'un sous-ensemble de réactions biochimiques allant du substrat au produit final [249], et faisant abstraction parfois de nombreuses interconnexions métaboliques possibles. Aujourd'hui, de nouvelles approches permettent d'étudier de façon plus globale les voies

métaboliques [250], notamment par l'utilisation d'outils tels que MetExplore [251] ou Metscape de Cytoscape [252]. Ces outils permettent une visualisation plus globale et une compréhension plus approfondie des phénomènes biologiques mis en jeu. Ils restent cependant limités et nécessitent souvent l'intervention d'experts du métabolisme (généralement des médecins, des biologistes en science intégrative) pour obtenir une interprétation biologique.

Après avoir parcouru l'ensemble du processus de l'approche métabolomique par LC/HRMS utilisé lors de ce travail de thèse, nous allons présenter l'état de l'art des études réalisées en métabolomique sur des échantillons de méconiums.

B. Analyses métabolomiques réalisées sur le méconium

Nous présentons dans la suite les différentes analyses métabolomiques ciblées et non ciblées réalisées sur des échantillons de méconium, mais également de premières selles de nourrissons (permettant d'avoir un comparatif). Les techniques utilisées incluent aussi bien la LC/HRMS, la GC/MS que la RMN.

1. Les analyses métabolomiques non ciblées réalisées sur le méconium

Peng et al ont analysé (par LC/HRMS), les profils métaboliques de méconiums collectés au sein d'une étude cas-témoins, de mères atteintes de diabète gestationnel (GDM). Cette étude a permis d'identifier 14 biomarqueurs du GDM dans le méconium, incluant l'acide urique, la méthylguanosine, l'acide taurodeoxycholique ou encore la méthylxanthine. Dix des biomarqueurs identifiés correspondaient à des métabolites endogènes, révélant que le GDM peut induire des perturbations dans le métabolisme des lipides, des aminoacides et de la purine [253].

Les principaux acides biliaires présents dans le méconium varient en fonction de l'âge gestationnel [254]. Dans ce contexte, Kumagai et al ont étudié par GC/MS la composition en acides biliaires de méconiums de nouveau-nés prématurés (n = 7) ou nés à terme (n = 9). Les échantillons ont été prélevés à la naissance puis à 6 jours post-partum, permettant de suivre l'évolution de la composition en acides biliaires durant les premiers jours de vie. Une composition similaire en acides biliaires primaires et secondaires est observée dans les échantillons de méconiums de nouveau-nés à terme et dans les fèces collectées à 6 jours chez des nouveau-nés prématurés. Les auteurs ont supposé que la voie principale de biosynthèse d'acides biliaires chez les prématurés est la voie acide, jusqu'à ce qu'ils atteignent environ 7 jours. Cette composition est dominée par les acides chénodésoxycholique et hyocholique, mais de grandes quantités d'acide désoxycholique et d'ursodésoxycholique sont également présents.

Après trente semaines de gestation, le liquide amniotique avalé par le fœtus est principalement composé de l'urine de celui-ci. Ainsi, la composition en acides biliaires du liquide amniotique à une période avancée de la grossesse est difficile à distinguer de la composition en acides biliaires urinaires du fœtus. Celle-ci contient des quantités élevées d'acides biliaires cétoniques insaturés. Les auteurs ont supposé retrouver ces acides biliaires en grandes quantités dans le méconium, or, ceux-ci n'ont pas été détectés, supposant qu'ils soient absorbés par l'intestin du fœtus au cours des dernières semaines de gestation. Les auteurs suggèrent également que ces acides biliaires peuvent être transférés de la mère par le lien placentaire, car de plus grandes teneurs en acides biliaires cétoniques insaturés sont détectés dans les urines des femmes en fin de gestation ou juste après l'accouchement [255].

Dans la continuité de ces travaux, Naritaka et al ont comparé la composition en acides biliaires de la vésicule biliaire et de méconiums prélevés chez neuf fœtus décédés à la naissance en comparaison à sept échantillons de méconium obtenus de nouveau-nés sains nés à terme. L'analyse réalisée par LC-MS/MS a permis d'identifier dans les méconiums des fœtus décédés et des nouveau-nés sains, la présence principalement d'acides biliaires primaires, mais également secondaires, tels que les acides chénodésoxycholique et cholique conjugués à la taurine, à la glycine et au sulfate et l'acide lithocholique. Ces mêmes acides biliaires ont été retrouvés dans la vésicule biliaire, à l'exception des acides conjugués au sulfate qui n'ont pas été détectés, suggérant que ceux-ci provenaient du liquide amniotique avalé contenant l'urine [254].

2. Les analyses métabolomiques ciblées réalisées sur le méconium

Righetti et al ont réalisé une analyse qualitative et quantitative par RMN d'une famille de 16 métabolites endogènes dans les extraits aqueux et organiques d'échantillons de méconiums excrétés au jour 1 et 3 suivant la naissance [256]. Dans la phase aqueuse, une augmentation globale en métabolites est observée entre le jour 1 et le jour 3. Dans la phase organique, une augmentation significative des concentrations en stérols et en acides gras libres est observée dans les méconiums collectés le 3^{ème} jour [256]. Les auteurs suggèrent que cette évolution est le reflet de l'alimentation des nouveau-nés par le lait maternel contenant environ 0,5 % de lipides [257].

Bearer et al ont analysé par GC/MS un des marqueurs biologiques connus du syndrome d'alcoolisme fœtal : le linoléate d'éthyle, à partir de 248 échantillons de méconium [258]. Les auteurs ont observé des teneurs de linoléate d'éthyle corrélées avec la quantité de boissons alcoolisées consommées par les mères. Ils ont conclu que le linoléate d'éthyle peut être utile en tant que biomarqueur de l'exposition

prénatale à l'éthanol, permettant d'aider à la pratique clinique et à la mise en place de prise en charge des nouveau-nés [258].

Des recherches de biomarqueurs d'expositions fœtales à des médicaments et des drogues ont également été réalisées par LC/HRMS dans différentes études. Ristimaa et al ont notamment réalisé une analyse ciblée et quantitative de 12 drogues et une analyse non ciblée des médicaments retrouvés chez 209 échantillons de méconium. Grâce à une base de données interne contenant les masses monoisotopiques et les formules moléculaires de 869 drogues et médicaments, ainsi que les temps de rétention pour environ la moitié de ces composés. Les auteurs ont pu identifier 77 médicaments et leurs métabolites incluant des opiacés, des stimulants de type amphétamines ainsi que des benzodiazépines. Le développement d'une méthode quantitative des drogues dans ce contexte permet de détecter rapidement et avec une très bonne sensibilité des drogues en néonatalogie. Ceci permettrait de diagnostiquer précocément l'exposition du nourrisson et d'assurer sa prise en charge [259].

Ostrea et al ont analysé et quantifié par GC/MS des métabolites de nicotine, de cotinine et de trans-3'-hydroxycotinine dans le méconium de nouveau-nés issus de mères exposées au tabac de manière active ou passive. Cette étude a été réalisée sur 55 échantillons de méconium et a démontré des concentrations en métabolites corrélées au degré de tabagisme déclaré par la mère [79].

L'impact des expositions fœtales aux pesticides et leurs métabolites sur l'apparition de malformations congénitales a beaucoup été étudié notamment au sein du projet PENEW (de l'anglais, Pregnancy and Environment Newborn Malformations) [47], [260], [261], ou encore du projet MecoExpo [262] en Picardie. Au sein de l'étude PENEW, des analyses ciblées de pesticides ont été réalisées par LC-MS/MS sur 700 échantillons de méconiums, permettant de quantifier 21 pesticides dont des carbamates et organophosphates ainsi que des herbicides [260]. Ces résultats ont montré une exposition significative du fœtus aux dithiocarbamates et aux pyréthroides. Le taux de détection le plus élevé a été observé pour les métabolites des organophosphates et des dithiocarbamates (probablement le mancozèbe), principal pesticide utilisé en Picardie. Ces différents résultats mettent en évidence l'exposition *in utero* à ces molécules, dont l'impact direct sur la santé de l'enfant doit être évalué [260].

3. Les analyses multi-omiques réalisées sur le méconium

Différentes études dans la littérature ont analysé le méconium ou de selles précoces par différentes approches omiques. Dans leur étude Wandro et al ont réalisé une analyse globale par GC/MS des métabolites ainsi qu'une analyse des communautés bactériennes par séquençage des ARN 16S dans des

échantillons de fèces de nourrissons collectés à partir du 7^{ème} jour suivant la naissance. Cette étude, portant sur l'analyse de 77 échantillons collectés à différents jours chez 32 nouveau-nés prématurés, visait à identifier une signature métabolique associée à l'apparition de septicémie ou d'entérocolite nécrosante. Au total, 224 métabolites endogènes ont été identifiés incluant des aminoacides, des acides biliaires, des acides gras, des nucléotides et des sucres. Cependant, aucune signature spécifique de ces pathologies n'a pu être révélée. Parallèlement, l'analyse de la composition du microbiote des fèces de nourrissons a mis en évidence des communautés microbiennes intestinales spécifiques aux nourrissons prématurés en comparaison avec des naissances à terme. Les résultats ont démontré une moindre richesse en *Bifidobacterium* et une dominance de *Staphylococcus* et des *Enterobacteriaceae* chez les enfants prématurés, dysbiose résultant de l'administration d'antibiotiques. De façon intéressante, les métabolomes fécaux étaient uniques à l'individu et n'étaient que faiblement associés à la composition bactérienne [263].

L'analyse conjointe du microbiote et du métabolome intestinal durant le développement périnatal a été réalisée par Del Chierico et al sur 70 échantillons de fèces issues de 21 nouveau-nés, prélevés entre le jour 1 et 30 suivant la naissance. Le but de cette étude est de corrélérer l'évolution de la composition en métabolites au cours des premiers jours de vie avec l'apparition d'espèces bactériennes intestinales. L'analyse métabolomique réalisée par RMN a permis d'identifier et de quantifier 33 métabolites incluant l'acide isocaproïque, la leucine, le malonate, la créatinine... Des analyses de corrélations entre la composition bactérienne et le métabolome ont permis d'associer la présence de certains métabolites à des bactéries (ou réciproquement). La Figure 24 montre notamment les corrélations positives ou négatives observées entre les concentrations des 33 métabolites quantifiés et l'abondance au sein du phylum des Actinobactéries. Par exemple, parmi les OTUs (de l'anglais, Operational Taxonomic Unit) définies au sein des Actinobactéries, l'abondance des *Propionibacterium* était positivement corrélée aux concentrations d'isovalérate, d'isoleucine, de la leucine, du glutamate, du propionate, du diméthylamine et de l'aspartate, et négativement corrélée à la concentration de l'acétate [115].

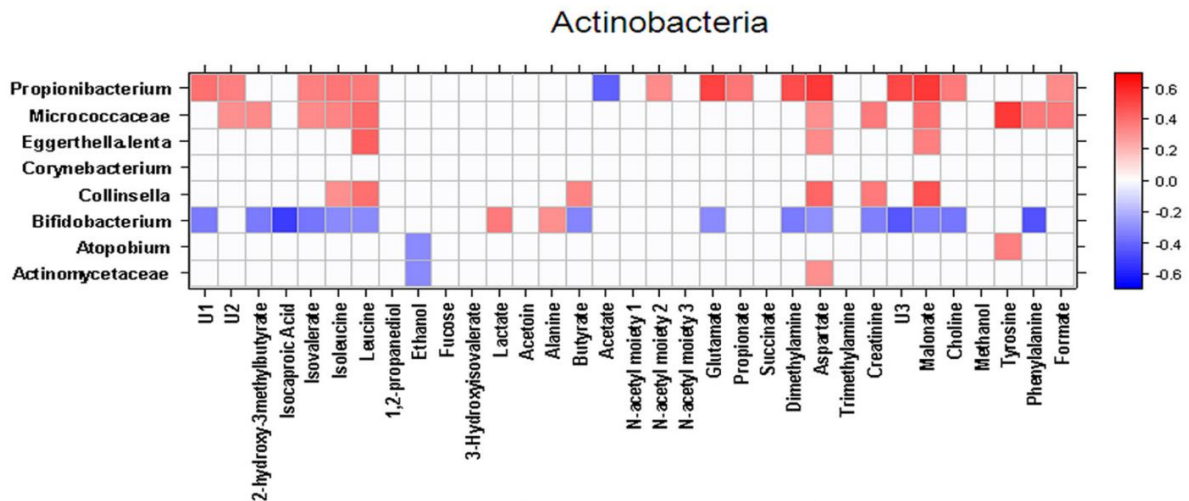


Figure 24 : Corrélations (heatmap) entre les concentrations des 33 métabolites et l'abondance des OTUs du phylum des Actinobactéries au sein d'échantillons de fèces de nourissons. Les codes couleurs permettent de distinguer les corrélations positives (rouge) des négatives (bleu) au sein des métabolites et des OTUs [115].

Dans une étude récente, Bittinger et al ont réalisé une étude multi-omique sur 88 échantillons de méconiums collectés entre le jour 1 et 7 suivant la naissance. L'objectif étant d'analyser la colonisation bactérienne au cours des premiers jours de vie. Lors de cette étude, trois approches omiques sont combinées : la métagénomique, la métabolomique et la métabolomique (LC/HRMS) – approches identiques à celles que nous mettons en place. Une analyse par qPCR quantitative a permis d'observer un taux d'ADN d'origine humaine plus élevé au sein des méconiums collectés à des temps précoces, comparés aux méconiums collectés plus tardivement. Les auteurs ont indiqué un taux d'ADN humain supérieur à 75 % dans les méconiums prélevés avant 16 heures, et inférieur à 75 % au-delà de 16 heures (Figure 25). Cette transition se faisant plus tôt chez les naissances par voie basse comparée aux naissances par césarienne [264]. Les profils métaboliques sont similaires dans les échantillons prélevés dans les 16h suivant la naissance. A l'inverse, 45 métabolites présentent des profils différents au sein des échantillons collectés au-delà de 16h après l'accouchement, où plus de bactéries sont détectées, comparé aux temps plus précoces. Parmi ces 45 métabolites, plusieurs acides aminés (par exemple, la sérine ou la thréonine) voient leurs concentrations diminuer, tandis que d'autres métabolites (tels que l'acétate ou le succinate) voient leurs concentrations augmenter. L'analyse du protéome du méconium dans cette étude a mis en évidence une majorité de protéines d'origine humaine (1071 *versus* 163 protéines d'origine bactérienne). La composition protéique est similaire entre les échantillons de méconiums prélevés au cours des 16 premières heures après la naissance, et ceux prélevés au delà. Les auteurs ont indiqué que les abondances des protéines étaient influencées par la présence bactérienne qui est davantage représentée au-delà de 16h après la naissance. Ils ont également rapporté que

l'abondance des protéines serait plutôt influencée par la colonisation du tube digestif du nouveau-né, et non pas seulement liée au temps écoulé post- naissance. L'ensemble des résultats de cette étude permettent de rendre compte que la colonisation bactérienne conditionne une part significative du métabolome et du protéome intestinal chez les nouveau-nés [264].

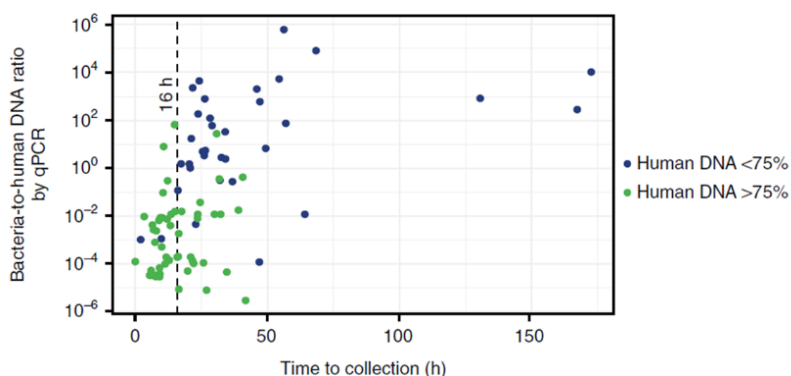


Figure 25 : Estimation du rapport de quantités de l'ADN humain par rapport à l'ADN bactérien détectés dans les échantillons de méconium (analyse par qPCR) [264].

Conclusion

Après avoir parcouru différentes études d'analyses métabolomiques ciblées et non ciblées réalisées sur des échantillons de méconium ainsi que sur des échantillons de premières selles, nous présentons une vue d'ensemble des principales études publiées dans la littérature (Tableau 5). Ce tableau indique l'approche suivie pour chaque étude, le jour de collecte des échantillons de méconium/premières selles, le nombre et l'origine des métabolites identifiés ainsi que la technique analytique utilisée pour chaque étude.

Tableau 5. Tableau récapitulatif des principales études métabolomiques réalisées sur du méconium et/ou premières selles, menées par des approches ciblées ou non ciblées.

Approche métabolomique	Jours de collecte post-naissance	Nombre de métabolites annotés	Technique analytique utilisée	Auteurs
Ciblée	Jours 1 à 30	33 métabolites endogènes	RMN	Del Chierico et al [115]
Ciblée	Jours 1 à 3	16 métabolites endogènes	RMN	Righetti et al [256]
Ciblée	Jours 1 à 6	14 métabolites endogènes	GC/MS	Kumagai et al [255]

Non ciblée	Non définis	77 métabolites exogènes	LC/MS (RP-C18 et Luna PFP)	Ristimaa et al [259]
Non ciblée	Jours 1 à 2	10 métabolites endogènes 4 métabolites exogènes	LC/MS (RP-C18)	Peng et al [253]
Non ciblée	Jours 1 à 7	45 métabolites endogènes	LC/MS (RP-C18)	Bittinger et al [264]
Non ciblée	Jours 7 à 75 (matières fécales)	224 métabolites endogènes	GC/MS	Wandro et al [263]

Présentation du projet de recherche

Au cours de cette introduction, nous avons décrit les allergies alimentaires et les facteurs pouvant prédisposer à cette pathologie. Les facteurs périnataux énoncés ont montré l'intérêt d'étudier le méconium, matrice fœtale la plus accumulative, que nous avons ensuite décrite. Un descriptif des différentes sciences « omiques » a ensuite été réalisé, mettant en avant l'intérêt de chacune des approches de « métagénomique ciblée de l'ADN 16S » de « métaprotéomique » et de « métabolomique », dans la démarche d'une analyse globale de la composition du méconium.

Nous nous sommes attachés lors de cette partie introductive à détailler les différentes étapes de l'approche métabolomique, faisant l'objet principal de ce travail de thèse. Nous avons d'abord présenté le déroulement d'une analyse métabolomique non ciblée, présentant les différentes techniques analytiques utilisées. Nous avons ensuite détaillé les étapes du « workflow » métabolomique par LC/HRMS, en s'intéressant particulièrement aux étapes de préparation des échantillons, aux traitements des données et à l'annotation des métabolites, étapes primordiales dans le cadre de ce travail de thèse.

Le premier objectif de ma thèse a été de mettre en place une méthodologie d'analyse pour l'exploration du métabolome du méconium, matrice jamais étudiée au niveau du laboratoire et étant finalement peu décrite dans la littérature. Ces développements et optimisations méthodologiques ont été réalisés sur 2 échantillons de méconiums disponibles au laboratoire. Les résultats obtenus font l'objet du premier chapitre de ce manuscrit, décrivant une première cartographie de la composition métabolique du méconium, à travers la méthodologie développée.

Le deuxième objectif de ma thèse a été de valider l'approche méthodologique développée en l'appliquant à 35 échantillons de méconium collectés longitudinalement auprès de 11 nouveau-nés issus de naissances vaginales. Ces différents prélèvements permettront d'approcher la variabilité interindividuelle et temporelle de la composition du méconium durant les 3 premiers jours de vie. Des analyses métagénomiques ciblées et métaprotéomiques ont été réalisées en parallèle sur ces mêmes échantillons, en collaboration respectivement avec l'Institut Micalis de l'INRAe de Jouy en Josas et le laboratoire LI2D du CEA de Marcoule, experts dans ces domaines. Ces analyses omiques indépendantes puis multiomiques (en intégrant les différentes données), permettraient d'avoir la description la plus complète du méconium humain à ce jour. Ces échantillons de méconiums, très précieux, ont également permis de démontrer les expositions prénatales aux nanoparticules de TiO₂, dans le cadre du travail de recherche de Mme Guillard auquel j'ai été associée [77].

L'objectif final de ce projet de recherche est d'appliquer les trois approches omiques dans l'analyse des méconiums collectés au sein d'une cohorte française mère-enfant EDEN (Etude des Déterminants pré et post natals du développement et de la santé de l'Enfant, www.eden.vjf.inserm.fr). Cette cohorte de par son effectif, du nombre d'informations collectées auprès des mères durant le dernier trimestre de grossesse, sa durée de suivi des enfants, et de sa biobanque, représente la plus vaste étude portant sur les facteurs pré/post natals en France. Pour mon projet de thèse, 308 échantillons de méconiums sont disponibles permettant *in fine* de tester l'association entre la composition du méconium et des facteurs d'expositions tels que l'alimentation de la mère, leur lieu de vie (rural ou urbain), leur exposition à des polluants, ainsi que le rôle de certains facteurs socio-économiques. L'association entre cette composition et la survenue d'allergies alimentaires au cours des 5 premières années de vie de l'enfant sera ensuite évaluée, et les signatures de prédispositions à l'allergie identifiées.

RESULTATS ET DISCUSSION

CHAPITRE I : MISE EN PLACE D'UNE METHODOLOGIE D'ANALYSE POUR L'EXPLORATION DU METABOLOME DU MECONIUM

I. Introduction

Le méconium représente les matières fécales excrétées au cours des 24 heures (dans 90 % des cas) à 48 heures suivant la naissance des nouveau-nés [265]. Comme décrit dans la partie introductive, cette matrice est biologiquement complexe, composée d'eau, de métabolites, de protéines, de sécrétions intestinales, de cheveux néonataux... en proportions potentiellement très variables, rendant l'extraction et l'analyse de ses composants difficile [266]. L'analyse métabolomique globale du méconium nécessite d'extraire de façon la moins sélective possible ses métabolites aux propriétés physico-chimiques très variées. De plus, afin de limiter au maximum toute source de variabilité ou de dégradation des métabolites, la procédure d'extraction devrait être réalisée avec un minimum d'étapes. Au-delà de l'étape de préparation de l'échantillon, les différentes étapes de l'approche métabolomique globale décrites précédemment doivent être maîtrisées : de l'acquisition aux traitements des données (annotations, analyses statistiques...).

Avant de commencer quelconque développement méthodologique, une recherche bibliographique approfondie a été menée, permettant de saisir les connaissances, les méthodologies et les données déjà disponibles. Comme présenté en introduction, la grande majorité des articles publiés en analyses métabolomiques du méconium concernent des analyses ciblées. Ainsi, afin de définir et d'optimiser les différentes étapes de préparation de l'échantillon de méconium, nous nous sommes basés sur les données de la littérature visant à l'analyse ciblée et non ciblée du méconium mais aussi de matières fécales (de nourrisson et d'adulte).

Ce chapitre est dédié à la mise en place des procédures expérimentales nécessaires à l'analyse métabolomique non ciblée du méconium, de la préparation de l'échantillon jusqu'à l'acquisition des données spectrales. Le « workflow » utilisé pour l'analyse des données obtenues (filtres, nettoyages, annotations), sera également décrit. Nous nous focaliserons essentiellement sur les techniques analytiques disponibles au niveau de la plateforme du SPI (LC-HRMS), et de l'institut parisien de chimie moléculaire au sein duquel des analyses ont été menées en parallèle sur un FT-ICR.

A. Préparation des échantillons

La préparation de l'échantillon comprend la collecte, le transport, le stockage, un éventuel prétraitement puis l'extraction des métabolites en elle-même. Concernant la récolte et le stockage, la plupart des études précisent que les échantillons de méconium sont récupérés directement des couches, placés dans des tubes stériles et conservés à -20 °C ou à -80 °C dès leur collecte [70], [256]. En effet, ceci permet de prévenir toute dégradation des métabolites liée à la labilité de certains composés ou à une activité bactériologique pouvant modifier la composition du métabolome. Il est à noter que le prélèvement sur la couche peut conduire à une contamination par les urines du nouveau-né, qui ne sera pas forcément relevée, ainsi que par des composés issus de la couche en elle-même.

1. Prétraitement des méconiums

Dans certaines études, des prétraitements de l'échantillon de méconium ont été réalisés. Le méconium est alors considéré à l'état brut (sans prétraitement), séché ou lyophilisé [51], [267]. L'élimination de la phase aqueuse a plusieurs intérêts : la possibilité de standardiser les échantillons par rapport à leur masse sèche, ainsi que de diminuer les variabilités interindividuelles au niveau de la teneur en eau. Meyer et al ont également rapporté que les molécules subissant la plus grande suppression de signal dans les méconiums analysés par spectrométrie de masse étaient les molécules polaires [46]. Dans ce contexte, plusieurs techniques peuvent être envisagées pour éliminer la phase aqueuse. La technique de séchage à haute température, comme utilisée par Berton et al sur des échantillons de méconiums, est à mettre en regard des molécules analysées. En effet, un séchage au-delà de 50 °C peut impacter significativement la composition en métabolites thermolabiles [51]. La lyophilisation semble ainsi plus adaptée dans le cas d'analyses globales d'échantillons de méconiums. Afin de vérifier que l'ajout de l'étape de lyophilisation n'impacte pas la composition globale des matières fécales, Moosmang et al ont comparé le nombre de variables détectées par LC/HRMS sur des fèces brutes et des fèces lyophilisées. Ces deux conditions ont permis de détecter respectivement 7289 et 7280 variables dans l'échantillon brut et lyophilisé. Ces deux méthodes permettent donc de couvrir une plage métabolique comparable [268]. Cheng et al ont obtenu des résultats similaires lors de l'analyse ciblée de 7 acides gras à chaînes courtes dans des matières fécales brutes ou lyophilisées [267]. Nous avons ainsi retenu la méthode de lyophilisation comme méthode de prétraitement des échantillons de méconium.

2. Extraction des métabolites : pré-extraction et choix du solvant

L'extraction des métabolites doit requérir un minimum d'étapes, afin de conserver des concentrations suffisantes en métabolites pour permettre leur détection et identification [269]. Pour les analyses non ciblées, l'extraction des métabolites du méconium (ou des fèces) est souvent réalisée par l'utilisation de différents solvants organiques et/ou aqueux. Dans leur étude, Wandro et al ont réalisé l'extraction des métabolites par un mélange de solvants (isopropanol, acétonitrile et eau, 3:3:2, v/v/v), chez 77 échantillons de fèces de nourrissons. Cette étude a permis d'identifier 224 métabolites endogènes par GC/MS, incluant des familles telles que des aminoacides, des acides gras ou biliaires, des sucres ou des lipides [263]. Peng et al ont comparé l'efficacité d'extraction des métabolites sur des échantillons de méconiums lyophilisés par différents solvants : le méthanol, l'acétonitrile, le chloroforme et l'eau. Cette étude montre que le MeOH pur est le solvant optimal pour l'extraction des métabolites de type aminoacides, acides biliaires et lipides [253]. D'un autre côté, Moosmang et al ont testé ces mêmes solvants pour l'extraction métabolique à partir de matières fécales lyophilisées. Les auteurs ont démontré un meilleur rendement d'extraction et une meilleure couverture métabolique avec de l'eau pure. Dans une autre étude sur des matières fécales, une extraction des métabolites par une précipitation des protéines au méthanol pur puis une analyse par LC/MS sur des colonnes HILIC et C18 a permis l'identification de 667 métabolites endogènes et exogènes [270].

Au vu de ces résultats, et en s'aidant des protocoles mis en place au laboratoire sur d'autres matrices biologiques telles que le lait maternel [271] ou le sérum humain [197]. Nous avons fait le choix de réaliser l'extraction à l'aide d'un mélange de solvants : méthanol/eau, permettant en théorie d'extraire un maximum de métabolites aux différentes propriétés physico-chimiques.

3. Extraction des métabolites : homogénéisation de la matrice

Au-delà du solvant d'extraction utilisé, différentes techniques aidant à l'homogénéisation et à l'extraction des métabolites du méconium ou des fèces ont été utilisées dans la littérature. Parmi ces techniques, Berton et al ont utilisé une agitation « vigoureuse » pendant 45 minutes [272], tandis qu'un bain à ultrasons pendant 30 minutes a été utilisé par Peng et al [253]. Des méthodes plus drastiques comme de la lyse mécanique par des billes en céramique ont également été utilisées [264]. Cette dernière méthode de lyse est déjà utilisée au laboratoire sur des fèces de souris, ou sur des matrices biologiques difficiles à dissocier tels que des tissus ou des biopsies d'organes. Au cours de notre étude, nous comparerons les méthodes d'homogénéisation du méconium par une lyse mécanique et une sonde à ultrasons.

4. Séchage et reprise des métabolites extraits

A la suite de l'étape d'extraction des métabolites, une étape d'évaporation sous un flux d'azote est souvent réalisée pour l'analyse par LC/HRMS, permettant : i) de solubiliser les culots secs contenant les métabolites dans les phases mobiles spécifiques au système analytique, ainsi que ii) d'ajuster la concentration/dilution finale de l'échantillon [253], [270]. Le volume de reprise à cette étape est notamment à optimiser, afin de trouver le bon équilibre entre effet de matrice et effet de dilution. Ces deux effets peuvent directement impacter le nombre et la qualité des signaux détectés. Des volumes de reprise très variables ont été utilisés dans la littérature, allant de 300 μL à 5 mL pour une masse initiale équivalente à 10 mg de méconium ou de fèces [253], [273], [274]. Cette étape sera optimisée dans nos conditions d'extraction et d'analyse.

La mise en place du protocole de préparation de l'échantillon du méconium s'est donc basée sur l'ensemble de ces études, en optimisant les étapes de prétraitement des méconiums (séchage et homogénéisation) ainsi que le volume de reprise des culots métaboliques secs en amont de l'analyse. Ces développements méthodologiques ont été réalisés sur deux échantillons de méconiums disponibles au laboratoire.

B. Séparation chromatographique en amont des acquisitions de données spectrales

Le développement méthodologique est réalisé par des analyses par LC/HRMS, du fait des appareils disponibles au laboratoire. L'analyse sur le FT-ICR est réalisée à la fin de la mise en place de la méthodologie de préparation de l'échantillon, permettant de comparer les deux techniques analytiques.

En LC/HRMS, différentes colonnes chromatographiques aux propriétés complémentaires sont utilisées dans la littérature, permettant de couvrir l'ensemble du métabolome à partir d'un même extrait. Boudah et al ont comparé trois colonnes chromatographiques pour l'analyse globale du métabolome du sérum humain : la ZIC-*p*HILIC (zwitterionic polymer hydrophilic column), la RP-C18 (reverse phase) et la PFPP (pentafluorophenyl propyl). Les auteurs ont décrit que la colonne ZIC-*p*HILIC a permis à elle seule de couvrir plus de 80 % des métabolites détectés et annotés (soit 219 sur 266 métabolites décrits) [197]. Ainsi, le développement méthodologique présenté dans ce chapitre a été réalisé uniquement sur la colonne ZIC-*p*HILIC, permettant la séparation et la détection de la grande majorité des métabolites extraits [198]. Afin d'augmenter la couverture métabolique dans le méconium, des séparations utilisant la colonne RP-C18 seront évaluées en complément dans le chapitre suivant.

C. Conclusion

Au vu de ces différentes données, nous avons établi le protocole présenté en Figure 26 pour le développement et l'optimisation de notre démarche expérimentale. Ce protocole inclut : i) une étape de préparation de l'échantillon (séchage, homogénéisation ainsi que l'optimisation des volumes de reprise), ii) une étape d'acquisition des données par LC/HRMS, iii) une étape de traitement des données et d'annotation des variables à partir d'une base de données interne et iv) une dernière étape des expériences MS/MS réalisées sur les annotations obtenues, permettant d'avoir une identification formelle des métabolites [210]. Au-delà du développement méthodologique, ces travaux permettront de fournir une première cartographie du métabolome du méconium.

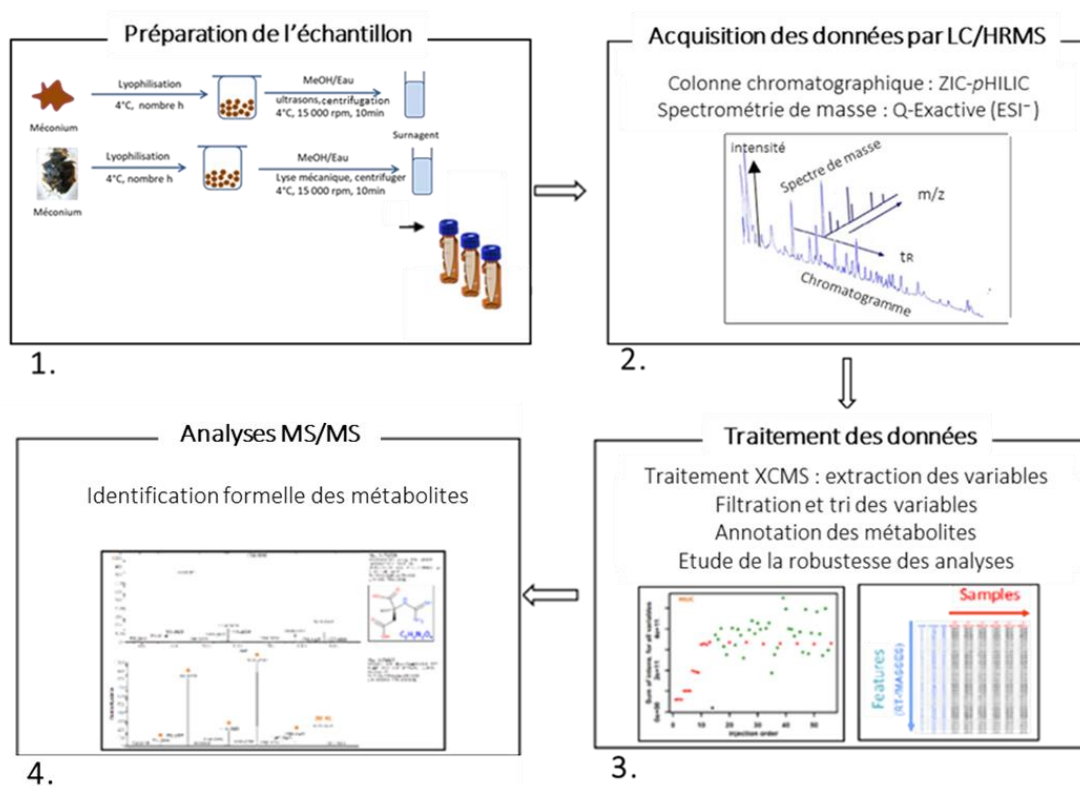


Figure 26 : Méthodologie suivie dans l'analyse métabolomique non ciblée du méconium.

II. Matériels et méthodes

A. Réactifs et produits chimiques

Le carbonate d'ammonium, l'eau de haute pureté, l'hydroxyde d'ammonium et le méthanol proviennent de chez Honeywell-Riedel-deHaënTM (Seelze, Allemagne). L'acétonitrile (ACN) provient de chez VWR chemicals (Fontenay-sous-Bois, France). L'acétate d'ammonium et l'acide acétique proviennent de chez Sigma Aldrich (Saint Quentin Fallavier, France). Les 1200 composés de référence de la chimiothèque interne, présentant le plus haut niveau de pureté, proviennent de chez Sigma Aldrich (Saint Quentin Fallavier, France). Les solutions mères des étalons internes utilisés sont préparées à 1 mg/mL dans les solvants recommandés, puis dilués dans les solvants de reprises adaptés au mode de séparation chromatographique, à la concentration de 5 µg/mL par composé de référence (Annexe 3).

B. Echantillons de méconiums

Les échantillons de méconiums utilisés pour ces phases d'optimisations ont été obtenus chez deux naissances par voie basse, de sexe féminin (échantillon A) et masculin (échantillon B). La collecte des échantillons a été réalisée par les parents, consentants à cette étude. Les échantillons sont collectés directement sur la couche au moment du change, à l'aide de tubes stériles de 50 mL. Les tubes ont été congelés aussitôt à - 20 °C, puis transférés et stockés à - 80 °C au laboratoire jusqu'à leur analyse. Plusieurs prélèvements ont été réalisés par nouveau-né, et chaque prélèvement est réalisé dans un nouveau tube et étiqueté à la date et l'heure d'excrétion. Les informations concernant ces deux échantillons sont regroupées dans le Tableau 6.

Tableau 6 : Description des échantillons de méconium A et B utilisés pour les phases d'optimisations de l'analyse métabolomique.

	A	B
Naissance	Avril 2017	Décembre 2017
Jour et heure de collecte post-naissance	J1 (5h20, 10h et 14h30) J2 (32h30)	J1 (< 24h) (Heure non disponible)
Sexe	Féminin	Masculin

Au début de cette étude, seul l'échantillon A était disponible. Pour cette raison et afin d'avoir une quantité suffisante et homogène de matériel pour les développements méthodologiques, tous les prélèvements issus de cet enfant et présentant des textures et aspects similaires ont été décongelés, regroupés et homogénéisés (les quatre échantillons inclus dans le tableau). Afin d'éviter les cycles de

congélation et de décongélation, des aliquots de 200 à 700 mg ont été transférés dans des tubes cryogéniques et stockés à - 80 °C jusqu'à leur analyse. Les aliquots ont volontairement été préparés avec des masses variables permettant de mimer les échantillons disponibles au sein de la cohorte française mère/enfant EDEN, contenant des masses variables (et inconnues).

C. Préparation des échantillons de méconium

1. Lyophilisation des méconiums

Des aliquots de méconiums ont été lyophilisés à l'aide de l'instrument lyophilisateur Triad™ Labconco. Les températures du plateau et du piège ont été fixées respectivement à + 4 °C et - 83 °C, pour une pression de 0,180 mbar. Les échantillons congelés sont disposés sur un portoir préalablement placé à - 80 °C pendant 10 à 15 minutes. Les bouchons des tubes sont à moitié ouverts avant leur mise en place dans le lyophilisateur. Différents temps de lyophilisation ont été évalués, variant de 17 à 48 heures. Après optimisation de ce temps, les méconiums lyophilisés ont été regroupés dans un seul tube Eppendorf et homogénéisés à l'aide de spatules stériles.

2. Homogénéisation des méconiums : test comparatif de la sonde à ultrasons et de la lyse mécanique

a) Test de différentes conditions d'homogénéisation

Une masse de 10 mg de méconium lyophilisé est pesée respectivement dans des tubes Eppendorf (pour l'homogénéisation par la sonde à ultrasons) et des tubes « Precellys » de 1,5 mL (tubes CK14, contenant des billes en céramiques de 1,4 mm de diamètre), pour l'homogénéisation par la lyse mécanique. Une première étape de solubilisation du méconium lyophilisé est réalisée en ajoutant un volume de 150 µL d'H₂O suivie de 30 secondes d'agitation. Les échantillons de méconium sont placés dans de la glace dès solubilisation. Un volume de 600 µL de MeOH est ensuite ajouté puis le mélange est agité pendant 30 secondes. L'homogénéisation des échantillons de méconium placés dans de la glace est soit réalisée par 3 cycles de 10 secondes à l'aide d'une sonde à ultrasons (Mettler Toledo, Viroflay, France), soit par 3 cycles de 30 secondes d'une lyse mécanique (6500 tr/min à 4°C) à l'aide d'un appareil Precellys 24® (Bertin Technologies, Montigny-le-Bretonneux, France). Pour favoriser la précipitation des protéines, les échantillons sont maintenus dans de la glace durant 1h30. Une centrifugation est réalisée à 20000 x g, à 4 °C durant 15 minutes. Sept cents microlitres de surnageant sont récupérés (en évitant la couche lipidique supérieure et le culot protéique au fond du tube), puis transférés dans un nouveau tube (tube Eppendorf de 1,5 mL). Les surnageants sont ensuite séchés sous un courant d'azote à 30 °C durant 3

heures (TurboVap®, NC, Etats Unis). Les culots métaboliques secs sont stockés à - 80 °C jusqu'à leur analyse.

En parallèle, des homogénéisations de 10 mg de méconium lyophilisé dans 220 µL d'H₂O ont été réalisées selon les mêmes conditions (sonde à ultrasons vs. lyse mécanique ; 5 réplicats/condition). Les surnageants sont ensuite récupérés par centrifugation pour une analyse quantitative et qualitative des protéines solubles extraites.

b) Analyse des protéines extraites

b.1) Dosage des protéines totales dans les extraits de méconium

Un dosage des protéines a été réalisé à partir des solutions issues de chaque condition d'homogénéisation réalisée dans l'eau. Le dosage des protéines a été réalisé par la méthode BCA (BiCinchoninic acid Assay), à l'aide d'un kit commercial (BCA Protein Assay Kit, Pierce, Thermo Fisher Scientific). Cette méthode combine la réduction de l'ion cuivrique Cu²⁺ en ion cuivreux Cu⁺ par les protéines dans un milieu alcalin (réaction de Biuret). Une chélation de deux molécules d'acide bichinchoninique BCA avec un ion cuivreux est ensuite réalisée. Le complexe ainsi formé absorbe fortement à 562 nm, avec une absorbance directement proportionnelle à la concentration en protéines présentes dans l'échantillon (gamme de concentration linéaire entre 0,5 µg/mL et 2 mg/mL). Une gamme étalon (0 à 1 mg/mL, 8 points de gamme en duplicats) est réalisée avec une protéine standard, la sérum albumine bovine (BSA), fournie dans le kit.

b.2) Analyse des protéines extraites par électrophorèse SDS-PAGE

Une analyse électrophorétique des protéines présentes dans les solutions issues de chaque condition d'homogénéisation en H₂O a également été réalisée. Le mode opératoire préconisé par le constructeur (Invitrogen) a été suivi. Il est utilisé un gel de polyacrylamide pré-coulé Bis-Tris, NuPAGE™ Novex™ (4-12 %, épaisseur 1,0 mm), et le tampon de migration fourni. Dix microlitres d'une solution ajustée à 1 mg/mL de protéines sont mélangés à 10 µL de solution de dépôt (tampon réducteur et dénaturant). Ce mélange est chauffé à 70 °C sous agitation à 300 rpm pendant 10 minutes, permettant de dénaturer les protéines. Vingt microlitres d'échantillon sont ensuite déposés dans chaque puit du gel, et la migration est réalisée à 200 V, 120 mA, 25 W durant 35 minutes. Les réactifs et conditions de migration appliquées permettent de séparer les protéines de masses s'échelonnant de 3,5 à 260 kDa. A la fin de la migration, le gel est récupéré, rincé 3 fois pendant 5 minutes dans l'eau, puis coloré avec du bleu de Coomassie (Imperial protein stain, Thermo Fisher Scientific) selon les recommandations du fournisseur.

3. Reprise des échantillons de méconium pour l'analyse en LC/HRMS

Le culot métabolique sec est solubilisé dans 320 µL d'une solution aqueuse de carbonate d'ammonium à 10 mM, pH 10,5. Après 30 secondes d'agitation au Vortex™, une centrifugation est réalisée (20000 *g*, 4 °C, durant 15 minutes). Le surnageant est transféré dans un nouvel Eppendorf, puis 480 µL d'ACN sont ajoutés. Une seconde étape de centrifugation est réalisée et 100 µL du surnageant final sont transférés dans un vial d'injection. L'échantillon de QC est préparé en mélangeant 50 µL de chaque échantillon à analyser. Cent µL du mélange de QC sont redistribués dans chaque vial d'injection de QC, puis des échantillons de QC dilués sont préparés (dilués au 1/2, 1/4 et 1/8) dans une solution de carbonate d'ammonium/ACN (40:60, v/v). Cent µL de ce tampon sont placés dans un flacon d'injection, afin de réaliser un blanc analytique. Enfin, 5 µL du mélange d'étalons internes sont ajoutés aux échantillons biologiques et aux QC et dilutions de QC (Annexe 3).

D. Analyse des métabolites par LC/HRMS

1. Conditions chromatographiques

La séparation chromatographique a été réalisée avec un système de chromatographie de type Dionex Ultimate 3000. La colonne SeQuant ZIC-*p*HILIC (5 µm ; 2,1 × 150 mm ; Merck, Darmstadt, Allemagne) est utilisée avec des phases mobiles composées d'une phase aqueuse de carbonate d'ammonium à 10 mM et pH 10,5 en A, et d'une phase organique à 100 % d'ACN en B. La séparation chromatographique dure 42 minutes, débutant par une élution à 80 % de B durant 2 minutes, suivie d'un gradient linéaire allant de 80 à 40 % de B durant 12 minutes. Le système chromatographique est ensuite rincé pendant 5 minutes à 100 % en tampon A. Enfin, un reconditionnement de la colonne est réalisé dans les conditions d'élutions initiales durant 25 minutes. Le débit appliqué est de 0,2 mL/min et un volume de 10 µL d'échantillon est injecté. Les températures du système HPLC sont fixées à 4 °C pour l'injecteur automatique, et 15 °C pour la colonne chromatographique.

2. Analyse par spectrométrie de masse

L'acquisition a été réalisée sur un spectromètre de masse de type Q-Exactive (Thermo Fisher Scientific, Les Ulis, France), utilisant une source electrospray en mode d'ionisation négative avec des tension et température de capillaire respectivement de - 3000 V et 280 °C. Les pressions de gaz de nébulisation et de gaz auxiliaire (azote) sont respectivement fixées à 60 et 10 unités arbitraires. Le pouvoir résolutif du spectromètre de masse est fixé à 50000 pour une largeur à mi-hauteur (FWHM) à *m/z* 200. La détection est réalisée sur une gamme de *m/z* de 75 à 1000. Une calibration externe du spectromètre de masse est réalisée chaque début de semaine et/ou début de séquence en utilisant une solution d'acétate

d'ammonium à 2 mM préparée au laboratoire. Les expériences MS/MS ont été réalisées en mode d'activation non résonnante à des valeurs d'énergies de collision normalisées de 10, 20 et 40 %.

3. Traitement des données LC/HRMS

L'extraction automatique des variables est réalisée sur l'XCMS.3.0.0 du Workflow4Metabolomics de Galaxy [211]. Les paramètres utilisés sur l'XCMS.3.0.0 sont détaillés en Annexe 4.

L'annotation des variables est réalisée à partir d'une base de données interne au laboratoire (base de données précédemment décrite par Mikail Berdi [41]). L'annotation est réalisée en utilisant un script développé sous R au sein du laboratoire. La tolérance en temps de rétention fixée suit la fonction : $\Delta RT = X \pm RT^Y$ (avec X et Y fixés à 5 et 0,8 respectivement). La tolérance dans la mesure de masse est fixée à 10 ppm.

E. Analyse des métabolites par FT-ICR

L'analyse des culots métaboliques secs (obtenus à la suite du séchage des surnageants sous azote, cf Partie II.C.2.a) a été réalisée sur un instrument FT-ICR-7-T (Solarix-XR, Bruker, Brême, Allemagne). Celui-ci est équipé d'une cellule ICR (résonnance cyclotronique ionique) de dernière génération dite « dynamically harmonized cell » (Para Cell™), et d'une source electrospray opérant dans le mode d'ionisation négatif (ESI⁻). Toutes les caractéristiques du spectromètre de masse ont été décrites par Habchi et al [204].

L'extraction des variables est réalisée manuellement à l'aide du logiciel « Data analysis » (Bruker Daltonics, Bremen, Germany), générant un fichier en format « .xlsx ».

L'annotation des variables est réalisée manuellement en comparant les masses expérimentales précises aux masses théoriques enregistrées dans la même base de données que celle utilisée en LC/HRMS [41]. La tolérance dans la mesure de masse est fixée à 1 ppm.

III. Résultats expérimentaux et discussion

A. Développement et optimisation de la procédure de préparation des échantillons

1. Lyophilisation du méconium

Dans un premier temps, nous avons optimisé le temps de lyophilisation nécessaire pour éliminer la totalité de la phase aqueuse des échantillons. Six aliquots d'échantillons de méconium bruts, de masses variant entre 190 et 420 mg, ont été inclus. Comme indiqué dans la partie II.B, les aliquots de méconium ont volontairement été préparés avec des masses variables, permettant de mimer les échantillons disponibles de la cohorte EDEN, contenant des masses variables en échantillons. Les six tubes sont placés dans le lyophilisateur et deux tubes sont sortis aléatoirement au bout de 17, 24 et 48 heures de lyophilisation, pesés puis réservés à - 80 °C. Le pourcentage de la masse perdue a été calculé au sein de chaque réplicat et est représenté en Figure 27.A. Nous observons des pourcentages proches de la masse perdue au sein des 2 réplicats (17h, 24h et 48h), avec néanmoins plus de variabilité au fur et à mesure où le temps augmente. Au sein de chaque réplicat, l'échantillon qui perd le moins d'eau ne correspond pas à un échantillon présentant une masse en méconium plus élevée de départ. En globalité, les résultats indiquent un pourcentage stable de la masse d'eau perdue au cours du temps. Ceci peut prédire à une élimination totale de l'eau dès 17 heures de lyophilisation, et ce quelle que soit la masse de l'échantillon de départ. Différents temps de lyophilisations ont été utilisés dans la littérature, de 15 heures [267], d'une nuit (environ 18 heures) [275], ou de 36 heures [268], sans indication de la masse totale perdue au cours de ce temps. La perte massique moyenne enregistrée dans notre étude est d'environ 76 %, représentative de la teneur en eau dans le méconium décrite dans la littérature (entre 70 et 80 % [3]).

En conclusion, étant donné que la quantité de méconium lyophilisée reste stable de 17 à 48 heures et que 76 % de la masse du méconium brut est perdue. Nous considérons que toute la quantité d'eau est éliminée au bout de 17 heures de lyophilisation, temps qui sera retenu dans la suite. Enfin les échantillons de méconiums lyophilisés sont regroupés, broyés à l'aide d'une spatule stérile puis réservés à - 80 °C jusqu'à leur analyse (Figure 27.B).

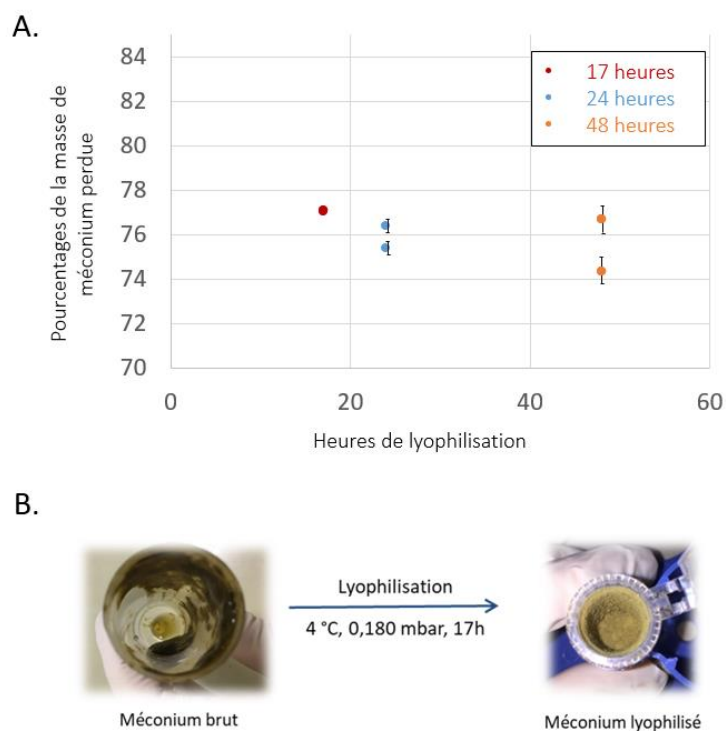


Figure 27 : A. Pourcentage de la masse en méconium brut perdue au cours du temps de lyophilisation. B. Conditions de lyophilisation retenues pour cette étude.

2. Méthode d'homogénéisation du méconium lyophilisé

La solubilisation du méconium lyophilisé ne se faisant pas complètement par simple ajout de solvant et par agitation, même vigoureuse (type vortex). Nous avons testé deux techniques d'homogénéisation : la sonde à ultrasons et la lyse mécanique. Afin de comparer le potentiel d'extraction global autorisé par ces deux méthodes, nous nous sommes basés sur l'analyse quantitative et qualitative des protéines extraites. Ceci, en émettant l'hypothèse que l'efficacité d'extraction des protéines constitue un reflet de l'efficacité d'extraction globale des molécules de la matrice (y compris celle des métabolites). Il était en effet difficile de comparer les profils métaboliques par LC/HRMS à cette étape, étant donné que la suite du protocole de préparation de l'échantillon est en cours de développement.

a) Analyse quantitative des protéines extraites de la matrice méconium

L'analyse des protéines solubles extraites de la matrice méconium est réalisée sur cinq réplicats pour chaque condition d'homogénéisation (sonde à ultrasons et lyse mécanique). Les masses en protéines quantifiées varient de 2,9 à 3,1 mg pour la sonde à ultrasons et de 3,2 à 3,9 mg pour la lyse mécanique, pour une masse initiale en méconium lyophilisé de 10 mg (Figure 28). L'utilisation de la sonde à ultrasons montre une meilleure répétabilité au sein des cinq réplicats, avec un coefficient de variation de 4 % (vs.

8 % pour la méthode de la lyse mécanique). Néanmoins la quantité moyenne en protéines extraites par la lyse mécanique est en moyenne 1,15 fois plus élevée que la quantité moyenne extraite par la sonde à ultrasons. Afin d'évaluer si l'ensemble des protéines présentes dans le méconium ont été extraites, nous avons procédé par une seconde extraction des protéines à partir du culot obtenu après centrifugation. A ce culot sont ajoutés 220 μ L d'eau, et une nouvelle homogénéisation est réalisée selon la même méthodologie que précédemment. Les quantités des protéines retenues lors de cette seconde extraction sont de 0,7 et 0,6 mg respectivement pour la méthode de la sonde à ultrasons et de la lyse mécanique.

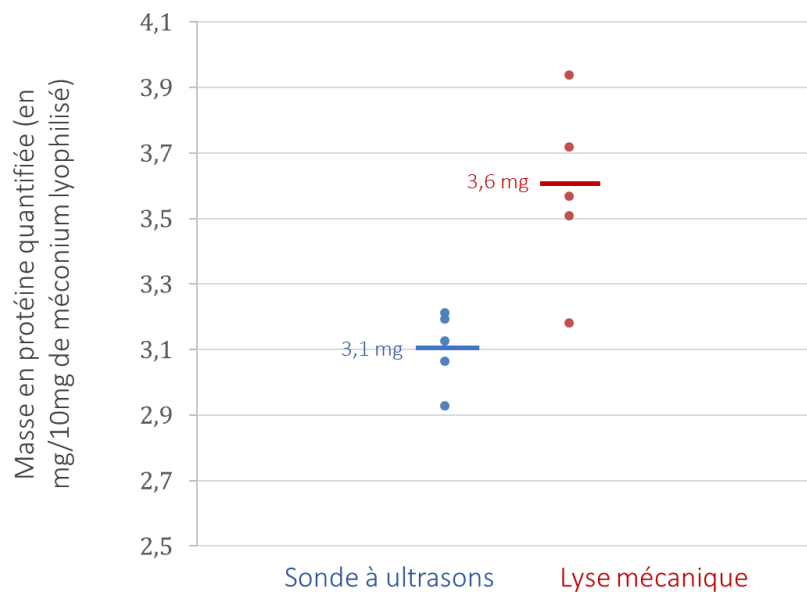


Figure 28 : Résultats de quantification des protéines (mg) réalisée sur 5 réplicats de méconiums homogénéisés à l'aide d'une sonde à ultrasons (bleu) ou d'une lyse mécanique (rouge). Les moyennes des masses en protéines quantifiées sont représentées par des traits horizontaux pour les deux conditions.

b) Analyse qualitative des protéines extraites de la matrice méconium

Nous nous sommes ensuite intéressés à comparer les profils électrophorétiques des protéines extraites selon la technique d'homogénéisation appliquée. Une séparation par SDS-PAGE a été réalisée, à partir des extraits de la première et de la seconde extraction (Figure 29). Les deux extraits obtenus lors de la première extraction par la lyse mécanique ou par la sonde à ultrasons montrent des profils protéiques comparables (cadre bleu). Une bande majoritaire et très intense est observée vers les 50 kDa, ainsi que de très nombreuses bandes protéiques plus ou moins visibles et résolues entre 25 et 160 kDa. Certaines de ces bandes apparaissent cependant plus nettement suite à l'homogénéisation par ultrasons (en exemple, les bandes protéiques à 160 kDa et \sim 30 kDa). Peu de protéines/polypeptides de faibles masses

moléculaires sont observés, suggérant l'absence de dégradation majeure des protéines au cours de l'extraction. A l'inverse, l'analyse électrophorétique des extraits obtenus suite à la seconde extraction, montre une extraction relativement plus importante des protéines/peptides de faibles masses moléculaires (< 5 kDa), notamment au sein de l'homogénéisation par la lyse mécanique, pouvant refléter une dégradation des protéines dans ces conditions (cadre rouge).

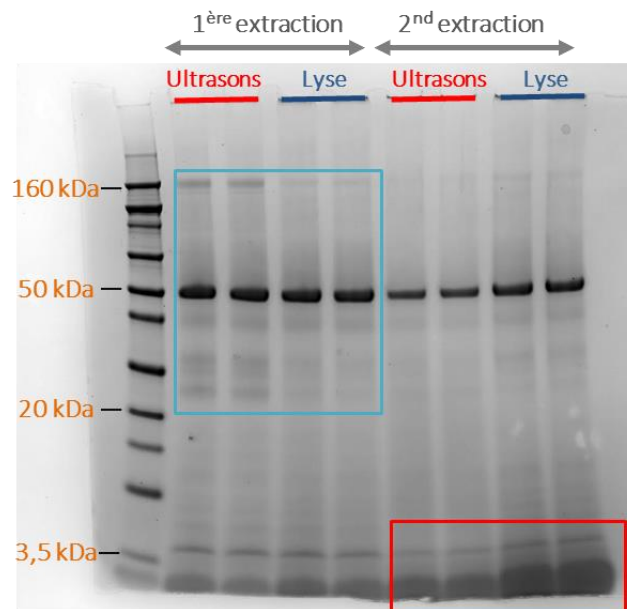
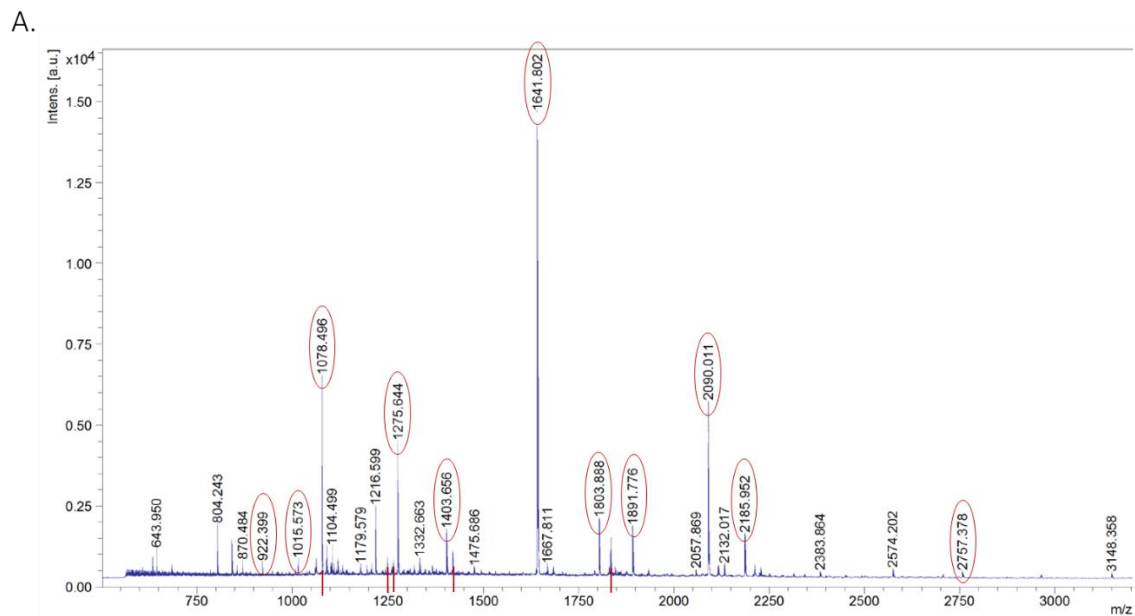
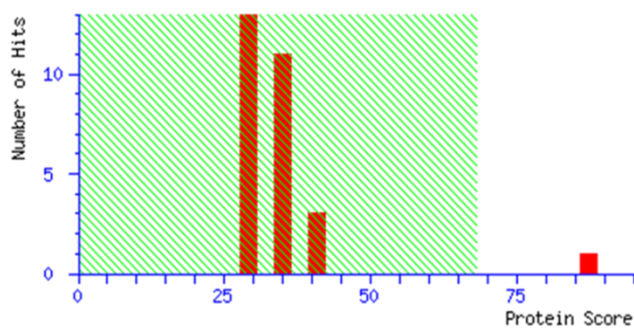


Figure 29 : Séparation des protéines hydrosolubles extraites de la matrice méconium par un gel SDS-PAGE. Le premier dépôt de gauche correspond à notre gamme de référence couvrant les masses allant de 3,5 à 160 kDa. Au sein de chacune des deux extractions, deux réplicats expérimentaux sont déposés et sont notés en rouge pour la sonde à ultrasons et en bleu pour la lyse mécanique. Le rectangle bleu montre la gamme où une concentration des bandes protéiques est observée lors de la première extraction. Le rectangle rouge montre des bandes intenses observées pour les faibles masses (< 3,5 kDa) lors de la seconde extraction.

La bande à 50 kDa a été excisée puis soumise à une digestion enzymatique « in-gel » par la trypsine. Les peptides résultants ont ensuite été analysés par spectrométrie de masse MALDI-ToF, afin de permettre l'identification de la ou des protéines correspondantes. Le spectre de masse des peptides obtenus est représenté en Figure 30.A. La correspondance des séquences détectées a été recherchée grâce au logiciel Mascot en utilisant la base de données de protéines humaines Swissprot. Cette analyse a permis d'identifier la protéine α -1-antitrypsine (A1AT) avec un score significatif de 81 (Figure 30.B). La protéine A1AT identifiée a pour masse théorique : 46737 Dalton, cohérente à la masse observée en SDS-PAGE. Cette protéine a beaucoup été rapportée dans la littérature notamment par Lisowska-Myjak et al, la décrivant comme l'une des protéines majoritaires dans le méconium [138], [276]. Nos résultats sont donc cohérents avec la littérature concernant la prépondérance de cette protéine.



B. α -1-antitrypsine



Base de données	SwissProt
Score	81
Masse expérimentale	46878
Taxonomie	Homo sapiens

Figure 30 : A. Spectre de masse des peptides issus de la digestion trypsique de la protéine à 50 KDa analysée en MALDI-ToF. Les rapports m/z entourés en rouge sont en accord avec les spectres des peptides de l' α -1-antitrypsine dans la base de données. B. Attribution de la protéine à 50 KDa à l' α -1-antitrypsine par le logiciel Mascot, en interrogeant la base de données Swissprot [131].

c) Conclusion

La lyse mécanique associée à l'appareil « Precellys » a permis d'extraire de plus fortes quantités de protéines, avec des profils qualitatifs au sein des réplicats relativement équivalents à ceux observés avec la technique de la sonde à ultrasons. La lyse mécanique présente également une bonne répétabilité au sein des cinq réplicats, avec un coefficient de variation acceptable de ~ 8 %. Bien qu'augmentant le rendement d'extraction, la seconde extraction semble conduire à une dégradation des protéines, ce qui pourrait également affecter les métabolites. De plus, elle induit à une dilution des composés extraits lors du rassemblement des deux surnageants. Wu et al ont notamment démontré que les multiples reprises et pools réalisés lors d'homogénéisations successives, diminuaient les concentrations notamment en métabolites, pouvant impacter significativement leur détection par LC/HRMS [161].

Dans une autre étude, l'homogénéisation de matières fécales a été testée par une sonde à ultrasons et par la lyse mécanique, chacune considérée séparément puis les deux combinées. L'analyse ciblée des métabolites par RMN a montré une extraction métabolique équivalente avec les 3 conditions expérimentales. Les auteurs recommandent cependant l'utilisation de la lyse mécanique étant plus facile d'utilisation et permettant de traiter un grand nombre d'échantillons (indispensable dans l'analyse de cohortes) [269]. Il est à noter que la technique de la lyse mécanique est de plus automatisée limitant l'introduction de biais entre échantillons, et qu'elle utilise des tubes à usage unique, limitant ainsi le risque de contamination entre échantillons. Son coût est par contre plus élevé.

A partir des résultats présentés et des données de la littérature, nous avons décidé de retenir pour la suite de nos analyses : un traitement des échantillons de méconium par une lyse mécanique à l'aide de tubes pré-remplis de billes en céramique en réalisant une seule étape d'extraction (3×30 secondes à 6500 tr/min, 4°C). Ces conditions de lyse sont fournies par le constructeur pour la lyse de matières fécales.

3. Optimisation du volume de reprise pour l'analyse par LC/HRMS

L'extraction des métabolites est réalisée dans un mélange de solvants eau/méthanol (1:4, v/v), nécessitant son évaporation ultérieure et une reprise dans le solvant adapté pour l'analyse LC/HRMS. A la suite de l'homogénéisation par la lyse mécanique, l'extrait métabolique est récupéré par simple centrifugation puis séché sous un flux d'azote. Le volume de reprise du culot de métabolites est à optimiser afin de trouver un compromis limitant l'effet de matrice et l'effet de dilution, qui peuvent directement impacter le nombre de signaux détectés. Comme indiqué dans la partie introductive de ce chapitre, des volumes de reprise très variables ont été utilisés dans la littérature allant de 300 μL et jusqu'à 1 voire 5 mL pour une masse initiale de méconium ou de fèces équivalente à 10 mg [253], [273],

[274]. Nous avons donc testé différents volumes de reprise, puis comparé le nombre de signaux détectés et de métabolites finalement annotés. Les volumes de reprise testés sont de 200, 400, 800, 1600, 3200 et 6400 μL , tous analysés en triplicats. La séquence analytique débute par un blanc, suivi de 3 à 5 réplicats de QC (surchargés avec de l'ATP), permettant de conditionner la colonne chromatographique. Les échantillons biologiques sont ensuite analysés en triplicats du moins concentré au plus concentré. Chaque volume de reprise est séparé du suivant par un échantillon de QC (QC préparé à partir d'un mélange d'échantillon au volume de reprise de 1600 μL , valeur moyenne des volumes de reprises considérés). Les QC dilués ont ensuite découlé de ce mélange (QC 1, QC $\frac{1}{2}$, QC $\frac{1}{4}$, et QC $\frac{1}{8}$) Les valeurs du courant total ionique (TIC) saturant aux volumes de reprise de 200 et de 400 μL , ils ont été exclus de la suite des traitements des données.

a) Nombre de variables détectées selon les conditions de reprise

Afin de comparer les profils métaboliques obtenus dans les conditions HILIC-ESI⁻ au sein des différents volumes de reprises, nous avons réalisé une extraction automatique des signaux. Celle-ci est réalisée à l'aide de l'outil XCMS de la plateforme Workflow4Metabolomics de Galaxy [277]. Comme présenté en introduction, chaque volume de reprise doit être considéré séparément lors du traitement XCMS, permettant de retenir l'ensemble des signaux détectés (présence du paramètre « minfrac » fixé à 70 %). Ainsi, le traitement XCMS a généré entre 14239 et 14863 variables au sein des différents volumes de reprises (Figure 31.A). Chaque variable n'est pas représentative d'un métabolite et peut correspondre à du bruit de fond généré par différentes sources de contaminations (tampons, tubulures...). Ainsi, des étapes de filtration des données ont été réalisées en considérant pour chaque variable :

- i. Le rapport du signal obtenu dans les échantillons biologiques sur celui obtenu dans les blancs analytiques (> 3)
- ii. Le coefficient de variation du QC ($< 30\%$)
- iii. La linéarité de réponse en fonction de la dilution, évaluée à partir du coefficient de corrélation des QC dilués ($> 70\%$)

A la suite de ces filtres, entre 6826 et 8034 variables ont été retenues (en moyenne (- 55 %) des nombres initiaux), avec un nombre de variables retenues décroissant avec la dilution (Figure 31.B).

b) Annotation des variables extraites et filtrées

Nous avons procédé ensuite à l'annotation de ces signaux à l'aide d'une banque de données interne au laboratoire, regroupant 1200 métabolites majoritairement endogènes [41]. Les détails sur les

métabolites identifiés dans la matrice méconium feront l'objet d'une partie indépendante. L'objectif ici d'abord est de comparer la qualité et la quantité des métabolites annotés en fonction des conditions de reprise.

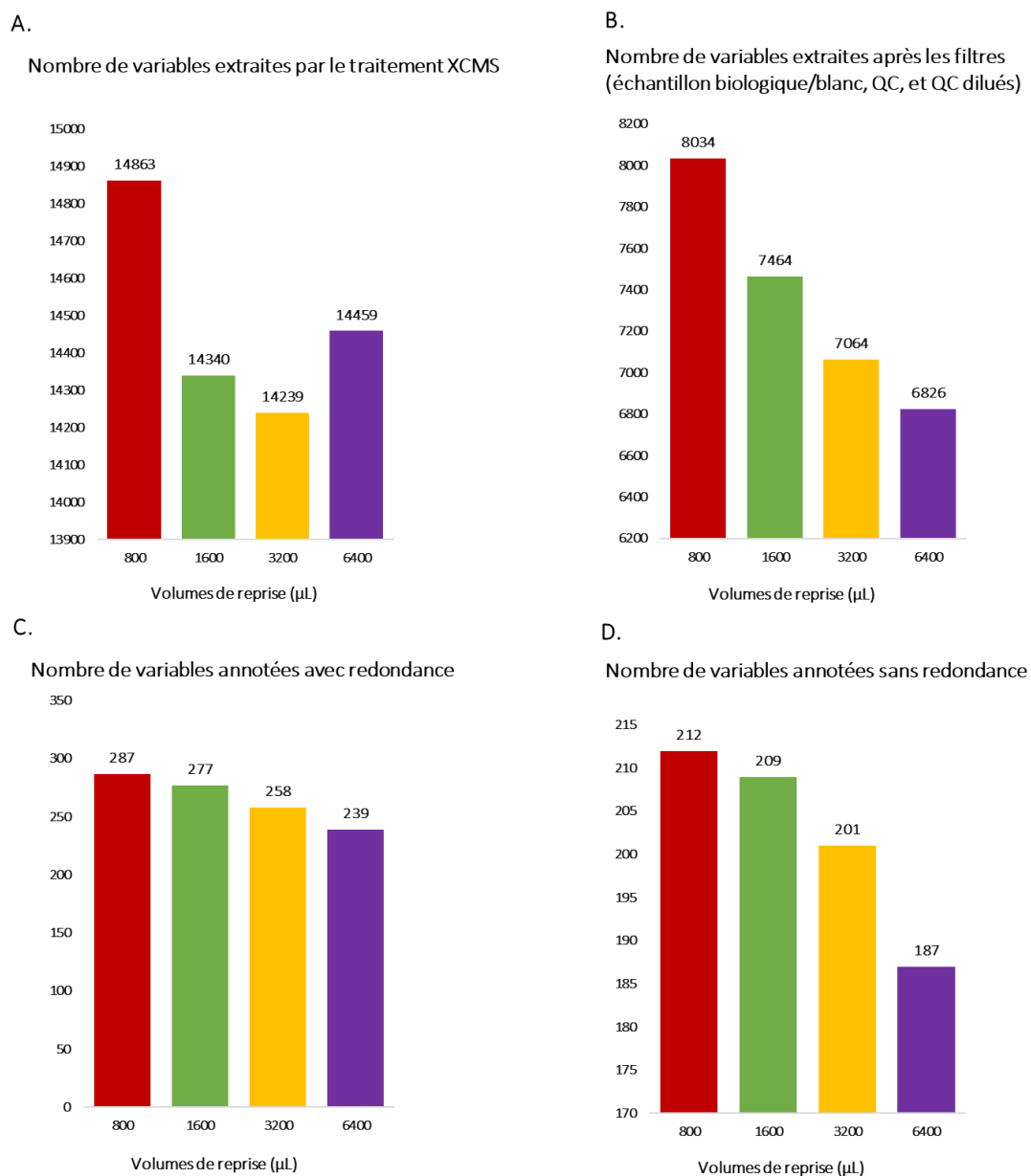


Figure 31 : Evaluation du nombre de variables extraites/filtrées/annotées au sein des solutions résultant des différents volumes de reprise, variant de 800 à 6400 µL. A. Le nombre de variables extraites après le prétraitement XCMS. B. Le nombre de variables retenues à la suite des trois filtres appliqués. C. Le nombre de variables annotées avec redondance (chimique). D. Le nombre de variables annotées sans redondance.

L'attribution des variables a permis d'annoter 238 à 287 métabolites dans les solutions résultant des différents volumes de reprises appliqués (Figure 31.C). Nous observons un nombre de variables

annotées décroissant lorsque le volume de reprise croit. Au sein de ces annotations, il est parfois observé que le même métabolite puisse être annoté plusieurs fois, apportant des redondances (soit par un métabolite résultant en plusieurs pics, soit par la détection d'isomères ou d'adduits... etc). Ainsi, un filtre manuel est nécessaire à cette étape, retenant le métabolite : i) ayant le temps de rétention et le m/z les plus proches des valeurs théoriques de la base de données, et ii) présentant la plus forte intensité au sein des redondances. Ce filtre a permis de retenir entre 187 et 212 métabolites uniques au sein des différentes solutions (Figure 31.D), dont 171 métabolites sont communs aux différents volumes de reprises considérés (Figure 32). Ces chiffres sont comparables à ceux obtenus lors d'analyses réalisées au laboratoire sur d'autres matrices biologiques, dans les mêmes conditions HILIC-ESI⁻ et utilisant la même base de données. En effet, malgré la complexité de la matrice méconium étudiée, nous avons pu annoter un nombre similaire à ce qui a été décrit au même niveau d'identification dans le plasma humain (219 métabolites) [198], ou encore le lait maternel (153 métabolites) [41].

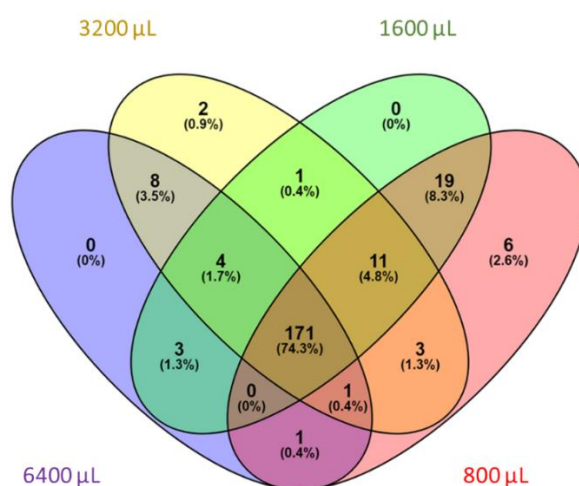


Figure 32 : Diagramme de Venn représentant l'ensemble des annotations des métabolites détectés au sein des différents volumes de reprise allant de 800 à 6400 µL.

Le diagramme de Venn représenté ci-dessus montre qu'un plus grand nombre de métabolites est annoté au sein des solutions aux faibles volumes de reprise : 25 métabolites ne sont retrouvés qu'au sein des volumes de reprise de 800 et de 1600 µL, et 6 composés sont spécifiques au volume de reprise de 800 µL. Ces 6 métabolites correspondent aux acides caproïque et aminobenzoïque, au D-Glutamine et D-glucosamine-phosphate, à l'adénosine et au semialdéhyde succinique. Ces métabolites peuvent être présents à de faibles concentrations dans la matrice méconium, passant sous les LODs (limite de détection) au cours des dilutions (effet de dilution). A l'inverse, d'autres métabolites ne sont détectés qu'au sein des solutions résultant des plus grands volumes de reprise. Ainsi 8 métabolites, dont le cellobiose, la desthiobiotine ou encore la méthionine, sont détectés qu'au sein des volumes de reprise

de 3200 et de 6400 μL . L'absence de ces métabolites dans les extraits les plus concentrés peut s'expliquer par un effet de compétition lors de l'ionisation, impactant la détection de certains métabolites (effet de matrice).

Comme décrit précédemment, le méconium contient des composés aux propriétés physico-chimiques très variées, balayant une large gamme de concentration. Cette caractéristique ne permet souvent pas une détection simultanée de l'ensemble des composants. Ainsi un compromis est souvent à faire, comme le cas est présent ici entre « effet de matrice » et « effet de dilution ».

Dans nos expériences, les volumes de reprise de 800 et de 1600 μL semblent être les plus pertinents avec l'annotation respective de 212 et 209 métabolites. Nous nous concentrons pour la suite sur les 212 métabolites annotés au sein du volume de reprise de 800 μL , générant le plus grand nombre de signaux extraits et de métabolites finalement annotés.

4. Etude de la stabilité des analyses de méconium par LC/HRMS pour l'analyse de grandes séries d'échantillons

Afin de valider les étapes du protocole de préparation de l'échantillon dont le volume de reprise de 800 μL , il est important d'étudier la stabilité du système analytique lors d'une séquence équivalente à celle de la cohorte d'intérêt. Au début de mes travaux de thèse, nous ne pensions analyser que les méconiums excrétés au cours du jour 1 au sein de la cohorte EDEN, afin de refléter au mieux les expositions *in utero* indépendamment des expositions néonatales. Ceci correspondait à environ 80 échantillons de méconium. Nous avons ainsi réalisé une étude de stabilité et de répétabilité des analyses LC/HRMS sur une séquence de plus de 120 injections de méconium. Celle-ci inclue 96 injections d'échantillons de méconium, 10 injections de QC (1 injection tous les 10 échantillons), 12 échantillons de QC dilués (4 dilutions, chacune analysée en triplicat) et 3 blancs analytiques (Figure 33).

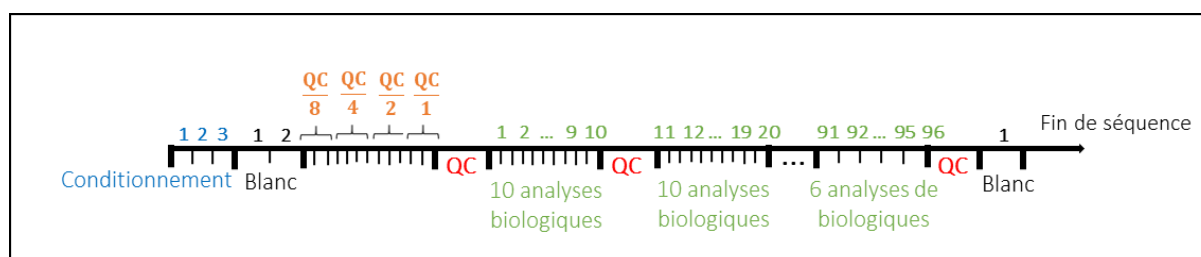


Figure 33 : Séquence analytique réalisée pour le suivi de la répétabilité du processus analytique dans les conditions HILIC-ESI'.

Pour cette étude, quatre répliqués expérimentaux de l'échantillon de méconium B collecté au jour 1 (équivalent à la cohorte EDEN) ont été réalisés en parallèle (lyophilisation, homogénéisation...). Ceci permet également d'évaluer la bonne reproductibilité de l'étape de préparation de l'échantillon. Chaque culot sec de métabolites est repris dans 800 μL , et l'échantillon de QC est préparé à partir de 150 μL prélevés de chacun des 4 répliqués (les dilutions de QC découlent de ce mélange). Les analyses des échantillons biologiques sont réalisées aléatoirement dans la séquence. En amont des analyses des échantillons biologiques, la séquence débute par 3 analyses de QC avec de l'ATP permettant le conditionnement de la colonne HILIC ()

Le suivi de la stabilité du système analytique au cours de la séquence a été réalisé sur : i) les intensités des standards internes ajoutés dans chacun des tubes d'injections des QC et des échantillons biologiques, ii) la somme des intensités de toutes les variables détectées dans les QC.

a) Suivi des intensités des étalons internes le long de la séquence d'analyse

Afin de pouvoir suivre et de vérifier la répétabilité du processus d'acquisition des données, les signaux correspondant aux étalons internes sont suivis tout au long de la séquence analytique. Les étalons internes utilisés au laboratoire sont soigneusement choisis de façon à couvrir l'ensemble des gammes de masses (m/z) et des temps de rétention des métabolites. Nous avons suivi dans cette étude 4 étalons internes balayant une large gamme de m/z ainsi qu'une large fenêtre des temps de rétentions. Ces composés incluent l'acide éthylmalonique (m/z 131,03498 et RT = 7,65 min), l'amiloride (m/z 228,04061 et RT = 6,7 min), le dinoseb (m/z 239,06734 et RT = 1,7 min), et la prednisone (m/z 393,14742 et RT = 1,95 min) (Annexe 3).

La Figure 34.A , montre les intensités des signaux détectés pour chacun de ces composés au sein de l'ensemble des échantillons biologiques et des QC analysés (représentés en fonction de leur ordre d'injection). Les intensités d'un même étalon interne au sein des différentes injections montrent un profil stable au cours de la séquence. Afin d'apprécier au mieux cette stabilité, les signaux obtenus sur l'ensemble de la séquence sont représentés en Figure 34.B sous forme de boîtes à moustaches et les coefficients de variation (CV) ont été calculés. Les résultats de cette figure montrent des valeurs de CV allant de 9 à 17 % pour l'acide éthylmalonique et la prednisone respectivement. Cette variabilité est tout à fait acceptable (< 20 %) pour une séquence de 120 injections d'extraits de méconium [278].

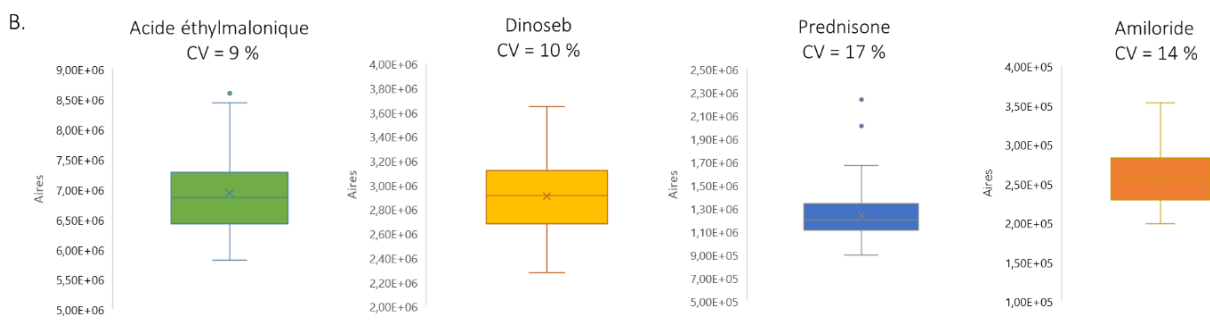
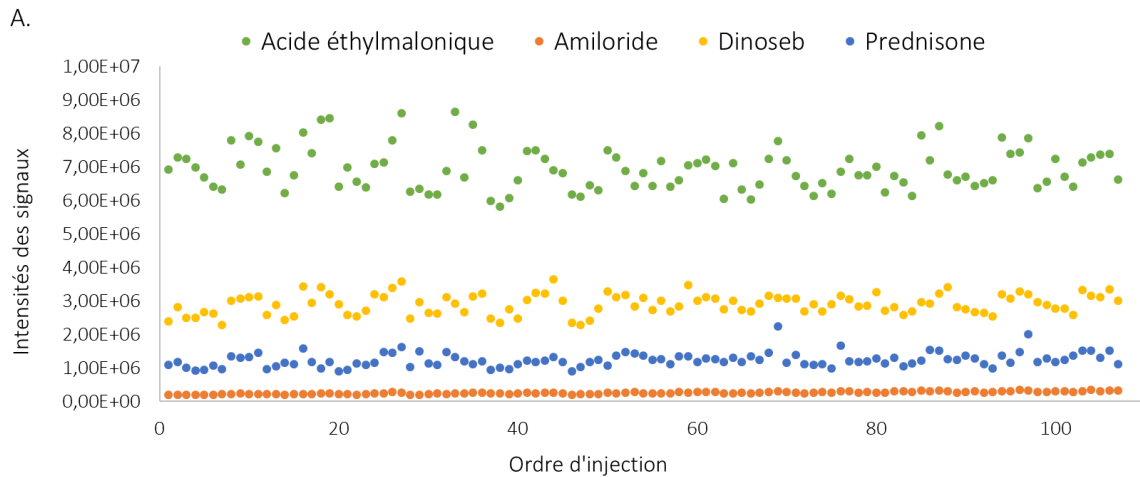


Figure 34 : Evaluation de la répétabilité des analyses au cours de 120 injections d'extraits de méconium

A. Suivi des intensités des étalons internes (acide éthylmalonique, amiloride, prednisone et dinoseb) dans les échantillons biologiques et des QC. B. Evaluation de la répétabilité des analyses représentée par des boîtes à moustaches. Ces boîtes à moustaches nous informent sur : 1. La valeur de la médiane pour chacun des composés de références représentée par le trait au milieu de chaque boîte, 2. La valeur de la moyenne représentée par une croix dans la boîte. 3. La valeur minimale de la distribution représentée par l'extrémité basse du quartile. 4. La valeur maximale de la distribution représentée par l'extrémité haute du quartile. 5. La valeur du premier quartile (25 %) représentée par la valeur en bas de chaque boîte, ainsi que 6. La valeur du troisième quartile (75 %) représentée par la valeur en haut de chaque boîte. Uniquement les points en dehors de la distribution sont représentés (cas de l'acide éthylmalonique et de la prednisone, représentés en bleu et en vert respectivement).

b) Suivi de la somme des intensités des variables des QC au cours de la séquence analytique

Parallèlement, la somme des intensités de toutes les variables au sein des QC et de l'ensemble des échantillons a été tracée en fonction de l'ordre d'injection (Figure 35). Cette figure est obtenue à partir de l'outil « QualityMetrics » de W4M. Elle permet de suivre visuellement l'ensemble des intensités des variables et de leur variation le long de la séquence d'analyse. Lors de notre expérience, il y a eu une erreur de programmation des échantillons : les deux premières injections de QC ont été prélevées dans le tube du QC dilué au 1/4 (entourés en rouge sur la Figure 35). Celles-ci permettent cependant de

constater une bonne stabilité des analyses lorsque ces deux injections sont comparées aux échantillons correspondants, injectés en début de séquence (QC/4, représentés par des triangles oranges). Ces deux injections n'ont donc pas été considérées lors des étapes de traitement des données.

Une première observation permet de rendre compte que les intensités des signaux mesurés dans les QC présentent une valeur moyenne comparées à celles mesurées dans les échantillons biologiques. Ce constat est cohérent étant donné que ces QC sont préparés à partir d'un mélange équivalent de l'ensemble de ces échantillons biologiques. L'analyse de la somme des intensités des signaux obtenus au sein des QC injectés régulièrement le long de la séquence, a permis de rendre compte d'un système stable. En effet, le coefficient de variation de la somme des intensités mesurées pour les 8 échantillons de QC exploitables est de 4 %.

Les intensités des signaux des QC dilués montrent des valeurs proportionnellement décroissantes, suivant l'ordre de dilution (deux en deux) avec une bonne reproductibilité observée pour les triplicats au sein de chaque dilution. Ces résultats permettent également d'indiquer que les étapes de filtrage des variables ont permis de ne retenir et de ne considérer que les variables biologiquement pertinentes. Enfin l'intensité du blanc est représentée à une basse échelle indiquant sa pureté. Les réplicats biologiques sont analysés aléatoirement en injectant 2 fois de suite le même échantillon, jusqu'à atteindre 24 injections pour chaque réplicat. Les coefficients de variations calculés intra-réplicats biologiques représentent de faibles valeurs de variation entre 3 et 5 %. Le coefficient de variation inter-réplicats (pour l'ensemble des analyses d'échantillons biologiques = 96) est faible également, égale à 4 %. Au regard de l'ensemble de ces résultats, nous pouvons conclure à la stabilité du système analytique après 120 injections de méconium, permettant de valider le protocole expérimental utilisé.

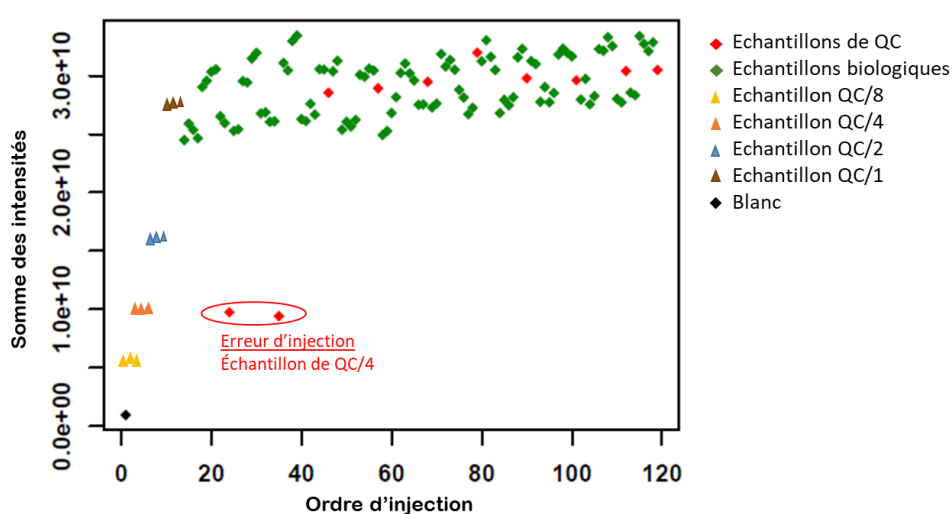


Figure 35 : Etude de la stabilité et de la reproductibilité du processus analytique dans les conditions HILIC-ESI. « QualityMetrics » : tracé des sommes des intensités des variables au sein de tous les échantillons en fonction de l'ordre d'injection.

5. Bilan des optimisations et des développements réalisés sur le protocole de préparation de l'échantillon, vers une première description du méconium

Après ces étapes d'optimisations, le protocole final retenu est le suivant : une lyophilisation des échantillons de méconiums bruts pendant 17 heures suivie d'une étape de reprise et d'homogénéisation dans de l'eau/méthanol assistée par une lyse mécanique. L'extrait métabolique est récupéré par centrifugation et est séché sous un flux d'azote durant 2 heures. Le culot sec est ensuite repris dans un volume de 800 µL de phases mobiles adéquates pour l'analyse par LC/HRMS (Figure 36). Quelques analyses complémentaires seront réalisées pour vérifier la compatibilité de ce protocole avec l'analyse sur une colonne chromatographique RP-C18 et une détection des ions dans le mode positif.

De façon exploratoire et complémentaire, nous avons évalué les performances d'un autre système analytique le FT-ICR (de l'anglais, Fourier Transform-Ion Cyclotron Resonance), et l'avons comparé aux résultats obtenus par LC/HRMS. Nous présentons les données obtenues dans ce qui suit.

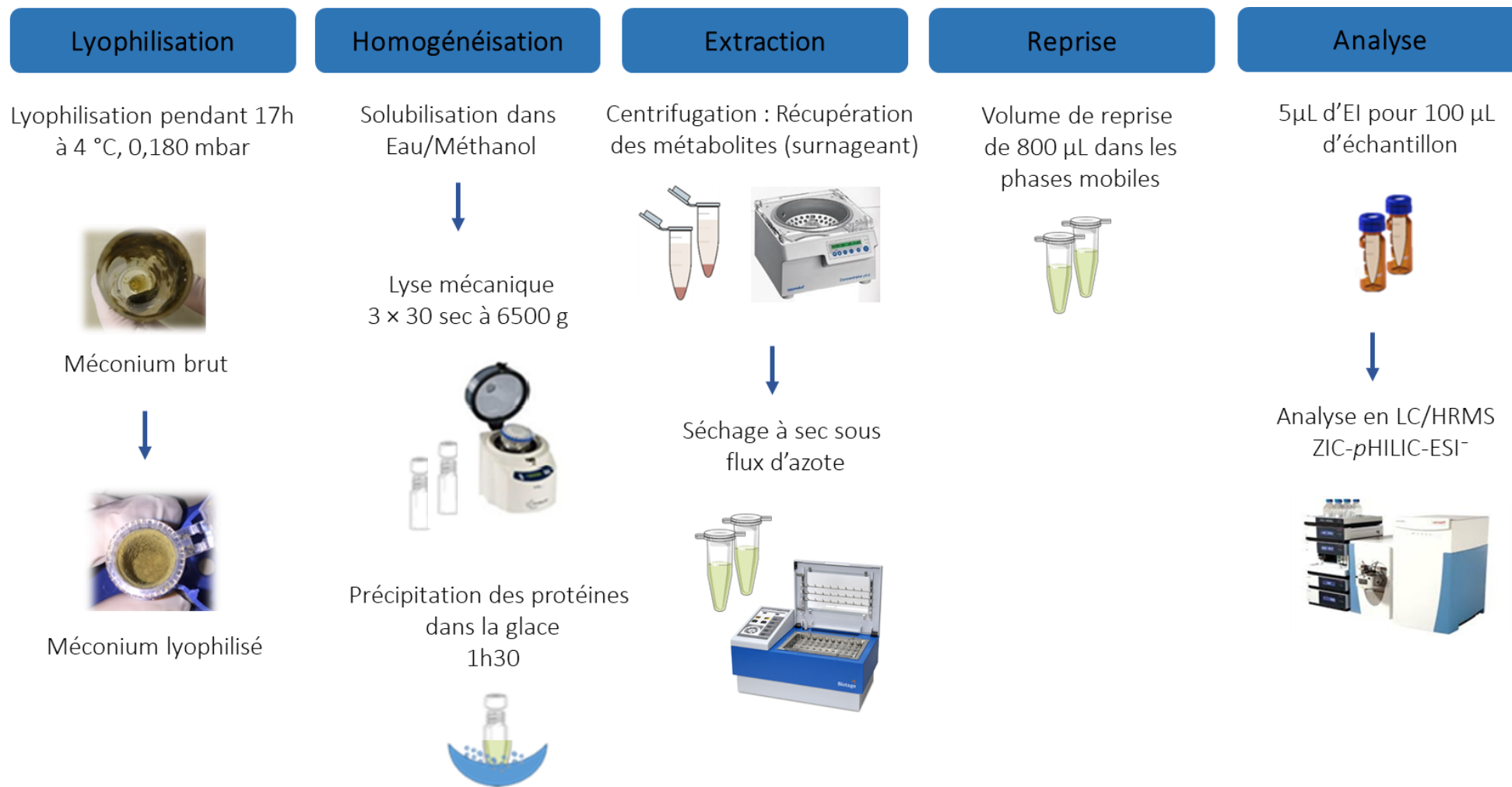


Figure 36 : Protocole de préparation de l'échantillon de méconium final.

6. Analyses métabolomiques réalisées par DIMS (Direct Introduction Mass Spectrometry) sur un FT-ICR

En parallèle des développements méthodologiques réalisés en LC/HRMS au SPI, nous avons exploré la possibilité de réaliser des analyses à haut débit par introduction directe des échantillons sur un instrument de type FT-ICR du laboratoire spectrométrie de masse CSOB (Institut Parisien de Chimie Moléculaire (IPCM), Sorbonne Université), en collaboration avec Estelle Rathahao-Paris, membre du LIAA).

Comme présenté dans la partie introductive, l'avantage de la technique DIMS (de l'anglais, Direct Introduction Mass Spectrometry) est de réduire considérablement les temps d'analyses. Ceci étant un facteur important dans l'étude de cohortes, qui est l'objectif final de ce projet de thèse. L'introduction directe sur ce type d'appareil est facile à mettre en place et permet de réaliser des analyses rapides (< 5 minutes), par rapport à l'analyse LC/MS réalisée avec une colonne ZIC-*p*HILIC pour laquelle l'analyse dure 42 minutes. De plus, l'instrument FT-ICR permet d'atteindre un pouvoir de résolution en masse très élevé (> 500 000 voire 1 million à un m/z de 200), ainsi que de fournir des mesures de masses très précises (<1 ppm). La principale limite de la technique DIMS réside dans les phénomènes d'effets de matrice qui résultent de la compétition d'ionisation entre différents composés arrivant simultanément au niveau de la source d'ionisation. En couplage LC/HRMS, les composés sont fractionnés selon leur affinités vis-à-vis de la colonne en amont de l'analyseur en masse, ce qui minimise ces effets de matrice.

La réalisation d'une analyse métabolomique par l'approche DIMS sur les mêmes échantillons de méconium que ceux précédemment analysés en LC/MS permettra d'évaluer les spécificités et complémentarités potentielles de ces deux méthodes analytiques *via* la comparaison du nombre de variables détectées et de métabolites annotés.

a) Optimisation des conditions analytiques

Une étude préliminaire est nécessaire pour définir le meilleur facteur de dilution des échantillons de méconium. Pour cela, nous avons dans un premier temps voulu déterminer la concentration des composés de référence à surcharger dans l'extrait de méconium, afin d'évaluer les effets de matrice. Comme indiqué dans le paragraphe précédent, l'analyse par introduction directe présente d'importantes limitations au niveau des phénomènes de suppression d'ions. Parmi les 10 composés standards répondant dans les conditions d'ionisation négative en LC/HRMS (Annexe 5). Nous avons choisi de tester 5 d'entre eux, couvrant une gamme de m/z suffisante. Ces composés incluent : l'acide éthylmalonique (m/z 131,03498), l'amiloride (m/z 228,04061), la prednisone (m/z 393,14742), l'ATP (adénosine triphosphate, (m/z 505,9879)) et la dihydrostreptomycine (m/z 618,25072). Tous ces

composés ont été introduits par injection en flux continu (FIA) et analysés en mode d'ionisation négatif à différentes concentrations (de 0,01 à 1 ng/μL). L'objectif ici est de travailler aux concentrations les plus faibles afin de ne pas surcharger la matrice méconium par la suite. Le solvant utilisé pour les dilutions est composé d'H₂O/MeOH (50:50, v/v). Parmi les composés analysés, seul l'acide éthylmalonique a été détecté de façon répétable et à une faible concentration (C = 0,04 ng/μL). Le dihydrostreptomycine n'a pas été détecté (jusqu'à 1 ng/μL). Les trois autres composés sont détectés de façon non répétable, nécessitant de plus forte concentrations. Nous avons ainsi décidé de ne pas les inclure afin de ne pas surcharger notre matrice.

A cette étape les culots métaboliques secs (identiques aux culots utilisés en LC/HRMS, décrits dans le Chapitre I, Partie II.C .2.a) sont repris à différentes dilutions représentées en Figure 37. Ces solutions ont été analysées de la moins concentrée à la plus concentrée (du fd5000 au fd50) dans la séquence analytique. Chacune de ces solutions contient 0,04 ng/μL du composé standard de référence décrit précédemment. Dans la séquence d'analyse, des injections de deux à quatre blancs analytiques sont réalisées entre chaque échantillon biologique afin d'éviter des effets de mémoire et de mieux évaluer le signal obtenu pour chaque facteur de dilution.

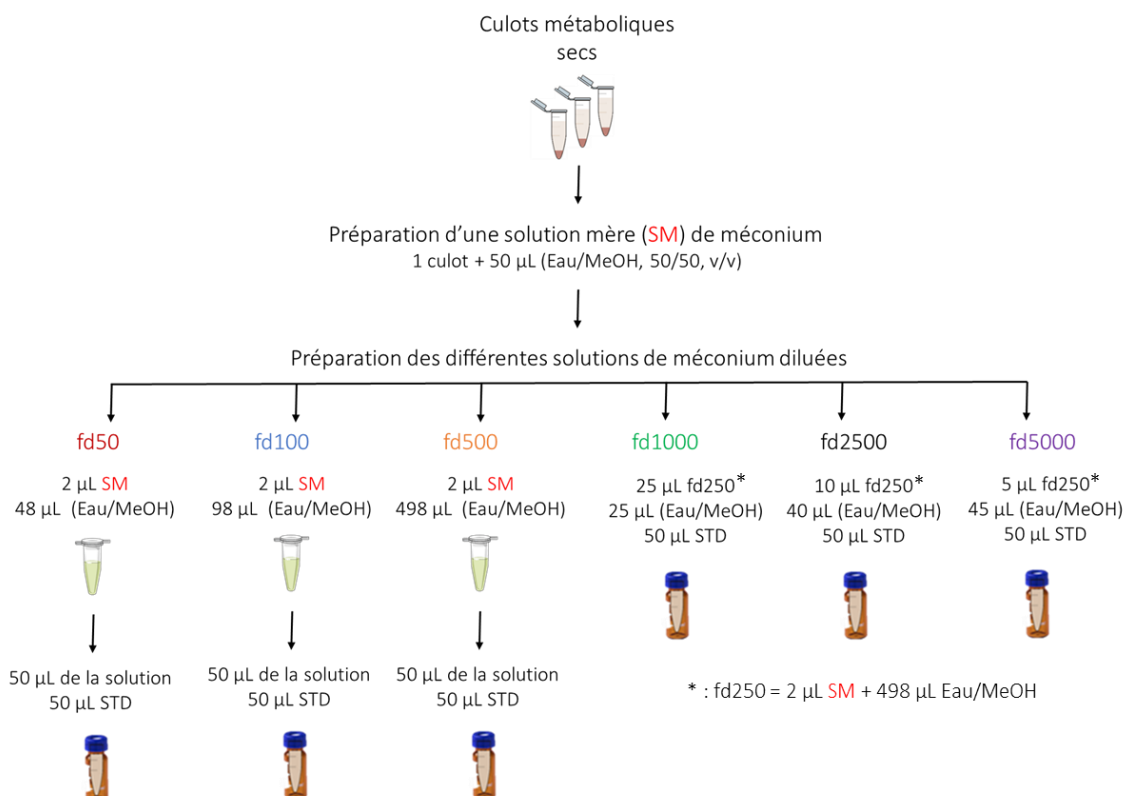
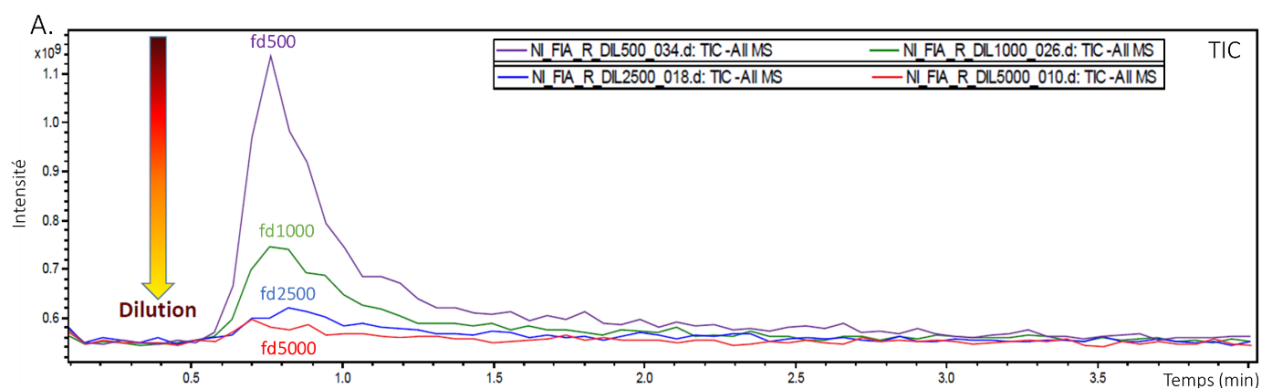


Figure 37 : Protocole de préparation des échantillons de méconium pour l'analyse en DIMS dans le mode d'ionisation négatif. STD : solution du standard de l'acide éthylmalonique à une concentration de 0,08 ng/μL. Proportions eau/MeOH (50:50, v/v).

Nous avons observé des effets de matrice importants pour les échantillons les plus concentrés (fd50 et fd100, où nous n'avons pas de retour du pic d'injection à la base). Ces facteurs de dilutions ne sont donc pas retenus pour la suite des analyses. Les profils du courant ionique total (TIC) pour les quatre autres solutions d'extraits métaboliques (de fd500 à fd5000) sont représentés en Figure 38.A. Nous observons un étalement du TIC pour la solution au fd500, pour laquelle le retour à la ligne de base se fait plus tardivement comparée aux autres dilutions. A l'inverse, la solution préparée à un facteur de dilution de 5000 ne génère qu'une très faible intensité du TIC, suggérant un effet de dilution qui se traduit par un nombre important de composés passant sous leur seuil de détection. Afin de choisir le facteur de dilution à utiliser et de trouver un compromis entre les effets de matrice et l'effet de dilution, nous avons analysé les spectres de masse obtenus pour les solutions aux fd500, fd1000 et fd2500 ainsi que le profil d'un blanc analytique, représentés en Figure 38.B.

Nous observons en premier lieu, le pic du standard de référence (étoile rouge), détecté dans la matrice méconium aux différents facteurs de dilutions. La comparaison entre les spectres de masse aux différents facteurs de dilutions, indique l'apparition de pics spécifiques des constituants du méconium dans la zone encadrée en rouge, correspondant à une gamme de m/z entre 350 et 800. Plus l'échantillon est concentré, plus le nombre et l'intensité des pics situés dans cette zone sont importants. Peu de signaux sont détectables dans cette zone pour la solution la plus diluée, soit fd2500. Dans ces conditions, le meilleur compromis semble être l'utilisation d'un facteur de dilution de 1000.



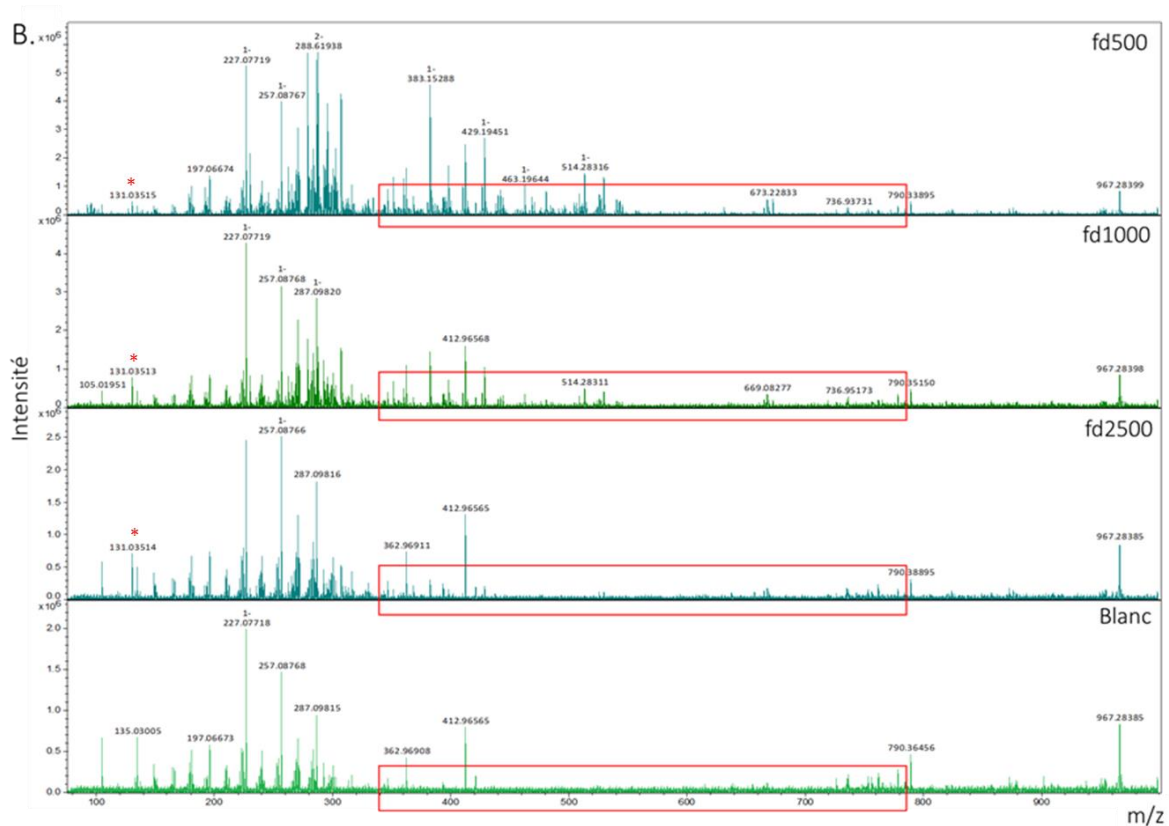


Figure 38 : A. Profils du courant total ionique (TIC) obtenus pour les différentes solutions considérées pour le méconium. B. Spectres de masse correspondants aux solutions de méconium aux fd500 à fd2500 ainsi que le spectre de masse du blanc analytique. Les rectangles rouges indiquent la région variant le plus entre les différents facteurs de dilution. Les étoiles rouges indiquent la détection du composé standard de référence dans la matrice méconium au sein des différentes solutions.

b) Evaluation de la répétabilité des analyses par l'approche DIMS

Afin de valider le facteur de dilution choisi pour l'analyse des échantillons de méconium, nous avons réalisé des répliquats analytiques sur un même extrait de méconium, permettant de rendre compte de la répétabilité des analyses. Pour cela, une quinzaine d'analyses d'un extrait de méconium dilué au 1:1000 ont été réalisées. Nous avons précédemment observé que 3 blancs analytiques entre chaque injection étaient suffisants (obtention de signaux de blancs très répétables). Ainsi chaque injection d'extrait de méconium est précédée par 3 injections de blancs, permettant d'éviter des phénomènes d'effet mémoire. Nous avons ensuite évalué la répétabilité des analyses au cours de la séquence d'injections par le suivi : i) des profils des TIC au sein des différentes injections, ii) de la répétabilité du signal d'un composé standard de référence au cours des injections, et iii) des spectres de masse obtenus pour ces différentes injections. Nous observons sur la Figure 39.A des profils de TIC répétables au cours des injections, avec un coefficient de variation < 20 %, de plus il y a un bon retour du pic à la ligne de

base, et ce le long des différentes analyses réalisées. L'analyse de la variation de l'intensité du composé de référence présent en quantités équivalentes au sein des différentes injections a également été réalisée. Ceci a permis de rendre compte d'une bonne stabilité et répétabilité du standard de référence avec un coefficient de variation de 12 % (Figure 39.B). Nous représentons sur la Figure 39.C, les spectres de masses obtenus pour trois injections réalisées en début, en milieu et en fin de séquence. Cette figure démontre une empreinte spectrale répétable au cours des différentes injections, reflétant une bonne stabilité des analyses le long de la séquence.

