



# THÈSE

en vue de l'obtention du grade de  
DOCTEUR DE L'UNIVERSITE PARIS XIII

Ecole doctorale Galilée

Spécialité : Chimie

Présentée et soutenue publiquement par

**Corentine GOOSSENS**

Le 10 décembre 2015

**Approche métabolomique par résonance magnétique nucléaire du proton dans l'évaluation des hépatopathies stéatosiques non alcooliques et dans le suivi d'un traitement curatif du carcinome hépatocellulaire**

Directeur de thèse : Pr. Philippe Savarin

## **JURY**

Rapporteurs	Dr Christophe Junot Dr Pascal de Tullio
Examineurs	Dr Lydie Nadal-Desbarats Pr Laurence Le Moyec Dr Jean-Marie Colet Dr Pierre Nahon Dr Mohamed Nawfal Triba Pr Philippe Savarin

*Manuscrit digne de « foie » ...*



# Remerciements

---

En premier lieu, je tiens à remercier le laboratoire CSPBAT, dirigé par le Pr. Véronique Migonney, pour l'accueil et la confiance qu'elle a manifesté à mon égard, et notamment en m'ouvrant l'accès aux différents instruments analytiques de pointe.

Je remercie Christophe Junot ainsi que Pascal de Tullio d'avoir accepté de juger mon travail et d'en avoir été les rapporteurs. Merci également à Lydie Nadal-Desbarats ainsi que Jean-Marie Colet pour leur participation au jury de cette thèse.

Je tiens à remercier tout particulièrement mon directeur de thèse, Philippe Savarin, pour son soutien sans faille durant ces trois années de thèse. Il m'a garanti un encadrement constant, me donnant la possibilité de travailler dans un environnement idéal, suivant de près l'évolution et la progression de mon travail, tout en me laissant une grande indépendance de choix. Je le remercie pour sa disponibilité constante, sa patience infinie, en un mot son aide en toutes circonstances.

Merci à tous les autres membres du groupe RMN métabolomique pour leur présence et leur bonne humeur : Edith, pour tes conseils expérimentés, tes nombreuses discussions engagées et tes délicieuses recettes de biryani de poulet ; Nadia, pour ton aide précieuse, ton expérience et expertise en RMN, tes conseils bien sûr, et les heures nombreuses que tu m'a consacrées ; Mohamed, merci d'avoir eu la patience d'expliquer et de ré-expliquer les statistiques aux « dummies ». Merci aussi à Roland pour tes différents clins d'œil, valant autant d'encouragements tout au long des trois ans. Agnès, Asma, Marianne et Zhicheng, je vous souhaite une bonne continuation dans vos projets futurs.

A vous tous, membres du groupe, je vous exprime ma sincère reconnaissance d'avoir accepté de distraire une partie du budget attribuée au groupe, au bénéfice des différentes formations et présentations de mes travaux que j'ai eu ainsi l'occasion d'exposer dans les congrès internationaux auxquels j'ai pris part. Ce ne fut pas toujours un exercice évident, mais, à la vérité incontestablement très formateur et enrichissant. Je vous remercie vivement de m'avoir encouragée dans cette voie.

J'associe à ces remerciements, Carina Prip-Buus ainsi que Pierre Nahon pour les collaborations que nous avons pu mettre en place ainsi que pour l'apport de leur point de vue clinico-biologique lors des différentes discussions que nous avons eues.

Je tiens également à exprimer ma gratitude envers Laurence, pour le temps consacré et le suivi qu'elle m'a accordé tout au long de la thèse, sans oublier de mentionner l'hospitalité offerte à l'occasion de réunions parfois tardives qui se sont déroulées dans son « antre ».

Merci aux autres doctorants et docteurs en particulier Max, Sophie et Margery, pour leur présence quotidienne durant ces trois années. Notre soutien réciproque, notre entraide spontanée, et nos moments de délire ont fait naître une amitié sincère.

Merci à « chouke » d'avoir diplomatiquement passé quelques (parfois) longues heures à animer nos discussions scientifiques, à balayer mes craintes, approuver mes joies, mais surtout à me soutenir au quotidien.

Et, une fois n'est pas coutume, à ma famille et mes amis belges !