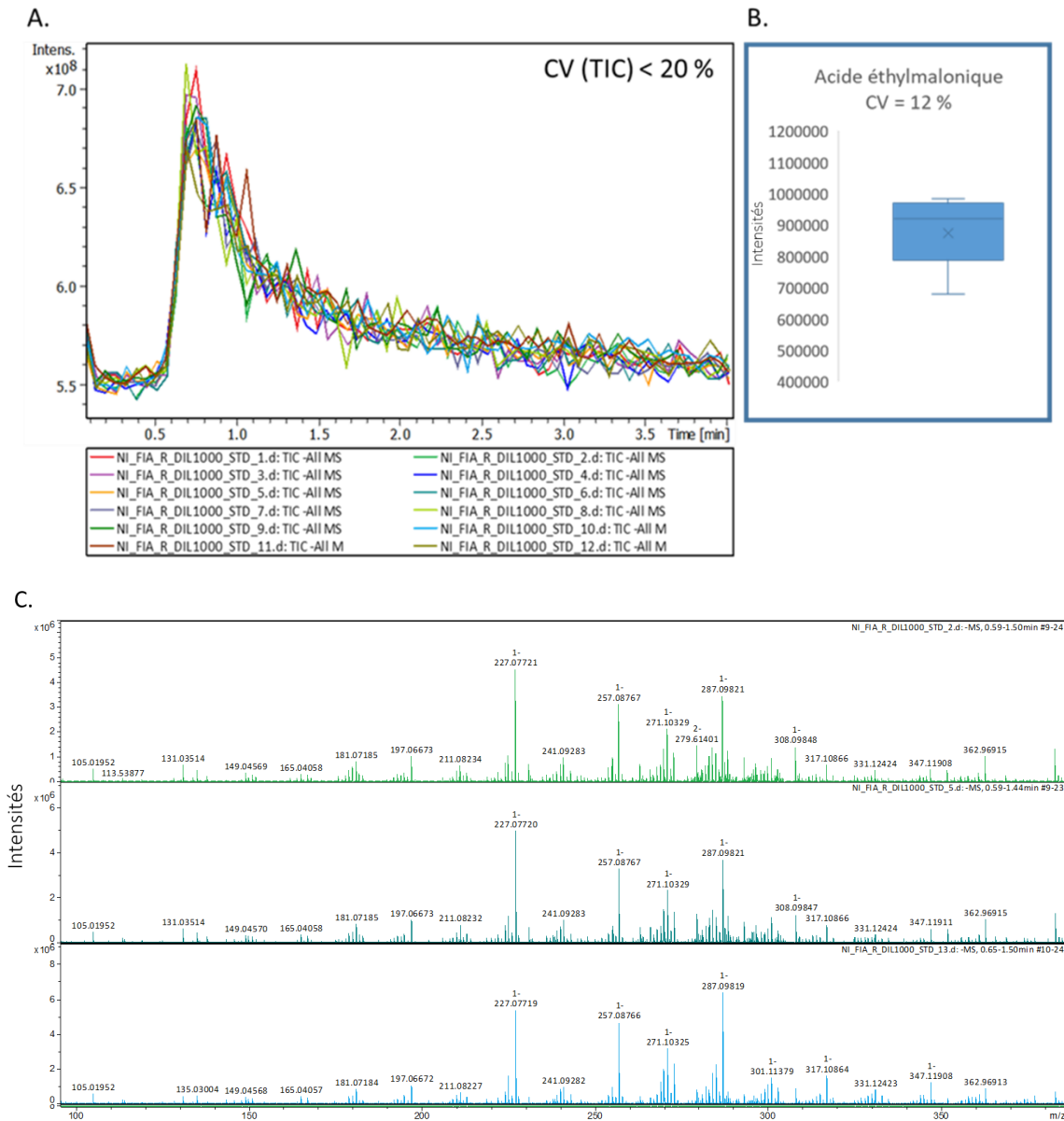


Figure 39 : Evaluation de la répétabilité des analyses au cours de la séquence d'injections par : A. Les profils des courants ioniques totaux (TIC) obtenus au sein des différents réplicats d'injections d'un extrait de méconiums au fd1000. B. La représentation de la variation du signal de réponse du composé standard de référence utilisé (acide éthylmalonique) par une boite à moustache. C. Les spectres de masse obtenus pour 3 analyses répétées, considérées dans l'ordre d'injection (en début, en milieu et en fin de séquence analytique).

Nous avons ensuite extrait les signaux afin d'analyser le nombre de variables détectées, et de les annoter. L'extraction manuelle est réalisée sur « Data analysis » générant un fichier en format « .xlsx » et permettant d'extraire 451 variables pour un signal sur bruit supérieur à 2. A titre de comparaison, Habchi et al. ont extrait 440 variables au sein d'échantillons d'urines analysées dans ces mêmes conditions et sur ce même instrument [204]. En LC/HRMS, le nombre de variables extraites était d'environ 8000 variables retenues après les filtres, ceci réalisé à partir du même échantillon de méconium (Figure 31, Partie III.3).

c) Annotation des variables extraites

L'annotation des variables extraites a été réalisée manuellement à partir de la même base de données que celle utilisée en LC/HRMS, afin d'être comparable aux analyses précédentes. L'attribution des métabolites est obtenue en comparant les masses expérimentales précises aux masses théoriques enregistrées. La méthode du DIMS a permis d'annoter cinq métabolites avec une précision $\leq 1\text{ppm}$, incluant le N-acetyl-glucosamine, l'adénine, la méthylxanthine, l'arabinose et l'arabitol. Ceux-ci sont communs à l'annotation obtenue en LC/HRMS.

En termes de nombre de variables extraites et de métabolites finalement annotés, le DIMS n'apporte pas ou peu d'informations complémentaires par rapport aux analyses réalisées par LC/HRMS. Celui-ci permet en revanche de gagner un temps d'analyse considérable. Dans leur étude, Lin et al ont comparé les deux techniques de LC/HRMS et de DIMS au sein d'une analyse métabolomique de sérums de patients atteints de cancer et des sujets sains. Leurs analyses ont permis de détecter 23 biomarqueurs potentiels de la maladie par la technique impliquant l'introduction directe contre 48 pour l'analyse par LC/HRMS. Les analyses par la méthode DIMS ont certes été réalisées en seulement 5 % du temps nécessaire pour les analyses LC/HRMS, mais moins de 50 % des composés ont été détectés par cette méthode. Les auteurs ont recommandé la méthode du DIMS dans le cas d'un diagnostic rapide, tandis que la LC/HRMS est nécessaire lorsqu'un dépistage complet des biomarqueurs est requis [279]. Ainsi, selon l'objectif de chaque étude, l'une ou l'autre des techniques peut se montrer plus pertinente.

d) Conclusion

Cette étude préliminaire a permis de déterminer le facteur de dilution optimal pour l'analyse du méconium par l'approche DIMS, permettant de minimiser les effets de matrice et de dilution. Au sein des spectres obtenus pour ce facteur de dilution, 450 variables ont été extraites. Parmi ces variables, 5 métabolites ont pu être annotés avec une erreur $\leq 1\text{ppm}$ en utilisant la même base de données que celle utilisée en LC/MS [41]. A partir des données LC/HRMS (analyses dans les conditions HILIC-ESI-) de ces mêmes échantillons de méconium, nous avons pu annoter 212 métabolites à partir de la même

base de donnée. Dans le cadre de notre projet visant à établir une cartographie métabolique globale du méconium, il a semblé plus pertinent d'utiliser le couplage LC/HRMS afin d'obtenir la couverture la plus large possible. Néanmoins, il pourrait être intéressant d'évaluer l'apport et le potentiel de stratification d'une analyse par DIMS dans le cas de larges cohortes d'échantillons comme EDEN.

B. Cartographie métabolique du méconium, obtenue dans les conditions HILIC-ESI⁻

A partir du développement méthodologique retenu dans les conditions analytiques HILIC-ESI⁻ présenté précédemment, 212 métabolites ont été putativement annotés. Nous avons par la suite cherché à affiner ces annotations en vérifiant manuellement et successivement la qualité et l'intensité des signaux à l'aide du module Qual-Browser de Xcalibur (logiciel Thermo Fisher Scientific). Trois paramètres ont été suivis : i) l'allure gaussienne des pics chromatographiques, qui doivent présenter une aire supérieure à 10⁴, ii) la présence de l'isotope ¹³C (différence de 1,0033 Da par rapport au ¹²C), ainsi que iii) la détection du signal dans au moins 5 spectres consécutifs. A cette étape, 183 métabolites sur les 212 annotés ont été retenus, respectant les 3 critères cités. L'exemple de l'acide muramique est présenté, montrant le chromatogramme de l'ion extrait (Figure 40.A, RT = 7,68 min), ainsi que le spectre correspondant (Figure 40.B), avec l'ion moléculaire à *m/z* 250,0936 et l'isotope ¹³C à *m/z* 251,0970, soit une différence de 1,0034 Da.

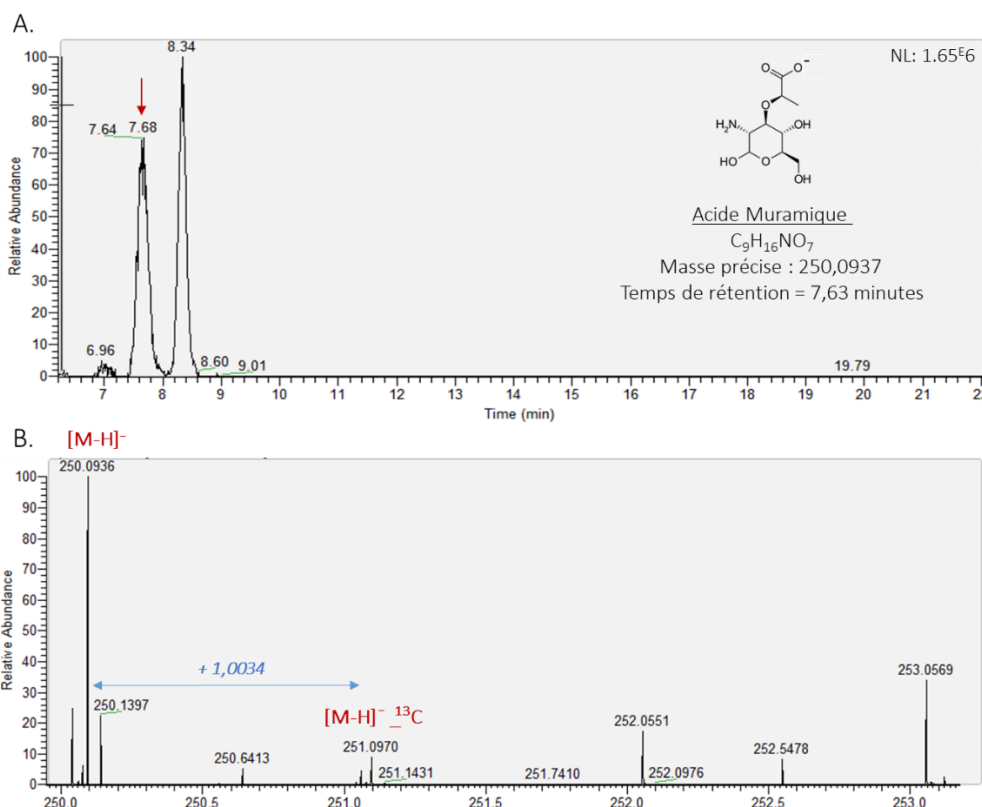


Figure 40 : Chromatogramme de l'ion extrait de l'acide muramique annoté dans le méconium en (A) et le spectre de masse correspondant en (B). NL indique l'intensité normalisée du pic de base, (de l'anglais Normalization Level).

Afin d'aller plus loin dans l'identification des métabolites du méconium, nous avons réalisé des expériences de fragmentation MS/MS sur les échantillons de QC, ceci sur l'ensemble des 183 métabolites retenus à la suite des vérifications sur le logiciel Xcalibur. Les spectres de fragmentation obtenus sont comparés aux spectres des composés de référence analysés dans les mêmes conditions analytiques. A la suite de ces expériences, nous avons pu confirmer l'identité par MS/MS de 132 métabolites. Ces métabolites sont identifiés sur la base de trois paramètres physico-chimiques orthogonaux, le temps de rétention, la mesure de masse précise et les spectres MS/MS, et ce par homologie avec des composés de référence issus de la chimiothèque et analysés dans les mêmes conditions expérimentales. Ces métabolites sont donc validés au plus haut niveau de confiance (niveau 1 selon Sumner et al [210]). L'exemple de l'acide N-acétylaspartique validé est donné en Figure 41. Nous observons la présence notamment de l'ion parent à un m/z 174,041, ainsi que des 7 ions fils les plus intenses. En plus de ces 132 métabolites, cinq autres métabolites n'ont été identifiés qu'à partir du temps de rétention et de la masse précise, sans que les spectres MS/MS obtenus permettent de les valider. Il s'agit des acides caprique, nervonique et tridécanoïque ainsi que de l'uracil et de l'hexose-phosphate.

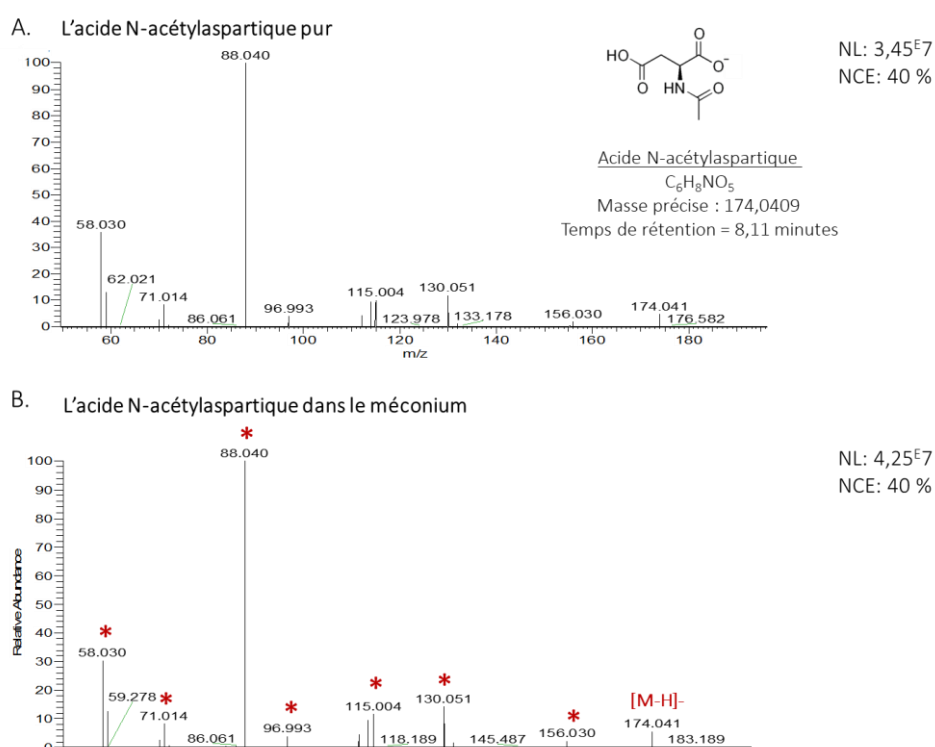


Figure 41 : Spectres MS/MS de l'acide N-acétylaspartique pur (A) et dans l'extrait de méconium (B). NL indique l'intensité normalisée du pic de base (de l'anglais : Normalization Level) et NCE indique la valeur de l'énergie de collision normalisée (de l'anglais : Normalized Collision Energy).

Les 137 métabolites identifiés dans le méconium sont regroupés en Annexe 5 et seront analysés plus en profondeur à la fin de ce chapitre. En termes de nombre de métabolites annotés, nous obtenons des résultats comparables avec ce qui a précédemment été réalisé au laboratoire sur d'autres matrices biologiques. Par exemple, Boudah et al ont identifié 130 métabolites avec le même niveau de confiance dans du sérum humain dans les mêmes conditions HILIC-ESI⁻ [236]. Dans une autre étude sur des laits maternels, Berdi et al ont identifié 125 métabolites, en combinant les conditions HILIC-ESI⁻ et C18-ESI⁺/ESI⁻ [333].

Au total, 8000 variables ont été extraites et filtrées conduisant à 137 métabolites annotés avec un niveau élevé de confiance, à partir de deux échantillons de méconium. Soit 2 % « seulement » de l'ensemble des données filtrées. Il faut néanmoins noter qu'il existe une grande redondance d'information au sein des 8000 variables retenues (exemple : redondances chimiques). Il pourrait être intéressant de réaliser des expériences MS/MS complémentaires afin de confirmer/infirmer certaines annotations obtenues par interrogation des bases de données externes, ce qui permettrait d'élargir significativement la couverture métabolique du méconium.

Même si les composés les plus apolaires sont probablement sous représentés du fait de nos conditions d'extraction et d'analyse, les composés annotés présentent des propriétés physico-chimiques très variées, permettant de rendre compte d'un protocole de préparation de l'échantillon très peu sélectif. La Figure 42.A indique le pourcentage et le nombre de métabolites identifiés selon leurs origines (endogène ou exogène). La Figure 42.B indique leurs différentes familles chimiques. L'ensemble de ces informations proviennent de la base de données sur le métabolome humain (HMDB) [234].

La chimiothèque utilisée contient principalement des composés endogènes, ainsi une majorité (88 %) des métabolites identifiés dans le méconium sont endogènes. Au sein de ces métabolites, nous retrouvons essentiellement des aminoacides (29 %), des lipides (20 %), des carbohydrates (14 %) et des nucléosides (4 %). Des pourcentages équivalents ont été décrits au laboratoire au sein du sérum humain ou du lait maternel, avec à titre d'exemple pour le lait maternel 24 % d'acides aminés, 22 % de lipides et 7 % de nucléosides [32]. Un certain nombre de métabolites exogènes sont identifiés, incluant des métabolites de nature alimentaire, médicamenteuse ou cosmétique. Les métabolites microbiens quant à eux peuvent être aussi bien endogènes qu'exogènes.

Parmi les composés identifiés, nous retrouvons par exemple des acides aminés essentiels provenant de l'alimentation, incluant l'histidine, l'isoleucine ou encore la phenylalanine. Nous retrouvons de l'eugenol, métabolite provenant de l'essence de clou de girofle, souvent retrouvé dans des produits cosmétiques, utilisée en tant qu'huile essentielle. Des métabolites de médicaments tels que l'acétaminophène-glucuronide, un métabolite du paracétamol ou encore le salicylamide utilisé comme

antalgique et anti-inflammatoire, sont également identifiés. Parmi les métabolites microbiens nous retrouvons de l'acide muramique généralement présent dans les parois bactériennes ou encore l'acide isopropylmalique retrouvé chez les *Acetobacter*. Parmi les acides biliaires identifiés dans le méconium, nous retrouvons les acides cholique et glycocholique, ainsi que le tauroolithocholic-3-sulfate. Dans leur étude, Naritaka et al ont indiqué la présence d'acides biliaires chez le fœtus dès la 12^{ème} semaine de gestation, retrouvés dans l'analyse du méconium.



Figure 42. Origines (A) et familles (B) des métabolites identifiés dans le méconium.

Les auteurs ont indiqué que 80 à 90 % de la composition globale en acides biliaires dans le méconium était représentée par les acides taurocholique, taurochenodeoxycholique ainsi que par les acides glycholique et glycochenodeoxycholique. Les acides biliaires fœtaux sont principalement conjugués à la taurine, et peuvent être produits par la bile du fœtus ou provenant de la mère à travers le liquide amniotique avalé [254]. Nous observons également la présence de dérivés du tryptophane notamment la kynurénine et l'acide quinolique, issus de la dégradation du tryptophane par les bactéries [280], ces composés ont déjà été décrits dans des études précédentes réalisées sur le méconium [264].

Wandro et al ont étudié la composition métabolique de 77 échantillons de fèces excrétées à partir du 7^{ème} et jusqu'au 75^{ème} jour suivant la naissance. Les auteurs ont pu identifier par GC/MS la présence de 224 métabolites primaires de diverses familles, incluant des aminoacides, des acides gras et biliaires, des nucléotides et des lipides dont une cinquantaine sont communs à notre cartographie, incluant le tryptophane, la taurine, l'hexose-phosphate ou encore l'acide maléique [263]. Cette étude est l'une des seules réalisant une cartographie d'échantillons de premières selles, même si celles-ci ne sont plus considérées comme des méconiums. En effet, ces échantillons impliquent un début de contamination par l'alimentation, les processus physiologiques du nouveau-né et le microbiote intestinal [264].

Notre approche non ciblée permet donc de fournir la première cartographie du métabolome du méconium précoce, i.e. de la matière fécale excrétée au cours du premier et du deuxième jour de vie. Celle-ci fournit le plus grand nombre de métabolites identifiés dans le méconium à ce jour, avec le plus haut niveau de confiance [210]. La couverture du métabolome réalisée pourrait être élargie avec la détection de composés plus apolaires. Elle sera ainsi complétée dans le Chapitre suivant par des analyses complémentaires dans les conditions C18-ESI⁺. De plus, le prochain Chapitre abordera l'analyse d'échantillons de méconium issus d'une petite cohorte d'une dizaine de nouveau-nés, permettant potentiellement d'avoir accès à une plus grande diversité métabolique.

IV. Conclusions du chapitre I et perspectives

Les développements méthodologiques réalisés sur les deux échantillons de méconium disponibles au laboratoire nous ont permis de mettre en place une méthodologie robuste de préparation de l'échantillon. Nous avons à travers les deux échantillons de méconiums analysés dans les conditions HILIC-ESI⁻, réalisé une première annotation et identification des métabolites présents dans cette matrice foétale. Dans ces conditions, 137 métabolites ont été identifiés avec un haut niveau de confiance. Cette cartographie reste néanmoins non exhaustive et sera complétée en réalisant en parallèle une séparation des composés plus apolaires sur une colonne C18 couplée à une détection des molécules dans le mode d'ionisation positif (C18-ESI⁺), offrant une plus grande couverture chimique [197]. Ces analyses seront d'abord réalisées sur une cohorte de 35 échantillons de méconium issus de 11 nouveau-nés, permettant : i) de valider notre méthodologie de préparation de l'échantillon, ii) d'apprécier la variabilité interindividuelle inter-échantillons, et iii) d'annoter potentiellement un plus grand nombre de métabolites. Par ces différents prélèvements longitudinaux, cette cohorte permettra d'avoir un premier aperçu de l'évolution du méconium au cours des 3 premiers jours de vie. Les résultats des analyses de la cohorte des 35 échantillons, appelée « cohorte clinique » sont décrits dans le Chapitre qui suit.

CHAPITRE II : Réalisation d'une cartographie de la composition du méconium et étude de sa variabilité interindividuelle et de son évolution durant les premiers jours de vie par des approches multi-omiques : métabolomique, métaprotéomique et métagénomique ciblée

I. Introduction

Les progrès récents des sciences « omiques » (génomique, épigénomique, transcriptomique, métaprotéomique, métabolomique...), permettent de suivre quantitativement et à haut débit l'abondance d'un grand nombre de molécules au sein d'un système biologique donné. La combinaison de ces approches omiques représente un outil puissant dans la compréhension de la biologie des systèmes et de leur dynamique fonctionnelle, en interaction avec leur environnement. Dans cette optique, nous avons cherché à combiner plusieurs approches « omiques », réalisées sur les mêmes échantillons de méconiums, permettant d'avoir une vision la plus intégrée et globale possible. Les approches omiques de l'étude incluent : la métabolomique, la métaprotéomique et la métagénomique ciblée (ADNr-16S) ».

Cette étude a été réalisée sur des échantillons de méconium collectés longitudinalement chez 11 nouveau-nés. Les échantillons (n = 35) ont été collectés entre le jour 1 (jour de naissance) et le jour 3 post-partum, et constituent ce que nous appellerons par la suite « cohorte clinique ». Un premier objectif est de valider les outils méthodologiques préalablement développés pour les analyses métabolomiques et de vérifier la pertinence de la couverture métabolique précédemment obtenue (Chapitre I). De plus, nous avons souhaité combiner ces données aux approches de l'analyse globale du microbiote et des protéines, afin de fournir une cartographie la plus exhaustive et intégrée possible du méconium. Un troisième objectif est d'apprécier de façon préliminaire la variabilité interindividuelle de la composition du méconium, et de suivre son évolution au cours des trois premiers jours de vie. Les outils mis en place seront appliqués ensuite à la cohorte EDEN, intégrant les échantillons de méconiums issus de 308 nouveau-nés, qui regroupe de larges informations sur l'environnement de la mère durant la grossesse (alimentation, lieu de vie...) (Chapitre III).

II. Matériels et méthodes

A. Echantillons de méconiums

Les échantillons de méconiums ont été collectés au cours du mois de février 2019 au sein de la clinique Sainte Thérèse (Paris 17^{ème}). Un consentement a été signé par les deux parents en amont des prélèvements, après que toutes les informations sur les objectifs de la recherche et les analyses réalisées aient été expliquées. Le projet a été validé en amont par le conseil administratif et scientifique de la clinique, et l'équipe médicale a également été informée du projet et de ses objectifs. J'ai réalisé l'ensemble des prélèvements avec l'aide de l'équipe de cette maternité. Afin d'anonymiser les échantillons et en accord avec la clinique, aucune « métadonnée » n'a été collectée (âge et morphologie des mères, caractéristiques socio-économiques, antécédents d'atopie ou d'autres problèmes de santé, médicaments de la mère ou de l'enfant...).

Les méconiums sont issus de 11 nouveau-nés, dont deux de sexe féminin et neuf de sexe masculin. Tous sont nés à terme et par voie basse. Les différentes excréments émises par le nouveau-né ont été collectées, jusqu'à ce qu'un changement de couleur et de texture de la matière fécale apparaisse, traduisant l'apparition des premières selles. Le nombre d'excréments et le moment de leur survenue, très variables d'un nouveau-né à l'autre, ont été précisément relevés. Le Tableau 7 renseigne le code des nouveau-nés (NBx), leur sexe et le temps de collecte des échantillons en minutes post-naissance (ainsi que l'équivalent en heure). Les échantillons seront par la suite codifiés selon ces deux données : NBx.temps de prélèvement en minutes (par exemple, NB1.390 pour l'échantillon du nouveau-né #1 collecté à 390 minutes post-partum).

Les couches « pleines » étaient placées à + 4°C au moment du change du nouveau-né par la mère ou la puéricultrice, qui notait également l'heure du change. Les échantillons de méconiums ont ensuite été collectés à l'aide d'une spatule et d'un tube stériles, sans gratter la surface de la couche. Ils ont été homogénéisés manuellement dans un tube stérile à l'aide de la spatule, puis séparés en 3 aliquots pour les 3 approches d'analyses. Ces aliquots sont immédiatement placés à - 20 °C jusqu'à leur transport au laboratoire où ils ont été transférés à - 80°C. Le délai entre le change et la collecte était au maximum de 12 heures. Afin d'homogénéiser les prélèvements et de limiter un biais expérimental lié à une contamination par des substances chimiques issues de la couche, nous avons fourni les mêmes couches à tous les participants. L'ensemble des échantillons collectés n'étaient pas disponibles en quantités suffisantes pour réaliser les trois approches omiques, nous avons ainsi prioriser l'analyse métabolomique faisant l'objet principal de mon travail de thèse, puis réaliser l'analyse métagénomique et ensuite l'analyse métagénomique ciblée (Tableau 7). Suite à un problème

expérimental lors de la préparation des échantillons NB1.1875 et NB4.3780, ces échantillons n'ont pas pu être analysés en métabolomique.

Tableau 7 : Codification et analyses réalisées sur les 35 échantillons de méconium collectés.

Données sur les échantillons				Analyses réalisées		
Nouveauté	Sexe	Temps de collecte Minutes (heures)	Code	Métabolomique (n = 33)	Métabolomique (n = 30)	Métagénomique ciblée (n = 24)
NB1	F	390 (6H30)	NB1.390	X	X	X
		915 (15H15)	NB1.915	X	X	X
		1875 (31H15)	NB1.1875	-	X	X
		3060 (51H)	NB1.3060	X	X	X
NB2	M	570 (9H30)	NB2.570	X	X	X
		1005 (16H45)	NB2.1005	X	-	-
		2640 (44H)	NB2.2640	X	X	X
NB3	M	420 (7H)	NB3.420	X	X	X
		1440 (24H)	NB3.1440	X	X	X
		4740 (79H)	NB3.4740	X	X	X
NB4	M	1170 (19H30)	NB4.1170	X	X	X
		1620 (27H)	NB4.1620	X	-	-
		3780 (63H)	NB4.3780	-	X	X
NB5	M	420 (7H)	NB5.420	X	X	X
		930 (15H30)	NB5.930	X	X	X
NB6	M	960 (16H)	NB6.960	X	X	X
		2520 (42H)	NB6.2520	X	X	X
NB7	M	1140 (19H)	NB7.1140	X	X	X
		1350 (22H30)	NB7.1350	X	X	X
		2580 (43H)	NB7.2580	X	X	X
NB8	M	270 (4H30)	NB8.270	X	X	-
		1020 (17H)	NB8.1020	X	X	X
		1200 (20H)	NB8.1200	X	-	-
		2640 (44H)	NB8.2640	X	X	-
		2880 (48H)	NB8.2880	X	X	X
NB9	F	195 (3H15)	NB9.195	X	X	X
		1215 (20H15)	NB9.1215	X	X	-
		1305 (21H45)	NB9.1305	X	-	-
		1920 (32H)	NB9.1920	X	X	X
		2700 (45H)	NB9.2700	X	X	-
		2820 (47H)	NB9.2820	X	X	-
NB10	M	90 (1H30)	NB10.90	X	X	-
		1020 (17H)	NB10.1020	X	-	X
		1065 (17H45)	NB10.1065	X	X	-
NB11	M	795 (13H15)	NB11.795	X	X	X

B. Analyses métabolomiques

1. « Workflow » suivi pour l'analyse métabolomique des échantillons de la cohorte clinique

La procédure suivie pour les analyses métabolomiques est très similaire à celle décrite dans le Chapitre précédent et est représentée en six étapes dans la Figure 43, incluant : (1) la collecte des échantillons de méconium, (2) la préparation des échantillons, (3) l'acquisition des données LC/HRMS (HILIC-ESI⁻ et C18-ESI⁺), (4) le traitement des données, (5) l'analyse MS/MS et enfin (6) l'analyse statistique et l'interprétation biologique des données.

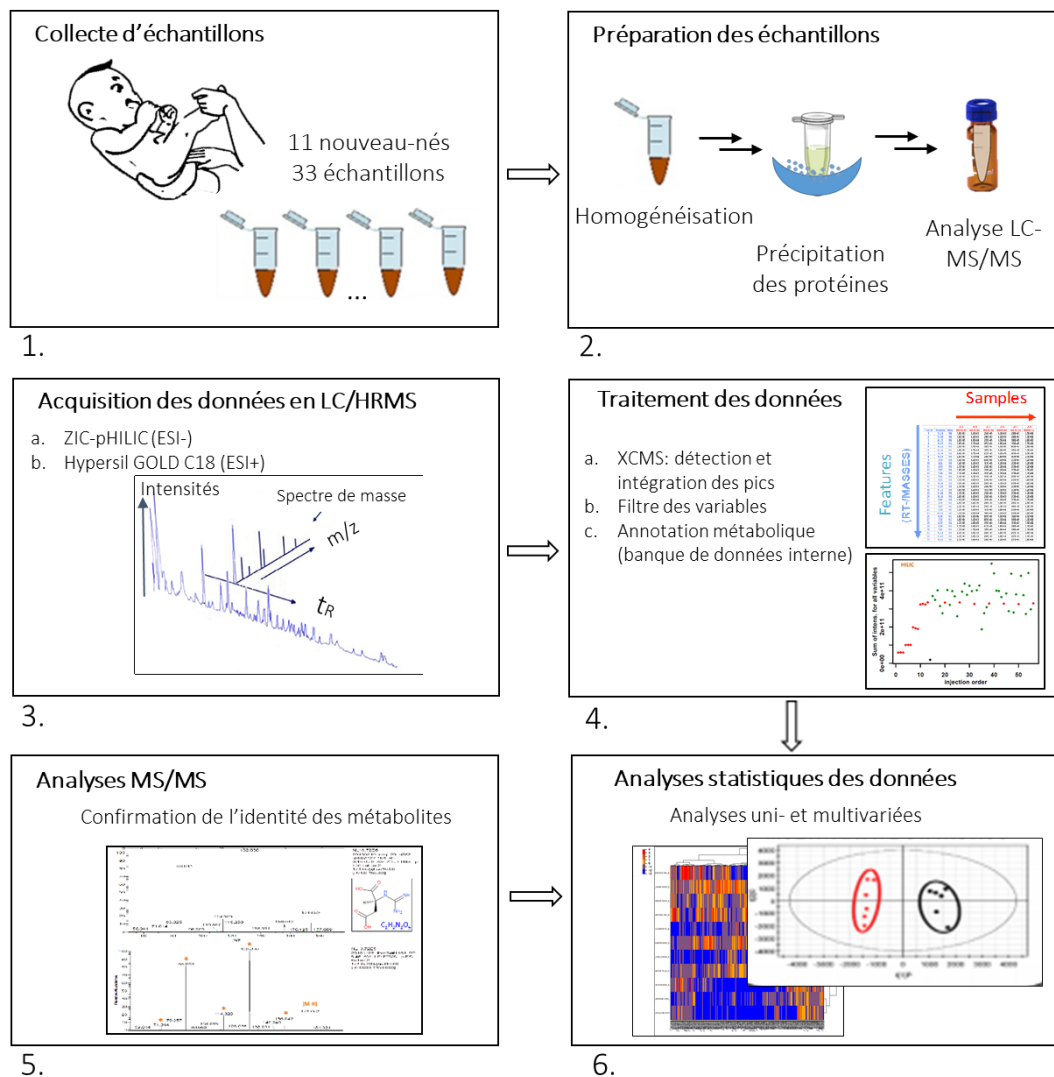


Figure 43 : Procédure suivie pour l'analyse métabolomique globale des échantillons de méconium de la cohorte clinique.

2. Préparation des échantillons de méconium

Le même protocole de prétraitement de l'échantillon de méconium décrit précédemment dans les conditions analytiques HILIC-ESI⁻ (chapitre I. Partie II.C) est appliqué pour l'analyse dans les conditions C18-ESI⁺. La préparation de l'échantillon pour les analyses HILIC-ESI⁻ est identique au Chapitre I.

Pour les analyses en C18-ESI⁺, le culot métabolique sec est solubilisé dans 800 µL de solvant de reprise constitué d'un mélange eau / Acide formique (AF) / Acétonitrile (ACN) (Eau + 0,1 % AF : ACN + 0,1 % AF, 95:5, v/v). Après 30 secondes d'agitation, une centrifugation est réalisée (20000 x g, 4 °C, 15 minutes). Cent µL du surnageant sont ensuite transférés dans un vial d'injection. L'échantillon de QC est préparé par le mélange de 50 µL de chaque échantillon de méconium. Cent microlitres du mélange sont redistribués dans chaque tube de QC puis des échantillons de QC dilués sont préparés (dilutions au 1/2, 1/4 et 1/8) dans le solvant de reprise. Cent microlitres de ce tampon seul sont également distribués dans un vial d'injection qui servira de blanc analytique. Enfin, 5 µL du mélange d'étalons internes sont ajoutés à chaque tube, excepté dans le blanc analytique (Annexe 3).

3. Acquisition des données LC/HRMS

a) La séquence analytique

Afin d'éviter les facteurs confondants liés par exemple à des paramètres physiologiques (sexe, échantillons appartenant au même nouveau-né, ou à des temps de prélèvements proches...), les échantillons ont été analysés dans un ordre aléatoire. Au vu du nombre modeste d'échantillons de méconiums de cette cohorte (33 échantillons), un QC a été injecté tous les cinq échantillons de méconiums analysés. Les séquences analytiques lancées dans les deux conditions d'analyses HILIC-ESI⁻ et C18-ESI⁺ sont représentées dans la Figure 44.

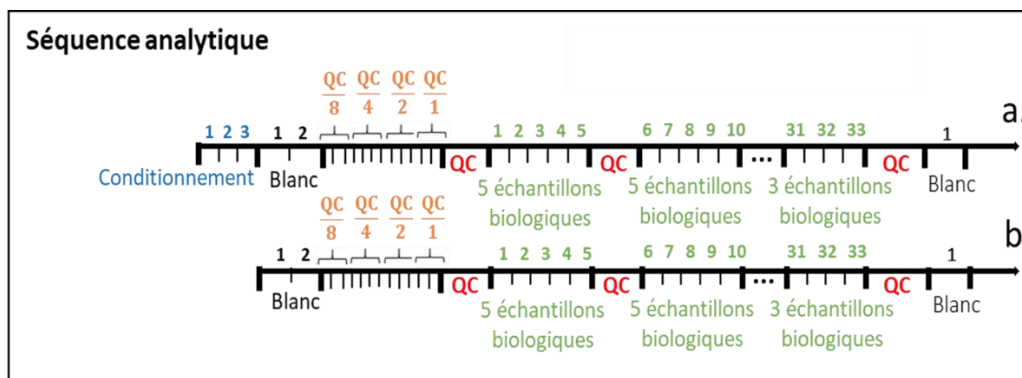


Figure 44 : Séquences analytiques réalisées dans les conditions LC/MS HILIC-ESI⁻ (a) et C18-ESI⁺ (b).

b) Acquisition des données dans les conditions analytiques HILIC-ESI⁻

Les mêmes conditions analytiques que celles décrites dans le chapitre I. Partie II.D sont utilisées.

c) Acquisition des données dans les conditions analytiques C18-ESI⁺

c.1) Conditions chromatographiques

La séparation chromatographique a été réalisée avec un système de type Dionex Ultimate 3000. Une colonne Hypersil-GOLD-C18 (1,9 μm ; 2,1 \times 150 mm ; Thermo Fisher Scientific, Les Ulis, France) est utilisée, avec des phases mobiles constituées d'une phase aqueuse en A ($\text{H}_2\text{O} + 0,1\%$ AF) et d'une phase organique en B (ACN + 0,1 % AF). L'élution chromatographique débute avec une phase isocratique à 5 % de la phase mobile B et 95 % de la phase mobile A durant 2 minutes, suivie d'un gradient linéaire durant 11 minutes, allant de 5 à 100 % de la phase mobile B. Le système chromatographique est ensuite rincé pendant 12 minutes par la phase organique (100 % de B). La colonne est ensuite rééquilibrée dans conditions initiales (5 % de B) durant 5 minutes. Le débit appliqué est de 0,5 mL/min et un volume de 10 μL d'échantillon est injecté. Les températures du système HPLC sont fixées à 4 °C pour l'injecteur automatique, et 30 °C pour la colonne chromatographique.

c.2) Analyse par spectrométrie de masse

L'acquisition spectrale a été réalisée comme pour l'analyse en conditions HILIC-ESI⁻ sur un spectromètre de masse de type Q-Exactive (Thermo Fisher Scientific, Les Ulis, France), utilisant une source electrospray opérant dans le mode d'ionisation positif avec des tension et température de capillaire respectivement de + 5000 V et 280 °C. Les pressions de gaz de nébulisation et de gaz auxiliaire (azote) sont respectivement fixées à 60 et 10 unités arbitraires. Le pouvoir résolutif du spectromètre de masse est fixé à 50000 pour une largeur à mi-hauteur (FWHM) à m/z 200. Une calibration externe du spectromètre de masse est réalisée chaque début de semaine et/ou début de séquence en utilisant une solution commerciale (Calmix positif, Thermo Fisher Scientific). La détection est réalisée sur une gamme de m/z de 75 à 1000. Les expériences MS/MS ont été réalisées en mode d'activation non résonnante à des valeurs d'énergies de collision normalisées de 10, 20 et 40 %.

4. Traitement des données

a) Extraction automatique des données par XCMS

L'outil XCMS de Workflow4Metabolomics de Galaxy a connu une mise à jour en amont de ces traitements de données (XCMS.3.2.1). Les différentes étapes de l'XCMS à travers cette nouvelle version sont décrites en Annexe 7. Les paramètres de l'XCMS utilisé pour chacune des conditions HILIC-ESI⁻ et C18-ESI⁺ sont décrits en Annexe 8 et Annexe 9 respectivement.

b) Annotation des variables

Un nouveau script sous langage R a été développé au sein du laboratoire pour l'annotation des variables extraites et sera utilisé dans la suite des analyses métabolomiques de cette thèse (script « mlomic »). Ce script a été développé dans un environnement « shiny », permettant d'intervenir sur davantage de paramètres d'annotation, mais également d'effectuer simultanément des analyses statistiques avec des graphiques interactifs. Pour les annotations, les tolérances en temps de rétention acceptés sont fixées respectivement à +/- 90 et +/- 15 secondes pour les colonnes HILIC et C18. Une tolérance dans la mesure de la masse précise est fixée à 10 ppm pour les deux modes d'ionisations. Enfin, un filtre est réalisé ne retenant que les variables ayant des aires supérieures respectivement à 30000 et 10000 pour les conditions HILIC-ESI⁻ et C18-ESI⁺.

Les voies métaboliques représentées dans le jeu de données ont été attribuées à l'aide de la banque KEGG sur l'outil « pathway enrichment » de MetaboAnalyst 5.0 (www.metaboanalyst.ca) [281].

c) Analyses statistiques

L'ensemble des analyses statistiques ont été réalisées sur W4M de Galaxy. Pour les analyses univariées (test de Wilcoxon), une correction pour tests multiples de type FDR est appliquée. Pour les analyses multivariées (ACP, PLS-DA), les données sont centrées réduites et une transformation log₁₀ est appliquée. Une classification hiérarchique non supervisée des échantillons et des variables (centrées et réduites) est également réalisée, et visualisée sous la forme d'une « carte de chaleur » (« heatmap »).

C. Analyses métabolomiques

1. « Workflow » suivi pour l'analyse métabolomique des échantillons de la cohorte clinique

La procédure suivie pour l'analyse métabolomique des échantillons de méconium de la cohorte clinique est décrite en quatre étapes dans la Figure 45, incluant : (1) la collecte des échantillons de méconium (la même que précédemment, Partie II.A.), (2) la préparation des échantillons, (3) l'acquisition des données par LC-MS/MS (nanoLC-MS/MS dans le mode d'ionisation positif), (4) le traitement et exploitation des données. Ces analyses ont été réalisées au Laboratoire Innovations technologiques pour la Détection et le Diagnostic (LI2D) du CEA de Marcoule par Guylaine Miotello, Lucia Grenga et Jean Armengaud.

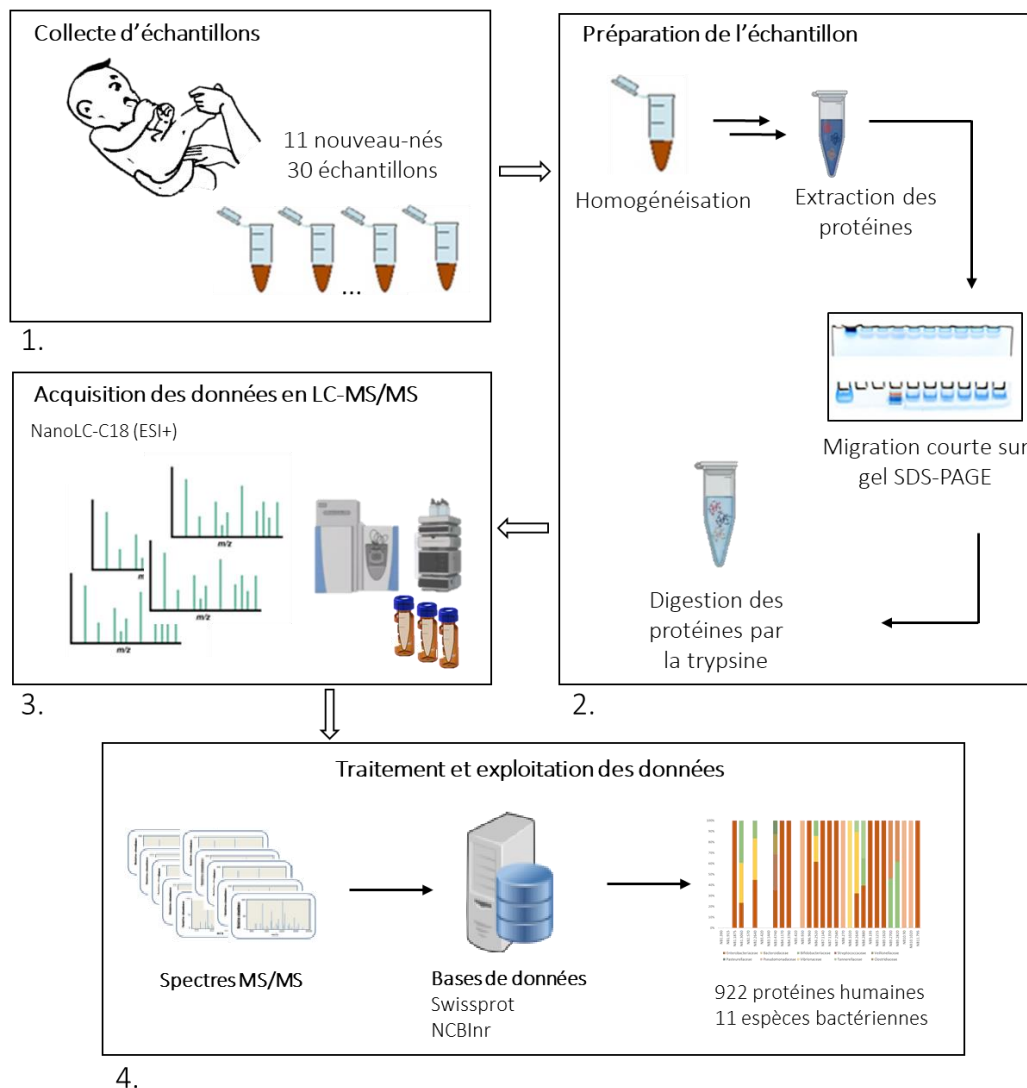


Figure 45 : Procédure suivie pour l'analyse métabolomique des échantillons de méconium de la cohorte clinique.

2. Préparation des échantillons

Des tests préalables ont été réalisés sur le méconium permettant de montrer que l'on pouvait se passer de l'étape de lyophilisation, nécessaire à l'analyse métabolomique des mêmes échantillons. Ainsi des aliquotes de 30 mg de méconium brut ont été pesés dans un tube de 1,5 mL (Eppendorf), puis repris dans 1 mL d'eau distillée stérile et incubés pendant 5 minutes à 90 °C dans un thermomixeur. L'extraction des protéines est réalisée telle que décrit par Hayoun et al. [291]. Les échantillons sont transférés dans des tubes Precellys de 2 mL (Sarstedt), contenant 200 mg d'un mélange de billes de silice (0,1 mm) et verre (0,5 et 0,1 mm) en proportions égales. Le broyage cellulaire est réalisé par le Precellys Evolution (Bertin Technologies), avec les paramètres suivants : 10000 tr/min, 10 cycles de 30 secondes et 30 secondes de pause entre chaque cycle. Après la lyse et une centrifugation à 16000 × *g* pendant 5 minutes, le surnageant est transféré dans des nouveaux tubes de 1,5 mL et les protéines sont précipitées par ajout d'acide trichloracétique (TCA) froid à une concentration finale de 10 % (p/v). Après une incubation de 10 minutes à 4 °C, le précipité contenant les protéines est récupéré par centrifugation à 16000 × *g* pendant 5 minutes. Les protéines sont solubilisées dans 30 µL de tampon LDS 1X (Lithium dodecyl sulfate, Invitrogen), contenant 26,5 mM Tris/HCl, 35,25 mM Tris base, 0,5 % de lithium dodecyl sulfate (p/v), 2,5 % de glycérol (p/v), 0,13 mM d'EDTA, 0,06 mM de SERVA Blue G-250, et 0,04 mM de rouge de phénol, tamponné à pH 8,5 et complété par 5 % de β-mercaptoéthanol (v/v). La purification et digestion des protéines est réalisée à partir du protocole décrit par Hartmann et al. [282]. Un volume de 25 µL est déposé sur un gel NuPAGE 4-12 % Bis-Tris et soumis à une courte migration SDS-PAGE. Les protéines sont colorées pendant 30 min avec du Coomassie SimplyBlue SafeStain (Thermo Fisher Scientific). Après découpe, les protéines sont directement digérées au sein des bandes de gel par l'ajout d'une solution enzymatique composée de 0,1 µg/µL de Trypsine Gold (Promega) et supplémentée par 0,1% de ProtéaseMax (Promega). Les peptides générés se remettent en solution et sont récupérés dans le surnageant de la solution enzymatique.

3. Acquisition des données de métaprotéomique en LC-MS/MS

Les peptides ont été analysés sur un système de nanoLC Ultimate 3000 (Thermo Fisher Scientific) couplé à un spectromètre de masse Q-Exacte HF (Thermo Fisher Scientific), comme décrit par Grenga et al. [283]. Les peptides sont déposés sur une micro-précolonne PepMap 100 C18 en phase inversée (5 µm, 100 Å, 300 µm i.d. × 5 mm, Thermo Fisher Scientific) avant la séparation des peptides sur une colonne nanoLC-PepMap 100 C18 (3 µm, 100 Å, 75 µm i.d. × 50 cm, Thermo Fisher Scientific) à un débit de 0,2 µL/min et en utilisant un gradient de 120 min d'ACN + 0,1 % d'AF. L'acquisition spectrale a été réalisée sur la gamme *m/z* 350-1800, et les 20 ions précurseurs les plus abondants ont été sélectionnés pour la

fragmentation avec un temps d'exclusion dynamique de 10 secondes. Les ions de charge (+2) et (+3) ont été sélectionnés pour l'analyse MS/MS.

4. Traitement des données, annotation et quantification des protéines

L'extraction des données associées aux peptides (temps de rétention, m/z de l'ion précurseur, ions fragments, etc...) est réalisée grâce au logiciel Proteome Discoverer (Thermo Fischer Scientific). Celui-ci génère des fichiers de type (« MGF », Mascot Generic Format) qui sont ensuite traités par le logiciel Mascot version 2.6.0 (Matrix Science). L'attribution des séquences peptidiques associées aux spectres est réalisée à l'aide de la base de données Swissprot Human (téléchargés en 2016, 20123 séquences et 1115742 résidus) pour l'identification des protéines humaines, et NCBI nr (National Center for Biotechnology Information, non-redundant ; téléchargés en 2018, 108307546 séquences et 41817980956 résidus) pour l'identification des protéines bactériennes. Les paramètres suivants ont été utilisés sur Mascot : spécificité complète de la trypsine, maximum de deux clivages manqués, tolérance de masse sur l'ion parent de 5 ppm et 0,02 Da pour les ions fragments, modification fixe des cystéines par carboxyamidométhylation (+ 57,0215), en considérant l'oxydation des méthionines (+ 15,9949) comme modification possible. Le traitement de ces données fournit deux tableaux :

- Un premier tableau regroupant le nombre de spectres MS/MS qui ont été attribués à des séquences de peptides (nombre appelé « PSMs », Peptide-Spectrum Matches). Ce premier tableau contient un grand nombre d'informations sur les peptides telles que leurs m/z , l'erreur d'attribution en ppm, le score Mascot du peptide, leur attribution aux différentes protéines.
- Un deuxième tableau des protéines identifiées à partir des peptides détectés. Ce tableau regroupe : i) la masse des protéines, ii) le nombre de spectres spécifiques détectés pour chaque protéine au sein de chacun des échantillons (nombre appelé « SC », Spectral Count), reflétant l'abondance des protéines, iii) leur abondance relative, par le calcul du NSAF (Normalized Spectral Abundance Factor) obtenu en divisant le nombre de SC de chaque protéine par la masse théorique du polypeptide en kDa ; ce facteur est exprimé en pourcentage une fois considéré toutes les valeurs de NSAF, comme décrit précédemment par Christie-Oleza et al. [284].

Les protéines sont regroupées si elles partagent au moins un peptide, et le comptage des PSMs de chaque protéine est basé sur le principe de parcimonie. Ce principe regroupe les peptides partagés avec la protéine la plus détectée telle qu'évaluée par les peptides spécifiques plutôt que de les attribuer aux différentes protéines possibles [122]. L'attribution avec un seul peptide spécifique peut générer de nombreuses attributions dues au hasard. Ainsi l'identification des protéines à travers deux peptides

distincts diminue cette probabilité d'erreur, et sera considérée dans notre étude [122]. L'évaluation du taux de faux positifs (FDR, False Discovery Rate) est ensuite réalisée grâce à l'utilisation d'une base de données leurre correspondant aux séquences inversées pour filtrer les fausses attributions. Une $FDR \leq 1\%$ est appliquée dans cette étude.

5. Analyse des données

Les analyses en composantes principales et les classifications hiérarchiques ont été réalisées sur Python, en utilisant les scripts décrits en Annexe 10.

D. Analyses métagénomiques ciblées

1. Workflow suivi pour l'analyse métagénomique ciblée des échantillons de la cohorte clinique

La procédure suivie pour l'analyse métagénomique ciblée des échantillons de la cohorte clinique est décrite en quatre étapes dans la Figure 46, incluant : (1) la collecte des échantillons de méconium (la même que précédemment, Partie II.A.), (2) la préparation des échantillons, (3) l'acquisition des données et (4) le traitement des données de séquençage. Cette approche a été effectuée au sein de l'institut Micalis (microbiologie de l'alimentation au service de la santé) de l'INRAE de Jouy en Josas, spécialisé dans le domaine de l'analyse du microbiote, par Claire Cherbuy et Aurélia Bruneau.

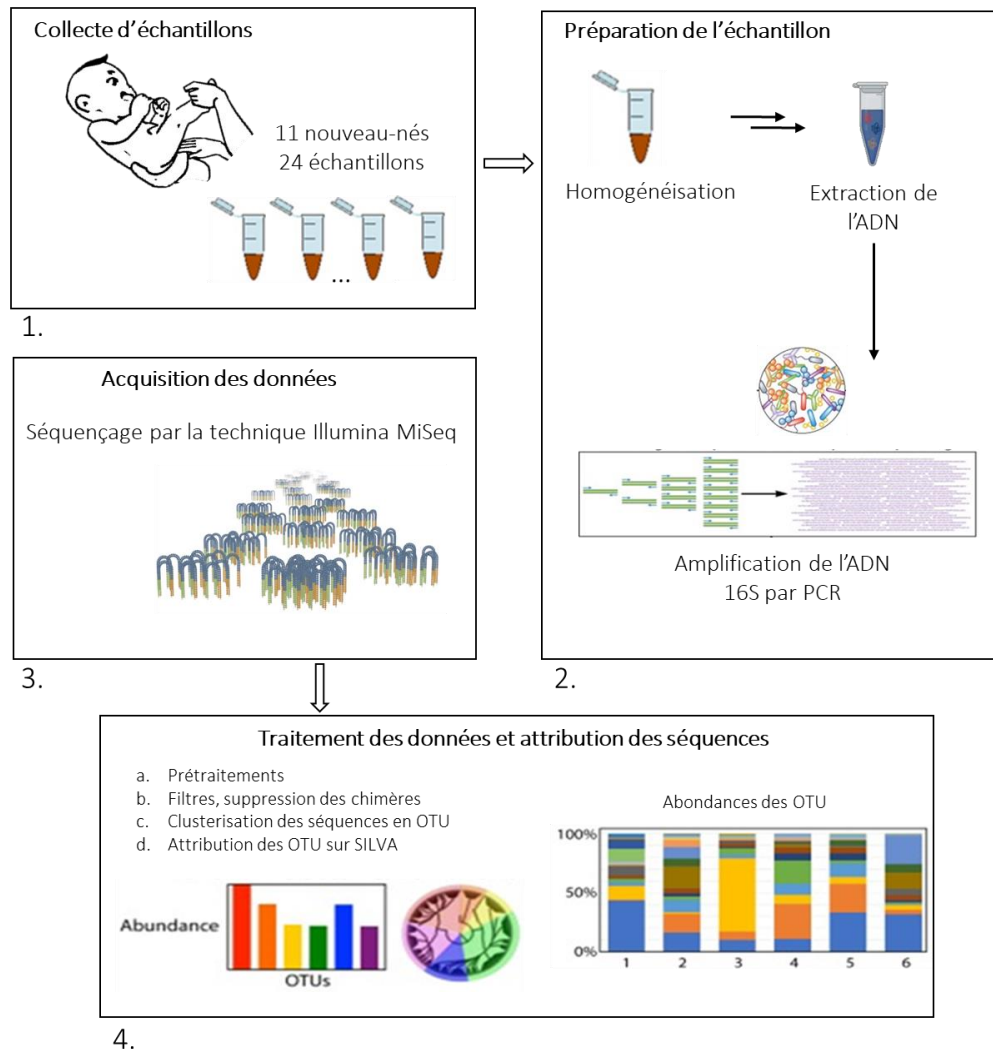


Figure 46 : Procédure suivie pour l'analyse du microbiote des échantillons de méconium de la cohorte clinique. OTU (de l'anglais, Operational Taxonomic Unit), PCR (de l'anglais, Polymerase Chain Reaction), SILVA : base de données des gènes codant pour l'ARN ribosomique [285].

2. Préparation des échantillons

Des tests préalables ont été réalisés pour optimiser les conditions d'extraction de l'ADN des échantillons de méconium, les conditions optimales sont présentées dans ce qui suit.

Les extractions de l'ADN génomique sont réalisées à l'aide du kit QIAamp PowerFecal DNA Kit (Qiagen), selon les recommandations du fournisseur et à l'aide de différentes solutions fournies (dont les compositions sont inconnues). Les aliquots de méconium bruts, non lyophilisés et non décongelés, stockés à $-80\text{ }^{\circ}\text{C}$ sont utilisés. Lors de la préparation des échantillons, les 24 aliquots disponibles sont sortis du congélateur et immédiatement placés dans de la carboglace pour éviter leur décongélation. Ces aliquots sont placés dans des tubes du kit « Dry Bead Tube » (Qiagen), et l'ensemble du protocole fourni par le fournisseur est suivi.

A la suite de l'extraction de l'ADN génomique, l'amplification des régions V3-V4 codant pour l'ARN 16S est réalisée. Cette amplification est réalisée par PCR (de l'anglais, Polymerase Chain Reaction) par la méthode « amplicon sequencing », à l'aide d'amorces permettant d'amplifier spécifiquement les séquences codant pour l'ARNr-16S procaryote. A la suite de l'amplification, les produits de PCR sont analysés sur gel d'agarose afin de visualiser les amplifications et leur qualité, puis les échantillons sont envoyés en sous-traitance (Plateforme @BRIDGe, INRAe) pour l'étape de séquençage par la technique « Illumina MiSeq » produisant 30000 lectures par échantillon.

3. Traitement des données de séquençage

Le traitement des données générées à la suite du séquençage est réalisé sur l'outil FROGS de la plateforme Galaxy [286].

a) Le « preprocessing » des données

Une première étape de dé-multiplexation des séquences en fonction de leur code barre est réalisée. Le code barre décrit les séquences appartenant à un même échantillon, car souvent plusieurs échantillons sont séquencés simultanément.

b) Le « clustering »

A cette étape, les fragments de séquences proches sont regroupés en clusters appelés OTUs (de l'anglais, Operational Taxonomic Unit), par une méthode récente appelée SWARM [287]. Les clusters sont regroupés dans une table avec des informations sur leurs abondances, en ne retenant que les clusters ayant une abondance supérieure à 0,05 % au sein de l'ensemble des OTUs.

c) L'étape « chimera »

Cette étape permet de filtrer les séquences de chimères générées lors de l'étape de l'amplification. Ces chimères sont à peu près de la même longueur que l'amplicon non chimérique avec une séquence d'amorçage à chaque extrémité de l'amplicon, comme décrit par Fichot et Norman [288].

d) L'affiliation des OTUs

A la suite des différentes étapes de filtres, l'affiliation des clusters retenus est réalisée par attribution d'une famille/classe ou encore d'une espèce taxonomique par homologie des séquences par rapport à celles contenues dans la base de donnée SILVA [285]. Cette base de données contient plus de 7 millions de séquences de gènes codants pour l'ARN ribosomique 16S, permettant l'identification des espèces bactériennes présentes.

Les représentations phylogéniques sont réalisées avec le package Phyloseq sur logiciel R.

III. Résultats expérimentaux et discussion

A. Analyses métabolomiques (article en préparation, Annexe 1).

1. Traitement XCMS des données générées par les analyses HILIC-ESI⁻ et C18-ESI⁺

Le traitement XCMS a extrait 9274 et 12397 variables dans les conditions HILIC-ESI⁻ et C18-ESI⁺ respectivement. Le nettoyage de ces données réalisé selon les étapes de filtres présentées en introduction et en Chapitre I, a permis de retenir respectivement 6843 et 8555 variables analytiquement pertinentes pour les conditions HILIC-ESI⁻ et C18-ESI⁺. Au sein de ces variables, plusieurs espèces moléculaires peuvent représenter un même analyte (ion monoisotopique, adduits, dimères ou fragments en source... etc). Ces redondances chimiques sont manuellement retirées lors de l'annotation décrite plus loin dans cette partie.

2. Suivi de la stabilité du système analytique dans les conditions HILIC-ESI⁻ et C18-ESI⁺

Le suivi de la stabilité des systèmes analytiques a été réalisé en calculant et en représentant la somme des intensités de toutes les variables détectées et filtrées au sein de chaque échantillon, en fonction de l'ordre d'injection. L'objectif d'une telle représentation est de permettre de visualiser les valeurs (et la stabilité) des intensités : i) des QC injectés régulièrement au sein de la séquence, ii) des QC dilués (qui doivent diminuer proportionnellement au facteur de dilution appliqué) et iii) des blancs analytiques (qui doivent générer peu de signal), ainsi que d'observer la dispersion des échantillons biologiques.

Les Figure 47.A (HILIC-ESI⁻) et Figure 47.B (C18-ESI⁺) permettent d'observer une très bonne stabilité des analyses au cours de la séquence. Celle-ci est évaluée à travers la somme des intensités des variables détectées au sein des QC (représentés en rouge), ayant un coefficient de variation de moins de 2 % pour les deux conditions analytiques. Les intensités au sein des QC dilués suivent l'ordre de dilution (divisées de deux en deux). La somme des intensités des variables au sein des échantillons biologiques sont dispersées autour des QC (constitués d'un mélange de l'ensemble de ces échantillons biologiques), avec une variabilité interindividuelle attendue. Ces observations permettent d'indiquer que le processus analytique est robuste et stable au cours de la séquence mais également que les variables extraites et filtrées semblent analytiquement pertinentes. A la vue de la représentation stable des QC (pas de dérive analytique observée), nous n'avons pas réalisé de correction ou de normalisation des signaux.

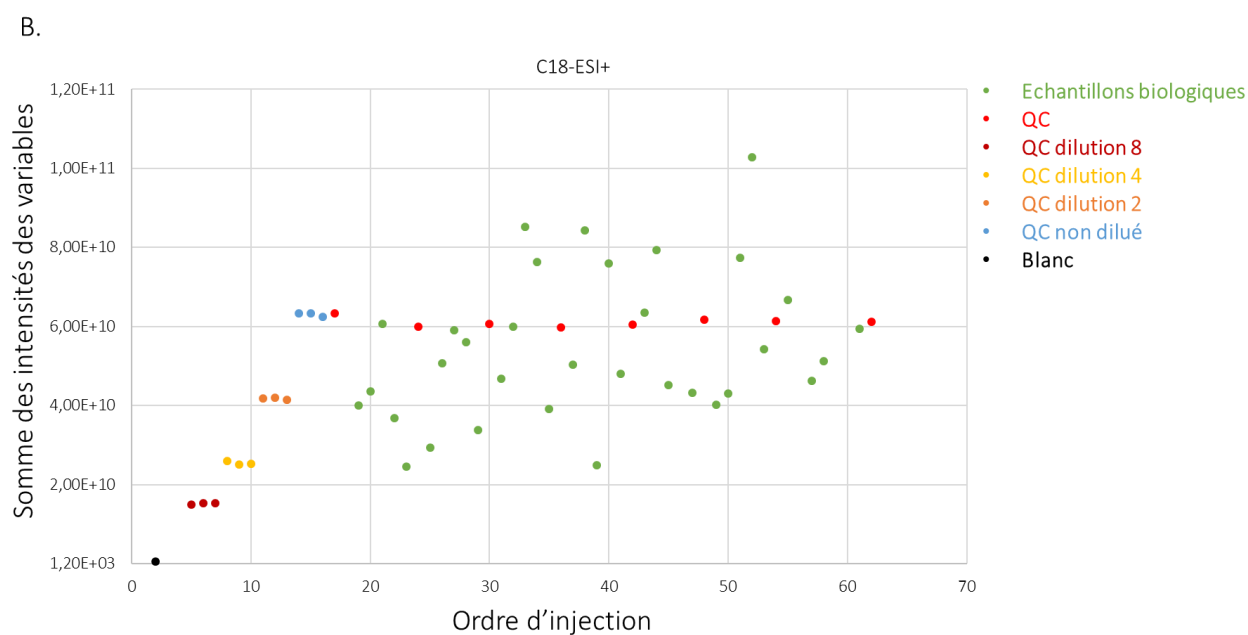
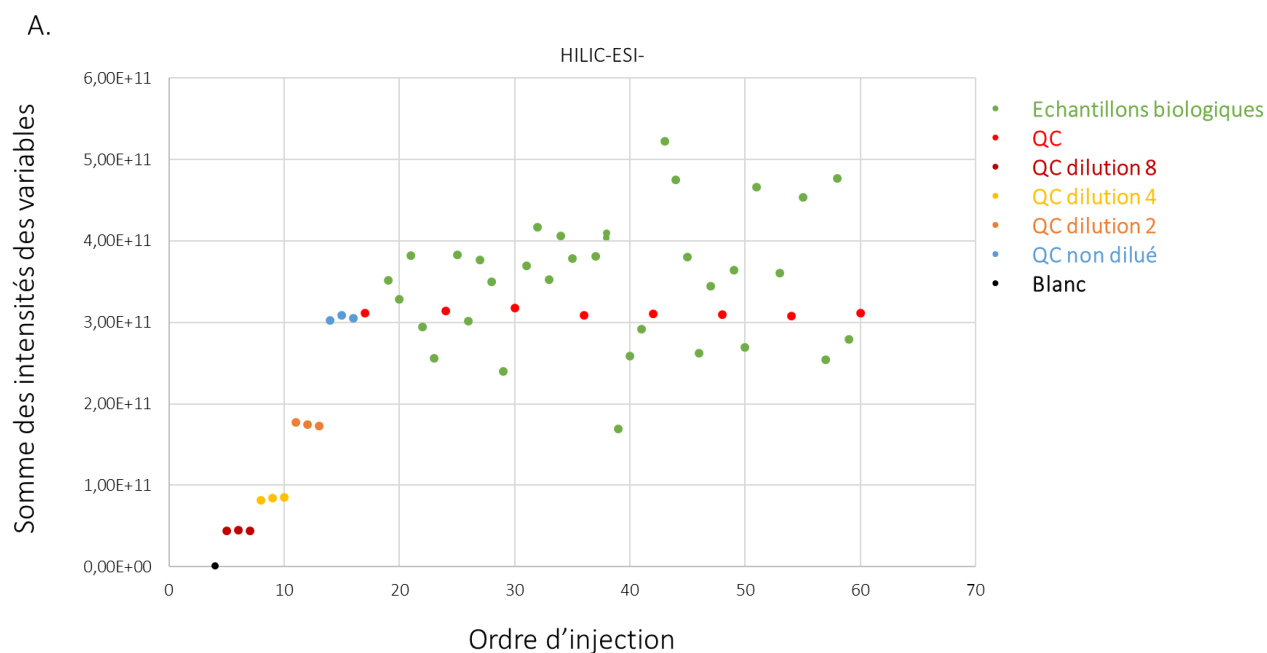


Figure 47 : Somme des intensités des variables détectées et retenues après filtres en fonction de l'ordre d'injection des échantillons dans la séquence. Conditions HILIC-ESI⁻ en (A) et C18-ESI⁺ en (B). Les échantillons de QC sont représentés en rouge, les échantillons biologiques en vert, les blancs analytiques en noir et les QC dilués en bleu (QC non dilué), orange (QC ½), jaune (QC ¼) et bordeaux (QC 1/8).

3. Annotation des variables

a) Annotation des variables et comparaison avec les résultats obtenus sur les échantillons A et B

L'annotation des variables et l'élimination de la redondance chimique au sein des variables annotées a permis d'identifier respectivement 194 et 163 métabolites pour les conditions analytiques HILIC-ESI⁻ et C18-ESI⁺. Dans un premier temps, nous avons comparé qualitativement ces annotations à celles obtenues au sein du chapitre I sur les échantillons A et B dans les conditions analytiques HILIC-ESI⁻, pour lesquelles nous avons identifiés 137 métabolites. Parmi ces métabolites identifiés dans les deux échantillons A et B, 30 métabolites n'ont pas été retrouvés dans la cohorte clinique. A l'inverse, près de 90 « nouveaux » métabolites ont été identifiés au sein de la cohorte clinique. Ce dernier résultat est cohérent étant donné que la cohorte clinique contient plus d'échantillons, et donc potentiellement un plus grand nombre de métabolites pouvant refléter une certaine variabilité interindividuelle. De plus, il peut exister une réelle différence de la composition des méconiums étudiés, étant donné qu'ils ont été collectés dans des environnements différents (clinique Parisienne *versus* milieu plus rural/périurbain pour les échantillons A et B), ce qui est dans la logique de notre hypothèse initiale. Nous avons cependant vérifié que ces différences n'étaient pas dues à un biais analytique du fait des nouveaux outils de traitement de données utilisés pour les échantillons de la cohorte clinique.

b) Evaluation de l'impact des nouveaux outils utilisés (XCMS.3.2.1 et annotation sur « mlomic ») sur les variables annotées

Afin de permettre de rendre compte si un biais analytique intervient par l'utilisation de ces nouveaux outils, nous avons comparé les paramètres appliqués au sein des deux méthodologies. En effet, ni la méthode de pré-traitement automatique des données « XCMS » ni la méthode d'annotation ne sont les mêmes entre ce chapitre et le précédent (Tableau 8).

Tableau 8 : Outils utilisés dans le pré-traitement XCMS et l'annotation des données obtenues dans les conditions HILIC-ESI⁻ au sein des chapitre I et II.

	Chapitre I (Echantillons A et B)	Chapitre II (Cohorte clinique – 33 échantillons)
Pré-traitement des données : Versions XCMS utilisées	XCMS.3.0.0	XCMS.3.2.1
Annotation des données : Tolérances acceptées au sein des temps de rétention (RT)	Script R $\Delta RT = X \pm RT^Y$ Où X = 5 et Y = 0,8	Mlomic $\Delta RT = \pm 45$ secondes
Seuil d'intensité	Aucun	« Cutoff » > 30000

Dans un premier temps, tous les paramètres pouvant être fixés par l'opérateur dans les versions 3.0.0 et 3.2.1 de XCMS ont été appliqués de façon comparable. Nous avons constaté que les étapes qui se rajoutent dans la nouvelle version du logiciel de détection automatique des pics n'ont *a priori* pas d'impact sur l'extraction ou sur l'intégration des pics, ni sur le nombre de variables détectées au final. En effet, nous avons de nouveau traité les données des échantillons A et B avec la nouvelle version XCMS.3.2.1, où nous avons obtenus un nombre de variables extraites comparable (respectivement 14863 et 13486 variables extraites à l'aide des versions 3.0.0 et 3.2.1, et respectivement 8034 et 8942 variables filtrées).

A l'inverse, la procédure d'annotation se montre très différente entre les deux méthodes utilisées, notamment au niveau des tolérances acceptées pour les temps de rétentions (RT). La tolérance acceptée pour les annotations des métabolites des échantillons A et B est définie par la fonction $\Delta RT = X \pm RT^Y$. Cette tolérance est donc dépendante du temps de rétention considéré : elle est d'autant plus importante que le temps de rétention est élevé. Dans la méthode d'annotation appliquée pour la cohorte clinique, cette tolérance est fixée à ± 45 secondes quel que soit le RT. Nous avons donc analysé à nouveau les données obtenues pour les échantillons A et B traitées par XCMS.3.0.0, en réalisant une annotation avec le nouvel outil « mlomic », avec une tolérance en temps de rétention fixée à ± 45 secondes. Cette méthode a conduit à l'annotation de 184 métabolites *versus* 210 métabolites précédemment annotés. Les 26 métabolites « manquants » nous ont donc semblé intéressants à suivre pour optimiser les paramètres d'annotation sur le nouvel outil utilisé.

Deux paramètres sont particulièrement pertinents à étudier au sein du nouvel outil d'annotation, à savoir i) les tolérances acceptées en temps de rétention et ii) le seuil d'intensité appliqué par défaut dans la nouvelle procédure. Ce seuil peut effectivement impacter le nombre de métabolites annotés, car le filtre supprime toutes les variables dont les aires sont inférieures à 30000 (considérées dans le bruit de fond) (Tableau 8).

Nous avons dans un premier temps étudié l'impact de la tolérance acceptée dans les temps de rétentions. Nous avons pour cela comparé différentes tolérances en temps de rétention, à savoir ± 45 , ± 60 , ± 90 et ± 120 secondes, ceci en maintenant le seuil d'intensité à 30000. Nous avons ensuite suivi l'ensemble des 26 métabolites manquants dans ces différentes conditions. Les résultats représentés en Figure 48.A, montrent que tous les composés manquants ont été annotés dès une tolérance supérieure ou égale à ± 90 secondes. La tolérance en temps de rétention de ± 120 secondes est cependant considérée comme trop élevée, pouvant générer beaucoup de redondances et potentiellement d'erreurs d'annotations.

Nous avons ensuite étudié l'impact du seuil d'intensité appliqué sur le nombre de métabolites finalement annotés. Pour cela nous avons réalisé et comparé différents tests, en paramétrant ce seuil à 0 (conditions du chapitre I), 5000, 8000 ou 30000, ceci en maintenant la tolérance en temps de rétention initiale, soit à +/- 45 secondes. Les résultats obtenus représentés en Figure 48.B, montrent des modifications mineures dans l'annotation des variables suite à l'application de ces différents seuils. En effet, uniquement 3 métabolites sur les 26 annotés manquants ont été retrouvés pour des valeurs de seuils variant de 0 à 8000 en intensité, incluant la sérine, le N-acetyl-asparagine, et l'acide urocanique.

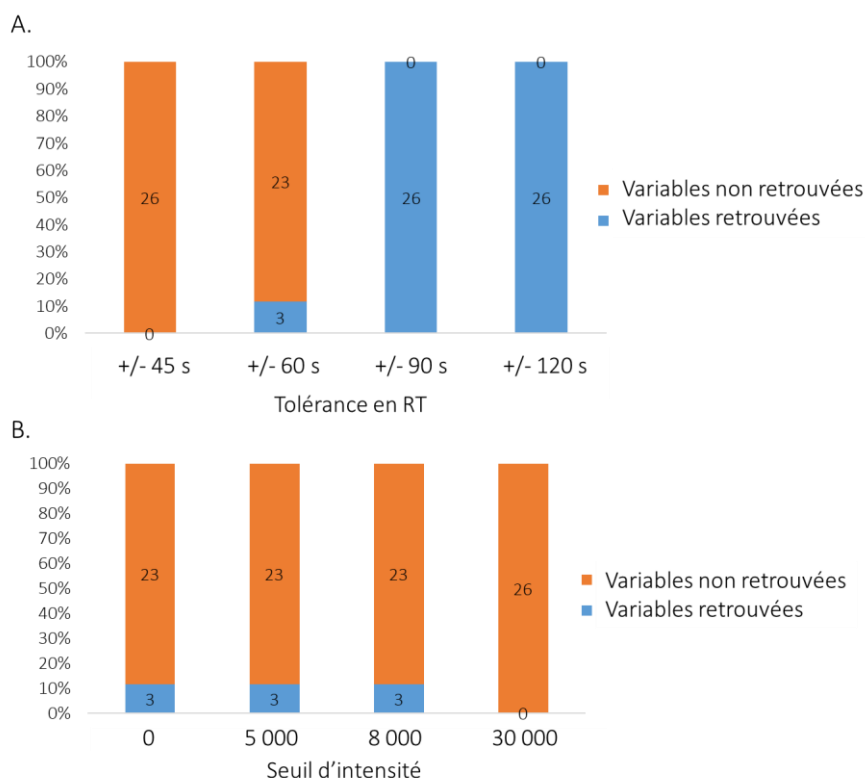


Figure 48 : Nombre de variables retrouvées (en bleu) parmi les 26 variables annotées au sein des échantillons A et B et manquantes au sein de la cohorte clinique. Nous faisons varier A. La tolérance en temps de rétention (+/- 45, 60, 90, 120 secondes ; avec un seuil d'intensité fixé à 30000) ou B. Le paramètre « seuil » d'intensité (0, 5000, 8000, 30000, avec une tolérance en temps de rétention fixée à +/- 45 secondes).

c) Re-annotation des variables de la cohorte clinique suivant le nouveau paramètre de tolérance au sein des RT

Nous avons ainsi réalisé l'annotation des variables détectées au sein des échantillons de la cohorte clinique, en appliquant les paramètres suivants : les seuils sont fixés à 30000 et à 10000, et les tolérances en temps de rétention sont fixées à +/- 90 et +/- 15 secondes dans les conditions HILIC-ESI⁻ et C18-ESI⁺ respectivement. Dans ces conditions, nous annotons 257 et 163 métabolites (avec

redondance). Nous observons qu'avec les nouveaux paramètres d'annotation, 63 métabolites supplémentaires sont annotés dans les conditions HILIC-ESI⁻ (194 vs. 257 annotations pour des tolérances en RT de +/- 45 sec vs. +/- 90 sec). L'ensemble des métabolites identifiés chez A et B et manquants au sein de la cohorte clinique ont été retrouvés avec ce nouveau paramètre. Après suppression manuelle des redondances chimiques au sein des deux jeux de variables annotées, nous retenons respectivement 224 et 149 métabolites annotés dans les conditions analytiques HILIC-ESI⁻ et C18-ESI⁺. Nous avons vérifié la pertinence de ces signaux sur le module QualBrowser d'Xcalibur selon les critères précédemment décrits (chapitre I. Partie III.A.3.a). Ces vérifications ont permis de retenir 212 et 131 métabolites annotés au sein des deux conditions HILIC-ESI⁻ et C18-ESI⁺ respectivement. Les annotations précédemment validées dans les conditions HILIC-ESI⁻ n'ont pas été analysées à nouveau en MS/MS. Ainsi 75 et 131 métabolites sont retenus respectivement dans les conditions analytiques HILIC-ESI⁻ et C18-ESI⁺ et ont été analysées par MS/MS. A la suite de ces analyses, 60 nouveaux métabolites ont été validés dans les conditions analytiques HILIC-ESI⁻ (en plus des 137 identifiés et vérifiés préalablement), et 89 métabolites dans les conditions analytiques C18-ESI⁺. L'ensemble de ces métabolites sont décrits en Annexe 11.

Nous retrouvons au sein des deux conditions analytiques utilisées, 56 métabolites communs incluant l'histidine, l'acétaminophène-glucuronide, l'acide maléique ou encore le 2-O-méthylinosine. Ainsi, en combinant les deux conditions analytiques, 230 métabolites uniques sont identifiés et validés dans le méconium.

4. Cartographie globale des métabolites identifiés dans le méconium

a) Familles et origine des métabolites identifiés

Les familles et les origines des 230 métabolites identifiés ont été attribuées à partir de la base de donnée HMDB (de l'anglais, Human Metabolome DataBase) et sont représentées en Figure 49. Parmi les familles les plus représentées, nous observons une distribution similaire à ce qui a été observé au laboratoire au sein d'autres matrices biologiques humaines telles que le plasma [198] ou le lait maternel [41], avec une forte représentation de la famille des acides aminés, peptides et analogues (27 %), des lipides (18 %) et des carbohydrates (14 %), représentant 59 % de la totalité des familles identifiées (Figure 49.A). Au sein des 40 % restants, nous décrivons majoritairement des acides organiques et dérivés (9 %), des composés hétéropolycycliques aromatiques (9 %) et des composés hétéromonocycliques aromatiques (7 %). Ces résultats restent cependant dépendants de la chimiothèque interne utilisée au laboratoire.

Parmi les 230 composés identifiés, nous retrouvons 87 % de métabolites endogènes et 13 % de métabolites exogènes (Figure 49.B). Parmi les métabolites exogènes, nous observons des métabolites

médicamenteux, cosmétiques, ou encore alimentaires. Au sein des métabolites microbiens, nous retrouvons l'agmatine ou la cadavérine résultant de réactions de décarboxylation respectivement des acides aminés arginine et lysine par les bactéries. D'autres composés tels que des acides quinoléiques et kynuréniques, issus de la métabolisation du tryptophane, sont également retrouvés au sein de notre cartographie.

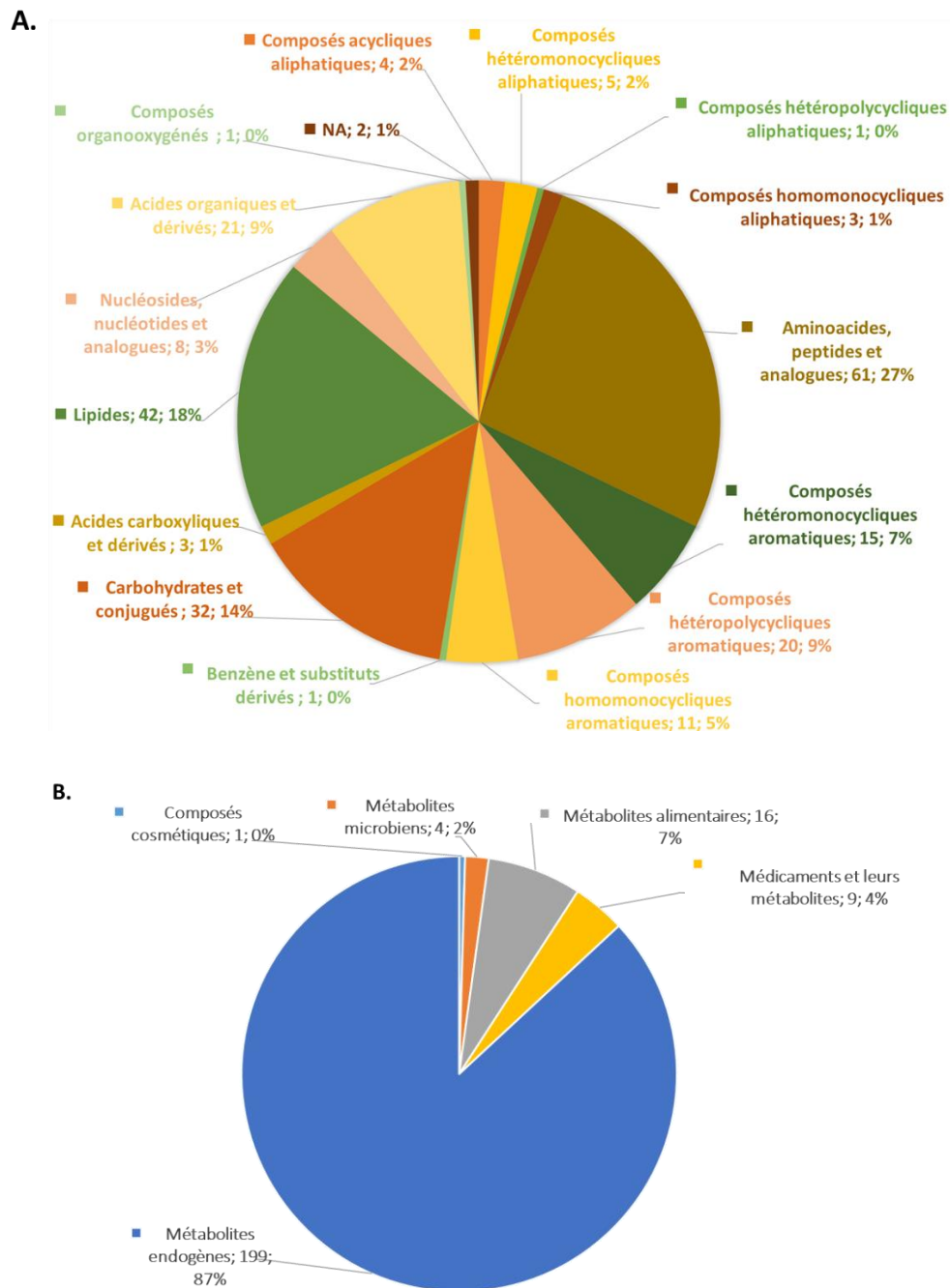


Figure 49.A. Répartition des 230 métabolites identifiés selon leurs familles. Le nombre ainsi que le pourcentage des métabolites au sein de chaque classe chimique sont indiqués. B. Origine des 230 métabolites identifiés.

Ces différentes identifications permettent de confirmer et de compléter les résultats obtenus de façon préliminaire avec les échantillons A et B, et permettent de rendre compte que la méthodologie mise en place a permis d'extraire, d'annoter et d'identifier un grand nombre de métabolites aux propriétés et classes chimiques très variées et de façon non ciblée.

b) Comparaison du métabolome du méconium au métabolome fécal

Les métabolites identifiés dans le méconium ont été comparés à ceux décrits dans les fèces, par recherche au sein de la base de données du métabolome fécal humain (Human Fecal Metabolome) [234]. La quasi-totalité des métabolites identifiés dans le méconium ont été précédemment décrits dans les fèces, sauf 9 métabolites, principalement des lipides, qui semblent spécifiques du méconium (Tableau 9). Parmi ces métabolites, certains ont été décrits dans l'urine, tel que l'estrone glucuronide [289]. Cependant, l'estrone glucuronide est retrouvé dans tous les échantillons de méconium, avec de fortes intensités (de l'ordre de 10^8), suggérant que sa présence ne résulte pas d'une contamination urinaire. Par ailleurs, nous avons identifié l'acide muconique dans tous les échantillons (intensités de l'ordre de 10^7) déjà décrit dans le méconium par Meyer et al, analysant la présence de métabolites issus d'expositions aux solvants BTEX (Benzène – Toluène – Ethylbenzène – Xylènes) [3], [47]. L'acide muconique dans sa forme « trans-trans », présente un fort intérêt biologique étant donné qu'il s'agit d'un composé résultant de la métabolisation du benzène et que sa détection est un marqueur de l'exposition de l'individu au benzène [290].

Tableau 9 : Liste des 9 métabolites retrouvés dans le méconium, non décrits dans les fèces de la base de données « Human Fecal Metabolome » .

Métabolites	Familles	Origines
2-O-Methylinosine	Nucléosides, Nucléotides, et Analogues	Endogène
4-Hydroxy-3-methoxyphenylglycol sulfate	Composés Homomonocycliques Aromatiques	Médicament
d-Desthiobiotin	Lipides	Endogène
Estrone glucuronide	Lipides	Médicament
Acide mésaconique	Lipides	Endogène
Acide muconique	Lipides	Endogène
N-Acetyl-carnosine	Aminoacides, Peptides et dérivés	Endogène
Succinic semialdehyde	Lipides	Endogène
Acide tiglique	Lipides	Endogène

c) Voies métaboliques impliquées dans notre cartographie

Nous avons ensuite repositionné l'ensemble des 230 métabolites identifiés dans les voies métaboliques en utilisant « KEGG human metabolic pathways » (mise à jour, en octobre 2019) (KEGG, de l'anglais, Kyoto Encyclopedia of Genes and Genomes). Au total, 52 voies métaboliques sont représentées à l'aide de l'outil « pathway enrichment » sur MetaboAnalyst 5.0. L'attribution des métabolites identifiés dans le méconium dans les différentes voies métaboliques est présentée Figure 50. Le ratio d'enrichissement en abscisse représente le nombre de métabolites attribués par voie métabolique dans notre jeu de données, sur le nombre total des métabolites appartenant à la voie étudiée. Selon ce critère, nous retrouvons de nombreuses voies impliquées dans la biosynthèse et le métabolisme des acides aminés et des protéines (arginine, glutamine, histidine, valine, leucine, isoleucine, phenylalanine, tryptophane... ; biosynthèse amino-acyl-tRNA), ainsi que le métabolisme de la caféine (paraxanthine, théobromine, caféine...). Ces résultats rejoignent des études précédentes décrivant ces voies métaboliques dans le méconium [80], [291], et sont en lien avec le fait que la majorité des métabolites retrouvés dans le méconium correspondent à des amino-acides, peptides ou analogues (Figure 49.A).

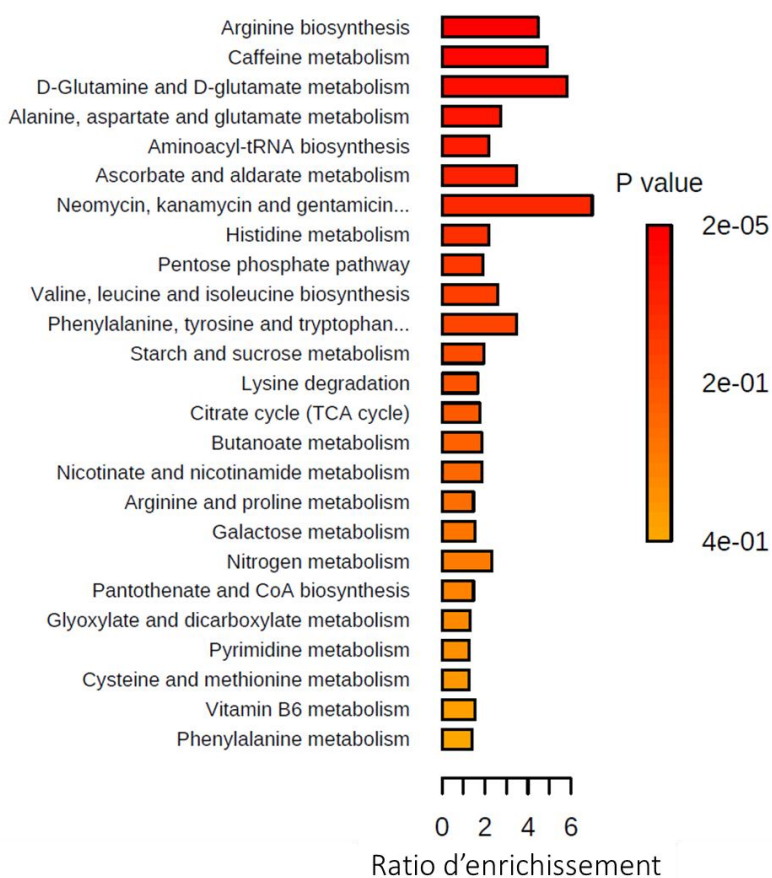


Figure 50 : Vue d'ensemble des 25 voies métaboliques les plus enrichies classées selon leur p-values au sein des 230 métabolites identifiés dans le méconium.

Les différents résultats présentés ont permis de décrire la cartographie la plus exhaustive du méconium en termes de couverture métabolique. A la suite, et par la comparaison des métabolomes du méconium et des fèces, nous avons défini des métabolites potentiellement spécifiques du méconium. La détection de certains métabolites, tels que l'acide muconique, qui sous sa forme trans-trans, est un marqueur d'exposition de l'individu au benzène.

5. Etude de l'évolution de la composition métabolique du méconium au cours des trois premiers jours de vie

a) Analyses non supervisées réalisées sur l'ensemble des variables extraites et filtrées

a.1) Analyses en composantes principales

Nous avons ensuite cherché à analyser l'évolution de la composition du méconium en métabolites au cours des trois premiers jours de vie, ainsi que d'apprécier sa variabilité interindividuelle. Pour ceci, nous avons utilisé différents outils statistiques permettant d'analyser et de représenter les données de façon supervisée et non supervisée.

Dans un premier temps, nous avons réalisé des analyses non supervisées de type ACP (analyses en composantes principales), permettant : i) de visualiser la répartition des échantillons sans *a priori*, ii) d'identifier certains échantillons ou variables extrêmes, iii) d'apprécier la structuration du jeu de données par des paramètres indépendants tels que le jour de prélèvement ou l'individu (seules métadonnées disponibles pour cette cohorte anonyme – le facteur sexe ne sera pas considéré étant donné le déséquilibre de genre dans notre cohorte, cf Tableau 7).

Lors d'une première analyse, les ACP ont été réalisées sur l'ensemble des variables détectées et retenues après les différents filtres appliqués, correspondant respectivement à 6843 et 8555 variables pour les conditions analytiques HILIC-ESI⁻ et C18-ESI⁺. Les graphes des individus obtenus dans ces conditions sont représentés en Figure 51. Au sein de ces représentations, 45 % (HILIC-ESI⁻) et 41 % (C18-ESI⁺) de la variabilité totale est expliquée à travers les deux premières dimensions de l'ACP (composantes principales t1 et t2). La troisième composante ne représentant que 10 et 8 % de l'inertie totale au sein des deux conditions respectives, nous décidons de ne retenir que les deux premiers axes pour la suite de notre analyse.

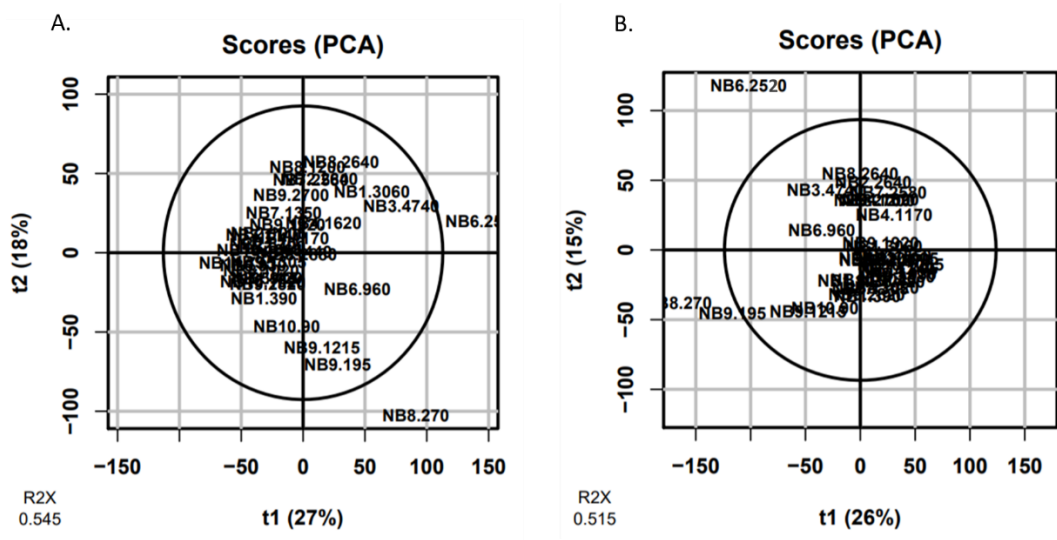


Figure 51 : Représentation des échantillons sur les deux premières composantes de l’analyse en composantes principales réalisée sur l’ensemble des variables extraites et filtrées, correspondant respectivement à 6843 et 8555 variables dans les conditions analytiques HILIC-ESI⁺ (A) et C18-ESI⁺ (B). Les données sont centrées-réduites et transformées (transformation log₁₀).

Les figures obtenues montrent la présence de deux échantillons extrêmes « outliers », les mêmes dans les deux conditions analytiques (échantillons NB8.270 et NB6.2520). Ces deux échantillons sont très différents en termes de temps de prélèvements (270 *versus* 2520 minutes), ne correspondent pas à des temps de prélèvements extrêmes (temps variant entre 90 et 4740 minutes pour l’ensemble de la cohorte), n’appartiennent pas au même nouveau-né, et ne représentent pas sexe en particulier (excrétés par deux garçons, qui représentent la population majoritaire de notre cohorte, 9/11 des nouveau-nés). Ces échantillons n’étaient pas proches dans la séquence analytique, éliminant la possibilité d’un biais analytique au moment de l’acquisition des données. Les autres échantillons collectés de ces deux mêmes nouveau-nés sont regroupés avec les autres échantillons analysés. Les deux échantillons NB8.270 et NB6.2520 présentent soit une composition réellement différente à ces temps de collectes, ou alors nous supposons qu’une contamination de ces échantillons est possible (par de l’urine ou lors de leur récupération sur la couche), un problème lors de leur conservation avant aliquotage (moins probable car beaucoup d’échantillons partageaient les mêmes conditions indiquées en partie.II.A), et/ou lors de la préparation pour l’analyse en LC/MS. Etant donné l’anonymat de cette collecte, nous n’avons pas d’autres informations qui permettraient d’expliquer la particularité de ces deux échantillons.

Etant donné le faible nombre d’échantillons disponibles, l’exclusion de ces échantillons des analyses fait ressortir deux autres échantillons extrêmes (le NB9.195 et NB3.4740). Nous avons donc décidé de

garder ces deux échantillons dans notre jeu de données, tout en gardant un regard critique dans les analyses qui suivent.

Afin de faciliter la visualisation de la distribution de nos échantillons au sein de l'ACP, nous avons coloré les échantillons en fonction du temps de prélèvement et en fonction du nouveau-né (Figure 52). La coloration des échantillons est indiquée en bleu pour les temps de prélèvements les plus précoces, et évolue jusqu'au rouge pour les temps les plus tardifs (Figure 52.A et Figure 52.B pour les conditions HILIC-ESI⁻ et C18-ESI⁺ respectivement). Ces représentations montrent que le facteur temps semble être très structurant pour le jeu de données, ceci dans les deux conditions analytiques, avec un regroupement des échantillons collectés à des temps proches. De plus, cette représentation met en évidence une évolution non-linéaire de la composition globale du méconium en métabolites en considérant ces 2 composantes principales, avec une évolution suivant une courbe en forme de « C ». Cette évolution non linéaire peut traduire la complexité de l'évolution de la composition du méconium au cours des premiers jours de vie, avec la dégradation de certains métabolites et l'apparition de nouveaux. Ce même constat a été décrit dans de nombreux articles, qui ont relié cette évolution à des processus multiples tels que la mise en place du microbiote intestinal de l'enfant impactant la composition métabolique dans l'intestin [264], le début de l'alimentation (allaitement maternel, formule infantile) [256]... etc. Nous analyserons l'évolution de cette composition plus en détail dans la suite et discuterons les résultats observés.

La coloration des échantillons de l'ACP selon le nouveau-né (Figure 52.C et Figure 52.D pour les conditions HILIC-ESI⁻ et C18-ESI⁺ respectivement), démontre un moindre regroupement des échantillons selon ce paramètre. En effet, la distribution semble être plus impactée par le temps de prélèvement que par l'appartenance au même nouveau-né.

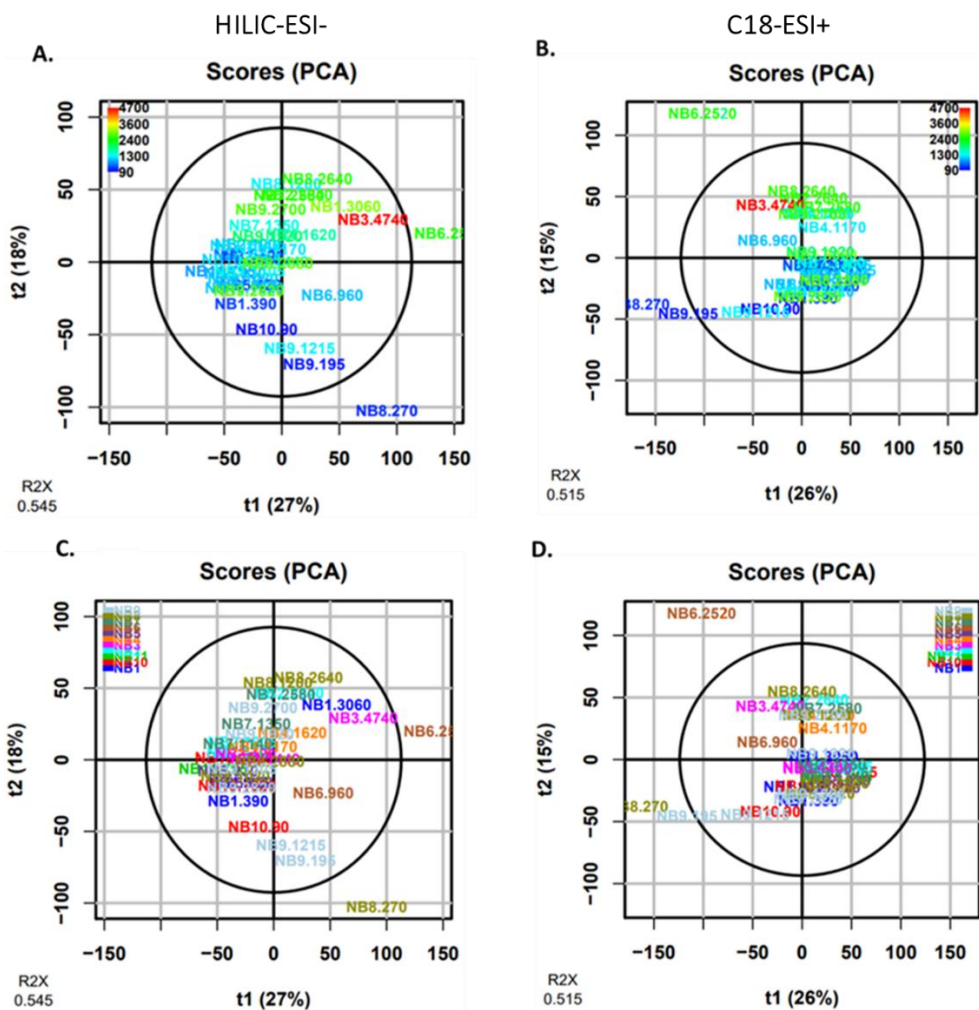


Figure 52 : Représentation des deux premières composantes de l'ACP réalisée sur l'ensemble des variables extraites et filtrées, correspondant respectivement à 6843 et 8555 variables dans les conditions analytiques HILIC-ESI⁻ (A, C) et C18-ESI⁺ (B, D). A et B : coloration en fonction des temps de prélèvements. Différents codes couleurs sont utilisés, correspondant à des temps de prélèvements continus représentés en haut à gauche de chaque figure (exprimé en minutes). C et D : coloration des échantillons en fonction du nouveau-né (de NB1 à NB11).

a.2) Structuration des échantillons et des variables par une clusterisation hiérarchique

Afin d'avoir une visualisation globale du regroupement des échantillons et des variables, nous avons réalisé une classification hiérarchique non supervisée. Cette analyse a été réalisée avec l'ensemble des échantillons et des variables filtrées dans les deux conditions analytiques HILIC-ESI⁻ et C18-ESI⁺ (respectivement Figure 53.A et Figure 53.B). Cette représentation permet d'avoir une visualisation de la proximité entre les échantillons (en colonnes) et entre les variables (en lignes), ceux-ci étant regroupés en « clusters » formés sans *a priori*. La hauteur des dendrogrammes reflète la distance des différents clusters et sous-clusters identifiés les uns par rapport aux autres. A travers ces

représentations, nous pouvons directement visualiser si une variable est plus ou moins intense dans un échantillon donné (en comparaison à la moyenne de l'intensité de cette variable dans l'ensemble des échantillons).

Ces représentations permettent de mettre en avant la proximité d'échantillons appartenant à des temps de prélèvements proches tels que NB10.90, NB9.195 et NB8.270 (cadres des échantillons en vert), représentant les plus faibles temps de prélèvements de notre cohorte. Ces échantillons sont caractérisés par de plus faibles intensités de nombreuses variables, notamment dans les conditions analytiques C18-ESI⁺. Nous observons dans d'autres cas, des échantillons regroupés appartenant au au même nouveau-né, tels que NB4, NB7 et NB9 (ces deux derniers sont encadrés en rouge et en bleu respectivement), et ce malgré des temps de prélèvements relativement éloignés. Nous observons notamment au sein de ces échantillons un profil métabolique assez similaire au sein des 3 échantillons du nouveau-né NB7 (encadrés en rouge), au cours des temps de prélèvements variant entre 1140 et 2580 minutes post-naissance. A l'inverse, nous observons dans les conditions analytiques HILIC-ESI⁻ pour les échantillons du nouveau-né NB9, une modification progressive de la composition du méconium au cours du temps (entre 195 et 2820 minutes post-naissance), avec un groupe de variables présentant une faible intensité dans les temps précoces puis augmentant au cours du temps (cluster des variables en bas du graphe). A l'inverse des variables présentant de fortes intensités dans les temps précoces qui se normalisent (cluster des variables en haut du graphe).

Les échantillons extrêmes identifiés précédemment en ACP (NB8.270 et NB6.2520), sont marqués par une étoile rouge et montrent un profil similaire, notamment dans les conditions HILIC-ESI⁻ (Figure 53.A). Un nombre important de variables sont présentes en très faibles intensités au sein de ces deux échantillons comparativement au reste. Cette similarité entre ces deux échantillons est cependant moins flagrante dans les conditions analytiques C18-ESI⁺ (Figure 53.B), où le nombre de variables de faibles intensités semble moins important. Il se peut que cette observation puisse expliquer en partie la distribution de ces échantillons, par notamment la présence de composés polaires pouvant traduire une contamination urinaire.

Afin de permettre de mieux visualiser et interpréter ces différences/similarités de compositions au sein des différents échantillons, nous allons travailler dans la suite sur les variables identifiées.

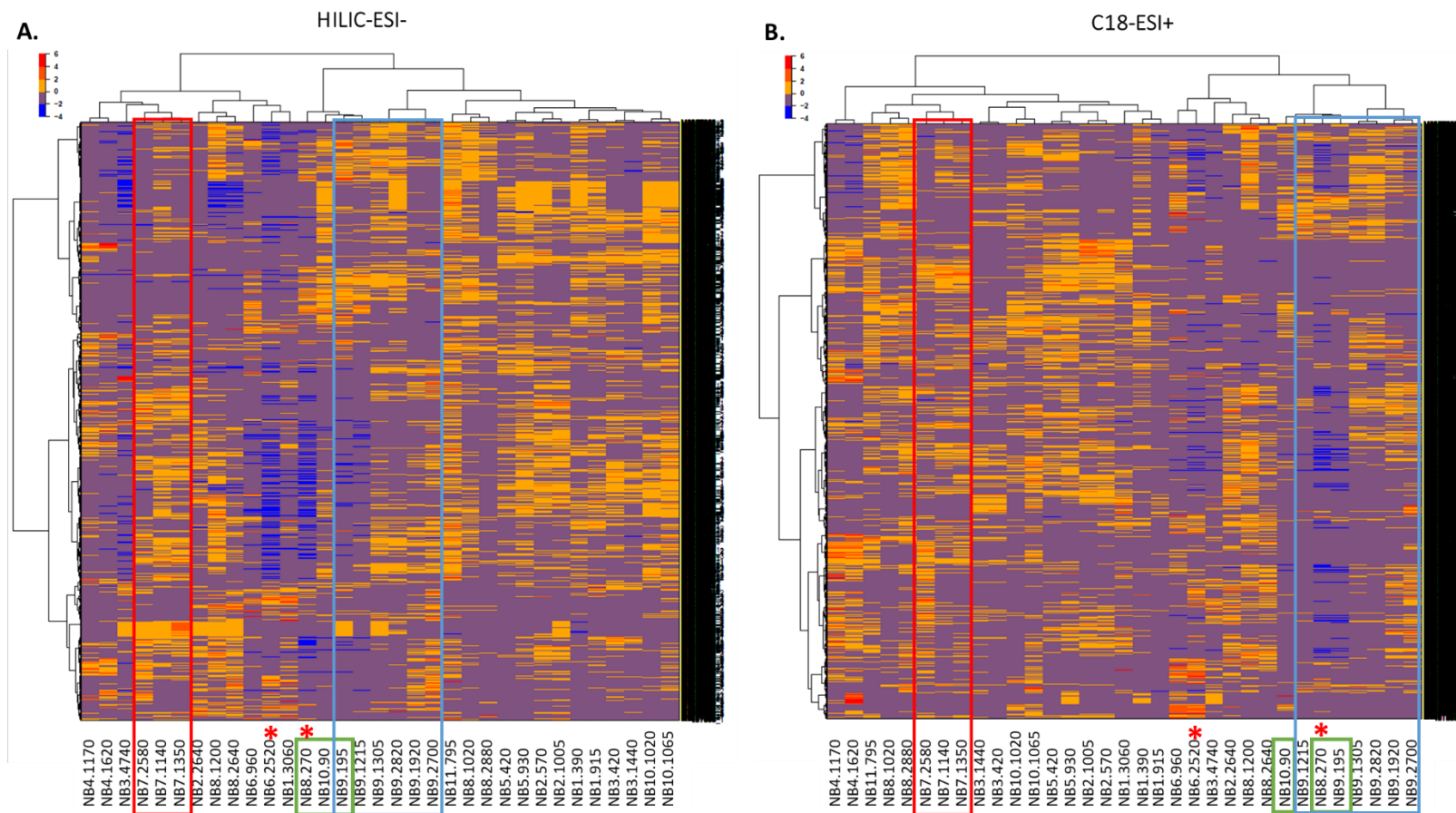


Figure 53 : Classification hiérarchique non supervisée des échantillons et des variables obtenues dans les conditions HILIC-ESI⁻ (A) et C18-ESI⁺ (B) (6843 et 8555 respectivement). Les échantillons de méconiums sont représentés en colonnes et les variables en lignes. La coloration indique si une variable est sous-représentée (du violet au bleu) ou sur-représentée (de l'orange au rouge) dans un échantillon donné, en comparaison à la moyenne de l'intensité de cette variable dans l'ensemble des échantillons. La couleur jaune indique des variables proches de la moyenne. Les valeurs des intensités des variables ont été au préalable centrées et réduites. Des exemples de la proximité d'échantillons appartenant à des temps de prélèvements proches sont encadrés en verts, et d'échantillons appartenant au même nouveau-né sont encadrés en rouge pour le NB7 et en bleu pour le NB9. Les échantillons extrêmes précédemment décrits sont marqués par une étoile rouge.

b) Analyses non supervisées réalisées sur les variables annotées

Nous avons réalisé les mêmes analyses non supervisées que précédemment, cette fois en ne considérant que les variables uniques annotées.

b.1) Analyses en composantes principales

Les représentations d'ACP non colorée (Figure 54.A), colorée en fonction du temps de prélèvement (Figure 54.B), ou du nouveau-né (Figure 54.C) ont été réalisées sur les 230 métabolites uniques. Les deux premières composantes (t1 et t2) de l'ACP représentent 40 % de la variance totale du jeu de données. La troisième dimension résume 7 % de la variance résiduelle (celle-ci ne sera pas considérée dans la suite, n'apportant pas d'informations supplémentaires).

Il semble que les variables annotées reproduisent globalement la structuration du jeu de données total. En effet, nous observons notamment les deux échantillons NB8.270 et NB6.2520 qui apparaissent comme extrêmes comme précédemment, ainsi qu'une évolution non linéaire de la composition globale des échantillons au cours des temps de collecte, suivant une courbe en forme de « C » (Figure 54.A et Figure 54.B). Les variables annotées, résumerait donc relativement bien l'ensemble de l'information.

Comme observé lors des ACP réalisées sur l'ensemble des variables extraites et filtrées, l'effet nouveau-né semble être moins structurant, avec peu de regroupements apparents des échantillons appartenant au même nouveau-né (Figure 54.C).

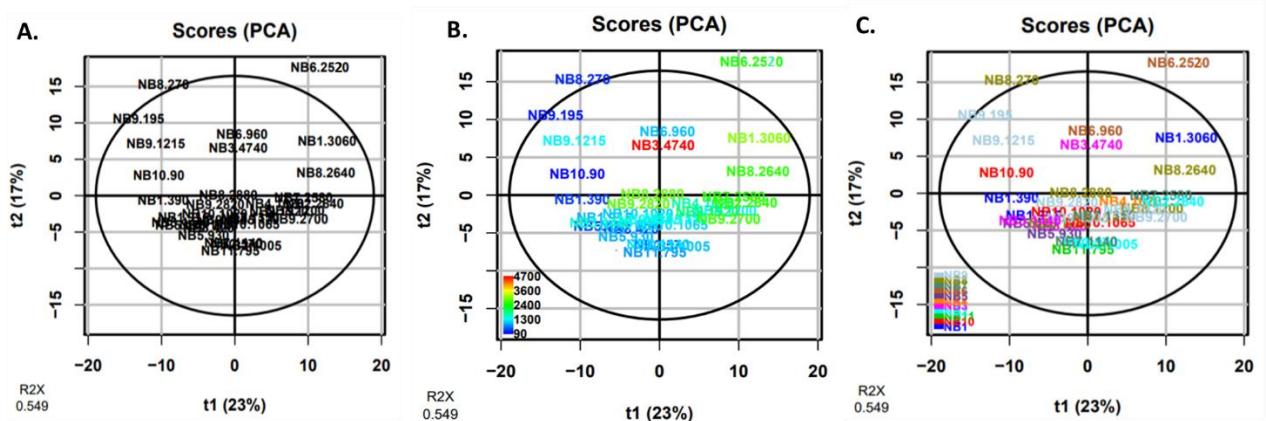


Figure 54 : A-C : Représentation des deux premières composantes de l'ACP réalisée sur les 230 variables uniques dont l'annotation a été confirmée par MS/MS . B. Coloration réalisée en fonction du temps de prélèvement (considéré comme une variable continue, code couleur en bas à gauche de la figure), et C. Coloration en fonction des individus (de NB1 à NB11, code couleur en bas à gauche de la figure).

b.2) Classification hiérarchique non supervisée des échantillons et des métabolites annotés

La visualisation globale du regroupement des échantillons et des variables annotées est représentée en Figure 55. Nous avons appelé les clusters (formés sans *a priori*) : 1 et 2 pour les échantillons, et A et B pour les métabolites annotés. Les clusters d'échantillons 1 et 2 sont respectivement formés de 22 et de 11 échantillons. Le cluster 1 comprend des échantillons collectés à des temps précoces tandis que le cluster 2 est plus représentatif d'échantillons collectés à des temps plus tardifs (> 960 minutes). Ce constat n'est néanmoins pas absolu (certains temps de collecte tardifs sont dans le cluster 1).

Nous observons à partir de cette représentation une évolution flagrante au cours du temps au sein des différents prélèvements appartenant à un même nouveau-né. Ceci est le cas par exemple de NB1, avec un passage du cluster 1 (NB1.390 et NB1.915) au cluster 2 (NB1.3060) (encadrés en vert dans la figure). Pour ces échantillons, certaines variables voient leurs intensités augmenter au cours des temps de prélèvements (cluster B), tandis que d'autres diminuent (cluster A). Ces différentes constatations peuvent être expliquées par : i) la dégradation des métabolites accumulés au cours de la vie *in utero* par la mise en route des différents processus métaboliques du nouveau-né et de son microbiote [256], ii) l'apport de l'alimentation.

En fixant un temps de prélèvement « seuil » de 1440 minutes (24h), nous retrouvons 90 % des échantillons prélevés aux temps précoces dans le cluster 1 (20/22, exceptés NB9.2820 et NB8.2880, marqués par une étoile verte sur la figure), et plus de 80 % des échantillons tardifs dans le cluster 2 (9/11 exceptés NB6.960 et NB4.1170, marqués par une étoile rouge sur la figure). Néanmoins, tous les échantillons collectés aux temps tardifs (cluster 2) ne présentent pas le même profil. Nous observons notamment pour le NB3.4740 (encadré en rouge), représentant le temps le plus tardif de notre collecte, des intensités beaucoup moins fortes que dans les autres échantillons collectés au cours du jour 2 ou du jour 3 post-naissance. Néanmoins, ceci semble être spécifique au nouveau-né NB3 qui présente de plus faibles intensités de variables même à des temps précoces (NB3.420 et NB3.1440 – à l'extrême droite du cluster 1), comparés aux prélèvements aux temps proches. Nous approfondirons l'analyse en fonction de ce seuil de 1440 minutes par la suite.

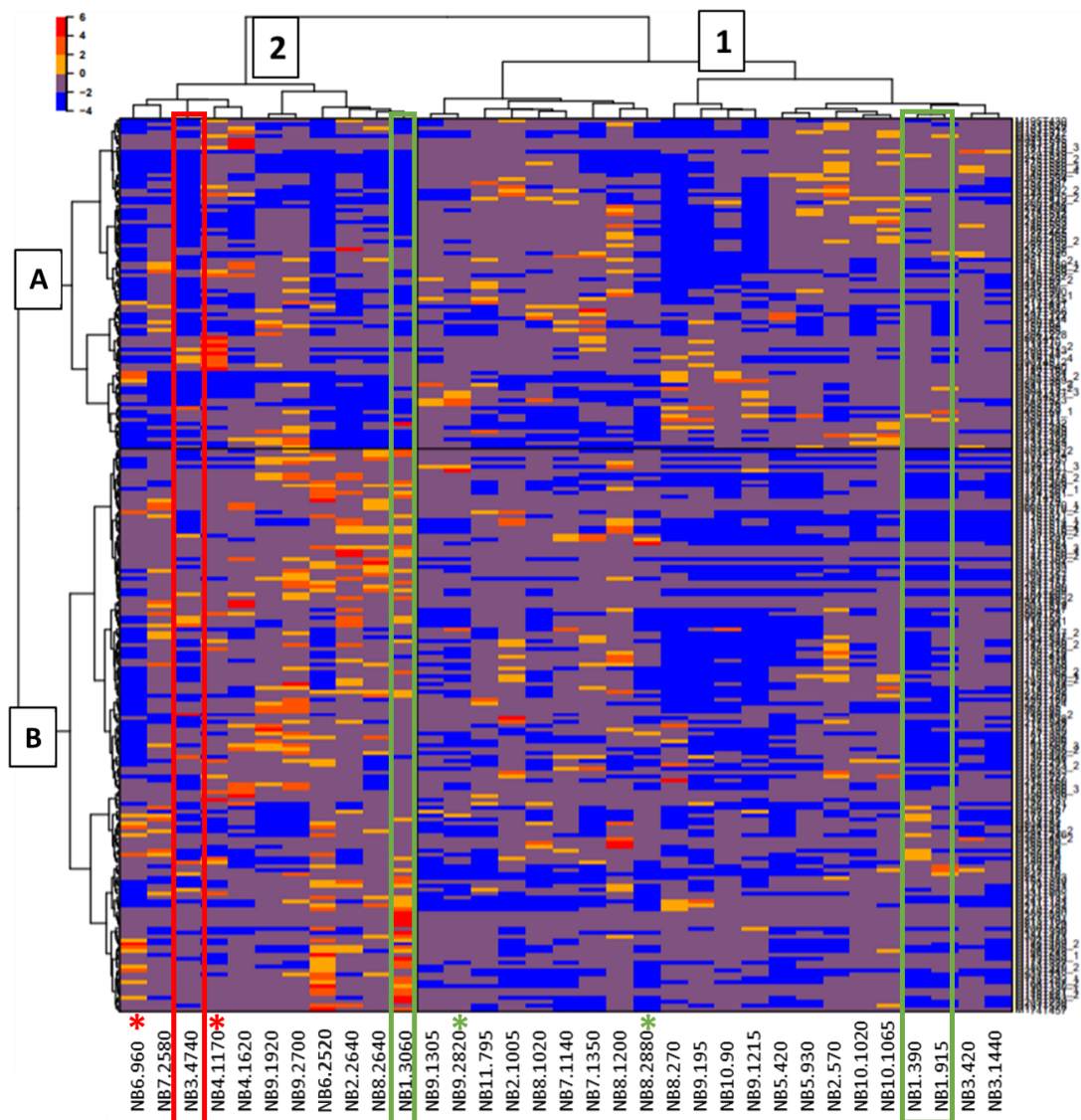


Figure 55 : Classification hiérarchique non supervisée des échantillons et des 230 métabolites annotées dans les conditions HILIC-ESI⁻ et C18-ESI⁺. Les échantillons sont représentés en colonnes (clusters 1 et 2) et les variables en lignes (clusters A et B). La coloration indique si une variable est sous-représentée (du violet au bleu) ou sur-représentée (de l'orange au rouge) dans un échantillon donné, en comparaison à la moyenne de l'intensité de cette variable dans l'ensemble des échantillons. La couleur jaune indique des variables proches de la moyenne. Les valeurs des intensités des variables ont été préalablement centrées et réduites. Les étoiles indiquent les deux échantillons qui n'appartiennent pas au seuil de 1440 minutes fixés au sein de chacun des clusters (rouge, cluster temps > 1440 minutes et en vert, temps < 1440 minutes). Les cadres (rouge, représente l'échantillon au temps de collecte le plus tardif : 4740 minutes, et en vert les échantillons appartenant au nouveau-né NB1).

b.3) Séparation des échantillons de méconium excrétés avant et après 24 heures (1440 minutes)

A la vue du regroupement des échantillons en fonction du temps de prélèvement, et notamment du seuil de 1440 min qui semble être structurant, le graphe des échantillons de l'ACP réalisée avec les 230 variables annotées, a été coloré en fonction de l'appartenance des échantillons aux deux groupes :

groupe 1 (< 1440 minutes, en bleu) et groupe 2 (> 1440 minutes, en rouge). Nous observons à partir de cette représentation, une séparation nette des deux groupes considérés, confirmant l'effet structurant du facteur « temps de prélèvement » sur la distribution du jeu de données (Figure 56).

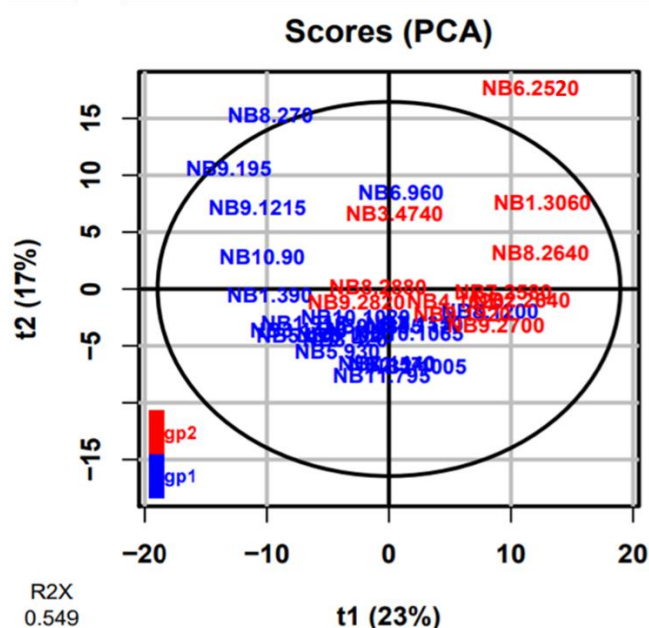


Figure 56 : Représentation des deux premières composantes de l'analyse en composantes principales réalisée sur les 230 variables annotées en HILIC-ESI⁺ et en C18-ESI⁺ colorée selon le temps de prélèvement des échantillons. Les temps précoces (< 1440 minutes, groupe 1) sont représentés en bleu et les temps tardifs (> 1440 min, groupe 2) sont représentés en rouge.

A la suite de ces résultats, nous avons réalisé des analyses univariées et multivariées supervisées, en considérant ces deux groupes d'échantillons, permettant : i) d'identifier les variables discriminant ces deux groupes, ii) de décrire les métabolites spécifiques des temps précoces vs des temps tardifs, et iii) de suivre l'évolution de ces métabolites au cours des trois premiers jours de vie.

c) Analyse des métabolites annotés selon le temps de prélèvement des échantillons

c.1) Analyses univariées

Nous avons dans un premier temps réalisé une analyse univariée de type Wilcoxon sur les 230 variables annotées, en utilisant une correction multiple de type FDR (de l'anglais, False Discovery Rate). Cette analyse a été réalisée entre les temps de prélèvements (gp1 : temps < 1440 minutes et gp2 : temps > 1440 minutes). Au total, 67 métabolites annotés présentent des intensités significativement différentes entre les deux groupes (p-values corrigées < 0,05, table en Annexe 12). Parmi ces 67 métabolites, 50 ont des intensités significativement plus élevées dans les échantillons collectés aux temps tardifs,

incluant par exemple le N-acétylglycine, l'acide L-thréonique ou encore le méthylguanine (Figure 57.A). Les 17 autres métabolites sont significativement plus intenses dans les échantillons collectés aux temps précoces, tels que le glycerol-phosphate, la xanthoptérine ou encore l'acide N-acétylneuraminique (Figure 57.B).

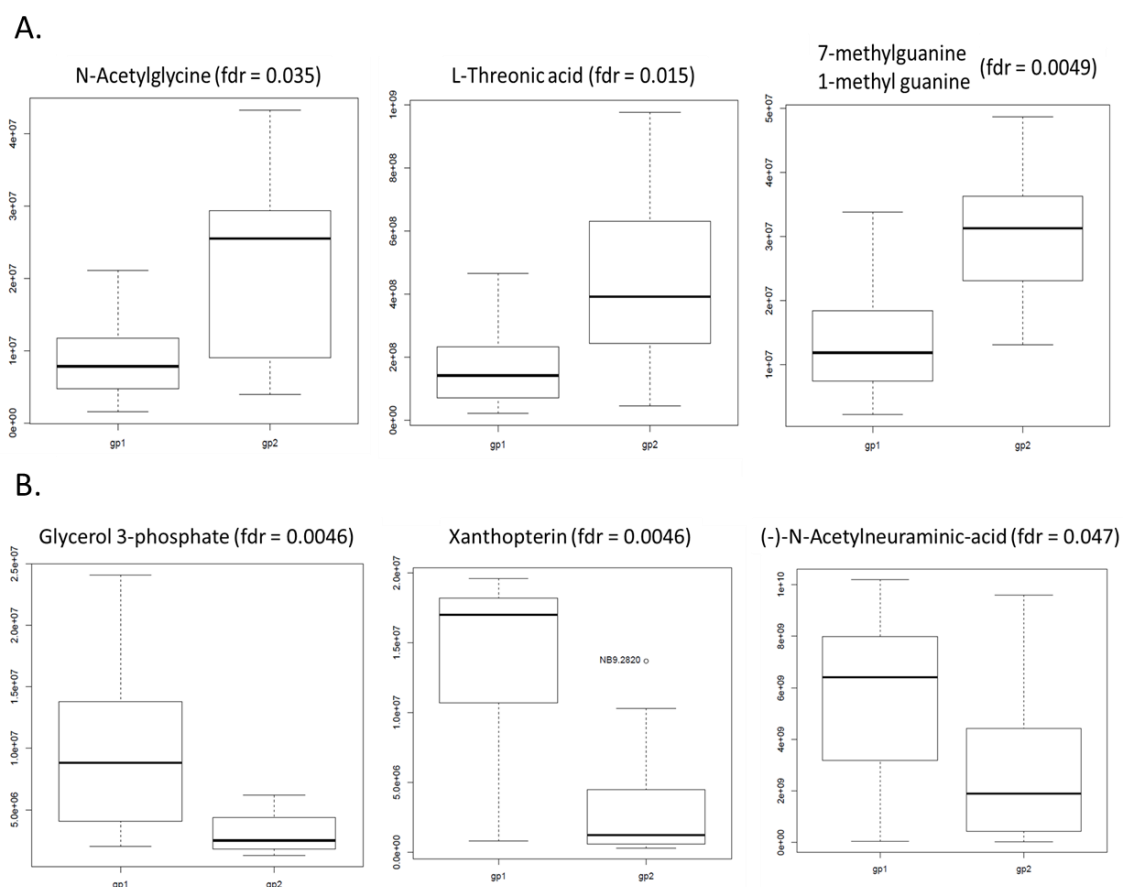


Figure 57 : Représentation en boîtes à moustaches de quelques exemples de variables significativement différentes entre les groupes 1 et 2. A. Métabolites « représentatifs » des temps tardifs (gp2). B. Métabolites « représentatifs » des temps précoces (gp1).

c.2) Analyses multivariées supervisées de type PLS-DA (Partial Least Square – Discriminant Analysis)

A la suite de ces résultats, nous avons réalisé une analyse multivariée supervisée de type PLS-DA sur ces deux groupes d'échantillons et sur les 230 variables annotées. Le graphe PLS-DA n'a pas été réalisé à travers l'analyse sur W4M, du fait de la présence d'une seule composante. Le graphe de pR2Y et de pQ2 (respectivement = 0,05) montre la construction d'un modèle juste, nous permettant d'interpréter les valeurs de VIP générées par ce dernier (Variable Influence on Projection). L'analyse de ces VIP permettrait d'avoir une idée supplémentaire (en plus de l'analyse univariée réalisée), sur les variables décrivant les temps précoces ou les temps tardifs.

Selon les références, la contribution d'une variable dans la construction d'un modèle est considérée comme significative, si la valeur de VIP est supérieure ou égale à 0,8 [211] à 1 [247], [292] à 1,5 [293] ou encore à 2 [294]. Nous avons fait le choix pour cette analyse de considérer une valeur de VIP $\geq 1,5$, permettant de définir les variables les plus discriminantes (sur les 230 annotées), entre les groupes d'échantillons précoces et tardifs. Nous définissons à travers cette analyse, 34 variables discriminantes entre les deux groupes d'échantillons analysés. L'ensemble de ces 34 variables ont montré des différences significatives par les analyses univariées réalisées précédemment entre ces mêmes groupes.

La Figure 58 décrit l'ensemble des valeurs de VIP obtenues sur les 230 variables annotées par l'analyse multivariée supervisée (PLS-DA), représentées en fonction des p-values obtenues pour ces mêmes variables à travers un test univarié de Wilcoxon. Les seuils fixés pour les p-values ($p < 0,05$) et les VIP ($VIP > 1,5$) sont représentés par deux traits noirs (respectivement vertical et horizontal). Les variables représentées en vert comptent les 34 variables statistiquement pertinentes décrites à travers les deux analyses uni et multivariée. Les variables représentées en rouge se rajoutent à celles-ci pour définir l'ensemble des variables significatives obtenues par les analyses univariées (soit 67 variables).

Les métabolites discriminant significativement nos deux groupes d'échantillons obtenus à travers les deux analyses réalisées sont décrits dans le

Tableau 10 10. Ces métabolites appartiennent à différentes familles chimiques (aminoacides, lipides, carbohydrates...), incluant la proline, la valine, des acides tels que l'acide cholique ou caprique, l'hexose-phosphate ou encore la taurine. Parmi les voies impliquées, nous retrouvons des voies métaboliques spécifiques telles que le métabolisme de l'ascorbate et l'aldarate en premier lieu (impliquant l'acide glucuronique gamma-lactone et l'acide glucuronique) ou des voies plus générales telles que le métabolisme de l'alanine, du galactose et de la biosynthèse du tryptophane. Nous observons que l'abondance de la plupart des métabolites augmente au cours du temps, avec 28/34 métabolites présentant une intensité plus forte dans les échantillons collectés après 1440 minutes. Ces résultats rejoignent ce qui a précédemment été décrit dans la littérature sur l'évolution de la composition métabolique dans le méconium au cours des premiers jours de vie [115], [256], [264].

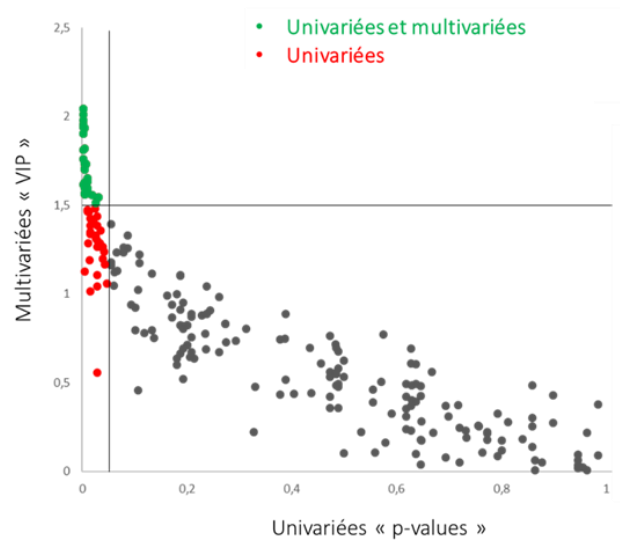


Figure 58 : Combinaison des résultats obtenus par les analyses univariées (test de Wilcoxon) et multivariées (PLS-DA) réalisées sur l'ensemble des 230 variables identifiées. Les seuils de p-value = 0,05 et de VIP = 1,5 sont représentés par deux traits noirs vertical et horizontal respectivement. Les variables présentant une différence significative par les tests univariés entre les groupes « temps précoces » (< 1440 minutes) et « temps tardifs » (> 1440 minutes) regroupent les points représentés en vert et en rouge. Les variables représentées en vert sont également identifiées comme discriminantes dans l'analyse supervisées de type PLS-DA (VIP > 1,5). Le restant des variables en gris, ne présentant pas de caractère significativement différent entre les groupes analysés.

Tableau 10 : Liste des 34 variables présentant une différence significative de leur intensité ($p < 0,05$, Test Wilcoxon) et identifiées comme discriminantes (VIP > 1,5 ; PLS-DA) entre les groupes d'échantillons collectés aux temps précoces ou aux temps tardifs (seuil de 1440 minutes).

Familles Chimiques	Composés	Origines	VIP_PLS-DA	pvalues_test Wilcoxon	Plus fortes intensités
Composés hétéromonocycliques aliphatiques	L-Pyroglutamic-acid	Endogène	1,5963	0,0107	> 1440 min
Aminoacides, peptides et analogues	N acetyl L glutamine	Endogène	1,9049	0,0018	> 1440 min
	N-Acetyl-L-alanine	Endogène	1,7612	0,0024	> 1440 min
	N-Acetyl-L-aspartic acid	Endogène	1,8118	0,0022	> 1440 min
	O-Acetyl-L-serine/N-Acetyl-DL-serine	Endogène	1,6312	0,0106	> 1440 min
	Orotic acid	Endogène	1,6518	0,0106	> 1440 min
	Proline	Endogène	1,565	0,0094	> 1440 min
	Sarcosine	Endogène	1,5987	0,0047	> 1440 min
	trans-4-hydroxy-L-proline	Endogène	1,585	0,0076	> 1440 min
Composés hétéropolycycliques aromatiques	7-methylguanine	Endogène	1,5603	0,0049	> 1440 min
	Adenine	Endogène	1,954	0,0024	> 1440 min
	Indolelactic acid	Endogène	1,6112	0,0049	> 1440 min
Composés homomonocycliques aromatiques	Acetyl-N-formyl-5-methoxykynurenamine/Phenylacetyl-L-glutamine	Endogène	1,5679	0,0046	> 1440 min
	Homovanillic acid	Endogène	1,9367	0,0046	> 1440 min
Carbohydrates et conjugués	D-Galacturonic acid	Endogène	1,6292	0,0047	< 1440 min
	D-Glucuronic acid	Endogène	1,6292	0,0047	< 1440 min
	D-Glyceric acid	Endogène	1,9051	0,0018	> 1440 min
	D-Ribose 5-phosphate	Endogène	1,6016	0,0052	< 1440 min
	D-Threitol	Endogène	1,9811	0,0024	> 1440 min
	Glycerol 3-phosphate	Endogène	1,7418	0,0046	< 1440 min
	Hexose-6-phosphate	Endogène	1,7009	0,0046	> 1440 min
	Maltotetraose	Endogène	2,0446	0,0018	> 1440 min
	Stachyose	Endogène	2,0446	0,0018	> 1440 min
Lipides	2-Hydroxy 3-methylbutyric acid (α -Hydroxyisovaleric acid)	Endogène	1,7068	0,0054	> 1440 min
	Capric acid	Endogène	2,0119	0,0028	> 1440 min
	Cholic acid	Endogène	1,6182	0,0018	> 1440 min
	Ethylmethylacetic acid	Endogène	1,5141	0,0257	> 1440 min
Acides organiques et dérivés	(S)-3-Hydroxybutyric acid	Endogène	1,5551	0,0194	> 1440 min
	2-Isopropylmalic acid	Microbien	1,5744	0,0106	< 1440 min
	Azelaic acid	Endogène	1,8193	0,0044	> 1440 min
	L-(+)-Tartaric acid	Alimentaire	1,5433	0,031	< 1440 min
	Methyl-succinic acid/mono-Methyl hydrogen succinate	Endogène	1,7327	0,0072	> 1440 min
	Mevalonic acid	Endogène	1,5141	0,0257	> 1440 min
	Taurine	Endogène	1,9384	0,0018	> 1440 min

Les intensités des 163 autres métabolites annotés sont comparables entre les deux groupes d'échantillons de méconium, c'est-à-dire entre des échantillons prélevés avant et après 1440 minutes (Tableau 11). Ces 163 métabolites pourraient représenter le « noyau » métabolique du méconium, avec des moyennes d'intensités stables entre les deux groupes de temps considérés, même si des variabilités entre individus peuvent être observés. Ces 163 métabolites appartiennent à différentes familles de composés chimiques. Ils incluent aussi bien des métabolites endogènes que exogènes issus de l'alimentation (ou de la supplémentation alimentaire) de la mère (telle que la présence de l'acide nicotinique ou de l'acide quinique), de l'activité de ses microbiotes (présence de l'acide muramique, d'acide biliaire secondaire : acide lithocolique-3-sulfate, ou des dérivés du tryptophane : acide kynurénique), ou encore de son exposition durant la grossesse à des médicaments (la présence de salicylamide ou d'acétaminophène). Parmi les voies métaboliques retrouvées, nous décrivons le métabolisme de l'arginine, de la proline et du tryptophane, le métabolisme des acides biliaires et de la purine...

Les résultats présentés ci-dessus, obtenus par différentes analyses statistiques, ont permis : i) de visualiser un regroupement des échantillons par temps de prélèvements (< ou > 1440 minutes), à travers des analyses non supervisées (de type ACP ainsi que par clusterisation hiérarchique), ii) d'identifier 67 variables significativement différentes entre les groupes d'échantillons prélevés précocément (< 1440 minutes) et plus tardivement (> 1440 minutes).

Ces résultats seront plus amplement discutés à la fin de ce chapitre, en regroupant les autres données « omiques » réalisées sur cette même cohorte d'échantillons, incluant la métagénomique ciblée et la métabolomique, que nous décrivons ci-après.

Tableau 11 : Liste des 163 métabolites non significativement différents entre les temps de prélèvements précoces et tardifs (seuil de temps 1440 minutes).

Familles	Composés	Origines
Composés acycliques aliphatiques	Carnitine	Endogène
	Pantothenic acid	Endogène
	Agmatine	Endogène
	Cadaverin	Endogène
Composés hétéromonocycliques aliphatiques	Allantoin	Endogène
	Creatinine	Endogène
	L-Threonic-acid-1,4-lactone	Endogène
	Stachydrine	Endogène
	Myo-inositol	Endogène
	D-(-)-Quinic acid	Alimentaire
	Shikimic acid	Endogène
Aminoacides, peptides et analogues	L-Glutamic acid	Endogène
	Alanine	Endogène
	L-Glutamine	Endogène
	L-Aspartic-acid	Endogène
	Asparagine	Endogène
	Serine	Endogène
	Histidine	Aminoacide essentiel
	L-Citrulline	Endogène
	Creatine	Endogène
	Ornithine	Endogène
	Arginine	Aminoacide essentiel
	S-Adenosyl-homocysteine	Endogène
	Methionine	Aminoacide essentiel
	Lysine	Aminoacide essentiel
	1-Methyluric acid	Endogène
	N6,N6,N6-Trimethyl-L-lysine	Endogène
	Tyrosine	Endogène
	5-Aminolevulinic acid	Endogène
	Alpha-D-Aminoadipic acid	Endogène
	L-Saccharopine	Endogène
	L-Homoserine	Endogène
	alpha-Hydroxyhippuric-acid	Endogène
	Aspartylglycosamine (N-As)	Endogène
	DL-Tryptophan	Endogène
	Glycyl-L-leucine	Endogène
	Gly-Pro	Endogène
	Guanidinosuccinic acid	Endogène
	Hydroxykynurenine	Endogène
	L-Cysteine S-sulfate	Endogène
	L- α -aminobutyric acid	Endogène
	N6-Acetyl-L-lysine	Endogène
	N-Acetyl-Asp-Glu	Endogène

	N-Acetyl-DL-serine	Endogène
	N-acetyl-DL-tryptophan	Endogène
	N-Acetyl-L-carnosine	Endogène
	Nicotinuric acid	Endogène
	N-Methyl-D-aspartic acid (NMDA)	Endogène
	N- α -acetyl-L-arginine	Endogène
	N α -Acetyl-L-asparagine	Endogène
	Phenylacetyl-L-glutamine	Endogène
	S-ethyl-L-cysteine	Endogène
	Quinolinic acid	Endogène
	5-Aminovaleric acid	Endogène
	Isoleucine	Aminoacide essentiel
	Kynurenic acid	Endogène
	L-Kynurenine	Endogène
	Xanthurenic-acid	Endogène
Composés hétéromonocycliques aromatiques	4-Imidazoleacrylic-acid-(urocanic-acid)	Endogène
	Uracil	Endogène
	Imidazoleacetic acid	Endogène
	2,5-Furandicarboxylic acid	Endogène
	6-Hydroxypyridine-3-carboxylic acid (6-hydroxynicotinic acid)	Endogène
	Cytosine	Endogène
	Imidazolelactic acid	Endogène
	Nicotinamide N-oxide	Endogène
	1-Methylnicotinamide	Médicament
	Nicotinic acid	Alimentaire
	4-Pyridoxic acid	Endogène
	Pyridoxine	Endogène
Composés hétéropolycycliques aromatiques	7-Methylxanthine	Endogène
	Caffeine	Alimentaire
	Theobromine	Endogène
	(\pm)-Salsolinol	Endogène
	1,3-Dimethyluric acid	Endogène
	1,7-Dimethyluric acid	Endogène
	3,7-Dimethyluric acid	Endogène
	Equol (phytoestrogen)	Endogène
	Indole-3-carboxylic acid	Alimentaire
	Indoxyl sulfate	Endogène
	paraxanthine	Endogène
	D-(+)-Neopterin	Endogène
	Hypoxanthine	Endogène
	Xanthine	Endogène
	Indoleacetic acid	Endogène
	4-Hydroxy-3-methoxyphenylglycol sulfate	Endogène
4-Hydroxymandelic acid	Endogène	

	Acetaminophen-(4-Acetamidophenol)	Médicament
	Benzoic acid	Alimentaire
	Eugenol (pdt cosmétique)	Cosmétique
	Salicylamide	Médicament
	DL-vanillomandelic acid	Endogène
Benzène et dérivés	Hydroxynicotinic acid	Endogène
Carbohydrates et conjugués	Dehydroascorbic acid	Endogène
	L-Fucose	Endogène
	Maltotriose	Endogène
	Melibiose	Microbial
	Hexose	Endogène
	Acetaminophen-glucuronide	Médicament
	beta-Pseudouridine	Endogène
	D-(+)-Arabitol	Endogène
	D-(+)-Cellobiose	Microbial
	D-Xylulose	Endogène
	Galactonic acid	Endogène
	Gluconic acid	Endogène
	Lactulose	Médicament
	L-Rhamnose	Endogène
	Muramic acid	Microbial
	N-Acetylgalactosamine	Endogène
	D-(-)-Ribose	Endogène
	D-(+)-Gluconic acid D-lactone (Gluconolactone)	Endogène
	Amylose	Endogène
Acides carboxyliques et dérivés	Aspartyl-Phenylalanine	Alimentaire
	Leu-Pro	Endogène
Lipides	Androsterone sulphate	Endogène
	17a-Hydroxypregnenolone	Endogène
	Fumaric acid	Endogène
	Acetyl-L-carnitin	Endogène
	Glycocholic acid	Endogène
	glycodeoxycholate	Endogène
	Taurochenodeoxycholic acid	Endogène
	2-cis,4-trans-Abcsic acid	Endogène
	2-Hydroxyhexadecanoic acid (2-Hydroxypalmitic acid)	Endogène
	3-Methylglutaric acid	Endogène
	7 α -Hydroxytestosterone	NA
	Behenic acid	Endogène
	beta-estradiol-17beta-glucuronide	Endogène
	cis,cis-Muconic acid	Endogène
	d-Desthiobiotin	Endogène
	Dehydroisoandrosterone 3-sulfate	Endogène
	Deoxycorticosterone-21-glucoside	Endogène

	D-erythro-Dihydrosphingosine (Sphinganine)	Endogène
	Ethylmalonic acid	Endogène
	Estrone glucuronide	Endogène
	Lithocholic acid 3-sulfate	Endogène
	Maleic-acid	Endogène
	Medroxyprogesterone	Médicament
	Mesaconic acid/Citraconic acid	Endogène
	Nervonic acid	Endogène
	Pregnenolone sulfate	Endogène
	Tiglic acid	Endogène
	Tridecanoic acid	Endogène
	Val-Pro	Endogène
	Propionylcarnitine	Endogène
	4-Methyl-2-oxovaleric acid	Endogène
NA	5-Sulfosalicylic acid	Endogène
	Diglycolic acid	Endogène
Nucléosides, nucléotides et analogues	5-Deoxy-5-(methylthio)adenosine	Endogène
	2-O-Methylinosine	Endogène
	Orotidine	Endogène
	Guanosine	Endogène
	Deoxycytidine	Endogène
	Thymidine	Endogène
	Uridine	Endogène
Acides organiques et dérivés	α -Ketoglutaric acid	Endogène
	Malonic acid	Endogène
	Isocitric-acid	Endogène
	Malic acid	Endogène
	2,2'-Thiodiacetic acid	Endogène
	2-Oxovaleric acid	Endogène
	3-Hydroxy-3-methylglutaric acid	Endogène
	Adipic acid	Alimentaire
	D-2-Hydroxyglutaric acid	Endogène

B. Analyses protéomiques

1. Analyse des protéines d'origine humaine

a) Traitement des données et identification des protéines humaines

L'analyse des protéines humaines a été réalisée selon une méthodologie bien décrite par Grenga et al [283], sur 30 échantillons de méconium collectés longitudinalement auprès de 11 nouveau-nés (Tableau 7, décrit en Partie II.A). Le traitement des données générées est automatisé et est décrit en Partie II.C.4. Au sein de ces analyses, 1018819 de spectres MS/MS ont été générés, parmi lesquels 230065 PSMs ont été attribués (soit un ratio de 23 % de spectres assignés). L'ensemble de ces peptides ont permis d'identifier 1525 protéines humaines avec un peptide spécifique, dont 922 protéines sont validées par au moins deux peptides spécifiques. Dans leur étude, Lisowska et al ont décrit 946 protéines humaines identifiées avec 1 peptide spécifique et 430 protéines identifiées par au moins 2 peptides spécifiques dans le méconium [136]. Notre apport est donc très significatif par rapport à l'existant. La liste des 922 protéines humaines validées est en Annexe 13.

a.1) Evaluation du nombre de protéines détectées dans chacun des échantillons de méconiums analysés

Afin d'évaluer la variabilité interindividuelle au sein des méconiums analysés, nous avons représenté en Figure 59 le nombre de protéines détectées pour chacun des échantillons. Les échantillons appartenant au même nouveau-né sont représentés par la même couleur, et dans l'ordre de temps de prélèvement croissant. Nous observons un nombre de protéines identifiées moyen de 370, et une médiane de 340. Le nombre de protéines humaines identifiées varie entre 240 et 618, respectivement identifiées au sein des échantillons NB5.930 et NB6.2520 (marqués par une étoile rouge). L'échantillon NB6.2520 était un échantillon extrême en métabolomique, et il semble présenter également un profil spécifique en protéomique, avec la plus grande richesse protéique observée. Cette richesse semble être spécifique à ce nouveau-né (NB6), pour lequel un grand nombre de protéines est également détecté au temps précoce (NB6.960, 604 protéines). Ceci suggère que le nouveau-né NB6 a excrété des méconiums montrant une composition protéique (et métabolique) particulière, sans que nous puissions le relier à des facteurs particuliers étant donné que nous ne disposons pas de données sur les mères. L'échantillon NB5.420 présentant la plus faible richesse protéique ne démontrait cependant pas de profil spécifique en analyses métabolomiques. Comme indiqué précédemment, au vu de la modeste taille de notre cohorte d'étude, nous retiendrons l'ensemble des échantillons pour la suite des analyses (tout en tenant compte de leurs différences).

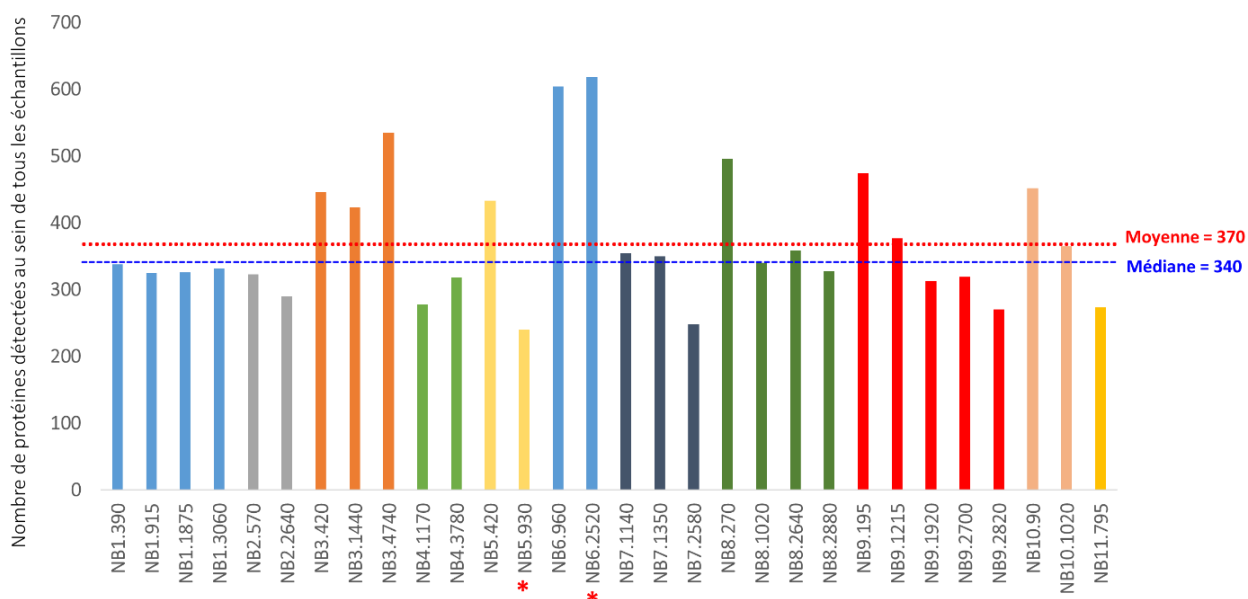


Figure 59 : Représentation du nombre de protéines humaines détectées au sein des échantillons de méconium analysés. Les échantillons appartenant au même nouveau-né sont représenté par la même couleur (de NB1 à NB11). Les pointillés rouges représentent la valeur moyenne du nombre de protéines détectées, et les tirets bleus représentent la médiane.

Mis à part le nouveau-né NB11.795 pour lequel un seul prélèvement est disponible, nous observons au sein de 60 % des nouveau-nés un nombre de protéines humaines décroissant au cours du temps (incluant : NB2, NB5, NB7, NB8, NB9 et NB10). Ces résultats sont similaires à ce qui a été décrit précédemment, indiquant une altération des profils protéiques au cours des temps de collectes d'échantillons de méconium [264]. Cependant, les échantillons les plus pauvres en protéines identifiées, ne sont pas toujours issus de prélèvements tardifs, comme observé pour le NB5.930 avec 240 protéines détectées. De la même façon, les échantillons les plus riches en protéines, ne sont pas toujours représentatifs de temps de prélèvements précoces, tel qu'observé pour le NB3.4740 avec 535 protéines identifiées. Pour les deux nouveau-nés NB1 et NB6, le nombre total de protéines humaines n'évolue pas au cours du temps, malgré une large période couverte pour le NB1 (4 prélèvements entre 390 et 3060 minutes). Pour NB3, cette quantité augmente au cours du temps. Nous reviendrons par la suite sur cette variabilité interindividuelle.

Nous avons dans premier temps investigué les similarités existantes au sein de nos échantillons, en décrivant les protéines les plus abondantes ainsi que le « core protéome » qui correspond aux protéines systématiquement détectées dans l'ensemble des échantillons de méconiums analysés.

a.2) Description des protéines les plus abondantes dans notre jeu de données

Nous décrivons en Figure 60, les 15 protéines les plus abondantes, tel qu'estimé par leur pourcentage de NSAF. Celles-ci représentent à elles seules 34 % de l'abondance totale de l'ensemble des protéines humaines détectées. Nous observons parmi les protéines les plus abondantes : des antiprotéases (telles que la α -1-antitrypsine, 7,5 % - protéine majoritaire, et la α -1-antichymotrypsine, 1,8 %), des calprotectines (telles que S100-A8/A9, 4,6 % chacune), des fragments d'immunoglobulines (régions constantes Ig kappa, 2,6 % et Ig lambda, 1,4%), des protéines nucléaires (telles que l'histone H4, 1,7 %), des protéines de structures tissulaires (telles que la kératine, 1,2 % pouvant traduire la présence de cheveux néonataux dans les échantillons), mais aussi des protéines de transport (telles que la transthyréline 1,2 % et la sérum albumine 1 %).

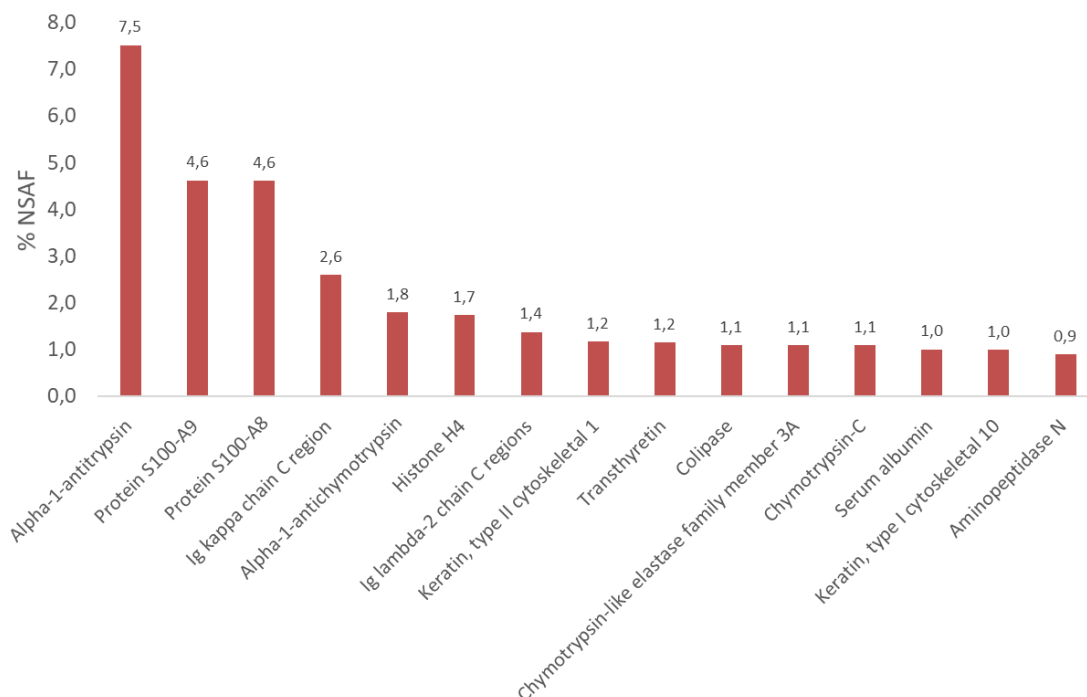


Figure 60 : Représentation des 15 protéines les plus abondantes parmi les 922 protéines humaines identifiées dans le méconium. Les protéines sont représentées dans leur ordre d'abondance relative (% NSAF).

La plus forte abondance est observée pour la α -1-antitrypsine (A1AT), et est en lien avec ce que nous avons visualisé en gel d'électrophorèse lors des tests d'homogénéisation du méconium (cf Figure 30, Chapitre I, Partie III.A.2.b). Cette protéine a largement été décrite dans la matrice méconium [136], [137], [295], ainsi que d'autres antiprotéases [295], [296] tel qu'observé ici pour l'antichymotrypsine. Les études décrivent une perméabilité élevée de la barrière épithéliale intestinale chez le nouveau-né, conduisant à un transport trans-épithélial plus important des macromolécules présentes dans la lumière intestinale. Les inhibiteurs de protéase permettraient d'inactiver les protéases intestinales produites par le nouveau-né, et ainsi de protéger les tissus et notamment le foie durant la période périnatale. Ces antiprotéases permettent également de maintenir actives et de transférer au nouveau-né les protéines

essentielles contenues dans le lait maternel, notamment dans le colostrum, telles que les cytokines et les facteurs de croissance [138], [276]. Des concentrations élevées de ces antiprotéases sont observées jusqu'à ce que la barrière intestinale mature et devienne imperméable aux macromolécules (« gut closure ») [300]. Il est également décrit que les protéases et antiprotéases s'accumulent dans le méconium exclusivement pendant la vie intra-utérine et sont excrétées après la naissance [76].

Les deuxièmes protéines les plus abondantes sont les protéines S100A8/A9, dont les rôles sont multiples : régulation de l'homéostasie immunitaire, des lésions post-traumatiques, de l'apoptose et de l'inflammation [301]. La forte concentration de ces protéines dans la lumière intestinale des nouveau-nés semble traduire leur rôle dans le maintien global de l'homéostasie intestinale. Cependant, la naissance est un processus stressant pour le fœtus, et peut conduire à une augmentation des concentrations de calprotectines à la naissance [75]. Une augmentation de la concentration de ces protéines dans les méconiums pourrait permettre de mettre en évidence une inflammation ou un stress périnatal.

De grandes quantités de fragments d'immunoglobulines (Ig kappa, lambda...) sont également mises en évidence. Des taux élevés en lactoferrines, principales composantes du lait maternel (possédant des propriétés anti-microbiennes et immunomodulatrices) ont également été observés. La littérature indique une corrélation entre les concentrations d'immunoglobulines et de lactoferrines dans le lait maternel et dans le méconium, suggérant que la présence de ces protéines à des taux élevés dans le méconium provient du lait maternel [297].

a.3) Description d'un « core protéome » dans les échantillons de méconiums analysés

La comparaison des données obtenues précédemment (Figure 59), a permis de définir un core protéome commun à tous les échantillons de méconiums analysés. Celui-ci est décrit par 79 protéines malgré des fluctuations importantes de l'abondance de certaines de ces protéines au sein des échantillons. Ces 79 protéines couvrent 9 % des protéines humaines identifiées en nombre, et représentant 53 % de l'abondance totale des protéines humaines détectées (Liste en Annexe 14, triée par ordre d'abondance décroissante). Parmi les protéines les plus abondantes de ce cœur protéome, nous retrouvons celles décrites précédemment, telles que la A1AT, les S100-A8/A9 et les Ig kappa.

Afin de connaître les voies biologiques les plus impliquées au sein de ce « core protéome », nous avons réalisé une analyse d'enrichissement de l'ontologie du gène sur les 79 protéines identifiées. Cette analyse a été réalisée sur l'outil Metascape en utilisant l'annotation et les termes d'ontologie trouvés dans les applications « Gene Ontology, GO » et « KEGG pathways » [298]. Les voies biologiques les plus représentées ont été regroupées et décrites par ordre décroissant d'enrichissement. Parmi les 18 voies

fonctionnelles les plus représentées, nous retrouvons des protéines liées à la localisation cellulaire et au système immunitaire, puis celles liées aux processus de régulation biologique et métabolique, et à la biologie du développement (Figure 61). Lisowska et al ont décrit l'implication des protéines humaines détectées dans le méconium au sein de différents processus, par exemple liés au développement, à la signalisation, au transport, au processus métaboliques et à la réponse aux stimuli... Les auteurs suggèrent que la détection de cet ensemble de protéines pourrait permettre d'évaluer de façon non invasive l'environnement intra-utérin [136], [137].

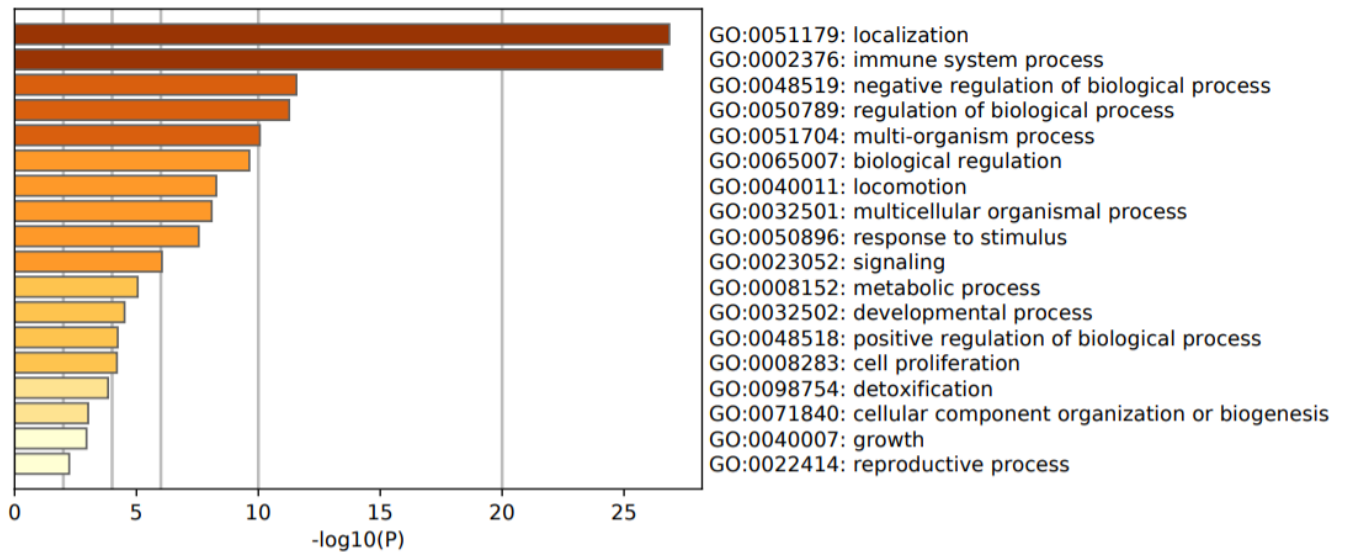


Figure 61 : représentation des voies fonctionnelles les plus enrichies par les 79 protéines présentes chez l'ensemble des échantillons de méconium analysés.

Un large panel des fonctions des protéines que nous avons identifié dans le méconium est décrit dans le Tableau 12. Celui-ci est réalisé à partir des données décrites dans la littérature [136], [137], [140], [141].

Tableau 12 : Description de quelques fonctions biologiques attribuées aux protéines détectées dans nos échantillons de méconiums [136], [137], [140], [141].

Rôle biologique	Protéines et Fonctions
Transport	<u>Albumine sérique</u> : lie l'eau, Ca ²⁺ , Na ⁺ , K ⁺ , le zinc, les acides gras, les hormones, et les médicaments.
	<u>Transthyretine</u> protéine de liaison à l'hormone thyroïdienne.
	<u>Lactotransferrine, melanotransferrine</u> protéines impliquées dans l'absorption cellulaire du fer, protéines de stockage du fer intracellulaire, impliquées dans de multiples processus tels que l'apoptose, l'immunité innée et le développement rénal.
	<u>Protéines liant les acides gras</u> se lie aux acides gras à chaînes longues et à d'autres ligands hydrophobes.
	<u>Protéines FAM151A</u> protéine intracellulaire, transport du glucose et d'autres sucres, sels biliaires, acides organiques, ions métalliques, composés aminés, co-transport des ions sodium et des dicarboxylates tels que le succinate et le citrate.
Protéines de structure	<u>Famille des kératine</u> protéines structurelles hétéropolymères qui forment le filament intermédiaire.
	<u>Famille des histones</u> protéines nucléaires (structure chromatine)
	Fibronectine, de forme dimère ou multimère à la surface de la cellule et dans la matrice extracellulaire.
Différents rôles impliqués dans différentes cellules	<u>Famille des Actines</u> , impliquées dans divers types de motilité cellulaire.
	<u>Famille des Annexins</u> joue un rôle dans la régulation de la croissance cellulaire et intervenant dans l'inflammation et l'apoptose.
	<u>Famille S100</u> impliqué dans la régulation d'un certain nombre de processus cellulaires tels que la progression et la différenciation du cycle cellulaire.
	<u>Famille des Cadhérines</u> protéines d'adhésions cellulaires dépendantes du calcium.
Protéines spécifiques dans les fonctions intestinales du fœtus	<u>Famille des Mucines</u> joue un rôle essentiel dans la formation de barrières muqueuses protectrices sur les surfaces épithéliales et forme une barrière muqueuse insoluble qui protège la lumière intestinale.
	<u>Molécules d'adhésions cellulaires glycoprotéines</u> impliquées dans l'adhésion cellulaire, normalement produites dans le tissu gastro-intestinal au cours du développement du fœtus.
	<u>Villin-1</u> régule la morphologie des cellules épithéliales intestinales, l'invasion cellulaire, la migration cellulaire et l'apoptose.
	<u>Famille des Trefoil</u> protéines sécrétoires stables exprimées dans la muqueuse gastro-intestinale, protègent la muqueuse contre les agressions et stabilisent la couche de mucus.
	<u>cGMP-dependent protein kinase 2</u> appartient à la famille des protéines sérine/thréonine protéine kinase, joue un rôle dans la régulation de l'équilibre des fluides dans l'intestin.
	<u>Transport du zinc ZIP4</u> nécessaire pour l'absorption du zinc dans l'intestin.
	<u>Echangeur de sodium/hydrogène</u> rôle dans l'absorption intestinale du sodium.

b) Analyse de la variabilité interindividuelle et de l'évolution temporelle de la composition du méconium en protéines humaines

b.1) Analyses en composantes principales réalisées sur les 922 protéines humaines identifiées

Après avoir investigué les similitudes en composition protéique dans nos échantillons, nous nous sommes intéressé : i) d'analyser leurs différences, ii) d'évaluer la variabilité interindividuelle ainsi que iii) de suivre l'évolution de la composition en protéines humaines au cours des trois premiers jours de vie.

Nous avons réalisé dans un premier temps des analyses non supervisées de type ACP. La représentation obtenue lors d'une première analyse permet de visualiser 39 % de la variance totale du jeu de données à travers les deux premières composantes : PC1 (25 %) et PC2 (14 %) (Figure 62.A). Nous avons au sein de la troisième composante 9 % de l'inertie résiduelle qui est expliquée (n'apportant pas d'informations complémentaires, celle-ci ne sera pas considérée). La représentation obtenue montre la présence de trois échantillons extrêmes, incluant NB6.960 et 2520 ainsi que NB3.4740. Une spécificité liée au nouveau-né NB6 serait possible. Dans le cas du NB6.2520, son profil extrême rejoint ce qui a précédemment été observé en analyses métabolomiques, représenté également en dehors du lot. Nous avons relancé cette analyse sans l'échantillon NB6.2520 et apercevons que les échantillons NB3.4740 et NB6.960 sont à leur tour identifiés très extrêmes du reste de la distribution. Nous retiendrons donc l'ensemble des échantillons pour la suite des analyses, au vu de la petite taille de notre cohorte. Au sein du restant des échantillons, nous observons un profil évoluant des temps de collecte les plus précoces aux plus tardifs (exemple du NB1.195 jusqu'à NB4.3780). Au vu des résultats obtenus précédemment en métabolomique (mise en place de profils métaboliques différents dès 1440 minutes post-naissance), nous avons représenté l'ACP en fonction des jours de prélèvements (Figure 62.B). Nous observons que les échantillons collectés avant 2880 minutes suivant la naissance sont relativement groupés, mis à part trois échantillons représentatifs des temps les plus précoces (NB10.90, NB9.195, NB8.270). La coloration des échantillons en fonction des nouveau-nés (Figure 62.C), montre un effet plus marqué qu'en métabolomique, avec le regroupement de nombreux échantillons issus du même nouveau-né, tel que NB1, NB2, NB4 et NB5.

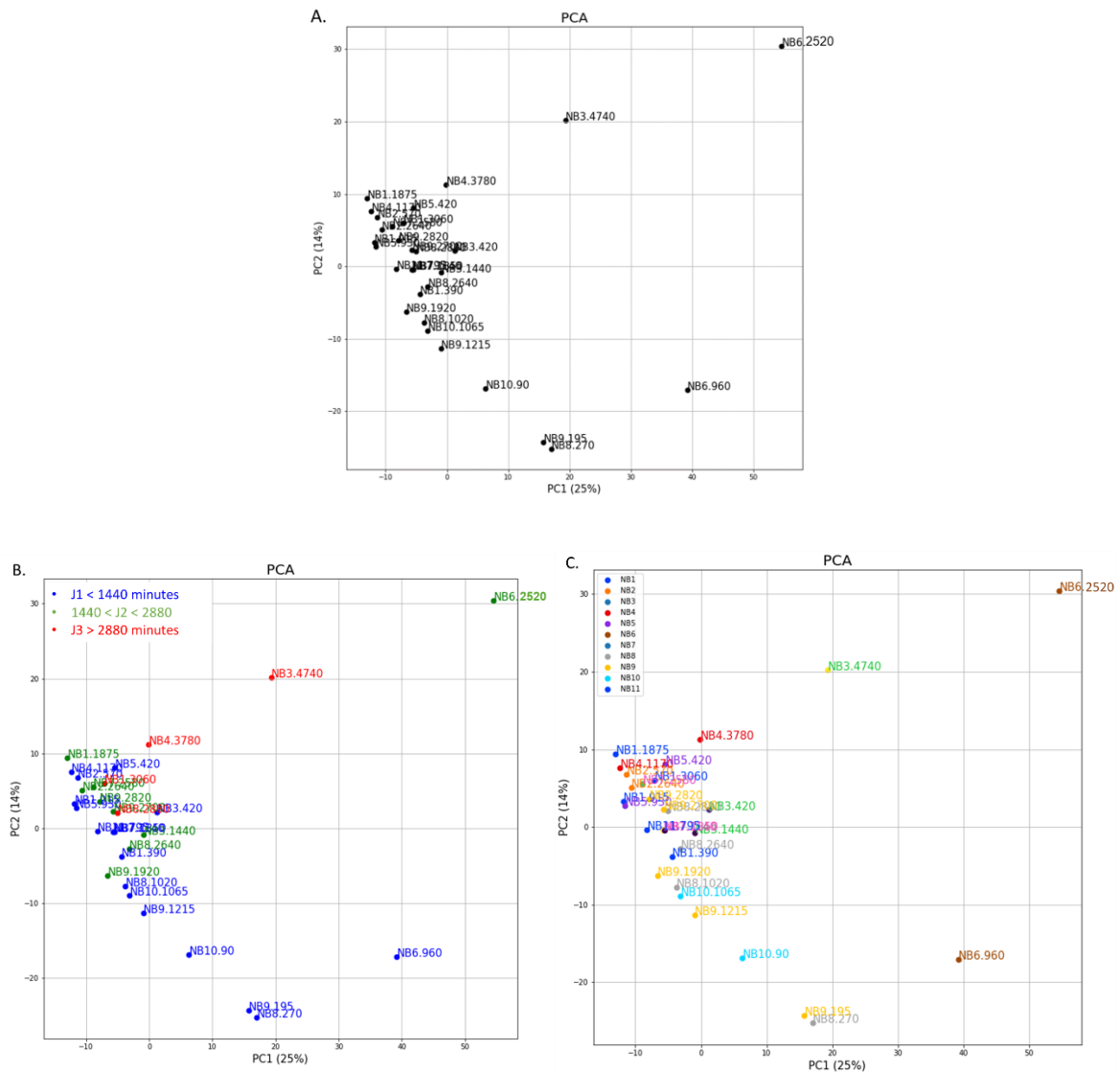


Figure 62 : Représentation des deux premières composantes de l'ACP réalisée sur les 922 protéines humaines identifiées dans le méconium. A. Sans coloration. B. Coloration des échantillons selon le jour de prélèvement. En bleu sont représentés les échantillons collectés avant 1440 minutes suivant la naissance (J1), en vert les échantillons collectés entre 1440 et 2880 minutes (J2), et en rouge au-delà de 2880 minutes (J3). C. Coloration des échantillons selon le nouveau né (de NB1 à NB11).

b.2) Classification non supervisée des échantillons et des protéines

Nous avons réalisé une clusterisation hiérarchique, permettant à la fois : i) d'évaluer la proximité des échantillons (en colonnes) et des protéines (en lignes), ii) de visualiser les abondances relatives des protéines au sein des échantillons (% NSAF) (Figure 63). La coloration ne permet pas de visualiser les différences en termes de quantités relatives entre les différents échantillons, du fait d'une majorité de protéines présentes à de faibles abondances et de façon plus ou moins équivalente au sein des différents échantillons de méconiums. De fortes abondances en bas de la figure sont observées (peu

visibles au vu du nombre de protéines représentées). D'un autre côté, nous observons la formation sans *a priori* de 3 clusters d'échantillons. Nous observons au sein du cluster 1 que 4 des 5 échantillons sont issus de temps de collectes > 2600 minutes (représentant les temps les plus tardifs de notre cohorte). Cependant, d'autres échantillons représentatifs de ces temps tardifs sont présents dans d'autres clusters (tels que NB9.2700 et NB8.2640/2880 représentés dans le cluster 2). Au sein des clusters 2 et 3, les échantillons ne sont pas regroupés par temps de prélèvement mais plus par nouveau-nés tels qu'observé pour NB7, NB8 et NB10 dans le cluster 2, et NB1 dans le cluster 3.

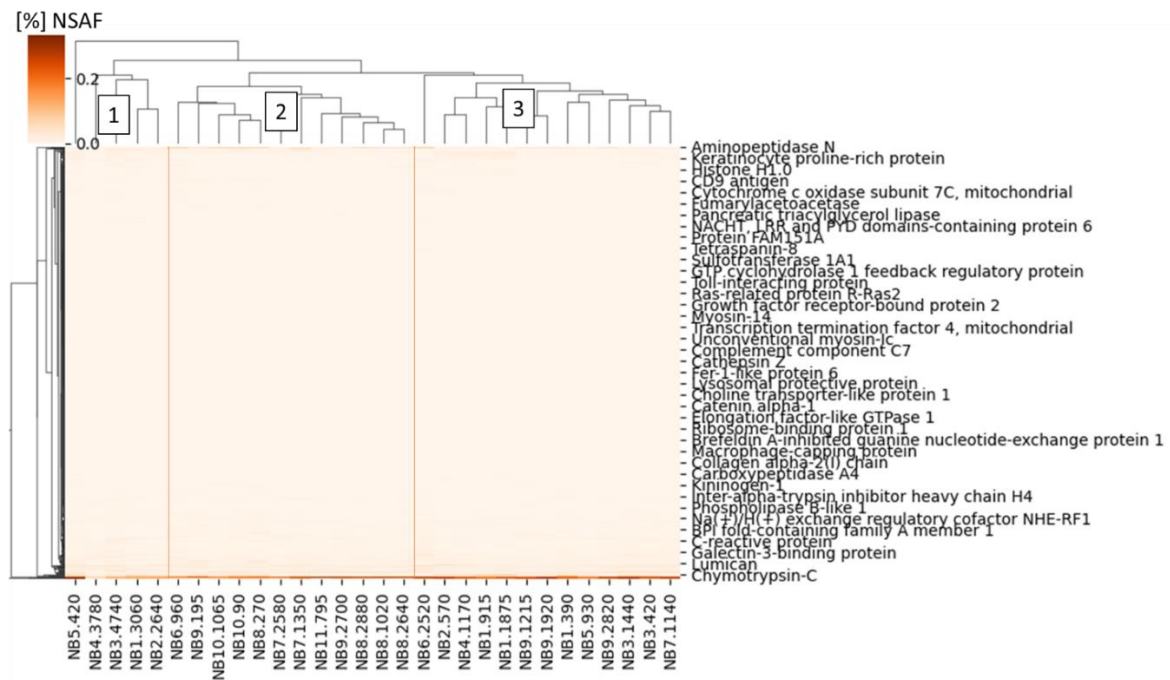


Figure 63 : Classification hiérarchique des échantillons en partant de l'ensemble des protéines identifiées (922 protéines). Représentation des échantillons en colonnes (clusters 1, 2 et 3) et des protéines humaines identifiées en lignes (toutes les protéines ne sont pas écrites (à droite de la figure)).

b.3) Classification non supervisée des protéines les plus variantes dans notre jeu de données

Afin d'avoir une visualisation sur les protéines variant le plus au sein de nos échantillons à travers cette représentation hiérarchique. Nous avons calculé la variance de chacune des 922 protéines chez l'ensemble des échantillons, et avons relancé cette représentation sur les 10, 30, 50 et 100 protéines variant le plus dans le jeu de données.

La classification hiérarchique des échantillons et des protéines (et de leurs abondances) permet d'avoir une première visualisation sur les différences au sein des individus/temps de collectes. La Figure 64 est représentée sur les 10 protéines montrant le plus de variabilité dans notre jeu de données (très peu de variabilité observée sur les 30, 50 et 100 protéines). Nous décrivons au sein de ces 10 protéines : les S100-A8/A9, la α -1-antitrypsine, la kératine type II, la protéine Ig kappa, l'histone H4, la colipase, le

sérum albumine... Cette figure montre que les protéines les (α -1-antitrypsine et S100-A8/A9) semblent plus abondantes dans les temps les plus précoces (encadrés en rose) que les temps les plus tardifs (encadrés en violets). Ces constatations sont en accord avec ce qui a été décrit dans la littérature [138], [295]. Certains échantillons précoces tels que NB10.90, NB8.270 et NB5.930 (marqués par une étoile rouge) montrent un déficit en protéines S100 et A1AT comparativement aux abondances observées chez les autres individus aux temps proches. Nous observons en globalité des abondances plus fortes en protéines telles que le sérum albumine, la colipase et l'histone au sein de temps de collectes précoces (cadre vert, intervalle de temps [90-1065] minutes). Le même constat est observé pour des protéines telles que la kératine, des protéines B associées au surfactant pulmonaire (cadre rouge, intervalle de temps moyennement précoce [570-1875], temps maximal de notre collecte = 4740 minutes). Concernant la protéine Ig kappa, nous n'observons pas de profil spécifique, celle-ci étant très abondante dans deux temps tardifs NB3.4740 et NB9.2820 ainsi qu'à un temps très précoce, NB5.420.

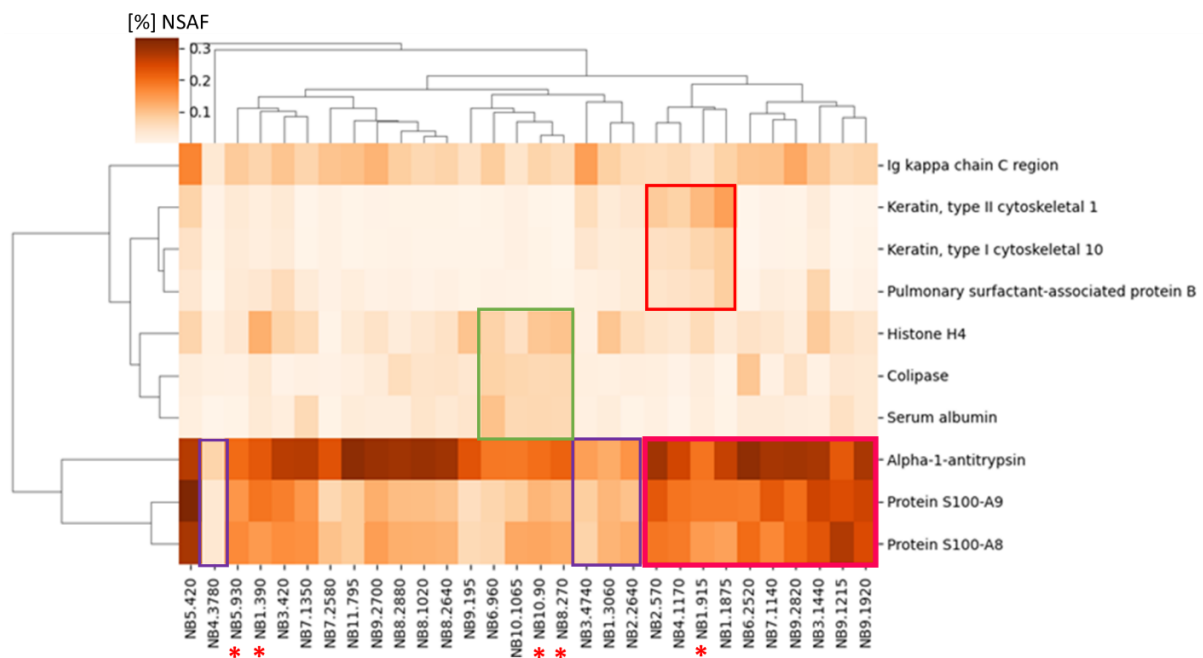


Figure 64 : Classification hiérarchique des individus et des variables, en considérant les 10 protéines montrant le plus de variabilité dans le jeu de données. La coloration indique l'abondance relative des protéines dans les différents échantillons.

Afin de suivre l'évolution des protéines au cours des trois premiers jours de vie, et d'identifier celles qui seraient représentatives des temps précoces *versus* des temps tardifs, nous avons réalisé des représentations en forme de boxplots, en regroupant les échantillons suivant leur jour d'excrétion. La Figure 65.A regroupe les protéines qui montrent une diminution de l'abondance relative au cours du temps, à savoir la α -1-antitrypsine, les S100-A8/A9, l'éosinophil peroxidase, l'histone H4 ou encore la sérum albumine. Lisowska et al ont également décrit une baisse des taux de l' α -1-antitrypsine au cours des premiers jours de vie [299]. Des taux élevés en calprotectines (protéines S100-A8/A9) à la naissance

ont également été rapportés, décrivant des teneurs jusqu'à dix fois plus élevées chez le nouveau-né que chez l'enfant à l'âge de 1 an [140].

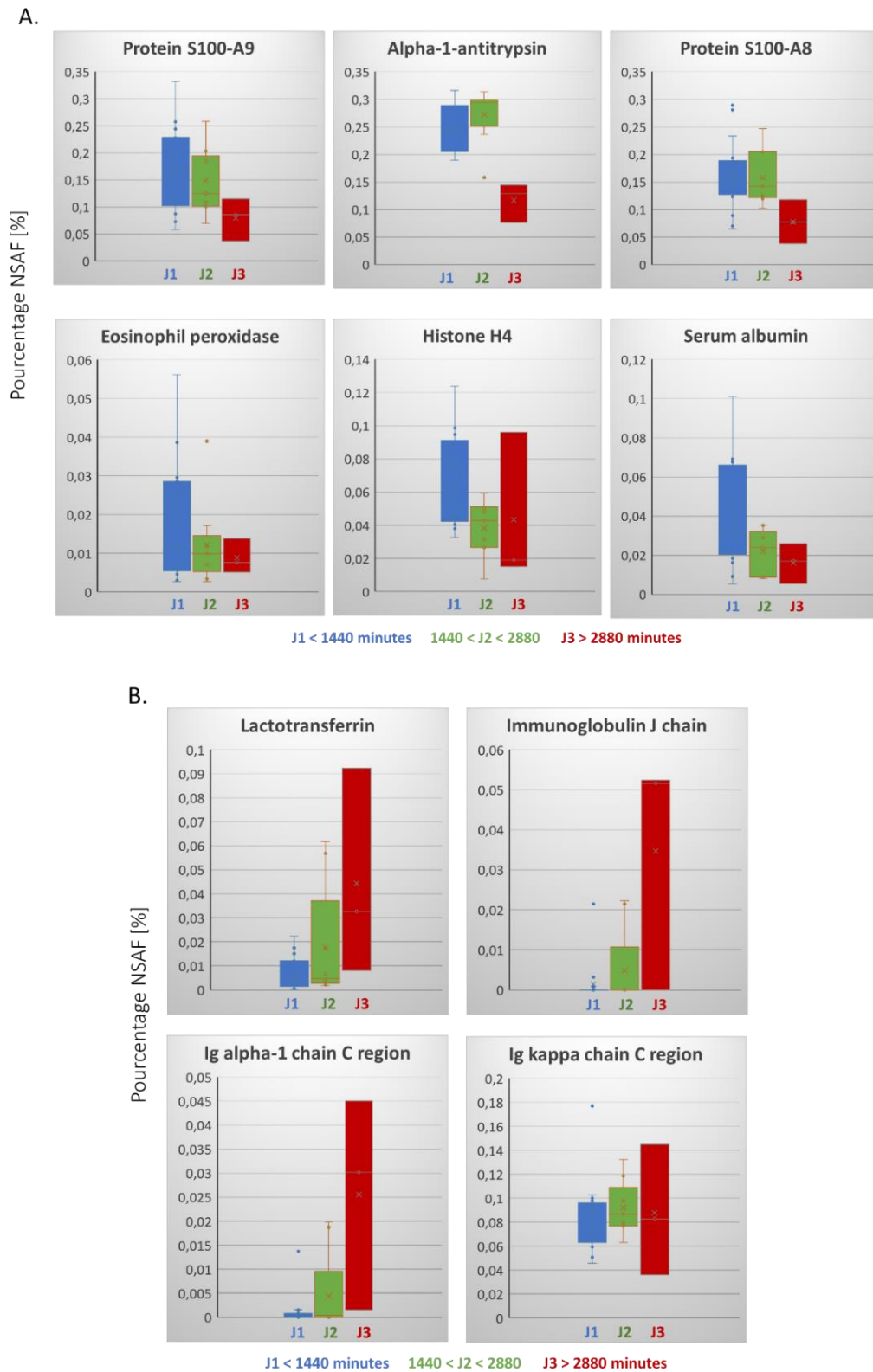


Figure 65 : Boxplots des protéines diminuant (A) *versus* augmentant (B) au cours du temps. En bleu sont représentés les échantillons excrétés le jour 1 (nombre d'échantillons = 19, nombre de nouveau-nés = 11), le jour 2 (nombre d'échantillons = 7, nombre de nouveau-nés = 6) et le jour 3 (nombre d'échantillons = 4, nombre de nouveau-nés = 4).

La Figure 65.B regroupe les protéines qui montre une augmentation de l'abondance relative au cours du temps, à savoir les régions constantes des chaînes légères λ et κ , ainsi que la chaîne J (liant les polymères d'IgA ou d'IgM sécrétés) et les lactotransferrines. La littérature indique que la présence d'immunoglobulines (notamment du type A) et de lactotransferrines à des taux élevés dans le méconium provenait du lait maternel [297]. En effet, il est décrit des teneurs très élevées en IgA dans le lait maternel, pouvant traduire une « contamination » du méconium par l'alimentation aux jour 2 et 3 suivant la naissance [41].

Après avoir analysé les spectres attribués aux protéines humaines, nous nous sommes intéressés d'analyser les spectres attribués aux taxons bactériens sur ce même jeu de données. Les résultats de ces analyses sont décrits ci-dessous.

2. Analyses métabolomiques

a) Traitement des données et attribution des protéines bactériennes

L'attribution des spectres MS/MS est cette fois réalisée à partir de la base de données NCBI nr (National Center for Biotechnology Information, non redondant), contenant les séquences de tous les organismes connus [300]. Nous générons en sortie de ces analyses une table avec le nombre de PSMs (Peptid-to-Spectrum-Matches) attribués aux protéines eucaryotes (principalement humaines) et procaryotes. Nous nous intéressons dans cette partie uniquement à l'analyse des PSMs attribués aux taxons bactériens, étant donné que l'étude des protéines humaines a été réalisée à travers la base de données Swissprot. Pible et al ont notamment proposé d'appeler ces spectres MS/MS attribués aux taxons par TSMs au lieu de PSMs, (pour Taxon-to-Spectrum-Matches), que nous utiliserons dans la suite [301].

Nous avons dans un premier temps analysé le taux de représentation du signal lié à des protéines bactériennes par rapport à celui lié à des protéines humaines au sein de l'ensemble de nos échantillons. Dans certains cas, il n'est observé aucun signal bactérien tels que pour les échantillons NB1.390, NB2.570, NB3.420 et NB5.420. Dans le cas où celui-ci est détecté, il présente globalement une augmentation progressive au cours des temps de prélèvements pour un nouveau-né donné (Figure 66). Cette représentation permet de rendre compte de la détection de TSMs très précocément chez la majorité des nouveau-nés (dès 90 minutes pour NB10 ou 195 minutes pour NB9). A l'inverse, le nouveau-né NB3 semble avoir une colonisation taxonomique plus tardive (> 1440 minutes). Ces résultats sont cohérents avec les données de la littérature où il est décrit la présence de signaux bactériens dans le méconium excrété dès le premier jour de naissance, bien qu'en faibles proportions [302].

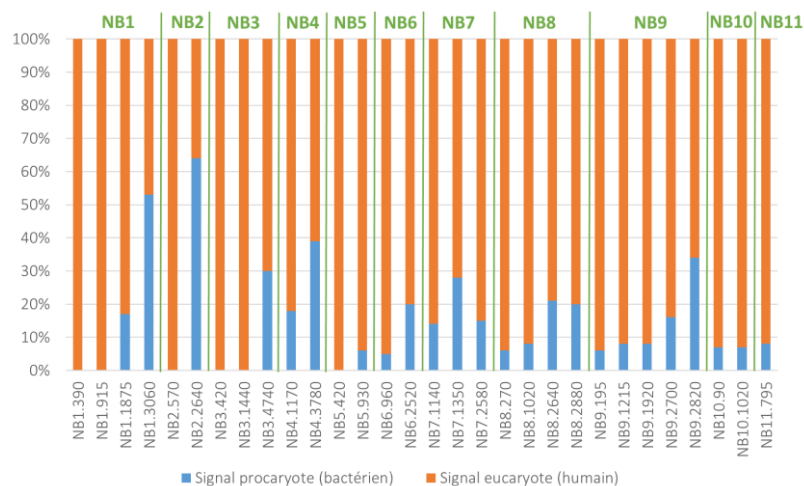


Figure 66 : Représentation du taux des signaux de protéines bactériennes et humaines détectées au sein de l'ensemble de nos échantillons. En bleu est représenté le signal bactérien, en orange le signal

humain. Les traits verticaux verts permettent de grouper les échantillons appartenant au même nouveau-né, afin de mieux les visualiser.

b) Description taxonomique au niveau phylum, et suivi de leur évolution temporelle chez les différents nouveau-nés

Nous avons ensuite cherché à décrire les différents rangs taxonomiques détectés dans notre jeu de données. Nous analysons dans cette première partie les TSMs qui ont été associés au niveau des phyla. Au sein de nos échantillons de méconiums, quatre phyla bactériens sont détectés : les Protéobactéries, les Actinobactéries, les Bacteroidetes et les Firmicutes (connus pour être des phyla abondant dans les fèces humains [2]). Les abondances relatives de chaque phylum chez les différents nouveau-nés et les différents temps de prélèvements ont ensuite été analysées. L'abondance relative est calculée à partir du nombre de TSMs attribués pour chaque phylum sur le nombre total de TSMs détectés dans chaque échantillon. Nous observons à partir de la Figure 67 que les Protéobactéries sont les premières colonies détectées dans le méconium, celles-ci étant détectables dès 90 minutes post-partum (NB10.90). Ce phylum représente également le seul présent dans tous les échantillons excrétés au cours des premiers jours de vie, excepté le NB9 aux temps plus tardifs (> 2700 min). L'abondance relative des TSMs associés aux Protéobactéries diminue au cours du temps chez la plupart des échantillons, ceci au profit de nouveaux phyla détectés. Ces phyla incluent des Actinobactéries, des Bacteroidetes et des Firmicutes, détectés dès 2520 minutes post-partum. Les Firmicutes n'apparaissent quant à elles qu'à des temps très tardifs (au-delà de 2700 minutes) et ne sont pas détectées chez tous les échantillons.

Nous observons à travers cette représentation deux types de profils. Un premier profil représenté par la détection des trois phyla des Protéobactéries, des Actinobactéries et des Bacteroidetes (par exemple, NB1.3060, NB2.2640, NB6.2520 ou encore NB8.2880). Un second profil est caractérisé par la détection des phyla Firmicutes et Actinobactéries (par exemple, NB9.2700 et 2820). L'échantillon appartenant au temps le plus tardif de notre collecte, soit le NB3.4740, présente un taux élevé de Firmicutes, rejoignant des résultats similaires décrits dans la littérature, sur l'augmentation des taux de Firmicutes, et ce dès le 3^{ème} jour post-naissance [115]. Au sein de cet échantillon, aucun TSM issu du phylum des Actinobactéries n'est détecté. Celui-ci étant décrit à de fortes abondances dans le méconium de nouveau-nés allaités (présence de *Bifidobacteria* issues de ce phylum), supposerait que le nouveau-né NB3 ne serait peut être pas allaité [302].

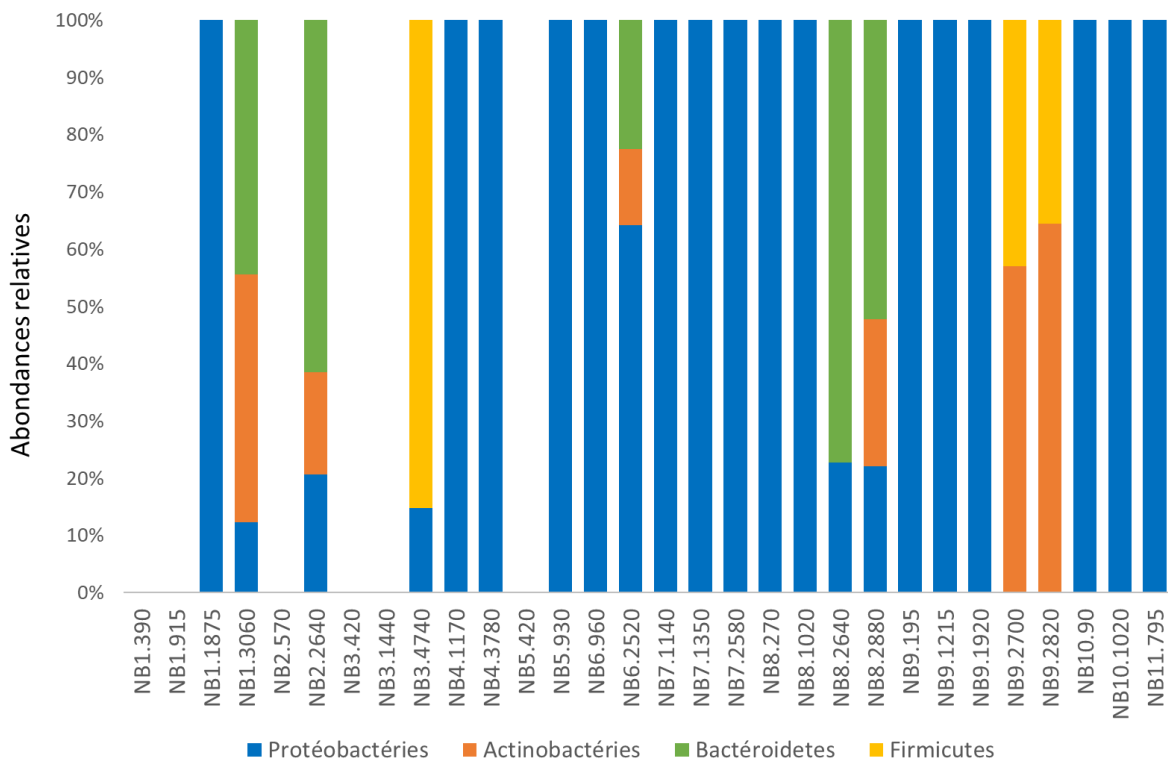


Figure 67 : Abondances relatives des différents phyla bactériens détectés dans les 30 échantillons de méconiums analysés. Les profils « blancs » correspondent aux échantillons où aucun signal bactérien n'est détecté.

Afin de suivre l'évolution du nombre de TSMs assignés à l'échelle des phyla et au cours des jours de prélèvements, nous avons représenté le pourcentage de chacun des phyla détectés au sein du jour 1, du jour 2 et du jour 3 post-partum (Figure 68). Ces représentations permettent de mieux visualiser ce qui a été décrit précédemment, avec la présence unique de TSMs issus du phylum des Protéobactéries au 1^{er} jour. A partir du 2^{ème} jour, les phyla Bacteroidetes (38 %), Actinobactéries (20 %) et Firmicutes (6 %) sont détectés. Les TSMs liés au phylum des Protéobactéries restent toujours présents avec 36 % du taux taxonomique total détecté au sein de ce jour. Enfin, à partir du 3^{ème} jour, nous observons une augmentation de la proportion des TSMs assignés aux Firmicutes, représentant 18 % du taux total (avec néanmoins une forte hétérogénéité dans les échantillons, selon les nouveau-nés). Ces résultats sont cohérents avec ce qui a précédemment été rapporté par Reinhardt et al, démontrant une primo-colonisation par les Protéobactéries dont la proportion diminue significativement dès les premiers jours de vie. Les auteurs décrivent les Actinobactéries comme étant le deuxième phylum bactérien détecté chez le nouveau-né suivi par les Bacteroidetes et les Firmicutes, devenant par la suite le phylum majoritaire à l'âge adulte [303].

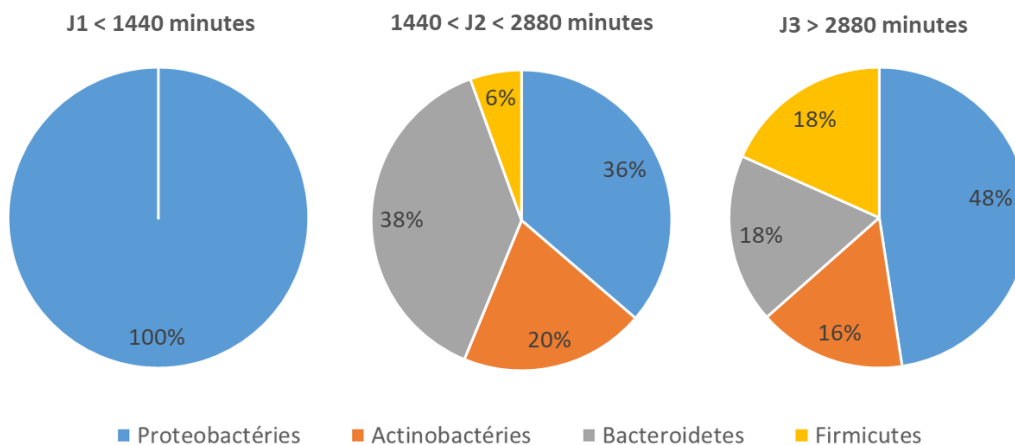
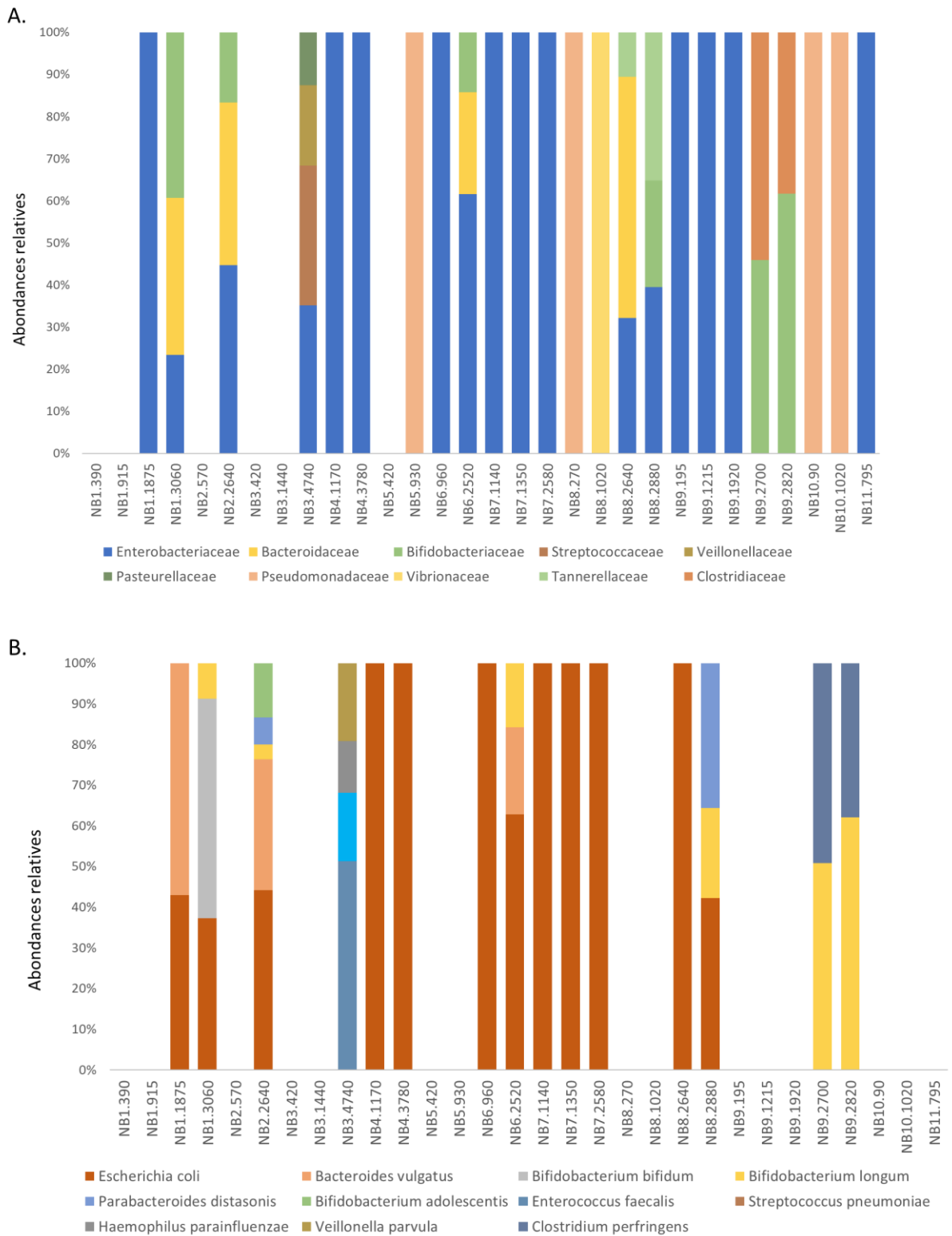


Figure 68 : Taux des TSMs attribués aux 4 phyla détectés dans le méconium, représentés par jour de prélèvement.

c) Description taxonomique aux niveaux des familles et des espèces, et suivi de leur évolution temporelle chez les différents nouveau-nés

Les différents niveaux de classification phylogénétique obtenus au sein de chaque TSM attribué a ensuite été analysé. Cette étude est réalisée sur le nombre de TSMs attribués à des peptides spécifiques des taxons et au sein de chaque rang taxonomique (aux niveaux de la famille et de l'espèce, présentées ici). Lorsque ce nombre est significatif et que l'attribution est correcte, alors le rang taxonomique peut être validé. Lorsque la famille (Figure 69.A) ou l'espèce (Figure 69.B) n'est pas validée à travers l'analyse des TSMs attribués, l'échantillon concerné est représenté en blanc dans ces figures. L'analyse des familles attribuées au sein des échantillons montre la présence d'*Enterobacteriaceae* (représentées en bleu) dans la quasi totalité des échantillons. Cette famille est prédominante chez des échantillons issus de nouveau-nés et de temps de collecte différents tels que NB6.960, NB1.1875 et NB4.3780. Dans la littérature, de fortes abondances relatives au niveau du genre *Enterobacter* chez des échantillons de méconiums collectés au cours des premiers jours de vie a également été décrite [302]. La famille des Pseudomonadaceae (phylum des Protéobactéries) est également prédominante dans quatre des échantillons étudiés, incluant principalement des temps précoces : NB5.930, NB8.270, NB10.90 et NB10.1020. Parmi les autres familles retrouvées, nous décrivons : les *Bifidobacteriaceae* (principalement représentatives des temps tardifs, tels que NB1.3060 et NB9.2820), les *Bacteroidaceae*, les *Streptococcaceae* ou encore les *Clostridiaceae*.



Au niveau de l'espèce, nous avons pu identifier 11 espèces bactériennes dans nos échantillons de méconiums (Liste dans le Tableau 13). L'espèce la plus abondante est *Escherichia coli* (représentée en marron dans la Figure 69.B). De nombreux articles ont décrit cette bactérie très précocement dans le méconium, la citant comme étant l'une des espèces les plus abondantes et les plus communes [302], [304], [305]. Les auteurs ont également rapporté que quelques échantillons de méconium hébergeaient jusqu'à trois souches ou de génomes uniques de *E. coli* [264]. Les autres espèces identifiées et relativement abondantes sont *Bifidobacterium longum*, *Bacteroides vulgatus*, *Clostridium perfringens*, ou encore *Bifidobacterium bifidum*. Bittinger et al. ont décrit trois espèces majoritaires au sein de leurs échantillons de méconiums, impliquant *Escherichia coli*, *Enterococcus faecalis* et *Bacteroides vulgatus* [264]. Dans leur étude, Zwittink et al. ont décrit un core de genre bactérien incluant les *Staphylococcus* et les *Enterococcus* (issus du phylum des Firmicutes), détectés dans des échantillons de méconiums et de premières selles (collectés au cours des 6 premières semaines post-naissance) [302].

Tableau 13 : Liste des 11 espèces bactériennes détectées dans le méconium.

Phyla	Familles	Espèces
Actinobactéries	<i>Bifidobacteriaceae</i>	<i>Bifidobacterium adolescentis</i>
		<i>Bifidobacterium bifidum</i>
		<i>Bifidobacterium longum</i>
Bacteroidetes	<i>Bacteroidaceae</i>	<i>Bacteroides vulgatus</i>
	<i>Tannerellaceae</i>	<i>Parabacteroides distasonis</i>
Firmicutes	<i>Clostridiaceae</i>	<i>Clostridium perfringens</i>
	<i>Enterococcaceae</i>	<i>Enterococcus faecalis</i>
	<i>Streptococcaceae</i>	<i>Streptococcus pneumoniae</i>
	<i>Veillonellaceae</i>	<i>Veillonella parvula</i>
Protéobactéries	<i>Enterobacteriaceae</i>	<i>Escherichia coli</i>
	<i>Pasteurellaceae</i>	<i>Haemophilus parainfluenzae</i>

Nos résultats sont cohérents avec la littérature, qui décrit que les premiers colonisateurs sont des bactéries aérobies, suivies par des bactéries anaérobies facultatives, principalement *Staphylococcus*, *Streptococcus*, *Enterococcus* et *Enterobacter*. Ces colonisateurs créent un environnement anaérobie conduisant à la colonisation par des anaérobies obligatoires ensuite, telles que *Bifidobacterium*, *Bacteroides* et *Clostridium* [306]. Ceci est indicateur d'un changement de microflore au cours des trois premiers jours de vie de nourrisson.

L'analyse métagénomique réalisée sur les échantillons de méconiums dans les premiers jours de vie post-natale a ainsi permis de détecter la présence de protéines bactériennes, et ce dès 90 minutes post-partum. Cette étude a été complétée par des analyses de métagénomiques ciblées réalisées sur ces mêmes échantillons de méconiums.

C. Analyses métagénomiques ciblées par séquençage des gènes codant pour les ARNr-16S

1. Traitement des données et attribution des OTUs (de l'anglais, Operational Taxonomic Unit)

Les données générées à la suite du séquençage sont traitées sur l'outil « FROGS » de la plateforme Galaxy (cf Partie II.D.3) [286]. Nous obtenons à cette étape, 10 millions de séquences qui sont par la suite filtrées, puis clusterisées par séquences proches (OTU). Nous avons ainsi défini 600 OTUs sur lesquelles un filtre de représentativité est appliqué, permettant de ne retenir que celles représentant des abondances supérieures à 0,05 % de l'ensemble des clusters. Dans ces conditions, 165 OTUs ont été retenues et identifiées (Annexe 15). Ces résultats sont cohérents à ce qui a précédemment été décrit dans la littérature par Del Chiericco et al, qui a définit 130 OTUs, détectées au sein de 121 échantillons de méconium [115].

La table des OTUs générée contient les noms des genres voire des espèces bactériennes identifiés ainsi que l'abondance de chacune de ces OTUs dans chaque échantillon. Une première « macro-analyse » permet de constater que *Escherichia coli*, *Bacteroides vulgatus* et *Clostridium perfringens* sont parmi les OTUs les plus abondantes. Ces résultats sont cohérents avec ce que nous avons précédemment décrit lors de l'analyse des protéines bactériennes, ainsi qu'avec la littérature [264]. *Clostridium perfringens* a notamment beaucoup été décrit dans le méconium, sa sur-abondance étant un marqueur de l'entérocolite nécrosante, [307]. *Escherichia coli* est décrite selon différentes souches, et est rapportée comme étant l'espèce bactérienne la plus détectée dans le méconium, et ce très tôt après la naissance [305], comme précédemment mentionné.

2. Etude de l'évolution de la composition bactérienne du méconium au cours des trois premiers jours de vie

a) Analyse de l' α -diversité et de l'évolution des abondances des OTUs entre nouveau-nés et au cours du temps

Le nombre ainsi que l'abondance de chacune des OTUs identifiées dans les différents échantillons sont représentés respectivement en Figure 70.A et Figure 70.B. Contrairement aux résultats de l'approche métagénomique précédemment décrite, nous détectons par l'approche métagénomique ciblée de l'ADN 16S, des bactéries chez tous les échantillons, pouvant suggérer une plus grande sensibilité de cette dernière approche. Nous détectons en moyenne 34 OTUs par échantillon, et notons une médiane de 30. De façon globale, nous observons un nombre croissant d'OTUs au cours des temps de prélèvements pour un même nouveau-né, hormis pour le NB8. L'échantillon présentant la plus faible α -

diversité (richesse spécifique, ou biodiversité, traduisant le nombre d'espèces présentes au sein d'un même échantillon) est l'échantillon NB6.960 avec 12 OTUs détectées. Celui qui présente la plus grande richesse est l'échantillon NB3.1440, avec 88 OTUs détectées. L'analyse métagénomique n'avait pas pu permettre de détecter des protéines bactériennes dans cet échantillon. Celui-ci présente un profil d'OTU très différent des autres échantillons, y compris de ceux prélevés à d'autres temps chez le même nouveau-né (Figure 70.B). Nous analyserons plus en détail dans la suite la composition phylogénique ainsi que les OTUs spécifiques à cet échantillon.

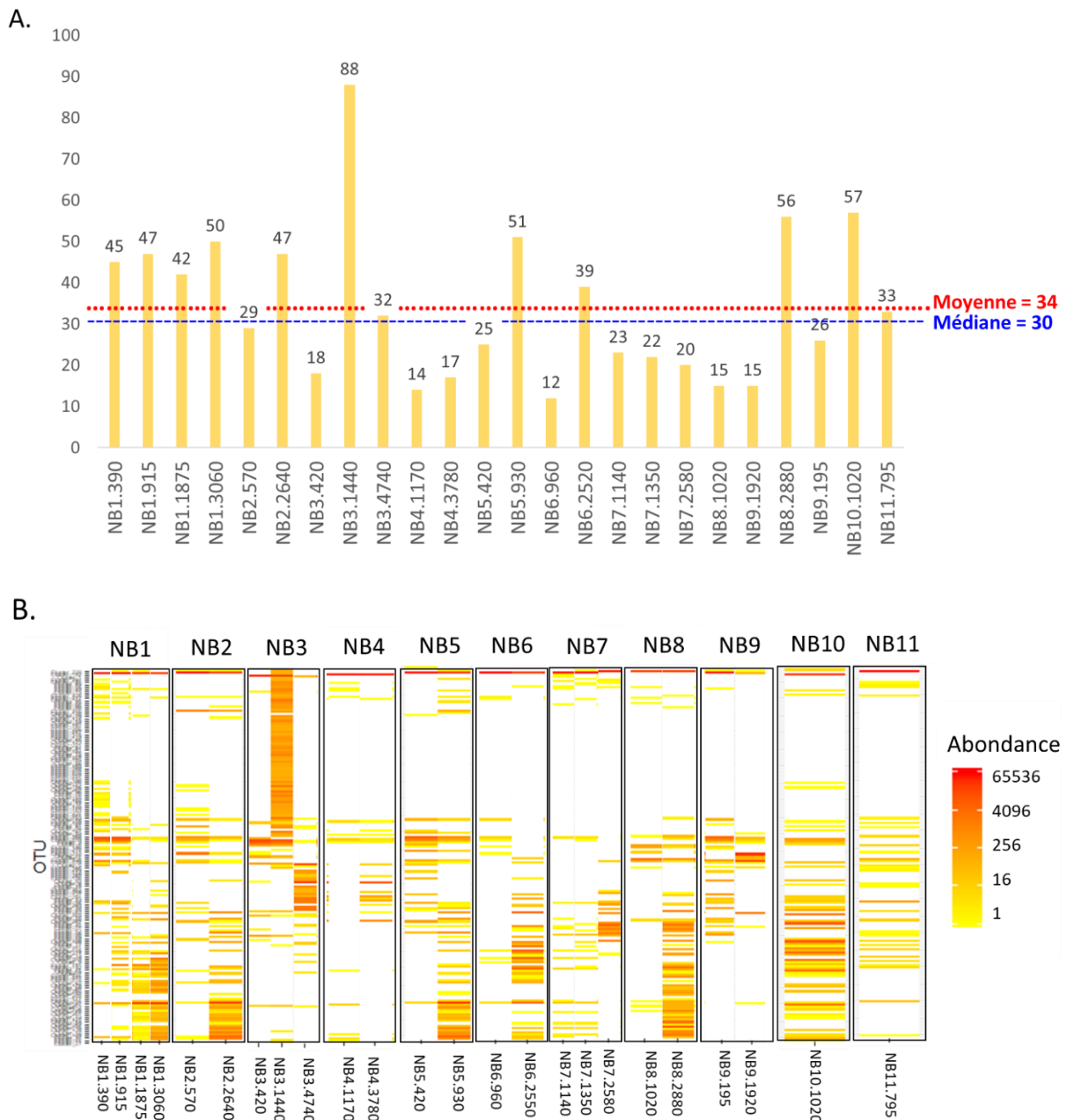


Figure 70.A. Représentation du nombre d'OTU détectées au sein de chacun des échantillons de méconiums analysés (en colonnes). B. Vue d'ensemble des OTUs identifiées et de leurs abondances au sein des différents échantillons de méconium. L'échelle d'abondance représente le nombre de séquences attribuées pour chaque OTU.

Les profils des OTUs présentent une grande variabilité interindividuelle, et pour un même nouveau-né, une forte évolution temporelle. A des temps de prélèvements équivalents, nous observons une α -diversité très variable d'un nouveau-né à un autre. En effet, en prenant comme exemple, les échantillons collectés aux alentours de 1000 minutes post-partum (16-17h) : nous observons chez le NB10 une diversité et richesse plus élevée que chez des échantillons collectés a des temps proches, tels que NB4, NB6, NB7 ou NB8. Nous n'avons pas pu identifier un « core OTU », commun à tous les échantillons : uniquement un OTU commun est détecté, correspondant à *Escherichia coli*. Ce constat rejoint ce qui a été décrit par Bittinger et al dans leur analyse du méconium [149]. Cependant certains profils d'OTUs sont en partie similaires, si on compare entre les échantillons précoces des différents nouveau-nés. Il en est de même entre échantillons aux temps de prélèvements tardifs (Figure 70.B).

b) Analyse de l'évolution des phyla bactériens

Les analyses métagénomiques ciblées décrivent la présence de cinq phyla dans les échantillons de méconium analysés : les Protéobactéries, les Bacteroidetes, les Actinobactéries, les Firmicutes et les Deferribactéries (par ordre d'abondance globale). Les quatres phyla les plus abondants sont en accord avec les résultats obtenus par l'analyse protéomique et avec la littérature.