# Table des matières

---

<b>Liste des abréviations</b>	<b>7</b>
<b>Liste des figures</b>	<b>9</b>
<b>Liste des tableaux</b>	<b>11</b>
<b>Introduction générale</b>	<b>13</b>
<b>Méthodologies</b>	<b>17</b>
<b>Chapitre I La métabolomique</b>	<b>17</b>
I Définition	17
II Démarche et cheminement d'une analyse métabolomique	18
III Domaines d'application	20
IV Vers la biologie des systèmes	21
IV.1 Croisée des sciences « omiques »	21
IV.2 Reconstruction des réseaux métaboliques	22
V Techniques de spectrométrie utilisées en métabolomique	22
V.1 La spectrométrie de masse	23
V.2 La Résonance Magnétique Nucléaire	23
<b>Chapitre II La RMN en métabolomique</b>	<b>25</b>
I Principe	25
I.1 Propriétés magnétiques du noyau atomique	25
I.2 Les états énergétiques	25
I.3 L'équation de Larmor	26
I.4 Détection du signal	28
I.5 Instrumentation	30
I.6 Interprétation des spectres	30
I.7 Optimisation des paramètres d'acquisition	31
II La technique « High Resolution Magic Angle Spinning » (HRMAS)	33
III Application à la métabolomique	33
III.1 Principales séquences utilisées en métabolomique	33
III.2 RMN multidimensionnelle	36
III.3 Méthodes de traitement des données RMN	38
III.4 Méthodes de prétraitement statistique	40
III.5 Interprétations des spectres en métabolomique	43
III.6 Métabolites présents dans le sérum	45
<b>Chapitre III Traitement statistique des données RMN</b>	<b>47</b>
I Notions de statistique et méthodes univariées	47
I.1 Ecart-type et variance d'échantillons	47
I.2 Paramètres de relation entre deux variables	48
I.3 Test de Student	48
I.4 ANOVA (ANalysis Of VAriance)	49
II Méthodes multivariées	49
II.1 Méthodes « Non supervisées »	49
II.2 Méthodes « Supervisées »	51
III Validation	52
III.1 Evaluation de la performance du modèle	52
III.2 Autres techniques de validation	54
IV Intérêt des modèles appariés	55
V Technique complémentaire pour l'interprétation des données : la STOCASY	56

<b>Métabolomique et atteintes hépatiques</b>	<b>57</b>
<b>Chapitre I Le Foie</b>	<b>57</b>
I Fonction du foie	57
II Les atteintes hépatiques	58
II.1 Symptômes	58
II.2 Gradation des atteintes hépatiques	59
II.3 Principales causes	61
II.4 Stéatose hépatique non alcoolique (NAFLD) et stéato-hépatite non alcoolique (NASH)	61
II.5 Moyens diagnostiques et biomarqueurs cliniques	62
II.6 Utilisation de scores pronostiques	63
III Le carcinome hépatocellulaire (CHC)	64
III.1 Dépistage	64
III.2 Traitements	64
<b>Chapitre II Etat de l'art des études métabolomiques dans les atteintes hépatiques</b>	<b>69</b>
I Approches métabolomiques des atteintes hépatiques par spectrométrie de masse	70
I.1 Infection par le VHB/VHC	70
I.2 Insuffisance hépatique	70
I.3 Evaluation des atteintes « NAFLD »	70
I.4 Atteintes chroniques du foie menant au CHC	72
II Approches métabolomiques des atteintes hépatiques par <sup>1</sup> H RMN	76
<b>Travaux de recherche</b>	<b>107</b>
<b>Chapitre I Stéatose hépatique non alcoolique et stéato-hépatite non alcoolique : Approche diagnostique par étude métabolomique RMN à partir d'extractions de tissus de foies humains.</b>	<b>107</b>
I Contexte introductif	107
II Matériel et Méthode	109
II.1 Constitution de la cohorte	109
II.2 Méthodes d'extraction des métabolites	110
II.3 Analyses RMN	111
II.4 Traitement des spectres	111
II.5 Analyses statistiques	111
III Résultats & Discussion	112
IV Conclusion	123
<b>Chapitre II Etude métabolomique du carcinome hépatocellulaire précoce : Approche pronostique et suivi de traitement par ablation par radiofréquence</b>	<b>125</b>
<b>Chapitre III Validation des modèles statistiques : impact d'une permutation des individus dans la matrice des données sur les paramètres de qualité du modèle</b>	<b>157</b>
<b>Conclusions &amp; Perspectives</b>	<b>167</b>
<b>Annexes</b>	<b>171</b>
<b>Bibliographie</b>	<b>173</b>

## Liste des abréviations

---

AA	Acides Aminés
ACLF	<i>Acute on Chronic Liver Failure</i>
ACP	Analyse en Composantes Principales
ADN	Acide Désoxyribonucléique
AFP	Alpha-foetoprotéine
ALAT	Alanine Aminotransférase
ANOVA	<i>Analysis Of Variance</i>
ARN	Acide Ribonucléique
ASAT	Aspartate Aminotransférase
AUROC	<i>Area Under Receiver Operating Characteristic</i>
BCLC	<i>Barcelona Clinic Liver Cancer</i>
CE	<i>Capillary Electrophoresis</i>
CHC	Carcinome Hépatocellulaire
CP	Composante Principale
CPMG	« <i>Carr-Purcell-Meiboom-Gill</i> »
Cor	Corrélation
Cov	Covariance
CV	<i>Cross Validation</i>
2CV	<i>Double Cross Validation</i>
Da	Dalton
DSS	<i>3-(trimethylsilyl)propane-1-sulfonic acid</i>
ExpASy	<i>Expert Protein Analysis System</i>
FID	<i>Free Induction Decay</i>
GC	<i>Gas Chromatography</i>
GGT	Gamma-glutamyl transférases
H <sub>0</sub>	Hypothèse nulle
HDL	<i>High Density Lipoprotein</i>
HMDB	<i>Human Metabolomic Data Basis</i>
HPLC	High Performance Liquid Chromatography
HR-MAS	<i>High Resolution Magic Angle Spinning</i>
IMC	Indice de Masse Corporel
INR	<i>International Normalized Ratio</i>
IRM	Imagerie par Résonance Magnétique
KEGG	<i>Kyoto Encyclopedia of Genes and Genomes</i>
LC	<i>Liquid chromatography</i>
LDL	<i>Low Density Lipoprotein</i>
LPC	<i>Lysophosphatidylcholines</i>
MCCV	<i>Monte Carlo Cross Validation</i>
MELD	<i>Model for End-Stage Liver Disease</i>
MHz	MegaHertz
MS	<i>Mass Spectrometry</i>
MUFA	Monounsaturated Fatty Acid
NAFLD	<i>Non-Alcoholic Fatty Liver Disease</i>

---

NASH	Nonalcoholic Steatohepatitis
NOESY	<i>Nuclear Overhauser Effect Spectroscopy</i>
OMS	Organisation Mondiale de la Santé
OPLS	<i>Orthogonal Partial Least Squares</i>
OPLS-DA	<i>Orthogonal Partial Least Squares-Discriminant Analysis</i>
OSC	<i>Orthogonal Signal Correlation</i>
PC	phosphatidylcholine
PC 1	<i>Principal Component 1</i>
PC 2	<i>Principal Component 2</i>
PEI	<i>Percutaneous Ethanol Injection</i>
PLS	<i>Partial Least Squares</i>
PLS-DA	<i>Partial Least Squares-Discriminant Analysis</i>
Ppm	Parties Par Million
PRESS	<i>Predicted Residual Sum of Squares</i>
PST	<i>Performance Status Test</i>
PUFA	<i>Polyunsaturated Fatty Acid</i>
RF	Radiofréquence
RFA	<i>Radiofrequency Ablation</i>
RMN	Résonance Magnétique Nucléaire
ROC	<i>Receiver Operating Characteristic</i>
SRM	Spectrométrie par Résonance Magnétique
SS	<i>Sum of Squares</i>
STOCSY	<i>Statistical Total Correlation Spectroscopy</i>
TACE	<i>Transarterial Chemoembolization</i>
TCA	<i>Tricarboxylic acid cycle</i>
TG	Triglycérides
TMAO	<i>Trimethylamine N-oxide</i>
TOCSY	<i>Total Spectroscopy Correlation</i>
TOF	<i>Time Of Flight</i>
TSP	<i>Trimethylsilylpropionic acid</i>
UPLC	<i>Ultra Performant Liquid Chromatography</i>
UV	<i>Unit Variance</i>
VHB	Virus Hépatite B
VHC	Virus Hépatite C
VLDL	<i>Very Low Density Lipoprotein</i>

## Liste des figures

---

**Figure 1 :** Représentation schématique des différents niveaux de caractérisation des systèmes biologiques

**Figure 2 :** Schéma d'une expérience type de métabolomique

**Figure 3 :** Illustration du concept de biologie des systèmes par une vision intégrée des différents niveaux des sciences « omiques » et de l'interaction des facteurs environnementaux

**Figure. 4 :** Représentation de l'effet Zeeman pour un noyau  $I=1/2$ , avec  $B_0$ , le champ magnétique initial et  $\Delta E$ , la différence d'énergie entre les 2 états  $\alpha$  et  $\beta$

**Figure. 5 :** Représentation du mouvement de précession autour de  $B_0$

**Figure. 6 :** Basculement de l'aimantation macroscopique  $M_0$  sous l'effet d'une impulsion de radiofréquence

**Figure. 7 :** Retour à l'équilibre de l'aimantation macroscopique

**Figure. 8 :** Application de la transformée de Fourier permettant de transposer le FID en signal de fréquence

**Figure. 9 :** Représentation schématique de l'appareillage RMN

**Figure. 10 :** Séquence noesypr1d telle que décrite par le constructeur Bruker

**Figure. 11 :** Séquence CPMG telle que décrite par le constructeur Bruker

**Figure. 12 :** Spectres RMN de sérum obtenus, en rouge, avec application de la séquence CPMG et en bleu, avec application de la séquence nosygppr1D