La Figure 71.A permet de rendre compte de la variation de l'abondance de ces phyla bactériens en fonction des individus et au cours du temps. Une représentation par intervalle de temps, indépendamment des nouveaux-nés est également fournie (Figure 71.B). Globalement, nous observons peu de variabilité interindividuelle au niveau des phyla au sein des différents échantillons analysés. Le phylum le plus abondant est représenté par les Protéobactéries, qui représentent à elles seules plus de 75 % de l'abondance totale au sein de la grande majorité des échantillons. Pour certains individus et essentiellement dans les échantillons les plus précoces (tels que, NB2.570, NB4.1170, NB6.960, NB7.1140 et NB7.1350, NB11.795), ce phylum représente jusqu'à 95 % voire 100 % des phyla détectés. Celui-ci reste abondant à des temps plus tardifs pour la majorité des individus, représentant 30 à 80 % de l'abondance totale. La diminution de l'abondance relative de ce phylum se fait au profit d'une augmentation de celle des Bacteroidetes, puis des Actinobactéries aux temps plus tardifs. L'évolution de l'abondance relative des Firmicutes est plus variable en fonction des individus, tant dans son temps d'apparition que dans son abondance (ce qui se traduit par une forte variabilité, notamment dans la période C3). Les Deferribactéries sont uniquement détectées chez NB4.1440, échantillon qui présentait la plus grande richesse en OTU de notre jeu de données. L'échantillon NB9.1920, particulièrement riche en Firmicutes, présente un profil similaire à celui précédemment décrit par l'analyse des protéines

bactériennes (présence de deux phyla : les Firmicutes et les Actinobactéries). Ces résultats sont cohérents à ce qui été décrit par Reinhardt et al, [303].

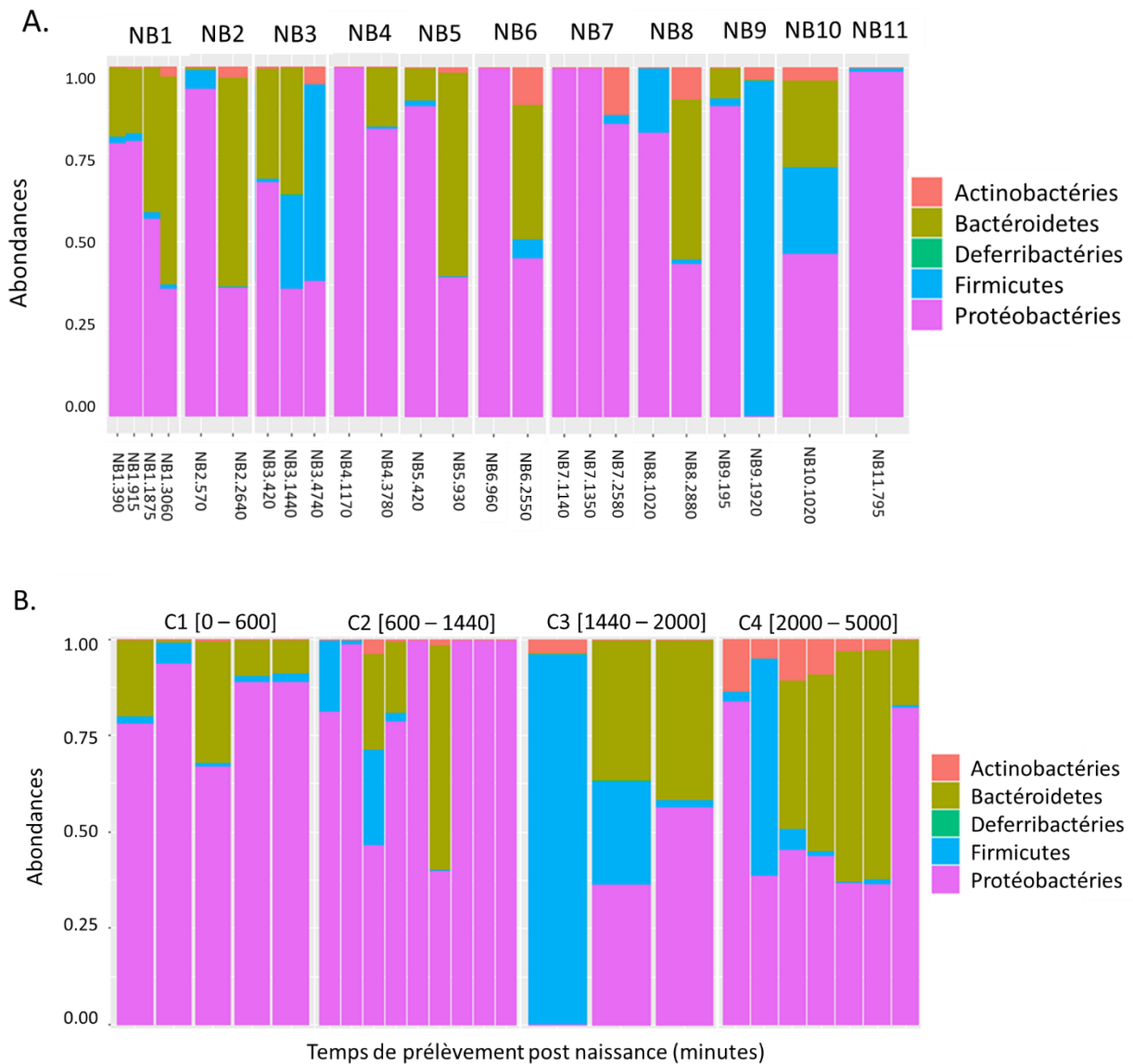


Figure 71. Représentation des abondances des phyla dans les différents échantillons de méconiums analysés. A. Représentation par individus et par temps croissant (de gauche vers la droite) ; B. Représentation des phyla bactériens au sein des échantillons regroupés par période de temps de prélèvements : C1 [0-600], C2 [600-1440], C3 [1440-2000] et C4 [2000-5000] minutes.

c) Analyse de l'évolution des familles bactériennes

Les familles taxonomiques les plus abondantes au sein des quatre phyla majoritaires identifiées sont représentées par intervalle de temps en Figure 72. Nous observons au sein des Protéobactéries, un même profil au sein des intervalles de temps C1 et C2, avec une prépondérance de la famille des

Enterobacteriaceae. Ces résultats rejoignent ce que nous avons précédemment décrit par les approches métagénomiques. On note ensuite l'apparition de la famille des *Burkholderiaceae* dans un échantillon de l'intervalle de temps C3, puis de la famille des *Pasteurellaceae* au sein de deux échantillons de l'intervalle de temps C4. Les autres familles sont très peu observées.

La famille des *Bacteroidaceae* est l'une des principales familles représentées au sein du phyla des Bacteroidetes. La famille des *Tannerellaceae* est détectée à partir des temps du groupe C2. Une plus grande variabilité est observée au sein de la période C3, avec l'apparition de la famille des *Muribaculaceae* et *Rikenellaceae* dans un des échantillons. Au sein de l'intervalle de temps C4, le phylum des Bacteroidetes est représenté que par deux familles, les *Bacteroidaceae* et des *Tannerellaceae*.

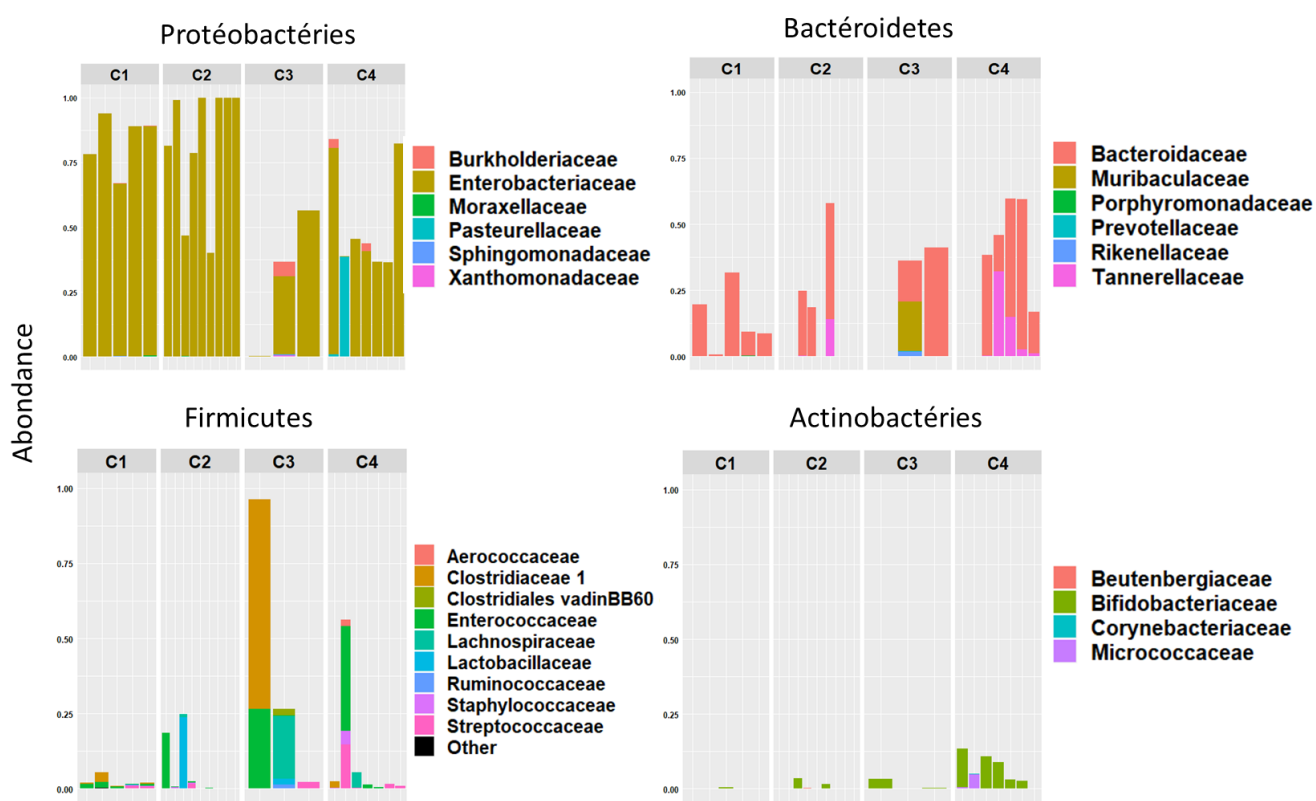


Figure 72 : Représentation des familles taxonomiques les plus abondantes au sein des différents phyla observés dans chacun des intervalles de temps considérés : C1 [0-600], C2 [600-1440], C3 [1440-2000] et C4 [2000-5000] minutes.

Une plus grande diversité et une plus grande variabilité interindividuelle est observée au sein du phyla des Firmicutes, avec la détection de différentes familles dès les temps précoces (C1). Les familles des *Enterococcaceae* et des *Lactobacillaceae* sont les plus abondantes au sein d'un des échantillons du groupe C2. Une variabilité plus importante est observée ensuite au sein des intervalles de temps C3 et C4, avec la présence de fortes abondances des familles *Clostridiaceae*, *Enterococcaceae* et

Lachnospiraceae. Une plus faible abondance est observée pour la famille des *Streptococcaceae*. Enfin le groupe C4 est principalement représenté par les *Enterococcaceae* et *Streptococcaceae*.

Parmi les Actinobactéries, des abondances très faibles sont observées au sein des quatre périodes analysées. Au sein des intervalles de temps C1, C2, et C3 et pour quelques échantillons, nous détectons des bactéries de la famille des *Bifidobacteriaceae*. Au sein de l'intervalle de temps C4, se rajoute la famille des *Micrococcaceae* par rapport aux précédents groupes (détectée dans un échantillon du groupe).

IV. Discussion générale des analyses « mutli-omiques » réalisées sur le méconium

Nous souhaitons à travers les analyses de métabolomique, métagénomique et métaprotéomique ciblée, réalisées sur les échantillons de méconiums de notre cohorte clinique, i) décrire une cartographie globale du méconium en métabolites, en protéines et en bactéries, ii) avoir une première idée de la variabilité interindividuelle de cette composition et iii) apprécier son évolution potentielle au cours des trois premiers jours de vie. Nous proposons désormais d'apporter une analyse plus intégrée des résultats, tout en gardant en mémoire la petite taille de cette cohorte et le caractère anonyme de la collecte. En effet, nous n'avons pas d'informations quant à des événements de santé de la mère ou du nouveau-né...

Des études omiques ont été réalisées auparavant sur du méconium ou des premières selles, incluant des analyses du microbiome, du protéome et/ou du métabolome [115], [264], [302]. Au sein de ces études, le traitement des données de chaque approche a été réalisé séparément, puis les données ont été mises en relation par la suite, permettant d'étudier les liens entre la composition microbienne et les profils protéiques et/ou métaboliques au cours des premiers jours de vie.

A travers les analyses réalisées, nous avons identifié 230 métabolites (au plus haut niveau de confiance), 922 protéines humaines, 11 espèces bactériennes par métaprotéomique et 165 OTUs bactériennes, représentant à notre connaissance, la description la plus exhaustive du méconium humain. Nous avons également décrit un « cœur ou core » métabolome et protéome, incluant 163 métabolites et 79 protéines humaines présents dans l'ensemble des échantillons analysés, quel que soit le temps de prélèvement et le nouveau-né. Ce « noyau » n'a pu être décrit au niveau des bactéries, que ce soit par les analyses métaprotéomique ou de métagénomique ciblée. Au sein des analyses métagénomiques ciblées, un seul OTU est présent dans l'ensemble des échantillons, correspondant à *E. coli*, en adéquation avec de précédentes observations [38], [115], [264].

Les analyses statistiques ont révélé que les compositions métaboliques, protéiques et bactériennes des méconiums évoluaient de façon significative au cours du temps, en accord avec de nombreuses études

[115], [264], [302]. Au-delà des « cores » décrits, nous avons mis en évidence la présence de profils métaboliques, protéiques et de phyla bactériens, représentatifs des temps de collecte précoces ou tardifs. Dans leur étude portant sur l'analyse de 88 échantillons excrétés de la naissance jusqu'au 7^{ème} jour post-partum, Bittinger et al ont décrit, 45 métabolites et 53 protéines humaines présentant des différences significatives au cours du temps. Nous avons défini 67 métabolites discriminants entre des temps de collecte inférieurs ou supérieurs à 1440 minutes (24h). Parmi ceux-ci, 50 métabolites montraient une augmentation de leur intensité, tandis que l'intensité des 17 autres diminuait significativement. Les métabolites dont l'intensité est plus importante dans les temps précoces concernent essentiellement les voies métaboliques de la purine, du glutamate ou des sucres. A l'inverse, les métabolites impliqués dans les voies telles que la biosynthèse des acides biliaires, de l'arginine et de la proline sont plus représentatifs des temps tardifs. Ces augmentations peuvent être partiellement mises en relation avec la mise en place du microbiote intestinal qui intervient au sein de ces voies [264]. Nous retrouvons des résultats similaires à ceux de la littérature concernant l'augmentation des métabolites tels que les acides glycérique, pyroglutamique, acétylaspartique et l'acétylalanine, ainsi que la diminution de l'adénosine, au cours du temps [264]. Ces résultats ont également été observés au sein de l'analyse ciblée réalisée par Righetti et al, démontrant une augmentation globale des concentrations des métabolites analysés dans des échantillons de méconiums issus du jour 1 et du jour 3 post-partum. Les auteurs ont mis en lien cette évolution avec le début de l'allaitement [256]. Del Chierrico et al ont également décrit de plus faibles quantités globales de métabolites fécaux au cours des 3 premiers jours de vie par rapport aux jours 7, 15 et 30 suivant la naissance. Cette augmentation a été expliquée par l'alimentation du nouveau-né, impliquant une augmentation du transit de composants du lait, ainsi que l'augmentation de la biomasse microbienne de l'intestin aidant à la digestion de certains de ces composants [115]. Par des analyses en composantes principales, ils ont défini 55 profils métaboliques évoluant au cours du temps. Des corrélations positives ont été observées entre les HMO et le lactate avec des espèces bactériennes telles que *Clostridium sphenoides* et *Bacteroides fragilis*.

Parmi les protéines humaines, nous avons également décrit, celles qui étaient les plus abondantes dans notre jeu de données, incluant la famille des antiprotéases (telles que la α -1-antitrypsine), accumulées *in utero* et excrétées en grandes proportions à la naissance. Celles-ci permettent en grande partie d'empêcher la protéolyse des protéines nécessaires pour la croissance du nourrisson, qui sont apportées par l'allaitement [276], [299]. La famille des calprotectines est également beaucoup représentée, ou encore des fragments d'immunoglobulines ou de lactoferrines apportés par l'allaitement [297].

Grâce aux analyses statistiques que nous avons réalisé ainsi qu'aux différentes représentations comparatives au sein des échantillons (par exemple nombre de protéines ou d'espèces bactériennes

détectées), nous avons relevé des échantillons extrêmes dans notre jeu de données. C'est le cas notamment de l'échantillon NB6.2520, qui présentait de faibles abondances métaboliques et un grand nombre de protéines détectées. Par contre, cet échantillon ne présentait pas un profil spécifique en termes de nombre d'OTUs. En parallèle, nous retrouvons en métaprotéomique et en métagénomique ciblée le même échantillon extrême NB6.960, qui présente un grand nombre de protéines humaines détectées et une très faible α -diversité (i.e. très faible nombre d'OTUs détectées). Bittinger et al ont rapporté à ce sujet une altération des abondances protéiques dans l'intestin lorsqu'un taux élevé en bactéries était détecté dans le méconium [264].

Un lien direct existe entre les métabolites, les protéines et les bactéries de la lumière intestinale. Par exemple, des acides aminés peuvent être libérés par l'hydrolyse des protéines (par les enzymes de l'hôte ou par les bactéries). Ces mêmes acides aminés peuvent être utilisés pour la synthèse de nouvelles protéines au sein des cellules bactériennes ou de l'hôte [308]. A la naissance, la colonisation microbienne du tractus gastro-intestinal humain évolue rapidement jusqu'à atteindre une population microbienne stable, ressemblant à celle à l'âge adulte. Les premiers colonisateurs sont des bactéries aérobies, puis des bactéries anaérobies facultatives, principalement *Staphylococcus*, *Streptococcus*, *Enterococcus* et *Enterobacter*. Ces colonisateurs créent un environnement anaérobie conduisant à la colonisation par des anaérobies obligatoires, telles que *Bifidobacterium*, *Bacteroides* et *Clostridium* [306].

Il est souvent produit, lors de la métabolisation des aminoacides par les bactéries, des acides gras saturés qui sont à leur tour métabolisés par des bactéries anaérobies, conduisant finalement à la production d'acides organiques ou d'acides gras à chaînes courtes (AGCC, acétate, propionate et butyrate) [309]. Les Bacteroidetes produisent essentiellement de l'acétate et du propionate, tandis que les Firmicutes produisent essentiellement du butyrate [2], [305]. Nos méthodes analytiques ne permettent pas de détecter ces AGCC. Par contre, nous retrouvons dans les méconiums des métabolites bactériens tels que les acides gras à chaînes branchées (AGCB), incluant principalement l'isobutyrate (acide aminé précurseur : valine) et l'isovalérate (acide aminé précurseur : leucine), qui sont les produits finaux majeurs du métabolisme bactérien des protéines et des acides aminés [308]. La présence de l'isobutyrate dans notre cartographie métabolique et sa présence au sein des 50 métabolites « représentatifs » des temps de prélèvements tardifs, traduit une colonisation et une activité métabolique bactériennes croissantes dans l'intestin du nourrisson. Par ailleurs, nous retrouvons au sein de notre cartographie le tryptophane, acide aminé essentiel qui peut être soit utilisé pour la synthèse des protéines, soit être métabolisé par les bactéries intestinales permettant de produire des métabolites essentiels sur le plan immunologique. Les principaux métabolites microbiens du

tryptophane dans l'intestin incluent l'indole, l'acide indolique et ses dérivés [305], que nous détectons dès des temps précoces dans nos échantillons.

Nous mettons en évidence d'autres métabolites bactériens dans le méconium, tels que des polyamines (agmatine et cadavérine) issues d'acides aminés (respectivement de l'arginine et de la lysine) par des réactions de décarboxylation bactérienne. Parmi les acides biliaires retrouvés, nous décrivons des acides biliaires primaires endogènes (tels que les acides cholique, glycocholique, taurochenodeoxycholique), ainsi que des acides biliaires secondaires bactériens (tels que les acides lithocholique et désoxycholique). D'autres acides biliaires conjugués au sulfate tels que le lithocholique-3-sulfate, et le tauroolitholique-3 sulfate sont également détectés. Ces acides biliaires ont déjà été décrits dans le méconium [253], [254]. Naritaka et al ont notamment démontré que les acides biliaires conjugués au sulfate sont issus du liquide amniotique avalé par le fœtus, et probablement issus de la métabolisation par le microbiote de la mère [254].

Nous avons regroupé dans le tableau ci-dessous, les métabolites bactériens retrouvés dans le méconium au sein de notre cartographie. Ces métabolites sont produits pas des espèces bactériennes spécifiques du microbiote intestinal. Nous citons notamment les produits dérivés du tryptophane retrouvés dans le méconium (tels que l'indoleacétique, l'indoleacrylique ou l'indolelactique), pouvant traduire la présence de bactéries telles que les *Bacteroides fragilis* ou *Bifidum longum* (Tableau 14).

Tableau 14 : Description des métabolites bactériens retrouvés au sein de nos échantillons de méconium, pouvant traduire la présence de certaines bactéries les produisant.

Substrats	Bactéries métabolisantes	Métabolites	Références
Tryptophane	<i>Bacteroides fragilis</i> , <i>Bifidobacterium longum</i> , <i>Clostridium perfringens</i>	Indoleacétique	[310], [311]
	<i>Clostridium</i>	Indoleacrylique	[312]
	<i>Bacteroides fragilis</i> , <i>Bifidobacterium longum</i> , <i>Clostridium perfringens</i> , <i>Lactobacillus</i>	Indolelactique	[310]
Acides biliaires	<i>Bacteroides fragilis</i> , <i>Clostridium perfringens</i> , <i>Bifidobacterium</i> , <i>Lactobacillus</i>	Acides biliaires secondaires	[313], [314]
Arginine	<i>Bacteroides fragilis</i> , <i>Streptococcus</i>	Polyamines	[315]

L'analyse de l'ADN bactérien ainsi que des protéines bactériennes du méconium se rejoignent et décrivent la présence de 4 phyla principaux dans le méconium : les Protéobactéries, les Actinobactéries,

les Bacteroidetes et les Firmicutes. Différentes études décrivent des résultats comparables aux nôtres [115], [305]. Dans leur étude, Del Chierrico et al ont également souligné une diminution significative (d'environ 50 %) de l'abondance en Protéobactéries dans le méconium dès le 2^e jour post-naissance, tandis qu'une abondance plus élevée est observée au sein des Firmicutes, tendance confirmée au 3^e jour [115]. Reinhardt et al ont notamment décrit une forte abondance en Protéobactéries à la naissance, suivie par une colonisation par des Bacteroidetes, des Firmicutes et enfin des Actinobactéries. Les auteurs décrivent une diminution des proportions des Protéobactéries à partir du 3^e jour suivant la naissance (Figure 73).

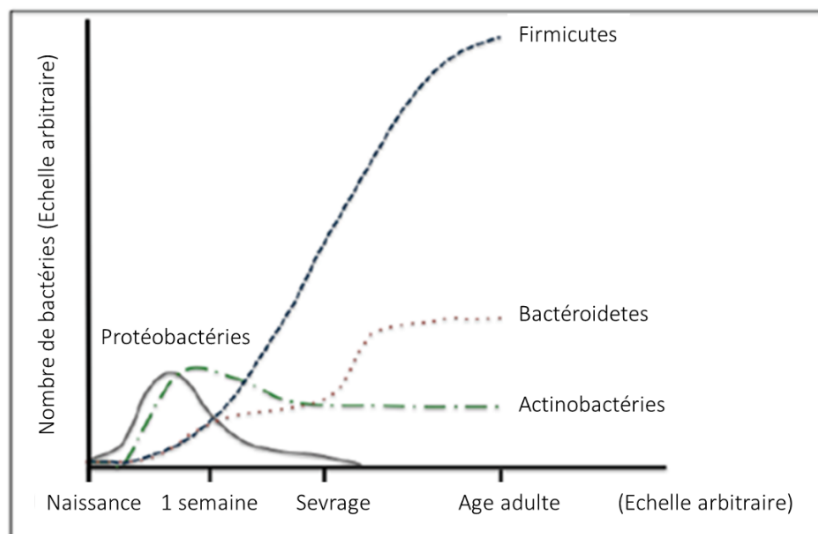


Figure 73: Evolution de la composition du microbiote intestinal dès la naissance, figure adaptée de Reinhardt et al [303].

Nos résultats confirment la présence d'une diversité microbienne dès les premiers jours de vie, bien que plus faible que celle observée dans des selles adultes. Au sein du méconium, il est décrit la présence de plus fortes proportions en Protéobactéries (71 % dans le méconium vs. 3,1 % dans les fèces adultes), et de plus faibles proportions de Bacteroidetes (2,5 % vs. 42,8 %), et de Firmicutes (20 % vs. 44,7 %) [114]. Les taux ont été calculés sur des échantillons de méconiums prélevés jusqu'à 48 heures suivant la naissance. Ainsi, afin de permettre de comparer ce que nous avons obtenu au sein de nos résultats, nous avons calculé les taux de représentation de ces phyla en partant du jour 1 et du jour 2 suivant la naissance. Nous retrouvons également des taux élevés et comparables en Protéobactéries avec 68 % de représentation, cependant nous observons des taux plus élevés en Bactéroidetes avec 19 % de représentation, ainsi qu'un plus faible taux des Firmicutes, avec 3 % de représentation au sein des deux premiers jours de vie.

Des études ont démontré que l'évolution des compositions métaboliques et protéiques du méconium sont directement associées avec l'augmentation des taux bactériens détectés [115], [264], [302], [305].

En effet, la principale activité métabolique du microbiote est liée à la fermentation des glucides et des protéines en aval du processus digestif [308]. Parmi les espèces bactériennes les plus détectées dans le méconium, Bittinger et al décrivent *E. coli*, *Enterococcus faecalis* et *Bacteroides vulgatus*. Au sein de leur étude, les auteurs ont indiqué qu'à partir de 16 heures (soit 960 minutes) suivant la naissance, de grandes différences dans la composition bactérienne étaient observées, impactant ainsi les profils protéiques et métaboliques, et supposant le début de l'activité métabolique de l'intestin [264]. La colonisation bactérienne se montre très variable d'un individu à un autre, lors des premiers jours de vie, et semble immédiatement s'enrichir à la naissance selon l'environnement rencontré (mode de naissance, mode d'allaitement, environnement de vie...) [2]. Cette colonisation évolue encore pendant des mois voire des années jusqu'à atteindre la composition ressemblant au microbiote adulte [303]. Les mises en relation de ces métadonnées seront notamment possibles au sein de notre cohorte EDEN, enregistrant de nombreuses informations sur la mère et l'enfant (Chapitre III. Partie II.A).

L'intégration de plusieurs niveaux d'informations générées par les approches "omiques" est nécessaire pour acquérir une image précise des organismes vivants. De précédentes études ont décrit le protéome ainsi que le microbiome du méconium, en revanche, les données en métabolomique sont moins nombreuses. Au sein de notre étude, nous avons pu décrire 230 métabolites, définissant le plus grand nombre de métabolites du méconium décrits à ce jour. Les études de la bibliographie rapportent uniquement des métabolites statistiquement pertinents pour une étude donnée (sans réaliser de cartographie globale), ou alors la grande majorité des études métabolomiques réalisées sur le méconium sont par des approches ciblées (cf, Tableau 5. Introduction. Chapitre III. Conclusion).

Les analyses statistiques ont mis en évidence des différences significatives de la composition globale du méconium au cours des trois premiers jours de vie. Des différences significatives étaient observées aussi bien en métabolites (dès le jour 1), qu'en bactéries et en protéines (dès le jour 1 ou significativement dès le jour 2). Les analyses multivariées de type ACP, ainsi que les représentations hiérarchiques et leur coloration ont permis d'apprécier la variabilité entre les individus.

Les résultats obtenus au sein de notre étude ainsi que des études bibliographiques décrites, montrent que le jour de prélèvement est un paramètre important à considérer lors de l'analyse des expositions *in utero* (notamment le cas de la cohorte EDEN). Dans ce contexte, il se montre plus judicieux de considérer des échantillons de méconiums les plus précoces, retenant des informations transmises durant la période gestationnelle, période non impactée par les expositions post-natales.

V. Conclusions du chapitre II et perspectives

L'analyse du méconium présente d'importants avantages en fournissant notamment la fenêtre d'observation la plus large de « l'exposome » fœtal. L'exploration de cette matrice est complexe mais présente un fort potentiel dans l'étude des transmissions de la mère vers le fœtus. Pour étudier cet exposome, nous avons considéré trois approches « omiques » : la métagénomique ciblée, la métaprotéomique et la métabolomique, réalisées sur 33 échantillons de méconiums excrétés par 11 nouveau-nés, durant les trois premiers jours post-partum.

Au niveau de l'analyse métabolomique, nous avons appliqué le protocole d'extraction des métabolites précédemment développé au sein du Chapitre I. Nous avons élargi la cartographie initiale en métabolites décrite chez les échantillons A et B, en incluant une méthode analytique complémentaire (C18-ESI⁺). La cartographie du méconium décrite à travers cette étude a permis d'explorer un large panel de classes chimiques avec 230 métabolites identifiés avec le plus haut niveau de confiance [210]. L'analyse LC/MS de 120 injections d'échantillons de méconiums, a démontrée la robustesse des méthodes de production des données mises en place. L'ensemble de ces résultats a également permis de rendre compte d'un protocole de préparation de l'échantillon robuste et d'une extraction des métabolites réalisée sans *a priori*.

La cartographie métabolique des échantillons de méconiums a été complétée par des analyses protéiques et bactériennes, définissant 922 protéines humaines, 11 espèces bactériennes par l'analyse métaprotéomique et 165 OTUs. Nous retrouvons au sein des analyses bactériennes des résultats concordants, avec la définition de quatre phyla majoritaires rejoignant les résultats de la littérature [115], [264].

Nous avons également pu apprécier la variabilité interindividuelle des échantillons analysés au sein des 3 approches, ainsi qu'une évolution dans la composition du méconium et ce, très tôt après la naissance. Les analyses statistiques réalisées ont permis d'observer une évolution dans la composition des métabolites dès 1440 minutes (24h post-partum), avec la définition de 67 métabolites significativement différents (sur les 230 identifiés) entre les groupes (< et > à 1440 minutes). En parallèle, nous observons une baisse globale du nombre de protéines humaines et de leurs abondances au cours du temps, corrélant avec la colonisation bactérienne et le début des processus digestif du nouveau-né [264], [302]. La colonisation bactérienne est révélée par l'évolution de la composition des protéines bactériennes et des OTUs dans le méconium permettant de mettre en évidence l'apparition et l'enrichissement progressifs en nouveaux phyla au cours des jours de prélèvements, et ce dès 1440 minutes. Ces résultats

viennent confirmer différentes études présentant des différences significatives dans les compositions des méconiums au cours des premiers jours de vie [115], [264], [302].

Nos résultats ont permis de mettre en évidence et de préciser la richesse de composition du méconium. Ces résultats ont également permis de révéler des temps seuils (par exemple de 1440 minutes pour le métabolome) au-delà duquel des différences significatives sont observées dans la composition du méconium. Ces différences peuvent traduire le début des processus métaboliques du nouveau-né et du microbiote, ainsi que l'apport de l'alimentation. Cette observation est à prendre en considération dans le prochain chapitre s'intéressant à l'exposome prénatal, et notamment à relier la composition du méconium aux informations sur la mère pendant la grossesse (cohorte EDEN).

CHAPITRE III : Analyse métabolomique des échantillons de méconium de la cohorte EDEN

I. Introduction

Au sein de ce chapitre, nous appliquerons la méthodologie d'analyse métabolomique globale développée et optimisée dans les chapitres précédents aux échantillons de méconium collectés au sein de la cohorte EDEN (Etude des Déterminants pré et post natals du développement et de la santé de l'Enfant). Cette étude nous permettra d'appréhender la variabilité interindividuelle de la composition métabolique globale du méconium, et d'étudier l'impact de différents facteurs environnementaux sur cette composition. Nous évaluerons ensuite la relation entre la composition métabolique du méconium et le devenir allergique des enfants.

Ces analyses seront par la suite reliées à des analyses complémentaires en métaprotéomique et en métagénomique ciblée (expériences en cours), comme réalisé pour les échantillons de la cohorte clinique. La combinaison de ces différentes approches donnera une image holistique permettant : i) de mettre en évidence des biomarqueurs d'expositions à certains facteurs environnementaux/maternels, ii) d'étudier leur relation avec le développement d'une allergie alimentaire dans l'enfance et, iii) d'approcher les mécanismes pathogénétiques sous-jacents.

L'étude métabolomique au sein de ce chapitre concerne 308 échantillons de méconiums issus de la cohorte française mère-enfant EDEN. L'analyse de l'ensemble de ces échantillons représente un défi analytique, afin de générer des données fiables et biologiquement pertinentes. En effet, il est décrit dans la littérature des soucis de répétabilité lors d'études de cohortes et suggèrent de diviser les analyses en différents lots, dans le but d'entretenir l'appareillage et d'éviter toute perte de signal [164]. D'après l'expérience du laboratoire, scinder l'analyse métabolomique en différents lots génère des problèmes lors de la fusion des données, tels que des décalages des temps de rétention ou des mesures des m/z . Ces décalages posent de grandes difficultés lors du traitement des données, compromettant les résultats obtenus. Par ailleurs, différentes publications suggèrent d'utiliser des « contrôles qualité » (QC) tout le long de la séquence analytique et à des intervalles réguliers, permettant de surveiller toute dérive et de la corriger [164], [224], [316]. Dans ce contexte, Dunn et al ont indiqué que chaque signal pouvait fluctuer différemment au sein de la séquence analytique, et qu'il était nécessaire de réaliser des corrections spécifiques à chaque signal et non une normalisation globale [164].

Pour les différentes raisons décrites au-dessus, et en considérant nos résultats montrant une bonne stabilité du système analytique lors d'une séquence constituée de 120 injections d'échantillons de

méconiums (Chapitre I, Partie III.A.4). Nous avons décidé de réaliser l'analyse de tous nos échantillons en un seul lot. Des QC seront introduits régulièrement tout au long de la séquence, permettant : i) de surveiller le signal au cours des analyses, ii) de filtrer les variables (élimination des variables présentant des CV trop importants), et iii) de réaliser la normalisation des données si celle-ci est nécessaire.

Nous présentons dans ce chapitre les premiers résultats obtenus suite à l'analyse métabolomique des échantillons de la cohorte EDEN. Ces analyses ont été réalisées au cours du stage de M2 de Vincent Marie, encadré par Blanche Guillon (IE au sein du LIAA) et moi-même. Les analyses métagénomiques ciblées et métagénomiques sont réalisées en collaboration avec l'institut Micalis de l'INRAE de Jouy en Josas et le laboratoire LI2D du CEA de Marcoule respectivement.

II. Matériels et méthodes

A. Description de la cohorte EDEN

La cohorte française mère-enfant EDEN est une cohorte prospective bi-centrique constituée de 2002 couples mères-enfants. Les mères participant à l'étude ont été recrutées à leur troisième mois de grossesse dans deux villes, Nancy et Poitiers. Les trois principaux objectifs de cette cohorte sont d'étudier l'impact/le rôle de différents facteurs sur le développement et la santé globale de l'enfant :

- i. Etude de l'impact de l'alimentation de la mère pendant la période gestationnelle.
- ii. Etude de l'impact des expositions environnementales (polluants, toxiques, ...) durant la période gestationnelle et au cours des premières années de vie.
- iii. Elucidation du rôle des facteurs socio-économiques/psycho-affectifs et leurs interactions avec les facteurs biologiques.

Afin de permettre de répondre à ces questions, des questionnaires ont été fournis aux parents participant à l'étude, permettant d'enregistrer un certain nombre d'informations sur la mère. Ces informations incluent : son alimentation (avant et pendant la grossesse), son lieu et son mode de vie, son état de santé, sa situation socio-économique... Certaines informations sont également collectées auprès du père, incluant sa situation socio-économique, son état de santé et son mode de vie. Les parents ont ensuite fourni des informations quant à la croissance de l'enfant, sa santé, son comportement et son développement à travers des questionnaires réguliers. Différentes informations ont également été enregistrées lors de visites dans le centre de recrutement, telles que le poids, la taille, la pression artérielle, la capacité respiratoire et les fonctions cognitives de l'enfant à l'âge de 5 ans.

En parallèle, différents prélèvements biologiques ont été recueillis en maternité, incluant des urines, du sang et du colostrum des mères, du méconium, des cheveux, de la salive et du sang de cordon du nouveau-né. Les protocoles de prélèvements et de conservation étaient définis et communs aux deux centres. Ces échantillons ont pu servir dans de nombreuses études, telles que l'analyse de l'environnement des mères sur le développement de l'enfant [22] et l'apparition d'asthme [23], ou encore l'analyse de l'impact de la consommation de caféine ou de tabac [317], [318] ou du type d'alimentation de la mère sur la croissance de l'enfant [319].

Au sein de cette cohorte, nous avons pu avoir accès aux prélèvements de colostrum et de méconium appartenant aux mêmes couples mères-enfants. Les échantillons de colostrum ont permis de réaliser une description des métabolites, des lipides, des oligosaccharides, et des facteurs immunitaires et de croissance présents dans le lait précoce. Ces échantillons ont également permis d'étudier leur

association avec différents facteurs environnementaux des mères, et de les relier au développement de l'allergie alimentaire dans l'enfance (thèse soutenue par Mikail Berdi en 2017, [41]).

Les échantillons de méconium serviront dans cette présente étude, et sont décrits ci-dessous.

B. Echantillons de méconiums de la cohorte EDEN

Au sein des 2002 couples mères-enfants de la cohorte française EDEN, nous avons pu avoir accès à 308 échantillons de méconiums (dont un seul prélèvement est collecté par nouveau-né) (tableau 15).

Initialement, un sous-échantillonnage aléatoire de 300 couples mères-enfant (n=150/centre), pour lesquels le méconium et le lait maternel avaient été collectés, a été réalisé au sein de la cohorte EDEN. Selon la disponibilité des échantillons de méconium au sein de cet échantillonnage, 257 échantillons de méconiums ont été finalement inclus, constituant notre « cohorte initiale ». Ces échantillons permettront de mettre en relation la composition du méconium avec les facteurs environnementaux et de santé des mères. Parmi les nouveau-nés ayant excrété ces méconiums, 25 ont développé une allergie alimentaire (AA) au cours de leur 5 premières années de vie. Les cas d'allergies alimentaires reposent sur la déclaration par les parents ou lors d'un diagnostic d'allergie alimentaire posée par le pédiatre. Aucun dosage d'IgE, de tests cutanés ou de tests de provocations n'ont confirmé ce diagnostic [271].

Afin d'avoir une puissance statistique suffisante pour étudier l'association entre la composition du méconium et le développement d'une AA dans l'enfance. Nous avons rajouté tous les échantillons de méconiums issus d'enfants qui développeront une AA et qui étaient disponibles dans la cohorte (soit 51 échantillons supplémentaires : « case-cohort »). Le centre de Nancy disposait de plus d'échantillons disponibles dans la biobanque : l'ensemble des cas rajoutés proviennent de ce centre.

Au total, nous avons accès à 308 échantillons de méconiums de la cohorte EDEN. La majorité des échantillons ont été collectés au cours du premier et du deuxième jour post-partum, ce qui est rassurant au vu de nos conclusions précédentes sur l'évolution de la composition du méconium au cours des premiers jours de vie. Un aliquot de ces échantillons, stockés dans les biobanques des centres de recrutements, avaient été reçus au sein du laboratoire et conservés à -80°C jusqu'à leur analyse. L'ensemble des informations concernant les sous-cohortes de notre étude sont décrites dans le Tableau 15.

Tableau 15 : Description des nouveau-nés et des échantillons de méconiums étudiés, issus de la cohorte française mère-enfant EDEN.

Cohorte de l'étude (308 échantillons)	Cohorte initiale (257 échantillons)		« Case-cohort » (51 échantillons)
Centre	Nancy	130 (50,6 %)	51 (100 %)
	Poitiers	127 (49,4 %)	0 (0 %)
Sexe	Fille	125 (48,6 %)	21 (41,2 %)
	Garçon	132 (51,4 %)	30 (58,8 %)
Jours de prélèvements	J1	142 (55,3 %)	26 (51 %)
	J2	38 (14,8 %)	15 (29,4 %)
	J3	23 (8,9 %)	5 (9,8 %)
	J4	7 (2,7 %)	4 (7,8 %)
	J5	4 (1,6 %)	- (0 %)
	J6	1 (0,4 %)	- (0 %)
	Non défini	42 (16,3 %)	J11 : 1 (2 %)
Allergie alimentaire	Oui	25 (9,7 %)	51 (100 %)
	Non	232 (92,3 %)	0 (0 %)

C. Acquisition des données

L'analyse des échantillons de méconiums de la cohorte EDEN a été réalisée selon le workflow décrit dans le chapitre II, partie II.A.1, et a été réalisée dans les deux conditions analytiques HILIC-ESI⁻ et C18-ESI⁺. Les séquences analytiques exécutées comprennent plus de 350 injections, incluant 308 échantillons biologiques, 32 QC, 12 dilutions de QC, 7 blancs analytiques (5 en début et 2 en fin de séquence) et 6 injections de conditionnement de la colonne dans les conditions HILIC-ESI⁻. Lors de l'analyse de cohortes, nous utilisons une pompe annexe, permettant de faire circuler un mélange eau/ACN (1:1 v/v) dans la source lors de l'étape de rééquilibrage de la colonne chromatographique, limitant ainsi l'encrassement de la source [41]. L'ensemble des protocoles, des réactifs et des appareillages utilisés sont décrits dans le chapitre II. Partie II.A.

Les échantillons ont été introduits dans la séquence de façon aléatoire, afin d'éviter les facteurs confondants liés à des paramètres physiologiques, environnementaux ou liés à la collecte (sexe, cas AA, centre ou jour de prélèvement... etc) [320]. La randomisation a été réalisée par la fonction « =ALEA » d'Excel à partir de l'identifiant fourni par les biobanques.

D. Traitement des données

L'ensemble des étapes depuis l'extraction à l'annotation des variables et aux analyses statistiques sont identiques que celles décrites dans le Chapitre II. Partie II.B.3.

III. Résultats expérimentaux et discussion

A. Traitement XCMS des données générées dans les conditions analytiques HILIC-ESI⁻ et C18-ESI⁺ et suivi de la stabilité du système analytique

1. Nombre de variable extraites et suivi de la stabilité du système analytique

Nous travaillons dans un premier temps uniquement sur les données obtenues sur les échantillons de la cohorte initiale, soit 257 échantillons de méconium. L'ensemble des 308 échantillons de méconium seront utilisés pour répondre à la question de l'allergie alimentaire (« case-cohort », étudiée à la fin de ce chapitre).

Le prétraitement XCMS des données générées pour les échantillons de la cohorte initiale utilise les mêmes paramètres que ceux du Chapitre II (décrits en Annexe 8 et en Annexe 9). Cette étape a permis d'extraire 6225 et 12029 variables, respectivement dans les conditions analytiques HILIC-ESI⁻ et C18-ESI⁺. Afin de suivre la performance de ces analyses au cours de la séquence d'injection, nous avons tracé la somme des intensités de l'ensemble des variables extraites et avons observé que, dans les conditions HILIC-ESI⁻, 77 % des variables détectées dans le QC avaient des intensités montrant un coefficient de variation (CV) inférieur à 30 %. Dans les conditions C18-ESI⁺, cela ne concernait que 48 % des variables détectées dans le QC.

Nous avons ensuite supprimé une partie des variables non pertinentes par l'application de 2 filtres, reposant sur : i) les aires des variables détectées dans les échantillons de QC/celles détectées dans les blancs analytiques et ii) le coefficient de corrélation des intensités des variables avec le facteur de dilution dans les QC dilués. A cette étape, le troisième filtre portant sur les CV des intensités des variables au sein des QC n'est pas réalisé, au vu des variations observées notamment dans les conditions C18-ESI⁺. Ces deux filtres ont permis de retenir 4447 et 7727 variables dans les conditions HILIC-ESI⁻ et C18-ESI⁺ respectivement. A la suite de ces filtres, nous observons 88 % et 67 % des variables des QC ayant un CV < 30 % au sein de ces deux conditions respectives.

Lors d'études métabolomiques sur des cohortes, une normalisation des intensités basées sur le suivi des QC inter ou intra batchs permet d'assurer une bonne répétabilité des mesures pour l'ensemble de la séquence [316]. Au laboratoire, nous réalisons une correction du signal lorsque moins de 70 % des variables au sein des QC ont un CV inférieur à 30 %. Ce qui est le cas dans les conditions C18-ESI⁺. Dunn et al ont rapporté une correction du signal utilisant une fonction robuste de régression locale (LOESS, de l'anglais : LOcally wEighted Scatterplot Smoother) [164]. Cette fonction permet de minimiser la décroissance du signal au cours de la séquence en corrigeant de manière sélective et spécifique les ions

détectés. Dans les conditions analytiques C18-ESI⁺, la normalisation LOESS réalisée sur W4M de Galaxy [218], a permis d'augmenter de 67 à 85 % le nombre de variables ayant un CV < 30 % au sein des QC, (Figure 74). Cette correction n'a pas été appliquée aux données générées dans les conditions analytiques HILIC-ESI⁻, celles-ci montrant une bonne qualité au niveau des CV des variables au sein des QC. En effet, cette correction peut parfois générer des surestimations ou sous-estimations des intensités de certaines variables, comme observé lors de précédentes études au laboratoire [41]. Celle-ci doit donc être appliquée avec parcimonie. Lors de l'étude des laits maternels de la même cohorte EDEN, la fonction LOESS a également dû être appliquée pour les conditions analytiques C18-ESI⁺, et a permis d'améliorer les coefficients de variations des variables des QC, passant de 32 à 50 % des variables ayant un CV < 30 % [41].

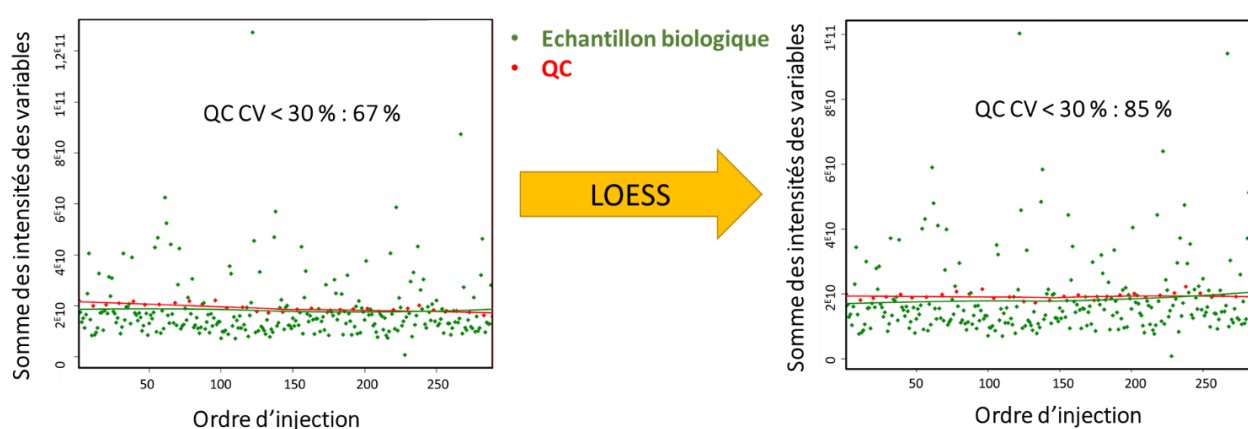


Figure 74 : Normalisation du signal « LOESS » réalisé sur les échantillons de QC et les échantillons biologiques (cohorte initiale), analysés dans les conditions C18-ESI⁺.

Nous avons à la suite réalisé le troisième filtre sur les variables (au sein du CV dans les QC), permettant de finalement retenir 4208 et 6544 variables « filtrées », respectivement dans les conditions analytiques HILIC-ESI⁻ et C18-ESI⁺.

2. Nécessité de modifier les filtres analytiques classiquement utilisés pour répondre à notre question scientifique

Afin de rendre compte de la pertinence du nombre de variables extraites au sein de nos données, nous les avons quantitativement comparées aux résultats précédemment obtenus au sein du chapitre I (2 échantillons de méconium A et B), et du chapitre II (33 échantillons de méconium « cohorte clinique »). Nous avons observé une baisse du nombre de signaux extraits avec l'augmentation du nombre d'échantillons analysés (Figure 75). Nous observons ainsi dans les conditions analytiques HILIC-ESI⁻, un passage de 14863 variables extraites dans les échantillons A et B, à 9274 variables extraites dans les échantillons de la cohorte clinique, et finalement à 6225 variables dans les échantillons de la cohorte

EDEN. Le nombre de données extraites est par contre comparable pour les échantillons de la cohorte clinique et de la cohorte EDEN dans les conditions analytiques C18-ESI⁺. Pour rappel, les échantillons A et B n'ont été analysés que dans les conditions HILIC-ESI⁻, lors de la mise en place de la méthodologie de préparation de l'échantillon.

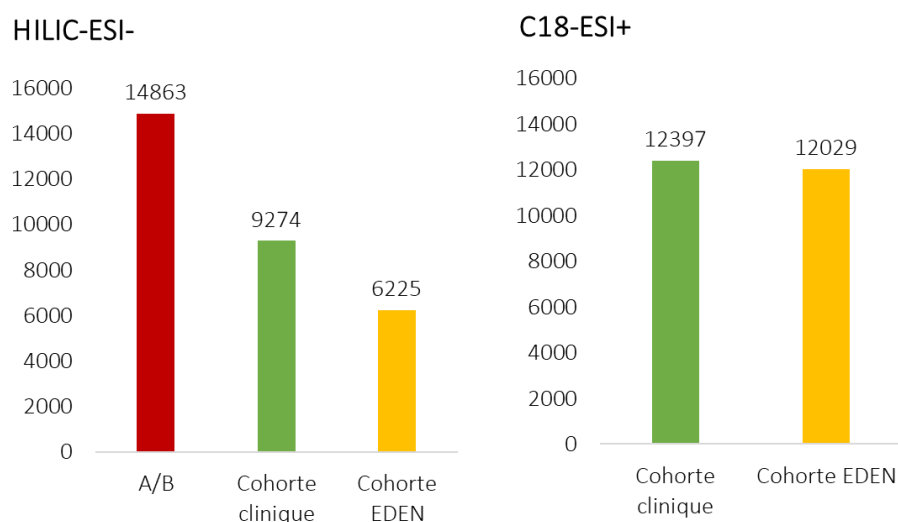


Figure 75 : Comparaison du nombre de variables extraites par XCMS dans les conditions analytiques HILIC-ESI⁻ (graphe de gauche) et C18-ESI⁺ (graphe de droite), obtenues sur les deux échantillons A/B (rouge), sur les 33 échantillons de la cohorte clinique (vert), et sur les 257 échantillons de la cohorte EDEN (jaune).

Etant donné que la cohorte EDEN contient un grand nombre d'échantillons (plus de 250), une forte variabilité des signaux détectés est attendue, aussi bien qualitativement (nombre de signaux détectés) que quantitativement (intensités des signaux). Ceci est en lien avec notre hypothèse de recherche, visant à analyser la variabilité interindividuelle de la composition du méconium en lien avec les facteurs environnementaux. Cependant, cette forte variabilité se traduit par un moindre nombre de variables retenues au moment du prétraitement des données par XCMS. Tout d'abord, l'effet de dilution au sein des QC des variables faiblement concentrées dans le méconium impactera leur sélection parmi les variables extraites par XCMS. En effet, celles-ci risquent d'être éliminées par le filtre appliqué sur le ratio des intensités dans les échantillons QC/blanc analytique. De plus, un filtre automatique est appliqué lors du prétraitement des données XCMS, incluant la fonction « minfrac » (de « Minimum Fraction »), fixée par défaut à une valeur de 0,7. Ce paramètre conduit à n'extraire que les variables dont les signaux sont détectés dans au moins 70 % des échantillons du même groupe (groupe de QC, de blancs analytiques ou d'échantillons biologiques). Ce paramètre peut avoir un impact important lorsque le nombre d'échantillons est conséquent et qu'il existe des variables détectées uniquement dans certains échantillons. De plus, ce paramètre intervient à deux reprises au cours du prétraitement XCMS, lors des

deux étapes du « group », workflow détaillé en Annexe 7. Ce paramètre peut cependant être modifié par l'opérateur. Nous avons par exemple utilisé des valeurs plus faibles pour ce paramètre, tel qu'un « minfrac » à 0,3 utilisé lors de l'analyse des échantillons de laits maternels de la cohorte EDEN [41]. Cependant, cette valeur conduisait à retenir de nombreuses variables dont les signaux se confondaient à ceux issus du bruit de fond, raison pour laquelle elle a été modifiée par la suite. Nous pensons qu'il serait intéressant de reconsidérer la valeur de ce paramètre notamment dans l'analyse de grandes cohortes, comme le cas de notre étude.

Nous avons ainsi testé différentes valeurs du paramètre « minfrac », en le fixant à : 0 ; 0,3 0,5 et 0,7. Ces tests ont été réalisés sur les données obtenues en conditions analytiques HILIC-ESI⁻ (i.e. les données les plus impactées par le nombre d'échantillons analysés). Une augmentation progressive du nombre de variables extraites est observée au fur et à mesure où la valeur du « minfrac » est diminuée, jusqu'à atteindre près de 45000 variables extraites pour une valeur de « minfrac » fixée à 0 (Figure 76). Cependant, comme indiqué précédemment, une faible valeur pour ce paramètre, notamment fixée à 0, génère beaucoup de redondances ainsi que beaucoup de signaux non pertinents issus du bruit de fond, qui rallongent de façon non acceptable le temps de traitement des données en aval.

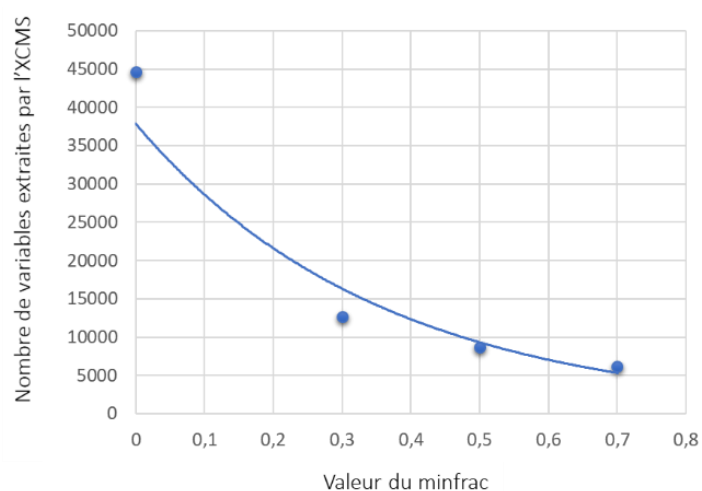


Figure 76 : Nombre de variables extraites lors du prétraitement des données dans XCMS, en fonction des valeurs de « minfrac » testées .

Il apparaît donc que, pour répondre à notre question scientifique liée à l'exposome, l'ensemble du processus de traitement des données par XCMS et post-XCMS utilisé actuellement au laboratoire devrait être réétudié. Une réflexion approfondie est nécessaire, ainsi que la réalisation de différents tests et optimisations (par exemple, tester des valeurs différentes de « minfrac » aux étapes 1 et 2 du workflow dans XCMS). Ceci, en évaluant la pertinence des variables retenues à chaque test et en mettant en place une stratégie robuste pour traiter et filtrer ces données en aval de l'XCMS.

N'ayant pas eu le temps de réaliser ces optimisations à la fin de mon travail de thèse, nous avons continué les analyses des données en maintenant le facteur du « minfrac » comme utilisé habituellement au laboratoire, c'est-à-dire à 0,7. Ceci afin de comparer les résultats avec ceux déjà obtenus avec notre petite « cohorte clinique », puis de répondre à la question liée à l'allergie alimentaire (analyse de la « case-cohort »).

B. Annotation des variables

1. Analyse comparative du nombre de variables annotées au sein de la cohorte clinique et de la cohorte EDEN

Au sein des cohortes clinique et EDEN (cohorte initiale), les mêmes outils et procédures analytiques ont donc été utilisés (version XCMS.3.2.1, méthode d'annotation « mlomic », conditions d'analyses HILIC-ESI⁻ et C18-ESI⁺). Nous avons ainsi pu comparer les annotations obtenues au sein de ces deux cohortes. Dans ces conditions, 274 et 148 métabolites ont été annotés avec redondance dans les échantillons de méconium de la cohorte EDEN, respectivement pour les conditions analytiques HILIC-ESI⁻ et C18-ESI⁺. A la suite du filtre manuel des redondances, nous retenons respectivement 228 et 137 métabolites annotés. Ce nombre de métabolites annotés est très comparable à ce que nous avons précédemment obtenu : soit 224 vs. 228 en conditions analytiques HILIC-ESI⁻ pour la cohorte clinique vs. la cohorte EDEN respectivement (Figure 77.A et Figure 77.C), et 149 vs. 137 en conditions analytiques C18-ESI⁺ (Figure 77.B et Figure 77.D), ceci malgré un nombre de variables extraites moins important pour la cohorte EDEN. En regroupant les annotations au sein des conditions HILIC-ESI⁻ et C18-ESI⁺, nous définissons au total une cartographie de 365 métabolites, incluant 64 métabolites détectés dans les deux conditions analytiques, soit 301 métabolites uniques.

Ce nombre de métabolites annotés est comparable à celui observé au cours d'autres études menées au laboratoire sur différentes cohortes et différentes matrices, avec par exemple 325 métabolites annotés dans le lait maternel [41], ou 223 métabolites annotés dans des sérums humains [198]. Nous nous intéresserons dans la suite à évaluer qualitativement ces annotations et à les comparer à ce qui a précédemment été décrit au sein de la cohorte clinique.

Cohorte clinique

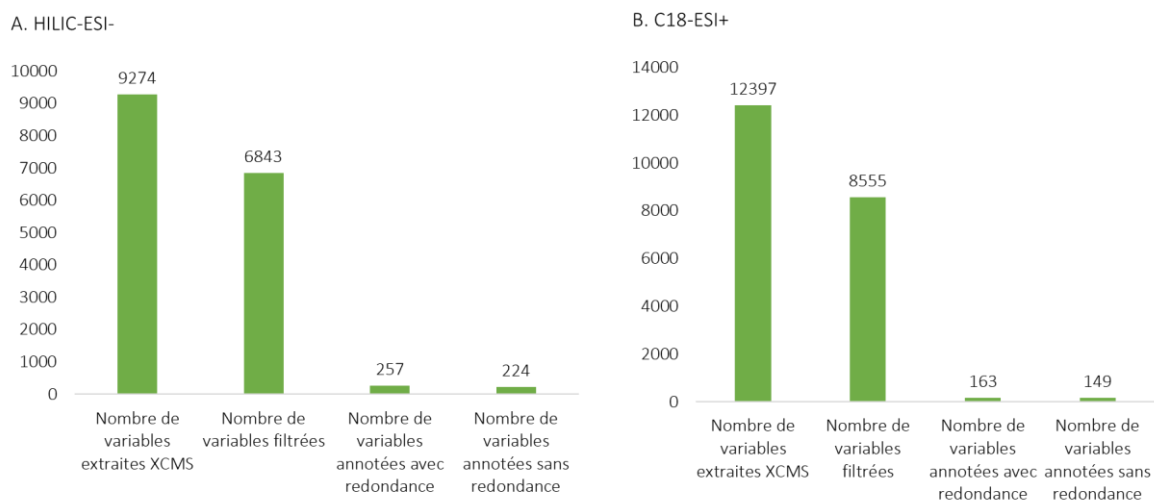


Figure 77 : Nombre de variables extraites, filtrées, annotées avec redondance et sans redondance au sein de la cohorte clinique (A-B) et de la sous-cohorte initiale EDEN (C-D), dans les conditions analytiques HILIC-ESI⁻ (panels de gauche) et C18-ESI⁺ (panels de droite).

2. Analyse qualitative comparative des variables annotées au sein des échantillons de la cohorte clinique et de la cohorte EDEN

Nous avons ensuite comparé les annotations obtenues grâce aux échantillons de la cohorte initiale de EDEN avec ce que nous avons précédemment détecté dans les échantillons de la cohorte clinique. Nous observons 67 métabolites communs dans les conditions C18-ESI⁺ et 175 métabolites communs dans les conditions HILIC-ESI⁻. Soit au total plus de 60 % de métabolites communs aux deux cohortes, dont plus de la moitié a été formellement identifiée suite à des expériences MS/MS. De nombreux nouveaux

métabolites (59) sont détectés au sein de la cohorte EDEN, incluant l'acide urique, le panthénol ou encore le rutin (liste en Annexe 16). Comme précédemment observé, parmi les familles chimiques les plus représentées, nous décrivons les aminoacides, peptides et dérivés (29 %), suivis par les familles des carbohydrates et des lipides (chacun représentant 14 % des variables) (Figure 78.A). Ces résultats sont comparables à ce qui a précédemment été décrit au sein de la cohorte clinique, avec une représentation de la famille des acides aminés, peptides et analogues à 27 %, des lipides à 18 % et des carbohydrates à 14 % (Chapitre II. Partie III.A.4). Nous définissons également d'autres familles moins représentées telles que les acides organiques et dérivés (10 %), ou les nucléosides, nucléotides et dérivés (5 %). L'ensemble de ces annotations permettent encore une fois de rendre compte d'un protocole d'extraction des métabolites réalisé sans *a priori*, avec la couverture de composés présentant des propriétés physico-chimiques très différentes.

Nous observons à partir de la Figure 78.B que la grande majorité des métabolites annotés sont des métabolites endogènes (90 %). Les métabolites d'origines alimentaire ou des médicaments représentent chacun 4 % du nombre de variables annotées, et les métabolites microbiens en représentent 2 %. Ces résultats sont cohérents avec notre précédente cartographie, décrivant 87 % de métabolites endogènes, et des taux comparables de métabolites d'origine alimentaire et médicamenteuse.

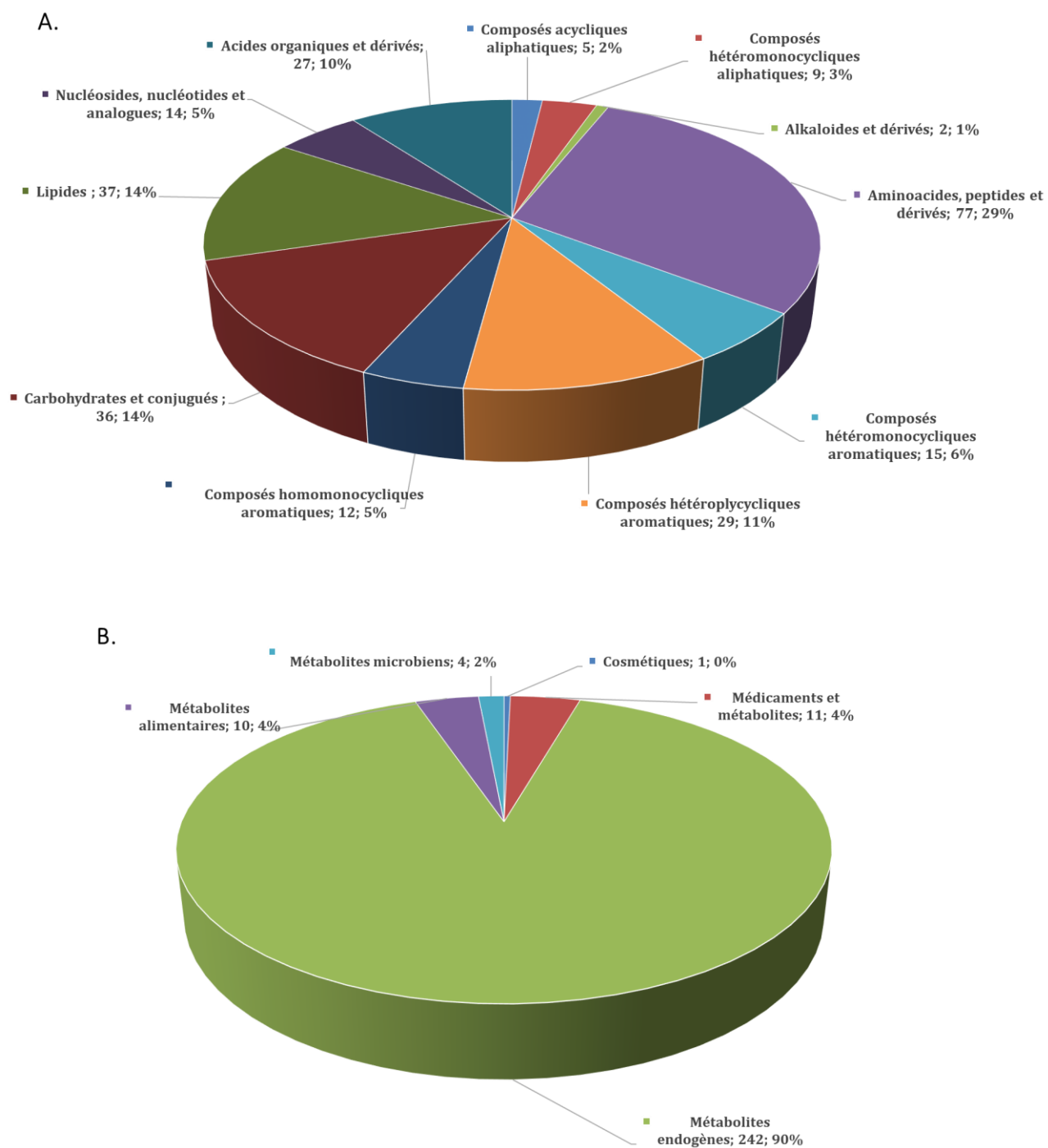


Figure 78 : Description des familles (A) et des origines (B) des métabolites annotés au sein des échantillons de méconium de la cohorte initiale EDEN, dans les deux conditions analytiques HILIC-ESI⁻ et C18-ESI⁺. Les métabolites qui n'ont pas pu être attribués, ne sont pas représentés sur ces figures.

C. Analyses statistiques des données de la cohorte EDEN : effet du temps, du centre et du sexe

Nous avons par la suite analysé l'évolution de la composition des méconiums au cours des jours de prélèvements (du jour 1 au jour 6 suivant la naissance). L'analyse de la distribution des données en fonction d'autres paramètres tels que le centre de prélèvement ou le sexe a également été menée. Pour cela, nous avons utilisé différents outils statistiques uni- et multivariés.

1. Analyses non supervisées réalisées sur l'ensemble des variables extraites au sein des conditions HILIC-ESI⁻

Nous avons dans un premier temps étudié la structuration du jeu de données globale, puis en fonction des facteurs jour, centre de collecte et sexe. Les résultats des analyses non supervisées donnant des résultats similaires dans les conditions HILIC-ESI⁻ et C18-ESI⁺, nous présentons dans cette première partie uniquement les résultats obtenus dans les conditions analytiques HILIC-ESI⁻.

Une analyse multivariée non supervisée de type ACP est réalisée en considérant les 4788 variables retenues après les étapes de filtres dans les conditions HILIC-ESI⁻ (Figure 79.A). Les deux premières composantes t1 (34 %) et t2 (13 %) de l'ACP représentent 47 % de la variabilité totale du jeu de données. Une chute importante de l'inertie totale est ensuite observée, avec seulement 4 % de la variance résiduelle portée par la troisième composante. De ce fait, seules les deux premières composantes seront analysées par la suite.

La coloration des échantillons par jour de prélèvement permet de mettre en évidence une évolution non-linéaire de la composition globale des échantillons en fonction du temps (Figure 79.B). Globalement, une courbe en forme de « C » est représentée, comme précédemment observée avec les données de la cohorte clinique. Quatre échantillons apparaissent comme extrêmes dans l'ACP (X222, X17, X198 et X254), correspondant à différents jours de prélèvements (respectivement J2, J3 et J5, le X254 n'est pas renseigné).

Suite à ce qui a été observé dans le chapitre précédent concernant un temps « seuil » de 1440 minutes discriminant la composition globale des méconiums, nous avons représenté l'ACP en colorant les échantillons collectés dans les 24h post-partum (J1 : rouge) *versus* les autres temps (> J1 : bleu) (les jours de prélèvements non connus sont représentés en vert) (Figure 79.C). Nous observons un effet jour (J1 et >1) beaucoup moins marqué comparativement à ce qui était observé sur la cohorte clinique. En effet, une grande partie chevauchante est observée entre les deux populations.

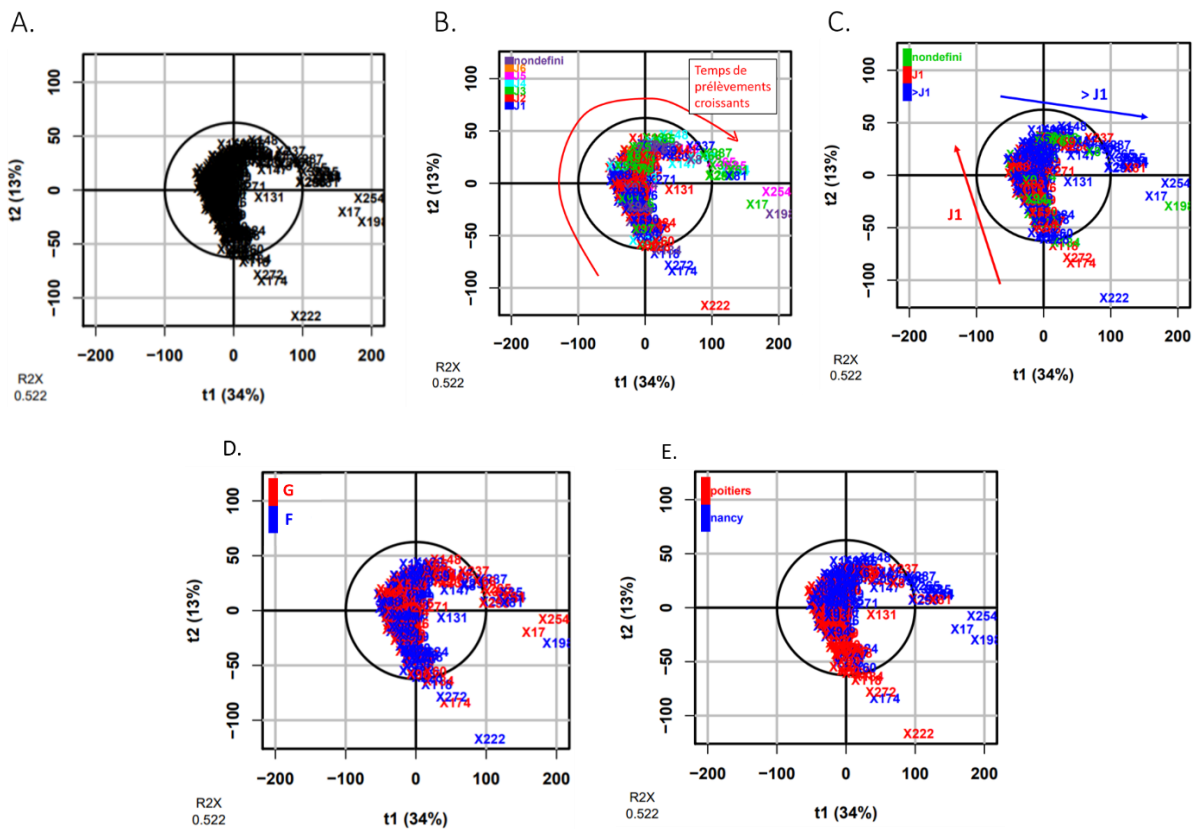


Figure 79 : Représentation des deux premières composantes de l'ACP réalisée sur les 4788 variables obtenues dans les conditions analytiques HILIC-ESI⁻ (A). Les échantillons sont colorés selon leur jour de prélèvement : (B) : J1, J2, J3, J4, J5 et J6 ; (C) : J1 vs. > J1. Le jour de prélèvement de certains échantillons n'est pas renseigné (« nondefini »). (D) : le centre de prélèvement (A, Nancy en bleu et Poitiers en rouge) et (E) : le sexe du nouveau-né (B, fille en bleu, et garçon en rouge).

La représentation de l'ACP sur le genre (Figure 79.D) ne montre pas d'effet structurant du sexe du nouveau-né sur la composition métabolique et sur la distribution des données. Nous avons ensuite analysé l'effet du centre sur cette distribution, représenté en Figure 79.E. La représentation de l'ACP en fonction du centre de prélèvement se montre très marquée, avec la distinction des échantillons issus des centres de Nancy (bleu) et de Poitiers (rouge).

Cependant, une sur-représentation d'échantillons collectés au jour 1 est observé à Poitiers versus Nancy (47/127, vs 33/130 échantillons). Nous avons ainsi représenté l'ACP uniquement sur les échantillons collectés au J1 au sein de ces deux centres (Figure 80.A). La sélection de ces échantillons conduit à une discrimination visuelle moindre du centre de prélèvement. L'effet « centre » observé sur l'ensemble des échantillons semble donc au moins partiellement lié à un effet « temps de prélèvement ».

En parallèle, nous avons vérifié si l'effet jour ne serait pas également lié à un « effet centre ». Nous avons ainsi représenté l'ACP de chacun des centres séparément, et coloré les échantillons en fonction des jours de prélèvements. Pour le centre de Nancy, nous observons un regroupement des échantillons

en fonction du temps de prélèvement (J1 (rouge) versus >1 (bleu), Figure 80.B). A l'inverse, ces regroupements ne sont pas observés pour le centre de Poitiers (Figure 80.C). L'analyse de l'effet de chacun de ces deux facteurs sera analysé plus finement.

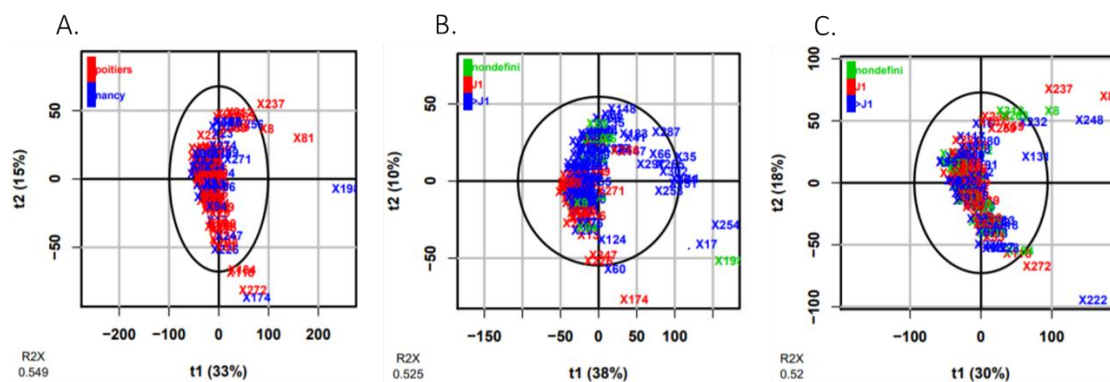


Figure 80 : Représentation des deux premières composantes de l'ACP réalisée sur les 4788 variables obtenues dans les conditions analytiques HILIC-ESI⁻. A. Représentation des échantillons collectés à J1 dans les centres de Nancy (bleu) et de Poitiers (rouge). B-C : Représentation des échantillons avant ou après 24h à Nancy (B) et Poitiers (C). J1 (rouge), > J1 (bleu), les échantillons dont le temps de collecte n'est pas renseigné sont représentés en vert.

2. Analyses non supervisées réalisées sur les métabolites annotés

Ces mêmes analyses ont été réalisées en ne considérant cette fois que les 301 variables uniques annotées, obtenues en combinant les deux conditions analytiques HILIC-ESI⁻ et C18-ESI⁺. L'analyse en composantes principales sur ce jeu de données a permis de résumer 43 % de la variabilité totale sur les deux premières composantes t1 (25 %) et t2 (18 %) (Figure 81.A). Une chute de l'inertie est observée à partir de la troisième composante, qui explique seulement 6 % de la variance totale. Seules les composantes 1 et 2 seront analysées par la suite. Trois des quatre échantillons extrêmes précédemment identifiés apparaissent également comme extrêmes au sein de ce jeu de données réduit (X222, X198 et X254, issus de différents jours de prélèvements). Un autre échantillon extrême, le X281, collecté au jour 2 est identifié.

Nous avons ensuite représenté l'ACP par jour de prélèvement (allant du jour 1 au jour 6 suivant la naissance, Figure 81.B). Nous observons une évolution temporelle plus marquée de la composition métabolique des échantillons que celle mise en évidence précédemment sur l'ensemble des données extraites. En effet, on observe l'évolution de la courbe en forme de « C » allant du bleu (J1), aux rouge, vert et bleu ciel respectivement pour J2, J3 et J4. Les échantillons collectés à J1 (< 1440 min) se distinguent du reste des échantillons (regroupés en bleu, à gauche de l'axe t1). Nous avons ainsi coloré l'ACP en fonction du temps de collecte des échantillons, en considérant un seuil à 24h (échantillons

collectés avant ou après 1440 minutes, Figure 81.C). Cette coloration montre une bonne distinction des deux groupes d'échantillons (< J1 en bleu et > J1 en rouge).

Nous avons ensuite coloré l'ACP en fonction du centre de prélèvement (Figure 81.D) et du sexe du nouveau-né (Figure 81.E). Les représentations correspondantes montrent des résultats similaires à ce qui a été observé précédemment sur l'ensemble du jeu de données. Il faut cependant noter que l'effet centre et temps à Poitiers sont relativement liés. De la même façon que précédemment, aucun effet du genre n'est visible sur l'ACP colorée correspondante.

En globalité, les 301 variables annotées semblent plutôt bien résumer l'ensemble du jeu de données. A titre comparatif, des effets semblables ont été observés lors de l'analyse multivariée non supervisée des laits maternels issus de cette même cohorte EDEN. Un effet temps de prélèvement marqué était observé sur la composition métabolomique globale des laits maternels, mais également sur les concentrations d'autres facteurs bioactifs tels que les cytokines et facteurs de croissance [41].

Ces différentes représentations permettent de rendre compte de deux effets principaux sur la structuration sans *a priori* des données : un effet temps et un effet centre, observés aussi bien sur l'ensemble des données (soit 4788 variables) que sur le jeu de données réduit (soit 301 variables), qu'il sera important à considérer dans la suite.

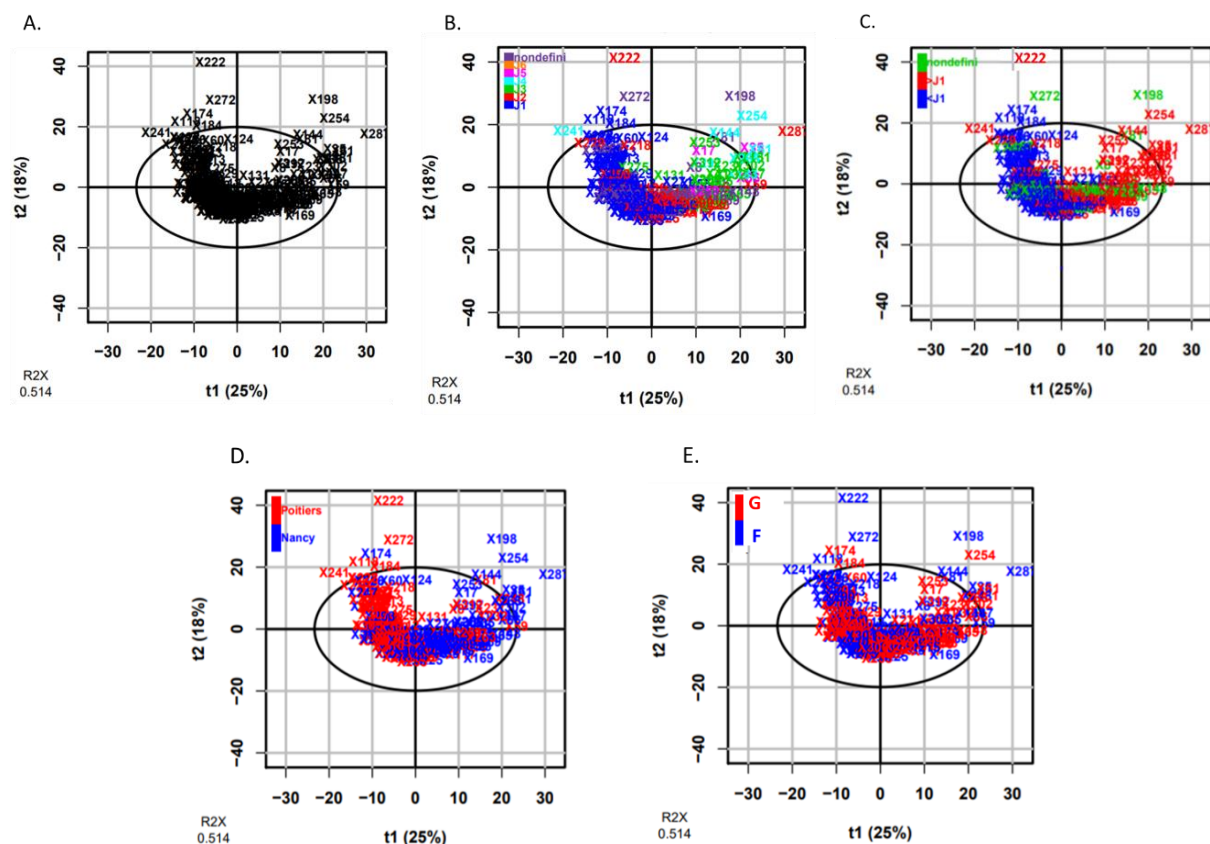


Figure 81 : Représentation des deux premières composantes de l'ACP réalisée sur les 301 variables annotées au sein des échantillons de la cohorte EDEN. A. Non colorée. B. Les échantillons sont colorés

selon le jour de prélèvement : J1, J2, J3, J4, J5 et J6. C. Selon un seuil de temps de prélèvement à J1 (< ou > J1). D. Selon le centre de prélèvement (Poitiers et Nancy). E. Selon le sexe du nouveau-né. Les échantillons pour lesquels le jour de prélèvement n'a pas été renseigné sont notés « nondefini ».

3. Analyses supervisées de l'effet temps et de l'effet centre sur la composition métabolique des méconiums

a) Analyse de l'effet temps

Au vu des résultats précédemment décrits, nous avons réalisé des analyses supervisées de type PLS-DA sur les 301 variables annotées, en définissant comme variable explicative le temps (< 1440 min *versus* > 1440 min) ou le centre de prélèvement (Nancy *versus* Poitiers).

Un modèle est construit selon le facteur temps, mais ne génère pas de représentation graphique (une seule composante dans le modèle). Cependant, le graphe des pR2Y et de pQ2 généré a permis de rendre compte d'un modèle juste (Annexe 17), nous permettant de considérer les valeurs de VIP calculées par ce dernier. En considérant une valeur de VIP > 1,5, 46 variables sont discriminantes entre les deux intervalles de temps considérés (Liste en Annexe 18). Parmi ces composés, 10 sont plus abondants dans les temps précoces (< J1), incluant des voies métaboliques telles que le métabolisme de l'arginine et de la proline, tandis que 36 sont plus abondants dans les temps tardifs (> J1), incluant les voies métaboliques de la valine, des purines ou du galactose. Ces résultats rejoignent ce qui a précédemment été décrit dans la cohorte clinique ainsi que dans la littérature, rapportant un enrichissement de la composition métabolique dans le méconium au cours des premiers jours de vie [115], [256], [264]. De façon intéressante, parmi ces 46 métabolites, nous retrouvons 11 métabolites communs à ceux identifiés comme discriminant sur ce même seuil de temps lors de l'analyse supervisée des données de la cohorte clinique, dont la proline, la taurine ou encore l'acide isopropylmalique.

b) Analyse de l'effet centre

L'analyse PLS-DA sur les 301 variables annotées n'ayant pas généré de modèle, nous avons refait l'analyse sur l'ensemble des variables extraites (soit 4788 dans les conditions HILIC-ESI⁺). L'observation de la carte factorielle de PLS-DA construite permet de visualiser une structuration des échantillons selon le centre de prélèvement (Figure 82). Cette analyse nous a permis d'identifier 274 variables discriminantes avec une valeur de VIP > 1,5. La base de données interne a permis d'annoter 38 de ces variables, incluant principalement des aminoacides ou des lipides. Parmi les composés discriminant les

deux centres, nous retrouvons ainsi les acides méthylmalonique et isopropylmalique, ou encore la taurine et la N-acétyl-glutamine... (Annexe 19).

Les différences observées entre les deux centres peuvent être reliées à différents paramètres. En effet, les deux zones géographiques étudiées incluent un milieu urbain (Nancy) et un milieu plus rural (Poitiers), associées à des « modes de vie » différents. Ceux-ci sont différents tant au niveau des habitudes alimentaires qu'aux niveaux des expositions environnementales (expositions à des polluants atmosphériques ou des composés agricoles). De précédentes études réalisées sur la cohorte EDEN ont notamment étudié l'impact du lieu de vie sur l'apparition de maladies respiratoires [23]. Ces maladies ont notamment été reliées à une plus forte exposition à des particules fines et des composés de type NOx à Nancy, pouvant expliquer le plus grand nombre de cas d'allergies déclarées au sein de cette population (90 % des cas d'allergies de la cohorte).

Par ailleurs, en analysant les métabolites discriminants identifiés lors de l'analyse supervisée de l'effet « centre » et de l'effet « temps », nous retrouvons 24 métabolites communs (soit quasiment 50 % des variables discriminantes identifiées). Différentes voies d'acides aminés sont impliquées, décrites dans les deux cas. Parmi les composés communs, nous retrouvons la glutamine, l'acide glycérique ou la taurine.

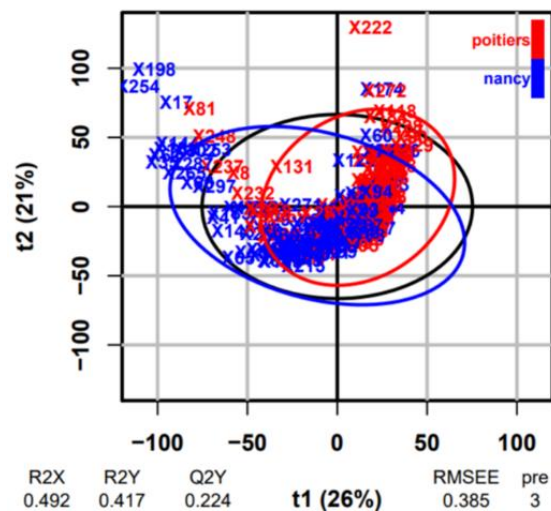


Figure 82 : Analyses discriminantes (PLS-DA) réalisées sur l'ensemble des échantillons de méconium de la cohorte initiale d'EDEN et l'ensemble des 4788 variables extraites dans les conditions HILIC-ESI.

Ce résultat suggère que le centre de prélèvement et le jour de collecte ont un effet significatif sur la composition métabolique globale du méconium, même si une partie des effets se recourent du fait d'un déséquilibre dans les temps de collecte au sein de ces deux centres. Ces facteurs seront donc à considérer et à modéliser pour pouvoir intégrer l'ensemble des échantillons et des données lors de l'analyse de l'impact des facteurs environnementaux/maternels sur la composition globale du

méconium. Ainsi que de l'association de cette composition avec le développement d'une allergie alimentaire dans l'enfance.

D. Analyses préliminaires de la relation entre la composition du méconium et le devenir allergique des enfants

1. Première analyse d'association sur les échantillons collectés à 24 heures (J1) au centre de Nancy

Afin d'éviter les facteurs confondants liés au jour et au centre de prélèvement décrits précédemment. Nous avons dans un premier temps évalué l'association entre la composition globale du méconium et la survenue d'allergie alimentaire dans l'enfance. Ceci en ne considérant que les échantillons collectés au cours des 24 premières heures post-partum, et provenant du centre de Nancy. En rajoutant, les cas supplémentaires d'allergies (< 24h, tous de Nancy), nous comptons au total 69 échantillons de méconium, dont 26 sont excrétés par des nouveau-nés qui déclareront une allergie alimentaire dans les 5 premières années de vie (soit 37 % des individus).

Nous présentons dans cette partie les analyses préliminaires réalisées sur l'ensemble des données extraites et filtrées, obtenues dans les conditions analytiques HILIC-ESI⁻ et C18-ESI⁺, ainsi que celles menées sur les variables annotées afin notamment d'en faciliter l'interprétation.

a) Traitement XCMS et extraction des variables dans les conditions HILIC-ESI⁻ et C18-ESI⁺

Pour réaliser cette étude, le prétraitement XCMS a été refait en ne considérant que les données obtenues sur les 69 échantillons sélectionnés dans les deux conditions analytiques HILIC-ESI⁻ et C18-ESI⁺. Cette étape a permis d'extraire respectivement 7294 et 15339 variables dans ces deux conditions. Nous observons par ailleurs, l'impact de la taille de la cohorte étudiée sur le nombre de variables extraites, comme rapporté précédemment (Figure 77). A la suite des étapes des filtres, nous avons retenu 4604 et 7369 variables respectivement dans les conditions HILIC-ESI⁻ et C18-ESI⁺.

b) Mise en relation de la composition du méconium avec le devenir allergique des enfants : Analyses sur l'ensemble des variables extraites et sur les variables annotées

Dans un premier temps, nous avons réalisé des analyses non supervisées de type ACP, permettant : i) de visualiser la répartition des échantillons sans *a priori*, ii) d'identifier certains échantillons ou variables extrêmes, iii) d'apprécier la structuration du jeu de données en considérant le facteur AA.

Lors d'une première analyse, nous avons réalisé les ACP sur 4604 variables (HILIC-ESI⁻), 7369 variables (C18-ESI⁺) et 301 variables annotées. Les graphes des individus obtenus à l'aide de ces trois jeux de données sont représentés en Figure 83 (panel de gauche). Ces représentations permettent d'expliquer 41 % (HILIC-ESI⁻), 33 % (C18-ESI⁺) et 41 % (variables annotées) de la variabilité totale à travers les deux premières composantes de l'ACP. Les troisièmes composantes décrivent respectivement 5, 5 et 3 % de la variabilité résiduelle du jeu de données dans ces trois conditions respectives. Nous observons les mêmes échantillons « extrêmes » dans les trois conditions, incluant X174, X126, X60, X169 ou encore X226. Ceci permet de rendre compte une fois de plus que le jeu de données réduit peut être très représentatif de l'ensemble des données, à la fois dans les conditions HILIC-ESI⁻ que C18-ESI⁺.

Les analyses en composantes principales colorées en fonction du devenir allergique des enfants sont également représentées en Figure 83 (panel de droite) pour les données extraites et filtrées obtenues dans les conditions analytiques HILIC-ESI⁻, C18-ESI⁺, et pour les variables annotées. Ces représentations ne montrent pas de structuration spécifique des jeux de données en fonction du devenir allergique des enfants.

Des analyses discriminantes de type PLS-DA réalisées avec le statut allergique en variable explicative, n'ont pas permis la construction de modèle, quel que soit le jeu de données considéré. En parallèle, des analyses univariées ont été menées (test paramétrique de Student, avec correction pour tests multiples), qui eux non plus, n'ont pas permis d'identifier de variables significativement différentes entre nos deux populations (cas allergiques et non allergiques).

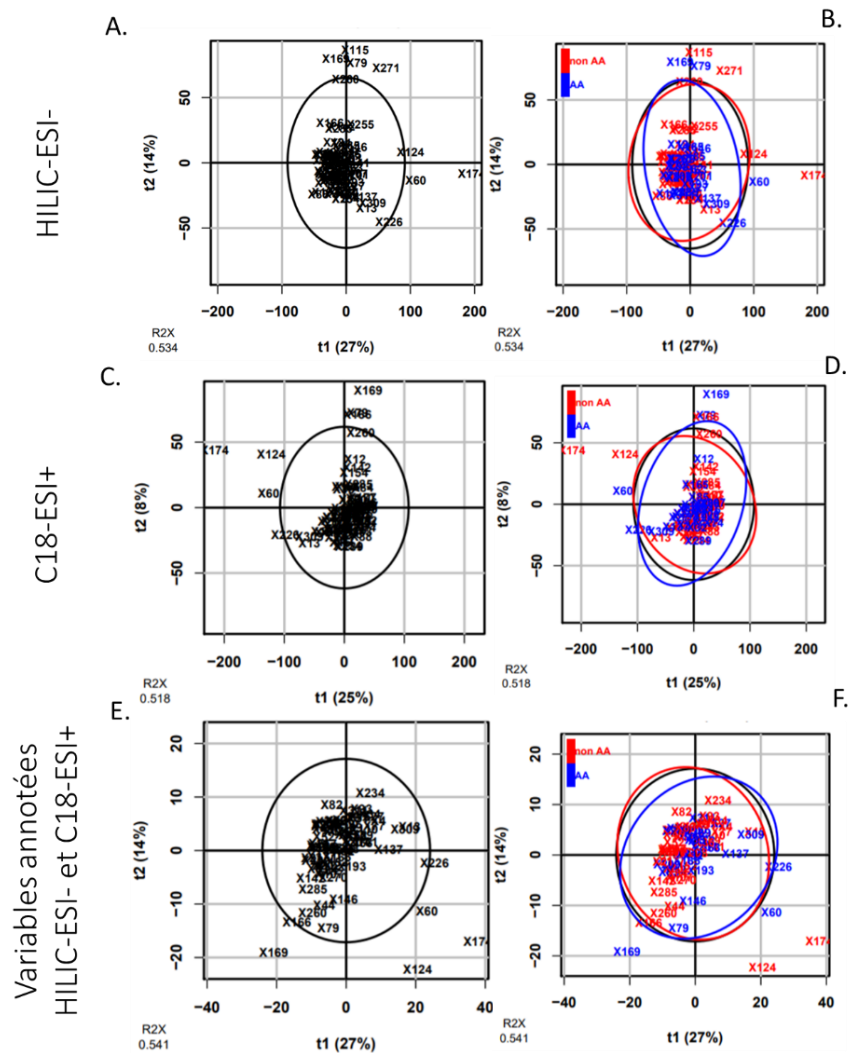


Figure 83 : Représentation des deux premières composantes de l'ACP réalisée sur les 4604, 7369 variables extraites et filtrées dans les conditions analytiques HILIC-ESI⁻ (A et B) et C18-ESI⁺ (C et D) et sur les 301 variables annotées dans ces deux conditions analytiques (E et F), obtenues sur les 69 échantillons sélectionnés. Les représentations en B, C et F respectives, sont colorées en fonction du facteur « AA » (cas allergiques en bleu et non allergiques en rouge).

2. Analyse d'association sur les échantillons collectés à 36 heures au centre de Nancy

Lors du stage de M2 de Vincent Marie, un « temps seuil » fixé à 36 heures a été considéré, temps qui semblait plus discriminant pour mettre en évidence une évolution métabolique dans les échantillons de la cohorte EDEN, et permettant d'inclure plus d'échantillons dans nos analyses d'associations.

Dans ces conditions 118 échantillons de méconiums issus du centre de Nancy sont considérés, dont 31 sont issus de nouveau-nés qui déclareront une allergie alimentaire (AA, soit 26 % des individus). Nous avons mené les mêmes analyses comme précédemment, sur l'ensemble des jeux de données (données extraites en conditions analytiques HILIC-ESI⁻ et C18-ESI⁺, et métabolites annotés).

Les analyses discriminantes (PLS-DA) n'ont pas permis de construire un modèle prédictif en fonction du devenir allergique des enfants. En revanche les analyses univariées des données obtenues dans les conditions analytiques C18-ESI⁺ ont permis d'identifier 42 variables dont les intensités sont significativement différentes entre les groupes « AA » et « non AA » (p-values < 0,05, correction pour test multiples de type fdr). La base de données interne n'a pas permis d'annoter ces variables. Cependant, une interrogation de la base de données HMDB [234], et en ne considérant que les ions monoisotopiques [M+H]⁺ et une erreur de la mesure de masse à 10 ppm, nous avons annoté putativement 6 de ces variables, décrites dans le Tableau 16.

Tableau 16 : Liste des composés annotés à travers la banque de données HMDB, ayant une p-value < 0,05 suite aux analyses univariées (test paramétrique ttest, avec correction de tests multiples) sur les échantillons prélevés à des temps < 36 heures dans le centre de Nancy.

Métabolites	p-values	Intensités plus élevé dans le groupe :
Madlongiside C	0,0086	Non AA
Amoxicillin	0,0277	Non AA
Garcimangosone C	0,0277	Non AA
β-Casomorphin	0,0277	Non AA
Niazicinin	0,0397	AA
Physalin D/Physalin E	0,0403	Non AA

Parmi les composés attribués, nous identifions putativement la Madlongiside C, l'amoxicilline, la Garcimangosone C, la β-casomorphine, la niazicinine, et la Physalin D/Physalin E. Tous ces composés hormis la niazicinine montrent des intensités plus élevées dans le groupe des « non AA ». Cinq exemples de boxplots sur les 42 variables identifiées sont représentés en Figure 84. Cette figure permet de rendre compte d'une forte hétérogénéité observée au sein des composés présentant des différences significatives, tel que représenté en Figure 84.A, Figure 84.D et Figure 84.E. La Figure 84.B représente le boxpot du seul composé annoté montrant une intensité plus forte dans le groupe des futurs allergiques (Niazicinin). Tous les autres composés montrent des intensités significativement plus élevés chez les futurs non-allergiques, pouvant traduire un effet protecteur de ces derniers.

L'identité de ces molécules devra être confirmée par des expériences MS/MS. Nous ne tirerons donc pas de conclusions trop précoces quant aux rôles et aux fonctions biologiques de ces composés. Notons cependant que la β-casomorphin, provient de la dégradation de la caséine-β du lait par les protéases digestives, pouvant traduire un allaitement maternel si elle cette caséine-β est d'origine maternelle. Ce résultat pourrait alors suggérer une effet protecteur de l'allaitement précoce [321]. La Garcimangosone C, pourrait être reliée à la consommation de fruits et légumes, peut être indicatif d'un régime diversifié

et riche en fibres de la mère, qui semblerait également protecteur. Les composés dérivés des plantes telles que la Madlongiside C ou les physalines, traduisent également ce régime riche en plantes, exerçant des activités anti-inflammatoires protectrices [322], [323].

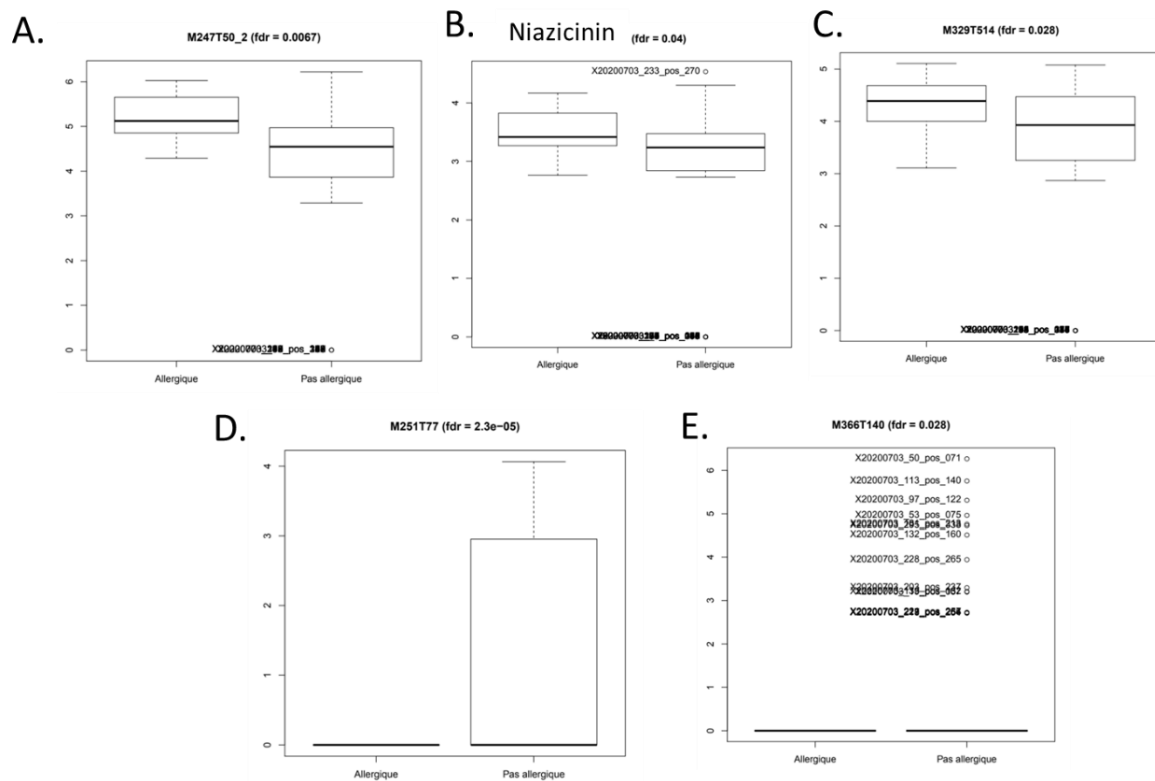


Figure 84 : Représentation sous forme de boxplots de 5/42 variables dont les intensités sont significativement différentes entre les cas allergiques et les cas non allergiques (analyse univariée de type ttest avec correction de tests multiples). Les échantillons analysés proviennent du centre de Nancy, et ont été collectés dans les 36h post-partum < 36h. Les noms au dessus des figures représentent le nom des variables qui n’ont pas été annotées, appelées (**Mm/zTRT**), ainsi que la valeur corrigée de la « p value ».

Lors de l’analyse du lait maternel de ces mêmes couples mères-enfants, Mikail Berdi a mis en évidence des différences significatives entre futurs allergiques *versus* non allergiques pour 36 composés (analyses univariées, test non paramétrique de Wilcoxon-Mann-Withney). Parmi ces composés, une majorité d’acides aminés sont identifiés, incluant la thréonine, la valine, la sérine, ou encore des carbohydrates tels que le rhamnose ou le fucose. Nombre de ces composés font notamment partie des métabolites annotés dans nos échantillons de méconiums. Cependant, les intensités de ces composés ne montrent pas de différences significatives entre futurs allergiques et non allergiques dans le méconium. Notons cependant que la majorité des laits maternels inclus dans les travaux de Mikail Berdi ont été collectés majoritairement 3 à 4 jours post-partum.

IV. Conclusions du chapitre III et perspectives

Le protocole d'extraction des métabolites précédemment développé a été appliqué aux 308 échantillons de méconium de la cohorte EDEN. L'analyse de ces métabolites a été réalisée en un seul lot au sein des deux conditions analytiques HILIC-ESI⁻ et C18-ESI⁺. Le suivi de la répétabilité de ces analyses au cours de la séquence a permis de rendre compte d'un système analytique stable, avec une légère dérive observée dans les conditions C18-ESI⁺. Celle-ci a pu être corrigée à l'aide de la fonction LOESS.

La comparaison du nombre de variables extraites au sein de la cohorte EDEN par rapport à la cohorte clinique a permis de rendre compte d'un éventuel impact du filtre « minfrac » réalisé lors du traitement par XCMS. Nous suggérons que ce paramètre soit optimisé dans le cas d'études de cohortes dédiées à l'étude des facteurs environnementaux sur la composition de la matrice analysée, où la variabilité entre échantillons peut être importante.

Les résultats d'annotation obtenus au sein de la cohorte EDEN ont permis de retrouver plus de 60 % de la signature métabolique décrite dans les chapitres précédents et de décrire plus de 50 nouveaux métabolites. Les analyses statistiques réalisées ont permis de suivre jour à jour l'évolution de la composition métabolique des méconiums et de définir plus de 46 métabolites discriminants entre les échantillons collectés avant ou après 24h. Les analyses réalisées sur un « facteur centre » a également montré un effet significatif de ce dernier, avec la définition de 38 métabolites discriminants entre Nancy et Poitiers par des analyses PLS-DA (VIP > 1,5). Cet effet semble confirmer notre hypothèse de départ quant à l'impact des facteurs environnementaux sur la composition métabolique du méconium. Elles seront approfondies une fois le traitement des données optimisé.

Les analyses statistiques supervisées réalisées selon la survenue ou non d'une allergie alimentaire dans l'enfance ont ensuite été menées. Dans un premier temps, nous avons considéré uniquement les échantillons prélevés avant 24h post-partum, et uniquement à Nancy. Aucune différence de la composition globale du méconium n'est mise en évidence entre nos deux groupes. En revanche la considération des prélèvements jusqu'à 36h post-partum, permettant d'inclure plus d'échantillons de méconiums, a permis d'identifier 40 variables significativement différentes entre les deux groupes par des analyses univariées (p-values < 0,05, correction pour les tests multiples). Ces variables n'ayant pas pu être annotées par notre base de données interne, nous avons interrogé la base de donnée HMDB, qui a permis d'annoter putativement 6 composés sur les 40. L'identité de ces composés devra être confirmée par des analyses MS/MS.

Les analyses métagénomiques ciblées et métaprotéomiques sont en cours de réalisation, au sein de l'institut Micalis de l'INRAE de Jouy en Josas et au LI2D du CEA de Marcoule respectivement pour les deux approches. L'intégration de l'ensemble de ces données permettrait de mieux appréhender la question posée sur le développement d'allergies alimentaires dans l'enfance.

CONCLUSIONS GENERALES ET PERSPECTIVES

La prévalence des allergies alimentaires est en constante augmentation depuis plus de 20 ans [1]. Ces allergies résultent d'une réponse immunitaire inappropriée contre des protéines alimentaires, touchant les enfants dès les premiers mois de vie. Cette observation suggère une maturation incorrecte du système immunitaire dès la vie *in utero*, résultant d'interactions néfastes du fœtus avec son environnement. Afin d'approcher ces interactions, nous avons proposé d'analyser le méconium. Celui-ci présente en effet le reflet d'une large fenêtre d'exposition, étant formé dès le deuxième mois de grossesse et s'accumulant dans l'intestin du fœtus jusqu'à la naissance [3].

Pour analyser la composition du méconium, nous avons choisi de réaliser une cartographie la plus globale possible en combinant différentes approches « omiques » : métabolomique, métagénomique et métaprotéomique. Les progrès récents des sciences technologiques « omiques » permettent de suivre qualitativement et quantitativement l'abondance de diverses molécules biologiques, et de surveiller leurs variations interindividuelles. L'exploration de cette matrice est complexe mais présente un fort potentiel dans l'étude des transmissions de la mère vers le fœtus, et de l'impact de cette fenêtre sur la santé de l'enfant.

Ce travail de thèse a été réalisé en trois grandes parties. Une première partie a porté sur la mise en place de la méthodologie d'analyse pour l'exploration du métabolome du méconium. Cette matrice n'avait jamais été étudiée au laboratoire et a nécessité de développer et d'optimiser les étapes de préparation de l'échantillon, afin i) d'extraire sans *a priori* le plus grand nombre de métabolites et, ii) d'assurer des analyses stables et répétables le long d'une grande série d'analyses. Le protocole optimisé sur deux échantillons de méconiums a permis de décrire une première cartographie, identifiant au plus haut niveau de confiance 137 métabolites suite à des analyses en conditions HILIC-ESI. Malgré la complexité de la matrice étudiée, nous avons pu extraire et identifier des composés aux propriétés physico-chimiques très variées. En effet, au-delà des acides aminés, des acides gras et des sucres observés, des familles telles que des acides organiques, aliphatiques ou aromatiques ont également été extraites, détectées et validées. Un test de répétabilité réalisé sur 120 injections d'échantillons de méconiums a permis de démontrer une bonne répétabilité et stabilité du système au cours de la séquence analytique. Nous avons réalisé des analyses supplémentaires à l'institut Parisien de Chimie Moléculaire (IPCM) sur un FT-ICR, concluant à un nombre de métabolites détectés moins important, notamment par l'effet de compétition des ions. Nous avons ainsi utilisé que la technique LC/HRMS dans la suite de l'étude.