**Figure. 13 :** Séquence dipsi2phpr telle que décrite par le constructeur Bruker

**Figure. 14 :** 2D-TOCSY (dipsi2phpr) comprise entre 0 et 5.5 ppm d'un échantillon de sérum illustrant la connectivité des protons correspondant à l' $\alpha$ -glucose

**Figure. 15 :** Effet du réalignement sur le signal du doublet de l'alanine : avant (a) et après traitement (b)

**Figure. 16 :** Découpage équidistant du spectre en régions appelées « buckets »

**Figure. 17 :** Exemple d'attribution du métabolite du glucose sur un spectre CPMG obtenu à partir de sérum à l'aide du logiciel « Chenomx NMR Suite ». En noir est représenté le spectre expérimental tandis que les signaux en bleu correspondent aux signaux du glucose pur issus de la banque de données

**Figure. 18 :** Représentation schématique de l'analyse du métabolome sérique obtenue par les différentes techniques de RMN, LC-MS, GC-MS et par la combinaison des deux techniques GC et LC-MS.

**Figure. 19 :** Représentation imagée du principe de l'analyse en composantes principales où un objet tridimensionnel est projeté dans un espace bidimensionnel

**Figure. 20 :** Illustration du principe de validation croisée pour  $k=3$

**Figure. 21 :** Exemple de corrélation STOCSY des signaux de la créatine (2 singulets à 3.02 ppm et 3.92 ppm) réalisé à partir d'un spectre NOESY-1D d'une extraction de métabolites polaires de tissu de foie humain

**Figure. 22 :** Représentation simplifiée du foie

**Figure. 23 :** Aspect macroscopique d'un foie : sain, avec fibrose, avec cirrhose et cancéreux

**Figure. 24 :** Représentation de la destruction tumorale par application de radiofréquences

**Figure. 25 :** Score plot de l'ACP obtenue à partir des extractions aqueuses des 106 échantillons dupliqués correspondant à 53 patients. Les 3 points entourés correspondent à une mauvaise manipulation lors de l'extraction

**Figure. 26 :** Score plot de l'ACP obtenue à partir des extractions organiques des 106 échantillons dupliqués correspondant à 53 patients

**Figure. 27 :** Score (A) et loading plot (B) de l'analyse OPLS-DA réalisée sur les fractions aqueuses des 32 échantillons ayant subi un cycle de décongélation/recongélation (en rouge) et des 40 autres échantillons préservés de cet événement (en bleu). Les variations significatives des métabolites sont représentées entre 1 et 4.5 ppm sur le loading plot. Un signal positif correspond aux métabolites présents à des concentrations plus élevées pour les échantillons ayant subi le cycle de décongélation/recongélation. À l'inverse, les signaux négatifs correspondent aux métabolites présents à des concentrations plus élevées pour les échantillons préservés du cycle décongélation/recongélation. Annotations : 1= lactate ; 2= glutamate ; 3= glutathion ; 4= pyruvate ; 5= glycine ; cor= corrélation

**Figure. 28 :** Loading plots de l'analyse OPLS-DA obtenue à partir des fractions aqueuses des échantillons (A) de foies normaux et stéatoses ; (B) de foies normaux et NASH. Les signaux des N-acétyl des glycoprotéines (1), du glutamate (2) et de la créatine (3) semblent être corrélés en univarié et augmentés chez les foies normaux comparés aux stéatoses et aux NASH

**Figure. 29 :** Représentation des valeurs en boîte à moustache de la créatine (A) et du groupement N-acétyl des glycoprotéines (B) pour les 3 groupes correspondant aux extraits aqueux des tissus de foies normaux, stéatoses et NASH. Moy= moyenne ; 1<sup>e</sup> q= 1<sup>er</sup> quartile ; 3<sup>e</sup> q= 3<sup>e</sup> quartile. Les croix rouges correspondent aux échantillons aberrants

**Figure. 30 :** Loading plots de l'analyse OPLS-DA obtenue à partir des fractions organiques des échantillons (A) de foies normaux et stéatoses ; (B) de foies normaux et NASH. L1= groupement lipidique  $\text{CH}_3\text{-CH}_2\text{-}$  (méthyle) ; L2= groupement lipidique :  $\text{-(CH}_2\text{)}_n\text{-}$  (méthylène) ; L3= groupement lipidique  $\text{-CH}_2\text{-CH}_2\text{CO-}$  (méthylène des L5) ; L4= groupement lipidique  $\text{-CH=CH-CH}_2\text{-CH}_2\text{-}$  (allylique) ; L5= groupement lipidique  $\text{-CH}_2\text{-CH}_2\text{-CO-}$  ( $\alpha$ -méthylène du carbonyle) ; L6= groupement lipidique  $\text{-CH=CH-CH}_2\text{-CH=CH-}$  (diallylique – acides gras polyinsaturés « PUFA ») ; L7= groupement lipidique  $\text{-CH=CH-}$  (n-ene – acides gras monoinsaturés « MUFA ») ; group. glycéryl= groupement glycéryl  $\text{CH}_2\text{OR}$