L'ensemble de ce processus analytique a été appliqué à une cohorte de 35 échantillons de méconiums collectés longitudinalement auprès de 11 nouveau-nés, grâce à une collaboration que j'ai établie au

cours de ma thèse avec la clinique Sainte Thérèse (Paris 17^{ème}). L'analyse métabolomique de ces échantillons a permis d'apprécier la variabilité interindividuelle de la composition globale du méconium, ainsi que de décrire une évolution temporelle de cette composition. Elle a de plus permis de compléter la cartographie métabolique précédemment décrite par des analyses complémentaires dans les conditions analytiques C18-ESI⁺. Nous avons ainsi réussi à définir une cartographie globale de 230 métabolites identifiés au niveau 1 selon Sumner et al [210]. Celle-ci décrit à notre connaissance la cartographie métabolique la plus complète du méconium, connue à ce jour. Par des analyses statistiques uni et multivariées supervisées, nous avons identifié 67 métabolites significativement discriminants entre les prélèvements collectés avant ou après 24 heures (1440 minutes).

En parallèle, des analyses métagénomiques et métagénomiques ciblées ont été réalisées sur les échantillons de la cohorte clinique, en collaboration avec le laboratoire LI2D du CEA de Marcoule et l'Institut Micalis de l'INRAE de Jouy en Josas, experts dans ces domaines respectifs. Les résultats obtenus ont permis de définir une cartographie plus globale de ces méconiums, en identifiant 922 protéines humaines et 165 OTUs bactériens.

En regroupant l'ensemble des données omiques obtenues au sein des échantillons de la cohorte clinique, nous avons pu observer un lien direct entre les différentes compositions métaboliques, protéiques et bactériennes. Nous observons globalement, au cours du temps, une augmentation des intensités des métabolites, une baisse des abondances des protéines, et une augmentation de la diversité et des abondances de certains phyla bactériens. Ces observations peuvent être expliquées par l'augmentation des activités digestives et métaboliques du nouveau-né, la colonisation de son tube digestif par un microbiote de plus en plus abondant et diversifié, ainsi que la « contamination » du méconium par le lait maternel [2], [271]. Bittinger et al ont notamment expliqué la baisse des abondances protéiques par l'apparition de nouvelles espèces bactériennes digérant ces protéines en acides aminés [264]. Des résultats similaires ont été décrits par Righetti et al quant à l'évolution des quantités de métabolites [256], et par Del Chierrico et al quant à l'évolution observée en phyla bactériens [115]. L'évolution de ces compositions est mise en évidence essentiellement à partir de 24 heures (1440 minutes) suivant la naissance.

L'ensemble de ces résultats ont permis de mettre en évidence la richesse de la composition du méconium en métabolites, en protéines et en bactéries. Les résultats des analyses de la cohorte clinique, aboutiront à la réalisation de deux publications scientifiques. Une première publication, présentant l'analyse métabolomique globale, définissant la plus large cartographie du méconium (Annexe 1). Une deuxième publication, regroupant les résultats obtenus au sein des différentes approches métabolomique, métagénomique ciblée et métagénomique (en cours de rédaction). Enfin,

ces précieux échantillons ont également permis de démontrer le transfert materno-fœtal de nanoparticules de TiO₂ provenant des aliments, et leur accumulation dans le méconium dans le cadre d'une collaboration avec l'UMR Toxalime du département Alimentation Humaine de l'INRAE (Annexe 2).

Les méthodologies mises en place ont finalement été appliquées à l'analyse des échantillons de la cohorte française mère-enfant EDEN, concernant 308 échantillons de méconiums. Les analyses réalisées en un seul lot dans les conditions HILIC-ESI⁻ et C18-ESI⁺ ont démontré des résultats stables au cours de la séquence. Seule une faible dérive a été observée dans les conditions C18-ESI⁺, qui a pu être corrigée grâce à nos échantillons de QC régulièrement injectés tout au long de la séquence. La comparaison des résultats obtenus au sein de la cohorte EDEN et des résultats de la cohorte clinique, bien que concordant sur les métabolites annotés, a permis de rendre compte d'un éventuel impact des filtres réalisés lors du prétraitement des données, sur notre question scientifique. Des optimisations sont en cours de réflexion au laboratoire, permettant de développer de nouvelles stratégies de traitement des données pour permettre de conserver la variabilité interindividuelle dans le cadre d'étude des facteurs environnementaux sur des cohortes d'individus, quelle que soit la matrice étudiée.

Les résultats d'annotations obtenus au sein de la cohorte EDEN ont permis de retrouver plus de 60 % de la signature métabolique décrite dans les chapitres précédents et de décrire plus de 50 nouveaux métabolites. Les analyses statistiques réalisées ont permis d'observer des effets du temps de prélèvement et du centre de collecte. L'analyse de l'association entre la composition du méconium et le devenir allergique de l'enfant a ainsi été réalisée à l'aide des prélèvements collectés dans les 24h post-naissance et issus du centre de Nancy (90 % des cas d'allergies de la cohorte), incluant 69 échantillons, dont 26 cas d'AA. Les analyses uni et multivariées supervisées dans ces conditions n'ont pas mis en évidence de différences entre les populations (allergiques *versus* non-allergiques). En revanche, la considération des prélèvements collectés dans les 36h post-partum, permettant d'intégrer 118 échantillons, a permis d'identifier 42 variables significativement différentes entre les deux groupes par des analyses univariées (p-values < 0,05). Ces variables n'ayant pas pu être annotées à l'aide de notre base de données interne, nous avons interrogé la base de donnée HMDB, qui a permis d'annoter 6 composés sur les 42, dont l'identité devra être confirmée.

La mise en relation des informations collectées auprès des mères (lieu de vie, alimentation, expositions environnementales, données socio-économiques et de santé...) avec la composition globale des méconiums est en train d'être menée, permettant d'affirmer ou non, si l'analyse du méconium peut permettre d'approcher l'exposome prénatal. Les analyses métagénomiques ciblées et métagénomiques sont en cours de réalisation au sein de l'institut Micalis de l'INRAE de Jouy en Josas

et du LI2D du CEA de Marcoule, spécialistes dans ces deux domaines respectifs. La combinaison de l'ensemble de ces données permettrait de mieux appréhender la question posée quant à l'impact des expositions prénatales sur la composition du méconium, et son association avec le développement d'allergies alimentaires dans l'enfance.

REFERENCES

- [1] D. Jaffuel, P. Demoly, and J. Bousquet, 'Les allergies alimentaires', *Rev. Fr. Allergol. Immunol. Clin.*, vol. 41, no. 2, pp. 169–186, Mar. 2001, doi: 10.1016/S0335-7457(01)00010-7.
- [2] C. Milani *et al.*, 'The First Microbial Colonizers of the Human Gut: Composition, Activities, and Health Implications of the Infant Gut Microbiota', *Microbiol. Mol. Biol. Rev.*, vol. 81, no. 4, Dec. 2017, doi: 10.1128/MMBR.00036-17.
- [3] M. Meyer-Monath, 'Thèse Marie Meyer, 2014, Université de Rennes 1'.
- [4] P. A. Eigenmann, S. H. Sicherer, T. A. Borkowski, B. A. Cohen, and H. A. Sampson, 'Prevalence of IgE-Mediated Food Allergy Among Children With Atopic Dermatitis', *Pediatrics*, vol. 101, no. 3, pp. e8–e8, Mar. 1998, doi: 10.1542/peds.101.3.e8.
- [5] J. M. James, 'Respiratory Manifestations of Food Allergy', *Pediatrics*, vol. 111, no. Supplement 3, pp. 1625–1630, Jun. 2003.
- [6] J. Wang and A. H. Liu, 'Food allergies and asthma', *Curr. Opin. Allergy Clin. Immunol.*, vol. 11, no. 3, pp. 249–254, Jun. 2011, doi: 10.1097/ACI.0b013e3283464c8e.
- [7] Y. Katz *et al.*, 'Early exposure to cow's milk protein is protective against IgE-mediated cow's milk protein allergy', *J. Allergy Clin. Immunol.*, vol. 126, no. 1, pp. 77-82.e1, Jul. 2010, doi: 10.1016/j.jaci.2010.04.020.
- [8] G. Prioult and C. Nagler-Anderson, 'Mucosal immunity and allergic responses: lack of regulation and/or lack of microbial stimulation?', *Immunol. Rev.*, vol. 206, no. 1, pp. 204–218, 2005, doi: <https://doi.org/10.1111/j.0105-2896.2005.00277.x>.
- [9] D. A. Moneret-Vautrin and G. Kanny, 'Food-induced anaphylaxis. A new French multicenter study', *Bull. Académie Natl. Médecine*, vol. 179, no. 1, pp. 161–172, 178–184; discussion 173-177, 1995.
- [10] S. H. Sicherer and H. A. Sampson, '9. Food allergy', *J. Allergy Clin. Immunol.*, vol. 117, no. 2, pp. S470–S475, Feb. 2006, doi: 10.1016/j.jaci.2005.05.048.
- [11] S. H. Sicherer and H. A. Sampson, 'Food allergy', *J. Allergy Clin. Immunol.*, vol. 125, no. 2, pp. S116–S125, Feb. 2010, doi: 10.1016/j.jaci.2009.08.028.
- [12] T. Liu, S. Navarro, and A. L. Lopata, 'Current advances of murine models for food allergy', *Mol. Immunol.*, vol. 70, pp. 104–117, Feb. 2016, doi: 10.1016/j.molimm.2015.11.011.
- [13] S. A. Lyons *et al.*, 'Prevalence of Food Sensitization and Food Allergy in Children Across Europe', *J. Allergy Clin. Immunol. Pract.*, vol. 8, no. 8, pp. 2736-2746.e9, Sep. 2020, doi: 10.1016/j.jaip.2020.04.020.
- [14] S. H. Sicherer and H. A. Sampson, 'Food allergy: Epidemiology, pathogenesis, diagnosis, and treatment', *J. Allergy Clin. Immunol.*, vol. 133, no. 2, pp. 291-307.e5, Feb. 2014, doi: 10.1016/j.jaci.2013.11.020.
- [15] J. Bousquet *et al.*, 'Scientific criteria and the selection of allergenic foods for product labelling', *Allergy Eur. J. Allergy Clin. Immunol. Suppl.*, vol. 53, no. 47, pp. 3–21, 1998.
- [16] R. W. Weber, 'Patterns of pollen cross-allergenicity', *J. Allergy Clin. Immunol.*, vol. 112, no. 2, pp. 229–239, Aug. 2003, doi: 10.1067/mai.2003.1683.
- [17] J. O. 'B. Hourihane, S. A. Kilburn, J. A. Nordlee, S. L. Hefle, S. L. Taylor, and J. O. Warner, 'An evaluation of the sensitivity of subjects with peanut allergy to very low doses of peanut protein: A randomized, double-blind, placebo-controlled food challenge study', *J. Allergy Clin. Immunol.*, vol. 100, no. 5, pp. 596–600, 1997, doi: 10.1016/S0091-6749(97)70161-1.
- [18] C. B. Daul, J. E. Morgan, J. Hughes, and S. B. Lehrer, 'Provocation-challenge studies in shrimp-sensitive individuals', *J. Allergy Clin. Immunol.*, vol. 81, no. 6, pp. 1180–1186, 1988, doi: 10.1016/0091-6749(88)90888-3.
- [19] J. E. Gern, E. Yang, H. A. Sampson, and H. M. Evrard, 'Allergic reactions to milk-contaminated nondairy products', *N. Engl. J. Med.*, vol. 324, no. 14, pp. 976–979, 1991, doi: 10.1056/NEJM199104043241407.
- [20] S. Lau, M. Thiemeier, R. Urbanek, M. Kemeny, and U. Wahn, 'Immediate hypersensitivity to ovalbumin in children with hen's egg white allergy', *Eur. J. Pediatr.*, vol. 147, no. 6, pp. 606–608, 1988, doi: 10.1007/BF00442473.
- [21] S. A. Bock, 'Natural history of severe reactions to foods in young children', *J. Pediatr.*, vol. 107, no. 5, pp. 676–680, 1985, doi: 10.1016/S0022-3476(85)80391-7.

- [22] N. Baiz *et al.*, 'Maternal exposure to air pollution before and during pregnancy related to changes in newborn's cord blood lymphocyte subpopulations. The EDEN study cohort', *BMC Pregnancy Childbirth*, vol. 11, p. 87, Nov. 2011, doi: 10.1186/1471-2393-11-87.
- [23] The EDEN Mother-Child Cohort Study Group, C. Zhou, N. Baiz, T. Zhang, S. Banerjee, and I. Annesi-Maesano, 'Modifiable exposures to air pollutants related to asthma phenotypes in the first year of life in children of the EDEN mother-child cohort study', *BMC Public Health*, vol. 13, no. 1, Dec. 2013, doi: 10.1186/1471-2458-13-506.
- [24] S. M. Willers *et al.*, 'Maternal food consumption during pregnancy and asthma, respiratory and atopic symptoms in 5-year-old children', *Thorax*, vol. 62, no. 9, p. 773, Sep. 2007, doi: 10.1136/thx.2006.074187.
- [25] M. G. Dominguez-Bello *et al.*, 'Partial restoration of the microbiota of cesarean-born infants via vaginal microbial transfer', *Nat. Med.*, vol. 22, no. 3, pp. 250–253, Mar. 2016, doi: 10.1038/nm.4039.
- [26] A. G. Hirsch *et al.*, 'Early-life antibiotic use and subsequent diagnosis of food allergy and allergic diseases', *Clin. Exp. Allergy*, vol. 47, no. 2, pp. 236–244, 2017, doi: <https://doi.org/10.1111/cea.12807>.
- [27] M. Eggesbø, G. Botten, H. Stigum, P. Nafstad, and P. Magnus, 'Is delivery by cesarean section a risk factor for food allergy?', *J. Allergy Clin. Immunol.*, vol. 112, no. 2, pp. 420–426, Aug. 2003, doi: 10.1067/mai.2003.1610.
- [28] E. Papathoma, M. Triga, S. Fouzas, and G. Dimitriou, 'Cesarean section delivery and development of food allergy and atopic dermatitis in early childhood', *Pediatr. Allergy Immunol.*, vol. 27, no. 4, pp. 419–424, 2016, doi: <https://doi.org/10.1111/pai.12552>.
- [29] J. Stokholm *et al.*, 'Delivery mode and gut microbial changes correlate with an increased risk of childhood asthma', *Sci. Transl. Med.*, vol. 12, no. 569, Nov. 2020, doi: 10.1126/scitranslmed.aax9929.
- [30] M. Gdalevich, D. Mimouni, and M. Mimouni, 'Breast-feeding and the risk of bronchial asthma in childhood: A systematic review with meta-analysis of prospective studies', *J. Pediatr.*, vol. 139, no. 2, pp. 261–266, Aug. 2001, doi: 10.1067/mpd.2001.117006.
- [31] M. Gdalevich, D. Mimouni, M. David, and M. Mimouni, 'Breast-feeding and the onset of atopic dermatitis in childhood: A systematic review and meta-analysis of prospective studies', *J. Am. Acad. Dermatol.*, vol. 45, no. 4, pp. 520–527, Oct. 2001, doi: 10.1067/mjd.2001.114741.
- [32] J. van Odijk *et al.*, 'Breastfeeding and allergic disease: a multidisciplinary review of the literature (1966–2001) on the mode of early feeding in infancy and its impact on later atopic manifestations', *Allergy*, vol. 58, no. 9, pp. 833–843, 2003, doi: <https://doi.org/10.1034/j.1398-9995.2003.00264.x>.
- [33] P. P. Echarri *et al.*, 'Assessment of intestinal microbiota of full-term breast-fed infants from two different geographical locations', *Early Hum. Dev.*, vol. 87, no. 7, pp. 511–513, Jul. 2011, doi: 10.1016/j.earlhumdev.2011.03.013.
- [34] S. M. Parigi, M. Eldh, P. Larssen, S. Gabrielsson, and E. J. Villablanca, 'Breast Milk and Solid Food Shaping Intestinal Immunity', *Front. Immunol.*, vol. 6, Aug. 2015, doi: 10.3389/fimmu.2015.00415.
- [35] M. Berdi *et al.*, 'Immune components of early breastmilk: Association with maternal factors and with reported food allergy in childhood', *Pediatr. Allergy Immunol.*, vol. 30, no. 1, pp. 107–116, 2019, doi: 10.1111/pai.12998.
- [36] V. J. Flaherman, N. R. Narayan, D. Hartigan-O'Connor, M. D. Cabana, C. E. McCulloch, and I. M. Paul, 'The Effect of Early Limited Formula on Breastfeeding, Readmission, and Intestinal Microbiota: A Randomized Clinical Trial', *J. Pediatr.*, vol. 196, pp. 84-90.e1, May 2018, doi: 10.1016/j.jpeds.2017.12.073.
- [37] M. Nuriel-Ohayon, H. Neuman, and O. Koren, 'Microbial Changes during Pregnancy, Birth, and Infancy', *Microbiology*, p. 1031, 2016, doi: 10.3389/fmicb.2016.01031.
- [38] M. Wang *et al.*, 'Fecal microbiota composition of breast-fed infants is correlated with human milk oligosaccharides consumed', *J. Pediatr. Gastroenterol. Nutr.*, vol. 60, no. 6, pp. 825–833, Jun. 2015, doi: 10.1097/MPG.0000000000000752.

- [39] T. Jost, C. Lacroix, C. Braegger, and C. Chassard, 'Impact of human milk bacteria and oligosaccharides on neonatal gut microbiota establishment and gut health', *Nutr. Rev.*, vol. 73, no. 7, pp. 426–437, Jul. 2015, doi: 10.1093/nutrit/nuu016.
- [40] K. Le Doare, B. Holder, A. Bassett, and P. S. Pannaraj, 'Mother's Milk: A Purposeful Contribution to the Development of the Infant Microbiota and Immunity', *Front. Immunol.*, vol. 9, p. 361, 2018, doi: 10.3389/fimmu.2018.00361.
- [41] 'Thèse Mikail BERDI, 2017, Université Paris-Descartes'.
- [42] R. J. Boyle *et al.*, 'Hydrolysed formula and risk of allergic or autoimmune disease: systematic review and meta-analysis', *BMJ*, vol. 352, Mar. 2016, doi: 10.1136/bmj.i974.
- [43] C. Davaise-Paturet *et al.*, 'Use of partially hydrolysed formula in infancy and incidence of eczema, respiratory symptoms or food allergies in toddlers from the ELFE cohort', *Pediatr. Allergy Immunol. Off. Publ. Eur. Soc. Pediatr. Allergy Immunol.*, vol. 30, no. 6, pp. 614–623, 2019, doi: 10.1111/pai.13094.
- [44] J. M. Swanson, S. Entringer, C. Buss, and P. D. Wadhwa, 'Developmental Origins of Health and Disease: Environmental Exposures', *Semin. Reprod. Med.*, vol. 27, no. 5, pp. 391–402, Sep. 2009, doi: 10.1055/s-0029-1237427.
- [45] D. A. Chernikova *et al.*, 'Fetal exposures and perinatal influences on the stool microbiota of premature infants', *J. Matern. Fetal Neonatal Med.*, vol. 29, no. 1, pp. 99–105, Jan. 2016, doi: 10.3109/14767058.2014.987748.
- [46] M. Meyer-Monath, C. Chatellier, F. Rouget, I. Morel, and F. Lestremau, 'Development of a multi-residue method in a fetal matrix: analysis of meconium', *Anal. Bioanal. Chem.*, vol. 406, no. 30, pp. 7785–7797, Dec. 2014, doi: 10.1007/s00216-014-8243-4.
- [47] M. Meyer-Monath, J. Beaumont, I. Morel, F. Rouget, K. Tack, and F. Lestremau, 'Analysis of BTEX and chlorinated solvents in meconium by headspace-solid-phase microextraction gas chromatography coupled with mass spectrometry', *Anal. Bioanal. Chem.*, vol. 406, no. 18, pp. 4481–4490, Jul. 2014, doi: 10.1007/s00216-014-7836-2.
- [48] O. Em *et al.*, 'Combined analysis of prenatal (maternal hair and blood) and neonatal (infant hair, cord blood and meconium) matrices to detect fetal exposure to environmental pesticides', *Environmental research*, Jan. 2009. <https://pubmed.ncbi.nlm.nih.gov/19019354/> (accessed Oct. 05, 2020).
- [49] O. Em *et al.*, 'A comparison of infant hair, cord blood and meconium analysis to detect fetal exposure to environmental pesticides', *Environmental research*, Feb. 2008. <https://pubmed.ncbi.nlm.nih.gov/17949707/> (accessed Oct. 05, 2020).
- [50] R. Gb, C. Mc, P. O, O. E, and D. C, 'The Tagum study I: analysis and clinical correlates of mercury in maternal and cord blood, breast milk, meconium, and infants' hair', *Pediatrics*, Oct. 2000. <https://pubmed.ncbi.nlm.nih.gov/11015522/> (accessed Oct. 05, 2020).
- [51] T. Berton *et al.*, 'Development of an analytical strategy based on LC–MS/MS for the measurement of different classes of pesticides and their metabolites in meconium: Application and characterisation of foetal exposure in France', *Environ. Res.*, vol. 132, pp. 311–320, Jul. 2014, doi: 10.1016/j.envres.2014.03.034.
- [52] J. Lozano, O. García-Algar, O. Vall, R. de la Torre, G. Scaravelli, and S. Pichini, 'Biological Matrices for the Evaluation of In Utero Exposure to Drugs of Abuse', *Ther. Drug Monit.*, vol. 29, no. 6, pp. 711–734, Dec. 2007, doi: 10.1097/FTD.0b013e31815c14ce.
- [53] L. S *et al.*, 'Assessment of tobacco, alcohol and cannabinoid metabolites in 645 meconium samples of newborns compared to maternal self-reports', *Journal of psychiatric research*, Jul. 2017. <https://pubmed.ncbi.nlm.nih.gov/28237885/> (accessed Sep. 17, 2020).
- [54] C. F. Bearer, 'Meconium as a Biological Marker of Prenatal Exposure', *Ambul. Pediatr.*, p. 4, 2003.
- [55] A. Leinonen *et al.*, 'Liquid-phase microextraction for sample preparation in analysis of unconjugated anabolic steroids in urine', *Anal. Chim. Acta*, vol. 559, no. 2, pp. 166–172, Feb. 2006, doi: 10.1016/j.aca.2005.12.004.

- [56] I. Marchi, S. Rudaz, and J.-L. Veuthey, 'Sample preparation development and matrix effects evaluation for multianalyte determination in urine', *J. Pharm. Biomed. Anal.*, vol. 49, no. 2, pp. 459–467, Feb. 2009, doi: 10.1016/j.jpba.2008.11.040.
- [57] F. D. Eyler, M. Behnke, K. Wobie, C. W. Garvan, and I. Tebbett, 'Relative ability of biologic specimens and interviews to detect prenatal cocaine use', *Neurotoxicol. Teratol.*, vol. 27, no. 4, pp. 677–687, Jul. 2005, doi: 10.1016/j.ntt.2005.04.001.
- [58] M. Villarreal, 'Detection of illicit drugs in urine in the Division of Neonatology Hospital Molas in La Pampa', p. 5.
- [59] S. Pichini *et al.*, 'Assessment of chronic exposure to cigarette smoke and its change during pregnancy by segmental analysis of maternal hair nicotine', *J. Expo. Sci. Environ. Epidemiol.*, vol. 13, no. 2, Art. no. 2, Mar. 2003, doi: 10.1038/sj.jea.7500264.
- [60] S.-P. M *et al.*, 'Diazepam and its metabolites in the mothers' and newborns' hair as a biomarker of prenatal exposure', *Journal of physiology and pharmacology: an official journal of the Polish Physiological Society*, Aug. 2013. <https://pubmed.ncbi.nlm.nih.gov/24101397/> (accessed Oct. 05, 2020).
- [61] B.-O. B, K. J, K. T, and K. G, 'Comparison of meconium and neonatal hair analysis for detection of gestational exposure to drugs of abuse', *Archives of disease in childhood. Fetal and neonatal edition*, Mar. 2003. <https://pubmed.ncbi.nlm.nih.gov/12598495/> (accessed Oct. 05, 2020).
- [62] M. F, P. L, and B. E, 'Nails of newborns in monitoring drug exposure during pregnancy', *Forensic science international*, Jun. 08, 2008. <https://pubmed.ncbi.nlm.nih.gov/18620824/> (accessed Oct. 05, 2020).
- [63] S. G and P. L, 'A case report on drug screening of nail clippings to detect prenatal drug exposure', *Therapeutic drug monitoring*, Aug. 1997. <https://pubmed.ncbi.nlm.nih.gov/9263377/> (accessed Oct. 05, 2020).
- [64] G. L and M. S, 'Lactoperoxidase activity in human milk and in saliva of newborn infants', *Infection and immunity*, Jun. 1975. <https://pubmed.ncbi.nlm.nih.gov/1140845/> (accessed Oct. 05, 2020).
- [65] O. H *et al.*, 'Noninvasive surrogate markers for plasma cortisol in newborn infants: utility of urine and saliva samples and caution for venipuncture blood samples', *The Journal of clinical endocrinology and metabolism*, Oct. 2014. <https://pubmed.ncbi.nlm.nih.gov/25078034/> (accessed Oct. 05, 2020).
- [66] E. M. Ostrea, 'Testing for exposure to illicit drugs and other agents in the neonate: A review of laboratory methods and the role of meconium analysis', *Curr. Probl. Pediatr.*, vol. 29, no. 2, pp. 41–56, 1999, doi: 10.1016/S0045-9380(99)80046-2.
- [67] J. Unsworth and S. Vause, 'Meconium in labour', *Obstet. Gynaecol. Reprod. Med.*, vol. 20, no. 10, pp. 289–294, Oct. 2010, doi: 10.1016/j.ogrm.2010.06.005.
- [68] K. A. Madej, 'Analysis of meconium, nails and tears for determination of medicines and drugs of abuse', *TrAC Trends Anal. Chem.*, vol. 29, no. 3, pp. 246–259, Mar. 2010, doi: 10.1016/j.trac.2010.01.005.
- [69] F. Mayhoub *et al.*, 'Self-Reported Parental Exposure to Pesticide during Pregnancy and Birth Outcomes: The MecoExpo Cohort Study', *PLoS ONE*, vol. 9, no. 6, p. e99090, Jun. 2014, doi: 10.1371/journal.pone.0099090.
- [70] M. Meyer-Monath, C. Chatellier, D. Cabooter, F. Rouget, I. Morel, and F. Lestremau, 'Development of liquid chromatography methods coupled to mass spectrometry for the analysis of substances with a wide variety of polarity in meconium', *Talanta*, vol. 138, pp. 231–239, Jun. 2015, doi: 10.1016/j.talanta.2015.02.058.
- [71] H. Ma and C. Re, 'Drug abuse's smallest victims: in utero drug exposure', *Forensic science international*, Aug. 14, 2002. <https://pubmed.ncbi.nlm.nih.gov/12208017/> (accessed Oct. 05, 2020).
- [72] L. R, L. A, and K. G, 'Pharmacokinetic changes during pregnancy and their clinical relevance', *Clinical pharmacokinetics*, Nov. 1997. <https://pubmed.ncbi.nlm.nih.gov/9391746/> (accessed Oct. 05, 2020).

- [73] K. G. N. I, C. Ae, and L. C, 'Fetal alcohol spectrum disorder', *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne*, Nov. 25, 2003. <https://pubmed.ncbi.nlm.nih.gov/14638655/> (accessed Oct. 05, 2020).
- [74] C. Moore, A. Negrusz, and D. Lewis, 'Determination of drugs of abuse in meconium', *J. Chromatogr. B. Biomed. Sci. App.*, vol. 713, no. 1, pp. 137–146, Aug. 1998, doi: 10.1016/S0378-4347(97)00479-9.
- [75] S. N. Ahanya, J. Lakshmanan, B. L. G. Morgan, and M. G. Ross, 'Meconium Passage in Utero: Mechanisms, Consequences, and Management', *Obstet. Gynecol. Surv.*, vol. 60, no. 1, pp. 45–56, Jan. 2005, doi: 10.1097/01.ogx.0000149659.89530.c2.
- [76] S. McDermott, M. K. Hailer, and J. R. Lead, 'Meconium identifies high levels of metals in newborns from a mining community in the U.S.', *Sci. Total Environ.*, p. 135528, Nov. 2019, doi: 10.1016/j.scitotenv.2019.135528.
- [77] A. Guillard *et al.*, 'Basal Ti Level in the human placenta and meconium and evidence of a materno-foetal transfer of food-grade TiO₂ nanoparticles in an ex vivo placental perfusion model', *Part. Fibre Toxicol.*, 2020, doi: 10.1186/s12989-020-00381-z.
- [78] D. Bielawski, E. Ostrea, N. Posecion, M. Corrión, and J. Seagraves, 'Detection of Several Classes of Pesticides and Metabolites in Meconium by Gas Chromatography-Mass Spectrometry', *Chromatographia*, vol. 62, no. 11, pp. 623–629, Dec. 2005, doi: 10.1365/s10337-005-0668-7.
- [79] E. M. Ostrea, 'Meconium analysis to assess fetal exposure to nicotine by active and passive maternal smoking', *J. Pediatr.*, vol. 124, no. 3, pp. 471–476, Mar. 1994, doi: 10.1016/S0022-3476(94)70378-7.
- [80] J. Baranowski, G. Pochopień, and I. Baranowska, 'Determination of nicotine, cotinine and caffeine in meconium using high-performance liquid chromatography', *J. Chromatogr. B. Biomed. Sci. App.*, vol. 707, no. 1–2, pp. 317–321, Apr. 1998, doi: 10.1016/S0378-4347(97)00619-1.
- [81] L. N. Bakhireva *et al.*, 'Prenatal alcohol exposure prevalence as measured by direct ethanol metabolites in meconium in a Native American tribe of the southwest', *Birth Defects Res.*, Dec. 2018, doi: 10.1002/bdr2.1427.
- [82] M. E. Cristescu, 'The concept of genome after one century of usage', *Genome*, vol. 62, no. 10, pp. iii–v, Oct. 2019, doi: 10.1139/gen-2019-0129.
- [83] V. McKusick and F. Ruddle, 'Toward a complete map of the human genome. - Abstract - Europe PMC', 1987, Accessed: Oct. 09, 2020. [Online]. Available: <https://europepmc.org/article/med/3480265>.
- [84] N. Gehlenborg *et al.*, 'Visualization of omics data for systems biology', *Nat. Methods*, vol. 7, no. S3, pp. S56–S68, Mar. 2010, doi: 10.1038/nmeth.1436.
- [85] Y. Hasin, M. Seldin, and A. Lusic, 'Multi-omics approaches to disease', *Genome Biol.*, vol. 18, no. 1, p. 83, May 2017, doi: 10.1186/s13059-017-1215-1.
- [86] K. J. Karczewski and M. P. Snyder, 'Integrative omics for health and disease', *Nat. Rev. Genet.*, vol. 19, no. 5, pp. 299–310, May 2018, doi: 10.1038/nrg.2018.4.
- [87] S. C. Schuster, 'Next-generation sequencing transforms today's biology', *Nat. Methods*, vol. 5, no. 1, pp. 16–18, Jan. 2008, doi: 10.1038/nmeth1156.
- [88] T. Fuhrer and N. Zamboni, 'High-throughput discovery metabolomics', *Curr. Opin. Biotechnol.*, vol. 31, pp. 73–78, Feb. 2015, doi: 10.1016/j.copbio.2014.08.006.
- [89] D. J. Lockhart and E. A. Winzeler, 'Genomics, gene expression and DNA arrays', *Nature*, vol. 405, no. 6788, pp. 827–836, Jun. 2000, doi: 10.1038/35015701.
- [90] C. R. Woese and G. E. Fox, 'Phylogenetic structure of the prokaryotic domain: The primary kingdoms', *Proc. Natl. Acad. Sci.*, vol. 74, no. 11, pp. 5088–5090, Nov. 1977, doi: 10.1073/pnas.74.11.5088.
- [91] F. D. Ciccarelli, T. Doerks, C. von Mering, C. J. Creevey, B. Snel, and P. Bork, 'Toward Automatic Reconstruction of a Highly Resolved Tree of Life', *Science*, vol. 311, no. 5765, pp. 1283–1287, Mar. 2006, doi: 10.1126/science.1123061.

- [92] A. Spor, O. Koren, and R. Ley, 'Unravelling the effects of the environment and host genotype on the gut microbiome', *Nat. Rev. Microbiol.*, vol. 9, no. 4, pp. 279–290, Apr. 2011, doi: 10.1038/nrmicro2540.
- [93] P. Gérard, 'Le microbiote intestinal: composition et fonctions', *Phytothérapie*, vol. 9, no. 2, pp. 72–75, Apr. 2011, doi: 10.1007/s10298-011-0615-8.
- [94] A. S. Neish, 'Microbes in Gastrointestinal Health and Disease', *Gastroenterology*, vol. 136, no. 1, pp. 65–80, Jan. 2009, doi: 10.1053/j.gastro.2008.10.080.
- [95] M. Bazanella *et al.*, 'Randomized controlled trial on the impact of early-life intervention with bifidobacteria on the healthy infant fecal microbiota and metabolome', *Am. J. Clin. Nutr.*, p. ajcn157529, Sep. 2017, doi: 10.3945/ajcn.117.157529.
- [96] N. Larsen *et al.*, 'Gut Microbiota in Human Adults with Type 2 Diabetes Differs from Non-Diabetic Adults', *PLOS ONE*, vol. 5, no. 2, p. e9085, Feb. 2010, doi: 10.1371/journal.pone.0009085.
- [97] L. Moles *et al.*, 'Bacterial Diversity in Meconium of Preterm Neonates and Evolution of Their Fecal Microbiota during the First Month of Life', *PLOS ONE*, vol. 8, no. 6, p. e66986, Jun. 2013, doi: 10.1371/journal.pone.0066986.
- [98] P. J. Turnbaugh *et al.*, 'A core gut microbiome in obese and lean twins', *Nature*, vol. 457, no. 7228, pp. 480–484, Jan. 2009, doi: 10.1038/nature07540.
- [99] A. Mc *et al.*, 'Early infancy microbial and metabolic alterations affect risk of childhood asthma', *Science translational medicine*, Sep. 30, 2015. <https://pubmed.ncbi.nlm.nih.gov/26424567/> (accessed Sep. 18, 2020).
- [100] L.-S. Ka *et al.*, 'Integrative analysis of the intestinal metabolome of childhood asthma', *The Journal of allergy and clinical immunology*, Aug. 2019. <https://pubmed.ncbi.nlm.nih.gov/30914378/> (accessed Sep. 17, 2020).
- [101] P. B. Eckburg *et al.*, 'Diversity of the Human Intestinal Microbial Flora', *Science*, vol. 308, no. 5728, pp. 1635–1638, Jun. 2005, doi: 10.1126/science.1110591.
- [102] R. Burcelin, 'Microbiote intestinale et dialogue immunitaire au cours de la maladie métabolique', *Biol. Aujourd'hui*, vol. 211, no. 1, pp. 1–18, 2017, doi: 10.1051/jbio/2017008.
- [103] R. B. Sartor, 'Microbial Influences in Inflammatory Bowel Diseases', *Gastroenterology*, vol. 134, no. 2, pp. 577–594, Feb. 2008, doi: 10.1053/j.gastro.2007.11.059.
- [104] E. Jiménez *et al.*, 'Is meconium from healthy newborns actually sterile?', *Res. Microbiol.*, vol. 159, no. 3, pp. 187–193, Apr. 2008, doi: 10.1016/j.resmic.2007.12.007.
- [105] T. M. Wassenaar and P. Panigrahi, 'Is a foetus developing in a sterile environment?', *Lett. Appl. Microbiol.*, vol. 59, no. 6, pp. 572–579, Dec. 2014, doi: 10.1111/lam.12334.
- [106] R. W. Walker, J. C. Clemente, I. Peter, and R. J. F. Loos, 'The prenatal gut microbiome: are we colonized with bacteria *in utero*? Colonization of the gut microbiome *in utero*', *Pediatr. Obes.*, vol. 12, pp. 3–17, Aug. 2017, doi: 10.1111/ijpo.12217.
- [107] E. Jiménez *et al.*, 'Isolation of Commensal Bacteria from Umbilical Cord Blood of Healthy Neonates Born by Cesarean Section', *Curr. Microbiol.*, vol. 51, no. 4, pp. 270–274, Oct. 2005, doi: 10.1007/s00284-005-0020-3.
- [108] D. B. DiGiulio *et al.*, 'Microbial Prevalence, Diversity and Abundance in Amniotic Fluid During Preterm Labor: A Molecular and Culture-Based Investigation', *PLoS ONE*, vol. 3, no. 8, p. e3056, Aug. 2008, doi: 10.1371/journal.pone.0003056.
- [109] K. Aagaard, J. Ma, K. M. Antony, R. Ganu, J. Petrosino, and J. Versalovic, 'The Placenta Harbors a Unique Microbiome', *Sci. Transl. Med.*, vol. 6, no. 237, pp. 237ra65–237ra65, May 2014, doi: 10.1126/scitranslmed.3008599.
- [110] A. N. Ardisson *et al.*, 'Meconium Microbiome Analysis Identifies Bacteria Correlated with Premature Birth', *PLOS ONE*, vol. 9, no. 3, p. e90784, Mar. 2014, doi: 10.1371/journal.pone.0090784.
- [111] M. E. Perez-Muñoz, M.-C. Arrieta, A. E. Ramer-Tait, and J. Walter, 'A critical assessment of the "sterile womb" and "in utero colonization" hypotheses: implications for research on the pioneer infant microbiome', *Microbiome*, vol. 5, no. 1, Dec. 2017, doi: 10.1186/s40168-017-0268-4.

- [112] L. A. Parnell, C. M. Briggs, B. Cao, O. Delannoy-Bruno, A. E. Schrieffer, and I. U. Mysorekar, 'Microbial communities in placentas from term normal pregnancy exhibit spatially variable profiles', *Sci. Rep.*, vol. 7, no. 1, Dec. 2017, doi: 10.1038/s41598-017-11514-4.
- [113] R. Hansen *et al.*, 'First-Pass Meconium Samples from Healthy Term Vaginally-Delivered Neonates: An Analysis of the Microbiota', *PLOS ONE*, vol. 10, no. 7, p. e0133320, Jul. 2015, doi: 10.1371/journal.pone.0133320.
- [114] J. Hu *et al.*, 'Diversified Microbiota of Meconium Is Affected by Maternal Diabetes Status', *PLoS ONE*, vol. 8, no. 11, p. e78257, Nov. 2013, doi: 10.1371/journal.pone.0078257.
- [115] F. Del Chierico *et al.*, 'Phylogenetic and Metabolic Tracking of Gut Microbiota during Perinatal Development', *PLOS ONE*, vol. 10, no. 9, p. e0137347, Sep. 2015, doi: 10.1371/journal.pone.0137347.
- [116] M. J. Gosalbes, S. Llop, Y. Vallès, A. Moya, F. Ballester, and M. P. Francino, 'Meconium microbiota types dominated by lactic acid or enteric bacteria are differentially associated with maternal eczema and respiratory problems in infants', *Clin. Exp. Allergy*, vol. 43, no. 2, pp. 198–211, Feb. 2013, doi: 10.1111/cea.12063.
- [117] H. Makino *et al.*, 'Mother-to-Infant Transmission of Intestinal Bifidobacterial Strains Has an Impact on the Early Development of Vaginally Delivered Infant's Microbiota', *PLoS ONE*, vol. 8, no. 11, p. e78331, Nov. 2013, doi: 10.1371/journal.pone.0078331.
- [118] M. C. Collado, S. Rautava, J. Aakko, E. Isolauri, and S. Salminen, 'Human gut colonisation may be initiated in utero by distinct microbial communities in the placenta and amniotic fluid', *Sci. Rep.*, vol. 6, no. 1, Mar. 2016, doi: 10.1038/srep23129.
- [119] M. S. Lim and K. S. J. Elenitoba-Johnson, 'Proteomics in pathology research', *Lab. Invest.*, vol. 84, no. 10, pp. 1227–1244, Oct. 2004, doi: 10.1038/labinvest.3700167.
- [120] C.-K. J. Cho, S. J. Shan, E. J. Winsor, and E. P. Diamandis, 'Proteomics Analysis of Human Amniotic Fluid', *Mol. Cell. Proteomics*, vol. 6, no. 8, pp. 1406–1415, Aug. 2007, doi: 10.1074/mcp.M700090-MCP200.
- [121] C. A. Kolmeder *et al.*, 'Comparative Metaproteomics and Diversity Analysis of Human Intestinal Microbiota Testifies for Its Temporal Stability and Expression of Core Functions', *PLoS ONE*, vol. 7, no. 1, p. e29913, Jan. 2012, doi: 10.1371/journal.pone.0029913.
- [122] J. Armengaud, 'Microbiology and proteomics, getting the best of both worlds!', *Environ. Microbiol.*, vol. 15, no. 1, pp. 12–23, Jan. 2013, doi: 10.1111/j.1462-2920.2012.02811.x.
- [123] S. D. Patterson and R. H. Aebersold, 'Proteomics: the first decade and beyond', *Nat. Genet.*, vol. 33, no. S3, pp. 311–323, Mar. 2003, doi: 10.1038/ng1106.
- [124] J. Armengaud, E. Marie Hartmann, and C. Bland, 'Proteogenomics for environmental microbiology', *PROTEOMICS*, vol. 13, no. 18–19, pp. 2731–2742, Oct. 2013, doi: 10.1002/pmic.201200576.
- [125] X. Liu *et al.*, 'Protein Identification Using Top-Down Spectra', *Mol. Cell. Proteomics*, vol. 11, no. 6, Jun. 2012, doi: 10.1074/mcp.M111.008524.
- [126] P. A. Demirev, A. B. Feldman, P. Kowalski, and J. S. Lin, 'Top-Down Proteomics for Rapid Identification of Intact Microorganisms', *Anal. Chem.*, vol. 77, no. 22, pp. 7455–7461, Nov. 2005, doi: 10.1021/ac051419g.
- [127] C. Bland and J. Armengaud, 'Proteogenomics: a New Integrative Approach for a Better Description of Protein Diversity Found in Soil Microflora'. <https://www.caister.com/hsp/abstracts/omics/08.html> (accessed Oct. 09, 2020).
- [128] E. Lange, R. Tautenhahn, S. Neumann, and C. Gröpl, 'Critical assessment of alignment procedures for LC-MS proteomics and metabolomics measurements', *BMC Bioinformatics*, vol. 9, no. 1, p. 375, Sep. 2008, doi: 10.1186/1471-2105-9-375.
- [129] W.-J. Qian *et al.*, 'Probability-Based Evaluation of Peptide and Protein Identifications from Tandem Mass Spectrometry and SEQUEST Analysis: The Human Proteome', *J. Proteome Res.*, vol. 4, no. 1, pp. 53–62, Feb. 2005, doi: 10.1021/pr0498638.
- [130] S. Tyanova, T. Temu, and J. Cox, 'The MaxQuant computational platform for mass spectrometry-based shotgun proteomics', p. 19, 2016.

- [131] A. Bairoch, 'The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000', *Nucleic Acids Res.*, vol. 28, no. 1, pp. 45–48, Jan. 2000, doi: 10.1093/nar/28.1.45.
- [132] L. Martens *et al.*, 'PRIDE: The proteomics identifications database', *PROTEOMICS*, vol. 5, no. 13, pp. 3537–3545, Aug. 2005, doi: 10.1002/pmic.200401303.
- [133] J. Peng *et al.*, 'A proteomics approach to understanding protein ubiquitination', *Nat. Biotechnol.*, vol. 21, no. 8, pp. 921–926, Aug. 2003, doi: 10.1038/nbt849.
- [134] J. K. Eng, A. L. McCormack, and J. R. Yates, 'An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database', *J. Am. Soc. Mass Spectrom.*, vol. 5, no. 11, pp. 976–989, Nov. 1994, doi: 10.1016/1044-0305(94)80016-2.
- [135] J. S. Cottrell, 'Protein identification using MS/MS data', *J. Proteomics*, vol. 74, no. 10, pp. 1842–1851, Sep. 2011, doi: 10.1016/j.jprot.2011.05.014.
- [136] B. Lisowska-Myjak, E. Skarżyńska, K. Wojdan, and A. Nasierowska-Guttmejer, 'Protein and peptide profiles in neonatal meconium', *J. Obstet. Gynaecol. Res.*, vol. 45, no. 3, pp. 556–564, Mar. 2019, doi: 10.1111/jog.13888.
- [137] B. Lisowska-Myjak, E. Skarżyńska, and M. Bakun, 'Meconium proteins as a source of biomarkers for the assessment of the intrauterine environment of the fetus', *J. Dev. Orig. Health Dis.*, vol. 9, no. 3, pp. 329–337, Jun. 2018, doi: 10.1017/S2040174418000028.
- [138] B. Lisowska-Myjak and J. Pachecka, 'Alpha-1-Antitrypsin and IgA in Serial Meconium and Faeces of Healthy Breast-Fed Newborns', *Fetal Diagn. Ther.*, vol. 22, no. 2, pp. 116–120, 2007, doi: 10.1159/000097108.
- [139] J. Langhorst, S. Elsenbruch, J. Koelzer, A. Rueffer, A. Michalsen, and G. J. Dobos, 'Noninvasive Markers in the Assessment of Intestinal Inflammation in Inflammatory Bowel Diseases: Performance of Fecal Lactoferrin, Calprotectin, and PMN-Elastase, CRP, and Clinical Indices', *Off. J. Am. Coll. Gastroenterol. ACG*, vol. 103, no. 1, pp. 162–169, Jan. 2008.
- [140] B. Lisowska-Myjak, J. Żytyńska-Daniluk, and E. Skarżyńska, 'Concentrations of neutrophil-derived proteins in meconium and their correlations', *Biomark. Med.*, vol. 10, no. 8, pp. 819–829, Aug. 2016, doi: 10.2217/bmm-2016-0034.
- [141] N. Laforgia *et al.*, 'Calprotectin levels in meconium', *Acta Paediatr.*, vol. 92, no. 4, pp. 463–466, Jan. 2007, doi: 10.1111/j.1651-2227.2003.tb00579.x.
- [142] O. Fiehn, J. Kopka, P. Dörmann, T. Altmann, R. N. Trethewey, and L. Willmitzer, 'Metabolite profiling for plant functional genomics', *Nat. Biotechnol.*, vol. 18, no. 11, pp. 1157–1161, Nov. 2000, doi: 10.1038/81137.
- [143] J. K. Nicholson, J. C. Lindon, and E. Holmes, "'Metabonomics": understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data', *Xenobiotica*, vol. 29, no. 11, pp. 1181–1189, Jan. 1999, doi: 10.1080/004982599238047.
- [144] S. G. Oliver, M. K. Winson, D. B. Kell, and F. Baganz, 'Systematic functional analysis of the yeast genome', *Trends Biotechnol.*, vol. 16, no. 9, pp. 373–378, Sep. 1998, doi: 10.1016/S0167-7799(98)01214-1.
- [145] S. Li *et al.*, 'Predicting Network Activity from High Throughput Metabolomics', *PLOS Comput. Biol.*, vol. 9, no. 7, p. e1003123, Jul. 2013, doi: 10.1371/journal.pcbi.1003123.
- [146] N. G. Mahieu, J. L. Genenbacher, and G. J. Patti, 'A roadmap for the XCMS family of software solutions in metabolomics', *Curr. Opin. Chem. Biol.*, vol. 30, pp. 87–93, Feb. 2016, doi: 10.1016/j.cbpa.2015.11.009.
- [147] G. J. Patti, O. Yanes, and G. Siuzdak, 'Metabolomics: the apogee of the omics trilogy', *Nat. Rev. Mol. Cell Biol.*, vol. 13, no. 4, pp. 263–269, Apr. 2012, doi: 10.1038/nrm3314.
- [148] O. Fiehn, 'Metabolomics — the link between genotypes and phenotypes', in *Functional Genomics*, C. Town, Ed. Dordrecht: Springer Netherlands, 2002, pp. 155–171.
- [149] L. D. Roberts, A. L. Souza, R. E. Gerszten, and C. B. Clish, 'Targeted Metabolomics', *Curr. Protoc. Mol. Biol.*, vol. 98, no. 1, p. 30.2.1–30.2.24, Apr. 2012, doi: 10.1002/0471142727.mb3002s98.
- [150] A. Alonso, S. Marsal, and A. Julià, 'Analytical methods in untargeted metabolomics: state of the art in 2015', *Bioeng. Biotechnol.*, vol. 3, p. 23, 2015, doi: 10.3389/fbioe.2015.00023.

- [151] S. Moco, J. Vervoort, S. Moco, R. J. Bino, R. C. H. De Vos, and R. Bino, 'Metabolomics technologies and metabolite identification', *TrAC Trends Anal. Chem.*, vol. 26, no. 9, pp. 855–866, Oct. 2007, doi: 10.1016/j.trac.2007.08.003.
- [152] E. J. Saude *et al.*, 'Metabolomic profiling of asthma: Diagnostic utility of urine nuclear magnetic resonance spectroscopy', *J. Allergy Clin. Immunol.*, vol. 127, no. 3, pp. 757–764.e6, Mar. 2011, doi: 10.1016/j.jaci.2010.12.1077.
- [153] E. A. Mathé *et al.*, 'Noninvasive Urinary Metabolomic Profiling Identifies Diagnostic and Prognostic Markers in Lung Cancer', *Cancer Res.*, vol. 74, no. 12, pp. 3259–3270, Jun. 2014, doi: 10.1158/0008-5472.CAN-14-0109.
- [154] W. Xu *et al.*, 'Development of High-Performance Chemical Isotope Labeling LC–MS for Profiling the Human Fecal Metabolome', *Anal. Chem.*, vol. 87, no. 2, pp. 829–836, Jan. 2015, doi: 10.1021/ac503619q.
- [155] S. Castillo, P. Gopalacharyulu, L. Yetukuri, and M. Orešič, 'Algorithms and tools for the preprocessing of LC–MS metabolomics data', *Chemom. Intell. Lab. Syst.*, vol. 108, no. 1, pp. 23–32, Aug. 2011, doi: 10.1016/j.chemolab.2011.03.010.
- [156] A. Luedemann, L. von Malotky, A. Erban, and J. Kopka, 'TagFinder: Preprocessing Software for the Fingerprinting and the Profiling of Gas Chromatography–Mass Spectrometry Based Metabolome Analyses', *Plant Metabolomics*, pp. 255–286, 2011, doi: 10.1007/978-1-61779-594-7_16.
- [157] L. R. Euceda, G. F. Giskeødegård, and T. F. Bathen, 'Preprocessing of NMR metabolomics data', *Scand. J. Clin. Lab. Invest.*, vol. 75, no. 3, pp. 193–203, Apr. 2015, doi: 10.3109/00365513.2014.1003593.
- [158] A. Koulman, G. Woffendin, V. K. Narayana, H. Welchman, C. Crone, and D. A. Volmer, 'High-resolution extracted ion chromatography, a new tool for metabolomics and lipidomics using a second-generation orbitrap mass spectrometer', *Rapid Commun. Mass Spectrom.*, vol. 23, no. 10, pp. 1411–1418, May 2009, doi: 10.1002/rcm.4015.
- [159] T. Soga, Y. Ohashi, Y. Ueno, H. Naraoka, M. Tomita, and T. Nishioka, 'Quantitative Metabolome Analysis Using Capillary Electrophoresis Mass Spectrometry', *J. Proteome Res.*, vol. 2, no. 5, pp. 488–494, Oct. 2003, doi: 10.1021/pr034020m.
- [160] P. Kiefer, J.-C. Portais, and J. A. Vorholt, 'Quantitative metabolome analysis using liquid chromatography–high-resolution mass spectrometry', *Anal. Biochem.*, vol. 382, no. 2, pp. 94–100, Nov. 2008, doi: 10.1016/j.ab.2008.07.010.
- [161] L. Wu *et al.*, 'Quantitative analysis of the microbial metabolome by isotope dilution mass spectrometry using uniformly ¹³C-labeled cell extracts as internal standards', *Anal. Biochem.*, vol. 336, no. 2, pp. 164–171, Jan. 2005, doi: 10.1016/j.ab.2004.09.001.
- [162] C. Junot, F. Fenaille, B. Colsch, and F. Bécher, 'High resolution mass spectrometry based techniques at the crossroads of metabolic pathways', *Mass Spectrom. Rev.*, vol. 33, no. 6, pp. 471–500, Nov. 2014, doi: 10.1002/mas.21401.
- [163] D. Vuckovic, 'Current trends and challenges in sample preparation for global metabolomics using liquid chromatography–mass spectrometry', *Anal. Bioanal. Chem.*, vol. 403, no. 6, pp. 1523–1548, Jun. 2012, doi: 10.1007/s00216-012-6039-y.
- [164] W. B. Dunn *et al.*, 'Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry', *Nat. Protoc.*, vol. 6, no. 7, pp. 1060–1083, Jul. 2011, doi: 10.1038/nprot.2011.335.
- [165] S. J. Bruce, I. Tavazzi, V. Parisod, S. Rezzi, S. Kochhar, and P. A. Guy, 'Investigation of Human Blood Plasma Sample Preparation for Performing Metabolomics Using Ultrahigh Performance Liquid Chromatography/Mass Spectrometry', *Anal. Chem.*, vol. 81, no. 9, pp. 3285–3296, May 2009, doi: 10.1021/ac8024569.
- [166] E. Zelena *et al.*, 'Development of a Robust and Repeatable UPLC–MS Method for the Long-Term Metabolomic Study of Human Serum', *Anal. Chem.*, vol. 81, no. 4, pp. 1357–1364, Feb. 2009, doi: 10.1021/ac8019366.

- [167] H. Tweeddale, L. Notley-McRobb, and T. Ferenci, 'Effect of Slow Growth on Metabolism of *Escherichia coli*, as Revealed by Global Metabolite Pool ("Metabolome") Analysis', *J. Bacteriol.*, vol. 180, no. 19, pp. 5109–5116, Oct. 1998, doi: 10.1128/JB.180.19.5109-5116.1998.
- [168] J. K. Nicholson and I. D. Wilson, 'High resolution proton magnetic resonance spectroscopy of biological fluids', *Prog. Nucl. Magn. Reson. Spectrosc.*, vol. 21, no. 4–5, pp. 449–501, Jan. 1989, doi: 10.1016/0079-6565(89)80008-1.
- [169] 'Thèse Kathleen Rousseau, 2020, Université Paris-Saclay'.
- [170] Q. N. Van *et al.*, 'Comparison of 1D and 2D NMR Spectroscopy for Metabolic Profiling', *J. Proteome Res.*, vol. 7, no. 2, pp. 630–639, Feb. 2008, doi: 10.1021/pr700594s.
- [171] B. Habchi *et al.*, 'Direct introduction Fourier transform mass spectrometry (FTMS) for metabolomic fingerprint', in *XX. Rencontres du Club Jeune de la Société Française de Spectrométrie de Masse*, La Bégude de Mazenc, France, May 2015, p. np, Accessed: Nov. 28, 2018. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-01435379>.
- [172] B. Habchi, S. Alves, A. Paris, D. N. Rutledge, and E. Rathahao-Paris, 'How to really perform high throughput metabolomic analyses efficiently?', *TrAC Trends Anal. Chem.*, vol. 85, pp. 128–139, Dec. 2016, doi: 10.1016/j.trac.2016.09.005.
- [173] R. Ramautar, A. Demirci, and G. J. de Jong, 'Capillary electrophoresis in metabolomics', *TrAC Trends Anal. Chem.*, vol. 25, no. 5, pp. 455–466, May 2006, doi: 10.1016/j.trac.2006.02.004.
- [174] B. Diémé *et al.*, 'Workflow methodology for rat brain metabolome exploration using NMR, LC–MS and GC–MS analytical platforms', *J. Pharm. Biomed. Anal.*, vol. 142, pp. 270–278, Aug. 2017, doi: 10.1016/j.jpba.2017.03.068.
- [175] D. J. Beale *et al.*, 'Review of recent developments in GC–MS approaches to metabolomics-based research', *Metabolomics*, vol. 14, no. 11, pp. 1–31, Nov. 2018, doi: 10.1007/s11306-018-1449-2.
- [176] G. A. Theodoridis, H. G. Gika, E. J. Want, and I. D. Wilson, 'Liquid chromatography–mass spectrometry based global metabolite profiling: A review', *Anal. Chim. Acta*, vol. 711, pp. 7–16, Jan. 2012, doi: 10.1016/j.aca.2011.09.042.
- [177] M. Moini and H. Huang, 'Application of capillary electrophoresis/ electrospray ionization-mass spectrometry to subcellular proteomics of *Escherichia coli* ribosomal proteins', *ELECTROPHORESIS*, vol. 25, no. 13, pp. 1981–1987, Jul. 2004, doi: 10.1002/elps.200305906.
- [178] J. W. Kim, J. H. Park, J. W. Park, H. J. Doh, G. S. Heo, and K. J. Lee, 'Quantitative analysis of serum proteins separated by capillary electrophoresis', *Clin. Chem.*, vol. 39, no. 4, pp. 689–692, Apr. 1993, doi: 10.1093/clinchem/39.4.689.
- [179] I. Garcia-Pérez, M. Vallejo, C. Legido-Quigley, and C. Barbas, 'Metabolic fingerprinting with capillary electrophoresis', *J. Chromatogr. A*, vol. 1204, no. 2, pp. 130–139, Sep. 2008, doi: 10.1016/j.chroma.2008.07.025.
- [180] E. C. Horning and M. G. Horning, 'Metabolic Profiles: Gas-Phase Methods for Analysis of Metabolites', *Clin. Chem.*, vol. 17, no. 8, pp. 802–809, Aug. 1971, doi: 10.1093/clinchem/17.8.802.
- [181] T. Shepherd *et al.*, 'Analysis of metabolites from *Solanum* species using gas chromatography–(Time-of-Flight) mass spectrometry and automated data analysis', *undefined*, 2003. <https://www.semanticscholar.org/paper/Analysis-of-metabolites-from-Solanum-species-using-Shepherd-Dobson/670465a9a8b29c73e5cbf6df6375ba01b6e69b64> (accessed Sep. 30, 2020).
- [182] M. Monteiro *et al.*, 'GC-MS metabolomics-based approach for the identification of a potential VOC-biomarker panel in the urine of renal cell carcinoma patients', *J. Cell. Mol. Med.*, vol. 21, no. 9, pp. 2092–2105, 2017, doi: <https://doi.org/10.1111/jcmm.13132>.
- [183] M. Vinaixa, E. L. Schymanski, S. Neumann, M. Navarro, R. M. Salek, and O. Yanes, 'Mass spectral databases for LC/MS- and GC/MS-based metabolomics: State of the field and future prospects', *TrAC Trends Anal. Chem.*, vol. 78, pp. 23–35, Apr. 2016, doi: 10.1016/j.trac.2015.09.005.
- [184] J. Hummel, J. Selbig, D. Walther, and J. Kopka, 'The Golm Metabolome Database: a database for GC-MS based metabolite profiling', *Metabolomics*, pp. 75–95, 2007, doi: 10.1007/4735_2007_0229.

- [185] 'Recent advances in liquid and gas chromatography methodology for extending coverage of the metabolome', *Curr. Opin. Biotechnol.*, vol. 43, pp. 77–85, Feb. 2017, doi: 10.1016/j.copbio.2016.09.006.
- [186] M. M. Koek, F. M. van der Kloet, R. Kleemann, T. Kooistra, E. R. Verheij, and T. Hankemeier, 'Semi-automated non-target processing in GC × GC–MS metabolomics analysis: applicability for biomedical studies', *Metabolomics*, vol. 7, no. 1, pp. 1–14, Mar. 2011, doi: 10.1007/s11306-010-0219-6.
- [187] X. Gao *et al.*, 'Metabolite analysis of human fecal water by gas chromatography/mass spectrometry with ethyl chloroformate derivatization', *Anal. Biochem.*, vol. 393, no. 2, pp. 163–175, Oct. 2009, doi: 10.1016/j.ab.2009.06.036.
- [188] S. Boudah *et al.*, 'Annotation of the human serum metabolome by coupling three liquid chromatography methods to high-resolution mass spectrometry', *J. Chromatogr. B*, vol. 966, pp. 34–47, Sep. 2014, doi: 10.1016/j.jchromb.2014.04.025.
- [189] S. Aros-Calt *et al.*, 'Annotation of the *Staphylococcus aureus* Metabolome Using Liquid Chromatography Coupled to High-Resolution Mass Spectrometry and Application to the Study of Methicillin Resistance', *J. Proteome Res.*, vol. 14, no. 11, pp. 4863–4875, Nov. 2015, doi: 10.1021/acs.jproteome.5b00697.
- [190] T. Cajka and O. Fiehn, 'Toward Merging Untargeted and Targeted Methods in Mass Spectrometry-Based Metabolomics and Lipidomics', *Anal. Chem.*, vol. 88, no. 1, pp. 524–545, Jan. 2016, doi: 10.1021/acs.analchem.5b04491.
- [191] K. Spagou, H. Tsoukali, N. Raikos, H. Gika, I. D. Wilson, and G. Theodoridis, 'Hydrophilic interaction chromatography coupled to MS for metabolomic/metabolomic studies', *J. Sep. Sci.*, vol. 33, no. 6–7, pp. 716–727, Mar. 2010, doi: 10.1002/jssc.200900803.
- [192] D.-Q. Tang, L. Zou, X.-X. Yin, and C. N. Ong, 'HILIC-MS for metabolomics: An attractive and complementary approach to RPLC-MS', *Mass Spectrom. Rev.*, vol. 35, no. 5, pp. 574–600, Sep. 2016, doi: 10.1002/mas.21445.
- [193] P. D. Rainville, G. Theodoridis, R. S. Plumb, and I. D. Wilson, 'Advances in liquid chromatography coupled to mass spectrometry for metabolic phenotyping', *TrAC Trends Anal. Chem.*, vol. 61, pp. 181–191, Oct. 2014, doi: 10.1016/j.trac.2014.06.005.
- [194] H. G. Gika, G. A. Theodoridis, and I. D. Wilson, 'Hydrophilic interaction and reversed-phase ultra-performance liquid chromatography TOF-MS for metabolomic analysis of Zucker rat urine', *J. Sep. Sci.*, vol. 31, no. 9, pp. 1598–1608, May 2008, doi: 10.1002/jssc.200700644.
- [195] T. Zhang, D. J. Creek, M. P. Barrett, G. Blackburn, and D. G. Watson, '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.*, vol. 84, no. 4, pp. 1994–2001, Feb. 2012, doi: 10.1021/ac2030738.
- [196] R. Moreau *et al.*, 'Blood metabolomics uncovers inflammation-associated mitochondrial dysfunction as a potential mechanism underlying ACLF☆', *J. Hepatol.*, vol. 72, no. 4, pp. 688–701, Apr. 2020, doi: 10.1016/j.jhep.2019.11.009.
- [197] S. Boudah *et al.*, 'Annotation of the human serum metabolome by coupling three liquid chromatography methods to high-resolution mass spectrometry', *J. Chromatogr. B*, vol. 966, pp. 34–47, Sep. 2014, doi: 10.1016/j.jchromb.2014.04.025.
- [198] 'Thèse Samia Boudah 2014, Université Paris-Saclay'.
- [199] R. A. Scheltema *et al.*, 'The Q Exactive HF, a Benchtop Mass Spectrometer with a Pre-filter, High-performance Quadrupole and an Ultra-high-field Orbitrap Analyzer', *Mol. Cell. Proteomics*, vol. 13, no. 12, pp. 3698–3708, Dec. 2014, doi: 10.1074/mcp.M114.043489.
- [200] A. Michalski *et al.*, 'Mass Spectrometry-based Proteomics Using Q Exactive, a High-performance Benchtop Quadrupole Orbitrap Mass Spectrometer', *Mol. Cell. Proteomics*, vol. 10, no. 9, Sep. 2011, doi: 10.1074/mcp.M111.011015.
- [201] X. Chen *et al.*, 'Metabolic alterations associated with polycystic ovary syndrome: A UPLC Q-Exactive based metabolomic study', *Clin. Chim. Acta*, vol. 502, pp. 280–286, Mar. 2020, doi: 10.1016/j.cca.2019.11.016.

- [202] M. Himmelsbach, '10years of MS instrumental developments – Impact on LC–MS/MS in clinical chemistry', *J. Chromatogr. B*, vol. 883–884, pp. 3–17, Feb. 2012, doi: 10.1016/j.jchromb.2011.11.038.
- [203] S. Beck *et al.*, 'The Impact II, a Very High-Resolution Quadrupole Time-of-Flight Instrument (QTOF) for Deep Shotgun Proteomics', *Mol. Cell. Proteomics*, vol. 14, no. 7, pp. 2014–2029, Jul. 2015, doi: 10.1074/mcp.M114.047407.
- [204] B. Habchi *et al.*, 'Potential of dynamically harmonized Fourier transform ion cyclotron resonance cell for high-throughput metabolomics fingerprinting: control of data quality', *Anal. Bioanal. Chem.*, vol. 410, no. 2, pp. 483–490, Jan. 2018, doi: 10.1007/s00216-017-0738-3.
- [205] J. M. Hagel and P. J. Facchini, 'Plant metabolomics: analytical platforms and integration with functional genomics', *Phytochem. Rev.*, vol. 7, no. 3, pp. 479–497, Oct. 2008, doi: 10.1007/s11101-007-9086-9.
- [206] M. B. Comisarow, V. Grassi, and G. Parisod, 'Fourier transform ion cyclotron double resonance', *Chem. Phys. Lett.*, vol. 57, no. 3, pp. 413–416, Aug. 1978, doi: 10.1016/0009-2614(78)85537-7.
- [207] A. M. Brunner *et al.*, 'Benchmarking Multiple Fragmentation Methods on an Orbitrap Fusion for Top-down Phospho-Proteome Characterization', *Anal. Chem.*, vol. 87, no. 8, pp. 4152–4158, Apr. 2015, doi: 10.1021/acs.analchem.5b00162.
- [208] B. L. Williamson, J. Marchese, and N. A. Morrice, 'Automated Identification and Quantification of Protein Phosphorylation Sites by LC/MS on a Hybrid Triple Quadrupole Linear Ion Trap Mass Spectrometer', *Mol. Cell. Proteomics*, vol. 5, no. 2, pp. 337–346, Feb. 2006, doi: 10.1074/mcp.M500210-MCP200.
- [209] F. Fenaille, P. Barbier Saint-Hilaire, K. Rousseau, and C. Junot, 'Data acquisition workflows in liquid chromatography coupled to high resolution mass spectrometry-based metabolomics: Where do we stand?', *J. Chromatogr. A*, vol. 1526, pp. 1–12, Dec. 2017, doi: 10.1016/j.chroma.2017.10.043.
- [210] L. W. Sumner *et al.*, 'Proposed minimum reporting standards for chemical analysis', *Metabolomics*, vol. 3, no. 3, pp. 211–221, Sep. 2007, doi: 10.1007/s11306-007-0082-2.
- [211] Y. Guitton *et al.*, 'Create, run, share, publish, and reference your LC–MS, FIA–MS, GC–MS, and NMR data analysis workflows with the Workflow4Metabolomics 3.0 Galaxy online infrastructure for metabolomics', *Int. J. Biochem. Cell Biol.*, vol. 93, pp. 89–101, Dec. 2017, doi: 10.1016/j.biocel.2017.07.002.
- [212] N. G. Mahieu, J. L. Genenbacher, and G. J. Patti, 'A roadmap for the XCMS family of software solutions in metabolomics', *Curr. Opin. Chem. Biol.*, vol. 30, pp. 87–93, Feb. 2016, doi: 10.1016/j.cbpa.2015.11.009.
- [213] T. Pluskal, S. Castillo, A. Villar-Briones, and M. Orešič, 'MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data', *BMC Bioinformatics*, vol. 11, no. 1, pp. 1–11, Dec. 2010, doi: 10.1186/1471-2105-11-395.
- [214] A. Lommen, 'MetAlign: Interface-Driven, Versatile Metabolomics Tool for Hyphenated Full-Scan Mass Spectrometry Data Preprocessing', *Anal. Chem.*, vol. 81, no. 8, pp. 3079–3086, Apr. 2009, doi: 10.1021/ac900036d.
- [215] C. A. Smith, E. J. Want, G. O'Maille, R. Abagyan, and G. Siuzdak, 'XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification', *Anal. Chem.*, vol. 78, no. 3, pp. 779–787, Feb. 2006, doi: 10.1021/ac051437y.
- [216] R. Tautenhahn, G. J. Patti, D. Rinehart, and G. Siuzdak, 'XCMS Online: A Web-Based Platform to Process Untargeted Metabolomic Data', *Anal. Chem.*, vol. 84, no. 11, pp. 5035–5039, Jun. 2012, doi: 10.1021/ac300698c.
- [217] J. Xia, I. V. Sinelnikov, B. Han, and D. S. Wishart, 'MetaboAnalyst 3.0—making metabolomics more meaningful', *Nucleic Acids Res.*, vol. 43, no. W1, pp. W251–W257, Jul. 2015, doi: 10.1093/nar/gkv380.
- [218] F. Giacomoni *et al.*, 'Workflow4Metabolomics: a collaborative research infrastructure for computational metabolomics', *Bioinformatics*, vol. 31, no. 9, pp. 1493–1495, May 2015, doi: 10.1093/bioinformatics/btu813.

- [219] S. Bijlsma *et al.*, 'Large-Scale Human Metabolomics Studies: A Strategy for Data (Pre-) Processing and Validation', *Anal. Chem.*, vol. 78, no. 2, pp. 567–574, Jan. 2006, doi: 10.1021/ac051495j.
- [220] R. Tautenhahn, C. Böttcher, and S. Neumann, 'Highly sensitive feature detection for high resolution LC/MS', *BMC Bioinformatics*, vol. 9, no. 1, pp. 1–16, Dec. 2008, doi: 10.1186/1471-2105-9-504.
- [221] M. Sindelar and G. J. Patti, 'Chemical Discovery in the Era of Metabolomics', *J. Am. Chem. Soc.*, vol. 142, no. 20, pp. 9097–9105, May 2020, doi: 10.1021/jacs.9b13198.
- [222] T. Barri and L. O. Dragsted, 'UPLC-ESI-QTOF/MS and multivariate data analysis for blood plasma and serum metabolomics: Effect of experimental artefacts and anticoagulant', *Anal. Chim. Acta*, vol. 768, pp. 118–128, Mar. 2013, doi: 10.1016/j.aca.2013.01.015.
- [223] R. Scheltema, S. Decuypere, J. Dujardin, D. Watson, R. Jansen, and R. Breitling, 'Simple data-reduction method for high-resolution LC–MS data in metabolomics', *Bioanalysis*, vol. 1, no. 9, pp. 1551–1557, Dec. 2009, doi: 10.4155/bio.09.146.
- [224] W. B. Dunn, I. D. Wilson, A. W. Nicholls, and D. Broadhurst, 'The importance of experimental design and QC samples in large-scale and MS-driven untargeted metabolomic studies of humans', *Bioanalysis*, vol. 4, no. 18, pp. 2249–2264, Sep. 2012, doi: 10.4155/bio.12.204.
- [225] O. Begou, H. G. Gika, G. A. Theodoridis, and I. D. Wilson, *Quality Control and Validation Issues in LC-MS Metabolomics*, vol. 1738. 2018.
- [226] M. R. Lewis *et al.*, 'Development and Application of Ultra-Performance Liquid Chromatography-TOF MS for Precision Large Scale Urinary Metabolic Phenotyping', *Anal. Chem.*, vol. 88, no. 18, pp. 9004–9013, Sep. 2016, doi: 10.1021/acs.analchem.6b01481.
- [227] M. Eliasson *et al.*, 'Strategy for Optimizing LC-MS Data Processing in Metabolomics: A Design of Experiments Approach', *Anal. Chem.*, vol. 84, no. 15, pp. 6869–6876, Aug. 2012, doi: 10.1021/ac301482k.
- [228] T. J. Demetrowitsch *et al.*, 'Validation of a two-step quality control approach for a large-scale human urine metabolomic study conducted in seven experimental batches with LC/QTOF-MS', *Bioanalysis*, vol. 7, no. 1, pp. 103–112, Jan. 2015, doi: 10.4155/bio.14.270.
- [229] A. Jankevics, M. E. Merlo, M. de Vries, R. J. Vonk, E. Takano, and R. Breitling, 'Separating the wheat from the chaff: a prioritisation pipeline for the analysis of metabolomics datasets', *Metabolomics*, vol. 8, no. 1, pp. 29–36, Jun. 2012, doi: 10.1007/s11306-011-0341-0.
- [230] A. Damont *et al.*, 'Proposal for a chemically consistent way to annotate ions arising from the analysis of reference compounds under ESI conditions: A prerequisite to proper mass spectral database constitution in metabolomics', *J. Mass Spectrom.*, vol. 54, no. 6, pp. 567–582, 2019, doi: <https://doi.org/10.1002/jms.4372>.
- [231] E. A. Thévenot, A. Roux, Y. Xu, E. Ezan, and C. Junot, 'Analysis of the Human Adult Urinary Metabolome Variations with Age, Body Mass Index, and Gender by Implementing a Comprehensive Workflow for Univariate and OPLS Statistical Analyses', *J. Proteome Res.*, vol. 14, no. 8, pp. 3322–3335, Aug. 2015, doi: 10.1021/acs.jproteome.5b00354.
- [232] M. R. Viant, I. J. Kurland, M. R. Jones, and W. B. Dunn, 'How close are we to complete annotation of metabolomes?', *Curr. Opin. Chem. Biol.*, vol. 36, pp. 64–69, Feb. 2017, doi: 10.1016/j.cbpa.2017.01.001.
- [233] R. Chaleckis, I. Meister, P. Zhang, and C. E. Wheelock, 'Challenges, progress and promises of metabolite annotation for LC–MS-based metabolomics', *Curr. Opin. Biotechnol.*, vol. 55, pp. 44–50, Feb. 2019, doi: 10.1016/j.copbio.2018.07.010.
- [234] D. S. Wishart *et al.*, 'HMDB 4.0: the human metabolome database for 2018', *Nucleic Acids Res.*, vol. 46, no. D1, pp. D608–D617, Jan. 2018, doi: 10.1093/nar/gkx1089.
- [235] I. Blaženović, T. Kind, J. Ji, and O. Fiehn, 'Software Tools and Approaches for Compound Identification of LC-MS/MS Data in Metabolomics', *Metabolites*, vol. 8, no. 2, p. 31, May 2018, doi: 10.3390/metabo8020031.
- [236] H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono, and M. Kanehisa, 'KEGG: Kyoto Encyclopedia of Genes and Genomes', *Nucleic Acids Res.*, vol. 27, no. 1, pp. 29–34, Jan. 1999, doi: 10.1093/nar/27.1.29.

- [237] D. S. Wishart *et al.*, 'HMDB: a knowledgebase for the human metabolome', *Nucleic Acids Res.*, vol. 37, no. suppl_1, pp. D603–D610, Jan. 2009, doi: 10.1093/nar/gkn810.
- [238] D. S. Wishart *et al.*, 'HMDB: the Human Metabolome Database', *Nucleic Acids Res.*, vol. 35, no. Database, pp. D521–D526, Jan. 2007, doi: 10.1093/nar/gkl923.
- [239] C. A. Smith *et al.*, 'METLIN: A Metabolite Mass Spectral Database', *Ther. Drug Monit.*, vol. 27, no. 6, pp. 747–751, Dec. 2005, doi: 10.1097/01.ftd.0000179845.53213.39.
- [240] C. Guijas *et al.*, 'METLIN: A Technology Platform for Identifying Knowns and Unknowns', *Anal. Chem.*, vol. 90, no. 5, pp. 3156–3164, Mar. 2018, doi: 10.1021/acs.analchem.7b04424.
- [241] A. D. Southam, R. J. M. Weber, J. Engel, M. R. Jones, and M. R. Viant, 'A complete workflow for high-resolution spectral-stitching nano-electrospray direct-infusion mass-spectrometry-based metabolomics and lipidomics', *Nat. Protoc.*, vol. 12, no. 2, pp. 310–328, Feb. 2017, doi: 10.1038/nprot.2016.156.
- [242] M. Vinaixa, S. Samino, I. Saez, J. Duran, J. J. Guinovart, and O. Yanes, 'A Guideline to Univariate Statistical Analysis for LC/MS-Based Untargeted Metabolomics-Derived Data', *Metabolites*, vol. 2, no. 4, pp. 775–795, Dec. 2012, doi: 10.3390/metabo2040775.
- [243] S. Ren, A. A. Hinzman, E. L. Kang, R. D. Szczesniak, and L. J. Lu, 'Computational and statistical analysis of metabolomics data', *Metabolomics*, vol. 11, no. 6, pp. 1492–1513, Dec. 2015, doi: 10.1007/s11306-015-0823-6.
- [244] B. Worley and R. Powers, 'Multivariate Analysis in Metabolomics', 2013. <https://www.ingentaconnect.com/content/ben/cmb/2013/00000001/00000001/art00008> (accessed Oct. 02, 2020).
- [245] T. Okada, F. Mochamad Afendi, M. Altaf-Ul-Amin, H. Takahashi, K. Nakamura, and S. Kanaya, 'Metabolomics of Medicinal Plants: The Importance of Multivariate Analysis of Analytical Chemistry Data', 2010. <https://www.ingentaconnect.com/content/ben/cad/2010/00000006/00000003/art00003> (accessed Oct. 02, 2020).
- [246] A. Roux, D. Lison, C. Junot, and J.-F. Heilier, 'Applications of liquid chromatography coupled to mass spectrometry-based metabolomics in clinical chemistry and toxicology: A review', *Clin. Biochem.*, vol. 44, no. 1, pp. 119–135, Jan. 2011, doi: 10.1016/j.clinbiochem.2010.08.016.
- [247] E. Saccenti, H. C. J. Hoefsloot, A. K. Smilde, J. A. Westerhuis, and M. M. W. B. Hendriks, 'Reflections on univariate and multivariate analysis of metabolomics data', *Metabolomics*, vol. 10, no. 3, pp. 361–374, Jun. 2014, doi: 10.1007/s11306-013-0598-6.
- [248] M. Kohl, S. Wiese, and B. Warscheid, 'Cytoscape: Software for Visualization and Analysis of Biological Networks', in *Data Mining in Proteomics*, Humana Press, 2011, pp. 291–303.
- [249] T. Shlomi, Y. Eisenberg, R. Sharan, and E. Ruppin, 'A genome-scale computational study of the interplay between transcriptional regulation and metabolism', *Mol. Syst. Biol.*, vol. 3, no. 1, p. 101, Jan. 2007, doi: 10.1038/msb4100141.
- [250] F. Jourdan *et al.*, 'Use of reconstituted metabolic networks to assist in metabolomic data visualization and mining', *Metabolomics*, vol. 6, no. 2, pp. 312–321, Jun. 2010, doi: 10.1007/s11306-009-0196-9.
- [251] L. Cottret *et al.*, 'MetExplore: a web server to link metabolomic experiments and genome-scale metabolic networks', *Nucleic Acids Res.*, vol. 38, no. suppl_2, pp. W132–W137, Jul. 2010, doi: 10.1093/nar/gkq312.
- [252] J. Gao *et al.*, 'Metscape: a Cytoscape plug-in for visualizing and interpreting metabolomic data in the context of human metabolic networks', *Bioinformatics*, vol. 26, no. 7, pp. 971–973, Apr. 2010, doi: 10.1093/bioinformatics/btq048.
- [253] S. Peng *et al.*, 'Newborn Meconium and Urinary Metabolome Response to Maternal Gestational Diabetes Mellitus: A Preliminary Case–Control Study', *J. Proteome Res.*, vol. 14, no. 4, pp. 1799–1809, Apr. 2015, doi: 10.1021/pr5011857.
- [254] N. Naritaka *et al.*, 'Profile of bile acids in fetal gallbladder and meconium using liquid chromatography-tandem mass spectrometry', *Clin. Chim. Acta*, vol. 446, pp. 76–81, Jun. 2015, doi: 10.1016/j.cca.2015.04.008.

- [255] M. Kumagai *et al.*, 'Perinatal bile acid metabolism: bile acid analysis of meconium of preterm and full-term infants', *J. Gastroenterol.*, vol. 42, no. 11, pp. 904–910, Nov. 2007, doi: 10.1007/s00535-007-2108-y.
- [256] C. Righetti, D. G. Peroni, A. Pietrobelli, and C. Zancanaro, 'Proton Nuclear Magnetic Resonance Analysis of Meconium... : Journal of Pediatric Gastroenterology and Nutrition', *LWW*, 2003, Accessed: Sep. 18, 2020. [Online]. Available: https://journals.lww.com/jpgn/Fulltext/2003/04000/Proton_Nuclear_Magnetic_Resonance_Analysis_of.17.aspx.
- [257] R. G. Jensen, A. M. Ferris, C. J. Lammi-Keefe, and R. A. Henderson, 'Lipids of Bovine and Human Milks: A Comparison', *J. Dairy Sci.*, vol. 73, no. 2, pp. 223–240, Feb. 1990, doi: 10.3168/jds.S0022-0302(90)78666-3.
- [258] C. F. Bearer *et al.*, 'Ethyl Linoleate in Meconium: A Biomarker for Prenatal Ethanol Exposure', *Alcohol. Clin. Exp. Res.*, vol. 23, no. 3, pp. 487–493, Mar. 1999, doi: 10.1111/j.1530-0277.1999.tb04142.x.
- [259] J. Ristimaa, M. Gergov, A. Pelander, E. Halmesmäki, and I. Ojanperä, 'Broad-spectrum drug screening of meconium by liquid chromatography with tandem mass spectrometry and time-of-flight mass spectrometry', *Anal. Bioanal. Chem.*, vol. 398, no. 2, pp. 925–935, Sep. 2010, doi: 10.1007/s00216-010-3942-y.
- [260] T. Berton, K. Chardon, F. Mayhoub, J. Caudeville, V. Bach, and K. Tack, 'Characterisation and cartography of foetal exposure to pesticides and theirs metabolites in meconium and maternal hair by UPLC-MS/MS', *23 SETAC Eur. Annu. Meet. "Building Better Future Responsible Innov. Environ. Prot.*, May 2013, [Online]. Available: <https://hal-ineris.archives-ouvertes.fr/ineris-00971153>.
- [261] R.-C. Duca, A. Floch-Barneaud, K. Chardon, F. Mayhoub, V. Bach, and K. Tack, 'In utero exposure to pesticides as measured in meconium - a preliminary study', presented at the Human Biomonitoring : Political Benefits - Scientific challenges, Sep. 2010, Accessed: Oct. 08, 2020. [Online]. Available: <https://hal-ineris.archives-ouvertes.fr/ineris-00970703>.
- [262] P. Tourneux *et al.*, 'Cohorte MecoExpo : utilisation du méconium pour estimer l'exposition in utero aux pesticides des nouveau-nés en Picardie', *Rev. Médecine Périnatale*, vol. 6, no. 2, pp. 122–133, Jun. 2014, doi: 10.1007/s12611-014-0277-5.
- [263] S. Wandro, S. Osborne, C. Enriquez, C. Bixby, A. Arrieta, and K. Whiteson, 'The Microbiome and Metabolome of Preterm Infant Stool Are Personalized and Not Driven by Health Outcomes, Including Necrotizing Enterocolitis and Late-Onset Sepsis', *mSphere*, vol. 3, no. 3, Jun. 2018, doi: 10.1128/mSphere.00104-18.
- [264] K. Bittinger *et al.*, 'Bacterial colonization reprograms the neonatal gut metabolome', *Nat. Microbiol.*, vol. 5, no. 6, Art. no. 6, Jun. 2020, doi: 10.1038/s41564-020-0694-0.
- [265] J. T. Harries, 'MECONIUM IN HEALTH AND DISEASE', *Br. Med. Bull.*, vol. 34, no. 1, pp. 75–78, Jan. 1978, doi: 10.1093/oxfordjournals.bmb.a071462.
- [266] E. M. Ostrea, D. M. Bielawski, and N. C. Posecion, 'Meconium analysis to detect fetal exposure to neurotoxicants', *Arch. Dis. Child.*, vol. 91, no. 8, pp. 628–629, Aug. 2006, doi: 10.1136/adc.2006.097956.
- [267] K. Cheng, C. Brunius, R. Fristedt, and R. Landberg, 'An LC-QToF MS based method for untargeted metabolomics of human fecal samples', *Metabolomics Off. J. Metabolomic Soc.*, vol. 16, no. 4, p. 46, 03 2020, doi: 10.1007/s11306-020-01669-z.
- [268] S. Moosmang, M. Pitscheider, S. Sturm, and C. Seger, 'Metabolomic analysis—Addressing NMR and LC-MS related problems in human feces sample preparation', *Clin. Chim. Acta*, vol. 489, pp. 169–176, Feb. 2019, doi: 10.1016/j.cca.2017.10.029.
- [269] J. Wu, Y. An, J. Yao, Y. Wang, and H. Tang, 'An optimised sample preparation method for NMR-based faecal metabonomic analysis', *Analyst*, vol. 135, no. 5, pp. 1023–1030, Apr. 2010, doi: 10.1039/B927543F.
- [270] J. Zierer *et al.*, 'The fecal metabolome as a functional readout of the gut microbiome', *Nat. Genet.*, vol. 50, no. 6, pp. 790–795, Jun. 2018, doi: 10.1038/s41588-018-0135-7.

- [271] M. Berdi *et al.*, 'Immune components of early breastmilk: Association with maternal factors and with reported food allergy in childhood', *Pediatr. Allergy Immunol.*, vol. 30, no. 1, pp. 107–116, Feb. 2019, doi: 10.1111/pai.12998.
- [272] T. Berton *et al.*, 'Development of an analytical strategy based on LC–MS/MS for the measurement of different classes of pesticides and their metabolites in meconium: Application and characterisation of foetal exposure in France', *Environ. Res.*, vol. 132, pp. 311–320, Jul. 2014, doi: 10.1016/j.envres.2014.03.034.
- [273] P. Prego-Meleiro, E. Lendoiro, M. Concheiro, A. Cruz, M. López-Rivadulla, and A. de Castro, 'Development and validation of a liquid chromatography tandem mass spectrometry method for the determination of cannabinoids and phase I and II metabolites in meconium', *J. Chromatogr. A*, vol. 1497, pp. 118–126, May 2017, doi: 10.1016/j.chroma.2017.03.066.
- [274] O. Deda, H. G. Gika, I. D. Wilson, and G. A. Theodoridis, 'An overview of fecal sample preparation for global metabolic profiling', *J. Pharm. Biomed. Anal.*, vol. 113, pp. 137–150, Sep. 2015, doi: 10.1016/j.jpba.2015.02.006.
- [275] B. K. Rubin, R. P. Tomkiewicz, M. E. Patrinos, and D. Easa, 'The Surface and Transport Properties of Meconium and Reconstituted Meconium Solutions', *Pediatr. Res.*, vol. 40, no. 6, Art. no. 6, Dec. 1996, doi: 10.1203/00006450-199612000-00010.
- [276] K. Klaus-Michael, K. Richard, and E. Klaus, 'Fecal α 1-Antitrypsin in Newborn Infants : Journal of Pediatric Gastroenterology and Nutrition', *LWW*, Accessed: Oct. 07, 2020. [Online]. Available: https://journals.lww.com/jpgn/Fulltext/1997/03000/Fecal__1_Antitrypsin_in_Newborn_Infants.7.aspx.
- [277] Y. Guitton, MarieTremblay-Franco, Gildas Le Corguillé, and Jean-François Martin, 'Create, run, share, publish, and reference your LC–MS, FIA–MS, GC–MS, and NMR data analysis workflows with the Workflow4Metabolomics 3.0 Galaxy online infrastructure for metabolomics', *Int. J. Biochem. Cell Biol.*, vol. 93, pp. 89–101, Dec. 2017, doi: 10.1016/j.biocel.2017.07.002.
- [278] R. t'Kindt, K. Morreel, D. Deforce, W. Boerjan, and J. Van Bocxlaer, 'Joint GC–MS and LC–MS platforms for comprehensive plant metabolomics: Repeatability and sample pre-treatment', *J. Chromatogr. B*, vol. 877, no. 29, pp. 3572–3580, Nov. 2009, doi: 10.1016/j.jchromb.2009.08.041.
- [279] L. Lin *et al.*, 'Direct infusion mass spectrometry or liquid chromatography mass spectrometry for human metabolomics? A serum metabolomic study of kidney cancer', *Analyst*, vol. 135, no. 11, pp. 2970–2978, 2010, doi: 10.1039/C0AN00265H.
- [280] N. Le Floc'h, W. Otten, and E. Merlot, 'Tryptophan metabolism, from nutrition to potential therapeutic applications', *Amino Acids*, vol. 41, no. 5, pp. 1195–1205, Nov. 2011, doi: 10.1007/s00726-010-0752-7.
- [281] J. Chong *et al.*, 'MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis', *Nucleic Acids Res.*, vol. 46, no. W1, pp. W486–W494, Jul. 2018, doi: 10.1093/nar/gky310.
- [282] E. M. Hartmann, F. Allain, J.-C. Gaillard, O. Pible, and J. Armengaud, 'Taking the Shortcut for High-Throughput Shotgun Proteomic Analysis of Bacteria', in *Host-Bacteria Interactions: Methods and Protocols*, A. C. Vergunst and D. O'Callaghan, Eds. New York, NY: Springer, 2014, pp. 275–285.
- [283] L. Grenga *et al.*, 'Shotgun proteomics analysis of SARS-CoV-2-infected cells and how it can optimize whole viral particle antigen production for vaccines', *Emerg. Microbes Infect.*, vol. 9, no. 1, pp. 1712–1721, Jan. 2020, doi: 10.1080/22221751.2020.1791737.
- [284] J. A. Christie-Oleza, J. M. Piña-Villalonga, R. Bosch, B. Nogales, and J. Armengaud, 'Comparative Proteogenomics of Twelve *Roseobacter* Exoproteomes Reveals Different Adaptive Strategies Among These Marine Bacteria', *Mol. Cell. Proteomics*, vol. 11, no. 2, p. M111.013110, Feb. 2012, doi: 10.1074/mcp.M111.013110.
- [285] C. Quast *et al.*, 'The SILVA ribosomal RNA gene database project: improved data processing and web-based tools', *Nucleic Acids Res.*, vol. 41, no. Database issue, pp. D590–D596, Jan. 2013, doi: 10.1093/nar/gks1219.
- [286] F. Escudié *et al.*, 'FROGS: Find, Rapidly, OTUs with Galaxy Solution', *Bioinformatics*, vol. 34, no. 8, pp. 1287–1294, Apr. 2018, doi: 10.1093/bioinformatics/btx791.

- [287] F. Mahé, T. Rognes, C. Quince, C. de Vargas, and M. Dunthorn, 'Swarm: robust and fast clustering method for amplicon-based studies', *PeerJ*, vol. 2, p. e593, Sep. 2014, doi: 10.7717/peerj.593.
- [288] E. B. Fichot and R. S. Norman, 'Microbial phylogenetic profiling with the Pacific Biosciences sequencing platform', *Microbiome*, vol. 1, no. 1, p. 10, Mar. 2013, doi: 10.1186/2049-2618-1-10.
- [289] L. F. Blackwell, P. Vigil, B. Gross, C. d'Arcangues, D. G. Cooke, and J. B. Brown, 'Monitoring of ovarian activity by measurement of urinary excretion rates of estrone glucuronide and pregnanediol glucuronide using the Ovarian Monitor, Part II: reliability of home testing', *Hum. Reprod.*, vol. 27, no. 2, pp. 550–557, Feb. 2012, doi: 10.1093/humrep/der409.
- [290] G. Scherer, T. Renner, and M. Meger, 'Analysis and evaluation of trans,trans-muconic acid as a biomarker for benzene exposure', *J. Chromatogr. B. Biomed. Sci. App.*, vol. 717, no. 1, pp. 179–199, Oct. 1998, doi: 10.1016/S0378-4347(98)00065-6.
- [291] Y. Wu *et al.*, 'Multiple elements related to metabolic markers in the context of gestational diabetes mellitus in meconium', *Environ. Int.*, vol. 121, pp. 1227–1234, Dec. 2018, doi: 10.1016/j.envint.2018.10.044.
- [292] R. Bujak, E. Dagher-Wojtkowiak, R. Kaliszan, and M. J. Markuszewski, 'PLS-Based and Regularization-Based Methods for the Selection of Relevant Variables in Non-targeted Metabolomics Data', *Mol. Biosci.*, p. 35, 2016, doi: 10.3389/fmolb.2016.00035.
- [293] A. L. C. Velho *et al.*, 'Metabolomic markers of fertility in bull seminal plasma', *PLOS ONE*, vol. 13, no. 4, p. e0195279, Apr. 2018, doi: 10.1371/journal.pone.0195279.
- [294] A. Kumar, T. Kroetsch, P. Blondin, and M. Anzar, 'Fertility-associated metabolites in bull seminal plasma and blood serum: ¹H nuclear magnetic resonance analysis: BULL SEMINAL PLASMA AND SERUM METABOLITES', *Mol. Reprod. Dev.*, vol. 82, no. 2, pp. 123–131, Feb. 2015, doi: 10.1002/mrd.22450.
- [295] E. Skarzyńska, B. Kiersztyn, P. Wilczyńska, A. Jakimiuk, and B. Lisowska-Myjak, 'Total proteolytic activity and concentration of alpha-1 antitrypsin in meconium for assessment of the protease/antiprotease balance', *Eur. J. Obstet. Gynecol. Reprod. Biol.*, vol. 223, pp. 133–138, Apr. 2018, doi: 10.1016/j.ejogrb.2017.10.006.
- [296] P. Wilczyńska, B. Lisowska-Myjak, and E. Skarzyńska, 'Meconium proteases and antiproteases as a potential source of biomarkers for the assessment of the intrauterine environment of the fetus', *J. Matern. Fetal Neonatal Med.*, vol. 0, no. 0, pp. 1–7, Nov. 2020, doi: 10.1080/14767058.2020.1846698.
- [297] L. A. Davidson and B. Lönnerdal, 'Persistence of human milk proteins in the breast-fed infant', *Acta Paediatr. Scand.*, vol. 76, no. 5, pp. 733–740, Sep. 1987, doi: 10.1111/j.1651-2227.1987.tb10557.x.
- [298] Y. Zhou *et al.*, 'Metascape provides a biologist-oriented resource for the analysis of systems-level datasets', *Nat. Commun.*, vol. 10, no. 1, Art. no. 1, Apr. 2019, doi: 10.1038/s41467-019-09234-6.
- [299] B. Lisowska-Myjak and J. Pachecka, 'Trypsin and antitrypsin activities and protein concentration in serial meconium and feces of healthy newborns', *J. Matern. Fetal Neonatal Med.*, vol. 19, no. 8, pp. 477–482, Jan. 2006, doi: 10.1080/14767050600746720.
- [300] K. D. Pruitt, T. Tatusova, and D. R. Maglott, 'NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins', *Nucleic Acids Res.*, vol. 33, no. suppl_1, pp. D501–D504, Jan. 2005, doi: 10.1093/nar/gki025.
- [301] O. Pible, F. Allain, V. Jouffret, K. Culotta, G. Miotello, and J. Armengaud, 'Estimating relative biomasses of organisms in microbiota using "phylopeptidomics"', *Microbiome*, vol. 8, no. 1, p. 30, Mar. 2020, doi: 10.1186/s40168-020-00797-x.
- [302] R. D. Zwiittink *et al.*, 'Metaproteomics reveals functional differences in intestinal microbiota development of preterm infants', p. 11.
- [303] C. Reinhardt, C. Reigstad, and F. Bäckhed, 'Intestinal Microbiota During Infancy and Its Implications for Obesity', *J. Pediatr. Gastroenterol. Nutr.*, vol. 48, pp. 249–56, Apr. 2009, doi: 10.1097/MPG.0b013e318183187c.
- [304] K. Bittinger, 'Bacterial colonization reprograms the neonatal gut metabolome', *Nat. Microbiol.*, p. 22.

- [305] G. Wang, 'Bridging intestinal immunity and gut microbiota by metabolites', p. 21.
- [306] J. Penders *et al.*, 'Factors influencing the composition of the intestinal microbiota in early infancy', *Pediatrics*, vol. 118, no. 2, pp. 511–521, Aug. 2006, doi: 10.1542/peds.2005-2824.
- [307] F. H. Heida *et al.*, 'A Necrotizing Enterocolitis-Associated Gut Microbiota Is Present in the Meconium: Results of a Prospective Study', *Clin. Infect. Dis.*, vol. 62, no. 7, pp. 863–870, Apr. 2016, doi: 10.1093/cid/ciw016.
- [308] X. Liu, 'Thèse Liu Xinxin, 2013, AgroParisTech'.
- [309] M. Gt and M. S, 'Bacteria, colonic fermentation, and gastrointestinal health.', *J. AOAC Int.*, vol. 95, no. 1, pp. 50–60, Jan. 2012, doi: 10.5740/jaoacint.sge_macfarlane.
- [310] E. A. Smith and G. T. Macfarlane, 'Enumeration of human colonic bacteria producing phenolic and indolic compounds: effects of pH, carbohydrate availability and retention time on dissimilatory aromatic amino acid metabolism', *J. Appl. Bacteriol.*, vol. 81, no. 3, pp. 288–302, 1996, doi: <https://doi.org/10.1111/j.1365-2672.1996.tb04331.x>.
- [311] S. R. Elsdén, M. G. Hilton, and J. M. Waller, 'The end products of the metabolism of aromatic amino acids by clostridia', *Arch. Microbiol.*, vol. 107, no. 3, pp. 283–288, Apr. 1976, doi: 10.1007/BF00425340.
- [312] D. Dodd *et al.*, 'A gut bacterial pathway metabolizes aromatic amino acids into nine circulating metabolites', *Nature*, vol. 551, no. 7682, Art. no. 7682, Nov. 2017, doi: 10.1038/nature24661.
- [313] B. V. Jones, M. Begley, C. Hill, C. G. M. Gahan, and J. R. Marchesi, 'Functional and comparative metagenomic analysis of bile salt hydrolase activity in the human gut microbiome', *Proc. Natl. Acad. Sci.*, vol. 105, no. 36, pp. 13580–13585, Sep. 2008, doi: 10.1073/pnas.0804437105.
- [314] W. Jia, G. Xie, and W. Jia, 'Bile acid–microbiota crosstalk in gastrointestinal inflammation and carcinogenesis', *Nat. Rev. Gastroenterol. Hepatol.*, vol. 15, no. 2, Art. no. 2, Feb. 2018, doi: 10.1038/nrgastro.2017.119.
- [315] M. L. Di Martino, R. Campilongo, M. Casalino, G. Micheli, B. Colonna, and G. Prosseda, 'Polyamines: Emerging players in bacteria–host interactions', *Int. J. Med. Microbiol.*, vol. 303, no. 8, pp. 484–491, Dec. 2013, doi: 10.1016/j.ijmm.2013.06.008.
- [316] Y. Gagnebin *et al.*, 'Metabolomic analysis of urine samples by UHPLC-QTOF-MS: Impact of normalization strategies', *Anal. Chim. Acta*, vol. 955, pp. 27–35, Feb. 2017, doi: 10.1016/j.aca.2016.12.029.
- [317] C. Galéra *et al.*, 'Prenatal Caffeine Exposure and Child IQ at Age 5.5 Years: The EDEN Mother-Child Cohort', *Biol. Psychiatry*, vol. 80, no. 9, pp. 720–726, Nov. 2016, doi: 10.1016/j.biopsych.2015.08.034.
- [318] M. Melchior *et al.*, 'Maternal tobacco smoking in pregnancy and children's socio-emotional development at age 5: The EDEN mother-child birth cohort study', *Eur. Psychiatry*, vol. 30, no. 5, pp. 562–568, Jul. 2015, doi: 10.1016/j.eurpsy.2015.03.005.
- [319] C. Galera *et al.*, 'Prenatal diet and children's trajectories of hyperactivity-inattention and conduct problems from 3 to 8 years: the EDEN mother-child cohort', *J. Child Psychol. Psychiatry*, vol. 59, no. 9, pp. 1003–1011, Sep. 2018, doi: 10.1111/jcpp.12898.
- [320] C. Junot, F. Fenaille, B. Colsch, and F. Bécher, 'High resolution mass spectrometry based techniques at the crossroads of metabolic pathways: HIGH RESOLUTION MASS SPECTROMETRY FOR METABOLOMICS', *Mass Spectrom. Rev.*, vol. 33, no. 6, pp. 471–500, Nov. 2014, doi: 10.1002/mas.21401.
- [321] 'Review of the potential health impact of β -casomorphins and related peptides', *EFSA J.*, no. EFSA Journal, doi: 10.2903/j.efsa.2009.231r.
- [322] H. He *et al.*, 'Physalin A induces apoptosis via p53-Noxa-mediated ROS generation, and autophagy plays a protective role against apoptosis through p38-NF- κ B survival pathway in A375-S2 cells', *J. Ethnopharmacol.*, vol. 148, no. 2, pp. 544–555, Jul. 2013, doi: 10.1016/j.jep.2013.04.051.
- [323] N. B. Pinto *et al.*, 'Topical anti-inflammatory potential of Physalin E from *Physalis angulata* on experimental dermatitis in mice', *Phytomedicine*, vol. 17, no. 10, pp. 740–743, Aug. 2010, doi: 10.1016/j.phymed.2010.01.006.

- [324] Madej et al., 'Analysis of meconium, nails and tears for determination of medicines and drugs of abuse', *TrAC Trends Anal. Chem.*, vol. 29, no. 3, pp. 246–259, Mar. 2010, doi: 10.1016/j.trac.2010.01.005.
- [325] M. K. Woźniak, E. Jaszczak, M. Wiergowski, Ż. Polkowska, J. Namieśnik, and M. Biziuk, 'Meconium analysis as a promising diagnostic tool for monitoring fetal exposure to toxic substances: Recent trends and perspectives', *TrAC Trends Anal. Chem.*, vol. 109, pp. 124–141, Dec. 2018, doi: 10.1016/j.trac.2018.09.025.
- [326] E. M. Ostrea, 'Testing for exposure to illicit drugs and other agents in the neonate: A review of laboratory methods and the role of meconium analysis', *Curr. Probl. Pediatr.*, vol. 29, no. 2, pp. 41–56, Feb. 1999, doi: 10.1016/S0045-9380(99)80046-2.
- [327] M. Oldiges, S. Lütz, S. Pflug, K. Schroer, N. Stein, and C. Wiendahl, 'Metabolomics: current state and evolving methodologies and tools', *Appl. Microbiol. Biotechnol.*, vol. 76, no. 3, pp. 495–511, Sep. 2007, doi: 10.1007/s00253-007-1029-2.
- [328] M. Suzuki, S. Nishiumi, A. Matsubara, T. Azuma, and M. Yoshida, 'Metabolome analysis for discovering biomarkers of gastroenterological cancer', *J. Chromatogr. B*, vol. 966, pp. 59–69, Sep. 2014, doi: 10.1016/j.jchromb.2014.02.042.
- [329] K. Adel-Patient *et al.*, 'Deep analysis of immune response and metabolic signature in children with food protein induced enterocolitis to cow's milk', *Clin. Transl. Allergy*, vol. 8, no. 1, p. 38, Sep. 2018, doi: 10.1186/s13601-018-0224-9.
- [330] Peng et al. *et al.*, 'Newborn Meconium and Urinary Metabolome Response to Maternal Gestational Diabetes Mellitus: A Preliminary Case–Control Study', *J. Proteome Res.*, vol. 14, no. 4, pp. 1799–1809, Apr. 2015, doi: 10.1021/pr5011857.
- [331] E. M. Ostrea, D. K. Knapp, A. Romero, M. Montes, and A. R. Ostrea, 'Meconium analysis to assess fetal exposure to nicotine by active and passive maternal smoking', *J. Pediatr.*, vol. 124, no. 3, pp. 471–476, Mar. 1994, doi: 10.1016/S0022-3476(94)70378-7.
- [332] J. Ristimaa, M. Gergov, A. Pelander, E. Halmesmäki, and I. Ojanperä, 'Broad-spectrum drug screening of meconium by liquid chromatography with tandem mass spectrometry and time-of-flight mass spectrometry', *Anal. Bioanal. Chem.*, vol. 398, no. 2, pp. 925–935, Sep. 2010, doi: 10.1007/s00216-010-3942-y.
- [333] A. Bakdash *et al.*, 'Quantification of fatty acid ethyl esters (FAEE) and ethyl glucuronide (EtG) in meconium from newborns for detection of alcohol abuse in a maternal health evaluation study', *Anal. Bioanal. Chem.*, vol. 396, no. 7, pp. 2469–2477, Apr. 2010, doi: 10.1007/s00216-010-3474-5.
- [334] P. Cabarcos *et al.*, 'Quantification of fatty acid ethyl esters (FAEE) and ethyl glucuronide (EtG) in meconium for detection of alcohol abuse during pregnancy: Correlation study between both biomarkers', *J. Pharm. Biomed. Anal.*, vol. 100, pp. 74–78, Nov. 2014, doi: 10.1016/j.jpba.2014.07.023.
- [335] L. N. Bakhireva *et al.*, 'Prenatal alcohol exposure prevalence as measured by direct ethanol metabolites in meconium in a Native American tribe of the southwest', *Birth Defects Res.*, p. bdr2.1427, Dec. 2018, doi: 10.1002/bdr2.1427.
- [336] S. Wandro, S. Osborne, C. Enriquez, C. Bixby, A. Arrieta, and K. Whiteson, 'The Microbiome and Metabolome of Preterm Infant Stool Are Personalized and Not Driven by Health Outcomes, Including Necrotizing Enterocolitis and Late-Onset Sepsis', *mSphere*, vol. 3, no. 3, pp. e00104-18, /msphere/3/3/mSphere104-18.atom, Jun. 2018, doi: 10.1128/mSphere.00104-18.
- [337] C. Junot, F. Fenaille, B. Colsch, and F. Bécher, 'High resolution mass spectrometry based techniques at the crossroads of metabolic pathways', *Mass Spectrom. Rev.*, vol. 33, no. 6, pp. 471–500, 2014, doi: 10.1002/mas.21401.
- [338] E. M. Ostrea *et al.*, 'Combined analysis of prenatal (maternal hair and blood) and neonatal (infant hair, cord blood and meconium) matrices to detect fetal exposure to environmental pesticides', *Environ. Res.*, vol. 109, no. 1, pp. 116–122, Jan. 2009, doi: 10.1016/j.envres.2008.09.004.

- [339] M. Vinaixa, S. Samino, I. Saez, J. Duran, J. J. Guinovart, and O. Yanes, 'A Guideline to Univariate Statistical Analysis for LC/MS-Based Untargeted Metabolomics-Derived Data', *Metabolites*, vol. 2, no. 4, pp. 775–795, Oct. 2012, doi: 10.3390/metabo2040775.
- [340] S. Peng *et al.*, 'Newborn Meconium and Urinary Metabolome Response to Maternal Gestational Diabetes Mellitus: A Preliminary Case–Control Study', *J. Proteome Res.*, vol. 14, no. 4, pp. 1799–1809, Apr. 2015, doi: 10.1021/pr5011857.
- [341] R. Cassoulet *et al.*, 'Monitoring of prenatal exposure to organic and inorganic contaminants using meconium from an Eastern Canada cohort', *Environ. Res.*, vol. 171, pp. 44–51, Apr. 2019, doi: 10.1016/j.envres.2018.12.044.
- [342] J. Baranowski, G. Pochopień, and I. Baranowska, 'Determination of nicotine, cotinine and caffeine in meconium using high-performance liquid chromatography', *J. Chromatogr. B. Biomed. Sci. App.*, vol. 707, no. 1, pp. 317–321, Apr. 1998, doi: 10.1016/S0378-4347(97)00619-1.
- [343] D. Bielawski, E. Ostrea, N. Posecion, M. Corrión, and J. Seagraves, 'Detection of Several Classes of Pesticides and Metabolites in Meconium by Gas Chromatography-Mass Spectrometry', *Chromatographia*, vol. 62, no. 11, pp. 623–629, Dec. 2005, doi: 10.1365/s10337-005-0668-7.

PRODUCTION SCIENTIFIQUE

PUBLICATIONS

- LC/HRMS-based metabolomics profiling of human meconium

Nihel Bekhti, Florence Castelli, Blanche Guillon, Christophe Junot, François Fenaille & Karine Adel-Patient.

- Basal Ti level in the human placenta and meconium and evidence of a maternofetal transfer of food-grade TiO₂ nanoparticles in an ex vivo placental perfusion model

A. Guillard, E. Gaultier, C. Cartier, L. Devoille, J. Noireaux, L. Chevalier, M. Morin, F. Grandin, M. Z. Lacroix, C. Coméra, A. Cazanave, A. de Place, V. Gayrard, V. Bach, K. Chardon, N. Bekhti, K. Adel-Patient, C. Vayssière, P. Fisicaro, N. Feltin, F. de la Farge, N. Picard-Hagen, B. Lamas & E. Houdeau.

COMMUNICATIONS ORALES

- Individual cartography of human meconium: a picture of the perinatal exposome correlated to health child development.