## Liste des tableaux

---

**Tableau 1.** Résumé des méthodes de normalisation des variables qui sont principalement utilisées au laboratoire ;  $\bar{x}_i$  correspond à la moyenne des variables ;  $s_i$  à l'écart-type des variables et  $\tilde{x}_{ij}$  représente la valeur des données après normalisation

**Tableau 2.** Classification BCLC pour l'évaluation de la prise en charge du CHC

**Tableau 3.** Résumé des études métabolomiques par MS des atteintes chroniques du foie menant au CHC

**Tableau 4.** Caractéristiques clinicopathologiques des patients inclus dans l'étude

**Tableau 5.** Echantillonnage pour l'extraction



# Introduction générale

---

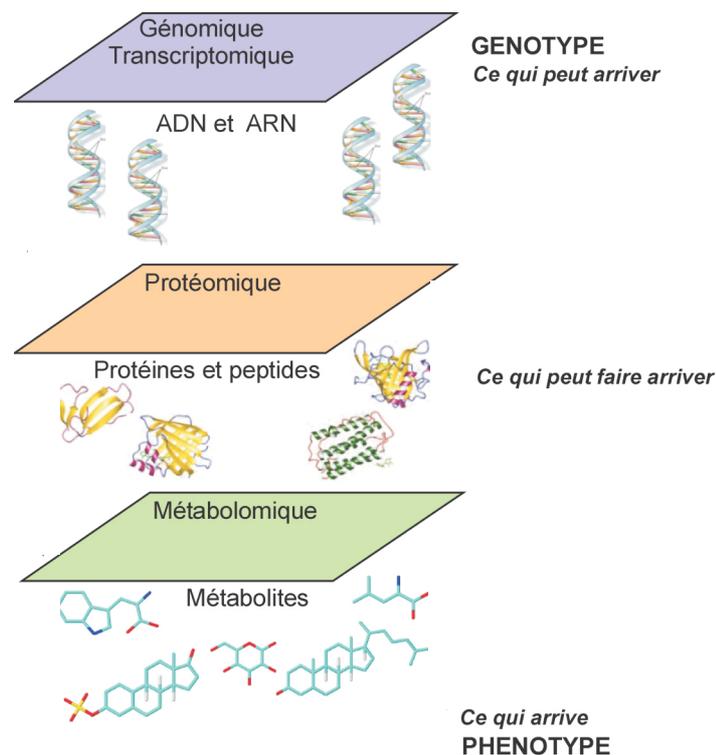
Les atteintes hépatiques souvent asymptomatiques et provoquées par différentes étiologies représentent aujourd'hui un enjeu majeur en recherche biomédicale. En France, environ 2 millions de personnes souffrent d'une maladie chronique du foie pouvant conduire à des complications sévères telles que le carcinome hépatocellulaire dont plus de 8000 français sont touchés chaque année (Données InVS et INCa, 2013).

Les enjeux actuels de la recherche biomédicale portent, par conséquent, sur la compréhension de l'interaction de l'environnement extérieur sur le métabolisme. Ils relèvent en hépatologie, d'une meilleure connaissance des mécanismes responsables du déclenchement et de la progression des atteintes hépatiques, de l'établissement d'une classification optimale des différents stades de ces atteintes ainsi que du développement de traitements ciblés voire personnalisés [1]. De plus, l'identification de biomarqueurs spécifiques des atteintes hépatiques est nécessaire pour permettre de meilleurs diagnostics mais également pour suivre l'évolution des maladies et les effets thérapeutiques. Ce concept de traitement appliqué suivant les caractéristiques du patient appelé médecine personnalisée représente une des voies les plus audacieuses dans la recherche de traitements médicaux [2,3].

Pour être capable de préciser le pronostic et d'adapter de nouvelles procédures thérapeutiques individualisées, des approches translationnelles, associant à la fois recherche clinique et recherche fondamentale deviennent incontournables. A cette fin, la détermination de nouveaux biomarqueurs est devenue un enjeu particulièrement important [4].

Parallèlement, le développement de nouvelles approches et méthodes de criblage à haut débit et plus spécifiquement l'avènement de la biologie systémique et des sciences « omiques » a permis d'accélérer la recherche de biomarqueurs d'exposition et/ou d'effets [5]. Au même titre que le génome, le transcriptome et le protéome étudient l'ensemble des

gènes, des transcrits, et des protéines, le métabolome fait référence à l'ensemble des métabolites contenus dans un système biologique donné. Le métabolome représente l'ultime réponse d'un organisme à une altération génétique, une pathologie, une exposition à un toxique ou à tout autre facteur susceptible de perturber son fonctionnement. A l'instar des autres sciences « omiques », la métabolomique (étude du métabolome) décrit par conséquent un état physiologique donné et consiste en l'étude des modifications des activités biologiques qui interviennent à tous les niveaux d'organisation du vivant (Fig.1) [6].



**Fig. 1** : Représentation schématique des différents niveaux de caractérisation des systèmes biologiques (Ipubli, Inserm, 2011)

La métabolomique est une approche capable de détecter et de quantifier des petites molécules impliquées dans les voies métaboliques des cellules, des organes et des tissus. Ces molécules appelées métabolites représentent les produits finaux des procédés de régulation cellulaire et leurs niveaux quantitatifs peuvent être non seulement la réponse des systèmes biologiques à des modifications génétiques, environnementales et thérapeutiques mais également peuvent servir de contrôle à une préservation de l'homéostasie et de la

spécificité des tissus [7]. L'avantage de la métabolomique est par conséquent de caractériser de façon assez globale et non ciblée ces perturbations biologiques. Cette technique combine ainsi profilages métaboliques et analyses statistiques pour déceler un ensemble de biomarqueurs impliqués préférentiellement dans la discrimination de différents sous-groupes de populations par exemple contrôles *versus* malades. Comme décrit par J. Nicholson dans *Nature* [8], la métabolomique fait partie des nouvelles techniques de médecine personnalisée qui dans un futur proche deviendra indispensable pour la recherche et la compréhension globale des voies d'apparition des maladies et pour l'élaboration de suivis thérapeutiques ciblés.

L'objectif général de cette thèse a été d'examiner à partir de problématiques hépatiques déterminées les apports des analyses métabolomiques par résonance magnétique nucléaire (RMN) du proton.

Il s'inscrit de cette manière dans un cadre plus global de recherches pluridisciplinaires qui met en place des collaborations étroites entre chercheurs et cliniciens. Ce projet donne suite aux collaborations fructueuses nouées entre l'équipe de recherche « RMN-métabolomique » du laboratoire CSPBAT et les services de chirurgie viscérale et d'hépatologie de l'hôpital Jean Verdier de Bondy ainsi qu'avec l'équipe « Mitochondries, bioénergétique, métabolisme et signalisation » de l'unité Inserm U1016.

La première partie de cette thèse porte sur l'étude métabolomique par RMN d'extractions de tissus de foies humains dans une approche diagnostique des hépatopathies stéatosiques non alcooliques (NAFLD). Caractérisées par une accumulation excessive de lipides dans le foie sous forme de triglycérides (stéatose), ces atteintes sont en augmentation croissante depuis ces dernières années principalement dans les pays industrialisés et mènent à des complications bien souvent diagnostiquées tardivement.

La seconde partie présentée dans ce manuscrit concerne une des complications les plus fréquentes de la cirrhose qui est le carcinome hépatocellulaire (CHC). Cette atteinte hépatique progressant de manière rapide et insidieuse est très souvent le résultat d'un pronostic peu optimiste avec un taux de survie à 5 ans ne dépassant pas 5%. Le dépistage précoce du CHC rend possible l'application de traitements curatifs portant ainsi la survie

globale de ces patients à 50% à 5 ans [9]. Ainsi, nous avons entrepris par une approche métabolomique par RMN, l'analyse du suivi sérique de patients atteints d'un CHC précoce et traités dans un but curatif par un traitement de type ablation par radiofréquence afin de pouvoir fournir de nouvelles pistes visant à améliorer la prise en charge thérapeutique du patient.

Enfin, la dernière partie des travaux de la thèse démontre l'importance de la validation lors de la détermination de modèles prédictifs en analyses statistiques.

# Méthodologies

---

## Chapitre I La métabolomique

### I Définition

A l'origine, deux définitions ont été établies pour l'étude du métabolisme :

- La métabolomique, qui consiste à décrire et à quantifier systématiquement la totalité des métabolites présents dans un système biologique (tel qu'un liquide biologique, une cellule ou un tissu) à un moment donné [10].

- La métabonomique, définie en 1999 par J. Nicholson comme étant « la mesure quantitative de la dynamique de la réponse métabolique des systèmes vivants suite aux stimuli pathophysiologiques ou aux modifications génétiques » [11].