Nihel Bekhti, Florence Castelli, Blanche Guillon, Christophe Junot, François Fenaille & Karine Adel-Patient.

JOURNEES DES DOCTORANTS DE L'INSTITUT DE JOLIOT, CEA-SACLAY, FRANCE, 2019

- Individual cartography of human meconium

Nihel Bekhti, Florence Castelli, Blanche Guillon, Christophe Junot, François Fenaille & Karine Adel-Patient.

JOURNEES DES DOCTORANTS DU DEPARTEMENT MEDICAMENTS ET TECHNOLOGIES POUR LA SANTE CEA-SACLAY, FRANCE, 2019

- Metabolomics profiling of meconium by LC/HRMS (TOP 3 des presentations)

Nihel Bekhti, Florence Castelli, Blanche Guillon, Christophe Junot, François Fenaille & Karine Adel-Patient.

ECOLE JEUNE DE LA SFSM, QUIBERON, FRANCE, 2019

- Metabolomics profiling of human meconium by LC/HRMS

Nihel Bekhti, Florence Castelli, Blanche Guillon, Christophe Junot, François Fenaille & Karine Adel-Patient.

COLLOQUE DES JEUNES CHERCHEURS DU CENTRE DE JOUY, INRAe, FRANCE, 2019

COMMUNICATIONS PAR AFFICHES

- Cartographie individuelle de la composition de méconiums humains : un reflet de l'exposome prénatal corrélé au développement d'une allergie alimentaire dans l'enfance? (2^{ème} prix de Poster)

Nihel Bekhti, Florence Castelli, Blanche Guillon, Christophe Junot, François Fenaille & Karine Adel-Patient.

COLLOQUE DES JEUNES CHERCHEURS DU CENTRE DE JOUY, INRAe, FRANCE, 2018

- Analyse métabolomique non ciblée du méconium humain: développement méthodologique

Nihel Bekhti, Florence Castelli, Blanche Guillon, Christophe Junot, François Fenaille & Karine Adel-Patient.

11^{ème} JOURNEES SCIENTIFIQUES DU RFMF, LIEGE, BELGIQUE, 2018

- Analyse métabolomique non ciblée du méconium humain: développement méthodologique

Nihel Bekhti, Florence Castelli, Blanche Guillon, Christophe Junot, François Fenaille & Karine Adel-Patient.

JOURNEES DES DOCTORANTS DE L'INSTITUT DE JOLIOT, CEA-SACLAY, FRANCE, 2018

- Metabolomics profiling of human meconium by LC/HRMS

Nihel Bekhti, Florence Castelli, Blanche Guillon, Estelle Paris, Christophe Junot, François Fenaille & Karine Adel-Patient.

METABOLOMICS SOCIETY, LA HAYE, PAYS-BAS, 2019

PRIX ET DISTINCTIONS

- Femme enceinte et santé de l'enfant (1^{er} prix du concours)

Nihel Bekhti, Florence Castelli, Blanche Guillon, Christophe Junot, Estelle Paris, Stephane Nano, François Fenaille & Karine Adel-Patient,

CONCOURS SCIENCETUBE, CEA FRANCE, <https://www.youtube.com/watch?v=OhnyxcFf8uE>

- Travel award

Nihel Bekhti, Florence Castelli, Blanche Guillon, Christophe Junot, Estelle Paris, Stephane Nano, François Fenaille & Karine Adel-Patient,

RESEAU FRANCOPHONE DE METABOLOMIQUE ET DE FLUXOMIQUE, 2019

ANNEXES

LC/HRMS-based metabolomics profiling of human meconium

Nihel Bekhti¹, Florence Castelli¹, Blanche Guillon¹, Christophe Junot¹, François Fenaille^{1, 2#} & Karine Adel-Patient^{1,2#}

¹ Département Médicaments et Technologies pour la Santé (DMTS), CEA, INRAE, Université Paris-Saclay, Gif-sur-Yvette, 91191, France.

² These authors jointly supervised this work.

Corresponding authors

Karine Adel-Patient, email: karine.patient@cea.fr

François Fenaille, email: francois.fenaille@cea.fr

Counts: 5650 words.

Keywords

Meconium, metabolome extraction, metabolites identification, prenatal exposure, liquid chromatography-ultra high mass spectrometry, workflow for metabolomics

Introduction

The exposome comprises all the individual genetic and non-genetic exposures encountered by an individual from conception to death (Paul Wild et al). In the neonatal window, these exposures could directly help elucidating some pathologies development. Detecting timely these exposures may help to provide treatments to the exposed neonates as early as possible to avoid any serious consequences. To investigate the knowledge on human exposure in utero (i.e., diet, lifestyle, environment...), newborn's meconium shows to be an interesting biological matrix. Meconium is the first new-born stool excreted during the first 24 hours [1] and until day 5 after birth [2]. Its starts forming from the 12th week of gestation and accumulates in the fetus gut until birth, and represents no metabolic functions [2], [3]. Thus, analyzing its composition seems to be a promising strategy as it concentrates a large information of components transfer trans-placentally from the mother to the fetus during the last months of gestation [2], [4]. Such components may derived from mother metabolism (both its own metabolism and her microbiota), but also from environmental exposure (contaminants, diet...). Studying this matrix may help to elucidate the impact of neonatal exposures on child development. For this purpose, we aim to provide a description of all metabolites within the meconium, as a first and reduced view of in utero "exposome".

Metabolomics analysis of meconium show to be a promising tool helping to understand the foetal exposure by monitoring and analyzing the meconium composition. Metabolomics have recorded a remarkable expansion over the last decade [5] helping to elucidate several human pathologies (i.e. asthma, cancers and gestational diabetes...) [6]–[8]. These analysis consist in identifying as much metabolites as possible including all small molecular mass compounds in meconium [3], [9]. This cartography is capable of capturing disease-relevant profile changes and may help to identify metabolic alteration.

To date, the two main technical approaches used for metabolic data generation of meconium are resonance magnetic nuclear (Righetti et al 2003) and high resolution mass spectrometry. Righetti et al have analysed both water-soluble and organic fractions of meconium and indicate a much higher content found in the water-soluble section. Several metabolites are detected in both fractions encountering fatty acids, sterols and amino-acids. A quantitative analysis of these metabolites

concentration in meconium samples from day 1 to day 3 after excretion is realized. A significant increase of concentration is noticed within several metabolites associated with the starting of breast-feeding. To uncover a biggest range of metabolic detection, mass spectrometry shows to be the technique of choice thanks to its high sensitivity. The LC/HRMS hyphenation is the most used technique for metabolic analysis of meconium. Nevertheless, only a few studies have described a global view of chemical compounds of meconium using this technique. These studies mostly described a particular class of compounds or even single metabolites compounds involving specific meconium preparation steps. Ristimaa et al used LC/TOF-MS technique to investigate a screening composition of drugs in meconium. 12 components were targeted, and 77 new drug components were identified in MS/MS [10]. Meyer et al focus on describing a family of pesticides, chlorinated solvents and xenobiotic in meconium helping to define the effect of such exposures on the fetus development [1], [11], [12]. Fatty acid ethyl esters found in meconium are defined as biomarkers of fetal ethanol exposure during pregnancy that several studies focus on [13]–[16]. Wandro et al made a global metabolic investigation of early fecal samples by GC/MS describing several metabolic families including amino-acids, fatty and organic acids, sugars, lipids, and nucleotides [17]. Performing such a view on meconium samples is interesting and still undescribed in the literature to date.

Metabolomics profiling need a detailed workflow to assure obtaining a significant and reproducible data [18]. A first step of meconium preparation is developed and represents a crucial step to guarantee reproducible measurements, normalize between the samples and assure a relevant metabolic extraction [19]. A second step of data acquisition is processed by LC/HRMS, owed to the diverse physico-chemical properties of metabolites, two complementary chromatography columns are used [20], [21]. A high volume of data is generated within metabolomics profiling and need to be processed, automatized and annotated using several tools and software [22]. A step of chemo-metric analysis is realized to help a better visualization of the data [18]. Finally, a metabolic identification is achieved using MS/MS analysis to guarantee high level of identification and help elucidating biomarkers.

In this context, this article aims to describe the global workflow used in meconium metabolic profiling. It includes the identification with a high level of confidence of the highest number possible of metabolites transferred from the mother to foetus during pregnancy. To the best of the authors' knowledge, this is the first article aiming to describe a global cartography of meconium composition.

Material and methods

Chemicals and Reagents

All analytical grade reference compounds were from Sigma (Saint Quentin Fallavier, France). The standards mixtures used for the external calibration of the MS instrument: Calmix positive, for the

positive ionization mode, consisting of caffeine, L-Methionyl-arginyl-phenylalanyl-alanine acetate was from Thermo Fisher Scientific (Courtaboeuf, France). For the negative ionization mode, the calibration mixture consists in a homemade solution of sodium acetate at 2mM in H₂O/ACN (50/50, v/v). Acetonitrile (ACN), formic acid (FA) were from VWR chemicals, methanol (MeOH) and water were from (Honeywell, riedel-de-Haën™), and ammonium carbonate from (Sigma Aldrich chemicals).

Sample collection

Meconium samples studied were obtained from Clinique Sainte Therese, Paris France from 11 anonymous and vaginal birth new-borns (Table 1). 33 meconium samples concerning multiple collections within the same new-born are obtained to: i. Describe the global meconium metabolome, ii. Evaluate the meconium composition evolution during the first days of life. The diapers were directly stored at 4°C until aliquoting in sterile tubes then frozen at -20°C in the maternity. Finally, the samples are kept at -80°C in the laboratory prior to analysis. It is important to note that only diapers containing meconium aspect have been sampled but need to be confirmed by studying its evolution day to day.

Table 1. Sampling time of the 33 specimens of the study.

New-born (NB) (Girl/Boy)	Sampling time (minutes after birth)	Number of samples per NB
NB1(G)	390 / 915 / 3060	3
NB2 (B)	570 / 1005 / 2640	3
NB3 (B)	420 / 1440 / 4740	3
NB4 (B)	1170 / 1620	2
NB5 (B)	420 / 930	2
NB6 (B)	960 / 2550	2
NB7 (B)	1140 / 1350 / 2580	3
NB8 (B)	270 / 1020 / 1200 / 2640 / 2880	5
NB9 (G)	195 / 1215 / 1305 / 1920 / 2700 / 2820	6
NB10 (B)	90 / 1020 / 1065	3
NB11 (B)	795	1

Sample preparation

A first step of meconium freeze-drying is achieved using a Triad™ Labconco lyophilizer with fixed temperatures of 4°C for the tray and -83°C for the trap; the vacuum is fixed at 0.180 mbar. The freeze-dried sample is homogenized and stored at -80°C until analysis. To perform protein precipitation, 10 mg freeze-dried meconium is suspended in 750 µL of methanol/H₂O (4:1 v/v). The sample is homogenized

using a Precellys 24® (Bertin Technologies, Montigny-le-Bretonneux, France), with CK14 ceramic beads 1.4 mm tubes (6500 tr/min; 4 °C; 3 x 30 seconds). The extract is kept on ice for 1h30 to complete protein precipitation. After a centrifugation (20000 g ; 4 °C ; 15 minutes), the supernatant is recovered and dried under nitrogen at 30°C using a TurboVap® concentration workstation from Biotage France. The pellet is stored dried at -80°C until further analysis.

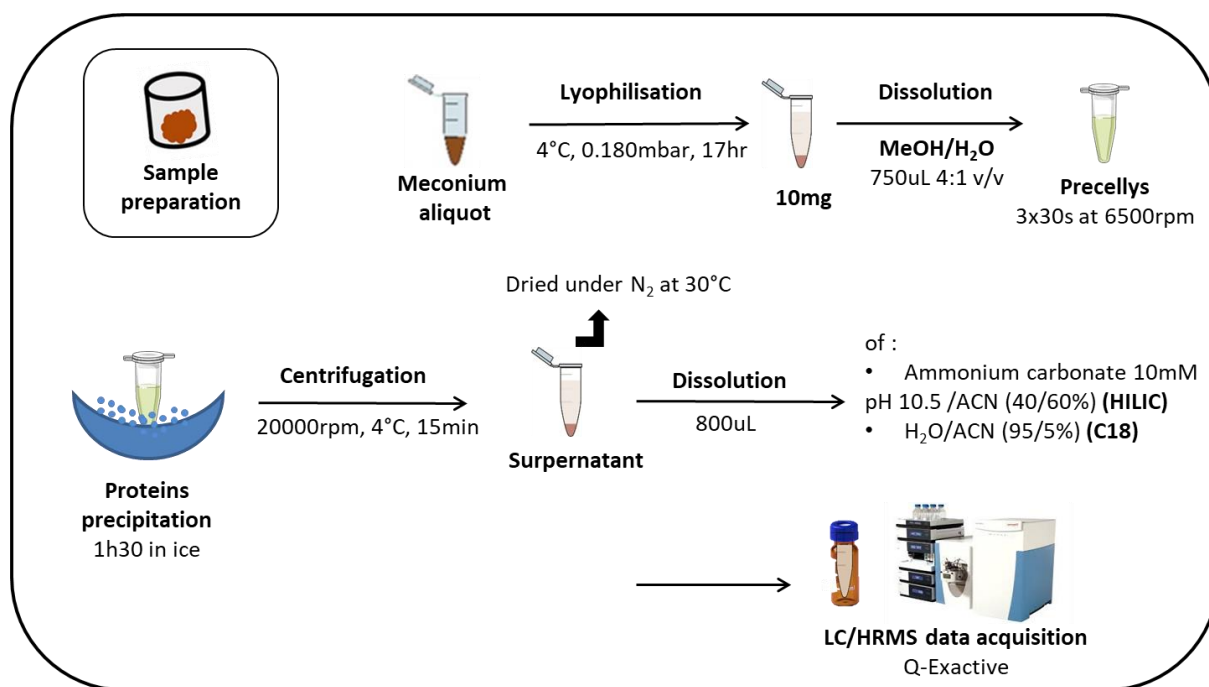


Figure 1. Protocol used for meconium sample preparation.

The pellet was resuspended in a volume of 800 μ L of ammonium carbonate (10 mM pH 10.5)/ACN (40/60%) or H₂O + 0.1 % FA/ACN + 0.1 % FA (95/5%) for ZIC-pHILIC and RP-C18 columns respectively. 100 μ L of each sample was spiked with 5 μ L of external standards (ES) mixture (appendix, table 1). Quality control (QC) samples were prepared by pooling the same volume of all the 33th specimens of the study. Multiple 100 μ L QC vials were aliquoted and spiked with 5 μ L of this same mixture allowing to follow the system stability along the sequence. A conditioning sample for ZIC-pHILIC column calibration was prepared consisting in a QC sample spiked with adenosine triphosphate (ATP).

Analytical sequence

The analytical sequence started by 2 blank runs spiked with external standards followed for ZIC-pHILIC column with 6 conditioning analysis. Triplicate runs of diluted QC to half, quarter and to eighth 3*(QC/2, QC/4, QC/8) followed. The meconium samples were manually randomised and a QC run was introduced each 10 biological samples (Figure 2.3). Finally, the sequence ended with a spiked blank run.

LC/HRMS data acquisition

Experiments were performed on a Dionex 3000 HPLC (refrigerated autosampler, + 4°C, and column heater, 15°C), coupled to a ThermoFisher Scientific Q-Exactive mass spectrometer fitted with an electrospray source operating with both positive and negative ionization modes. Solvents used for the reversed phase Hypersil GOLD C18 column (1.9 µm, 2.1 x 150 mm) consisted of water + 0.1 % formic acid in A, and ACN + 0.1 % of formic acid in B. 10 µL of each sample were injected, and a 500 µL/min flow rate was applied. The chromatographic elution was performed using an isocratic step during 2 minutes at 5 % of B followed by a linear gradient from 5 to 100 % of phase B for 11 min. Before returning to initial conditions, the proportions were kept constants for 12.5 minutes. For the C18 column, a positive ionisation mode was performed in the electrospray source (droplet evaporation temperature, 280 °C; capillary voltage, + 5000 V; sheath and auxiliary gas pressures were respectively 60 and 10 arbitrary units with nitrogen gas). The mass spectrometer resolution power was 50,000 $m/\Delta m$ for full width at half maximum (FWHM) m/z 200 and for singly charged ions, detection goes from m/z 85 to 1000.

For SeQuant ZIC-pHILIC (5 µm, 2.1 x 150 mm) separation, buffers A and B consisted of ammonium carbonate at 10 mM; pH 10.5 and 100 % of ACN as mobile phases respectively. The chromatographic elution was performed with an isocratic step starting with 80 % B until 2 minutes followed by a linear gradient from 80 to 40 % of B until 12 minutes. The flow rate applied was 200 µL/min with a 10 µL sample injection. Negative ionization mode for electrospray was applied (droplet evaporation temperature, 280 °C; capillary voltage, - 3000 V; sheath and auxiliary gas pressures were respectively 60 and 10 arbitrary units with nitrogen gas). The mass spectrometer resolution and the mass dynamic range were the same as for the positive mode.

At the end of each run, both C18 and HILIC columns were thoroughly washed with a strong eluotropic solvent consisting of 100% of B (ACN + 0.1 % of formic acid) and 100 % of A (aqueous ammonium carbonate) respectively, then stabilized under the initial conditions. An annexe pump from Waters Singapore was connected between the chromatographic device and MS instrument to allow solvent flow stability in the MS ion source while washing the columns (Figure 2.4).

Data processing

Raw data were manually checked using the Qual-browser module of xCalibur version 4.1 (Thermo Fisher Scientific, Courtaboeuf, France) then converted to mzXML format using MSconvert (ProteoWizard). Peak picking, alignment and integration were performed using the XCMS software on Workflow4Metabolomics platform. The list of the parameters applied is available as supplementary material. Two peak tables were generated from the data processing of each chromatographic condition.

First a data matrix table containing all the peak areas within the samples, and a variable metadata containing the retention times and exact masses for each feature.

Data were further filtered based on three conditions: i) Ratio of chromatographic peak area of biological to blank samples > 3, ii) Variation coefficient of the quality control samples < 30%, iii) Correlation between the diluted QC samples > 70%. The output files were analysed immediately (non-annotated variables) and were used for metabolite annotation thanks to an in-house chemical database concerning more than 1200 metabolites from different family classes of the human metabolome described in supplementary datas. This local database includes protonated and deprotonated ions, retention times, natural isotopes, adducts and in-source fragment molecules adapted to both of columns and MS instrument used obtained with commercial standard (Figure 2.5).

The following step consisted in the feature annotation, for this, a developed local software was used. The parameters used were: retention time tolerance of +/- 15 and +/- 90 seconds for C18 and HILIC columns respectively, m/z tolerance of 10 ppm was fixed for exact mass consideration and finally an area cutoff of 10 000 and 30 000 for C18 and HILIC columns respectively avoiding to consider background noises. Each annotated peak was manually checked on Qual-browser module of xCalibur by considering the peak shape, the presence of the natural isotopic form and the presence of the considered peak in at least 6 successive scans.

MS/MS experiments

MS/MS analysis was then achieved on the retained signals to confirm their annotation. Three energies of collision for each mass were tested to allow ion fragmentations: 20, 40 and 80. Retention time and exact masses of the parent and its fragments were compared to chemical standards analysed in the same conditions and same instruments which allows obtaining the highest degree of confidence (Figure 2.6).

Statistical analysis

Statistical analysis was conducted on the platform W4M of Galaxy version 3.4.4 including both univariate and multivariate analysis. Multivariate analysis allowing to observe more than one statistical outcome simultaneously focused on heatmap clustering and PCA in this paper. These two chosen methods helped to visualize the data's clustering with an unsupervised manner. PCA analysis was mean-centred and log10 transformed. A univariate analysis was launched suiting a Wilcoxon qualitative test and the p-values evaluated whether a variable was statically significant in the considered clustering. All features with p-values < 0.05 were considered statistically significant. A FDR multiple testing correction

(False Discovery Rate) was applied to figure out the number of false positives on significant varied metabolic features [23].

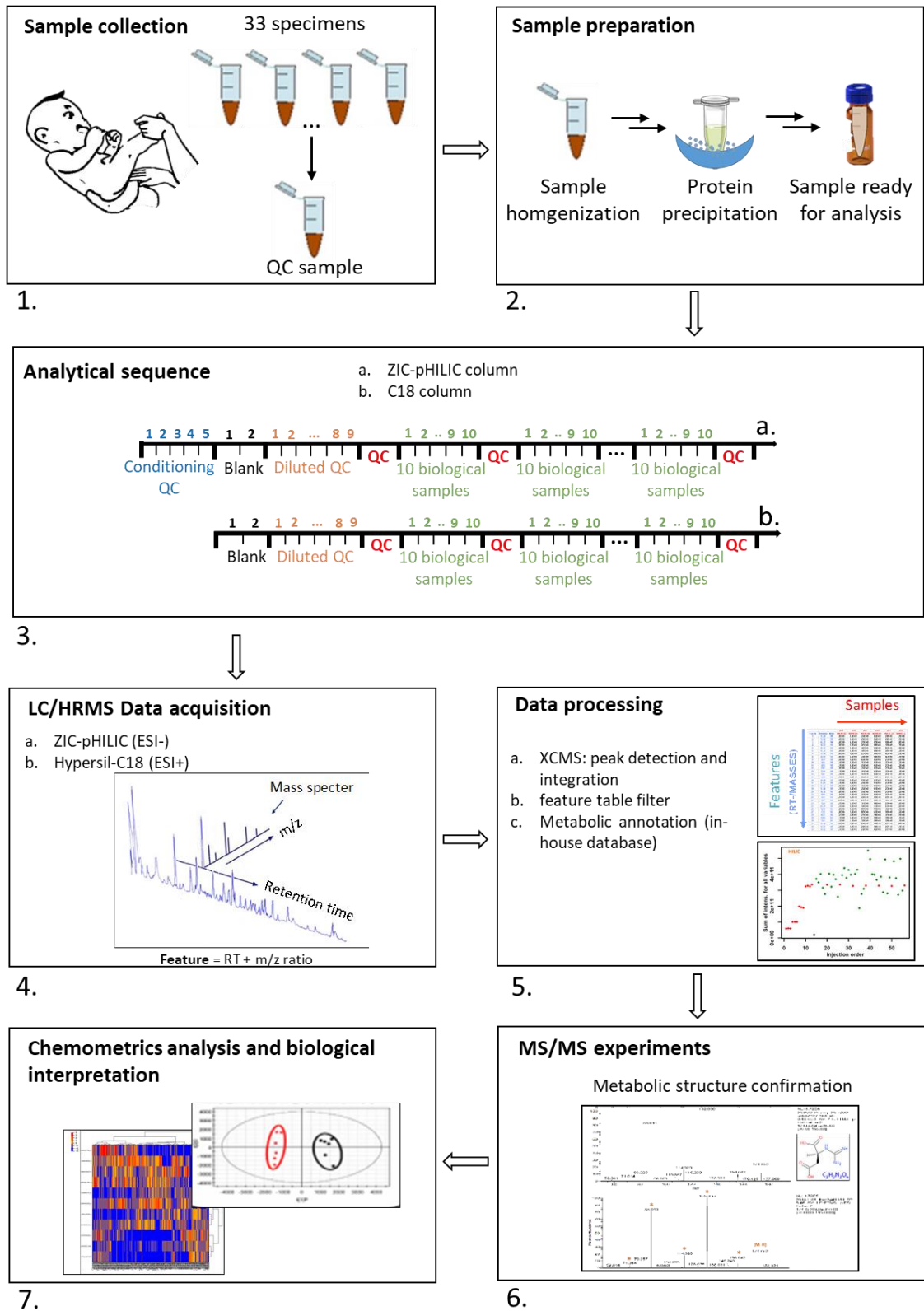


Figure 2. Experimental design used for meconium metabolic profiling.

Results and discussion

Workflow optimization & validation

Preliminary experiments consist in optimizing several steps of the sample preparation and the xcms peak picking (not shown). Cheng et al compared freeze-dried to fresh frozen fecal samples and observed higher number of features extracted in freeze-dried samples [24]. For this reason and to normalize between the samples, the meconium specimens were freeze-dried. Metabolite extraction was first optimized comparing manual sonication probes with automated homogenizer (Precellys apparatus and single-use tubes with ceramic beads) demonstrating a higher efficiency and reproducibility of the latter one. Methanol showed to be an optimal solvent maximizing the metabolome extraction [25] but we also observed that the solvent ratio significantly affected the number and total intensity of detected features, and was then also optimized. Finally, xcms parameters were adjusted to detect the largest amount of repeatable and significant features for all datasets. Validation of the sample preparation and the metabolic extraction steps were accomplished by following the external standard and the QCs evidencing an efficient and reproducible datasets. Both decreased signal correlation for QC dilutions and reproducibility of signal obtained from sequential injected QC demonstrated the system stability during the analytical sequence. Finally, no carry over or marked matrix effect interferences were observed (figure 3). We noticed some extreme biological samples from both early and late collection times represented either with a lower or higher total intensity for instance NB8.270 or NB9.2700 and NB6.2550 that are extreme in the negative and positive modes respectively.

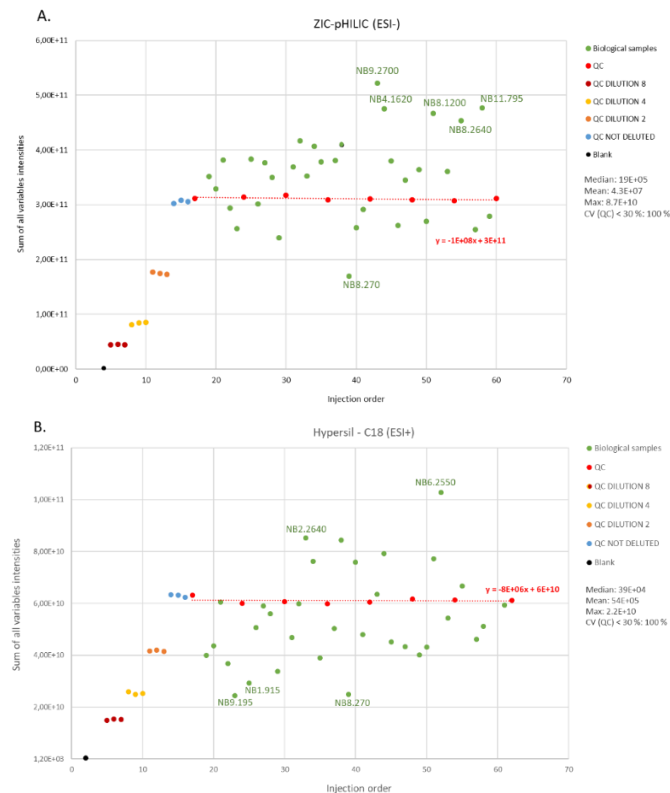


Figure 3. Evolution of the sum of all significant variables intensities (TIC) within the samples following the injection order in the sequence in both columns ZIC-pHILIC (ESI-) (A) and Hypersil-C18 (ESI+) (B).

Global description of meconium samples & day-to-day inter-individual variation

The actual work firstly aims to detect as many metabolites as possible in meconium samples. In total, 9274 and 12397 total features were extracted in the negative and positive ionization modes respectively, using xcms tools of W4M [22]. Of these, 6843 and 8555 features were respectively kept for further evaluation after data filtering based on blank samples, coefficient variation of the QC and correlation between the diluted QC. The filtering step on blank samples allows removing all the features generated from the elution solvent, the diluted QC helps keeping only representative features that suit the dilution, and finally the QC filtering provide the information on repeatable features along the sequence. It's important to note that each metabolite does not represent a unique variable. Actually, several features might represent a multiplicity of molecular species for the same analyte, i.e. monoisotopic form, adducts, dimers and fragments resulting in an overestimation of the real number of metabolites in the sample.

A first exploration of the data was realised by applying a multivariate analysis with a non-supervised manner to investigate the variability inter-samples. Firstly, a principal component analysis (PCA) was performed to all the extracted features in the positive and negative modes (figures 4.A. & 4.B.). These graphs illustrate 45% of the system variation in A. ZIC-pHILIC/ESI (-) and 41% in B. hypersil-C18/ESI (+). We globally notice a non-linear evolution from early to late collection sample times. Some of the extreme specimens observed in figure 3.A. and 3.B. NB8.270 and NB6.2550 appear extreme with these representations. For further investigation, we found substantial to visualize the weights and contributions of each feature in the sample classification. For this end, a multivariate analysis had been launched using clustering heatmap describing the samples and variables classification. Each illustration 4.C and 4.D show two clusters formed with an unsupervised manner. The clusters 1 and 2 in both modes are not equivalents. In fact, samples from the same newborn are more or less close in the classification but lower impact of the collection time is noticed. The extreme samples observed with PCA, NB8.270 and NB6.2550 have several features present at very low intensities which influence their classification. Low intensities are marked by the blue color in the heatmap figure 4.C. and 4.D. These samples are contributing more than others in the ACP representation, for instance in figure 4.B. NB6.2550 has more impact in axis 1 and NB8.270 in axis 2. The sample collection was fully anonymous, excepting the gender that was mentioned but does not explain this distribution. No analytical bias had been detected neither, this representation may explain an inter-individual variation reflecting the foetus exposition during pregnancy. For this aim, a global cartography had been achieved on all significant features to define the meconium composition and outlier metabolites differing from a sample to another.

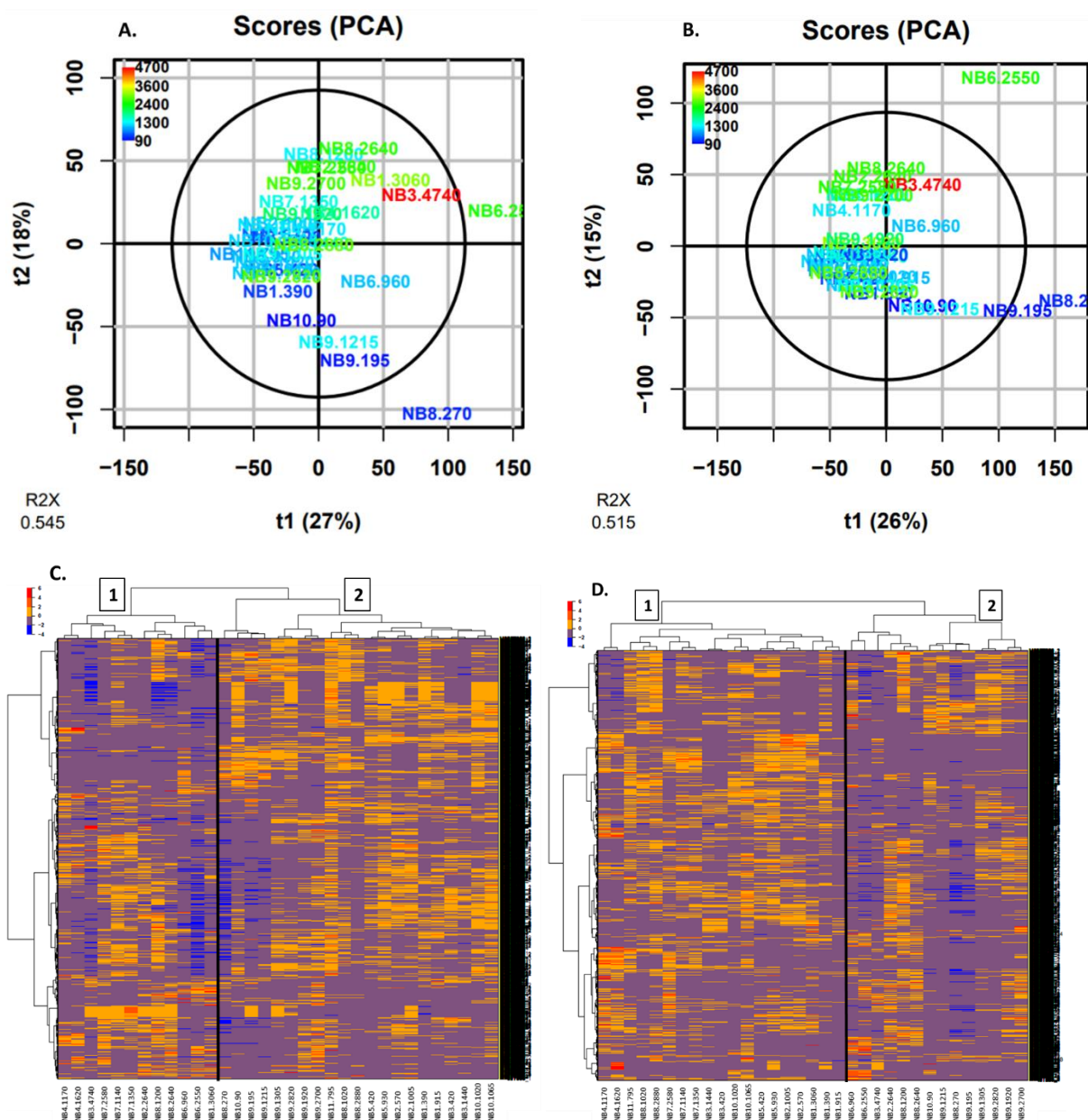


Figure 4. A two-dimensional PCA scores plot (PC1 vs PC2) are represented in A and B for negative and positive modes respectively. The colors used follow the sample collection time starting from dark blue for early times until red for late times. Clustering heatmap are illustrated in C and D for the same modes respectively. Meconium samples clustering (columns) and features are represented in rows. Colors show the intensity of each feature in every sample, going from blue to red through purple, yellow and orange from less to more intense features.

Meconium metabolites annotation

Within the > 17000 chemical features and using an in-house database and MS² analysis, the identification of 198 and 89 metabolites in the negative and positive modes respectively was achieved. 56 common metabolites were detected and annotated in both conditions. In total, 230 metabolites were uniquely identified in meconium samples. The corresponding list of metabolites is provided in the peak table 3 in appendices. Firstly, a putative annotation was achieved using shiny.app software followed by a filtration step on the Qual-browser module of xCalibur. The kept annotations were then

analysed in MS/MS positive and negative ionization modes. 1200 standards spectras analysed in the same conditions and same instruments helped the validation of 210 metabolites in meconium at level 1 according to Sumner et al [26].

Meconium evidences a rich metabolic composition covering diverse molecular species that encompass amino and organic acids, nucleosides, lipids and complex carbohydrates represented in figure 5.A. These metabolites cover mostly endogenous metabolites, but exogenous compounds were also detected (figure 5.B).

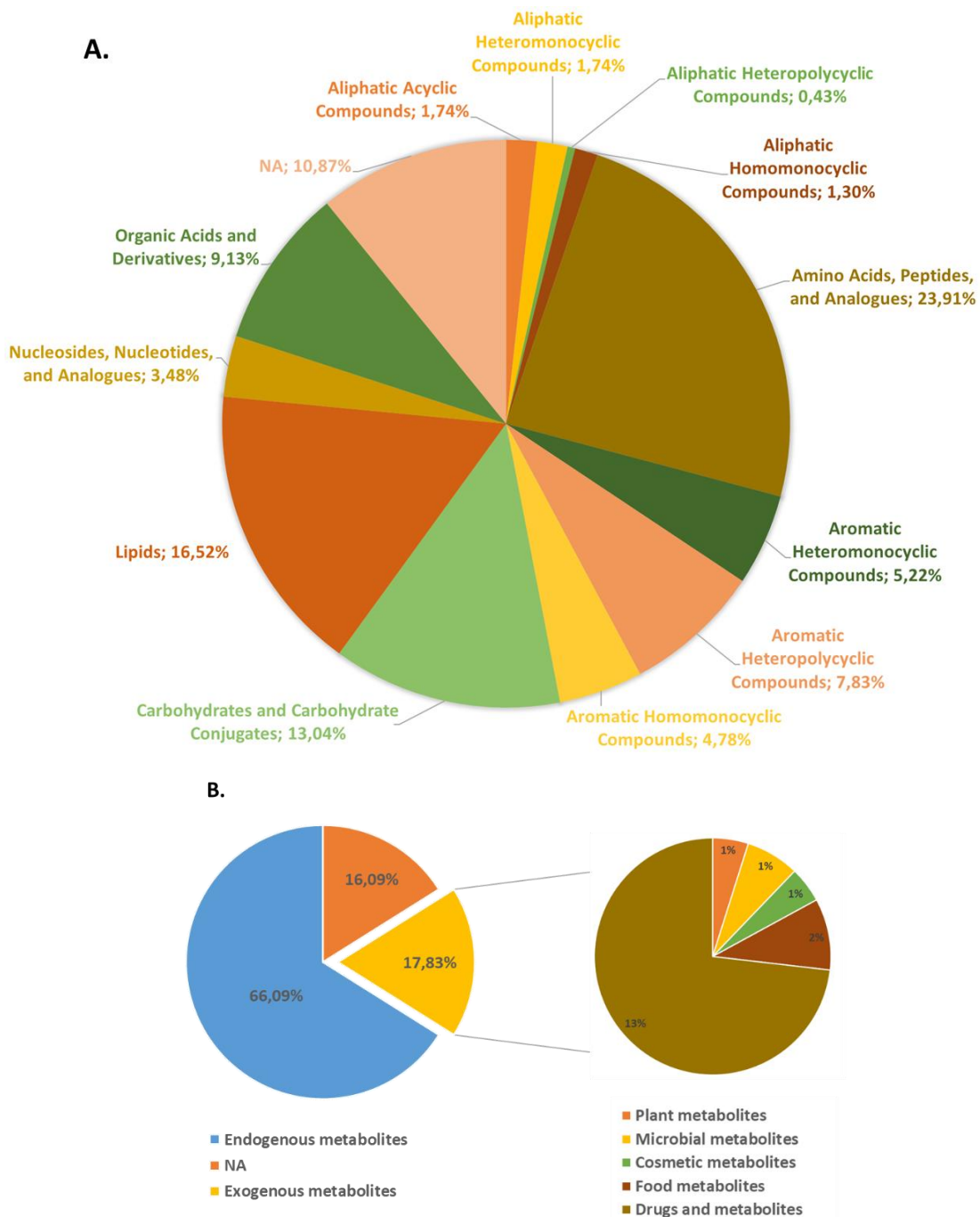


Figure 5: A. Graphic showing the diverse identified families and their percentage in the studied meconium samples. B. Origin of the identified metabolites.

A second multivariate investigation had been achieved within the annotated features in both positive and negative modes. Firstly, a heatmap representation had been launched to check the samples and annotated metabolites clusterisation. Figure 6.A shows two evident clusters within the identified metabolites and within the samples. The clusters are characterized either by a lower or higher intensities comparing cluster 1 to cluster 2 in both cases. The cluster 2/2 (sample's cluster/feature's cluster) shows higher feature intensities presented in orange and red colors. The sample's cluster 2 contains mostly late collected meconiums and evidences an increased concentration of several metabolites. In an other hand, the cluster 2/1 shows a decreasing in metabolites areas from early to late meconium. The represented heatmap may indicate the beginning of a metabolism activity in the new-born gut.

For this investigation and to study the meconium diversity, we decided to divide the samples into two groups according to their collection time (> and < to 1440 minutes) to evaluate the meconium composition evolution within time. Group 1 (gp1) characterized mostly early collection times, and group 2 (gp2) concerns mostly late ones. A multivariate PCA was performed to reveal any clustering in an unsupervised manner. The two mentioned groups are colored in blue and red in figure 6.B. Both of the groups are suiting a non-linear evolution within the collection time and are clustering separately. We also noticed that samples NB8.270 and NB6.2550 still outliers with only the 230 features considered.

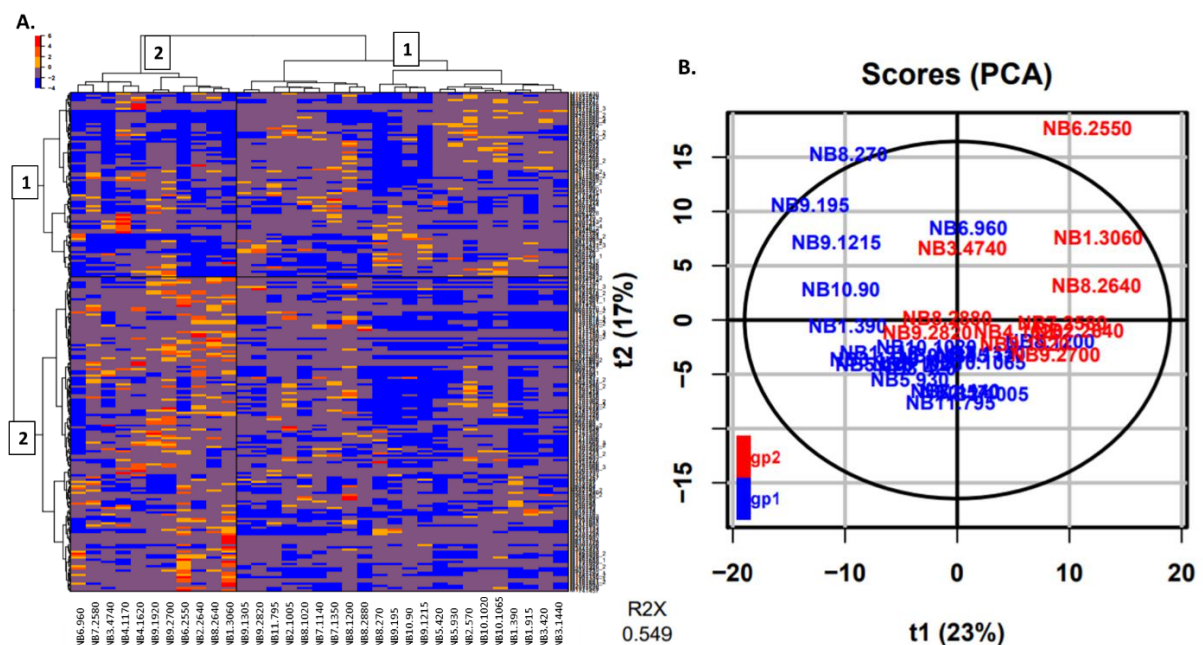


Figure 6: A clustering heatmap is illustrated in A on the 230 annotated features. Clustering of metabolites (rows) and individual meconium samples (columns). The colors represented show the intensity of each feature in every sample, going from blue to red through purple, yellow and orange from less to more intense features. A two-dimensional PCA scores plot (PC1 vs PC2) is represented in B for all samples on the same variables. Group 1 and 2 represent samples < and > to 1440 minutes respectively.

We then focused on knowing the discriminant features allowing the sample clusterisation in figure 6. For this aim, a univariate Wilcoxon-test had been launched. The p-values allow the identification of 67

discriminant features associated with the collection time (appendices table 3). Some of these are observed being more intense in group 1, others are pretty much presents in group 2. The boxplots in figure 7 show few examples of the identified discriminants features between the two groups. In (A) features representing late times for instance N-acetylglycine, L-threonic acid and 1-methylguanine/7-methylguanine that are increasing from group 1 to group 2. Glycerol-3-phosphate, xanthopterin, and (-)-N-acetylneuraminic acid in (B) are decreasing following the collection time and are more representative of early times. These 67 discriminants metabolites are showing the variability inter-meconium samples and the remaining 163 metabolites are more stable and common to all meconium samples. This core metabolome includes both endogenous and exogenous metabolites encompassing uracil, malonic acid, histidine, creatinine... (Appendices table 3).

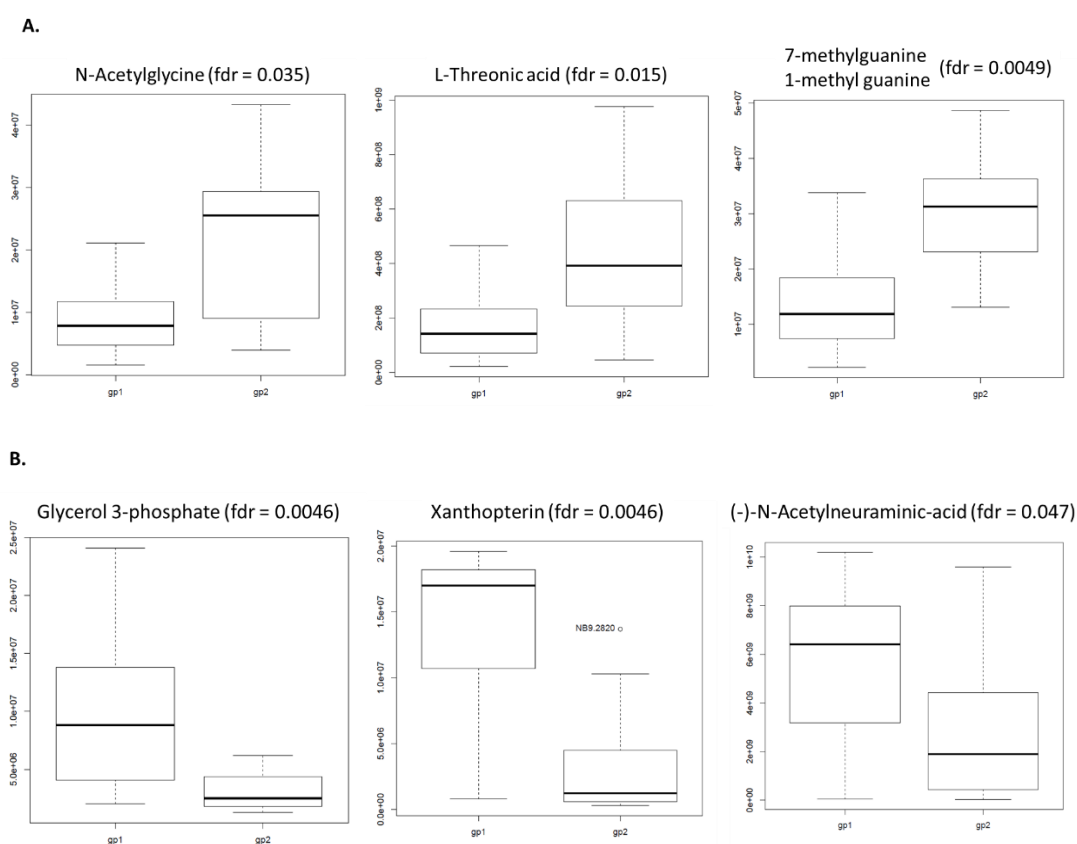


Figure 7: A clustering heatmap is illustrated in A on the 230 annotated features. Colors represented show the intensity of each feature in every sample, going from blue to red through purple, yellow and orange from less to more intense features. A two-dimensional PCA scores plot (PC1 vs PC2) is represented in B for all samples on the same variables. Group 1 and 2 represent samples < and > to 1440 minutes respectively.

Meconium analysis has advantages as it provides the longer historical record of fetal exposure. In this paper, we focused on providing a global meconium analysis focusing especially on endogenous metabolites and trying to explore the potential of meconium composition in metabolic deposition during gestation. Meconium had already been presented in the literature to indicate fetal exposure to

a selection of xenobiotics agents comprising alcohol and nicotine metabolites, food additives and drugs [27]–[29]... However, there is a lack of published papers regarding the meconium endogenous global composition. Until today and for the authors' knowledge, this is the first global cartography of meconium described.

Nevertheless, some published studies worked on meconium with an untargeted approach or also on endogenous metabolites. Peng et al worked on elucidating gestational diabetes mellitus by analysing the metabolome response on new-born meconium of GDM mothers against healthy cases. They potentially identified 14 GDM biomarkers belonging to various metabolic patterns [25]. 4 of these biomarkers had been identified in this study concerning: uridine, glycocholic acid, dimethyluric acid and methylxanthine issued from both endogenous and drug metabolisms. Righetti et al have studied by NMR the water soluble and organic fractions of meconium in day 1 and 3 after birth, allowing the identification and quantification of several metabolites from sterols and fatty acids (Righetti et al). 16 of these metabolites have been quantified and showed globally an increasing in concentration from day 1 to day 3 concurring with our univariate results. In fact, 2 of these metabolites had been identified and showed significant evolution from the Wilcoxon test with an increased value from group 1 to group 2 that are taurine and phenylalanine (appendices table 3). Other shared metabolites were also identified in meconium like aminobutyric acid, valine, tyrosine and alanine.

Conclusions and outlook

This study firstly aims to explore the differences in metabolic profiles of meconium following the collection times within 33 meconium samples. Secondly, we aimed to investigate the global composition of meconium allowing the detection of more than 17000 chemical features, comprising redundancy, from which 230 unique metabolites were identified with a high level of confidence. This work concludes in a noticeable evolution of the meconium composition from day 1 (< 1140 minutes) until day 3. This evolution might result from a metabolic activity starting in the neonate gut. This work suggest that the “*in utero* exposome” should focus on the meconium samples excreted within the first day to avoid any postnatal contamination.

This manuscript describes the highest coverage of the meconium metabolome known to date, using sensitive and robust methods. Nevertheless, the identified metabolome represents less than 5% of the extracted signals. Highlighting the use of online databases could be complementary to explore highest metabolome recovery of the meconium.

Acknowledgments

The French Alternative Energies and Atomic Energy Commission, and the National Institute of Agricultural, Food and Environment Research support this study. A special thanks to the Sainte Therese

Clinique for providing the meconium samples, the parents who accepted to participate in this study and all the nurses who offered their kind help with meconium collection.

References

- [1] M. Meyer-Monath, C. Chatellier, F. Rouget, I. Morel, and F. Lestremau, 'Development of a multi-residue method in a fetal matrix: analysis of meconium', *Anal Bioanal Chem*, vol. 406, no. 30, pp. 7785–7797, Dec. 2014, doi: 10.1007/s00216-014-8243-4.
- [2] Madej et al., 'Analysis of meconium, nails and tears for determination of medicines and drugs of abuse', *TrAC Trends in Analytical Chemistry*, vol. 29, no. 3, pp. 246–259, Mar. 2010, doi: 10.1016/j.trac.2010.01.005.
- [3] M. K. Woźniak, E. Jaszczak, M. Wiergowski, Ż. Polkowska, J. Namieśnik, and M. Biziuk, 'Meconium analysis as a promising diagnostic tool for monitoring fetal exposure to toxic substances: Recent trends and perspectives', *TrAC Trends in Analytical Chemistry*, vol. 109, pp. 124–141, Dec. 2018, doi: 10.1016/j.trac.2018.09.025.
- [4] E. M. Ostrea, 'Testing for exposure to illicit drugs and other agents in the neonate: A review of laboratory methods and the role of meconium analysis', *Current Problems in Pediatrics*, vol. 29, no. 2, pp. 41–56, Feb. 1999, doi: 10.1016/S0045-9380(99)80046-2.
- [5] M. Oldiges, S. Lütz, S. Pflug, K. Schroer, N. Stein, and C. Wiendahl, 'Metabolomics: current state and evolving methodologies and tools', *Appl Microbiol Biotechnol*, vol. 76, no. 3, pp. 495–511, Sep. 2007, doi: 10.1007/s00253-007-1029-2.
- [6] M. Suzuki, S. Nishiumi, A. Matsubara, T. Azuma, and M. Yoshida, 'Metabolome analysis for discovering biomarkers of gastroenterological cancer', *Journal of Chromatography B*, vol. 966, pp. 59–69, Sep. 2014, doi: 10.1016/j.jchromb.2014.02.042.
- [7] K. Adel-Patient *et al.*, 'Deep analysis of immune response and metabolic signature in children with food protein induced enterocolitis to cow's milk', *Clinical and Translational Allergy*, vol. 8, no. 1, p. 38, Sep. 2018, doi: 10.1186/s13601-018-0224-9.
- [8] Peng et al. *et al.*, 'Newborn Meconium and Urinary Metabolome Response to Maternal Gestational Diabetes Mellitus: A Preliminary Case–Control Study', *J. Proteome Res.*, vol. 14, no. 4, pp. 1799–1809, Apr. 2015, doi: 10.1021/pr5011857.
- [9] E. M. Ostrea, D. K. Knapp, A. Romero, M. Montes, and A. R. Ostrea, 'Meconium analysis to assess fetal exposure to nicotine by active and passive maternal smoking', *The Journal of Pediatrics*, vol. 124, no. 3, pp. 471–476, Mar. 1994, doi: 10.1016/S0022-3476(94)70378-7.
- [10] J. Ristimaa, M. Gergov, A. Pelander, E. Halmesmäki, and I. Ojanperä, 'Broad-spectrum drug screening of meconium by liquid chromatography with tandem mass spectrometry and time-of-flight mass spectrometry', *Anal Bioanal Chem*, vol. 398, no. 2, pp. 925–935, Sep. 2010, doi: 10.1007/s00216-010-3942-y.
- [11] M. Meyer-Monath, C. Chatellier, D. Cabooter, F. Rouget, I. Morel, and F. Lestremau, 'Development of liquid chromatography methods coupled to mass spectrometry for the analysis of substances with a wide variety of polarity in meconium', *Talanta*, vol. 138, pp. 231–239, Jun. 2015, doi: 10.1016/j.talanta.2015.02.058.
- [12] M. Meyer-Monath, J. Beaumont, I. Morel, F. Rouget, K. Tack, and F. Lestremau, 'Analysis of BTEX and chlorinated solvents in meconium by headspace-solid-phase microextraction gas chromatography coupled with mass spectrometry', *Anal Bioanal Chem*, vol. 406, no. 18, pp. 4481–4490, Jul. 2014, doi: 10.1007/s00216-014-7836-2.
- [13] A. Bakdash *et al.*, 'Quantification of fatty acid ethyl esters (FAEE) and ethyl glucuronide (EtG) in meconium from newborns for detection of alcohol abuse in a maternal health evaluation study', *Anal Bioanal Chem*, vol. 396, no. 7, pp. 2469–2477, Apr. 2010, doi: 10.1007/s00216-010-3474-5.
- [14] P. Cabarcos *et al.*, 'Quantification of fatty acid ethyl esters (FAEE) and ethyl glucuronide (EtG) in meconium for detection of alcohol abuse during pregnancy: Correlation study between both

- biomarkers', *Journal of Pharmaceutical and Biomedical Analysis*, vol. 100, pp. 74–78, Nov. 2014, doi: 10.1016/j.jpba.2014.07.023.
- [15] C. F. Bearer *et al.*, 'Ethyl Linoleate in Meconium: A Biomarker for Prenatal Ethanol Exposure', *Alcoholism: Clinical and Experimental Research*, vol. 23, no. 3, pp. 487–493, Mar. 1999, doi: 10.1111/j.1530-0277.1999.tb04142.x.
- [16] L. N. Bakhireva *et al.*, 'Prenatal alcohol exposure prevalence as measured by direct ethanol metabolites in meconium in a Native American tribe of the southwest', *Birth Defects Research*, p. bdr2.1427, Dec. 2018, doi: 10.1002/bdr2.1427.
- [17] S. Wandro, S. Osborne, C. Enriquez, C. Bixby, A. Arrieta, and K. Whiteson, 'The Microbiome and Metabolome of Preterm Infant Stool Are Personalized and Not Driven by Health Outcomes, Including Necrotizing Enterocolitis and Late-Onset Sepsis', *mSphere*, vol. 3, no. 3, pp. e00104-18, /msphere/3/3/mSphere104-18.atom, Jun. 2018, doi: 10.1128/mSphere.00104-18.
- [18] C. Junot, F. Fenaille, B. Colsch, and F. Bécher, 'High resolution mass spectrometry based techniques at the crossroads of metabolic pathways', *Mass Spectrometry Reviews*, vol. 33, no. 6, pp. 471–500, 2014, doi: 10.1002/mas.21401.
- [19] E. M. Ostrea *et al.*, 'Combined analysis of prenatal (maternal hair and blood) and neonatal (infant hair, cord blood and meconium) matrices to detect fetal exposure to environmental pesticides', *Environmental Research*, vol. 109, no. 1, pp. 116–122, Jan. 2009, doi: 10.1016/j.envres.2008.09.004.
- [20] S. Boudah *et al.*, 'Annotation of the human serum metabolome by coupling three liquid chromatography methods to high-resolution mass spectrometry', *Journal of Chromatography B*, vol. 966, pp. 34–47, Sep. 2014, doi: 10.1016/j.jchromb.2014.04.025.
- [21] S. Aros-Calt *et al.*, 'Annotation of the *Staphylococcus aureus* Metabolome Using Liquid Chromatography Coupled to High-Resolution Mass Spectrometry and Application to the Study of Methicillin Resistance', *J. Proteome Res.*, vol. 14, no. 11, pp. 4863–4875, Nov. 2015, doi: 10.1021/acs.jproteome.5b00697.
- [22] Y. Guitton *et al.*, 'Create, run, share, publish, and reference your LC–MS, FIA–MS, GC–MS, and NMR data analysis workflows with the Workflow4Metabolomics 3.0 Galaxy online infrastructure for metabolomics', *The International Journal of Biochemistry & Cell Biology*, vol. 93, pp. 89–101, Dec. 2017, doi: 10.1016/j.biocel.2017.07.002.
- [23] M. Vinaixa, S. Samino, I. Saez, J. Duran, J. J. Guinovart, and O. Yanes, 'A Guideline to Univariate Statistical Analysis for LC/MS-Based Untargeted Metabolomics-Derived Data', *Metabolites*, vol. 2, no. 4, pp. 775–795, Oct. 2012, doi: 10.3390/metabo2040775.
- [24] K. Cheng, C. Brunius, R. Fristedt, and R. Landberg, 'An LC-QToF MS based method for untargeted metabolomics of human fecal samples', *Metabolomics*, vol. 16, no. 4, p. 46, 03 2020, doi: 10.1007/s11306-020-01669-z.
- [25] S. Peng *et al.*, 'Newborn Meconium and Urinary Metabolome Response to Maternal Gestational Diabetes Mellitus: A Preliminary Case–Control Study', *J. Proteome Res.*, vol. 14, no. 4, pp. 1799–1809, Apr. 2015, doi: 10.1021/pr5011857.
- [26] L. W. Sumner *et al.*, 'Proposed minimum reporting standards for chemical analysis', *Metabolomics*, vol. 3, no. 3, pp. 211–221, Sep. 2007, doi: 10.1007/s11306-007-0082-2.
- [27] R. Cassoulet *et al.*, 'Monitoring of prenatal exposure to organic and inorganic contaminants using meconium from an Eastern Canada cohort', *Environmental Research*, vol. 171, pp. 44–51, Apr. 2019, doi: 10.1016/j.envres.2018.12.044.
- [28] J. Baranowski, G. Pochopień, and I. Baranowska, 'Determination of nicotine, cotinine and caffeine in meconium using high-performance liquid chromatography', *Journal of Chromatography B: Biomedical Sciences and Applications*, vol. 707, no. 1, pp. 317–321, Apr. 1998, doi: 10.1016/S0378-4347(97)00619-1.
- [29] D. Bielawski, E. Ostrea, N. Posecion, M. Corrión, and J. Seagraves, 'Detection of Several Classes of Pesticides and Metabolites in Meconium by Gas Chromatography–Mass Spectrometry', *Chroma*, vol. 62, no. 11, pp. 623–629, Dec. 2005, doi: 10.1365/s10337-005-0668-7.

Annexe 2 : Publication « Basal Ti level in the human placenta and meconium and evidence of a maternofoetal transfer of food-grade TiO₂ nanoparticles in an ex vivo placental perfusion model ».

RESEARCH

Open Access



Basal Ti level in the human placenta and meconium and evidence of a materno-foetal transfer of food-grade TiO₂ nanoparticles in an ex vivo placental perfusion model

A. Guillard¹, E. Gaultier¹, C. Cartier¹, L. Devoille², J. Noireaux³, L. Chevalier⁴, M. Morin⁵, F. Grandin¹, M. Z. Lacroix⁶, C. Coméra¹, A. Cazanave¹, A. de Place⁵, V. Gayraud¹, V. Bach⁷, K. Chardon⁷, N. Bekhti⁸, K. Adel-Patient⁸, C. Vayssière^{5,9}, P. Fiscaro³, N. Feltrin², F. de la Farge¹, N. Picard-Hagen¹, B. Lamas¹ and E. Houdeau^{1*} 

Abstract

Background: Titanium dioxide (TiO₂) is broadly used in common consumer goods, including as a food additive (E171 in Europe) for colouring and opacifying properties. The E171 additive contains TiO₂ nanoparticles (NPs), part of them being absorbed in the intestine and accumulated in several systemic organs. Exposure to TiO₂-NPs in rodents during pregnancy resulted in alteration of placental functions and a materno-foetal transfer of NPs, both with toxic effects on the foetus. However, no human data are available for pregnant women exposed to food-grade TiO₂-NPs and their potential transfer to the foetus. In this study, human placentae collected at term from normal pregnancies and meconium (the first stool of newborns) from unpaired mothers/children were analysed using inductively coupled plasma mass spectrometry (ICP-MS) and scanning transmission electron microscopy (STEM) coupled to energy-dispersive X-ray (EDX) spectroscopy for their titanium (Ti) contents and for analysis of TiO₂ particle deposition, respectively. Using an ex vivo placenta perfusion model, we also assessed the transplacental passage of food-grade TiO₂ particles.

Results: By ICP-MS analysis, we evidenced the presence of Ti in all placentae (basal level ranging from 0.01 to 0.48 mg/kg of tissue) and in 50% of the meconium samples (0.02–1.50 mg/kg), suggesting a materno-foetal passage of Ti. STEM-EDX observation of the placental tissues confirmed the presence of TiO₂-NPs in addition to iron (Fe), tin (Sn), aluminium (Al) and silicon (Si) as mixed or isolated particle deposits. TiO₂ particles, as well as Si, Al, Fe and zinc (Zn) particles were also recovered in the meconium. In placenta perfusion experiments, confocal imaging and SEM-EDX analysis of foetal exudate confirmed a low transfer of food-grade TiO₂ particles to the foetal side, which was barely quantifiable by ICP-MS. Diameter measurements showed that 70 to 100% of the TiO₂ particles recovered in the foetal exudate were nanosized.

(Continued on next page)

* Correspondence: eric.houdeau@inrae.fr

¹Toxalim UMR1331 (Research Centre in Food Toxicology), Toulouse University, INRAE, ENVT, INP-Purpan, UPS, Toulouse, France
Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

Conclusions: Altogether, these results show a materno-foetal transfer of TiO₂ particles during pregnancy, with food-grade TiO₂ as a potential source for foetal exposure to NPs. These data emphasize the need for risk assessment of chronic exposure to TiO₂-NPs during pregnancy.

Keywords: Titanium dioxide, Nanoparticles, Human placenta, E171 food additive, Foetus

Introduction

Titanium dioxide (TiO₂) is one of the most commonly produced nanomaterials used for various industrial applications (cosmetics, water and soil treatment, UV filter, medicine, food sector) based on its colouring and opacifying properties as well as its photocatalytic and biocidal activities [1, 2]. Due to these widespread applications, including in food, human exposure to TiO₂ nanoparticles (NPs, at least one dimension < 100 nm) occurs by inhalation, dermal exposure and the oral route. Translocation across biological barriers (lung, skin, intestine) depends on the crystal form (anatase or rutile), coating, size, surface area and aggregate/agglomerate formation [3]. A systemic passage for TiO₂ is documented in human [4–6] and resembles that reported in rodents [7, 8], with accumulation in the liver and spleen of nano- and submicronic particles [9], indicating a low but chronic distribution of TiO₂ particulate matter in the human organism. In vivo studies in rodents exposed to TiO₂ reported toxic effects such as inflammation, impairment of biological barrier functions (intestinal, placental, blood-testis), as well as the promotion of cancer development [7, 10, 11]. Overall, repeated exposure to these particles exhibiting cytotoxic, genotoxic and immunotoxic effects is considered to be an important health issue [12–14]. For the general population, there is accumulating evidence that the main uptake route is ingestion, since TiO₂ is produced at high volumes as a food grade pigment (E171 in EU) incorporated in many foodstuffs as well as in toothpaste, medicines and food supplements. Dietary exposure to TiO₂ ranges from 0.4 to 10.4 mg/kg BW/day [15], depending on the age group and food habits. E171 is a powder consisting of anatase and/or rutile TiO₂, exhibiting up to 55% NPs by number depending on the commercial supplier [7, 16]. Oral kinetic studies in human volunteers showed that a fraction of food-grade TiO₂ is absorbed and reaches the bloodstream a few hours after ingestion [5]. However, in this context, maternal exposure during pregnancy has not yet been evaluated for risk assessment in humans. Chronic oral intake of TiO₂-NPs in pregnant women could lead to placental dysfunction and/or transplacental passage, both of which may have deleterious impacts on foetal development [11, 17, 18].

Due to the separation of maternal and foetal circulations, the placenta acts as a barrier to environmental

toxicants [19, 20] and potential pathogens [21] that could be hazardous for the growing foetus. This barrier was nevertheless shown to be poorly effective when exposed to TiO₂-NPs, at least in mice and rats, with evidence of particle accumulation in the placenta and passage to the foetus [22]. Toxicity data indicate resorption of embryos and foetal growth restriction in mice exposed intravenously to TiO₂-NPs for two days at mid-pregnancy [17]. Moreover, low doses of TiO₂-NPs orally administered to dams during the first two gestational weeks were shown to impair placentation and induced dysregulation of vascularization and proliferation, possibly leading to alteration of nutrient and gas exchanges with the foetus [11]. Other studies in rodents exposed through oral route [23–25] or intravenous injection [17, 26] throughout gestation showed a materno-foetal transfer of TiO₂-NPs and their accumulation in several foetal organs, such as the liver, testis and brain, the latter distribution leading to an impairment of cognitive functions in the offspring. However, all these hazards cannot be directly transposed to humans for risk assessment due to interspecies differences in placentation and structure [27, 28]. Recently, one study reported a significant titanium (Ti) content in the term placenta and in the cord blood in paired mothers-infants, suggesting that a placental transfer of TiO₂ may also exist in humans [29]. Nevertheless, specific data concerning the human placental and meconium content of TiO₂ particles (and particularly NPs) and their transfer are currently lacking. The meconium could be the best matrix to analyse in utero exposure to TiO₂, as it is the first neonatal stool accumulated during the last 6 months of pregnancy in the foetal intestine. Analysis of meconium, which is naturally excreted within the first two days of life and usually discarded, is non-invasive and considered highly accurate to detect chronic foetal exposure to xenobiotics [30].

In humans, studies on the transplacental passage of inorganic particles have been conducted in vitro on trophoblastic cells and ex vivo using isolated perfused placenta. The size was reported an important contributing factor for particle transfer. Ex vivo, only polystyrene beads up to 240 nm crossed the syncytiotrophoblast that separates the foetal circulation from the maternal blood [31], with low transfer rates and bidirectional passage (i.e., materno-foetal and inversely) [32]. In vitro, 25 and 50 nm silicon (Si) NPs were reported to cross the placental barrier, an observation confirmed ex vivo using

perfused human placenta [33]. These findings suggest the capacity for the nanosized fraction (< 100 nm) of food-grade TiO₂ to reach the foetus. To date, in the absence of specific information on food-grade formulations, two studies have examined the capacity of TiO₂-NP models perfused with different surface charges to cross human placenta. These studies failed to observe a quantifiable increase in Ti levels in the foetal compartment using inductively coupled plasma mass spectrometry (ICP-MS) analysis after 6 h of perfusion, but authors do not exclude the possibility that a low amount of NPs may cross the placental barrier in humans [34, 35], which remains highly challenging to detect. Even in the case of low placental transfer rates, foetal accumulation of insoluble TiO₂-NPs may occur if the foetus is chronically exposed via its mother to TiO₂ of various environmental origins, including dietary sources.

In this study, we aimed to assess TiO₂ exposure in the human foetus by performing ICP-MS analysis of Ti contents in the human placenta collected at term from pregnancies and in the meconium. These matrices were also analysed with scanning transmission electron microscopy coupled to energy dispersive X-ray (STEM-EDX) analysis to ensure the presence of particles and their chemical nature. Second, by using the ex vivo human placenta perfusion model, we assessed whether TiO₂-NPs of dietary origin (E171 additive) may cross the placental barrier by combining ICP-MS, confocal microscopy, scanning electron microscopy (SEM) and STEM-EDX for compositional analysis and elemental mapping of the perfused samples in the foetal side and particle distribution into perfused placental tissues.

Results

Physico-chemical characteristics of food-grade TiO₂ particles

SEM-EDX analyses showed that E171 had TiO₂ particles with a mean particle size of 104.9 ± 44.9 nm and a particle size distribution ranging from 20 to 440 nm, with 55% of NPs by number (Table 1 and Fig. 1). The specific surface area (SSA) was shown to be 9.6 m²/g by BET, and the zeta potential in water suspension was -35.1 ± 3.22 mV (Table 1). DLS analysis showed increased hydrodynamic diameters of the TiO₂ particles in Earle's medium (Table 1), indicating particle agglomeration in the perfusion medium (PM) compared to ultrapure water.

Table 1 Physico-chemical characteristics of E171

Mean particle size (nm)	% of NPs	SSA (m ² /g)	Ultrapure water (pH = 8.92)			Perfusion medium (pH = 7.40)		
			Zeta potential (mV)	H. diam. (nm)	Pdl	Zeta potential (mV)	H. diam. (nm)	Pdl
104.9 ± 44.9	55	9.6	-35.1 ± 3.22	303.27 ± 134.27	0.223	-12.3 ± 0.00	383.3 ± 163.6	0.363

All data are presented as mean ± SD. SSA specific surface area, H. diam. hydrodynamic diameter, Pdl polydispersity index

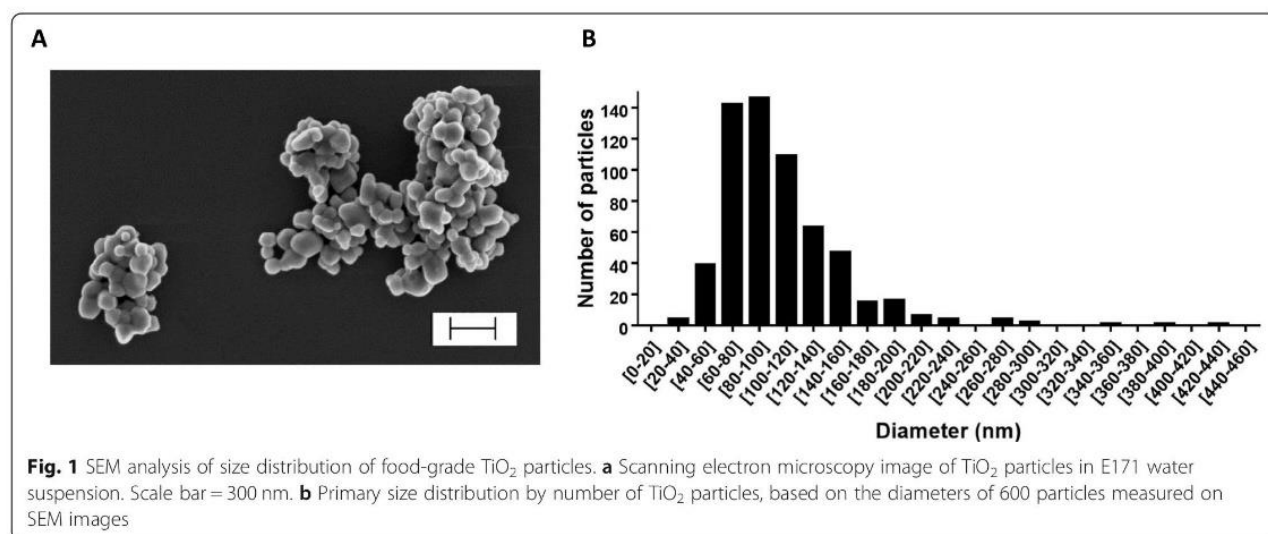
ICP-MS measurement of total Ti content in the human term placenta and the meconium, and particle analysis by STEM-EDX

As shown in Table 2, Ti was found in all placental samples (*n* = 22), with the total Ti content ranging from 0.01 to 0.48 mg/kg of tissue. Seven women displayed a total placental Ti content above 0.1 mg/kg of tissue, with 2 of them reaching 0.4–0.5 mg/kg of tissue. On TEM tissue sections, particulate matter was observed as isolated primary particles and/or as small aggregates in placental tissues. Chemical elemental mapping using STEM-EDX confirmed the presence of Ti and oxygen (O) onto particle deposits (Fig. 2a), often appearing as elongated crystal forms associated with aluminium (Al) and Si trace elements, and most of the analysed TiO₂ particles were below 100 nm in diameter (Fig. 2a). Among the 34 particles analysed by EDX, Si-, tin (Sn)-, Al-, iron (Fe)- and zinc (Zn)-containing particles were frequently found in addition to TiO₂ (Fig. S1). Size measurements of all particles (i.e., whatever their chemical nature) recovered during the TEM analysis showed that 32 of 34 particles exhibited diameters ranging from 10 to 225 nm, 16 of them (i.e., 50%) being below 100 nm (Fig. S2), and two large agglomerates of 320 and 380 nm (Fig. S2).

In the meconium (Table 2), Ti was detected by ICP-MS in 50% of samples (9 of 18), ranging from 0.02 to 1.50 mg/kg of material. In most of these samples (*n* = 8), the Ti levels were above 0.1 mg/kg of meconium (Table 2). TEM-EDX analysis of two representative meconium samples confirmed the presence of Ti and O elements in the particulate deposits (Fig. 2b). In addition, clustered elements of mainly Si, Al, Fe and Zn were observed as particulate matter (Fig. S3). Overall, size analysis of all particles indicated a diameter ranging from 5 to 194 nm (*n* = 33 particles), and 26 of them (i.e., 82%) were in the nanorange.

Transfer profile of food-grade TiO₂ particles across the human placenta ex vivo

A total of 7 placentae were validated based on antipyrine passage during ex vivo perfusion in an open circulating system with PM alone and E171 suspension. There was no difference in permeability (i.e., antipyrine transfer rate) between placentae collected after caesarean section or vaginal delivery (Fig. S4). After the perfusion of placentae with food-grade TiO₂ (E171), Ti was detected by ICP-MS in most foetal exudates (0.41 to 3.46 ng Ti/mL), but 92% were in the range of the blank level (0.33 to



1.92 ng Ti/mL) (Table S1). To assess whether a transfer occurred for TiO₂ particles, we examined the exudate samples collected every 5 min on the foetal side by confocal microscopy to detect laser-diffracting (TiO₂-like) particles, as shown in Fig. S5. An average blank value of 1.05 ± 0.32 laser-diffracting spots per microscopic field was calculated from 4 independent experiments with PM free of E171, and then subtracted from each time point in each placenta perfused with the E171 suspension. As shown in Fig. 3, the placental translocation of laser-diffracting TiO₂ particles increased from 10 min of E171 perfusion, reached a plateau at 20–30 min, and then decreased until the end of the perfusion.

SEM-EDX analysis of perfused particles recovered in the foetal exudate

Based on the above reported particle transfer profile with confocal imaging, pooled liquid fractions in the foetal side corresponding to the 20–30 min period of perfusion with the E171 suspension (see Fig. 3) were subjected to SEM-EDX analysis. After sample purification, SEM-EDX scanning showed the presence of many TiO₂ particles appearing as isolated particles or aggregates trapped in the matrix of the dried PM (Fig. 4). Furthermore, Al and Fe elements were also found in some particulate deposits, associated or not with Ti (Fig. S6). The current detection method by SEM-EDX did not provide quantitative information on the total number of TiO₂ particles recovered in the foetal circuit over the 20–30 min of E171 perfusion but allowed us to evaluate the size distribution of the particles crossing the placenta during this period. As a result, a total of 300 TiO₂ particles sampled and purified from the foetal side were measured by taking the smallest dimension observed for each TiO₂-positive particle in the recovered agglomerates (Fig. S7), and the results compared to the TiO₂

particle size distribution in the PM with E171 added from the maternal reservoir (Fig. 5). Except for 3 particles, TiO₂ particles recovered in the foetal side exhibited a diameter < 200 nm, with 70 and 100% of NPs in two perfusion experiments (Fig. 5).

STEM-EDX analyses of TiO₂ particles in the perfused placental tissues

In placental tissue, (S)TEM-EDX was used to investigate the distribution of TiO₂ particles into the perfused cotyledon. As illustrated in Fig. 6, enriched tissue areas with Ti + O showed isolated round shaped particles or small aggregates of TiO₂, typical of the food-grade batch used in this study. TiO₂ particles from the E171 suspension were recovered in the syncytiotrophoblast microvilli (Fig. 6a and S8D) and had translocated in deeper areas of the placental chorionic mesenchyme surrounding foetal vessels (Fig. 6b and S8D). As shown in Fig. S8, size measurement of the 33 TiO₂ (i.e., EDX-characterized) particles translocated into placental tissues showed 26 particles with a diameter below 250 nm, and 17 of them in the nanorange.

Discussion

The effects of prenatal exposure to xenobiotics (potentially toxic) on birth outcomes and child development are an area of concern for public health [36], which have been extended to NPs of anthropogenic origin, including TiO₂-NPs. For pregnant women, a recent opinion by the French High Council for Public Health [37] indicated that the risk of a transfer of TiO₂-NPs to the foetus and the health consequences for the newborn have not been documented despite animal studies showing harmful effects [22], with TiO₂-NPs recovered in foetal organs such as the brain, liver and testis [17, 23, 26]. A recent study in humans analysed the Ti content in maternal

Table 2 ICP-MS analysis of basal Ti content in human term placenta and meconium (unpaired samples)

	Total Ti (mg/kg)	
	Placenta (n = 22)	Meconium (n = 18)
1	0.04	1 0.19
2	0.02	2 0.09
3	0.01	3 < LOQ
4	0.05	4 0.17
5	0.20	5 < LOQ
6	0.15	6 < LOQ
7	0.03	7 0.34*
8	0.17	8 1.41*
9	0.11	9 < LOQ
10	0.05	10 < LOQ
11	0.03	11 0.02
12	0.01	12 < LOQ
13	0.16	13 < LOQ
14	0.02	14 0.25*
15	0.47	15 < LOQ
16	0.06	16 0.28*
17	0.09	17 1.50*
18	0.04	18 < LOQ
19	0.08	
20	0.48	
21	0.01	
22	0.03	
n > LOQ	100%	50%
average	0.10	0.47
SD	0.13	0.57
min	0.01	0.02
max	0.48	1.50
median	0.05	0.25
1st quartile	0.03	0.17
3rd quartile	0.14	0.34

Samples were obtained from unpaired mother-infants. All concentrations are corrected for total blank signals. *Weighted mean for 3 analyses. LOQ = 0.01 mg/kg for meconium and 0.003 mg/kg for placenta. Typical relative measurement uncertainties were 6.5% (k = 1) for meconium, and 8% (k = 1) for placenta

and cord blood and suggested a high placental transfer efficiency of Ti [38]. Titanium was also evidenced in the amniotic fluid, even if the prevalence of pregnant women exhibiting amniotic Ti load is quite low [39]. However, whether Ti signal originated from (nano) particulate matter was not specified and, to date, no analytical data exist on the potential accumulation of TiO₂ particles in the human placenta and the meconium, the latter being used herein as a biomarker of prenatal exposure of the newborn to TiO₂ from the mother. In this

study, a combination of two analytic methods with ICP-MS and STEM-EDX allowed us to determine the total Ti content in these biological matrices together with the morphology, size and chemical characterization of the particles identified in the placental tissue sections and in the meconium preparations for electron microscopy. We report that all collected placentae from term pregnancies had a quantifiable amount of Ti, with an average level of 0.10 mg/kg of tissue, which is consistent with the findings of Li et al. [29] (0.2 mg/kg) and close to the Ti concentrations under particulate form in the human liver and spleen, as previously measured by single particle (sp) ICP-HRMS in organs recovered post-mortem [9]. Furthermore, seven placentae exhibited Ti contents between 0.1 and 0.5 mg/kg of tissue, these amounts being comparable to the maximum Ti levels reported in the human spleen [9], suggesting a high capacity of the placenta to accumulate Ti(O₂) from the maternal blood. TEM imaging coupled to EDX analysis showed the presence of isolated and clustered TiO₂ particles; most of the constitutive particles were spherical and assumed to have an anatase crystal structure, while others consisted of elongated TiO₂ particles typical of the rutile forms. The rutile and anatase forms of TiO₂ are authorized in the formulations of the E171 food additive in the EU (EC Regulation n°231/2012), while rutile is also commonly used in sunscreens and cosmetics such as day creams and loose and pressed powders [40, 41], with possible inhalation and/or absorption through the skin, although very few studies have revealed dermal penetration through intact skin [3]. Furthermore, TiO₂ is used as an opacifying agent in indoor paints with particles released into domestic air with dust and pulmonary exposure [42–44]. However, all these different routes of exposure cannot be discriminated on the sole base of the crystal forms recovered in the tissue due to their different usages both in the environment and in consumer products. Furthermore, we consistently identified particle deposits of Al, Si, Fe, Sn and Zn in the term placentae. These accumulated particles may originate from dietary intake [45] and pharmaceutical products [5] along with other sources such as air pollution in an urban environment [46–48]. With the example of tin, our data showing deposition as particles in the human placenta are consistent with the exposure profiles to Sn measured in the maternal blood and the tin accumulation in the placenta as reported by ICP-MS measurements [49], which requires further research regarding placental transfer.

For meconium, 50 % of the samples had a quantifiable amount of Ti (range 0.02–1.50 mg/kg), with some of them exhibiting Ti contents well above those recovered in the placenta. Meconium reflects a long window of exposure and is considered a relevant matrix in toxicology

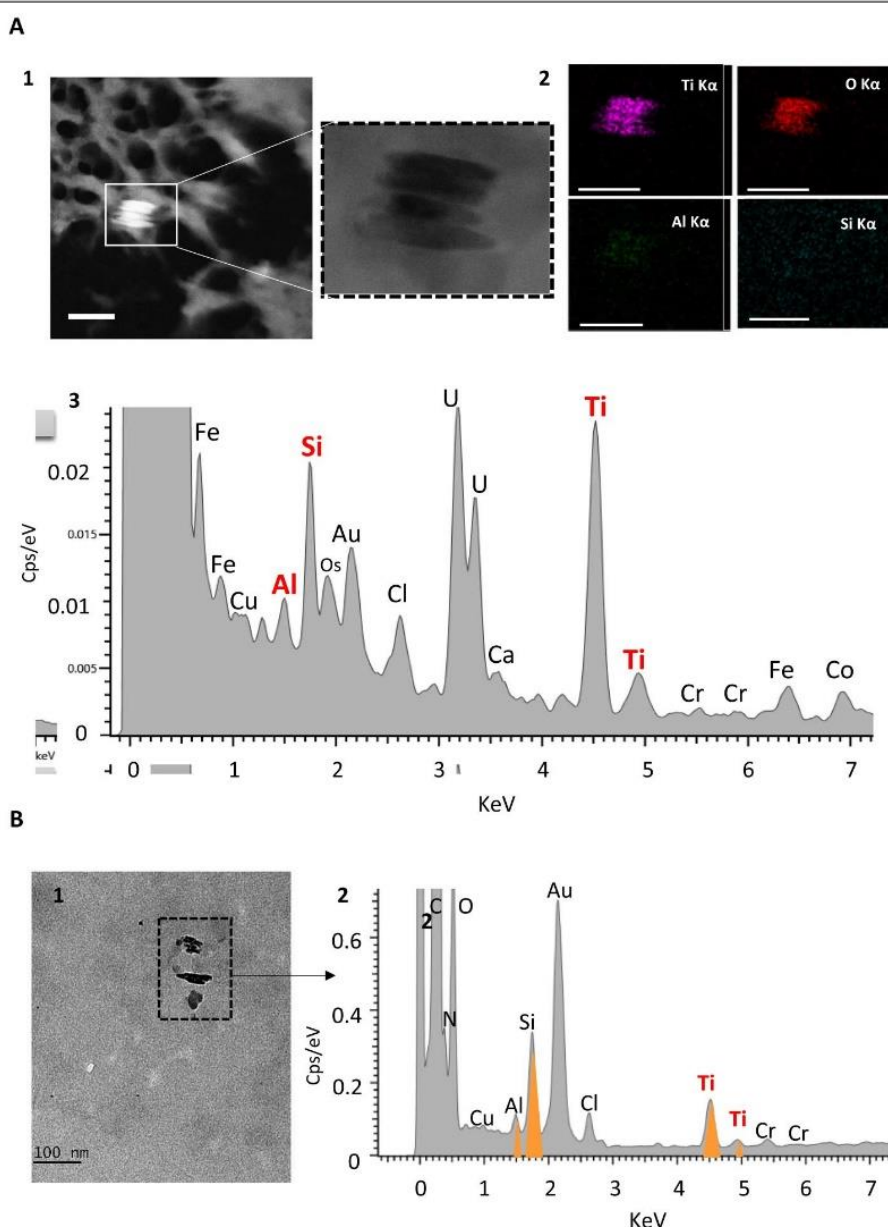
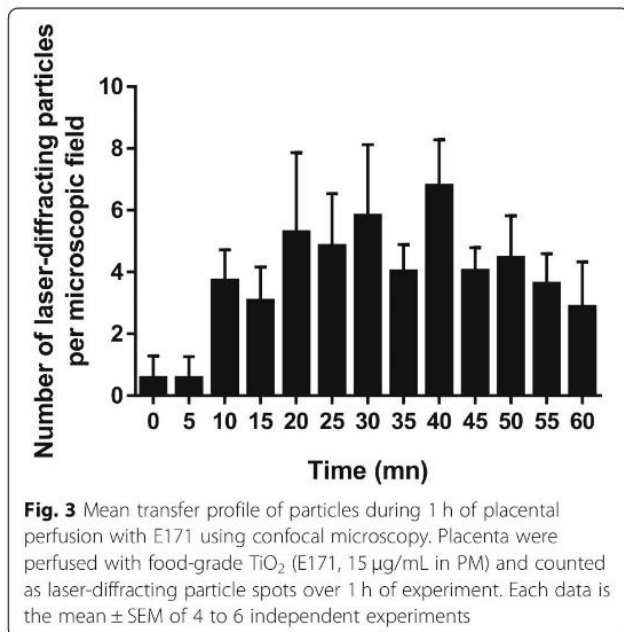


Fig. 2 STEM-EDX analysis of TiO₂ particles basal distribution in human placenta (a) and meconium (b). (A1) (S) TEM -micrographs showing cluster of four particles. Left hand corner micrograph is acquired in STEM-HAADF mode where particles appear brighter than biological matrix, due to the detection of elastic electron scattered by high Z number particles (chemical contrast). (A2) STEM-EDX elemental maps and sum spectrum identifying Ti, O, Al, and Si as main elements, and (A3) corresponding EDX analysis. All scale bars = 100 nm. (B1) TEM image illustrating particle content in meconium and coupled to EDX analysis (B2) with Ti, O, Al and Si identified as main elements over 3 particles

screening for detecting foetal exposure to various xenobiotics, such as airway pollutants, heavy metals, pesticides or pharmaceuticals, with which pregnant women are in contact [30, 50]. These chemicals can migrate into the foetus directly from the placental interface through the cord blood or come from the amniotic fluid of which formation partly results from transudation of maternal fluids together with secretions of the amnion epithelium [51]. Because meconium matrix contains amniotic fluid

mixed with foetal urines, all being continuously swallowed (as well as inhaled) by the foetus during pregnancy, and reabsorbed by the foetal intestine [51], such matrix may reflect global foetal body exposure to TiO₂. From the current ICP-MS analysis, the high variability of Ti content among the meconium samples, as shown in placenta, suggests inter-individual differences in mothers' exposure to TiO₂ and/or in basal permeability of biological barriers to TiO₂ that remain unexplored.

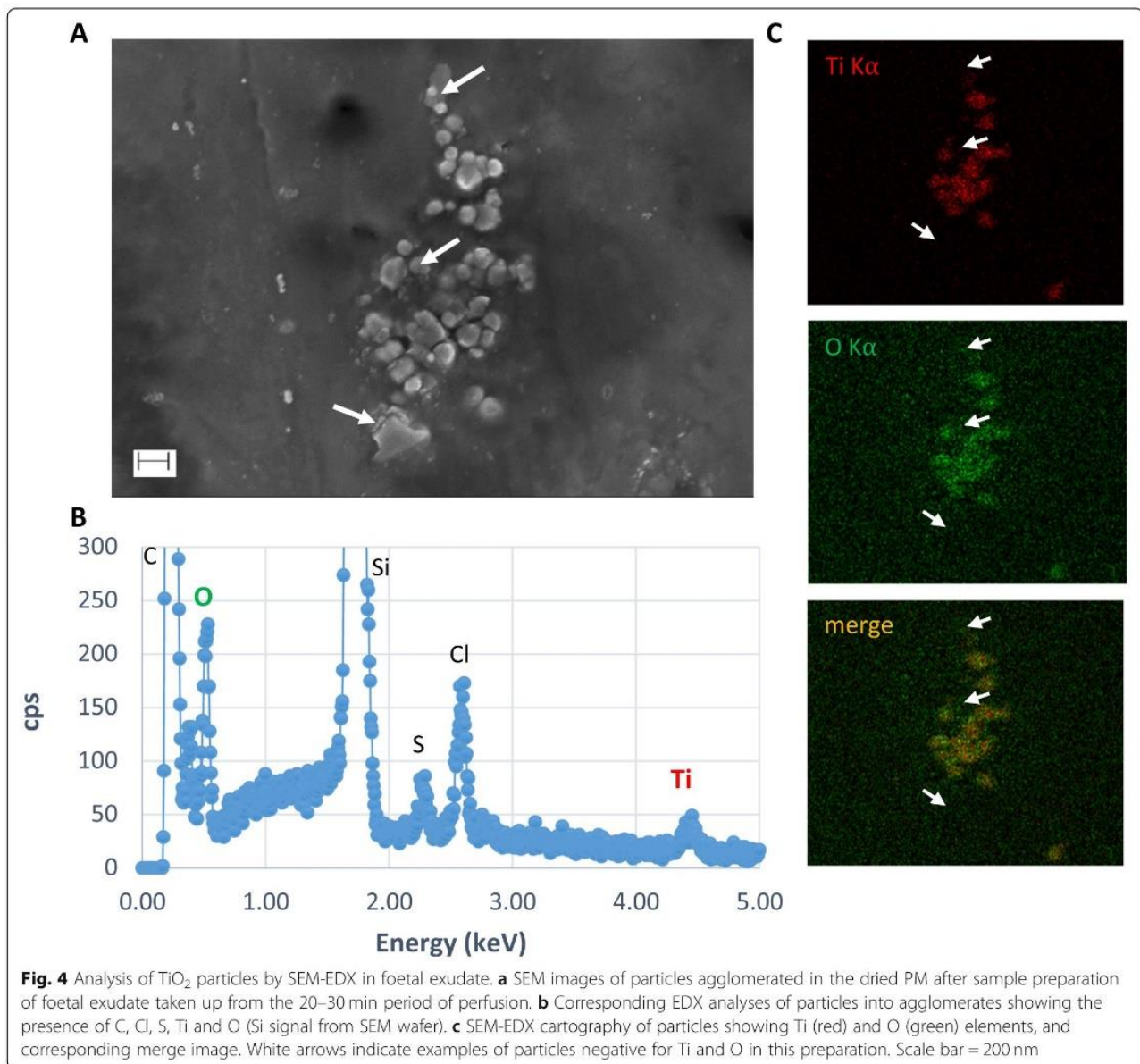


Given the maternal origin of xenobiotics in meconium [30], it is assumed that foetal exposure to TiO₂ directly relies on the mother's use of TiO₂-containing products and/or TiO₂ emission in her environment, which may have multiple origins on a household basis, as noted above. In our study, further EDX analysis confirmed TiO₂ particle deposition in the meconium. Furthermore, all the recovered TiO₂ particles were nanosized and often exhibited morphology resembling that of elongated TiO₂ (putative rutile) structures, as observed in the placental tissues. In addition, Fe particles and particulates deposits with Al and Si were found in the meconium as in the placenta. Notably, materno-foetal translocation of Si-NPs has been demonstrated *in vitro* in the placental BeWo cell line [33], which represents the rate-limiting barrier for maternal-foetal exchange, as well as *ex vivo* in perfused human placenta [33] and *in vivo* in mice [52]. For Al and Si particle deposits, as noted above, both elements are ubiquitous in the environment and diet [45, 53, 54]. Finally, for TiO₂, because the present study was not conducted on mother-child pairs, it is not possible to relate the Ti levels measured in the meconium to the amount found in the placenta. Nonetheless, altogether, the current findings provide evidence of a materno-foetal passage of TiO₂ nanoparticles in humans.

To confirm this passage, we conducted a study using the *ex vivo* human placenta perfusion model considered the gold standard for studying human materno-foetal transfer of xenobiotics [55], including nanoparticles [31, 32]. We focused on E171 food additive due to recent risk assessment studies highlighting oral ingestion as the major TiO₂ route of exposure in the general population

[9, 56]. This conclusion was based on the substantial daily ingestion of food-grade TiO₂ used as an E171 additive (up to several milligrams per kilogram of BW per day [15]) in the absence of an acceptable daily intake, and on toxicokinetic information for accumulation in organs and clearance from tissues [56]. We first showed poorly detectable Ti signals in the foetal exudate, suggesting a passage of particles too low to be distinguished from the blank levels after 1 h of E171 perfusion. Similar observations have been reported by other groups after 6 h of perfusion with TiO₂-NPs models, with Ti signals in the range of the background levels of the control perfusion medium [34, 35], as herein reported. In these studies, including the present, the TiO₂ concentration used for perfusion (10 to 25 µg/mL for toxicokinetic purposes) was approximately 1000 times higher than the basal Ti blood level in humans (10 µg/L) [4, 5] in order to optimize particle detection under a short time of perfusion (i.e., maximum 6 h) and given the short viability of the placenta *ex vivo* in a non-recirculating system. A high TiO₂ concentration in perfusion medium could rapidly lead to clogging of the intervillous space on the maternal side, limiting particle recovery on the foetal side. Indeed, in the study of Wick et al. (2010) using 25 µg/mL PS beads for a 6 h perfusion, nanosized beads were reported to cross the placental barrier only during the first hour, with no subsequent transfer when the PS beads were *de novo* added after 3 h in the maternal compartment. The authors suggested that the high dose of particles in the intervillous spaces created an agglomeration close to the placental villus on the maternal side, preventing the transfer of beads to the foetal compartment after a certain time [31]. These results suggested that a similar situation for TiO₂ particles could explain the low Ti levels on the foetal side, as reported in the above perfusion studies [34, 35]. In our study, we focused on a one-hour perfusion to limit the risk of placental obstruction by TiO₂ particles from the perfusion medium with E171.

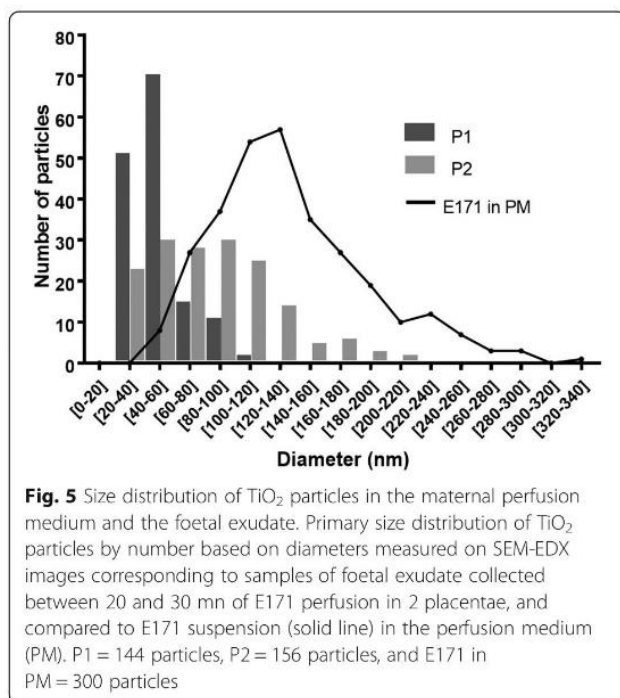
To avoid uncertainties in our study, we used confocal microscopy for particle visualization in the foetal effluent (i.e., appearing as laser-diffracting metal particles in the exudate) and SEM-EDX analysis to ascertain their chemical nature as TiO₂ particulate matter. Based on the confocal transfer profile, a progressive increase of the TiO₂ particles was recovered in the foetal side as soon as 10 min after the E171 suspension was added in the maternal compartment. Interestingly, a rapid passage was also reported for gold (Au) NPs (20 nm) across the rat placenta via *ex vivo* perfusion, i.e., within 20 min following material infusion [57]. In our study, given the poorly quantifiable level of Ti by ICP-MS in the foetal effluent following E171 perfusion, it is assumed that the mean transfer rate of the TiO₂ particles over time is very low.



However, given that our experiment was conducted for only 1 h, the overall transfer of particles during the whole duration of pregnancy could account for a non-negligible accumulation of TiO_2 (nano) particles in the foetal body. This assumption is well supported by our above demonstration of the presence of Ti in 50% of the meconium samples collected for this study. Furthermore, we showed that a large majority of the TiO_2 particles found in the foetal exudate after E171 perfusion was in the nanorange. Because the size distribution of TiO_2 particles recovered into tissues of a perfused cotyledon showed 50% composed of TiO_2 nanoforms, this suggested that the fraction of larger TiO_2 particles remained mostly trapped into the placental tissues and did not reach the foetal compartment. Both these findings

indicate that the placenta cannot be considered as an absolute barrier preventing the passage of TiO_2 NPs from the maternal blood, as previously reported using PS beads or Au-NPs with similar perfusion approaches [31, 58]. To date, the mechanisms for NP translocation are still unknown, but the main hypothesis for NP transport involves transtrophoblastic channels and/or endocytotic mechanisms across the placental barrier [32, 58, 59].

From a risk perspective, animal studies raised concerns about NPs entering the foetus in rodents exposed through inhalation or the transcutaneous or oral route. A large variety of effects on developmental processes have been reported for TiO_2 -NPs, particularly on brain functions due to translocation through the foetal blood–



brain barrier, with consequences on behaviour [22]. In addition, because immunotoxic and antibacterial properties have been reported for TiO₂ (nano) particles, including the food-grade form [7, 60–63], their accumulation in the foetal gut through the meconium, as shown herein, could affect the primary colonization of the intestine by the microbiota at birth as well as the maturation of the intestinal immune system, two perinatal events of which dysfunctions have long-term health consequences, as recently reviewed [64]. Accumulation in the placenta may also lead to placental dysfunction (e.g., dysregulation of vascularization, inhibition of cell proliferation, induction of apoptosis) and subsequent foetal growth restriction [11]. In addition to the findings in these rodent studies, TiO₂-NPs can trigger autophagy and mitochondrial dysfunction *in vitro* in human trophoblast cells [65, 66], and further impair the cell migration [67] that is needed for trophoblastic invasion in the uterine wall, permitting embryo implantation. Therefore, additional studies are needed to quantify and further characterize the human foetal exposure to TiO₂-NPs shown in our study, together with studies in animals chronically exposed to TiO₂ during pregnancy, including from the oral route, to determine the potential hazards for foetal development and newborn health.

Conclusion

By combining different methods for particle detection, element composition and size analysis together with Ti quantification, the present study highlighted the passage of TiO₂ particles across the human placenta with

potential local accumulation during pregnancy depending on individuals. Even if the placental transfer could not be quantified, a translocation to the foetal body seems to exist, as reflected by the Ti content and TiO₂ particle deposits in meconium samples, mainly of nano-sized forms, that suggest prenatal TiO₂ exposure from various sources. We further demonstrated TiO₂ particle transfer through isolated human placenta perfused with food-grade TiO₂ from the E171 additive, concluding that the human placental barrier is unable to completely prevent the passage of TiO₂-NPs from dietary sources and protect the foetus. Finally, given the current study showing *ex vivo* a materno-foetal passage of nanosized TiO₂ particles in humans, together with quantitative data from different cohorts showing placental and meconium Ti load at term of pregnancy, our findings emphasize the need to assess the risk of TiO₂-NP exposure in pregnant women and warrant specific attention for oral exposure to the nanosized fraction of the E171 food additive.

Methods

Particle characterization and preparation

The E171 food additive was purchased as powder from the website of a French commercial supplier of food colouring agents. These food-grade TiO₂ particles were prepared following the generic Nanogenotox dispersion protocol as previously described [7]. The E171 stock suspension (25.6 mg/mL) was sonicated in an ice bath for 16 min at 30% amplitude (VCX 750-230 V, Sonics Materials) to obtain a stable dispersion of TiO₂ particles and then stocked at 4 °C during 15 days maximum before use. Dynamic light scattering (DLS, ZetaSizer nano ZS; Malvern Instruments Ltd.) measurements were performed on TiO₂ particles in ultrapure water (pH = 8.92) and perfusion medium (PM = Earle + 2% BSA, pH 7.4). Five µL of E171 stock suspension were diluted in 3 mL of PM and the hydrodynamic diameter (Z-average), polydispersity index and zeta potential of the TiO₂ particles were measured. Particle diameters were measured by scanning electron microscopy (SEM, *n* = 600 particles) and are expressed as the percentage of NPs by number. The specific surface area (SSA) of the particles was assessed according to the Brunauer, Emmet and Teller method (BET).

Human placenta and meconium collection

The study was conducted in accordance with the Declaration of Helsinki and its later amendments. Placentae and meconium were collected from different sites and thus were not from mother-infant pairs. Term placentae were collected from 22 uncomplicated pregnancies in the CHU Paule de Viguiet (Toulouse, France; Institutional approval (DC-2013-1950). Signed informed consent was obtained from all the mothers. Meconium

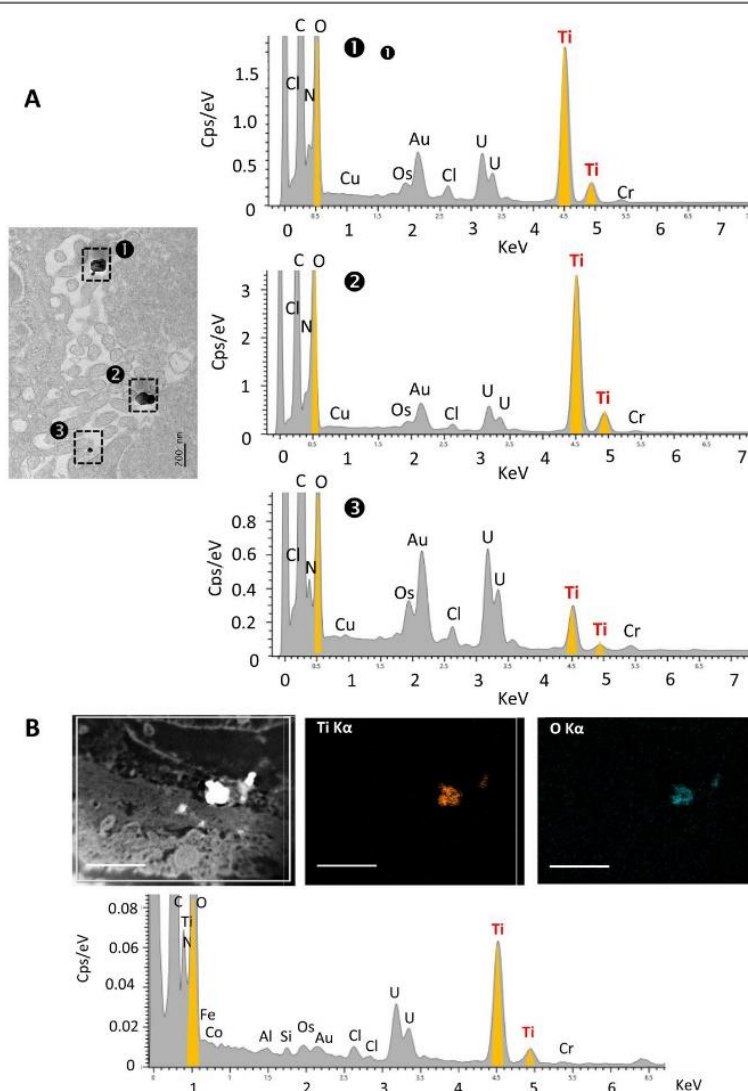


Fig. 6 Tissue distribution of TiO₂ particles in the E171 perfused placenta. **a** TEM-Bright-Field micrographs coupled to EDX analysis showing food-grade TiO₂ particles into the syncytiotrophoblast microvilli. **b** Chemical characterization of TiO₂ particles by STEM-EDX mapping based on STEM-HAADF micrographs for particle detection, and elemental maps and sum spectrum. Scale bars in (B) = 1 μm

samples (< 48 h post-partum) were collected in the maternity ward of Sainte-Thérèse Clinic (Paris, France), under an internal agreement of the scientific committee ($n = 11$), or from volunteer parents ($n = 7$). Informed signed consents were obtained from all parents. Meconium collection is non-invasive as it is directly recovered by scraping stained diaper with a sterile disposable spatula, taking care not touching nappy surface.

For basal Ti level determination in all placentae and meconium, a biopsy from one cotyledon per placenta and all meconium samples were weighed and stored at -80°C before analysis. Samples from 2 placentae and 2 meconium were also prepared for STEM-EDX analysis.

For ex vivo perfusion experiments, after a visual examination of organ integrity, 15 placentae (623 ± 186 g)

collected after caesarean section ($n = 8$) or normal vaginal delivery ($n = 7$) were used.

Ex vivo placental perfusion model

The placentae were transported in a thermostatically controlled container (37°C) and prepared for perfusion within 1 h of delivery in an open double circuit system as previously described [68]. Briefly, an intact peripheral cotyledon was chosen, and the truncal branch of the chorionic artery supplying the cotyledon and its associated vein were catheterized (Microtube Tygon S54-HL 1.02*1.78 mm, Courbevoie, France). The perfusion medium (PM) was Earle's medium (Euromedex, Souffel Weyersheim, France) supplemented with 2 g/L of BSA (Fraction V, PAA Laboratories, Vélizy-Villacoublay,

France), and the foetal flow rate was established at 6 mL/min. After a few minutes, the perfused cotyledon progressively whitened, allowing its isolation and was placed into a thermostatically controlled glass receptacle maintained at 37 °C, with maternal side facing upwards. Two hypodermic cannulas were used to perfuse the maternal side of the placenta with a flow rate established at 12 mL/min. Each placenta was perfused for 30 min with PM alone to flush the blood out of the maternal and foetal circulation and then for 1 h with ($n = 13$) or without ($n = 2$) PM with food-grade (E171) TiO₂ (15 µg/mL). To ensure proper dispersion, the E171 stock suspension was vigorously agitated using a vortex before addition in the maternal circulation. The perfusion medium was under constant agitation for oxygenation, allowing the homogeneous dispersion of TiO₂ particles.

During the whole perfusion time (90 min), the foetal and maternal pH values were continuously adjusted throughout the experiments to 7.27 ± 0.05 and 7.41 ± 0.01 (mean \pm SD), respectively, and the temperature and flow rates were checked. Antipyrine (Sigma-Aldrich), a reference control substance for passive diffusion and barrier integrity, was added to the maternal reservoir (20 µg/mL) at the beginning of the perfusion (t_0), and antipyrine concentrations were measured in foetal exudates collected every 15 min by high performance liquid chromatography coupled with UV detection. All placentae with a maternal to foetal transfer of antipyrine lower than 20% were excluded from the study.

For assessment of the materno-foetal transfer of TiO₂ particles, repeated samples (2 mL each) from the foetal flow through were collected every 5 min before and after E171 addition on the maternal side and stored at 4 °C after the addition of fungizone (3.5 µL/mL) and penicillin + streptomycin (10 µL/mL) to prevent from bacterial growth. These samples were used for confocal microscopy detection of laser-reflecting particulate matter that reached the foetal side. In addition, pools of foetal exudate (≈ 50 mL) were collected by 10- or 15-min periods over the whole perfusion time and then stored at -20 °C for subsequent ICP-MS analysis of the Ti concentration and SEM-EDX detection of the TiO₂ particles.

At the end of the experiment, the perfused cotyledon was washed for 20 min with PM only to flush out TiO₂ particles that did not penetrate the tissue. Tissue samples were collected close to the perfusion cannula and taken along the materno-foetal axis as representative areas for particle diffusion from the maternal side to the foetal circuit, while adjacent non-perfused cotyledons were also collected. All these tissue samples were prepared for STEM-EDX observations of particle distribution.

ICP-MS analysis of titanium

Placenta biopsies, foetal flow through collected during perfusion and meconium were analysed for Ti

concentration with a Thermo Element XR mass spectrometer (ThermoFischer, Bremen, Germany) operated in high resolution mode. Samples were injected with a Seaspray nebulizer in self-aspiration (Glass expansion, Melbourne, Australia) and a quartz double-Scott Peltier-cooled spray chamber. Measurements were performed on isotopes ⁴⁷Ti, ⁴⁸Ti and ⁴⁹Ti, and Ti concentrations were calculated based on the average signals of the three isotopes. The instrument was tuned daily according to the manufacturer's recommendations, and the Ti spectra were confirmed to be free of interference, particularly calcium interference, and its effect on ⁴⁸Ti was correctly resolved. For foetal exudate, due to the high Ca content (0.49 mM) of the Earle's medium used for perfusion, only signals on ⁴⁷Ti and ⁴⁹Ti were selected for calculations.