Aujourd'hui ces deux définitions sont confondues et l'utilisation des deux termes est considérée comme équivalente [12]. Nous adopterons, par facilité pour la suite de l'exposé, le terme « métabolomique ».

Tout comme la génomique, la transcriptomique ou encore la protéomique, la métabolomique fait partie de ces disciplines « omiques » qui a pour but de déterminer de manière extensive l'ensemble des métabolites. Elle consiste en l'analyse qualitative et quantitative du métabolome qui représente le résultat des réactions métaboliques se déroulant au sein d'une cellule, d'un tissu ou d'un organe. Son but est l'identification, la mesure et l'interprétation de la concentration des métabolites responsables d'un comportement biologique grâce à l'analyse spectrométrique de liquides biologiques et/ou de tissus et par l'interprétation des données par des méthodes chimiométriques. Elle dresse par conséquent le portrait complet et systématique des métabolites et de leurs changements temporels provoqués par des facteurs tels que l'alimentation, le mode de vie, l'environnement, les effets post-génétiques ou encore pharmaceutiques [13,14].

Les métabolites sont définis par des molécules de faible poids moléculaire (inférieur au kDa) participant au catabolisme et à l'anabolisme de l'organisme (sucres, acides aminés, acides gras,..). Ils représentent les produits des différents processus cellulaires et sont par conséquent les plus descriptifs du phénotype d'un organisme. La plupart d'entre eux sont connus et ont été mesurés par des méthodes biochimiques ou physicochimiques mais

certains restent encore non identifiés ou inexploités [13]. De plus, le nombre de métabolites diffère non seulement entre les différents types d'échantillons étudiés (liquides biologiques, tissus, type cellulaire) mais aussi entre les individus ou organismes étudiés. Chaque cellule, tissu ou organisme est fondamentalement unique mais peut néanmoins présenter quelques caractéristiques communes comme l'appartenance à une espèce, l'atteinte de la même maladie ou l'administration d'un même traitement. Le véritable challenge de la métabolomique est d'identifier l'ensemble de ces petites variations chimiques face à un ensemble large et complexe de métabolites définissant l'individu [3,7,15].

Deux approches peuvent être entreprises en métabolomique [6] :

- L'approche globale dite d'empreinte métabolique. Cette approche représente une analyse globale, sans a priori sur la composition de l'échantillon. Cette méthode est la méthode privilégiée pour établir une cartographie complète et celle que nous utilisons le plus fréquemment au sein du laboratoire.
- L'approche ciblée. Cette approche est utilisée pour une étude ciblée de voies métaboliques. Elle est plus souvent entreprise après exploration par une approche globale.

## II Démarche et cheminement d'une analyse métabolomique

Une étude métabolomique est constituée de plusieurs étapes dont le cheminement est schématisé en figure 2.

Toute démarche métabolomique est à l'origine initiée par une ou plusieurs question(s) biologique(s) ou clinique(s) à laquelle on souhaite répondre. Qu'il s'agisse de suivre les effets d'un traitement ou encore d'améliorer les outils de diagnostic, le design expérimental doit être consciencieusement réfléchi afin de réduire les biais et éviter l'introduction d'une trop grande variabilité. On peut être amené, lors de la conception de l'étude, à établir des recommandations plus spécifiques ou des critères d'inclusion déterminés afin d'avoir une population la plus homogène possible. En raison de l'influence de l'environnement exogène sur les voies métaboliques, il est indispensable d'établir un protocole expérimental bien déterminé et reproductible en ce qui concerne la collecte, la conservation ainsi que la préparation des échantillons [16,17].

Les échantillons étudiés en métabolomique sont des prélèvements de liquides biologiques, de cellules ou de tissus. Les prélèvements d'urine et de plasma ou de sérum sont les

prélèvements les plus étudiés parce qu'ils peuvent être recueillis sans trop de difficulté, en abondance suffisante et de manière non invasive [18]. Cependant, il est également possible de réaliser une étude métabolomique sur d'autres liquides biologiques tels que : le liquide sérial, amniotique, synovial, céphalo-rachidien, digestif [19-22] ou sur des pièces de biopsies ou excrèses cliniques [23-25].

Les échantillons sont ensuite analysés par l'une des deux techniques analytiques privilégiées en métabolomique : la Spectrométrie de Masse (MS) ou la Résonance Magnétique Nucléaire (RMN) qui seront détaillées par la suite. Différentes étapes de traitement des données sont alors entreprises avant de procéder à l'analyse statistique. La dernière étape de l'étude métabolomique constitue sans doute une des étapes les plus complexes qui est l'identification et l'interprétation des métabolites [13].