Prior to analysis, placental tissue samples were dried at 700 °C in a furnace to burn the organic content without damaging the mineral particles. For foetal exudate, each 10-min pool (50 mL) collected during placental perfusion (90 min) was evaporated at 120 °C to near dryness to pre-concentrate the sample. Placental tissue ashes and pre-concentrated perfusates were digested by the addition of 0.5 mL of HF (40% HF, Suprapur®, Merck) and 10 mL of HNO₃ (65% HNO₃, Suprapur®, Merck), evaporated at 120 °C, and suspended in ultrapure water to a final volume of 50 mL. The recovery rates of these procedures were assessed by spiking samples with E171, and yielded 104% for the placental tissues and 85% for the exudates. Digestion blanks and control perfusates (no E171 addition) were used to correct the Ti concentrations from blank contributions.

For meconium, approximately 0.5 mg was sampled and digested in 0.5 mL of HF and 10 mL of HNO₃ and maintained at room temperature (RT) for 2 days before evaporation at 60 °C. These steps were repeated once or twice until complete digestion. The samples were then diluted in 15 mL of ultrapure water and subsequently diluted in 2% HNO₃ for analysis. The recovery rates of this procedure were assessed against NIST 1566b and yielded $99.9 \pm 10.2\%$ on average ($n = 2$, 1 S.D.).

The LOD was calculated as 3 times the standard deviation in the results of a blank sample and the LOQ as 10 times the standard deviation, and the results are shown in mg/kg or ng/mL for each sample.

Particle detection in the perfusion medium by confocal microscopy

For each perfused placenta, a 35 µL volume droplet of the foetal exudate collected every 5 min from t_0 to t_{90} min of perfusion was placed on 8-well microscopy slides, dried, and mounted in Mowiol medium (glycerol 240 g/L, DABCO 25 g/L, Mowiol 4-88 96 g/L, Tris-HCl 15 g/L) within 24 h after the collect. The

slides were then examined under a confocal microscope (Leica SP8) at 488/BP 488–494 nm to detect light scattering (i.e., laser-diffracting) TiO₂ particles appearing as green fluorescence and at 514/BP 560–660 nm to monitor autofluorescence in the sample, as previously described [7, 8]. Three fields per well were examined, and laser-diffracting (particle) spots were counted using a 63X objective and a magnification factor of 1 pixel to 50 nm. Particles counted in the microscopic fields corresponding to exudate samples from the equilibrium period (from 10 to 30 min) and over the whole experiment with a control placenta perfused with PM alone were considered the background level. To obtain a transfer profile to the foetal circuit of TiO₂ particles during 1 h after E171 addition in the maternal reservoir, we calculated the background particle level as an average number of particles per observed field, and the value was subtracted from each corresponding value in all placentae perfused with the E171 suspension.

Preparation of foetal exudate samples for SEM-EDX analysis

To characterize the particles crossing the human placenta, we selected pooled samples from foetal exudate (50 mL) corresponding to the 20–30 min of perfusion after E171 addition in the maternal reservoir based on the particle transfer profile by confocal imaging. Samples were lyophilized to concentrate the particles, and 5 g of powder was suspended in 1 mL of HCl to dissolve organic compounds, sonicated (20 min, 75 W), and centrifuged (10 min, 9503 g) and the supernatant was eliminated. The sample was then washed 5 times as follows: addition of 1 mL of HCl, sonication (1 min, 150 W), centrifugation (10 min, 9503 g) and elimination of the supernatant. The preparation was suspended in 1 mL of HCl, and then, a 7.5 µL droplet was deposited onto a silica wafer by spin-coating according to a method developed by the LNE to perform NP dimensional metrology in complex matrices [69, 70]. Briefly, for good dispersion of the NPs on the silicon substrate while the solvent evaporated, the droplet was spread over the surface of the silica wafer using a low spin speed (1000 rpm, 5 min), and then dried rapidly for 10 s at a fast spin speed (8000 rpm). The samples were stored at RT under a controlled atmosphere until observation.

SEM imaging was performed using a Zeiss Ultra-Plus field emission (FE) microscope equipped with a Gemini optical column. Images were obtained through secondary electrons collected by an InLens detector at a voltage of 3 kV and with a working distance equal to 3 mm. In these working conditions, the provider claimed the resolution of the microscope was 1.7 nm. An EDX detector (Princeton Gamma-Tech Instruments, Princeton, USA)

was installed in the SEM chamber for a qualitative elemental analysis of the observed particles. The size distribution of the TiO₂ particles observed with SEM-EDX was assessed in a semiquantitative way with ImageJ software (NIH, USA), in PM containing 15 µg/mL of E171 and in foetal exudate from 2 placentae.

Tissue preparation for TEM

Samples from 2 placentae and 2 meconium were fixed in 2% paraformaldehyde-2.5% glutaraldehyde in 0.1 M cacodylate buffer (pH 7.4) for 1 h at 4 °C. For placentae, tissue blocks of 2.5 mm length were excised and immersed in the same fixative overnight at 4 °C. After several rinses in cacodylate buffer, the samples were in 1% OsO₄ (Osmium (VIII) oxide) for 1 h at 4 °C and rapidly rinsed again. Dehydration was carried out at 4 °C using a graded series of ethanol and acetone. The sections were impregnated with low viscosity epoxy resin (EMS) under a vacuum and polymerized at 60 °C for 48–72 h. Ultrathin sections (50–60 nm, ultra-cut UCT, Leica) were collected on gold-600 mesh grids and stained for 7 min with 0.5% uranyl acetate in methanol solution before TEM observations. The same protocol was used for meconium samples, except that osmium post-fixation and uranyl staining were not performed.

Scanning (S)TEM-EDX analysis, elemental mapping and particle size measurement

TEM-EDX analysis was performed on a JEM 2010 (JEOL, Tokyo, Japan) operating at 120 KV and equipped with a LaB6 cathode and two CCD cameras (Oriu 1000 and ES500W) driven by Digital Micrograph software (Gatan-Ametek, Pleasanton, United States). EDX-analysis was performed with a Silicon-drifted Detector (SDD) (Resolution: 125eV Mn_K) (Oxford-Instruments, Abington, Oxfordshire, England). Elemental mapping was performed on JEM-ARM200F HR-TEM ((JEOL, Tokyo, Japan) with analytical configurations. Elemental maps were acquired at 80 kV in STEM mode with an 8C probe size, a camera length of 8 cm, and a 50 µm condenser aperture with an SDD detector XMax TLE (Resolution: 127eV Mn_K 0.7 sr; Oxford-Instruments, Abington, England). Size measurement was determined from bright-field TEM images by using the image processing open-source software ImageJ (NIH, United States).

Data analysis

The data are presented as the mean ± SD for determination of the TiO₂ particle diameter in an E171 water suspension and for the Ti dosage in the placentae and the meconium (Table 1, Table 2) or as the mean ± SEM for particle counts and size measurements in the placenta and meconium TEM sections and in the foetal exudate (Figs. 3, and 5).

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12989-020-00381-z>.

Additional file 1 Table S1. ICP-MS analysis of Ti content in foetal exudate collected during control or E171 perfusion. Ti content of foetal exudates from 6 independent E171 perfusion experiments (15 µg/mL) and 1 control perfusion. Samples were collected by time fraction of 10- or (*)15-min during perfusion (first 30 min of 1 h for E5 and E6, total 60 min for E1, E2, E3, E4, and E7). All concentrations are corrected for total blank signals. LOD = 0.23 ng/mL. Typical relative measurement uncertainty was 5% (k = 1). **Fig. S1.** Basal particulate content in term human placenta. (A and B) Representative TEM-Bright Field micrographs showing particulate matter in the placental tissues. Elemental characterization was determined by TEM-EDX analysis. In addition to Carbon and Oxygen, the following elements were found: ⊙: Tin, Iron, Silicon; ⊖: Silicon; ⊕: Tin, Zinc, Iron, Manganese, Phosphorus, Silicon; ⊗: Silicon; ⊙: Iron, Aluminium, Silicon; ⊕: Silicon, Aluminium; ⊖: Silicon, Aluminium. (C) EDX-Blank spectrum of biological matrix: elements in deconvolution are Uranium (U), Osmium (Os), Gold (Au), Copper (Cu), Chromium (Cr) coming from grids, sample preparation and staining of TEM pieces. **Fig. S2.** Particle diameter distribution in term human placenta. Size measurement of particles recovered per field micrograph on ultrafine sections from 2 placentae collected at term of pregnancy. Dashed line represents the 100 nm limit. **Fig. S3.** Basal particulate content in human meconium. (A to C) TEM-Bright Field micrographs of meconium ultrafine sections showing particles in meconium, and combined TEM-EDX analysis for elemental characterization; ⊙: Iron (Fe), Silicon (Si), Magnesium (Mg), Calcium (Ca), Aluminium (Al); ⊖: Titanium (Ti), Aluminium (Al), Silicon (Si). Elements in deconvolution are Gold (Au), Copper (Cu), Chromium (Cr) coming from grids, sample preparation of TEM pieces. **Fig. S4.** Antipyrine rate transfer depending on the placental origin. Foeto-maternal rate transfer of antipyrine during ex vivo perfusions of placenta collected after vaginal delivery (n = 5) or caesarean section (n = 6). Note that there is no difference in basal permeability depending on the mode of delivery. Data are presented as mean ± SD. **Fig. S5.** Confocal imaging of laser-diffracting particles in perfusion medium with E171, foetal exudate and perfused placenta. (A) Laser-diffracting particles in the E171-containing PM added to maternal side. Scale bar = 50 µm. (B) Foetal exudate collected between 30 and 35 min of E171 perfusion. White arrows indicate laser-diffracting particles. Scale bars = 50 µm. (C) Tissue section of perfused placenta showing laser-diffracting particles spread in the intervillous spaces (ivs) of the maternal side close to the syncytiotrophoblast (sct). White arrows indicate foetal vessels. Scale bar = 100 µm. **Fig. S6.** Complementary EDX analysis of foetal exudates. Purified samples of particles recovered in foetal exudates and showing (A) aluminium (Al) and (B) iron (Fe) elements associated or not with titanium (Ti). Corresponding SEM micrographs of purified particles in upper right panels; Si signal from SEM wafer. **Fig. S7.** Method for size measurement of TiO₂ particles in the foetal exudate. (A) and (B) TiO₂ particles were identified by an EDX detector coupled to SEM chamber for Ti (green) and O (red) element analysis (upper right panels in A and B), and observed as agglomerates (A) or as isolated particles or small aggregates trapped into the matrix of dried PM after sample preparation, i.e., resulting from the sample deposition onto silicate wafer by spin-coating and evaporation of the solvent [62]. Only the diameters of particles showing Ti + O colocalization (yellow) were measured, here for some particles as examples. Scale bar = 200 nm. **Fig. S8.** Size measurement of TiO₂ particles in two representative E171-perfused placentae. (A) TEM image reconstruction showing tissue architecture across a placental villus along the materno-foetal axis, i.e., the syncytiotrophoblast microvilli and syncytiotrophoblast cells, the cytotrophoblast cells, then the chorionic mesenchyme composed of the basal lamina supporting trophoblast tissue, and the endothelial cells surrounding foetal capillaries. (B and C) Size distribution of TiO₂ (EDX-characterized) particles into perfused placenta, (B) trapped into microvilli of the syncytiotrophoblast on the maternal side, and (C) in deeper area until close to foetal vessels. (D) Representative TEM images of perfused TiO₂ particles (arrowheads) recovered in the

intervillous spaces (ivs) close to the syncytiotrophoblast microvilli (D1) and into the placental chorionic mesenchyme surrounding foetal capillaries (D2).

Acknowledgements

The authors wish to thank Sebastien Cambier (Luxembourg Institute of Science and Technology) for preliminary ICP-MS measurements. Nora Lam-beng and Caroline Oster (LNE) are also thanks for their help during SEM-EDX and ICP-MS analysis, respectively. All the nurses of the Toulouse University Hospital who participated to the placenta collection, and Clemence Moiron and Dr. Laurent Petit (Clinique Sainte Thérèse, Paris) for their interest in this project.

Authors' contributions

A.G., M.M., C.Co., V. G, K.C., C.V., F.de la F., N.P.-H., B.L. and E.H. designed the study; E.G., M.M., A.de P., V.B., K.C., N.B. and K.A.-P. contributed to sample collection; E.G., M.M., F.G., M.L., A.C. and A.G. conducted the perfusion experiments; A.G., C.Ca., L.D., L.C., and N.F. performed TEM, SEM-EDX and (S)TEM-EDX analysis; A.G. and J.N. performed ICP-MS measurements; M.L. performed antipyrine dosage; A.G., E.G., C.Ca., L.D., J.N., L.C., F.G., M.L., C.Co., V. G, V. B, K.C., K.A.-P., C.V., P.F., N.F., N.P.-H., B.L. and E.H. analyzed the data; A.G., L.D., J.N., L.C., V. G, K.A.-P., F.de la F., N.P.-H., B.L. and E.H. wrote the paper. All author(s) read and approved the final manuscript.

Funding

The study was supported by INRAE (Institut national de recherche pour l'agriculture, l'alimentation et l'environnement) and The Gerard Cazanave's Research Association. Adèle Guillard has fellowship from the French Ministry of Higher Education, Research and Innovation.

Availability of data and materials

All relevant data are included in the manuscript and supporting information, and available from the authors upon request.

Ethics approval and consent to participate

The collect of human placenta and use for ex vivo perfusion was performed following the principles of the Declaration of Helsinki and written informed consent was given by the mothers before delivery. The study was approved by the local ethics committee (DC-2013-1950). Meconium samples were obtained under an internal agreement of the scientific committee or from volunteer parents, and informed signed consents were obtained from all parents.

Consent for publication

All authors read and approved the final manuscript.

Competing interests

Authors declare that they have no competing interests.

Author details

¹Toxalim UMR1331 (Research Centre in Food Toxicology), Toulouse University, INRAE, ENVT, INP-Purpan, UPS, Toulouse, France. ²Department of materials, LNE, Trappes, France. ³Department for biomedical and inorganic chemistry, LNE, Paris, France. ⁴Group Physic of Materials, GPM-UMR6634, CNRS, Rouen University, Rouen, France. ⁵Department of Obstetrics and Gynecology, Paule de Viguier Hospital, CHU Toulouse, Toulouse, France. ⁶INHERES, UMR 1436 Toulouse University, INRAE, ENVT, Toulouse, France. ⁷Péritox UMR-I 01 (Perinatal and Toxic Risk), Jules Verne University, Amiens, France. ⁸Université Paris Saclay, CEA, INRAE, Département Médicaments et Technologies pour la Santé (DMTS), SPI, 91191 Gif-sur-Yvette, France. ⁹UMR 1027 INSERM, Team SPHERE, Toulouse III University, Toulouse, France.

Received: 6 May 2020 Accepted: 14 September 2020

Published online: 07 October 2020

References

- Skocaj M, Filipic M, Petkovic J, Novak S. Titanium dioxide in our everyday life; is it safe? *Radiol Oncol.* 2011;45:227-47.

2. Weir A, Westerhoff P, Fabricius L, Hristovski K, von Goetz N. Titanium dioxide nanoparticles in food and personal care products. *Environ Sci Technol*. 2012;46:2242–50.
3. Shakeel M, Jabeen F, Shabbir S, Asghar MS, Khan MS, Chaudhry AS. Toxicity of Nano-titanium dioxide (TiO₂-NP) through various routes of exposure: a review. *Biol Trace Elem Res*. 2016;172:1–36.
4. Böckmann J, Lahl H, Eckert T, Unterhalt B. Blood levels of titanium before and after oral administration of titanium dioxide. *Pharmazie*. 2000;55:140–3.
5. Pele LC, Thoree V, Brugggraber SF, Koller D, Thompson RP, Lomer MC, et al. Pharmaceutical/food grade titanium dioxide particles are absorbed into the bloodstream of human volunteers. *Part Fibre Toxicol*. 2015;12:26.
6. Jones K, Morton J, Smith I, Jurkschat K, Harding A-H, Evans G. Human in vivo and in vitro studies on gastrointestinal absorption of titanium dioxide nanoparticles. *Toxicol Lett*. 2015;233:95–101.
7. Bettini S, Boutet-Robinet E, Cartier C, Coméra C, Gaultier E, Dupuy J, et al. Food-grade TiO₂ impairs intestinal and systemic immune homeostasis, initiates preneoplastic lesions and promotes aberrant crypt development in the rat colon. *Sci Rep*. 2017;7:40373.
8. Coméra C, Cartier C, Gaultier E, Catrice O, Panouille Q, El Hamdi S, et al. Jejunal villus absorption and paracellular tight junction permeability are major routes for early intestinal uptake of food-grade TiO₂ particles: an in vivo and ex vivo study in mice. *Part Fibre Toxicol*. 2020;17:26.
9. Heringa MB, Peters RJB, Bleys RLAW, van der Lee MK, Tromp PC, van Kesteren PCE, et al. Detection of titanium particles in human liver and spleen and possible health implications. *Part Fibre Toxicol*. 2018;15:15.
10. Abbasi-Oshaghi E, Mirzaei F, Pourjafar M. NLRP3 inflammasome, oxidative stress, and apoptosis induced in the intestine and liver of rats treated with titanium dioxide nanoparticles: in vivo and in vitro study. *Int J Nanomedicine*. 2019;14:1919–36.
11. Zhang L, Xie X, Zhou Y, Yu D, Deng Y, Ouyang J, et al. Gestational exposure to titanium dioxide nanoparticles impairs the placentation through dysregulation of vascularization, proliferation and apoptosis in mice. *IJN*. 2018;13:777–89.
12. Chakrabarti S, Goyary D, Karmakar S, Chattopadhyay P. Exploration of cytotoxic and genotoxic endpoints following sub-chronic oral exposure to titanium dioxide nanoparticles. *Toxicol Ind Health*. 2019;35:577–92.
13. Dhupal M, Oh J-M, Tripathy DR, Kim S-K, Koh SB, Park K-S. Immunotoxicity of titanium dioxide nanoparticles via simultaneous induction of apoptosis and multiple toll-like receptors signaling through ROS-dependent SAPK/JNK and p38 MAPK activation. *Int J Nanomedicine*. 2018;13:6735–50.
14. Suzuki T, Miura N, Hojo R, Yanagiba Y, Suda M, Hasegawa T, et al. Genotoxicity assessment of titanium dioxide nanoparticle accumulation of 90 days in the liver of gpt delta transgenic mice. *Genes Environ*. 2020;42:7.
15. EFSA Panel on Food Additives and Nutrient Sources added to Food (ANS). Re-evaluation of titanium dioxide (E171) as a food additive. *EFSA J*. 2016;14(9):4545.
16. Dorier M, Béal D, Marie-Desvergne C, Dubosson M, Barreau F, Houdeau E, et al. Continuous *in vitro* exposure of intestinal epithelial cells to E171 food additive causes oxidative stress, inducing oxidation of DNA bases but no endoplasmic reticulum stress. *Nanotoxicology*. 2017;11(6):751–61.
17. Yamashita K, Yoshioka Y, Higashisaka K, Mimura K, Morishita Y, Nozaki M, et al. Silica and titanium dioxide nanoparticles cause pregnancy complications in mice. *Nature Nanotech*. 2011;6:321–8.
18. Umezawa M. Effect of fetal exposure to titanium dioxide nanoparticle on brain development - brain region information. *J Toxicol Sci*. 2012;37(6):1247–52.
19. Mölsä M, Heikkinen T, Hakkola J, Hakala K, Wallerman O, Wadelius M, et al. Functional role of P-glycoprotein in the human blood-placental barrier. *Clin Pharmacol Therapeutics*. John Wiley & Sons Ltd. 2005;78:123–31.
20. Hemauer SJ, Patrikeeva SL, Wang X, Abdelrahman DR, Hankins GDV, Ahmed MS, et al. Role of transporter-mediated efflux in the placental biodisposition of bupropion and its metabolite, OH-bupropion. *Biochem Pharmacol*. 2010;80:1080–6.
21. Delorme-Axford E, Bayer A, Sadovsky Y, Coyne CB. Autophagy as a mechanism of antiviral defense at the maternal-fetal interface. *Autophagy* Taylor Francis. 2013;9:2173–4.
22. Rollerova E, Tulinska J, Liskova A, Kuricova M, Kovriznych J, Mlynarcikova A, et al. Titanium Dioxide Nanoparticles. 2015;49:97–112.
23. Mohammadipour A, Fazel A, Haghiri H, Motejaded F, Rafatpanah H, Zabihi H, et al. Maternal exposure to titanium dioxide nanoparticles during pregnancy; impaired memory and decreased hippocampal cell proliferation in rat offspring. *Environ Toxicol Pharmacol*. 2014;37:617–25.
24. Ebrahimzadeh Bideskan A, Mohammadipour A, Fazel A, Haghiri H, Rafatpanah H, Hosseini M, et al. Maternal exposure to titanium dioxide nanoparticles during pregnancy and lactation alters offspring hippocampal mRNA BAX and Bcl-2 levels, induces apoptosis and decreases neurogenesis. *Exp Toxicol Pathol*. 2017;69:329–37.
25. Gao X, Yin S, Tang M, Chen J, Yang Z, Zhang W, et al. Effects of developmental exposure to TiO₂ nanoparticles on synaptic plasticity in hippocampal dentate Gyrus area: an in vivo study in anesthetized rats. *Biol Trace Elem Res*. 2011;143:1616–28.
26. Takeda K, Suzuki K, Ishihara A, Kubo-rie M, Fujimoto R, Tabata M, et al. Nanoparticles transferred from pregnant mice to their offspring can damage the genital and cranial nerve systems. *J Health Sci*. 2009;55:95–102.
27. Schmidt A, Morales-Prieto DM, Pastuschek J, Fröhlich K, Markert UR. Only humans have human placentas: molecular differences between mice and humans. *J Reprod Immunol*. 2015;108:65–71.
28. Enders AC, Blankenship TN. Comparative placental structure. *Adv Drug Deliv Rev*. 1999;38:3–15.
29. Li X, Li A, Zhang W, Liu X, Liang Y, Yao X, et al. A pilot study of mothers and infants reveals fetal sex differences in the placental transfer efficiency of heavy metals. *Ecotoxicol Environ Saf*. 2019;186:109755.
30. Woźniak MK, Jaszczak E, Wierowski M, Polkowska Z, Namieśnik J, Biziuk M. Meconium analysis as a promising diagnostic tool for monitoring fetal exposure to toxic substances: recent trends and perspectives. *TrAC Trends Anal Chem*. 2018;109:124–41.
31. Wick P, Malek A, Manser P, Meili D, Maeder-Althaus X, Diener L, et al. Barrier capacity of human placenta for Nanosized materials. *Environ Health Perspect*. 2010;118:432–6.
32. Grafmueller S, Manser P, Diener L, Diener P-A, Maeder-Althaus X, Maurizi L, et al. Bidirectional transfer study of polystyrene nanoparticles across the placental barrier in an *ex Vivo* human placental perfusion model. *Environ Health Perspect*. 2015;123:1280–6.
33. Poulsen MS, Mose T, Maroun LL, Mathiesen L, Knudsen LE, Rytting E. Kinetics of silica nanoparticles in the human placenta. *Nanotoxicology*. 2015;9:79–86.
34. Muoth C, Wichser A, Monopoli M, Correia M, Ehrlich N, Loeschner K, et al. A 3D co-culture microtissue model of the human placenta for nanotoxicity assessment. *Nanoscale*. 2016;8:17322–32.
35. Aengenheister L, Dugershaw BB, Manser P, Wichser A, Schoenenberger R, Wick P, et al. Investigating the accumulation and translocation of titanium dioxide nanoparticles with different surface modifications in static and dynamic human placental transfer models. *Eur J Pharm Biopharm*. 2019;142:488–97.
36. Iwai-Shimada M, Nakayama SF, Isobe T, Kobayashi Y, Suzuki G, Nomura K. Investigation of the effects of exposure to chemical substances on child health. *Nihon Eiseigaku Zasshi*. 2019;74.
37. Gaffet E, Marano F, Ferrari L, Flahaut E, Jouzel J-N, Madec L, et al. Bilan des connaissances relatives aux effets des nanoparticules de TiO₂ sur la santé humaine ; caractérisation de l'exposition des populations et mesures de gestion. 2018 Available from: https://www.researchgate.net/publication/326007817_Bilan_des_connaissances_relatives_aux_effets_des_nanoparticules_de_TiO2_sur_la_sante_humaine_caracterisation_de_l%27exposition_des_populations_et_mesures_de_gestion.
38. Li A, Zhuang T, Shi J, Liang Y, Song M. Heavy metals in maternal and cord blood in Beijing and their efficiency of placental transfer. *J Environ Sci*. 2019;80:99–106.
39. Raia-Barjat T, Prieux C, Leclerc L, Sarry G, Grimal L, Chaleur C, et al. Elemental fingerprint of human amniotic fluids and relationship with potential sources of maternal exposure. *J Trace Elem Med Biol*. 2020;60:126477.
40. Lu PJ, Fang SW, Cheng WL, Huang SC, Huang MC, Cheng HF. Characterization of titanium dioxide and zinc oxide nanoparticles in sunscreen powder by comparing different measurement methods. *J Food Drug Anal*. 2018;26:1192–200.
41. Lewicka ZA, Benedetto AF, Benoit DN, Yu WW, Fortner JD, Colvin VL. The structure, composition, and dimensions of TiO₂ and ZnO nanomaterials in commercial sunscreens. *J Nanopart Res*. 2011;13:3607.
42. Saber AT, Jacobsen NR, Mortensen A, Szarek J, Jackson P, Madsen AM, et al. Nanotitanium dioxide toxicity in mouse lung is reduced in sanding dust from paint. *Part Fibre Toxicol*. 2012;9:4.
43. Mikkelsen L, Jensen KA, Koponen IK, Saber AT, Wallin H, Loft S, et al. Cytotoxicity, oxidative stress and expression of adhesion molecules in

- human umbilical vein endothelial cells exposed to dust from paints with or without nanoparticles. *Nanotoxicology*. 2013;7:117–34.
44. Saber AT, Mortensen A, Szarek J, Jacobsen NR, Levin M, Koponen IK, et al. Toxicity of pristine and paint-embedded TiO₂ nanomaterials. *Hum Exp Toxicol*. 2019;38:11–24.
 45. Filippini T, Tancredi S, Malagoli C, Cilloni S, Malavolti M, Violi F, et al. Aluminum and tin: food contamination and dietary intake in an Italian population. *J Trace Elem Med Biol*. 2019;52:293–301.
 46. Grigoratos T, Martini G. Brake wear particle emissions: a review. *Environ Sci Pollut Res Int*. 2015;22:2491–504.
 47. Martin J, Bello D, Bunker K, Shafer M, Christiani D, Woskie S, et al. Occupational exposure to nanoparticles at commercial photocopy centers. *J Hazard Mater*. 2015;298:351–60.
 48. Gonet T, Maher BA. Airborne, vehicle-derived Fe-bearing nanoparticles in the urban environment: a review. *Environ Sci Technol*. 2019;53:9970–91.
 49. Iwai-Shimada M, Kameo S, Nakai K, Yaginuma-Sakurai K, Tatsuta N, Kurokawa N, et al. Exposure profile of mercury, lead, cadmium, arsenic, antimony, copper, selenium and zinc in maternal blood, cord blood and placenta: the Tohoku study of child development in Japan. *Environ Health Prev Med*. 2019;24:35.
 50. Berton T, Mayhoub F, Chardon K, Duca R-C, Lestremay F, Bach V, et al. Development of an analytical strategy based on LC-MS/MS for the measurement of different classes of pesticides and their metabolites in meconium: application and characterisation of foetal exposure in France. *Environ Res*. 2014;132:311–20.
 51. Beall MH, van den Wijngaard JPHM, van Gemert MJC, Ross MG. Amniotic fluid water dynamics. *Placenta*. 2007;28:816–23.
 52. Pietroiusti A, Vecchione L, Malvindi MA, Aru C, Massimiani M, Camaioni A, et al. Relevance to investigate different stages of pregnancy to highlight toxic effects of nanoparticles: the example of silica. *Toxicol Appl Pharmacol*. 2018;342:60–8.
 53. Food Safety Commission of Japan. Aluminium Ammonium Sulfate and Aluminium Potassium Sulfate (Food Additives). *Food Saf (Tokyo)*. 2019;7:79–82.
 54. Winkler HC, Suter M, Naegeli H. Critical review of the safety assessment of nano-structured silica additives in food. *J Nanobiotechnol*. 2016;14:44.
 55. Etwel F, Hutson JR, Madadi P, Gareri J, Koren G. Fetal and perinatal exposure to drugs and chemicals: novel biomarkers of risk. *Annu Rev Pharmacol Toxicol*. 2014;54:295–315.
 56. Heringa MB, Geraets L, van Eijkeren JCH, Vandebriel RJ, de Jong WH, Oomen AG. Risk assessment of titanium dioxide nanoparticles via oral exposure, including toxicokinetic considerations. *Nanotoxicology*. 2016;10:1515–25.
 57. D'Errico JN, Doherty C, Fournier SB, Renkel N, Kallontzi S, Goedken M, et al. Identification and quantification of gold engineered nanomaterials and impaired fluid transfer across the rat placenta via ex vivo perfusion. *Biomed Pharmacother*. 2019;117:109148.
 58. Semmler-Behnke M, Lipka J, Wenk A, Hirn S, Schäffler M, Tian F, et al. Size dependent translocation and fetal accumulation of gold nanoparticles from maternal blood in the rat. *Part Fibre Toxicol*. 2014;11:33.
 59. Buerki-Thurnherr T, von Ursula M, Wick P. Knocking at the door of the unborn child: engineered nanoparticles at the human placental barrier. *Swiss Med Wkly [Internet]*. 2012; doi: <https://doi.org/10.4414/smw.2012.13559>.
 60. Radziwill-Bienkowska JM, Talbot P, Kamphuis JBJ, Robert V, Cartier C, Fourquaux I, et al. Toxicity of food-grade TiO₂ to commensal intestinal and transient food-borne Bacteria: new insights using Nano-SIMS and synchrotron UV fluorescence imaging. *Front Microbiol*. 2018;9:794.
 61. Talbot P, Radziwill-Bienkowska JM, Kamphuis JBJ, Steenkeste K, Bettini S, Robert V, et al. Food-grade TiO₂ is trapped by intestinal mucus in vitro but does not impair mucin O-glycosylation and short-chain fatty acid synthesis in vivo: implications for gut barrier protection. *J Nanobiotechnol*. 2018;16:53.
 62. Li J, Yang S, Lei R, Gu W, Qin Y, Ma S, et al. Oral administration of rutile and anatase TiO₂ nanoparticles shifts mouse gut microbiota structure. *Nanoscale Royal Soc Chem*. 2018;10:7736–45.
 63. Chen H, Zhao R, Wang B, Cai C, Zheng L, Wang H, et al. The effects of orally administered Ag, TiO₂ and SiO₂ nanoparticles on gut microbiota composition and colitis induction in mice. *NanoImpact*. 2017;8:80–8.
 64. Lamas B, Martins Breyner N, Houdeau E. Impacts of foodborne inorganic nanoparticles on the gut microbiota-immune axis: potential consequences for host health. *Part Fibre Toxicol*. 2020;17:19.
 65. Zhang Y, Xu B, Yao M, Dong T, Mao Z, Hang B, et al. Titanium dioxide nanoparticles induce proteostasis disruption and autophagy in human trophoblast cells. *Chem Biol Interact*. 2018;296:124–33.
 66. Mao Z, Yao M, Li Y, Fu Z, Li S, Zhang L, et al. miR-96-5p and miR-101-3p as potential intervention targets to rescue TiO₂ NP-induced autophagy and migration impairment of human trophoblastic cells. *Biomater Sci*. 2018;6:3273–83.
 67. Mao Z, Guan Y, Li T, Zhang L, Liu M, Xing B, et al. Up regulation of miR-96-5p is responsible for TiO₂ NPs induced invasion dysfunction of human trophoblastic cells via disturbing Ezrin mediated cytoskeletons arrangement. *Biomed Pharmacother*. 2019;117:109125.
 68. Corbel T, Gayraud V, Puel S, Lacroix MZ, Berrebi A, Gil S, et al. Bidirectional placental transfer of Bisphenol a and its main metabolite, Bisphenol A-Glucuronide, in the isolated perfused human placenta. *Reprod Toxicol*. 2014;47:51–8.
 69. Delvallée A, Feltin N, Ducourtieux S, Trabelsi M, Hochepied JF. Direct comparison of AFM and SEM measurements on the same set of nanoparticles. *Meas Sci Technol*. 2015;26:085601.
 70. Devoille L, Revel M, Batana C, Feltin N, Giambérini L, Châtel A, et al. Combined influence of oxygenation and salinity on aggregation kinetics of the silver reference nanomaterial NM-300K. *Environ Toxicol Chem*. 2018;37:1007–13.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions



Annexe 3 : Liste des étalons internes utilisés en analyses métabolomiques.

		Non détecté						
		Non détecté ou forme de pic médiocre			UPLC-C18		ZICpHILIC	
Compound	Elemental formula	Exact Mass	EI+	EI-	Pos	Neg	Pos	Neg
					Retention time	Retention time	Retention time	Retention time
					Rt (min)	Rt (min)	Rt (min)	Rt (min)
Alanine ¹³ C	¹³ C ₁ ¹² C ₂ H ₇ N O ₂	90.05103	91.05831		0.88		7.08	
Metformin	C ₄ H ₁₁ N ₅	129.10145	130.10873		0.83		8.12 / 9.35 / 7	
Ethylmalonic acid	C ₅ H ₈ O ₄	132.04226		131.03498		2.82		7.65
Aspartate ¹⁵ N	C ₄ H ₇ ¹⁵ N ₁ O ₄	134.03454	135.04182	133.02726	0.81	0.92	8.06	8.08
Glucose ¹³ C	¹³ C ₁ ¹² C ₅ H ₁₂ O ₆	181.06674	204.05596 (+Na)	216.03614 (+HCl)	0.85	0.93	nd	7.2
2-Aminoanthracene	C ₁₄ H ₁₁ N	193.08915	194.09643		8.05		2.09	
Amiloride	C ₆ H ₈ Cl N ₇ O	229.04789	230.05517	228.04061 and 264.01729 (+HCl)	3.1	3.22	6.5 / 7.03	6.7
Imipramine	C ₁₉ H ₂₄ N ₂	280.19395	281.20123		7.5		1.9	
Atropine	C ₁₇ H ₂₃ N O ₃	289.16779	290.17507		5.2		2 à 3	
Prednisone	C ₂₁ H ₂₆ O ₅	358.17802	359.1853	403.17482 (+HCOOH) and 393.14742 (+HCl)	6.75 and 7.00	6.75 and 7.00	2	1.95
Colchicine	C ₂₂ H ₂₅ N O ₆	399.16819	400.17547		6.44		1.9	
Dihydrostreptomycin	C ₂₁ H ₄₁ N ₇ O ₁₂	583.28132	584.2886 et 292.64794	618.25072 (+HCl)	0.72	0.73	nd	nd
Roxithromycin (fragment)	C ₂₉ H ₅₄ O ₁₀ N ₂	590.37784	591.38512	625.34725 (+HCl)	5.87 and 6.08	5.91 and 6.11	1.86	1.8
Dimetridazole	C ₅ H ₇ N ₃ O ₂	141.05383	142.06111		3.3		2.1	
AMPA 2-amino-3-(3-hydroxy-5-methyl-isoxazol-4-yl)propanoic acid	C ₇ H ₁₀ N ₂ O ₄	186.06406	187.07134	185.05678	1.1	1.1	7.1	7.1
MCPA 2-methyl-4-chlorophenoxyacetic acid	C ₉ H ₉ ClO ₃	200.02402		199.01674		8.05		1.9
Dinoseb	C ₁₀ H ₁₂ N ₂ O ₅	240.07462		239.06734		10.15		1.7

Annexe 4 : Paramètres XCMS version 3.0.0 de W4M utilisé pour le traitement des données métabolomiques dans les conditions HILIC-ESI⁻ (Conditions chapitre I des résultats).

Step 1: Input dataset

PI-PRvsTe-HilicNeg.mzXML.zip

select at runtime

Step 2: xcms.xcmsSet

File(s) from your history containing your chromatograms

Output dataset 'output' from step 1

Scan range option

hide

Extraction method for peaks detection

centWave

Max tolerated ppm m/z deviation in consecutive scans in ppm

10

Min,Max peak width in seconds

15,90

Advanced options

show

Signal/Noise threshold

3

Minimum difference in m/z for peaks with overlapping retention times

0.01

peak limits method

peak limits based on smoothed 2nd derivative (less precise)

Prefilter step for the first phase

3,100

Noise filter

0

Step 3: xcms.group

xset RData file

Output dataset 'xsetRData' from step 2

Method to use for grouping

density

Bandwidth

5.0

Minimum fraction of samples necessary

0.7

Width of overlapping m/z slices

0.01

Advanced options

show

Maximum number of groups to identify in a single m/z slice

50

Get a Peak List

False

Step 4: xcms.retcor

xset RData file

Output dataset 'xsetRData' from step 3

Method to use for retention time correction

peakgroups

Smooth method

loess

Number of extra peaks to allow in retention time correction correction groups

1

Number of missing samples to allow in retention time correction groups

7

Advanced options

show

Degree of smoothing for local polynomial regression fitting

0.2

Family

gaussian

plotype

deviation

Resubmit your raw dataset or your zip file:

Resubmit your dataset or your zip file

no need

Step 5: xcms.group

xset RData file

Output dataset 'xsetRData' from step 4

Method to use for grouping

density

Bandwidth

2.5

Minimum fraction of samples necessary

0.7

Width of overlapping m/z slices

0.01

Advanced options

show

Maximum number of groups to identify in a single m/z slice

50

Get a Peak List

True

Convert retention time (seconds) into minutes

False

Number of decimal places for mass values reported in ions' identifiers.

4

Number of decimal places for retention time values reported in ions' identifiers.

2

Reported intensity values

into

Step 6: xcms.fillPeaks

xset RData file

Output dataset 'xsetRData' from step 5

Filling method

chrom

Get a Peak List

True

Convert retention time (seconds) into minutes

False

Number of decimal places for mass values reported in ions' identifiers.

4

Number of decimal places for retention time values reported in ions' identifiers.

2

Reported intensity values

into

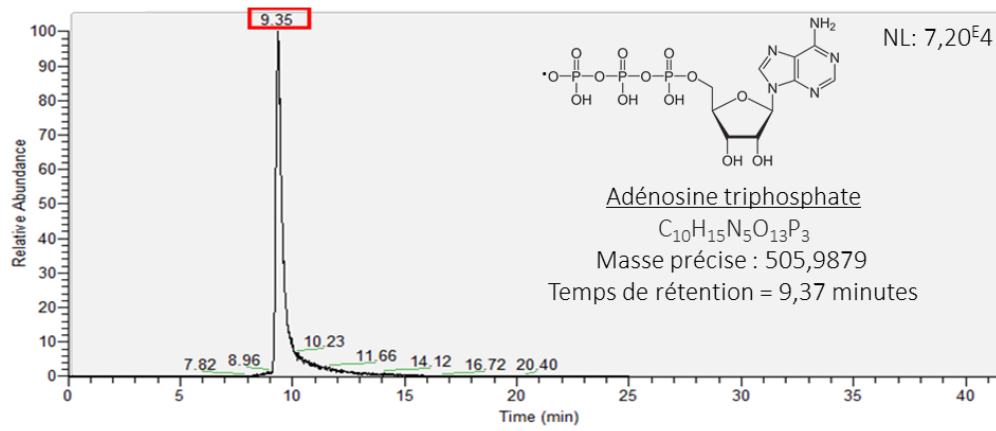
Resubmit your raw dataset or your zip file:

Resubmit your dataset or your zip file

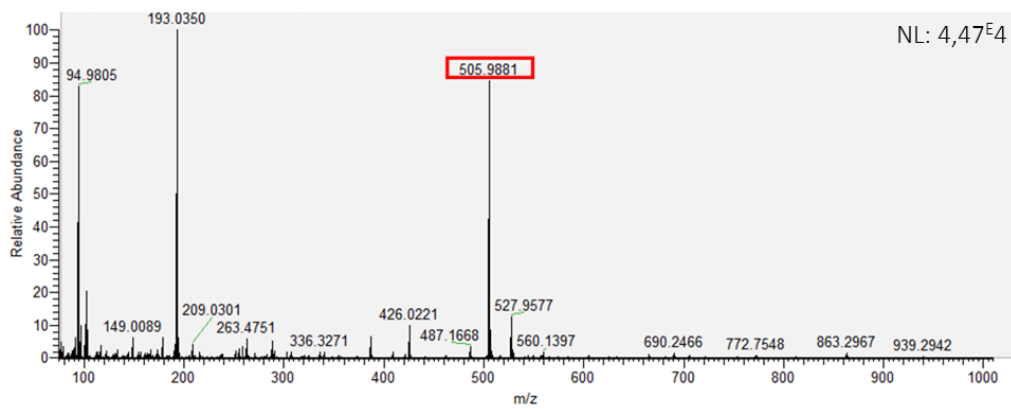
no need

Annexe 5 : Chromatogramme extrait de l'ATP en (A) et spectre de masse correspondant en (B) extraits du composé adénosine triphosphate analysé dans les conditions HILIC-ESI-.

A.



B.



Annexe 6 : Liste des 137 métabolites identifiés au niveau (a, b) et (a, b, d) dans les échantillons de méconium A et B sur la colonne SeQuant ZIC-pHILIC dans le mode d'ionisation négatif (ESI⁻) sur un Q-Exactive.

Les différents critères d'identification fixés sont :

- a. Identification sur la base de la masse précise mesurée
- b. Identification sur la base du temps de rétention
- d. Identification sur la base des spectres de fragmentations MS/MS

Composé identifié	Formule chimique	M	m/z (ESI ⁻)	RT (ZIC-pHILIC)	Statut d'identification	Familles chimiques	Origine
(-)-N-Acetylneuraminic acid	C ₁₁ H ₁₉ N ₁ O ₉	309,106	308,0988	7,3	a,b,d	Carbohydrates and Carbohydrate Conjugates	Endogenous
(S)-3-Hydroxybutyric acid	C ₄ H ₈ O ₃	104,0473	103,0402	2,98	a,b,d	Organic Acids and Derivatives	Endogenous
1,7-Dimethyluric acid	C ₇ H ₈ N ₄ O ₃	196,0596	195,0525	2,71	a,b,d	Aromatic Heteropolycyclic Compounds	Endogenous
1-Methyluric acid	C ₅ H ₁₀ N ₂ O ₃	146,0691	181,0369	4,26	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous
2,5-Furandicarboxylic acid	C ₆ H ₄ O ₅	156,0053	154,9988	9,91	a,b,d	Aromatic Heteromonocyclic Compounds	Endogenous
2-cis,4-trans-Abscisic acid	C ₁₅ H ₂₀ O ₄	264,1362	263,1308	1,38	a,b,d	Lipids	Endogenous

2-Hydroxy 3-methylbutyric acid (α -Hydroxyisovaleric acid)	C5H10O3	118,063	117,0558	2,31	a,b,d	Lipids	Endogenous
2-Hydroxyhexadecanoic acid (2-Hydroxypalmitic acid)	C16H32O3	272,2351	271,2284	1,24	a,b,d	Lipids	Endogenous
2-Isopropylmalic acid	C7H12O5	176,0685	175,0614	7,49	a,b,d	Organic Acids and Derivatives	Microbial
2-O-Methylinosine	C11H14N4O5	282,0964	281,0896	2,65	a,b,d	Nucleosides, Nucleotides, and Analogues	Endogenous
3-Hydroxy-3-methylglutaric acid	C6H10O5	162,0528	161,0459	9,01	a,b,d	Organic Acids and Derivatives	Endogenous
3-Methylglutaric acid	C6H10O4	146,0579	145,0508	7,65	a,b,d	Lipids	Endogenous
4-Acetamidobutanoic acid	C6H11NO3	145,0739	144,0668	3,39	a,b,d	Lipids	Endogenous
4-Hydroxy-3-methoxyphenylglycol sulfate	C9H12O7S	264,0304	263,0236	3,44	a,b,d	Aromatic Homomonocyclic Compounds	Endogenous
4-Hydroxybenzoic acid	C7H6O3	138,0317	137,0246	4,94	a,b,d	Aromatic Homomonocyclic Compounds	Endogenous
4-Imidazoleacrylic acid (urocanic acid)	C6H6N2O2	138,0429	137,0358	5,01	a,b,d	Aromatic Heteromonocyclic Compounds	Endogenous
4-Methyl-2-oxovaleric acid	C6H10O3	130,063	129,0558	1,58	a,b,d	Lipids	Endogenous

4-Pyridoxic acid	C8H9NO4	183,0532	182,0461	1,61	a,b,d	Aromatic Heteromonocyclic Compounds	Endogenous
5-Hydroxylysine	C6H14N2O3	162,1004	161,0933	8,66	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous
5-Sulfosalicylic acid	C7H6O6S	217,9885	215,1151	9,04	a,b,d	NA	Endogenous
6-Hydroxypyridine-3-carboxylic acid (6-hydroxynicotinic acid)	C6H5NO3	139,0269	138,0198	6,17	a,b,d	Aromatic Heteromonocyclic Compounds	Endogenous
7-methylguanine	C6H7N5O	165,0651	164,0582	3,44	a,b,d	Aromatic Heteropolycyclic Compounds	Endogenous
7-Methylxanthine	C6H6N4O2	166,0491	165,042	2,95	a,b,d	Aromatic Heteropolycyclic Compounds	Endogenous
Acetaminophen-glucuronide	C14H17NO8	327,0954	326,0886	5,77	a,b,d	Carbohydrates and Carbohydrate Conjugates	Drug metabolite
Adenine	C5H5N5	135,0545	134,0474	3,21	a,b,d	Aromatic Heteropolycyclic Compounds	Endogenous
Adenosine	C10H13N5O4	267,0967	266,0906	5,2	a,b,d	Nucleosides, Nucleotides, and Analogues	Endogenous
Allantoin	C4H6N4O3	158,044	157,0369	7,41	a,b,d	Aliphatic Heteromonocyclic Compounds	Endogenous

Alpha-D-Aminoadipic acid	C6H11NO4	161,0688	142,0511	3,07	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous
alpha-Hydroxyhippuric-acid	C9H9NO4	195,0526	194,0462	3,69	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous
Amylose	C6H10O5	162,0528	204,0668	8,51	a,b,d	Carbohydrates and Carbohydrate Conjugates	Endogenous
Androsterone sulphate	C19H30O5S	370,1814	369,1743	1,29	a,b,d	Lipids	Endogenous
Arginine	C6H14N4O2	174,1117	173,1046	9,67	a,b,d	Amino Acids, Peptides, and Analogues	essential amino-acid
Asparagine	C4H8N2O3	132,0535	131,0463	7,12	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous
Aspartylglycosamine (N-As)	C12H21N3O8	335,1329	334,1261	7,93	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous
Azelaic acid	C9H16O4	188,1049	187,0978	6,05	a,b,d	Organic Acids and Derivatives	Endogenous
Behenic acid	C22H44O2	340,3341	339,327	1,2	a,b,d	Lipids	Endogenous
beta-estradiol-17beta-glucuronide	C24H32O8	448,2097	447,206	1,5	a,b,d	Lipids	Endogenous
beta-Pseudouridine	C9H12N2O6	244,0695	243,0628	5,9	a,b,d	Carbohydrates and Carbohydrate Conjugates	Endogenous
Capric acid	C10H20O2	172,1463	171,1392	1,32	a,b	Lipids	Endogenous
Cholic acid	C24H40O5	408,2876	407,2806	1,52	a,b,d	Lipids	Endogenous

cis-5,8,11,14,17-Eicosapentaenoic acid	C20H30O2	302,2246	301,2174	1,23	a,b,d	Lipids	Food
Creatine	C4H9N3O2	131,0695	130,0623	7,43	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous
Cytosine	C4H5N3O	111,0433	110,0361	3,31	a,b,d	Aromatic Heteromonocyclic Compounds	Endogenous
D-(-)-Quinic acid	C7H12O6	192,0634	191,0565	7,47	a,b,d	Aliphatic Homomonocyclic Compounds	Food
D-(-)-Ribose	C5H10O5	150,0528	149,0457	4,52	a,b,d	Carbohydrates and Carbohydrate Conjugates	Endogenous
D-(+)-Arabitol	C5H12O5	152,0685	151,0614	6,18	a,b,d	Carbohydrates and Carbohydrate Conjugates	Endogenous
D-(+)-Gluconic acid D-lactone (Gluconolactone)	C6H10O6	178,0477	177,0407	7,85	a,b,d	Carbohydrates and Carbohydrate Conjugates	Endogenous
D-(+)-Glucuronic acid gamma-lactone	C6H8O6	176,0321	131,035	8,75	a,b,d	Aliphatic Heteropolycyclic Compounds	Endogenous
D-(+)-Neopterin	C9H11N5O4	253,0811	252,073	7,98	a,b,d	Aromatic Heteropolycyclic Compounds	Endogenous
D-2-Hydroxyglutaric acid	C5H8O5	148,0372	147,03	8,33	a,b,d	Organic Acids and Derivatives	Endogenous

d-Desthiobiotin	C10H18N2O3	214,1317	213,1248	2,14	a,b,d	Lipids	Endogenous
Dehydroisoandrosterone 3-sulfate	C19H28O5S	368,1657	367,1586	1,25	a,b,d	Lipids	Endogenous
Deoxycorticosterone-21-glucoside	C27H40O8	492,2723	491,2652	1,45	a,b,d	Lipids	Endogenous
Deoxycytidine	C9H13N3O4	227,0906	226,0837	3,85	a,b,d	Nucleosides, Nucleotides, and Analogues	Endogenous
D-Glucuronic acid	C6H10O7	194,0426	193,0354	8,75	a,b,d	Carbohydrates and Carbohydrate Conjugates	Endogenous
D-Glyceric acid	C3H6O4	106,0266	105,0194	6,08	a,b,d	Carbohydrates and Carbohydrate Conjugates	Endogenous
Diglycolic acid	C4H6O5	134,0215	133,0143	8,61	a,b,d	NA	NA
DL-Tryptophan	C11H12N2O2	204,0899	203,0566	3,86	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous
DL-vanillomandelic acid	C9H10O5	198,0528	195,0525	3,94	a,b,d	Aromatic Homomonocyclic Compounds	Endogenous
D-Mannitol	C6H14O6	182,079	181,0719	7	a,b,d	Lipids	Endogenous
D-Saccharic acid	C6H10O8	210,0376	207,0512	9,21	a,b,d	Carbohydrates and Carbohydrate Conjugates	Endogenous
D-Threitol	C4H10O4	122,0579	157,0257	5,23	a,b,d	Carbohydrates and Carbohydrate Conjugates	Endogenous
Ehtylmalonic acid	C5H8O4	132,0423	131,0351	8,66	a,b,d	Lipids	Endogenous
Estriol 17-sulphate	C18H24O6S	368,1294	367,125	1,61	a,b,d	Lipids	Endogenous

Estrone glucuronide	C24H30O8	446,1941	445,1903	1,49	a,b,d	Lipids	Endogenous
Eugenol (pdt cosmétique)	C10H12O2	164,0837	163,0766	1,42	a,b,d	Aromatic Homomonocyclic Compounds	Cosmetic
Gluconic acid	C6H12O7	196,0583	195,0509	7,63	a,b,d	Carbohydrates and Carbohydrate Conjugates	Endogenous
Glucose-6-PO ₄	C6H13O9P	260,0297	213,1862	6,36	a,b	Carbohydrates and Carbohydrate Conjugates	Endogenous
Glycerol 3-phosphate	C3H9O6P	172,0137	171,0067	7,63	a,b,d	Carbohydrates and Carbohydrate Conjugates	Endogenous
Glycocholic acid	C26H43NO6	465,3091	464,3018	1,48	a,b,d	Lipids	Endogenous
Gly-Pro	C7H12N2O3	172,0848	171,0777	6,17	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous
Guanidinosuccinic acid	C5H9N3O4	175,0593	174,0523	8,47	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous
Guanosine	C10H13N5O5	283,0917	282,0848	6,11	a,b,d	Nucleosides, Nucleotides, and Analogues	Endogenous
Histidine	C6H9N3O2	155,0695	154,0624	7,47	a,b,d	Amino Acids, Peptides, and Analogues	essential amino-acid
Imidazoleacetic acid	C5H6N2O2	126,0429	125,0356	5,66	a,b,d	Aromatic Heteromonocyclic Compounds	Endogenous

Indoleacetic acid	C10H9NO2	175,0633	174,0563	3,2	a,b,d	Aromatic Heteropolycyclic Compounds	Endogenous
Indoleacrylic acid	C11H9NO2	187,0633	186,0563	2,67	a,b,d	Aromatic Heteropolycyclic Compounds	Endogenous
Isocitric-acid	C6H8O7	192,027	191,02	9,91	a,b,d	Organic Acids and Derivatives	Endogenous
Isoleucine	C6H13NO2	131,0946	130,0875	3,43	a,b,d	Amino Acids, Peptides, and Analogues	essential amino-acid
Kynurenic acid	C10H7NO3	189,0426	188,0356	2,15	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous
L(+)-arabinose	C5H10O5	150,0528	149,0457	6,81	a,b,d	Carbohydrates and Carbohydrate Conjugates	Endogenous
L-(+)-Tartaric acid	C4H6O6	150,0164	149,0093	9,24	a,b,d	Organic Acids and Derivatives	Food
L-Alanyl-L-proline	C8H14N2O3	186,1004	185,0935	3,06	a,b,d	Carboxylic acids and derivatives	Endogenous
L-Aspartic-acid	C4H7NO4	133,0375	132,0304	9,11	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous
L-Citrulline	C6H13N3O3	175,0957	174,0886	7,64	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous
L-Cysteine S-sulfate	C3H7NO5S2	200,9766	197,0459	9,49	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous

L-Cystine	C6H12N2O4S2	240,0239	239,017	8,33	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous
Leu-Pro	C11H20N2O3	228,1474	227,1406	1,62	a,b,d	Carboxylic acids and derivatives	Endogenous
L-Glutamic acid	C5H9NO4	147,0532	146,046	10,62	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous
L-Glutamine	C5H10N2O3	146,0691	145,062	7,09	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous
L-Pyroglutamic-acid	C5H7NO3	129,0426	128,0355	4,66	a,b,d	Aliphatic Heteromonocyclic Compounds	Endogenous
L-Saccharopine	C11H20N2O6	276,1321	275,1255	8,62	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous
Lysine	C6H14N2O2	146,1055	145,0984	9,21	a,b,d	Amino Acids, Peptides, and Analogues	essential amino-acid
Maleic-acid	C4H4O4	116,011	115,0037	8,72	a,b,d	Lipids	Endogenous
Malonic acid	C3H4O4	104,011	103,0037	8,73	a,b,d	Organic Acids and Derivatives	Endogenous
Maltotetraose	C24H42O21	666,2219	665,2151	9,49	a,b,d	Carbohydrates and Carbohydrate Conjugates	Endogenous
Maltotriose	C18H32O16	504,169	503,1622	8,47	a,b,d	Carbohydrates and Carbohydrate Conjugates	Endogenous

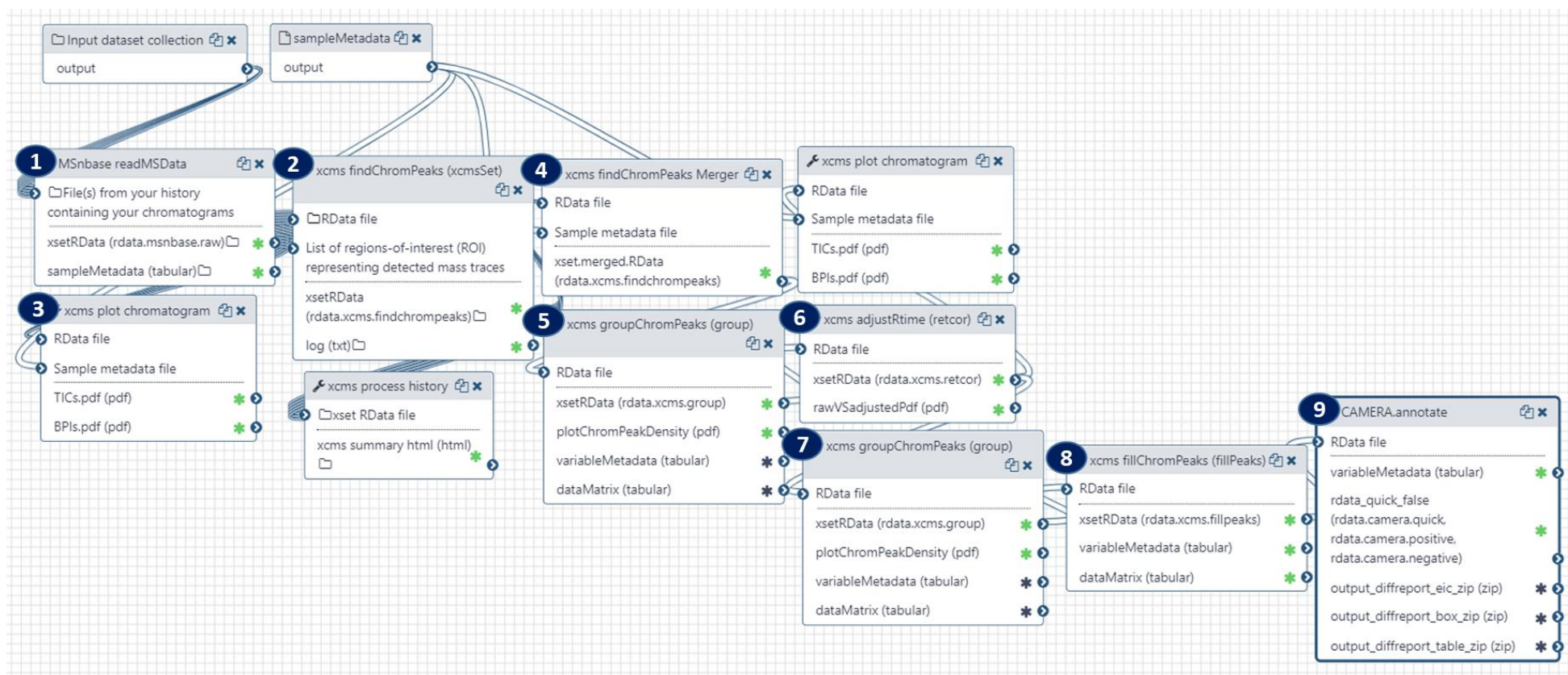
Medroxyprogesterone	C22H32O3	344,2351	343,228	1,24	a,b,d	Lipids	Drug
Mesaconic acid/Citraconic acid	C5H6O4	130,0266	129,0195	9,86	a,b,d	Lipids	Endogenous
Methionine	C5H11NO2S	149,0511	148,044	3,68	a,b,d	Amino Acids, Peptides, and Analogues	essential amino-acid
Methyl-succinic acid/mono-Methyl hydrogen succinate	C5H8O4	132,0423	131,0351	2,34	a,b,d	Organic Acids and Derivatives	Endogenous
Mevalonic acid	C6H12O4	148,0736	147,0664	2,51	a,b,d	Organic Acids and Derivatives	Endogenous
Muramic acid	C9H17NO7	251,1005	250,0937	7,63	a,b,d	Carbohydrates and Carbohydrate Conjugates	Microbial
N acetyl L glutamine	C7H12N2O4	188,0797	187,0728	4,99	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous
N-Acetyl-Asp-Glu	C11H16N2O8	304,0907	303,084	9,11	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous
N-Acetylgalactosamine	C8H15NO6	221,0899	199,9696	9,39	a,b,d	Carbohydrates and Carbohydrate Conjugates	Endogenous
N-Acetylglycine	C4H7NO3	117,0426	116,0355	5,81	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous
N-Acetyl-L-alanine	C5H9NO3	131,0582	130,051	3,53	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous
N-Acetyl-L-aspartic acid	C6H9NO5	175,0481	174,041	8,11	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous

N-Acetyl-L-glutamic acid	C7H11NO5	189,0637	188,0568	7,87	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous
Nervonic acid	C24H46O2	366,3498	365,3426	1,18	a,b	Lipids	Endogenous
N α -Acetyl-L-asparagine	C6H10N2O4	174,0641	173,0571	8,02	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous
Ornithine	C5H12N2O2	132,0899	131,0827	9,57	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous
Orotic acid	C5H4N2O4	156,0171	155,0101	5,32	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous
Orotidine	C10H12N2O8	288,0594	287,0527	7,43	a,b,d	Nucleosides, Nucleotides, and Analogues	Endogenous
Pantothenic acid	C9H17NO5	219,1107	216,9816	3,26	a,b,d	Aliphatic Acyclic Compounds	Endogenous
Phenylacetyl-L-glutamine	C13H16N2O4	264,111	263,1041	2,18	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous
Phenylalanine	C9H11NO2	165,079	164,072	2,82	a,b,d	Amino Acids, Peptides, and Analogues	essential amino-acid
Pregnenolone sulfate	C21H32O5S	396,197	395,1899	1,23	a,b,d	Lipids	Endogenous
Proline	C5H9NO2	115,0633	114,0562	6,02	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous

Pyridoxine	C8H11NO3	169,0739	168,0669	3,7	a,b,d	Aromatic Heteromonocyclic Compounds	Endogenous
Quinolinic acid	C7H5NO4	167,0219	166,0147	8,31	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous
Salicylamide	C7H7NO2	137,0477	136,0404	1,54	a,b,d	Aromatic Homomonocyclic Compounds	Drug
Serine	C3H7NO3	105,0426	104,0354	7,25	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous
Suberic acid	C8H14O4	174,0892	173,0821	6,82	a,b,d	Organic Acids and Derivatives	Endogenous
Succinic semialdehyde	C4H6O3	102,0317	101,0245	5,35	a,b,d	Lipids	Endogenous
Taurine	C2H7NO3S	125,0147	124,0074	8,19	a,b,d	Organic Acids and Derivatives	Endogenous
Taurolithocholic acid 3-sulfate	C26H45NO8S2	563,2587	562,2515	1,28	a,b,d	Lipids	Endogenous
Threonic acid	C4H8O5	136,0372	135,0301	6,77	a,b,d	Organooxygen compounds	Endogenous
Thymine	C5H6N2O2	126,0429	125,0357	2,21	a,b,d	Aromatic Heteromonocyclic Compounds	Endogenous

trans-4-hydroxy-L-proline	C5H9NO3	131,0582	130,0511	8,1	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous
Tridecanoic acid	C13H26O2	214,1933	213,1248	1,28	a,b	Lipids	Endogenous
Tyrosine	C9H11NO3	181,0739	180,0667	5,37	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous
Uracil	C4H4N2O2	112,0273	111,0201	2,67	a,b	Aromatic Heteromonocyclic Compounds	Endogenous
Xanthine	C5H4N4O2	152,0334	151,0264	5,81	a,b,d	Aromatic Heteropolycyclic Compounds	Endogenous
Xanthurenic-acid	C10H7NO4	205,0375	203,083	7,66	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous
α -Ketoglutaric acid	C5H6O5	146,0215	145,0145	8,81	a,b,d	Organic Acids and Derivatives	Endogenous

Annexe 7 : « Workflow » réunissant l'ensemble des procédures réalisées lors de l'extraction automatique des signaux (l'étape du XCMS).



- 1- L'étape 1 permet la lecture des spectres.
- 2- L'étape du 'xcmsSet' permet d'extraire les pics.
- 3- Le 'plotting' à l'étape 3 permet de tracer les BPI et TIC correspondants respectivement aux tracés des intensités des pics de base et des courants ioniques totaux chez tous les échantillons.
- 4- L'étape du 'merge' permet de fusionner les pics extraits à l'étape 2 dans un seul fichier facilitant l'étape du 'grouping' qui suit.

- 5- L'étape 5 du *'group'* comme son nom l'indique permet de grouper les ions par leur masse précise m/z et par leur temps de rétention en considérant les tolérances attribuées.
- 6- L'algorithme *'retcor'* à l'étape 6 permet de corriger et d'aligner les ions au sein des différents échantillons en se basant sur leur temps de rétention.
- 7- Suite à l'étape d'alignement, de nouveaux groupes d'ions peuvent être considérés et donc une nouvelle étape de *'grouping'* est réalisée.
- 8- Tous les ions à cette étape ne sont pas détectés au sein de tous les échantillons, pour cela une étape de *'fillPeaks'* est réalisée. Cette étape permet de rechercher et de réintégrer différents signaux non détectés dans certains échantillons.
- 9- Enfin, l'analyse XCMS se finit par une étape d'annotation à l'aide de l'algorithme *'camera'*. Celui-ci permet d'annoter les ions mono-isotopiques, les adduits ainsi que les fragments en questionnant différentes bases de données aussi bien internes que externes (HMDB, KEGG, Chempider, Massbank...). Cet algorithme reste néanmoins non-utilisé pour cette étude.

Pour l'étape d'annotation, un outil développé sous R au laboratoire est utilisé. Cet outil développé sur *shiny_app* permet d'intervenir sur davantage de paramètres d'annotations, permettant d'effectuer simultanément des analyses statistiques et d'avoir des graphiques interactifs. Les différents paramètres de tolérance en temps de rétention utilisés sont à +/- 90 et +/- 15 secondes pour les colonnes HILIC et C18 respectivement et une tolérance en masse fixée à 10 ppm dans les deux modes d'ionisations. Enfin un filtre retenant que les variables ayant des aires supérieures à 30 000 et 10 000 respectivement pour les deux conditions chromatographiques HILIC et C18.

Annexe 8 : Paramètres XCMS version 3.2.1 de W4M utilisée pour le traitement des données métabolomiques dans les conditions HILIC-ESI⁻ (conditions chapitres II et III des résultats).

Step 1: Input dataset collection

Raw data

select at runtime

Step 2: Input dataset

SampleMetadata

select at runtime

Step 3: MSnbase readMSData

File(s) from your history containing your chromatograms

Output dataset 'output' from step 1

Step 4: xcms findChromPeaks (xcmsSet)

RData file

Output dataset 'xsetRData' from step 3

Spectra Filters:

Filter on Acquisition Numbers

Empty.

Filter on Retention Time

Empty.

Filter on Mz

Empty.

Extraction method for peaks detection

CentWave - chromatographic peak detection using the centWave method

Max tolerated ppm m/z deviation in consecutive scans in ppm

10

Min,Max peak width in seconds

15,90

Advanced Options:

Signal to Noise ratio cutoff

10

Prefilter step for for the first analysis step (ROI detection)

3,100

Name of the function to calculate the m/z center of the chromatographic peak

intensity weighted mean of the peak's m/z values

Integration method

peak limits are found through descent on the mexican hat filtered data (more robust, but less exact)

Minimum difference in m/z for peaks with overlapping retention times

-0.001

fitgauss

False

Noise filter

0

verbose Columns

False

List of regions-of-interest (ROI):

List of regions-of-interest (ROI) representing detected mass traces

select at runtime

Is continuous data within regions of interest is checked to be above the first baseline.

True

Numeric vector defining the scale for each region of interest in 'roiList'

Empty.

Resubmit your raw dataset or your zip file:

Resubmit your dataset or your zip file

no need

Step 5: xcms plot chromatogram

RData file

Output dataset 'xsetRData' from step 3

Sample metadata file

Output dataset 'output' from step 2

Resubmit your raw dataset or your zip file:

Resubmit your dataset or your zip file

no need

Step 6: xcms findChromPeaks Merger

RData file

Output dataset 'xsetRData' from step 4

Sample metadata file

Output dataset 'output' from step 2

Resubmit your raw dataset or your zip file:

Resubmit your dataset or your zip file

no need

Step 7: xcms groupChromPeaks (group)

RData file

Output dataset 'xsetRData' from step 6

Method to use for grouping

PeakDensity - peak grouping based on time dimension peak densities

Bandwidth

5.0

Minimum fraction of samples

0.7

Minimum number of samples

3

Width of overlapping m/z slices

0.01

Advanced Options:

Maximum number of groups to identify in a single m/z slice

50

Get the Peak List

True

Convert retention time (seconds) into minutes

False

Number of decimal places for mass values reported in ions' identifiers.

4

Number of decimal places for retention time values reported in ions' identifiers.

2

Reported intensity values

into

Replace the remain NA by 0 in the dataMatrix

True

Resubmit your raw dataset or your zip file:

Resubmit your dataset or your zip file

no need

Step 8: xcms adjustRtime (retcor)

RData file

Output dataset 'xsetRData' from step 7

Method to use for retention time correction

PeakGroups - retention time correction based on alignment of features (peak groups) present in most/all samples.

Minimum required fraction of samples in which peaks for the peak group were identified

0.9

Maximal number of additional peaks for all samples to be assigned to a peak group for retention time correction

1

Smooth method

loess - non-linear alignment

Advanced Options:

Degree of smoothing for the loess fitting

0.2

Family

gaussian

Resubmit your raw dataset or your zip file:

Resubmit your dataset or your zip file

no need

Step 9: xcms plot chromatogram

RData file

Output dataset 'xsetRData' from step 8

Sample metadata file

Output dataset 'output' from step 2

Resubmit your raw dataset or your zip file:

Resubmit your dataset or your zip file

no need

Step 10: xcms groupChromPeaks (group)

RData file

Output dataset 'xsetRData' from step 8

Method to use for grouping

PeakDensity - peak grouping based on time dimension peak densities

Bandwidth

5.0

Minimum fraction of samples

0.7

Minimum number of samples

3

Width of overlapping m/z slices

0.01

Advanced Options:

Maximum number of groups to identify in a single m/z slice

50

Get the Peak List

True

Convert retention time (seconds) into minutes

False

Number of decimal places for mass values reported in ions' identifiers.

4

Number of decimal places for retention time values reported in ions' identifiers.

2

Reported intensity values

into

Replace the remain NA by 0 in the dataMatrix

True

Resubmit your raw dataset or your zip file:

Resubmit your dataset or your zip file

no need

Step 11: xcms fillChromPeaks (fillPeaks)

RData file

Output dataset 'xsetRData' from step 10

Advanced Options:

Value by which the m/z width of peaks should be expanded

0

Value by which the RT width of peaks should be expanded

0

Specifying a ppm by which the m/z width of the peak region should be expanded

5

Peak List:

None

True

Convert retention time (seconds) into minutes

False

Number of decimal places for mass values reported in ions' identifiers.

4

Number of decimal places for retention time values reported in ions' identifiers.

2

Reported intensity values

into

Replace the remain NA by 0 in the dataMatrix

True

Resubmit your raw dataset or your zip file:

Resubmit your dataset or your zip file

no need

Annexe 9 : Paramètres XCMS version 3.2.1 de W4M utilisée pour le traitement des données métabolomiques dans les conditions C18-ESI⁺ (conditions chapitres II et III des résultats).

Step 1: Input dataset collection

Raw data

select at runtime

Step 2: Input dataset

SampleMetadata

select at runtime

Step 3: MSnbase readMSData

File(s) from your history containing your chromatograms

Output dataset 'output' from step 1

Step 4: xcms findChromPeaks (xcmsSet)

RData file

Output dataset 'xsetRData' from step 3

Spectra Filters:

Filter on Acquisition Numbers

Empty.

Filter on Retention Time

0,1500

Filter on Mz

Empty.

Extraction method for peaks detection

CentWave - chromatographic peak detection using the centWave method

Max tolerated ppm m/z deviation in consecutive scans in ppm

10

Min,Max peak width in seconds

10,40

Advanced Options:**Signal to Noise ratio cutoff**

3

Prefilter step for for the first analysis step (ROI detection)

3,100

Name of the function to calculate the m/z center of the chromatographic peak

intensity weighted mean of the peak's m/z values

Integration method

peak limits are found through descent on the mexican hat filtered data (more robust, but less exact)

Minimum difference in m/z for peaks with overlapping retention times

0.01

fitgauss

False

Noise filter

0

verbose Columns

False

List of regions-of-interest (ROI):**List of regions-of-interest (ROI) representing detected mass traces**

select at runtime

Is continuous data within regions of interest is checked to be above the first baseline.

True

Numeric vector defining the scale for each region of interest in 'roiList'

Empty.

Resubmit your raw dataset or your zip file:

Resubmit your dataset or your zip file

no need

Step 5: xcms plot chromatogram

RData file

Output dataset 'xsetRData' from step 3

Sample metadata file

Output dataset 'output' from step 2

Resubmit your raw dataset or your zip file:

Resubmit your dataset or your zip file

no need

Step 6: xcms findChromPeaks Merger

RData file

Output dataset 'xsetRData' from step 4

Sample metadata file

Output dataset 'output' from step 2

Resubmit your raw dataset or your zip file:

Resubmit your dataset or your zip file

no need

Step 7: xcms groupChromPeaks (group)

RData file

Output dataset 'xsetRData' from step 6

Method to use for grouping

PeakDensity - peak grouping based on time dimension peak densities

Bandwidth

5.0

Minimum fraction of samples

0.7

Minimum number of samples

3

Width of overlapping m/z slices

0.01

Advanced Options:

Maximum number of groups to identify in a single m/z slice

50

Get the Peak List

True

Convert retention time (seconds) into minutes

False

Number of decimal places for mass values reported in ions' identifiers.

4

Number of decimal places for retention time values reported in ions' identifiers.

2

Reported intensity values

into

Replace the remain NA by 0 in the dataMatrix

True

Resubmit your raw dataset or your zip file:

Resubmit your dataset or your zip file

no need

Step 8: xcms adjustRtime (retcor)

RData file

Output dataset 'xsetRData' from step 7

Method to use for retention time correction

PeakGroups - retention time correction based on alignment of features (peak groups) present in most/all samples.

Minimum required fraction of samples in which peaks for the peak group were identified

0.9

Maximal number of additional peaks for all samples to be assigned to a peak group for retention time correction

1

Smooth method

loess - non-linear alignment

Advanced Options:

Degree of smoothing for the loess fitting

0.2

Family

gaussian

Resubmit your raw dataset or your zip file:

Resubmit your dataset or your zip file

no need

Step 9: xcms plot chromatogram

RData file

Output dataset 'xsetRData' from step 8

Sample metadata file

Output dataset 'output' from step 2

Resubmit your raw dataset or your zip file:

Resubmit your dataset or your zip file

no need

Step 10: xcms groupChromPeaks (group)

RData file

Output dataset 'xsetRData' from step 8

Method to use for grouping

PeakDensity - peak grouping based on time dimension peak densities

Bandwidth

5.0

Minimum fraction of samples

0.7

Minimum number of samples

3

Width of overlapping m/z slices

0.01

Advanced Options:

Maximum number of groups to identify in a single m/z slice

50

Get the Peak List

True

Convert retention time (seconds) into minutes

False

Number of decimal places for mass values reported in ions' identifiers.

4

Number of decimal places for retention time values reported in ions' identifiers.

2

Reported intensity values

into

Replace the remain NA by 0 in the dataMatrix

True

Resubmit your raw dataset or your zip file:

Resubmit your dataset or your zip file

no need

Step 11: xcms fillChromPeaks (fillPeaks)

RData file

Output dataset 'xsetRData' from step 10

Advanced Options:

Value by which the mz width of peaks should be expanded

0

Value by which the RT width of peaks should be expanded

0

Specifying a ppm by which the mz width of the peak region should be expanded

5

Peak List:

None

True

Convert retention time (seconds) into minutes

False

Number of decimal places for mass values reported in ions' identifiers.

4

Number of decimal places for retention time values reported in ions' identifiers.

2

Reported intensity values

into

Replace the remain NA by 0 in the dataMatrix

True

Resubmit your raw dataset or your zip file:

Resubmit your dataset or your zip file

no need

Annexe 10 : Scripts utilisés pour les analyses en composantes principales et pour les classifications hiérarchiques réalisées en protéomique.

```
1 import pandas as pd
2 from sklearn.decomposition import PCA
3 from sklearn.preprocessing import StandardScaler
4
5 # Read the data as a dataframe using pandas
6 df = pd.read_excel('my_data.xlsx')
7
8 # Create the features dataframe for PCA
9 features_PCA = df.copy()
10
11 # Scale the data
12 features_PCA = StandardScaler().fit_transform(features_PCA)
13
14 # Run PCA
15 pca = PCA(n_components=3)
16 principalComponents = pca.fit_transform(features_PCA)
```

```
1 import pandas as pd
2 import numpy as np
3 import seaborn as sns
4
5 # Read the data as a dataframe using pandas
6 df = pd.read_excel('my_data.xlsx')
7
8 # Plot the heatmap with the clusters
9 sns.clustermap(df, cmap='Oranges')
10
11 # Take the indices with the top 30 variance
12 indices_30 = np.std(df).sort_values(ascending=False)[:30].index
13 # Plot the heatmap with the clusters for the sub-sample
14 sns.clustermap(df.loc[indices_30], cmap='Oranges')
15
```


Annexe 11 : Liste des 149 métabolites identifiés au niveau (a,b), (a,c), (a,b,d) et (a,c,e) sur les échantillons de méconium de la cohorte clinique à partir des colonnes SeQuant ZIC-pHILIC et Hypersil GOLD C18 en mode d'ionisation négatif et positif respectivement sur un Q-Exactive

Les différents critères d'identification fixés sont :

- Identification sur la base de la masse précise mesurée
- Identification sur la base du temps de rétention sur la colonne ZIC-pHILIC
- Identification sur la base du temps de rétention sur la colonne RP-C18
- Identification sur la base des spectres de fragmentation MS/MS en mode d'ionisation négatif (ESI-)
- Identification sur la base des spectres de fragmentation MS/MS MS en mode d'ionisation négatif (ESI+)

Composés identifiés	Formules chimiques	M	m/z (ESI-)	m/z (ESI+)	RT (ZIC-pHILIC)	RT (RP-C18)	Niveau d'identification	Familles chimiques	Origines
(-)-N-Acetylneuraminic acid	C ₁₁ H ₁₉ NO ₉	309,106	308,0988	310,1121	7,3	0,78	a,c,e	Carbohydrates and Carbohydrate Conjugates	Endogenous
(±)-Salsolinol	C ₁₀ H ₁₃ NO ₂	179,0946	nd	180,1014	nd	1,02	a,c	Aromatic Heteropolycyclic Compounds	Endogenous
1,3-Dimethyluric acid	C ₇ H ₈ N ₄ O ₃	196,0596	nd	197,0663	nd	2,04	a,c,e	Aromatic Heteropolycyclic Compounds	Endogenous
17a-Hydroxypregnenolone	C ₂₁ H ₃₂ O ₃	332,2351	331,228	nd	1,33	nd	a,b,d	Lipids	Endogenous
1-Methylnicotinamide	C ₇ H ₉ N ₂ O	136,0637	nd	137,0705	nd	0,74	a,c	Aromatic Heteromonocyclic Compounds	Drug

2,2'-Thiodiacetic acid	C4H6O4S	149,9981	148,9915	nd	9,33	nd	a,b,d	Organic Acids and Derivatives	Endogenous
2,5-Furandicarboxylic acid	C6H4O5	156,0053	154,9988	157,0128	9,91	1,64	a,c,e	Aromatic Heteromonocyclic Compounds	Endogenous
2-Isopropylmalic acid	C7H12O5	176,0685	175,0614	199,0596	7,49	4,43	a,c	Organic Acids and Derivatives	Microbial
2-O-Methylinosine	C11H14N4O5	282,0964	281,0896	283,1027	2,65	1,53	a,c,e	Nucleosides, Nucleotides, and Analogues	Endogenous
2-Oxovaleric acid	C5H8O3	116,0473	115,0401	nd	1,85	nd	a,b,d	Organic Acids and Derivatives	Endogenous
3,7-Dimethyluric acid	C7H8N4O3	196,0596	nd	197,0663	nd	1,41	a,c,e	Aromatic Heteropolycyclic Compounds	Endogenous
3-Hydroxy-3-methylglutaric acid	C6H10O5	162,0528	161,0459	145,049	9,01	1,24	a,c	Organic Acids and Derivatives	Endogenous
4-Hydroxymandelic acid	C8H8O4	168,0423	167,0352	nd	4,89	nd	a,b,d	Aromatic Homomonocyclic Compounds	Endogenous
4-Imidazoleacrylic-acid-(urocanic-acid)	C6H6N2O2	138,0429	nd	139,0498	nd	0,83	a,c,e	Aromatic Heteromonocyclic Compounds	Endogenous
4-Pyridoxic acid	C8H9NO4	183,0532	182,0461	184,0597	1,61	1,26	a,c,e	Aromatic Heteromonocyclic Compounds	Endogenous
5-Aminolevulinic acid	C5H9NO3	131,0582	nd	114,0546	nd	1,02	a,c	Amino Acids, Peptides, and Analogues	Endogenous
5-Aminovaleric acid	C5H11NO2	117,079	nd	118,0858	nd	0,85	a,c,e	Amino Acids, Peptides, and Analogues	Endogenous

5-Deoxy-5-(methylthio)adenosine	C11H15N5O3S	297,0896	nd	298,0959	nd	4,27	a,c,e	Nucleosides, Nucleotides, and Analogues	Endogenous
5-Hydroxylysine	C6H14N2O3	162,1004	161,0933	163,1071	8,66	0,68	a,c,e	Amino Acids, Peptides, and Analogues	Endogenous
7 α -Hydroxytestosterone	C19H28O3	304,2038	nd	305,2099	nd	6,91	a,c,e	Lipids	Drug
Acetaminophen-(4-Acetamidophenol)	C8H9NO2	151,0633	150,0562	152,0701	1,76	2,59	a,b,c,d,e	Aromatic Homomonocyclic Compounds	Drug
Acetaminophen-glucuronide	C14H17NO8	327,0954	326,0886	328,1014	5,77	1,43	a,c,e	Carbohydrates and Carbohydrate Conjugates	Drug metabolite
Acetyl-L-carnitin	C9H17NO4	203,1158	nd	204,1224	nd	0,9	a,c,e	Lipids	Endogenous
Acetyl-N-formyl-5-methoxykynurenamine/P henylacetyl-L-glutamine	C13H16N2O4	264,111	263,1042	nd	1,82	nd	a,b,d	Aromatic Homomonocyclic Compounds	Endogenous
Adenosine	C10H13N5O4	267,0967	266,0906	268,103	5,2	1,17	a,c,e	Nucleosides, Nucleotides, and Analogues	Endogenous
Adipic acid	C6H10O4	146,0579	145,0508	nd	7,78	nd	a,b,d	Organic Acids and Derivatives	Food
Agmatine	C5H14N4	130,1218	nd	131,1287	nd	0,7	a,c,e	Aliphatic Acyclic Compounds	Endogenous
Alanine	C3H7NO2	89,0477	nd	90,0547	nd	0,76	a,c,e	Amino Acids, Peptides, and Analogues	Endogenous
Alpha-D-Aminoadipic acid	C6H11NO4	161,0688	142,0511	162,0756	3,07	0,8	a,c,e	Amino Acids, Peptides, and Analogues	Endogenous
alpha-Hydroxyhippuric-acid	C9H9NO4	195,0526	194,0462	196,0597	3,69	3,39	a,c,e	Amino Acids, Peptides, and Analogues	Endogenous

Arginine	C6H14N4O2	174,1117	173,1046	175,1184	9,67	0,72	a,c,e	Amino Acids, Peptides, and Analogues	essential amino-acid
Aspartyl-Phenylalanine	C13H16N2O5	280,1059	nd	281,1123	nd	4,1	a,c,e	Carboxylic acids and derivatives	Food
Benzoic acid	C7H6O2	122,0368	121,0296	nd	2,13	nd	a,b,d	Aromatic Homomonocyclic Compounds	Food
beta-Pseudouridine	C9H12N2O6	244,0695	243,0628	245,076	5,9	0,97	a,c,e	Carbohydrates and Carbohydrate Conjugates	Endogenous
Cadaverin	C5H14N2	102,1157	nd	86,0961	nd	0,67	a,c	Aliphatic Acyclic Compounds	Endogenous
Caffeine	C8H10N4O2	194,0804	nd	195,0873	nd	5	a,c,e	Aromatic Heteropolycyclic Compounds	Drug
Carnitine	C7H15NO3	161,1052	nd	162,1119	nd	0,75	a,c,e	Aliphatic Acyclic Compounds	Endogenous
cis,cis-Muconic acid	C6H6O4	142,0266	141,0196	nd	9,14	nd	a,b,d	Lipids	Endogenous
Creatinine	C4H7N3O	113,0589	112,0518	nd	3,28	nd	a,b,d	Aliphatic Heteromonocyclic Compounds	Endogenous
Curcumin	C21H20O6	368,126	367,1223	nd	1,37	nd	a,b,d	Aromatic Homomonocyclic Compounds	Food
D-(-)-Ribose	C5H10O5	150,0528	149,0457	173,0439	4,52	1,08	a,c	Carbohydrates and Carbohydrate Conjugates	Endogenous
D-(+)-Cellobiose	C12H22O11	342,1162	343,1248	nd	8,59	nd	a,b	Carbohydrates and Carbohydrate Conjugates	Microbial

D-(+)-Neopterin	C9H11N5O4	253,0811	252,073	254,0873	7,98	0,8	a,c	Aromatic Heteropolycyclic Compounds	Endogenous
Dehydroascorbic acid	C6H6O6	174,0164	173,0093	nd	10,66	nd	a,b,d	Carbohydrates and Carbohydrate Conjugates	Endogenous
Deoxycorticosterone-21-glucoside	C27H40O8	492,2723	491,2652	493,2769	1,45	7,61	a,c	Lipids	Endogenous
D-erythro-Dihydrosphingosine (Sphinganine)	C18H39NO2	301,2981	283,2644	nd	1,22	nd	a,b	Lipids	Endogenous
D-Galacturonic acid	C6H10O7	194,0426	193,0354	217,0311	8,75	0,77	a,b,c,d	Carbohydrates and Carbohydrate Conjugates	Endogenous
Diglycolic acid	C4H6O5	134,0215	133,0143	157,0127	8,61	1,03	a,c	NA	NA
D-Lactic acid	C3H6O3	90,0317	89,0245	nd	3,9	nd	a,b,d	Organic Acids and Derivatives	Endogenous
D-Ribose-phosphate	C5H11O8P	230,0192	229,0123	nd	8,58	nd	a,b,d	Carbohydrates and Carbohydrate Conjugates	Endogenous
D-Xylulose	C5H10O5	150,0528	nd	173,0414	nd	0,78	a,c	Carbohydrates and Carbohydrate Conjugates	Endogenous
Equol (phytoestrogen)	C15H14O3	242,0943	241,0864	nd	1,3	nd	a,b	Aromatic Heteropolycyclic Compounds	Endogenous
Ethylmethylacetic acid	C5H10O2	102,0681	147,0664	nd	2,58	nd	a,b,d	Lipids	Endogenous
Fumaric acid	C4H4O4	116,011	115,0037	nd	8,72	nd	a,b,d	Lipids	Endogenous

Galactonic acid	C6H12O7	196,0583	195,0509	nd	7,63	nd	a,b,d	Carbohydrates and Carbohydrate Conjugates	Endogenous
Gluconic acid	C6H12O7	196,0583	195,0509	219,0468	7,63	0,76	a,c	Carbohydrates and Carbohydrate Conjugates	Endogenous
glycodeoxycholate	C26H43NO5	449,3141	448,3071	nd	1,43	nd	a,b,d	Lipids	Endogenous
Glycyl-L-leucine	C8H16N2O3	188,1161	nd	189,1227	nd	2,65	a,c,e	Amino Acids, Peptides, and Analogues	Endogenous
Gly-Pro	C7H12N2O3	172,0848	171,0777	173,0915	6,17	0,84	a,c,e	Amino Acids, Peptides, and Analogues	Endogenous
Guanidinosuccinic acid	C5H9N3O4	175,0593	174,0523	176,066	8,47	1,02	a,c,e	Amino Acids, Peptides, and Analogues	Endogenous
Guanosine	C10H13N5O5	283,0917	282,0848	284,0979	6,11	1,09	a,c,e	Nucleosides, Nucleotides, and Analogues	Endogenous
Hexose	C6H12O6	180,0634	179,0565	nd	8,47	nd	a,b,d	Carbohydrates and Carbohydrate Conjugates	Endogenous
Histidine	C6H9N3O2	155,0695	154,0624	156,0763	7,47	0,71	a,c,e	Amino Acids, Peptides, and Analogues	essential amino-acid
Homovanillic acid	C9H10O4	182,0579	181,051	nd	3,43	nd	a,b,d	Aromatic Homomonocyclic Compounds	Endogenous
Hydroxykynurenine	C10H12N2O4	224,0797	179,0828	nd	2,59	nd	a,b	Amino Acids, Peptides, and Analogues	Endogenous
Hydroxynicotinic acid	C6H5NO3	139,0269	138,0197	nd	8,8	nd	a,b,d	Benzene and substituted derivatives	Endogenous

Hypoxanthine	C5H4N4O	136,0385	135,031	137,0453	5,01	1,22	a,c,e	Aromatic Heteropolycyclic Compounds	Endogenous
Imidazolelactic acid	C6H8N2O3	156,0535	155,0465	nd	5,16	nd	a,b,d	Aromatic Heteromonocyclic Compounds	Endogenous
Indole-3-carboxylic acid	C9H7NO2	161,0477	160,0407	nd	3,46	nd	a,b,d	Aromatic Heteropolycyclic Compounds	Food
Indolelactic acid	C11H11NO3	205,0739	204,0304	nd	2,62	nd	a,b,d	Aromatic Heteropolycyclic Compounds	Endogenous
Indoxyl sulfate	C8H7NO4S	213,009	209,0305	nd	2,5	nd	a,b,d	Aromatic Heteropolycyclic Compounds	Endogenous
Isobutyric acid	C4H8O2	88,0524	87,0452	nd	2,48	nd	a,b	Organic Acids and Derivatives	Endogenous
Isocitric-acid	C6H8O7	192,027	191,02	215,0155	9,91	0,94	a,c	Organic Acids and Derivatives	Endogenous
Lactulose	C12H22O11	342,1162	341,1093	nd	9,26	nd	a,b,d	Carbohydrates and Carbohydrate Conjugates	Drug
L-Alanyl-L-proline	C8H14N2O3	186,1004	185,0935	187,1071	3,06	0,92	a,c,e	Carboxylic acids and derivatives	Endogenous
L-Aspartic-acid	C4H7NO4	133,0375	132,0304	134,0444	9,11	0,76	a,c,e	Amino Acids, Peptides, and Analogues	Endogenous
L-Citrulline	C6H13N3O3	175,0957	174,0886	176,1023	7,64	0,76	a,c,e	Amino Acids, Peptides, and Analogues	Endogenous
L-Cystine	C6H12N2O4S2	240,0239	239,017	241,0303	8,33	0,76	a,c,e	Amino Acids, Peptides, and Analogues	Endogenous

L-Fucose	C6H12O5	164,0685	nd	187,0573	nd	0,8	a,c	Carbohydrates and Carbohydrate Conjugates	Endogenous
L-Glutamine	C5H10N2O3	146,0691	145,062	147,076	7,09	0,76	a,c,e	Amino Acids, Peptides, and Analogues	Endogenous
L-Homoserine	C4H9NO3	119,0582	118,0511	120,0651	6,5	0,76	a,b,c,d,e	Amino Acids, Peptides, and Analogues	Endogenous
Lithocholic acid 3-sulfate	C24H40O6S	456,2546	455,2474	nd	1,31	nd	a,b	Lipids	Endogenous
L-Kynurenine	C10H12N2O3	208,0848	nd	209,0914	nd	2,09	a,c,e	Amino Acids, Peptides, and Analogues	Endogenous
L-Pyroglutamic-acid	C5H7NO3	129,0426	128,0355	130,0496	4,66	1,01	a,c,e	Aliphatic Heteromonocyclic Compounds	Endogenous
L-Rhamnose	C6H12O5	164,0685	nd	165,0752	nd	0,77	a,c,e	Carbohydrates and Carbohydrate Conjugates	Endogenous
L-Saccharopine	C11H20N2O6	276,1321	275,1255	277,1383	8,62	0,77	a,c,e	Amino Acids, Peptides, and Analogues	Endogenous
L-Threonic-acid-1,4-lactone	C4H6O4	118,0266	nd	119,0336	nd	1,17	a,c,e	Aliphatic Heteromonocyclic Compounds	Endogenous
Lysine	C6H14N2O2	146,1055	145,0984	147,1123	9,21	0,69	a,c,e	Amino Acids, Peptides, and Analogues	essential amino-acid
L- α -aminobutyric acid	C4H9NO2	103,0633	102,0561	nd	8,14	nd	a,b	Amino Acids, Peptides, and Analogues	Endogenous
Maleic-acid	C4H4O4	116,011	115,0037	117,0179	8,72	1,08	a,c,e	Lipids	Endogenous
Malic acid	C4H6O5	134,0215	133,0143	nd	8,75	nd	a,b,d	Organic Acids and Derivatives	Endogenous

Maltotriose	C18H32O16	504,169	503,1622	527,1559	8,47	0,74	a,c	Carbohydrates and Carbohydrate Conjugates	Endogenous
Melibiose	C12H22O11	342,1162	nd	365,104	nd	0,74	a,c	Carbohydrates and Carbohydrate Conjugates	Microbial
Muramic acid	C9H17NO7	251,1005	250,0937	252,1067	7,63	0,76	a,c,e	Carbohydrates and Carbohydrate Conjugates	Microbial
Myo-inositol	C6H12O6	180,0634	179,0564	nd	9,01	nd	a,b,d	Aliphatic Homomonocyclic Compounds	Endogenous
Myristic acid	C14H28O2	228,2089	227,2019	nd	1,26	nd	a,b	Lipids	Endogenous
N acetyl L glutamine	C7H12N2O4	188,0797	187,0728	189,0863	4,99	1,01	a,c,e	Amino Acids, Peptides, and Analogues	Endogenous
N6,N6,N6-Trimethyl-L-lysine	C9H20N2O2	188,1525	nd	189,1591	nd	0,71	a,c,e	Amino Acids, Peptides, and Analogues	Endogenous
N6-Acetyl-L-lysine	C8H16N2O3	188,1161	nd	189,1227	nd	0,81	a,c,e	Amino Acids, Peptides, and Analogues	Endogenous
N-Acetyl-Asp-Glu	C11H16N2O8	304,0907	303,084	305,0967	9,11	1,07	a,c,e	Amino Acids, Peptides, and Analogues	Endogenous
N-Acetyl-DL-serine	C5H9NO4	147,0532	nd	148,0599	nd	0,76	a,c,e	Amino Acids, Peptides, and Analogues	Endogenous
N-acetyl-DL-tryptophan	C13H14N2O3	246,1004	nd	247,1069	nd	6,11	a,c,e	Amino Acids, Peptides, and Analogues	Endogenous
N-Acetylhistamine	C7H11N3O	153,0902	190,0546	nd	2,17	nd	a,b,d	Aromatic Heteromonocyclic Compounds	Endogenous
N-Acetyl-L-aspartic acid	C6H9NO5	175,0481	174,041	130,0496	8,11	0,76	a,c	Amino Acids, Peptides, and Analogues	Endogenous

N-Acetyl-L-carnosine	C11H16N4O4	268,1172	nd	269,1233	nd	0,8	a,c,e	Amino Acids, Peptides, and Analogues	Endogenous
N-Acetyl-L-glutamic acid	C7H11NO5	189,0637	188,0568	190,0703	7,87	1,33	a,c,e	Amino Acids, Peptides, and Analogues	Endogenous
Nicotinamide N-oxide	C6H6N2O2	138,0429	137,0357	nd	2,5	nd	a,b,d	Aromatic Heteromonocyclic Compounds	Endogenous
Nicotinic acid	C6H5NO2	123,032	nd	124,039	nd	0,93	a,c,e	Aromatic Heteromonocyclic Compounds	Drug
Nicotinuric acid	C8H8N2O3	180,0535	179,0466	nd	3,6	nd	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous
N-Methyl-D-aspartic acid (NMDA)	C5H9NO4	147,0532	146,0461	nd	8,24	nd	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous
N- α -acetyl-L-arginine	C8H16O3N4	216,1222	215,033	nd	7,67	nd	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous
N α -Acetyl-L-asparagine	C6H10N2O4	174,0641	173,0571	175,0708	8,02	1,02	a,c,e	Amino Acids, Peptides, and Analogues	Endogenous
O-Acetyl-L-serine/N-Acetyl-DL-serine	C5H9NO4	147,0532	146,0461	nd	5,23	nd	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous
Ornithine	C5H12N2O2	132,0899	131,0827	133,0967	9,57	0,69	a,c	Amino Acids, Peptides, and Analogues	Endogenous
Oxaloacetic acid	C4H4O5	132,0059	87,0085	nd	7,19	nd	a,b	Organic Acids and Derivatives	Endogenous
Pantothenic acid	C9H17NO5	219,1107	216,9816	220,1172	3,26	2,24	a,c,e	Aliphatic Acyclic Compounds	Endogenous
paraxanthine	C7H8N4O2	180,0647	179,0577	181,0714	2,02	3,55	a,b,c,d,e	Aromatic Heteropolycyclic Compounds	Endogenous

Phenylalanine	C9H11NO2	165,079	164,072	166,0856	2,82	1,88	a,c,e	Amino Acids, Peptides, and Analogues	essential amino-acid
Proline	C5H9NO2	115,0633	114,0562	116,0702	6,02	1,04	a,c,e	Amino Acids, Peptides, and Analogues	Endogenous
Propionylcarnitine	C10H19NO4	217,1314	nd	218,138	nd	1,34	a,c,e	Lipids	Endogenous
S-Adenosyl-homocysteine	C14H20N6O5S	384,1216	nd	385,1299	nd	0,75	a,c,e	Amino Acids, Peptides, and Analogues	Endogenous
Sarcosine	C3H7NO2	89,0477	88,0405	nd	6,89	nd	a,b,d	Amino Acids, Peptides, and Analogues	Endogenous
Serine	C3H7NO3	105,0426	104,0354	106,0495	7,25	0,97	a,c,e	Amino Acids, Peptides, and Analogues	Endogenous
S-ethyl-L-cysteine	C5H11NO2S	149,0511	148,044	133,0314	3,68	0,96	a,b,c,d	Amino Acids, Peptides, and Analogues	Endogenous
Shikimic acid	C7H10O5	174,0528	173,0457	nd	7,53	nd	a,b,d	Aliphatic Homomonocyclic Compounds	Endogenous
Stachydrine	C7H13NO2	143,0946	nd	144,1015	nd	0,79	a,c,e	Aliphatic Heteromonocyclic Compounds	Endogenous
Stachyose	C24H42O21	666,2219	665,2151	689,2083	9,69	0,75	a,b,c,d	Carbohydrates and Carbohydrate Conjugates	Endogenous
Succinic acid/Methylmalonic acid	C4H6O4	118,0266	117,0195	nd	9,92	nd	a,b,d	Organic Acids and Derivatives	Endogenous
Sucrose	C12H22O11	342,1162	341,1093	nd	7,81	nd	a,b,d	Carbohydrates and Carbohydrate Conjugates	Food
Taurochenodeoxycholic acid	C26H45NO6S	499,2968	498,2895	482,2912	1,36	8,17	a,b,c,d	Lipids	Endogenous

Theobromine	C7H8N4O2	180,0647	179,0577	181,0714	2,02	2,25	a,b,c,d,e	Aromatic Heteropolycyclic Compounds	Endogenous
Thymidine	C10H14N2O5	242,0903	241,0836	nd	2,18	nd	a,b,d	Nucleosides, Nucleotides, and Analogues	Endogenous
Tiglic acid	C5H8O2	100,0524	99,0452	nd	2,59	nd	a,b,d	Lipids	Endogenous
Tyrosine	C9H11NO3	181,0739	180,0667	182,0806	5,37	1,06	a,c,e	Amino Acids, Peptides, and Analogues	Endogenous
Uridine	C9H12N2O6	244,0695	243,0626	nd	3,93	nd	a,b,d	Nucleosides, Nucleotides, and Analogues	Endogenous
Valine	C5H11NO2	117,079	116,0718	nd	4,86	nd	a,b,d	Amino Acids, Peptides, and Analogues	essential amino-acid
Val-Pro	C10H18N2O3	214,1317	212,0026	215,1384	2,14	1,51	a,b,c,d,e	Lipids	Endogenous
Xanthine	C5H4N4O2	152,0334	151,0264	153,0402	5,81	0,97	a,c,e	Aromatic Heteropolycyclic Compounds	Endogenous
Xanthopterin	C6H5N5O2	179,0443	178,0359	nd	8,74	nd	a,b,d	Aromatic Heteropolycyclic Compounds	Endogenous
Xanthurenic-acid	C10H7NO4	205,0375	203,083	206,0441	7,66	3,81	a,c,e	Amino Acids, Peptides, and Analogues	Endogenous

Annexe 12 : Liste des 67 variables discriminantes obtenues par une analyse univariée de type wilcoxon ayant une pvalue < 0,05 entre les temps de prélèvements précoces (< 1440 minutes (groupe 1 = gp1)) et tardifs (> 1440 minutes (groupe 2 = gp2)) au sein des échantillons de méconium de la cohorte clinique.

Composés	Familles Chimiques	Origines	p-values	Evolution gp1-gp2
Succinic semialdehyde	Lipids	endogène	0,0218	diminue
4-Imidazoleacrylic acid (urocanic acid)	Aromatic Heteromonocyclic Compounds	endogène	0,0285	diminue
L-(+)-Tartaric acid	Organic Acids and Derivatives	alimentaire	0,0310	diminue
Glycerol 3-phosphate	Carbohydrates and Carbohydrate Conjugates	endogène	0,0046	diminue
D-(+)-Glucuronic acid gamma-lactone	Aliphatic Heteropolycyclic Compounds	endogène	0,0106	diminue
2-Isopropylmalic acid	Organic Acids and Derivatives	microbien	0,0106	diminue
D-Mannitol	Lipids	endogène	0,0293	diminue
D-Galacturonic acid	Carbohydrates and Carbohydrate Conjugates	endogène	0,0047	diminue
D-Glucuronic acid	Carbohydrates and Carbohydrate Conjugates	endogène	0,0047	diminue
D-Saccharic acid	Carbohydrates and Carbohydrate Conjugates	endogène	0,0049	diminue
Xanthopterin	Aromatic Heteropolycyclic Compounds	endogène	0,0046	diminue
D-Ribose 5-phosphate	Carbohydrates and Carbohydrate Conjugates	endogène	0,0052	diminue
L-Cystine	Amino Acids, Peptides, and Analogues	endogène	0,0145	diminue
Adenosine	Nucleosides, Nucleotides, and Analogues	endogène	0,0293	diminue
(-)-N-Acetylneuraminic acid	Carbohydrates and Carbohydrate Conjugates	endogène	0,0472	diminue
Taurolithocholic acid 3-sulfate	Lipids	endogène	0,0117	diminue
Oxaloacetic acid	Organic Acids and Derivatives	endogène	0,0391	diminue
(S)-3-Hydroxybutyric acid	Organic Acids and Derivatives	endogène	0,0194	augmente
D-Glyceric acid	Carbohydrates and Carbohydrate Conjugates	endogène	0,0018	augmente
Proline	Amino Acids, Peptides, and Analogues	endogène	0,0094	augmente
Valine	Amino Acids, Peptides, and Analogues	Aminoacide essentiel	0,0106	augmente
N-Acetylglycine	Amino Acids, Peptides, and Analogues	endogène	0,0352	augmente
2-Hydroxy 3-methylbutyric acid (α -Hydroxyisovaleric acid)	Lipids	endogène	0,0054	augmente
Succinic acid/Methylmalonic acid	Organic Acids and Derivatives	endogène	0,0431	augmente
D-Threitol	Carbohydrates and Carbohydrate Conjugates	endogène	0,0024	augmente
Taurine	Organic Acids and Derivatives	endogène	0,0018	augmente

Thymine	Aromatic Heteromonocyclic Compounds	endogène	0,0244	augmente
L-Pyroglutamic-acid	Aliphatic Heteromonocyclic Compounds	endogène	0,0107	augmente
N-Acetyl-L-alanine	Amino Acids, Peptides, and Analogues	endogène	0,0024	augmente
trans-4-hydroxy-L-proline	Amino Acids, Peptides, and Analogues	endogène	0,0076	augmente
Methyl-succinic acid/mono-Methyl hydrogen succinate	Organic Acids and Derivatives	endogène	0,0072	augmente
Adenine	Aromatic Heteropolycyclic Compounds	endogène	0,0024	augmente
Threonic acid	Organooxygen compounds	endogène	0,0153	augmente
4-Hydroxybenzoic acid	Aromatic Homomonocyclic Compounds	endogène	0,0341	augmente
4-Acetamidobutanoic acid	Lipids	endogène	0,0153	augmente
O-Acetyl-L-serine/N-Acetyl-DL-serine	Amino Acids, Peptides, and Analogues	endogène	0,0106	augmente
Ethylmethylacetic acid	Lipids	endogène	0,0257	augmente
Mevalonic acid	Organic Acids and Derivatives	endogène	0,0257	augmente
L(+)-arabinose	Carbohydrates and Carbohydrate Conjugates	endogène	0,0293	augmente
Orotic acid	Amino Acids, Peptides, and Analogues	endogène	0,0106	augmente
5-Hydroxylysine	Amino Acids, Peptides, and Analogues	endogène	0,0282	augmente
Phenylalanine	Amino Acids, Peptides, and Analogues	Aminoacide essentiel	0,0257	augmente
7-methylguanine	Aromatic Heteropolycyclic Compounds	endogène	0,0049	augmente
Capric acid	Lipids	endogène	0,0028	augmente
Suberic acid	Organic Acids and Derivatives	endogène	0,0132	augmente
N-Acetyl-L-aspartic acid	Amino Acids, Peptides, and Analogues	endogène	0,0022	augmente
Homovanillic acid	Aromatic Homomonocyclic Compounds	endogène	0,0046	augmente
L-Alanyl-L-proline	Carboxylic acids and derivatives	endogène	0,0282	augmente
N acetyl L glutamine	Amino Acids, Peptides, and Analogues	endogène	0,0018	augmente
Azelaic acid	Organic Acids and Derivatives	endogène	0,0044	augmente
N-Acetyl-L-glutamic acid	Amino Acids, Peptides, and Analogues	endogène	0,0153	augmente
N-Acetylhistamine	Aromatic Heteromonocyclic Compounds	endogène	0,0161	augmente
Indolelactic acid	Aromatic Heteropolycyclic Compounds	endogène	0,0049	augmente
Glucose-6-PO4-CO2	Carbohydrates and Carbohydrate Conjugates	endogène	0,0046	augmente

Indoleacrylic acid	Aromatic Heteropolycyclic Compounds	endogène	0,0106	augmente
Myristic acid	Lipids	endogène	0,0293	augmente
Acetyl-N-formyl-5-methoxykynurenamine/Phenylacetyl-L-glutamine	Aromatic Homomonocyclic Compounds	endogène	0,0046	augmente
cis-5,8,11,14,17-Eicosapentaenoic acid	Lipids	alimentaire	0,0352	augmente
Sucrose	Carbohydrates and Carbohydrate Conjugates	alimentaire	0,0153	augmente
Estriol 17-sulphate	Lipids	endogène	0,0257	augmente
Curcumin	Aromatic Homomonocyclic Compounds	alimentaire	0,0399	augmente
Cholic acid	Lipids	endogène	0,0018	augmente
Maltotetraose	Carbohydrates and Carbohydrate Conjugates	endogène	0,0018	augmente
Stachyose	Carbohydrates and Carbohydrate Conjugates	endogène	0,0018	augmente
Isobutyric acid	Organic Acids and Derivatives	endogène	0,0153	augmente
Sarcosine	Amino Acids, Peptides, and Analogues	endogène	0,0047	augmente
D-Lactic acid	Organic Acids and Derivatives	endogène	0,0414	augmente

Annexe 13 : Liste des 922 protéines identifiées avec au moins deux peptides spécifiques (au sein de tous les échantillons), décrites par ordre d'abondance décroissant dans le méconium (NSAF [%] décroissant).

N°	Accession [Humain]	Protéines	masses	Nombre de peptides spécifiques (tous les échantillons)	Couverture [%]	NSAF [%]
1	A1AT	Alpha-1-antitrypsin	46707	40	66,7	7,511
2	S10A9	Protein S100-A9	13234	14	70,2	4,620
3	S10A8	Protein S100-A8	10828	17	86	4,618
4	IGKC	Ig kappa chain C region	11602	8	93,4	2,605
5	AACT	Alpha-1-antichymotrypsin	47621	25	47,3	1,801
6	H4	Histone H4	11360	9	57,3	1,738
7	LAC2	Ig lambda-2 chain C regions	11287	4	93,4	1,375
8	K2C1	Keratin, type II cytokeletal 1	65999	46	57,1	1,174
9	TTHY	Transthyretin	15877	10	65,3	1,155
10	COL	Colipase	11946	5	56,2	1,098
11	CEL3A	Chymotrypsin-like elastase family member 3A	29470	10	45,6	1,093
12	CTRC	Chymotrypsin-C	29465	14	39,2	1,090
13	ALBU	Serum albumin	69321	57	69,8	1,007
14	K1C10	Keratin, type I cytokeletal 10	58792	33	64,7	1,001
15	AMPN	Aminopeptidase N	109471	59	49,8	0,897
16	PSPB	Pulmonary surfactant-associated protein B	42089	7	14,7	0,790
17	ANT3	Antithrombin-III	52569	28	46,3	0,718
18	H2B1C	Histone H2B type 1-C/E/F/G/I	13898	3	50,8	0,674
19	CLCA1	Calcium-activated chloride channel regulator 1	100163	48	43,9	0,661
20	HV307	Immunoglobulin heavy variable 3-7	12934	3	41,9	0,659
21	ITLN1	Intelectin-1	34939	11	47	0,632
22	TFF3	Trefoil factor 3	8635	4	47,5	0,630
23	H2B1B	Histone H2B type 1-B	13942	2	50,8	0,620

24	PPBI	Intestinal-type alkaline phosphatase	56776	14	40,5	0,619
25	SLPI	Antileukoproteinase	14316	11	47	0,597
26	PERE	Einophil peroxidase	80989	36	50,6	0,585
27	KV320	Immunoglobulin kappa variable 3-20	12549	2	27,6	0,585
28	CEL3B	Chymotrypsin-like elastase family member 3B	29244	5	32,2	0,584
29	IGLL5	Immunoglobulin lambda-like polypeptide 5	23049	4	50	0,581
30	APOD	Apolipoprotein D	21262	11	41,3	0,562
31	IC1	Plasma protease C1 inhibitor	55119	23	40,2	0,552
32	NEP	Neprilysin	85460	38	46,4	0,542
33	DPEP1	Dipeptidase 1	45645	16	42,1	0,525
34	ACTB	Actin, cytoplasmic 1	41710	9	52,5	0,500
35	SPR2D	Small proline-rich protein 2D	7900	2	79,2	0,494
36	TRFL	Lactotransferrin	78132	71	76,1	0,467
37	IGHG2	Ig gamma-2 chain C region	35878	3	32,8	0,461
38	TFF2	Trefoil factor 2	14274	6	41,9	0,455
39	ANXA2	Annexin A2	38580	30	72,9	0,447
40	IGHA1	Ig alpha-1 chain C region	37631	6	50,7	0,435
41	K22E	Keratin, type II cytoskeletal 2 epidermal	65393	33	73,7	0,431
42	KV401	Immunoglobulin kappa variable 4-1	13372	3	23,1	0,417
43	K1C14	Keratin, type I cytoskeletal 14	51529	4	39,2	0,396
44	K1C16	Keratin, type I cytoskeletal 16	51236	5	38,1	0,390
45	PRG2	Bone marrow proteoglycan	25189	10	42,8	0,388
46	IGHG3	Ig gamma-3 chain C region	41260	2	21,8	0,386
47	IGHG1	Ig gamma-1 chain C region	36083	6	51,5	0,379
48	S10A6	Protein S100-A6	10173	5	33,3	0,378
49	DEF1	Neutrophil defensin 1	10194	2	28,7	0,375
50	KVD20	Immunoglobulin kappa variable 3-20	12507	2	27,6	0,356
51	PA21B	Phospholipase A2	16349	6	44,6	0,354
52	CTRB2	Chymotrypsinogen B2	27905	2	44,9	0,349
53	ECP	Einophil cationic protein	18373	9	42,5	0,334

54	IGHA2	Ig alpha-2 chain C region	36503	3	39,7	0,331
55	PERM	Myeloperoxidase	83815	33	45,5	0,328
56	ITLN2	Intelectin-2	36188	8	36,6	0,323
57	SAMP	Serum amyloid P-component	25371	9	34,1	0,320
58	ANXA1	Annexin A1	38690	25	63	0,316
59	ANXA4	Annexin A4	35860	29	68,7	0,313
60	SPB6	Serpin B6	42594	15	58,5	0,305
61	LYSC	Lysozyme C	16526	10	50	0,288
62	K1C19	Keratin, type I cytokeletal 19	44079	6	25,8	0,280
63	K1C17	Keratin, type I cytokeletal 17	48076	8	38,9	0,279
64	K2C5	Keratin, type II cytokeletal 5	62340	26	58,1	0,273
65	S10AB	Protein S100-A11	11733	6	45,7	0,262
66	DMBT1	Deleted in malignant brain tumors 1 protein	260569	37	18,5	0,261
67	ZG16B	Zymogen granule protein 16 homolog B	22725	7	38,5	0,258
68	FRIL	Ferritin light chain	20007	12	43,4	0,257
69	DPP4	Dipeptidyl peptidase 4	88222	36	41,5	0,256
70	K2C8	Keratin, type II cytokeletal 8	53671	14	28,8	0,255
71	FABPL	Fatty acid-binding protein, liver	14199	7	37,8	0,254
72	HV315	Immunoglobulin heavy variable 3-15	12917	5	37	0,252
73	LUM	Lumican	38405	13	36,4	0,251
74	CBPA1	Carboxypeptidase A1	47111	18	48,7	0,250
75	SPB3	Serpin B3	44537	12	56,2	0,242
76	NPC2	Epididymal secretory protein E1	16559	8	51,7	0,242
77	K2C1B	Keratin, type II cytokeletal 1b	61864	30	45,8	0,226
78	RS27A	Ubiquitin-40S ribosomal protein S27a	17953	4	24,4	0,223
79	SODC	Superoxide dismutase [Cu-Zn]	15926	9	49,4	0,221
80	PA2GA	Phospholipase A2, membrane associated	16072	9	38,9	0,210
81	K2C7	Keratin, type II cytokeletal 7	51354	4	13,4	0,198
82	CEL2A	Chymotrypsin-like elastase family member 2A	28869	3	16	0,198
83	K1C13	Keratin, type I cytokeletal 13	49557	3	21,6	0,196

84	HBA	Hemoglobin subunit alpha	15248	8	50,7	0,193
85	G3P	Glyceraldehyde-3-phosphate dehydrogenase	36030	12	38,2	0,192
86	PPB1	Alkaline phosphatase, placental type	57917	2	15,5	0,191
87	MUC13	Mucin-13	54569	14	33,6	0,191
88	K1C9	Keratin, type I cytoskeletal 9	62027	16	32,3	0,190
89	FCGBP	IgG Fc-binding protein	571639	116	25,2	0,188
90	IPSP	Plasma serine protease inhibitor	45646	10	21,7	0,185
91	ANXA5	Annexin A5	35914	18	44,4	0,178
92	XPP2	Xaa-Pro aminopeptidase 2	75577	25	33,1	0,176
93	EZRI	Ezrin	69370	30	50	0,176
94	IGJ	Immunoglobulin J chain	18087	11	52,8	0,173
95	ANX13	Annexin A13	35393	22	64,2	0,172
96	K2C4	Keratin, type II cytoskeletal 4	57250	5	15,5	0,167
97	RET2	Retinol-binding protein 2	15697	7	44,8	0,165
98	SPB4	Serpin B4	44825	2	29,2	0,164
99	COF1	Cofilin-1	18491	12	39,8	0,160
100	CRP	C-reactive protein	25023	7	21	0,158
101	TRY2	Trypsin-2	26471	3	32	0,157
102	GDN	Glia-derived nexin	43974	20	43,2	0,155
103	FRIH	Ferritin heavy chain	21212	9	28,4	0,154
104	LG3BP	Galectin-3-binding protein	65289	16	30,9	0,153
105	CBPB1	Carboxypeptidase B	47338	14	37,9	0,151
106	K2C80	Keratin, type II cytoskeletal 80	50494	11	16,4	0,151
107	HV434	Immunoglobulin heavy variable 4-34	13806	4	26,8	0,150
108	TRY1	Trypsin-1	26541	5	34	0,148
109	K1H1	Keratin, type I cuticular Ha1	47207	3	25,2	0,146
110	1433Z	14-3-3 protein zeta/delta	27728	5	32,7	0,144
111	GP2	Pancreatic secretory granule membrane major glycoprotein GP2	59442	14	26,6	0,142
112	KV230	Immunoglobulin kappa variable 2-30	13177	3	16,7	0,141
113	CD63	CD63 antigen	25619	4	12,6	0,139

114	S10AE	Protein S100-A14	11655	5	46,2	0,138
115	SPB12	Serpin B12	46247	18	45,7	0,136
116	H31T	Histone H3.1t	15499	2	19,9	0,134
117	LEG4	Galectin-4	35918	15	36,8	0,131
118	K1C18	Keratin, type I cytokeletal 18	48029	8	18,8	0,131
119	ILEU	Leukocyte elastase inhibitor	42715	16	47,2	0,130
120	S10A7	Protein S100-A7	11464	5	35,6	0,129
121	K22O	Keratin, type II cytokeletal 2 oral	65800	7	23,8	0,127
122	PIGR	Polymeric immunoglobulin receptor	83232	35	40,1	0,126
123	KRT36	Keratin, type I cuticular Ha6	52214	7	26,1	0,125
124	CBG	Corticosteroid-binding globulin	45112	7	19,3	0,124
125	RETN	Resistin	11411	3	35,2	0,123
126	PTTG	Pituitary tumor-transforming gene 1 protein-interacting protein	20310	4	20,6	0,122
127	CYTA	Cystatin-A	11000	5	63,3	0,120
128	K2C79	Keratin, type II cytokeletal 79	57800	7	22,1	0,120
129	ENPP7	Ectonucleotide pyrophosphatase/phosphodiesterase family member 7	51461	10	21,6	0,119
130	DNAS1	Deoxyribonuclease-1	31414	4	18,1	0,119
131	MEP1A	Meprin A subunit alpha	84365	16	22,9	0,119
132	CERU	Ceruloplasmin	122128	26	23,2	0,118
133	CASPE	Caspase-14	27662	14	47,1	0,117
134	HV349	Immunoglobulin heavy variable 3-49	13048	3	26,1	0,117
135	STOM	Erythrocyte band 7 integral membrane protein	31711	11	48,3	0,116
136	PROF1	Profilin-1	15045	6	45,7	0,115
137	LDHA	L-lactate dehydrogenase A chain	36665	7	21,1	0,114
138	BST1	ADP-ribosyl cyclase/cyclic ADP-ribose hydrolase 2	35701	6	17,9	0,114
139	CBPA2	Carboxypeptidase A2	47001	11	18,4	0,113
140	ASAH1	Acid ceramidase	44631	13	16,2	0,113
141	SAP	Prapin	58074	10	17,4	0,111
142	CDHR2	Cadherin-related family member 2	141455	24	18,6	0,111
143	HBB	Hemoglobin subunit beta	15988	4	68,7	0,110

144	SIM24	Small integral membrane protein 24	14955	4	30	0,110
145	MOES	Moesin	67778	3	18	0,108
146	GGH	Gamma-glutamyl hydrolase	35941	8	22,3	0,107
147	PPIA	Peptidyl-prolyl cis-trans isomerase A	18001	9	26,7	0,106
148	TRFE	Serotransferrin	77014	34	50,3	0,105
149	GBG12	Guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-12	8001	2	25	0,104
150	CBPQ	Carboxypeptidase Q	51855	12	26,9	0,102
151	KCRB	Creatine kinase B-type	42617	14	44,6	0,100
152	CAP7	Azurocidin	26869	4	16,3	0,097
153	SBP1	Selenium-binding protein 1	52358	20	43,6	0,096
154	1433G	14-3-3 protein gamma	28285	4	15,8	0,095
155	1433F	14-3-3 protein eta	28201	3	22	0,095
156	ZG16	Zymogen granule membrane protein 16	18135	5	45,5	0,094
157	BPIA1	BPI fold-containing family A member 1	26696	4	21,1	0,093
158	CEAM8	Carcinoembryonic antigen-related cell adhesion molecule 8	38130	2	12,9	0,093
159	HEP2	Heparin cofactor 2	57034	11	16,8	0,092
160	CEAM5	Carcinoembryonic antigen-related cell adhesion molecule 5	76747	5	13,7	0,092
161	EF1A1	Elongation factor 1-alpha 1	50109	11	17,7	0,090
162	K2C78	Keratin, type II cytoskeletal 78	56830	15	30,2	0,090
163	ARG11	Arginase-1	34713	14	45,7	0,088
164	BGLR	Beta-glucuronidase	74685	16	21,4	0,088
165	S10A4	Protein S100-A4	11721	6	39,6	0,088
166	CATG	Cathepsin G	28819	15	43,1	0,087
167	1433E	14-3-3 protein epsilon	29155	5	20	0,087
168	ARF4	ADP-ribosylation factor 4	20498	2	15,6	0,086
169	RB27A	Ras-related protein Rab-27A	24852	3	19	0,086
170	GNS	N-acetylglucosaminase-6-sulfatase	62042	8	14,9	0,086
171	FABPI	Fatty acid-binding protein, intestinal	15198	7	48,5	0,084
172	THBG	Thyroxine-binding globulin	46295	10	23,1	0,084
173	PTGDS	Prostaglandin-H2 D-isomerase	21015	7	33,2	0,083

174	ES8L2	Epidermal growth factor receptor kinase substrate 8-like protein 2	80570	17	19,2	0,083
175	SPHM	N-sulphoglucamine sulphohydrolase	56659	9	16,7	0,082
176	HV118	Immunoglobulin heavy variable 1-18	12812	2	26,5	0,082
177	KAIN	Kallistatin	48511	15	30,9	0,082
178	CLCA4	Calcium-activated chloride channel regulator 4	101219	2	3,5	0,081
179	SG3A2	Secretoglobin family 3A member 2	10155	6	41,9	0,081
180	VAMP8	Vesicle-associated membrane protein 8	11431	4	41	0,080
181	CEAM1	Carcinoembryonic antigen-related cell adhesion molecule 1	57525	3	8,7	0,078
182	LAMP2	Lysome-associated membrane glycoprotein 2	44932	5	8	0,078
183	RB27B	Ras-related protein Rab-27B	24592	2	17,4	0,078
184	LV321	Immunoglobulin lambda variable 3-21	12438	2	23,1	0,077
185	PRDX6	Peroxiredoxin-6	25019	8	29,9	0,077
186	LORI	Loricrin	25744	4	10,3	0,075
187	KRT84	Keratin, type II cuticular Hb4	64801	19	46,3	0,073
188	S100P	Protein S100-P	10393	4	28,4	0,072
189	SORCN	Sorcin	21662	6	24,2	0,071
190	ZA2G	Zinc-alpha-2-glycoprotein	34237	10	31,9	0,070
191	RAC1	Ras-related C3 botulinum toxin substrate 1	21436	2	22,9	0,070
192	HEMO	Hemopexin	51643	6	11,3	0,070
193	LV325	Immunoglobulin lambda variable 3-25	12004	2	18,8	0,069
194	OLFM4	Olfactomedin-4	57244	18	32,2	0,067
195	GSTP1	Glutathione S-transferase P	23341	4	21	0,066
196	SPB8	Serpin B8	42739	2	11,2	0,065
197	HBD	Hemoglobin subunit delta	16045	2	49,7	0,065
198	XP32	Skin-specific protein 32	26219	7	25,2	0,065
199	PIP	Prolactin-inducible protein	16562	7	45,9	0,065
200	CALM	Calmodulin	16827	5	26,8	0,064
201	ASPG	N(4)-(beta-N-acetylglucaminy)-L-asparaginase	37184	8	15,3	0,063
202	MUC2	Mucin-2	539958	46	7,4	0,063
203	CEAM7	Carcinoembryonic antigen-related cell adhesion molecule 7	29361	5	17,4	0,063

204	ENOA	Alpha-enolase	47139	12	36,9	0,062
205	VILI	Villin-1	92637	36	35,9	0,061
206	ALDOB	Fructe-bisphphate aldolase B	39448	7	27,2	0,061
207	CATH	Pro-cathepsin H	37369	7	18,2	0,061
208	ASM3A	Acid sphingomyelinase-like phphodiesterase 3a	51228	10	12,6	0,060
209	CEL	Bile salt-activated lipase	79272	21	26,8	0,058
210	CLIC1	Chloride intracellular channel protein 1	26906	8	32	0,058
211	LDHB	L-lactate dehydrogenase B chain	36615	7	23,7	0,058
212	FABP6	Gastrotropin	14362	3	27,3	0,058
213	CYTB	Cystatin-B	11133	2	21,4	0,057
214	RAI3	Retinoic acid-induced protein 3	40225	3	5,9	0,056
215	NALDL	N-acetylated-alpha-linked acidic dipeptidase-like protein	80508	17	24,7	0,056
216	UROM	Uromodulin	69714	13	18,1	0,056
217	RAP1B	Ras-related protein Rap-1b	20812	2	42,4	0,055
218	LCE1A	Late cornified envelope protein 1A	10975	2	15,5	0,055
219	LEG3	Galectin-3	26136	6	23,2	0,055
220	IGHM	Ig mu chain C region	49276	5	35	0,055
221	ANX11	Annexin A11	54355	15	25,7	0,054
222	GNA11	Guanine nucleotide-binding protein subunit alpha-11	42097	9	31,2	0,054
223	PNPH	Purine nucleide phphorylase	32097	4	12,8	0,054
224	NICA	Nicastrin	78362	7	6,6	0,054
225	A2AP	Alpha-2-antiplasmin	54531	6	12,6	0,053
226	ASM3B	Acid sphingomyelinase-like phphodiesterase 3b	50781	6	11,9	0,053
227	SPRR3	Small proline-rich protein 3	18142	7	37,9	0,052
228	VNN1	Pantetheinase	56975	10	15,6	0,051
229	TPIS	Triphosphate isomerase	30772	11	36,7	0,051
230	H12	Histone H1.2	21352	3	24,9	0,050
231	S10AA	Protein S100-A10	11196	2	19,6	0,050
232	CATD	Cathepsin D	44524	9	19,2	0,048
233	GDPD3	Glycerophphodiester phphodiesterase domain-containing protein 3	36573	5	17,3	0,048

234	ELNE	Neutrophil elastase	28500	5	16,1	0,048
235	AMPL	Cytol aminopeptidase	56131	14	32	0,048
236	APOA2	Apolipoprotein A-II	11168	4	19	0,048
237	KPRP	Keratinocyte proline-rich protein	64093	12	17,1	0,047
238	KPYM	Pyruvate kinase PKM	57900	17	26,7	0,047
239	CTHR1	Collagen triple helix repeat-containing protein 1	26207	4	21	0,047
240	CLRN3	Clarin-3	25305	3	12,4	0,047
241	NHRF1	Na(+)/H(+) exchange regulatory cofactor NHE-RF1	38845	10	20,9	0,045
242	LV319	Immunoglobulin lambda variable 3-19	12035	2	17	0,045
243	GBB1	Guanine nucleotide-binding protein G(l)/G(S)/G(T) subunit beta-1	37353	2	22,1	0,045
244	KRT85	Keratin, type II cuticular Hb5	55766	9	31,4	0,045
245	CAH1	Carbonic anhydrase 1	28852	6	21,5	0,045
246	CBR1	Carbonyl reductase [NADPH] 1	30356	4	19,9	0,044
247	CTL4	Choline transporter-like protein 4	79201	7	6,1	0,044
248	GNAI2	Guanine nucleotide-binding protein G(i) subunit alpha-2	40425	3	25,6	0,044
249	NGAL	Neutrophil gelatinase-associated lipocalin	22574	9	55,1	0,044
250	MYO1A	Unconventional myosin-Ia	118326	35	33,9	0,044
251	SPB9	Serpin B9	42376	4	8	0,043
252	PGK1	Phosphoglycerate kinase 1	44586	16	27,1	0,043
253	GBB2	Guanine nucleotide-binding protein G(l)/G(S)/G(T) subunit beta-2	37307	3	24,1	0,043
254	FOLH1	Glutamate carboxypeptidase 2	84277	11	30,4	0,042
255	MGA	Maltase-glucoamylase, intestinal	209720	24	15,6	0,041
256	SAP3	Ganglioside GM2 activator	20825	6	28	0,041
257	PRTN3	Myeloblastin	27789	3	11,3	0,041
258	SUIS	Sucrase-isomaltase, intestinal	209320	51	30	0,041
259	ES8L3	Epidermal growth factor receptor kinase substrate 8-like protein 3	66819	17	23,4	0,041
260	PAI1	Plasminogen activator inhibitor 1	45031	10	13,9	0,040
261	DKK1	Dickkopf-related protein 1	28653	4	14,3	0,040
262	ADT2	ADP/ATP translocase 2	32831	2	23,8	0,040
263	GSHR	Glutathione reductase, mitochondrial	56221	8	14,9	0,040

264	SFTA1	Pulmonary surfactant-associated protein A1	26226	6	24,2	0,039
265	DHR11	Dehydrogenase/reductase SDR family member 11	28290	9	31,5	0,039
266	IST1	IST1 homolog	39725	6	15,4	0,039
267	FILA	Filaggrin	434922	82	20,9	0,038
268	HEXB	Beta-hexaminidase subunit beta	63071	14	22,1	0,038
269	ATPA	ATP synthase subunit alpha, mitochondrial	59714	12	21,2	0,038
270	NAPSA	Napsin-A	45358	4	9,5	0,038
271	PDC6I	Programmed cell death 6-interacting protein	95963	27	23,8	0,038
272	H10	Histone H1.0	20850	4	17,5	0,037
273	SLC31	Neutral and basic amino acid transport protein rBAT	78802	13	15	0,037
274	CO9	Complement component C9	63133	14	21,3	0,037
275	CYC	Cytochrome c	11741	3	31,4	0,037
276	GNAQ	Guanine nucleotide-binding protein G(q) subunit alpha	42115	3	8,9	0,035
277	MUC1	Mucin-1	122029	7	5,1	0,035
278	GRN	Granulins	63500	9	9,6	0,035
279	TGM1	Protein-glutamine gamma-glutamyltransferase K	89730	20	26,8	0,035
280	TFF1	Trefoil factor 1	9143	2	29,8	0,035
281	AK1A1	Alcohol dehydrogenase [NADP(+)]	36550	4	13,8	0,034
282	VDAC1	Voltage-dependent anion-selective channel protein 1	30754	9	41,7	0,034
283	GAPR1	Golgi-associated plant pathogenesis-related protein 1	17208	2	16,9	0,033
284	RAB5C	Ras-related protein Rab-5C	23468	3	22,7	0,033
285	RS13	40S ribosomal protein S13	17212	2	10,6	0,033
286	PLS1	Phospholipid scramblase 1	35026	8	29,6	0,032
287	CO3	Complement C3	187030	51	25,3	0,032
288	REG1A	Lithothine-1-alpha	18719	7	39,8	0,032
289	HSP7C	Heat shock cognate 71 kDa protein	70854	4	14,7	0,032
290	TGM3	Protein-glutamine gamma-glutamyltransferase E	76584	14	19,2	0,032
291	ARF6	ADP-ribosylation factor 6	20069	2	12	0,031
292	ALDOA	Fruite-bisphosphate aldolase A	39395	6	15,4	0,031
293	TPP1	Tripeptidyl-peptidase 1	61210	5	13,1	0,031

294	PLBL1	Phospholipase B-like 1	63214	8	16,5	0,030
295	HS71A	Heat shock 70 kDa protein 1A	70009	2	11,7	0,030
296	CEAM6	Carcinoembryonic antigen-related cell adhesion molecule 6	37172	3	16	0,030
297	NPM	Nucleophmin	32555	4	16,7	0,030
298	ARPC4	Actin-related protein 2/3 complex subunit 4	19654	4	22,6	0,030
299	A1AG1	Alpha-1-acid glycoprotein 1	23497	4	13,4	0,030
300	D103A	Beta-defensin 103	7692	2	22,4	0,030
301	UDB17	UDP-glucuronyltransferase 2B17	61055	5	17,4	0,030
302	ASAH2	Neutral ceramidase	85462	9	13,1	0,030
303	EHD1	EH domain-containing protein 1	60589	8	23	0,029
304	AMBP	Protein AMBP	38974	9	26,1	0,029
305	ATPB	ATP synthase subunit beta, mitochondrial	56525	10	19,3	0,029
306	CIB1	Calcium and integrin-binding protein 1	21690	6	24,1	0,028
307	CD9	CD9 antigen	25399	5	9,2	0,028
308	PSME1	Proteasome activator complex subunit 1	28705	3	13,3	0,028
309	EHD4	EH domain-containing protein 4	61137	7	24,8	0,027
310	S10AC	Protein S100-A12	10569	5	21,7	0,027
311	ITIH4	Inter-alpha-trypsin inhibitor heavy chain H4	103293	11	10,4	0,027
312	CAH2	Carbonic anhydrase 2	29228	4	13,5	0,027
313	LYAG	Lysosomal alpha-glucidase	105257	10	12,9	0,026
314	FILA2	Filaggrin-2	247928	40	18,7	0,026
315	IBP1	Insulin-like growth factor-binding protein 1	27885	5	19,7	0,026
316	THIO	Thioredoxin	11730	4	22,9	0,026
317	H15	Histone H1.5	22566	3	17,3	0,025
318	PARK7	Protein deglycase DJ-1	19878	2	7,4	0,025
319	HNRPC	Heterogeneous nuclear ribonucleoproteins C1/C2	33650	6	13,4	0,025
320	AGR2	Anterior gradient protein 2 homolog	19967	3	22,3	0,025
321	GGCT	Gamma-glutamylcyclotransferase	20994	7	42	0,025
322	ELAF	Elafin	12261	3	17,9	0,025
323	HUTH	Histidine ammonia-lyase	72652	17	23,4	0,025

324	KNG1	Kininogen-1	71912	7	5,6	0,024
325	CPVL	Probable serine carboxypeptidase CPVL	54129	5	9	0,024
326	GGT1	Gamma-glutamyltranspeptidase 1	61372	12	22,5	0,024
327	ANXA3	Annexin A3	36353	12	38,4	0,024
328	GSDMA	Gasdermin-A	49334	7	15,1	0,024
329	APOA1	Apolipoprotein A-I	30759	8	28,5	0,024
330	BDH2	3-hydroxybutyrate dehydrogenase type 2	26707	4	17,1	0,024
331	A2MG	Alpha-2-macroglobulin	163188	37	25,7	0,024
332	CALL5	Calmodulin-like protein 5	15883	3	21,9	0,023
333	CATC	Dipeptidyl peptidase 1	51820	6	14,9	0,023
334	RL9	60S ribosomal protein L9	21850	2	8,9	0,023
335	MYL6	Myosin light polypeptide 6	16919	2	27,2	0,023
336	PSA1	Proteasome subunit alpha type-1	29537	4	8,4	0,023
337	DR9C7	Short-chain dehydrogenase/reductase family 9C member 7	35240	9	23,3	0,023
338	CAH4	Carbonic anhydrase 4	35010	6	20,2	0,023
339	PRDX1	Peroxiredoxin-1	22096	4	15,6	0,022
340	ACY1	Aminoacylase-1	45856	11	26,7	0,022
341	VTDB	Vitamin D-binding protein	52929	9	15,8	0,021
342	ACE2	Angiotensin-converting enzyme 2	92404	17	21,9	0,021
343	KRT34	Keratin, type I cuticular Ha4	49392	2	18,6	0,021
344	CUZD1	CUB and zona pellucida-like domain-containing protein 1	68110	10	20,8	0,020
345	CHM4B	Charged multivesicular body protein 4b	24935	4	21,4	0,020
346	PPIB	Peptidyl-prolyl cis-trans isomerase B	23728	3	9,7	0,020
347	TRFM	Melanotransferrin	80163	16	20,9	0,020
348	CDC42	Cell division control protein 42 homolog	21245	4	20,9	0,020
349	KCRU	Creatine kinase U-type, mitochondrial	47007	7	16,8	0,020
350	CRNN	Cornulin	53502	3	5,1	0,019
351	A2ML1	Alpha-2-macroglobulin-like protein 1	161004	18	10,5	0,019
352	GPC5C	G-protein coupled receptor family C group 5 member C	48162	6	11,8	0,019
353	CAN1	Calpain-1 catalytic subunit	81838	13	12	0,019

354	CATA	Catalase	59719	14	23,3	0,019
355	PEBP1	Phphatidylethanolamine-binding protein 1	21044	3	15	0,018
356	AMPE	Glutamyl aminopeptidase	109176	16	11,9	0,018
357	RAB1A	Ras-related protein Rab-1A	22663	4	22,9	0,018
358	SC5A1	Sodium/glucose cotransporter 1	73449	11	13	0,018
359	SDCB1	Syntenin-1	32424	5	14,8	0,018
360	VIME	Vimentin	53619	3	6,7	0,017
361	GDIR1	Rho GDP-dissociation inhibitor 1	23193	3	11,3	0,017
362	K1C23	Keratin, type I cyckeletal 23	48101	7	23,9	0,017
363	LV743	Immunoglobulin lambda variable 7-43	12443	2	13,7	0,017
364	ACE	Angiotensin-converting enzyme	149620	26	14,4	0,017
365	PRDX2	Peroxiredoxin-2	21878	5	22,7	0,017
366	PSA4	Proteasome subunit alpha type-4	29465	3	10,3	0,017
367	CADH1	Cadherin-1	97396	7	6,5	0,017
368	STX3	Syntaxin-3	33134	6	13,5	0,017
369	CHM1B	Charged multivesicular body protein 1b	22095	2	8	0,017
370	MUC5A	Mucin-5AC	585199	24	3,5	0,017
371	BLMH	Bleomycin hydrolase	52528	11	13,4	0,017
372	RRAS2	Ras-related protein R-Ras2	23385	2	14,7	0,017
373	WNT7A	Protein Wnt-7a	38979	2	3,4	0,017
374	AMY1	Alpha-amylase 1	57731	2	19,2	0,016
375	ACTN4	Alpha-actinin-4	104788	8	15,8	0,016
376	CDHR5	Cadherin-related family member 5	88169	9	10,5	0,016
377	ARRD1	Arrestin domain-containing protein 1	45953	5	6,5	0,016
378	RAB10	Ras-related protein Rab-10	22527	2	9	0,016
379	GNAS1	Guanine nucleotide-binding protein G(s) subunit alpha isoforms XLas	110956	2	3,9	0,016
380	LIPP	Pancreatic triacylglycerol lipase	51124	6	15,1	0,015
381	MA2B1	Lysosomal alpha-mannidase	113672	9	6,1	0,015
382	UTRO	Utrophin	394220	2	0,5	0,015
383	RALB	Ras-related protein Ral-B	23394	3	9,7	0,015

384	CAN5	Calpain-5	73123	6	8,1	0,015
385	RISC	Retinoid-inducible serine carboxypeptidase	50798	7	11,3	0,015
386	PLB1	Phospholipase B1, membrane-associated	162977	12	9	0,015
387	CNDP2	Cytolic non-specific dipeptidase	52845	7	8,4	0,015
388	ANGT	Angiotensinogen	53121	4	10,5	0,015
389	SIAE	Sialate O-acetylerase	58277	5	8,6	0,015
390	FSTL1	Follistatin-related protein 1	34963	4	10,1	0,014
391	CE290	Centromal protein of 290 kDa	290207	3	0,6	0,014
392	GPX3	Glutathione peroxidase 3	25536	3	18,1	0,014
393	ST2A1	Bile salt sulfotransferase	33758	6	11,9	0,014
394	HSPB1	Heat shock protein beta-1	22768	2	13,2	0,014
395	LIPR2	Pancreatic lipase-related protein 2	51913	2	6,8	0,014
396	IF5A1	Eukaryotic translation initiation factor 5A-1	16821	2	12,3	0,014
397	RS25	40S ribosomal protein S25	13734	2	15,2	0,014
398	CLD3	Claudin-3	23303	2	9,1	0,014
399	CD47	Leukocyte surface antigen CD47	35191	3	8,7	0,013
400	ES8L1	Epidermal growth factor receptor kinase substrate 8-like protein 1	80202	2	2,5	0,013
401	G6PI	Glucose-6-phosphate isomerase	63107	8	10,2	0,013
402	S10AG	Protein S100-A16	11794	3	29,1	0,013
403	QPCT	Glutamyl-peptide cyclotransferase	40851	3	10,2	0,013
404	RAB8A	Ras-related protein Rab-8A	23653	2	17,9	0,013
405	THTR	Thiulfate sulfurtransferase	33408	3	12,1	0,013
406	SNG2	Synaptogyrin-2	24794	2	8	0,013
407	PAI2	Plasminogen activator inhibitor 2	46566	4	16,9	0,013
408	RS11	40S ribosomal protein S11	18419	2	5,1	0,013
409	IF4A1	Eukaryotic initiation factor 4A-I	46125	2	4,7	0,013
410	MYO1D	Unconventional myosin-Id	116129	16	14,4	0,013
411	RS3	40S ribosomal protein S3	26671	4	15,2	0,013
412	GPA33	Cell surface A33 antigen	35609	2	6	0,013
413	ASSY	Argininosuccinate synthase	46501	7	5,6	0,013

414	PSA3	Proteasome subunit alpha type-3	28415	6	22,7	0,013
415	KAP2	cAMP-dependent protein kinase type II-alpha regulatory subunit	45490	6	12,6	0,013
416	MYO7B	Unconventional myosin-VIIb	241445	43	13,5	0,013
417	BPIB4	BPI fold-containing family B member 4	65013	4	5,5	0,013
418	PSA6	Proteasome subunit alpha type-6	27382	4	12,2	0,013
419	BAIP2	Brain-specific angiogenesis inhibitor 1-associated protein 2	60830	7	7,1	0,013
420	PRG3	Proteoglycan 3	25389	3	8,4	0,012
421	LIPR1	Inactive pancreatic lipase-related protein 1	51815	3	5,4	0,012
422	TAGL2	Transgelin-2	22377	3	15,1	0,012
423	RS18	40S ribosomal protein S18	17708	2	13,2	0,012
424	PDIA3	Protein disulfide-isomerase A3	56747	7	10,1	0,012
425	PRDX3	Thioredoxin-dependent peroxide reductase, mitochondrial	27675	5	18	0,012
426	GALNS	N-acetylgalactamine-6-sulfatase	57989	3	5,4	0,012
427	FAAA	Fumarylacetoacetase	46344	8	11,2	0,012
428	EPS8	Epidermal growth factor receptor kinase substrate 8	91824	15	12,4	0,012
429	RNT2	Ribonuclease T2	29462	2	3,9	0,012
430	ARSF	Arylsulfatase F	65897	7	6,1	0,012
431	PCBP1	Poly(rC)-binding protein 1	37474	5	16,3	0,011
432	CALL4	Calmodulin-like protein 4	21869	4	19,9	0,011
433	K1C20	Keratin, type I cytoskeletal 20	48457	2	3,8	0,011
434	PBLD	Phenazine biosynthesis-like domain-containing protein	31765	6	13,5	0,011
435	CLIC3	Chloride intracellular channel protein 3	26632	2	8,5	0,011
436	GRAN	Grancalcin	23994	5	27,2	0,011
437	APOB	Apolipoprotein B-100	515283	43	6,3	0,011
438	ITB1	Integrin beta-1	88357	8	6	0,011
439	SAPL1	Proactivator polypeptide-like 1	56589	3	7,3	0,011
440	UPK3L	Uroplakin-3b-like protein	28368	2	11	0,011
441	VDAC2	Voltage-dependent anion-selective channel protein 2	31547	7	29,6	0,010
442	MYH14	Myosin-14	227732	15	8,1	0,010
443	PSA5	Proteasome subunit alpha type-5	26394	4	20,3	0,010

444	GMDS	GDP-manne 4,6 dehydratase	41923	3	7,3	0,010
445	PSB6	Proteasome subunit beta type-6	25341	3	13	0,010
446	CPNS1	Calpain small subunit 1	28298	3	12,3	0,010
447	MPCP	Phosphate carrier protein, mitochondrial	40069	2	4,1	0,010
448	RNAS6	Ribonuclease K6	17185	2	12	0,010
449	USH1C	Harmonin	62172	8	12,1	0,010
450	GDIB	Rab GDP dissociation inhibitor beta	50631	10	14,4	0,010
451	GKN2	Gastrokine-2	20474	3	14,7	0,010
452	TS101	Tumor susceptibility gene 101 protein	43916	3	7,4	0,010
453	VGFR1	Vascular endothelial growth factor receptor 1	150672	7	4,4	0,010
454	AQP1	Aquaporin-1	28508	2	3,3	0,010
455	TCPB	T-complex protein 1 subunit beta	57452	2	4,3	0,009
456	ELL	RNA polymerase II elongation factor ELL	68223	2	2,3	0,009
457	F151A	Protein FAM151A	63987	6	8,5	0,009
458	VPS4B	Vacuolar protein sorting-associated protein 4B	49271	2	5,2	0,009
459	CTL2	Choline transporter-like protein 2	80071	2	1,1	0,009
460	CLIC5	Chloride intracellular channel protein 5	46474	4	11,7	0,009
461	TKFC	Triokinase/FMN cyclase	58910	6	10,1	0,009
462	WDR1	WD repeat-containing protein 1	66152	12	14,7	0,009
463	IDHC	Isocitrate dehydrogenase [NADP] cytoplasmic	46630	8	8,7	0,009
464	GFRP	GTP cyclohydrolase 1 feedback regulatory protein	9692	2	9,5	0,009
465	QSOX1	Sulfhydryl oxidase 1	82525	6	4,6	0,009
466	STXB2	Syntaxin-binding protein 2	66410	9	9,1	0,009
467	AEGP	Apical endosomal glycoprotein	131416	13	7,9	0,009
468	S26A3	Chloride anion exchanger	84451	12	12,7	0,009
469	CPNE3	Copine-3	60092	12	18,8	0,009
470	RLA0L	60S acidic ribosomal protein P0-like	34343	2	7,9	0,009
471	TSP1	Thrombospondin-1	129300	16	8,3	0,009
472	LX12B	Arachidonate 12-lipoxygenase, 12R-type	80304	13	13,1	0,009
473	DLDH	Dihydrolipoyl dehydrogenase, mitochondrial	54143	6	13,8	0,009

474	RET4	Retinol-binding protein 4	22995	7	24,9	0,009
475	TSN1	Tetraspanin-1	26283	2	8,7	0,009
476	LPH	Lactase-phlorizin hydrolase	218450	29	11,5	0,009
477	RHCG	Ammonium transporter Rh type C	53144	2	4,2	0,009
478	NHLC3	NHL repeat-containing protein 3	38258	3	7,5	0,009
479	RAB25	Ras-related protein Rab-25	23481	3	8,5	0,009
480	GRP78	78 kDa glucose-regulated protein	72288	4	4,9	0,009
481	ANXA7	Annexin A7	52706	7	13,7	0,008
482	DIAC	Di-N-acetylchitobiase	43732	5	12,7	0,008
483	SDCB2	Syntenin-2	31574	7	20,2	0,008
484	TM11B	Transmembrane protease serine 11B	46307	4	7,7	0,008
485	COX7C	Cytochrome c oxidase subunit 7C, mitochondrial	7241	2	41,3	0,008
486	S26A2	Sulfate transporter	81609	6	8,1	0,008
487	SAHH	Adenylyhomocysteinase	47685	2	4,6	0,008
488	CAPZB	F-actin-capping protein subunit beta	31331	2	8,7	0,008
489	PLP2	Proteolipid protein 2	16680	2	9,9	0,008
490	CHP2	Calcineurin B homologous protein 2	22438	2	9,7	0,008
491	SNX3	Sorting nexin-3	18751	2	9,9	0,008
492	RAN	GTP-binding nuclear protein Ran	24408	2	9,7	0,008
493	COMP	Cartilage oligomeric matrix protein	82808	3	4	0,008
494	TXTP	Tricarboxylate transport protein, mitochondrial	33991	3	7,4	0,008
495	MDHC	Malate dehydrogenase, cytoplasmic	36403	6	13,5	0,007
496	GTR1	Solute carrier family 2, facilitated glucose transporter member 1	54049	2	3,7	0,007
497	B2MG	Beta-2-microglobulin	13706	2	16,8	0,007
498	CD36	Platelet glycoprotein 4	53019	6	12,1	0,007
499	HS90A	Heat shock protein HSP 90-alpha	84607	3	10,1	0,007
500	GSTO1	Glutathione S-transferase omega-1	27548	3	7,1	0,007
501	NLRP6	NACHT, LRR and PYD domains-containing protein 6	98705	14	12,2	0,007
502	TALDO	Transaldolase	37516	3	8,6	0,007
503	GELS	Gelsolin	85644	7	6,9	0,007

504	TSN8	Tetraspanin-8	26026	2	4,2	0,007
505	CDD	Cytidine deaminase	16174	3	32,2	0,007
506	GSTM3	Glutathione S-transferase Mu 3	26542	3	12,9	0,007
507	FINC	Fibronectin	262460	18	5,4	0,007
508	CHM2A	Charged multivesicular body protein 2a	25088	2	4,1	0,007
509	RAB35	Ras-related protein Rab-35	23011	2	5,5	0,007
510	MEP1B	Meprin A subunit beta	79521	8	12,8	0,007
511	CBPB2	Carboxypeptidase B2	48393	3	3,8	0,007
512	NHRF3	Na(+)/H(+) exchange regulatory cofactor NHE-RF3	57094	7	7,5	0,007
513	LOXE3	Hydroperoxide isomerase ALOXE3	80491	7	11	0,007
514	LASP1	LIM and SH3 domain protein 1	29698	4	16,1	0,007
515	2B11	HLA class II histocompatibility antigen, DRB1-1 beta chain	29895	3	9	0,007
516	BAT1	b(0,+)-type amino acid transporter 1	53446	2	3,1	0,007
517	MY15B	Unconventional myin-XVB	167013	18	9	0,007
518	FETUA	Alpha-2-HS-glycoprotein	39300	2	3,8	0,007
519	PEDF	Pigment epithelium-derived factor	46283	5	9,6	0,007
520	MTP	Micromal triglyceride transfer protein large subunit	99289	7	8,2	0,007
521	RS9	40S ribomal protein S9	22578	2	8,8	0,006
522	A1BG	Alpha-1B-glycoprotein	54220	3	5,3	0,006
523	ST1A1	Sulfotransferase 1A1	34143	5	11,2	0,006
524	MYH9	Myin-9	226392	6	4	0,006
525	AK1C1	Aldo-keto reductase family 1 member C1	36765	2	5	0,006
526	MUC5B	Mucin-5B	595960	20	3,9	0,006
527	FRK	Tyrine-protein kinase FRK	58217	5	7,5	0,006
528	FUCO	Tissue alpha-L-fucidase	53655	5	9,7	0,006
529	CAB39	Calcium-binding protein 39	39844	4	12,6	0,006
530	PSA2	Proteasome subunit alpha type-2	25882	3	10,3	0,006
531	S15A1	Solute carrier family 15 member 1	78754	6	7,1	0,006
532	MFGM	Lactadherin	43095	6	16,5	0,006
533	PYGM	Glycogen phphorylase, muscle form	97031	3	5,1	0,006

534	SCRB1	Scavenger receptor class B member 1	60838	5	9,1	0,006
535	PLCX1	PI-PLC X domain-containing protein 1	36644	5	14,2	0,006
536	RS15A	40S ribosomal protein S15a	14830	2	11,5	0,006
537	PLAK	Junction plakoglobin	81693	8	7,9	0,006
538	ABHEB	Protein ABHD14B	22332	3	11,4	0,006
539	PERL	Lactoperoxidase	80237	6	6,2	0,006
540	MNDA	Myeloid cell nuclear differentiation antigen	45807	8	16,2	0,006
541	TMC4	Transmembrane channel-like protein 4	79157	4	5,5	0,006
542	RL37A	60S ribosomal protein L37a	10268	2	17,4	0,006
543	FBX50	F-box only protein 50	30828	4	11,6	0,006
544	MGAL	Probable maltase-glucoamylase 2	277816	4	3	0,006
545	BI2L1	Brain-specific angiogenesis inhibitor 1-associated protein 2-like protein 1	56847	4	7,2	0,006
546	ATPO	ATP synthase subunit O, mitochondrial	23263	3	14,6	0,006
547	ANS4B	Ankyrin repeat and SAM domain-containing protein 4B	46568	6	9,8	0,006
548	STK24	Serine/threonine-protein kinase 24	49277	4	7,4	0,006
549	EFTU	Elongation factor Tu, mitochondrial	49510	10	21	0,006
550	RAB7A	Ras-related protein Rab-7a	23475	5	11,1	0,006
551	EF2	Elongation factor 2	95277	10	5,6	0,005
552	JAM1	Junctional adhesion molecule A	32562	4	7,7	0,005
553	6PGD	6-phosphogluconate dehydrogenase, decarboxylating	53106	4	3,9	0,005
554	CLH1	Clathrin heavy chain 1	191492	14	6,3	0,005
555	PSB5	Proteasome subunit beta type-5	28462	6	20,2	0,005
556	ATP5L	ATP synthase subunit g, mitochondrial	11421	2	21,4	0,005
557	GDIR2	Rho GDP-dissociation inhibitor 2	22974	4	16,4	0,005
558	AT1B1	Sodium/potassium-transporting ATPase subunit beta-1	35039	3	10,6	0,005
559	PKP1	Plakophilin-1	82808	6	8	0,005
560	ENPL	Endoplasmin	92411	8	5,6	0,005
561	ITA3	Integrin alpha-3	116538	5	2,5	0,005
562	THIM	3-ketoacyl-CoA thiolase, mitochondrial	41898	3	7,6	0,005

563	PCKGM	Phphoenolpyruvate carboxykinase [GTP], mitochondrial	70685	11	18,8	0,005
564	SODM	Superoxide dismutase [Mn], mitochondrial	24707	2	10,4	0,005
565	FBN1	Fibrillin-1	312022	16	5,3	0,005
566	CATB	Cathepsin B	37797	4	11,5	0,005
567	AMPB	Aminopeptidase B	72549	3	4,8	0,005
568	CBPA4	Carboxypeptidase A4	47322	3	6,2	0,005
569	ACSL5	Long-chain-fatty-acid--CoA ligase 5	75942	11	18,6	0,005
570	SMS	Somattatin	12727	2	11,2	0,005
571	CK054	Ester hydrolase C11orf54	35095	6	16,5	0,004
572	TGM2	Protein-glutamine gamma-glutamyltransferase 2	77280	9	13,4	0,004
573	AT1A1	Sodium/potassium-transporting ATPase subunit alpha-1	112824	13	17,4	0,004
574	CAP1	Adenylyl cyclase-associated protein 1	51869	6	10,5	0,004
575	SPON2	Spondin-2	35824	3	4,8	0,004
576	APT	Adenine phphoribyltransferase	19595	2	10	0,004
577	MGST3	Micromal glutathione S-transferase 3	16506	2	14,5	0,004
578	HV270	Immunoglobulin heavy variable 2-70	13252	2	11,8	0,004
579	SQRD	Sulfide:quinone oxidoreductase, mitochondrial	49929	5	12,9	0,004
580	5NTD	5'-nucleotidase	63327	6	9,9	0,004
581	F16P1	Fructe-1,6-bisphphatase 1	36819	4	16	0,004
582	ITA1	Integrin alpha-1	130765	5	2,6	0,004
583	MDR1	Multidrug resistance protein 1	141389	10	7,2	0,004
584	VTNC	Vitronectin	54271	3	5,2	0,004
585	PADI2	Protein-arginine deiminase type-2	75515	10	16,8	0,004
586	F10A1	Hsc70-interacting protein	41305	3	4,6	0,004
587	DSG1	Desmoglein-1	113676	6	5,1	0,004
588	LGUL	Lactoylglutathione lyase	20764	2	3,8	0,004
589	PLSI	Plastin-1	70209	5	7,8	0,004
590	KRA31	Keratin-associated protein 3-1	10532	2	24,5	0,004
591	HEBP1	Heme-binding protein 1	21084	2	10,6	0,004
592	ATRN	Attractin	158432	11	6,4	0,004

593	PHLD	Phphatidylinitol-glycan-specific phpholipase D	92278	2	2,1	0,004
594	ENPP3	Ectonucleotide pyrophphatase/phphodiesterase family member 3	100059	7	9,4	0,004
595	ANXA6	Annexin A6	75826	6	6,5	0,004
596	EPDR1	Mammalian ependymin-related protein 1	25421	3	9,4	0,004
597	ANGL7	Angiopietin-related protein 7	39993	2	4	0,004
598	MTEF4	Transcription termination factor 4, mitochondrial	43930	2	2,6	0,004
599	DJC14	DnaJ homolog subfamily C member 14	78519	2	2,1	0,004
600	AGAL	Alpha-galactidase A	48735	4	4,4	0,004
601	PROF2	Profilin-2	15036	2	15,7	0,004
602	ACOC	Cytoplasmic aconitate hydratase	98337	4	4,3	0,004
603	KGP2	cGMP-dependent protein kinase 2	87377	8	5	0,004
604	PSB2	Proteasome subunit beta type-2	22822	2	9	0,004
605	CH60	60 kDa heat shock protein, mitochondrial	61016	5	7,3	0,004
606	CAD17	Cadherin-17	92162	11	16,2	0,004
607	XDH	Xanthine dehydrogenase/oxidase	146330	11	6,5	0,004
608	PSA7	Proteasome subunit alpha type-7	27870	2	9,3	0,004
609	AL3B1	Aldehyde dehydrogenase family 3 member B1	51807	4	7,3	0,004
610	RS19	40S ribomal protein S19	16051	2	6,2	0,004
611	RL6	60S ribomal protein L6	32708	2	8	0,004
612	LEG10	Galectin-10	16442	2	4,9	0,003
613	LMNA	Prelamin-A/C	74095	2	2,6	0,003
614	AMD	Peptidyl-glycine alpha-amidating monooxygenase	108263	3	2,7	0,003
615	NALD2	N-acetylated-alpha-linked acidic dipeptidase 2	83539	2	3	0,003
616	GRB2	Growth factor receptor-bound protein 2	25190	2	7,4	0,003
617	CBPM	Carboxypeptidase M	50481	2	3,8	0,003
618	CPSM	Carbamoyl-phphate synthase [ammonia], mitochondrial	164835	9	6,1	0,003
619	CBPO	Carboxypeptidase O	42502	4	12,8	0,003
620	CO4A	Complement C4-A	192664	15	6	0,003
621	LXN	Latexin	25734	2	8,6	0,003
622	TOLIP	Toll-interacting protein	30262	2	6,2	0,003

623	KCTD5	BTB/POZ domain-containing protein KCTD5	26076	2	5,1	0,003
624	EPCAM	Epithelial cell adhesion molecule	34910	5	15,6	0,003
625	BPIB1	BPI fold-containing family B member 1	52408	6	8,9	0,003
626	ADEC1	ADAM DEC1	52742	3	7,4	0,003
627	NADC	Nicotinate-nucleotide pyrophosphorylase [carboxylating]	30826	4	12,1	0,003
628	RENBP	N-acylglucamine 2-epimerase	48799	2	4,4	0,003
629	ALAT2	Alanine aminotransferase 2	57866	2	2,9	0,003
630	CLD7	Claudin-7	22403	2	15,6	0,003
631	HPT	Haptoglobin	45177	2	12,1	0,003
632	PGM1	Phosphoglucomutase-1	61411	5	3	0,003
633	RGN	Regucalcin	33231	4	14,4	0,003
634	METK2	S-adenylmethionine synthase isoform type-2	43633	2	3,8	0,003
635	RL13	60S ribosomal protein L13	24247	2	5,2	0,003
636	NIBL1	Niban-like protein 1	84085	4	4,6	0,003
637	RS3A	40S ribosomal protein S3a	29926	3	6,4	0,003
638	MYO1B	Unconventional myosin-Ib	131902	2	4,1	0,003
639	ANAG	Alpha-N-acetylglucosaminidase	82214	5	3,6	0,003
640	SUCB2	Succinate--CoA ligase [GDP-forming] subunit beta, mitochondrial	46481	3	3,7	0,003
641	PRSS8	Proteinase 8	36408	2	6,1	0,003
642	TMC5	Transmembrane channel-like protein 5	114724	7	4,6	0,003
643	MYO1C	Unconventional myosin-Ic	121606	10	5,1	0,003
644	VILL	Villin-like protein	95847	3	3,2	0,003
645	DDC	Aromatic-L-amino-acid decarboxylase	53891	4	3,8	0,003
646	SCRB2	Lysosomal membrane protein 2	54255	3	4	0,003
647	PEPD	Xaa-Pro dipeptidase	54513	5	9,5	0,003
648	APOH	Beta-2-glycoprotein 1	38273	3	9,3	0,003
649	COR1A	Coronin-1A	50994	4	9,1	0,003
650	KLH41	Kelch-like protein 41	67994	2	2,6	0,003
651	ECHB	Trifunctional enzyme subunit beta, mitochondrial	51262	3	5,7	0,003
652	KLK1	Kallikrein-1	28871	2	8	0,002

653	AATC	Aspartate aminotransferase, cytoplasmic	46219	4	9,4	0,002
654	GSHB	Glutathione synthetase	52352	2	3,8	0,002
655	DNPEP	Aspartyl aminopeptidase	52395	3	9,3	0,002
656	A20A1	Ankyrin repeat domain-containing protein 20A1	93989	2	0,9	0,002
657	RAB2A	Ras-related protein Rab-2A	23531	2	6,1	0,002
658	IL1AP	Interleukin-1 receptor accessory protein	65377	2	1,6	0,002
659	TERA	Transitional endoplasmic reticulum ATPase	89266	3	2,1	0,002
660	TCPQ	T-complex protein 1 subunit theta	59582	2	1,3	0,002
661	DPP2	Dipeptidyl peptidase 2	54307	3	7,5	0,002
662	HNRPU	Heterogeneous nuclear ribonucleoprotein U	90528	6	4,8	0,002
663	PYGB	Glycogen phosphorylase, brain form	96635	2	2,6	0,002
664	AL1A1	Retinal dehydrogenase 1	54827	4	7,2	0,002
665	ROA2	Heterogeneous nuclear ribonucleoproteins A2/B1	37407	3	6,5	0,002
666	SC5A6	Sodium-dependent multivitamin transporter	68596	3	5,5	0,002
667	EF1G	Elongation factor 1-gamma	50087	2	2,3	0,002
668	IQGA1	Ras GTPase-activating-like protein IQGAP1	189134	8	4,3	0,002
669	PLST	Plastin-3	70766	2	3,8	0,002
670	LIPG	Gastric triacylglycerol lipase	45209	2	5	0,002
671	GFPT1	Glutamine--fructose-6-phosphate aminotransferase [isomerizing] 1	78756	4	3,7	0,002
672	RL3	60S ribosomal protein L3	46080	2	1,7	0,002
673	TCPH	T-complex protein 1 subunit eta	59329	2	2,6	0,002
674	ARK72	Aflatoxin B1 aldehyde reductase member 2	39564	2	9,5	0,002
675	ITIH1	Inter-alpha-trypsin inhibitor heavy chain H1	101326	2	2,9	0,002
676	ADA	Adenine deaminase	40739	4	6,9	0,002
677	CAYP1	Calcyphin	20954	2	6,9	0,002
678	PDIA1	Protein disulfide-isomerase	57081	6	8,3	0,002
679	FRPD1	FERM and PDZ domain-containing protein 1	173328	2	0,9	0,002
680	ERCC2	TFIIH basal transcription factor complex helicase XPD subunit	86854	2	0,9	0,002
681	ZPI	Protein Z-dependent protease inhibitor	50674	2	4,1	0,002

682	SUCA	Succinate--CoA ligase [ADP/GDP-forming] subunit alpha, mitochondrial	36227	2	7,2	0,002
683	HNRPK	Heterogeneous nuclear ribonucleoprotein K	50944	3	6,3	0,002
684	JPH4	Junctophilin-4	65821	2	2,2	0,002
685	RS4X	40S ribosomal protein S4, X isoform	29579	2	6,5	0,002
686	OLA1	Obg-like ATPase 1	44715	2	6,6	0,002
687	COMT	Catechol O-methyltransferase	30018	3	8,5	0,002
688	MUC6	Mucin-6	256887	2	0,7	0,002
689	APOA4	Apolipoprotein A-IV	45371	2	4,5	0,002
690	FABP5	Fatty acid-binding protein, epidermal	15155	2	18,5	0,002
691	LEG1H	Protein LEG1 homolog	37902	2	5,8	0,002
692	FETA	Alpha-fetoprotein	68633	2	2,6	0,002
693	A2GL	Leucine-rich alpha-2-glycoprotein	38154	3	7,2	0,002
694	PTPRH	Receptor-type tyrosine-protein phosphatase H	122276	2	2,1	0,002
695	TMCC3	Transmembrane and coiled-coil domains protein 3	53752	2	1,9	0,002
696	KPCD	Protein kinase C delta type	77455	4	3,3	0,002
697	NAGAB	Alpha-N-acetylgalactaminidase	46534	3	2,7	0,002
698	PLTP	Phospholipid transfer protein	54705	2	4,3	0,002
699	RAB14	Ras-related protein Rab-14	23882	2	3,7	0,002
700	ENTK	Enteropeptidase	112863	3	2,8	0,002
701	SDHA	Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial	72645	4	4,8	0,002
702	AGRG7	Adhesion G-protein coupled receptor G7	88852	4	4,9	0,002
703	DX39A	ATP-dependent RNA helicase DDX39A	49098	2	4,2	0,002
704	CHMP5	Charged multivesicular body protein 5	24555	3	4,6	0,002
705	SC5A9	Sodium/glucose cotransporter 4	74025	2	3,1	0,002
706	PLMN	Plasminogen	90510	2	2,3	0,002
707	SPB5	Serpin B5	42073	2	3,2	0,002
708	FIBG	Fibrinogen gamma chain	51479	2	2,6	0,002
709	TENA	Tenascin	240700	9	5,1	0,002

710	FIBA	Fibrinogen alpha chain	94914	4	4,3	0,002
711	FR1L6	Fer-1-like protein 6	209176	11	4,4	0,002
712	WSDU1	WD repeat, SAM and U-box domain-containing protein 1	52783	2	1,5	0,002
713	ACADS	Short-chain specific acyl-CoA dehydrogenase, mitochondrial	44269	2	4,6	0,002
714	ITIH2	Inter-alpha-trypsin inhibitor heavy chain H2	106397	2	2,2	0,002
715	QORX	Quinone oxidoreductase PIG3	35514	2	3	0,002
716	S6A19	Sodium-dependent neutral amino acid transporter B(0)AT1	71063	2	3,8	0,002
717	MUC4	Mucin-4	231373	6	3	0,002
718	AP2A1	AP-2 complex subunit alpha-1	107478	2	2,8	0,002
719	AK1BA	Aldo-keto reductase family 1 member B10	35997	3	3,5	0,002
720	SUSD2	Sushi domain-containing protein 2	90150	3	3,3	0,002
721	PPGB	Lysosomal protective protein	54431	3	4	0,002
722	PSME2	Proteasome activator complex subunit 2	27384	3	3,8	0,002
723	ETHE1	Persulfide dioxygenase ETHE1, mitochondrial	27855	3	11	0,002
724	PEPL1	Probable aminopeptidase NPEPL1	55825	2	4,6	0,002
725	CO7	Complement component C7	93457	4	2,1	0,002
726	CATS	Cathepsin S	37471	3	11,8	0,002
727	GPDA	Glycerol-3-phosphate dehydrogenase [NAD(+)], cytoplasmic	37543	2	6,9	0,002
728	ITB2	Integrin beta-2	84726	6	7,4	0,002
729	VINC	Vinculin	123722	7	5,5	0,002
730	GALE	UDP-glucose 4-epimerase	38257	2	6	0,002
731	MB12A	Multivesicular body subunit 12A	28765	2	5,5	0,001
732	CAPG	Macrophage-capping protein	38474	2	3,7	0,001
733	OAT	Ornithine aminotransferase, mitochondrial	48504	3	4,3	0,001
734	ERAP1	Endoplasmic reticulum aminopeptidase 1	107166	2	1,1	0,001
735	IMPA1	Initol monophosphatase 1	30169	2	7,9	0,001
736	DHE3	Glutamate dehydrogenase 1, mitochondrial	61359	3	1,4	0,001
737	RS2	40S ribosomal protein S2	31305	3	4,4	0,001
738	CTL1	Choline transporter-like protein 1	73253	3	1,8	0,001
739	MISP	Mitotic interactor and substrate of PLK1	75311	3	4,4	0,001

740	CO1A2	Collagen alpha-2(I) chain	129235	6	3,7	0,001
741	LOX12	Arachidonate 12-lipoxygenase, 12S-type	75645	5	7,1	0,001
742	KHK	Ketohexokinase	32502	2	4,7	0,001
743	DYN2	Dynamamin-2	98003	2	3,7	0,001
744	PDCD6	Programmed cell death protein 6	21855	2	5,2	0,001
745	ATPG	ATP synthase subunit gamma, mitochondrial	32975	3	8,1	0,001
746	PCTL	PCTP-like protein	33028	2	4,5	0,001
747	PLPL8	Calcium-independent phospholipase A2-gamma	88421	2	2,2	0,001
748	ARP2	Actin-related protein 2	44732	3	7,6	0,001
749	MANBA	Beta-mannidase	100831	4	5,3	0,001
750	POF1B	Protein POF1B	68022	2	3,6	0,001
751	BASP1	Brain acid soluble protein 1	22680	2	11,9	0,001
752	CAND1	Cullin-associated NEDD8-dissociated protein 1	136289	4	2	0,001
753	UGPA	UTP--glucose-1-phosphate uridylyltransferase	56905	4	3,5	0,001
754	HCDH	Hydroxyacyl-coenzyme A dehydrogenase, mitochondrial	34272	3	8,6	0,001
755	AL1B1	Aldehyde dehydrogenase X, mitochondrial	57170	3	7,5	0,001
756	SL9A3	Sodium/hydrogen exchanger 3	92796	4	4,1	0,001
757	BROX	BRO1 domain-containing protein BROX	46447	3	3,4	0,001
758	SCNBA	Sodium channel protein type 11 subunit alpha	204788	2	0,4	0,001
759	HEM2	Delta-aminolevulinic acid dehydratase	36271	2	6,1	0,001
760	QCR2	Cytochrome b-c1 complex subunit 2, mitochondrial	48413	2	5,3	0,001
761	LIMA1	LIM domain and actin-binding protein 1	85173	2	2,2	0,001
762	ACON	Aconitate hydratase, mitochondrial	85372	6	5,6	0,001
763	ANXA8	Annexin A8	36858	2	6,1	0,001
764	SC22B	Vesicle-trafficking protein SEC22b	24578	2	10,2	0,001
765	PCFT	Proton-coupled folate transporter	49738	2	2,8	0,001
766	DPYL2	Dihydropyrimidinase-related protein 2	62255	2	3	0,001
767	SFTPD	Pulmonary surfactant-associated protein D	37705	2	7,7	0,001
768	RHG01	Rho GTPase-activating protein 1	50404	2	1,6	0,001
769	GMPR2	GMP reductase 2	37850	3	3,4	0,001

770	ARLY	Argininuccinate lyase	51625	2	5,6	0,001
771	S23A1	Solute carrier family 23 member 1	64789	2	3,2	0,001
772	PTER	Phphotriesterase-related protein	38993	2	4	0,001
773	PLBL2	Putative phpholipase B-like 2	65430	3	2,5	0,001
774	QCR1	Cytochrome b-c1 complex subunit 1, mitochondrial	52612	2	2,5	0,001
775	MOD5	tRNA dimethylallyltransferase, mitochondrial	52692	2	1,5	0,001
776	LRC57	Leucine-rich repeat-containing protein 57	26737	2	5,4	0,001
777	CO8B	Complement component C8 beta chain	67003	2	2,5	0,001
778	TKT	Transketolase	67835	3	2,2	0,001
779	G6PT3	Glucose-6-phosphate exchanger SLC37A2	54401	2	2,2	0,001
780	IQCA1	IQ and AAA domain-containing protein 1	95281	2	1,9	0,001
781	PLD3	Phpholipase D3	54670	2	3,7	0,001
782	FIBB	Fibrinogen beta chain	55892	2	4,1	0,001
783	HDHD3	Haloacid dehalogenase-like hydrolase domain-containing protein 3	27982	2	2,8	0,001
784	ERMP1	Endoplasmic reticulum metalloproteinase 1	100167	3	2	0,001
785	DERL1	Derlin-1	28782	2	3,2	0,001
786	GRAA	Granzyme A	28980	2	3,8	0,001
787	RL7	60S ribosomal protein L7	29207	2	4,4	0,001
788	NLTP	Non-specific lipid-transfer protein	58956	2	3,5	0,001
789	PSA	Puromycin-sensitive aminopeptidase	103211	4	4,4	0,001
790	PEF1	Peflin	30361	2	3,9	0,001
791	URP2	Fermitin family homolog 3	75905	5	9,1	0,001
792	COPA	Coatomer subunit alpha	138258	3	2,1	0,001
793	CO1A1	Collagen alpha-1(I) chain	138857	2	0,8	0,001
794	SPA12	Serpin A12	47145	3	2,9	0,001
795	FBLN2	Fibulin-2	126489	4	2,5	0,001
796	FGD6	FYVE, RhoGEF and PH domain-containing protein 6	160714	2	1,2	0,001
797	LRRF2	Leucine-rich repeat flightless-interacting protein 2	82121	2	1	0,001
798	DESP	Desmoplakin	331569	12	1,9	0,001
799	ADA10	Disintegrin and metalloproteinase domain-containing protein 10	84088	3	3,6	0,001

800	CATZ	Cathepsin Z	33846	2	7,3	0,001
801	S39A4	Zinc transporter ZIP4	68348	2	3,7	0,001
802	CO5	Complement C5	188186	3	1,3	0,001
803	DYHC1	Cytoplasmic dynein 1 heavy chain 1	532072	12	1,5	0,001
804	ST1B1	Sulfotransferase family cytoloc 1B member 1	34877	2	3	0,001
805	CRYL1	Lambda-crystallin homolog	35396	2	2,8	0,001
806	ANKH1	Ankyrin repeat and KH domain-containing protein 1	269291	3	0,6	0,001
807	ABCG2	ATP-binding cassette sub-family G member 2	72267	3	3,1	0,001
808	ALAT1	Alanine aminotransferase 1	54602	2	1,8	0,001
809	UGDH	UDP-glucose 6-dehydrogenase	54989	2	1,8	0,001
810	VPS35	Vacuolar protein sorting-associated protein 35	91649	2	2,5	0,001
811	CPT2	Carnitine O-palmitoyltransferase 2, mitochondrial	73730	3	4,6	0,001
812	KAD9	Adenylate kinase 9	221273	2	0,8	0,001
813	VATB2	V-type proton ATPase subunit B, brain isoform	56465	2	4,9	0,001
814	PHLB1	Pleckstrin homology-like domain family B member 1	151068	2	1	0,001
815	TRIM4	E3 ubiquitin-protein ligase TRIM4	57424	2	2,2	0,001
816	KPCA	Protein kinase C alpha type	76700	3	5,8	0,001
817	MRP2	Canalicular multispecific organic anion transporter 1	174096	3	1,2	0,001
818	HAVR1	Hepatitis A virus cellular receptor 1	38696	2	2,2	0,001
819	ERLN1	Erlin-1	38901	2	5,5	0,001
820	AL7A1	Alpha-aminoadipic semialdehyde dehydrogenase	58450	2	2,6	0,001
821	KIF3A	Kinesin-like protein KIF3A	79991	2	1	0,001
822	CO6A3	Collagen alpha-3(VI) chain	343457	2	0,4	0,001
823	ECM1	Extracellular matrix protein 1	60635	2	4,1	0,001
824	PTPA	Serine/threonine-protein phosphatase 2A activator	40641	2	2	0,001
825	S22A9	Solute carrier family 22 member 9	62128	2	1,4	0,001
826	RBP17	Ran-binding protein 17	124295	2	1,5	0,001
827	PP1R7	Protein phosphatase 1 regulatory subunit 7	41539	2	6,9	0,001
828	VAT1	Synaptic vesicle membrane protein VAT-1 homolog	41893	2	3,6	0,001
829	PLXA4	Plexin-A4	212318	3	0,5	0,001

830	MYO6	Unconventional myin-VI	149596	5	2,4	0,001
831	PUR9	Bifunctional purine biynthesis protein PURH	64575	3	5,9	0,001
832	VMA5A	von Willebrand factor A domain-containing protein 5A	86434	3	2,3	0,001
833	PKP3	Plakophilin-3	87029	2	2	0,001
834	CP110	Centriolar coiled-coil protein of 110 kDa	113353	2	0,7	0,001
835	SASH1	SAM and SH3 domain-containing protein 1	136568	2	0,8	0,001
836	DGLA	Sn1-specific diacylglycerol lipase alpha	114879	3	1,4	0,001
837	FHAD1	Forkhead-associated domain-containing protein 1	161804	2	0,7	0,001
838	PLCB3	1-phphatidylnitol 4,5-bisphphate phphodiesterase beta-3	138713	4	1,5	0,001
839	FLNB	Filamin-B	277990	4	1,6	0,001
840	APMAP	Adipocyte plasma membrane-associated protein	46451	2	4,1	0,001
841	ACADV	Very long-chain specific acyl-CoA dehydrogenase, mitochondrial	70345	2	1,7	0,001
842	UBA1	Ubiquitin-like modifier-activating enzyme 1	117774	3	1,8	0,001
843	PLCH1	1-phphatidylnitol 4,5-bisphphate phphodiesterase eta-1	189104	2	0,4	0,001
844	CALR	Calreticulin	48112	2	5	0,001
845	SMYD3	Histone-lysine N-methyltransferase SMYD3	49065	2	1,6	0,001
846	CTNA1	Catenin alpha-1	100009	2	3,1	0,001
847	EFL1	Elongation factor-like GTPase 1	125350	2	0,7	0,001
848	FBLN5	Fibulin-5	50147	2	4,5	0,001
849	EVPL	Envoplakin	231463	4	0,6	0,001
850	NU4M	NADH-ubiquinone oxidoreductase chain 4	51547	2	2,2	0,001
851	DOP2	Protein dopey-2	258067	7	1,3	0,001
852	AP1B1	AP-1 complex subunit beta-1	104570	2	2	0,001
853	LPPRC	Leucine-rich PPR motif-containing protein, mitochondrial	157805	3	1,9	0,001
854	RFIP1	Rab11 family-interacting protein 1	137083	3	0,9	0,001
855	PI3R6	Phphoinitide 3-kinase regulatory subunit 6	84205	2	2	0,001
856	TRHY	Trichohyalin	253777	6	2	0,001
857	PCYOX	Prenylcysteine oxidase 1	56604	2	3	0,001
858	EAA3	Excitatory amino acid transporter 3	57063	2	2,1	0,001
859	TCPD	T-complex protein 1 subunit delta	57888	2	3	0,00050

860	TCPZ	T-complex protein 1 subunit zeta	57988	2	3,2	0,00050
861	P5CS	Delta-1-pyrroline-5-carboxylate synthase	87248	3	3	0,00049
862	PTPRJ	Receptor-type tyrosine-protein phosphatase eta	145851	3	2,7	0,00049
863	SAHH2	Adenylhomocysteinase 2	58913	2	1,5	0,00049
864	ISK5	Serine protease inhibitor Kazal-type 5	120637	2	0,7	0,00048
865	RRBP1	Ribosome-binding protein 1	152381	2	0,6	0,00047
866	NXPE4	NXPE family member 4	62223	2	1,5	0,00046
867	FZD2	Frizzled-2	63512	2	2,8	0,00045
868	LDLR	Low-density lipoprotein receptor	95314	2	2,1	0,00045
869	MAOX	NADP-dependent malic enzyme	64109	2	1,2	0,00045
870	NCKP1	Nck-associated protein 1	128707	2	1,7	0,00045
871	S13A2	Solute carrier family 13 member 2	64368	2	2,9	0,00045
872	HXK3	Hexokinase-3	98962	2	3	0,00044
873	CALX	Calnexin	67526	2	2,9	0,00043
874	LKHA4	Leukotriene A-4 hydrolase	69241	2	2,9	0,00041
875	COFA1	Collagen alpha-1(XV) chain	141632	2	1,6	0,00041
876	INT7	Integrator complex subunit 7	106766	2	0,8	0,00040
877	GANAB	Neutral alpha-glucosidase AB	106807	2	1	0,00040
878	NPCL1	Niemann-Pick C1-like protein 1	148632	3	1,8	0,00039
879	DAPLE	Protein Daple	228091	2	0,3	0,00038
880	ACSL4	Long-chain-fatty-acid--CoA ligase 4	79137	2	1,1	0,00036
881	MTA1	Metastasis-associated protein MTA1	80735	2	2,2	0,00036
882	PLEC	Plectin	531466	5	0,4	0,00035
883	KS6A1	Ribosomal protein S6 kinase alpha-1	82671	2	1,5	0,00035
884	BRAT1	BRCA1-associated ATM activator 1	88063	2	1	0,00033
885	CFAH	Complement factor H	139005	2	0,9	0,00031
886	MUC12	Mucin-12	557831	3	0,7	0,00031
887	MILK1	MICAL-like protein 1	93383	2	0,9	0,00031
888	DHX9	ATP-dependent RNA helicase A	140869	2	0,6	0,00031
889	DPOLQ	DNA polymerase theta	289437	2	0,5	0,00030

890	SYTL2	Synaptotagmin-like protein 2	104865	2	1,1	0,00027
891	NU160	Nuclear pore complex protein Nup160	162017	2	0,7	0,00027
892	OPA1	Dynamin-like 120 kDa protein, mitochondrial	111561	2	0,9	0,00026
893	ERCC6	DNA excision repair protein ERCC-6	168311	2	0,5	0,00026
894	TSH3	Teashirt homolog 3	118492	2	1,3	0,00024
895	TITIN	Titin	3813651	4	0	0,00024
896	K1614	Uncharacterized protein KIAA1614	126527	2	0,6	0,00023
897	MYO1E	Unconventional myosin-le	126982	2	1,8	0,00023
898	PCD12	Protocadherin-12	128915	2	0,6	0,00022
899	CUBN	Cubilin	398480	3	0,5	0,00022
900	DSRAD	Double-stranded RNA-specific adenine deaminase	135981	2	1,1	0,00021
901	AT8B1	Phospholipid-transporting ATPase IC	143604	2	1,8	0,00020
902	CYFP1	Cytoplasmic FMR1-interacting protein 1	145089	2	1	0,00020
903	K2026	Uncharacterized protein KIAA2026	227945	2	0,3	0,00019
904	MYOF	Myoferlin	234561	2	1,3	0,00018
905	FBN2	Fibrillin-2	314558	2	0,6	0,00018
906	MUC3A	Mucin-3A	344914	2	0,8	0,00017
907	AHNK	Neuroblast differentiation-associated protein AHNK	628699	3	0,5	0,00016
908	SPTA1	Spectrin alpha chain, erythrocytic 1	279842	2	0,3	0,00015
909	DYHC2	Cytoplasmic dynein 2 heavy chain 1	492311	2	0,2	0,00015
910	SYNE1	Nesprin-1	1010455	3	0,2	0,00014
911	BIG1	Brefeldin A-inhibited guanine nucleotide-exchange protein 1	208634	2	0,4	0,00014
912	KALRN	Kalirin	339960	2	0,2	0,00013
913	LRP1	Pro-low-density lipoprotein receptor-related protein 1	504276	2	0,5	0,00011
914	FSIP2	Fibrous sheath-interacting protein 2	780119	3	0,2	0,00011
915	MPRI	Cation-independent mannose-6-phosphate receptor	274199	2	0,3	0,00010
916	ABCAD	ATP-binding cassette sub-family A member 13	575791	3	0,1	0,00010
917	PGBM	Basement membrane-specific heparan sulfate proteoglycan core protein	468532	2	0,5	0,00009
918	DYH6	Dynein heavy chain 6, axonemal	475679	2	0,2	0,00009

919	OBSCN	Obscurin	867940	2	0,2	0,00008
920	MYO15	Unconventional myin-XV	395044	2	0,2	0,00007
921	SYNE2	Nesprin-2	795943	2	0,1	0,00005
922	NEBU	Nebulin	772428	2	0,1	0,00004

Annexe 14 : Liste des 79 protéines décrivant le « core protéome » identifiées au sein de l'ensemble des échantillons de méconium étudiés, incluant les temps de prélèvements précoces et tardifs (attribuées avec au moins 2 peptides spécifiques). Les protéines sont décrites par ordre d'abondance décroissant dans le méconium (NSAF [%] décroissant).

N°	Accession [Humain]	Protéines	Masses	Nombre de peptides spécifiques (tous les échantillons)	NSAF [%]
1	A1AT	Alpha-1-antitrypsin	46707	40	7,511
2	S10A9	Protein S100-A9	13234	14	4,620
3	S10A8	Protein S100-A8	10828	17	4,618
4	IGKC	Ig kappa chain C region	11602	8	2,605
5	AACT	Alpha-1-antichymotrypsin	47621	25	1,801
6	H4	Histone H4	11360	9	1,738
7	LAC2	Ig lambda-2 chain C regions	11287	4	1,375
8	K2C1	Keratin, type II cytokeletal 1	65999	46	1,174
9	TTHY	Transthyretin	15877	10	1,155
10	COL	Colipase	11946	5	1,098
11	CEL3A	Chymotrypsin-like elastase family member 3A	29470	10	1,093
12	CTRC	Chymotrypsin-C	29465	14	1,090
13	ALBU	Serum albumin	69321	57	1,007
14	K1C10	Keratin, type I cytokeletal 10	58792	33	1,001
15	AMPN	Aminopeptidase N	109471	59	0,897
16	PSPB	Pulmonary surfactant-associated protein B	42089	7	0,790
17	ANT3	Antithrombin-III	52569	28	0,718
18	CLCA1	Calcium-activated chloride channel regulator 1	100163	48	0,661
19	ITLN1	Intelectin-1	34939	11	0,632
20	TFF3	Trefoil factor 3	8635	4	0,630
21	PPBI	Intestinal-type alkaline phosphatase	56776	14	0,619
22	SLPI	Antileukoproteinase	14316	11	0,597

23	PERE	Einophil peroxidase	80989	36	0,585
24	KV320	Immunoglobulin kappa variable 3-20	12549	2	0,585
25	IGLL5	Immunoglobulin lambda-like polypeptide 5	23049	4	0,581
26	APOD	Apolipoprotein D	21262	11	0,562
27	IC1	Plasma protease C1 inhibitor	55119	23	0,552
28	NEP	Neprilysin	85460	38	0,542
29	DPEP1	Dipeptidase 1	45645	16	0,525
30	ACTB	Actin, cytoplasmic 1	41710	9	0,500
31	TRFL	Lactotransferrin	78132	71	0,467
32	TFF2	Trefoil factor 2	14274	6	0,455
33	ANXA2	Annexin A2	38580	30	0,447
34	KV401	Immunoglobulin kappa variable 4-1	13372	3	0,417
35	PRG2	Bone marrow proteoglycan	25189	10	0,388
36	IGHG1	Ig gamma-1 chain C region	36083	6	0,379
37	S10A6	Protein S100-A6	10173	5	0,378
38	PA21B	Phospholipase A2	16349	6	0,354
39	ECP	Einophil cationic protein	18373	9	0,334
40	ANXA1	Annexin A1	38690	25	0,316
41	ANXA4	Annexin A4	35860	29	0,313
42	SPB6	Serpin B6	42594	15	0,305
43	DMBT1	Deleted in malignant brain tumors 1 protein	260569	37	0,261
44	ZG16B	Zymogen granule protein 16 homolog B	22725	7	0,258
45	FRIL	Ferritin light chain	20007	12	0,257
46	DPP4	Dipeptidyl peptidase 4	88222	36	0,256
47	HV315	Immunoglobulin heavy variable 3-15	12917	5	0,252
48	CBPA1	Carboxypeptidase A1	47111	18	0,250
49	SPB3	Serpin B3	44537	12	0,242
50	NPC2	Epididymal secretory protein E1	16559	8	0,242
51	RS27A	Ubiquitin-40S ribosomal protein S27a	17953	4	0,223
52	SODC	Superoxide dismutase [Cu-Zn]	15926	9	0,221

53	G3P	Glyceraldehyde-3-phosphate dehydrogenase	36030	12	0,192
54	K1C9	Keratin, type I cytoskeletal 9	62027	16	0,190
55	FCGBP	IgG Fc-binding protein	571639	116	0,188
56	IPSP	Plasma serine protease inhibitor	45646	10	0,185
57	ANXA5	Annexin A5	35914	18	0,178
58	XPP2	Xaa-Pro aminopeptidase 2	75577	25	0,176
59	ANX13	Annexin A13	35393	22	0,172
60	RET2	Retinol-binding protein 2	15697	7	0,165
61	COF1	Cofilin-1	18491	12	0,160
62	FRIH	Ferritin heavy chain	21212	9	0,154
63	LG3BP	Galectin-3-binding protein	65289	16	0,153
64	HV434	Immunoglobulin heavy variable 4-34	13806	4	0,150
65	GP2	Pancreatic secretory granule membrane major glycoprotein GP2	59442	14	0,142
66	LEG4	Galectin-4	35918	15	0,131
67	CERU	Ceruloplasmin	122128	26	0,118
68	PROF1	Profilin-1	15045	6	0,115
69	LDHA	L-lactate dehydrogenase A chain	36665	7	0,114
70	CBPA2	Carboxypeptidase A2	47001	11	0,113
71	ASAH1	Acid ceramidase	44631	13	0,113
72	SAP	Prapin	58074	10	0,111
73	CDHR2	Cadherin-related family member 2	141455	24	0,111
74	GGH	Gamma-glutamyl hydrolase	35941	8	0,107
75	PPIA	Peptidyl-prolyl cis-trans isomerase A	18001	9	0,106
76	CBPQ	Carboxypeptidase Q	51855	12	0,102
77	CEAM5	Carcinoembryonic antigen-related cell adhesion molecule 5	76747	5	0,092
78	EF1A1	Elongation factor 1-alpha 1	50109	11	0,090
79	BGLR	Beta-glucuronidase	74685	16	0,088

Annexe 15 : Liste des 165 OTUs détectés au sein des 24 échantillons de méconiums analysés par l'approche métagénomique ciblée de l'ARN-16S.

PHYLUM	CLASSE	ORDRE	FAMILLE	GENRE	ESPECE		
Actinobacteria (1.0)	Actinobacteria (1.0)	Bifidobacteriales (1.0)	Bifidobacteriaceae (1.0)	Bifidobacterium (1.0)	Bifidobacterium adolescentis (1.0)		
					Bifidobacterium bifidum (1.0)		
					Bifidobacterium breve (1.0)		
					Bifidobacterium gallicum (0.41)		
					Bifidobacterium longum (0.62)		
					Bifidobacterium longum (1.0)		
					Bifidobacterium pseudocatenulatum (0.72)		
		Corynebacteriales (1.0)	Corynebacteriaceae (1.0)	Corynebacterium 1 (1.0)	Corynebacterium tuberculostearicum (1.0)		
		Micrococcales (1.0)	Beutenbergiaceae (0.84)	unknown genus (0.59)	Serinibacter sp. (0.59)		
Micrococcaceae (1.0)	Rothia (1.0)				Rothia mucilaginosa DY-18 (0.55)		
	Rothia mucilaginosa DY-18 (0.94)						
Bacteroidetes (1.0)	Bacteroidia (1.0)	Bacteroidales (1.0)	Bacteroidaceae (0.97)	Bacteroides (0.97)	Bacteroides massiliensis (0.6)		
					metagenome (0.37)		
					Bacteroidaceae (1.0)	Bacteroides (1.0)	bacterium NLAE-zl-P920 (0.03)
							Bacteroides acidifaciens (0.88)
							Bacteroides caccae (1.0)
							Bacteroides clarus (1.0)
		Bacteroides dorei (0.63)					
		Bacteroides eggerthii (0.76)					
		Bacteroides faecis (0.27)					
		Bacteroides fragilis (0.99)					
		Bacteroides fragilis HMW 610 (0.26)					
		Bacteroides massiliensis (0.75)					

					Bacteroides massiliensis (1.0)
					Bacteroides thetaiotaomicron (1.0)
					Bacteroides uniformis (0.62)
					Bacteroides uniformis (0.95)
					Bacteroides uniformis (1.0)
					Bacteroides vulgatus (0.42)
					Bacteroides vulgatus (0.61)
					Bacteroides vulgatus (0.7)
					Bacteroides vulgatus (0.98)
					Bacteroides vulgatus ATCC 8482 (0.19)
					Bacteroides vulgatus ATCC 8482 (0.29)
					Bacteroides vulgatus ATCC 8482 (0.73)
					unknown species (0.49)
					unknown species (0.63)
					unknown species (0.76)
					unknown species (0.81)
					unknown species (0.84)
					unknown species (0.84)
					unknown species (0.87)
					unknown species (0.88)
					unknown species (0.89)
					unknown species (0.92)
					unknown species (0.97)
			Muribaculaceae (1.0)	Muribaculum (1.0)	unknown species (1.0)
				unknown genus (1.0)	unknown species (0.88)
					unknown species (0.98)
					unknown species (0.99)
					unknown species (0.99)
					unknown species (1.0)
					unknown species (1.0)

					unknown species (1.0)
					unknown species (1.0)
					unknown species (1.0)
					unknown species (1.0)
					unknown species (1.0)
					unknown species (1.0)
					unknown species (1.0)
					unknown species (1.0)
					unknown species (1.0)
					unknown species (1.0)
			Porphyromonadaceae (1.0)	Porphyromonas (1.0)	Porphyromonas bennonis (1.0)
			Prevotellaceae (1.0)	Alloprevotella (1.0)	unknown species (1.0)
			Rikenellaceae (1.0)	Alistipes (1.0)	unknown species (0.59)
					unknown species (0.96)
					unknown species (1.0)
			Tannerellaceae (0.94)	Parabacteroides (0.94)	unknown species (0.53)
			Tannerellaceae (1.0)	Parabacteroides (1.0)	Parabacteroides distasonis (0.65)
					Parabacteroides distasonis ATCC 8503 (0.92)
					Parabacteroides johnsonii (1.0)
					Parabacteroides merdae (1.0)
					unknown species (0.97)
					unknown species (0.98)
Deferribacteres (1.0)	Deferribacteres (1.0)	Deferribacterales (1.0)	Deferribacteraceae (1.0)	Mucispirillum (1.0)	Mucispirillum schaedleri ASF457 (1.0)
Firmicutes (0.53)	Bacilli (0.53)	Lactobacillales (0.53)	Enterococcaceae (0.53)	Enterococcus (0.53)	Enterococcus faecalis EnGen0405 (0.01)
Firmicutes (1.0)	Bacilli (1.0)	Bacillales (0.6)	Staphylococcaceae (0.59)	Staphylococcus (0.58)	bacterium H17 (2009) (0.07)
		Bacillales (1.0)	Staphylococcaceae (1.0)	Staphylococcus (1.0)	Staphylococcus epidermidis (0.68)

					Staphylococcus epidermidis PM221 (0.16)
					Staphylococcus epidermidis RP62A phage SP-beta (0.2)
	Lactobacillales (1.0)	Aerococcaceae (1.0)	Aerococcus (1.0)		unknown species (0.95)
		Enterococcaceae (1.0)	Enterococcus (1.0)		Enterococcus durans (0.97)
					Enterococcus faecalis (0.97)
					Enterococcus faecalis ATCC 29200 (0.46)
					Enterococcus faecalis ATCC 29200 (0.47)
					human gut metagenome (0.35)
		Lactobacillaceae (1.0)	Lactobacillus (1.0)		Lactobacillus animalis (1.0)
					Lactobacillus coleohominis (1.0)
					Lactobacillus crispatus (0.6)
					Lactobacillus jensenii (1.0)
					Lactobacillus reuteri (1.0)
					Lactobacillus taiwanensis (0.48)
		Streptococcaceae (1.0)	Streptococcus (1.0)		Streptococcus anginosus (1.0)
					Streptococcus cristatus (0.47)
					Streptococcus gallolyticus (1.0)
					Streptococcus oralis (0.49)
					Streptococcus parasanguinis (0.88)
					Streptococcus salivarius (0.98)
					unknown species (0.4)
	Clostridia (0.98)	Clostridiales (0.98)	Clostridiaceae 1 (0.98)	Clostridium sensu stricto 1 (0.98)	Clostridium perfringens ATCC 13124 (0.69)
	Clostridia (1.0)	Clostridiales (1.0)	Clostridiaceae 1 (1.0)	Clostridium sensu stricto 1 (1.0)	Clostridium perfringens (0.73)
			Clostridiaceae 1 (1.0)	Clostridium sensu stricto 1 (1.0)	metagenome (1.0)
			Clostridiales vadinBB60 group (1.0)	unknown genus (1.0)	unknown species (1.0)

			Clostridiales vadinBB60 group (1.0)		unknown species (1.0)
			Clostridiales vadinBB60 group (1.0)		unknown species (1.0)
			Lachnospiraceae (1.0)	[Eubacterium] xylanophilum group (1.0)	unknown species (1.0)
				[Ruminococcus] gnavus group (1.0)	unknown species (1.0)
				A2 (1.0)	unknown species (1.0)
				ASF356 (1.0)	unknown species (1.0)
				ASF356 (1.0)	unknown species (1.0)
				Blautia (1.0)	Lachnospiraceae (0.65)
				Blautia (1.0)	Lachnospiraceae bacterium 615 (1.0)
				Lachnoclostridium (0.58)	unknown species (0.55)
				Lachnoclostridium (0.95)	Dorea sp. (0.72)
				Lachnoclostridium (0.99)	Dorea sp. (0.98)
				Lachnospiraceae NK4A136 group (0.87)	unknown species (0.87)
				Lachnospiraceae NK4A136 group (0.92)	unknown species (0.92)
				Lachnospiraceae NK4A136 group (0.97)	unknown species (0.97)
				Lachnospiraceae NK4A136 group (0.99)	Blautia coccoides (0.47)
				Lachnospiraceae NK4A136 group (0.99)	
				Lachnospiraceae NK4A136 group (1.0)	

				Lachnospiraceae UCG-001 (1.0)	
				Lachnospiraceae UCG-001 (1.0)	
				Roseburia (0.92)	
				Roseburia (0.99)	
				Roseburia (1.0)	
				unknown genus (0.26)	
				unknown genus (0.31)	
				unknown genus (0.61)	
				unknown genus (0.66)	
				unknown genus (0.94)	
			Ruminococcaceae (1.0)	Butyricoccus (1.0)	
				Ruminiclostridium (0.99)	
				Ruminiclostridium (1.0)	
				unknown genus (0.74)	
	Erysipelotrichia (1.0)	Erysipelotrichales (1.0)	Erysipelotrichaceae (1.0)	Turcibacter (1.0)	
Proteobacteria (0.97)	Gammaproteobacteria (0.97)	Enterobacteriales (0.97)	Enterobacteriaceae (0.97)	Escherichia-Shigella (0.87)	
Proteobacteria (1.0)	Alphaproteobacteria (1.0)	Sphingomonadales (1.0)	Sphingomonadaceae (1.0)	Sphingomonas (1.0)	
	Gammaproteobacteria (1.0)	Betaproteobacteriales (0.8)	Burkholderiaceae (0.8)	Sutterella (0.8)	
		Betaproteobacteriales (1.0)	Burkholderiaceae (1.0)	Massilia (1.0)	

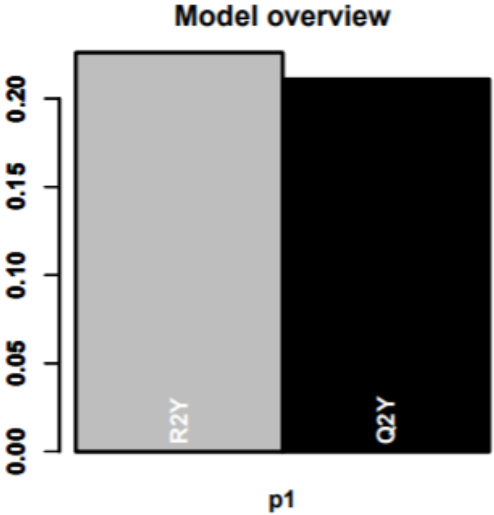
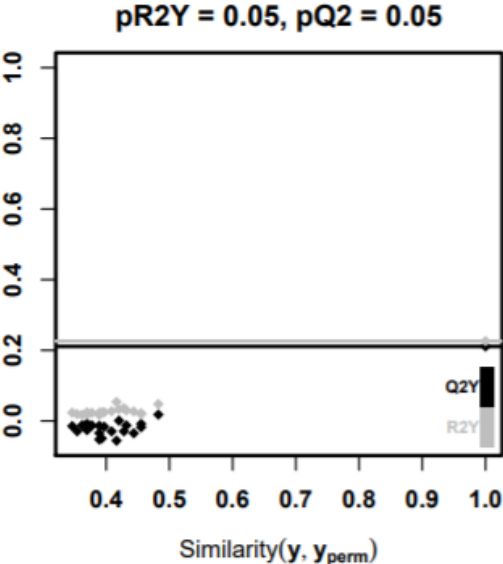
				Parasutterella (1.0)	
				Sutterella (1.0)	
				Undibacterium (1.0)	
		Enterobacteriales (0.63)	Enterobacteriaceae (0.63)	Escherichia-Shigella (0.58)	
		Enterobacteriales (1.0)	Enterobacteriaceae (1.0)	Escherichia-Shigella (1.0)	
		Pasteurellales (1.0)	Pasteurellaceae (1.0)	Haemophilus (1.0)	
		Pseudomonadales (1.0)	Moraxellaceae (1.0)	Acinetobacter (1.0)	
				Enhydrobacter (1.0)	
		Xanthomonadales (1.0)	Xanthomonadaceae (1.0)	Pseudoxanthomonas (1.0)	

Annexe 16 : Liste des 59 métabolites supplémentaires annotés au sein de la cohorte EDEN dans les conditions HILIC-ESI⁻ et C18-ESI⁺.

Métabolites	m/z (ESI+)	m/z (ESI-)	RT (RP-C18)	RT (ZIC-pHILIC)
13(S)-Hydroxyoctadeca-9Z,11E-dienoic acid	nd	295,2277	nd	1,43
2,3-Diaminopropionic acid	nd	103,0514	nd	7,91
2-hydroxycaproic acid	nd	131,0718	nd	1,96
4-Hydroxyphenylacetic acid/4-Hydroxy-2-methylbenzoic acid /3-Hydroxyphenyl acetate/4-Hydroxy-3-methylbenzoic acid	nd	151,0400	nd	4,17
5-Aminoimidazole-4-carboxamide	nd	125,0470	nd	3,89
5-hydroxy-dl-tryptophan	nd	219,0773	nd	5,96
Chlorogenic acid	nd	353,0850	nd	6,92
Cinnamic acid	nd	147,0452	nd	3,21
cis-8,11,14-Eicosatrienoic acid	nd	305,2484	nd	1,38
D-erythro-Dihydrosphingosine (Sphinganine)	nd	283,2645	nd	2,10
D-Sphingosine /Elaidic acid/trans-Vaccenic acid	nd	281,2488	nd	2,10
Gallic acid /mono-Methyl hydrogen succinate /Methyl-succinic acid	nd	169,0142	nd	2,84
gamma-Linolenic acid	nd	279,2332	nd	2,10
Indoleacetic acid/Dimethyl 3-hydroxyglutarate/Indoxyl acetate	nd	175,0599	nd	3,03
Isoxanthopterin/Xanthopterin	nd	178,0358	nd	7,27
L-Histidine methyl ester	nd	168,0778	nd	3,56
Methylimidazoleacetic acid	nd	139,0513	nd	4,85
N-Methylnicotinamide/6-Methylnicotinamide/2-Methylnicotinamide	nd	135,0565	nd	1,84
Octanoic acid	nd	143,1077	nd	1,59
Oxaloacetic acid	nd	87,0086	nd	7,64
Panthenol	nd	204,1242	nd	2,24
p-Coumaric acid /2-Hydroxycinnamic acid/trans-3-Hydroxycinnamic acid (m-Coumaric acid)/o-Tyrosine/Isohomovanillic acid (3-Hydroxy-4-methoxyphenylacetic acid)	nd	163,0401	nd	3,60
Phosphoserine	nd	184,0017	nd	8,55
Pyridoxamine-5prim-phosphate	nd	247,0468	nd	9,54
Rutin	nd	609,1454	nd	7,63
Trigonelline/4-Aminobenzoic acid/2-Aminobenzoic acid	nd	136,0404	nd	5,38
1-Methyladenosine	282,1180	nd	0,83	nd
1-Methyluric-acid	183,0512	nd	1,34	nd
3-Amino-3-(4-hydroxyphenyl)propanoic	165,0546	nd	1,12	nd
3-Hydroxy-3-methylglutaric-acid	145,0496	nd	1,31	nd
4-Acetamidobutanoic-acid	146,0812	nd	1,36	nd
4-Amino-nicotinic-acid/6-Aminopyridine-3-carboxylic-acid/4-Imidazoleacrylic-acid-(urocanic-acid)/5-	139,0502	nd	0,88	nd

Aminopyridine-3-carboxylic-acid/2-Aminopyridine-3-carboxylic-acid/Nicotinamide-N-oxide				
4-Hydroxy-D-phenylglycine/Pyridoxal	168,0654	nd	0,86	nd
4-Hydroxypyridine/3-Hydroxypyridine	96,0442	nd	0,94	nd
5-AMP/3-AMP	348,0687	nd	0,83	nd
5-Hydroxyindoleacetic-acid	192,0655	nd	4,52	nd
6beta-Hydroxytestosterone/7alpha-Hydroxytestosterone	305,2108	nd	7,09	nd
6-Biopterin	238,0933	nd	1,07	nd
Aalanyl-Tyrosine	253,1182	nd	1,56	nd
Acetyl-L-carnitin	204,1230	nd	0,92	nd
Argininosuccinic-acid	273,1189	nd	0,79	nd
Aspartyl-Phenylalanine	281,1132	nd	4,07	nd
Deoxycorticosterone-21-glucoside	493,2791	nd	7,74	nd
Deoxycytidine	228,0978	nd	0,82	nd
Deoxyuridine/2?-Deoxyuridine	229,0819	nd	1,02	nd
Dihydrothymine	129,0660	nd	1,35	nd
glycochenodeoxycholic-acid	450,3208	nd	8,88	nd
Inosine	269,0877	nd	1,22	nd
Isonicotinic-acid-N-oxide/Nicotinic-acid-N-oxide/3-Hydroxypicolinic-acid/6-Hydroxypyridine-2-carboxylic-acid/5-Hydroxynicotinic-acid	140,0343	nd	1,08	nd
Isopentyl-nitrite/Valine/5-Aminovaleric-acid/Betaine	118,0861	nd	0,89	nd
Methionine-sulfoxide	166,0531	nd	0,82	nd
Methylimidazoleacetic-acid	141,0659	nd	0,95	nd
Mevalonolactone	131,0705	nd	1,45	nd
N8-Acetylspermidine	188,1755	nd	0,77	nd
Propionylcarnitine	218,1387	nd	1,31	nd
S-ethyl-L-cysteine	133,0318	nd	1,00	nd
Testosterone	289,2157	nd	8,83	nd
Triethanolamine-ou-Trolamine	150,1125	nd	0,77	nd
Uric-acid	169,0355	nd	0,97	nd

Annexe 17 : Graphe des pR2Y et pQ2, obtenu lors de l'analyse des moindres carrés partiels (PLS-DA) de l'effet temps de collecte (J1, >J1) au sein des échantillons de la cohorte initiale EDEN, à partir des 301 variables annotées.



Annexe 18 : Liste des 46 métabolites discriminants entre les prélèvements d'échantillons de méconium de la cohorte EDEN réalisés au cours des premières 24 heures (< J1) et au-delà (> J1).

Métabolites	Origines	Evolution
gamma-Linolenic acid	Médicament	<J1
N6-Acetyl-L-lysine	Endogène	<J1
N-Acetylglycine	Endogène	<J1
N-Acetyl-L-alanine	Endogène	<J1
N-Acetyl-ornithine	Endogène	<J1
O-Acetyl-L-serine	Endogène	<J1
Orotic acid	Endogène	<J1
Proline	Endogène	<J1
Uridine	Endogène	<J1
Valine	Endogène	<J1
2-Aminoisobutyric acid (2-Methylalanine)	Endogène	>J1
2-hydroxycaproic acid	Endogène	>J1
2-Isopropylmalic acid	Microbien	>J1
2-O-Methylguanosine	Endogène	>J1
3-Hydroxy-3-methylglutaric acid	Endogène	>J1
3-Hydroxypentanoic acid	Endogène	>J1
5-Methoxyindoleacetate	Endogène	>J1
7-methylguanine	Endogène	>J1
Adenine	Endogène	>J1
Cytidine	Endogène	>J1
D-Glucuronic acid	Endogène	>J1
D-Glutamine	Endogène	>J1
D-Glyceric acid	Endogène	>J1
Dihydrouracil	Endogène	>J1
DL-Dihydroorotic acid	Endogène	>J1
DL-p-Hydroxyphenyllactic acid	Endogène	>J1
D-Mannitol	Endogène	>J1
D-Sphingosine	Endogène	>J1
Ethylmethylacetic acid (2-Methylbutyric acid)	Endogène	>J1
Imidazolelactic acid	Endogène	>J1
Isopentyl-nitrite	Médicament	>J1
L-2-Aminobutyric-acid	Endogène	>J1
Maleic acid	Endogène	>J1
Methylmalonic acid	Endogène	>J1
Mevalonic acid	Endogène	>J1
Mevalonolactone	Endogène	>J1
mono-Methyl hydrogen succinate	Endogène	>J1
N acetyl L glutamine	Endogène	>J1
N-acetyl-DL-Valine	Endogène	>J1
N-Acetyl-L-phenylalanine	Endogène	>J1
Nicotinic-acid	Médicament	>J1
p-Coumaric acid	Endogène	>J1

Sarcosine	Endogène	>J1
Taurine	Endogène	>J1
trans-3-Hydroxy-L-proline	Endogène	>J1
Trigonelline	Endogène	>J1

Annexe 19 : Liste des 38 métabolites discriminants au sein des échantillons des centres de Nancy et de Poitiers de la cohorte EDEN.

Métabolites	Origines	Centre_PLSDA_VIP
DL-Dihydroorotic acid	Endogène	1.98748816587029
Methylmalonic acid	Endogène	1.89910168536339
Ethylmethylacetic acid (2-Methylbutyric acid)	Endogène	1.88961622667155
Mevalonic acid	Endogène	1.84701967385396
N acetyl L glutamine	Endogène	1.83420340535662
Taurine	Endogène	1.8186683118688
4-Pyridoxic acid	Endogène	1.80585249766011
7-methylguanine	Endogène	1.80321569529647
L-Pyroglutamic acid	Endogène	1.77493028546069
Orotic acid	Endogène	1.77358659655857
D-Glyceric acid	Endogène	1.76643618394562
L-2-Aminobutyric-acid	Endogène	1.75818109640172
DL-p-Hydroxyphenyllactic acid	Endogène	1.7503244831972
2-hydroxycaproic acid	Endogène	1.74247999870742
2-Aminoisobutyric acid (2-Methylalanine)	Endogène	1.73546782735946
N-Acetyl-L-phenylalanine	Endogène	1.72973824661611
Succinic semialdehyde	Endogène	1.72609318365153
Mevalonolactone	Endogène	1.71606358021301
trans-3-Hydroxy-L-proline	Endogène	1.69772544845257
(-)-N-Acetylneuraminic acid	Endogène	1.68373333553774
N6-Acetyl-L-lysine	Endogène	1.67181611130154
D-Glutamine	Endogène	1.65815969899956
D-a-Hydroxyglutaric acid (ou D-2-Hydroxyglutaric acid)	Endogène	1.65249115536069
L-Threonic acid	Endogène	1.63814810896622
D-Threitol	Endogène	1.63509182815423
5-hydroxytryptophan	Endogène	1.62699392329126
Cholic acid	Endogène	1.62559401043514
N-acetyl-DL-Valine	Endogène	1.56878189488685
3-Hydroxypentanoic acid	Endogène	1.56743671256414
Pimelic acid	Endogène	1.55682908869571
Kynurenic-acid	Endogène	1.55060715461681
N acetyl DL glutamic acid	Endogène	1.54430496287937
4-Acetamidobutanoic acid	Endogène	1.52926048829088
Trigonelline	Endogène	1.52907280303543
D-Alanyl-D-Alanine	Endogène	1.52714624639676
N-Acetylglycine	Endogène	1.52084301724463
2-Isopropylmalic acid	Microbien	1.51369490488586
p-Coumaric acid	Endogène	1.51139484268282

Cartographie individuelle de la composition de méconiums humains : un reflet de l'exposome prénatal corrélé au développement d'une allergie alimentaire dans l'enfance ?

Métabolomique, LC/HRMS, Méconium, Cartographie globale, Approches multi-omiques, Allergies alimentaires

Le méconium est une matrice complexe, métaboliquement inactive qui s'accumule *in utero* pendant les 6 derniers mois de grossesse. Sa composition pourrait refléter les expositions du fœtus, en lien avec les facteurs environnementaux et de santé des mères, pouvant impacter la santé future de l'enfant. L'un des enjeux de cette thèse a été de mettre en place une méthodologie d'analyse métabolomique non ciblée du méconium par LC/HRMS. Les optimisations réalisées ont permis de définir une méthodologie robuste de préparation de l'échantillon et d'établir une cartographie globale du méconium réalisée sans *a priori*. Un second objectif a été d'appliquer cette méthodologie à 35 échantillons de méconiums, collectés longitudinalement au cours des trois premiers jours de vie chez 11 nouveau-nés (collaboration réalisée avec la clinique Ste Thérèse Paris 17^e). L'analyse métabolomique a été complétée par des analyses de (méta)protéomique (collaboration réalisée avec le LI2D du CEA-Marcoule) et de métagénomique ciblée (collaboration réalisée avec l'institut Micalis de l'INRAE de Jouy en Josas). Ces analyses ont permis de décrire la cartographie la plus exhaustive du méconium, connue à ce jour avec l'identification : de 230 métabolites au plus haut niveau de confiance, de 922 protéines humaines et de 165 OTUs bactériens. De plus, ces analyses ont permis d'approcher les variabilités interindividuelles et l'évolution temporelle de la composition globale des méconiums. L'analyse métabolomique a ensuite été menée sur 308 échantillons collectés au sein de la cohorte mère-enfant EDEN, une cohorte bi-centrique française. Cette analyse enrichit la cartographie métabolique préalablement décrite, et met en évidence un effet jour et un effet centre sur cette composition, confortant notre hypothèse quant à l'impact de l'environnement sur la composition du méconium. De plus, 40 variables non annotées montrent des intensités significativement différentes entre des méconiums excrétés par des enfants qui développeront une allergie alimentaire dans l'enfance comparé aux enfants qui n'en développeront pas. La combinaison de ces données, qui sont en train d'être complétées par les analyses de métagénomique et (méta)protéomique, avec les informations collectées auprès des mères (alimentation, santé, lieu de vie...), permettrait de mettre en lien les expositions *in utero* sur la composition du méconium, et de son association avec la santé future de l'enfant.

Individual mapping of human meconium: a prenatal exposure reflection correlated to food allergy development in childhood?

Metabolomics profiling, LC/HRMS, Meconium, Mapping, Multiomics approaches, Food Allergy

Meconium is a metabolically inactive matrix that accumulates in utero during the last 6 months of pregnancy. Its composition may reflect fetal exposures, related to environmental and maternal health factors that may affect the future child health. One of the challenges of my thesis was to implement a methodology for non-targeted metabolomics analysis of meconium by LC/HRMS. The optimizations carried out defined a robust methodology of sample preparation, and allowed to establish a global mapping of meconium with no *a priori*. A second objective was to apply this methodology to 35 meconium samples, collected longitudinally during the first three days of life within 11 newborns (Collaboration with Clinique Ste Thérèse, Paris 17th). Metabolomics analysis were completed by (meta)proteomics analysis (collaboration with the LI2D of CEA-Marcoule) and targeted metagenomics (collaboration with the Micalis Institute of INRAE of Jouy en Josas). These analyses allowed describing the most exhaustive mapping of meconium known to date by identifying: 230 metabolites at the highest level of confidence, 922 human proteins and 165 bacterial OTUs. In addition, they allowed approaching the interindividual variability and the temporal evolution of meconium composition. Metabolomics analysis were then conducted on 308 samples collected from the mother-child cohort EDEN, a French bi-centric cohort. These analyses enriched the previous metabolic cartography described and highlighted a day and a center effect on the meconium composition. These results allowed confirming our hypothesis on the impact of the environment on the meconium composition. In addition, 40 variables showed significantly different intensities between meconium excreted by children who will develop a food allergy comparing to those who will not. The combination of these data with the metagenomic and (meta)proteomic analyses with the collected information from mothers (diet, health, living area and environment...), would allow to link in utero exposures to the meconium composition and its association with the child's future health.