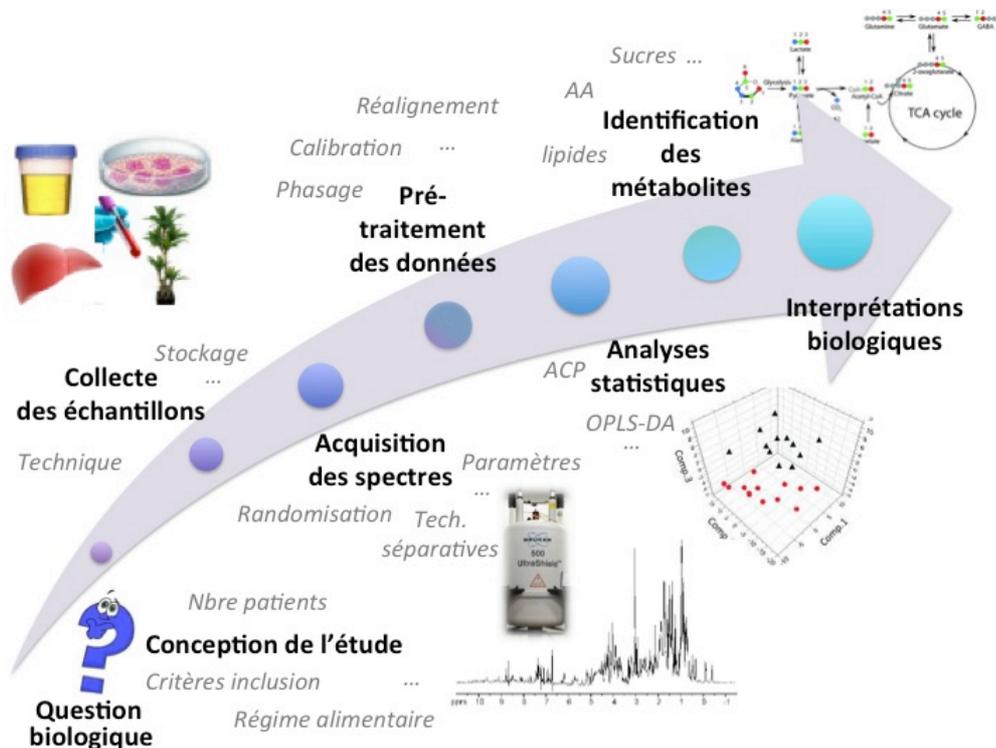


Fig. 2 : Schéma d'une expérience type de métabolomique

### III Domaines d'application

La métabolomique intervient dans des domaines de recherche très variés, tels que l'environnement, la toxicologie, la nutrition, la recherche biomédicale de la recherche en pharmacologie au dépistage de maladies et cancers.

A l'origine, les premiers domaines d'applications ont été entrepris en recherche biomédicale par l'étude de la RMN des liquides biologiques menant à la conceptualisation du terme métabon(l)omique introduit à la fin des années 1990 par J. Nicholson et J. Lindon. Un consortium pour l'évaluation de la toxicité des xénobiotiques en métabolomique (Consortium for Metabolomic Toxicology : COMET) avait initialement été formé par la collaboration de cinq grandes compagnies pharmaceutiques et de l'Imperial College de Londres en 1998 [26].

Depuis les applications se retrouvent de plus en plus nombreuses témoignant de l'essor de cette technique dans la communauté médicale. La métabolomique offre la possibilité de découvrir de nouveaux biomarqueurs tant pour le diagnostic que le pronostic et mène à une meilleure compréhension de l'évolution des maladies [3,27,28]. Duarte *et al.* [29] ont retracé, depuis ces 5 dernières années, un certain nombre d'études entreprises en RMN métabolomique appliquées dans la recherche de maladies. Elles portent sur différents types de cancers mais aussi sur de nouvelles investigations de troubles digestifs, nerveux et respiratoires.

En toxicologie, les impacts de certaines substances (phtalates, bisphénol A,..) ainsi que leurs biomarqueurs d'exposition ou d'effet sont régulièrement étudiés grâce aux analyses métabolomiques de systèmes biologiques donnés [30-34].

Le secteur de l'environnement est également un domaine d'application très souvent retrouvé en métabolomique. Les différentes études environnementales portent sur le fonctionnement, l'évolution et l'adaptation des organismes végétaux [35-38] mais s'étendent également dans des domaines de chimie verte, biotechnologies, biologie synthétique, biodiversité et écologie ainsi que dans l'amélioration génétique [39-41].

Au cours de ces dernières années, la métabolomique apparaît également de plus en plus prometteuse dans le domaine de la nutrition, particulièrement pour la caractérisation du rôle des aliments dans la régulation des voies métaboliques. Elle est utilisée pour évaluer

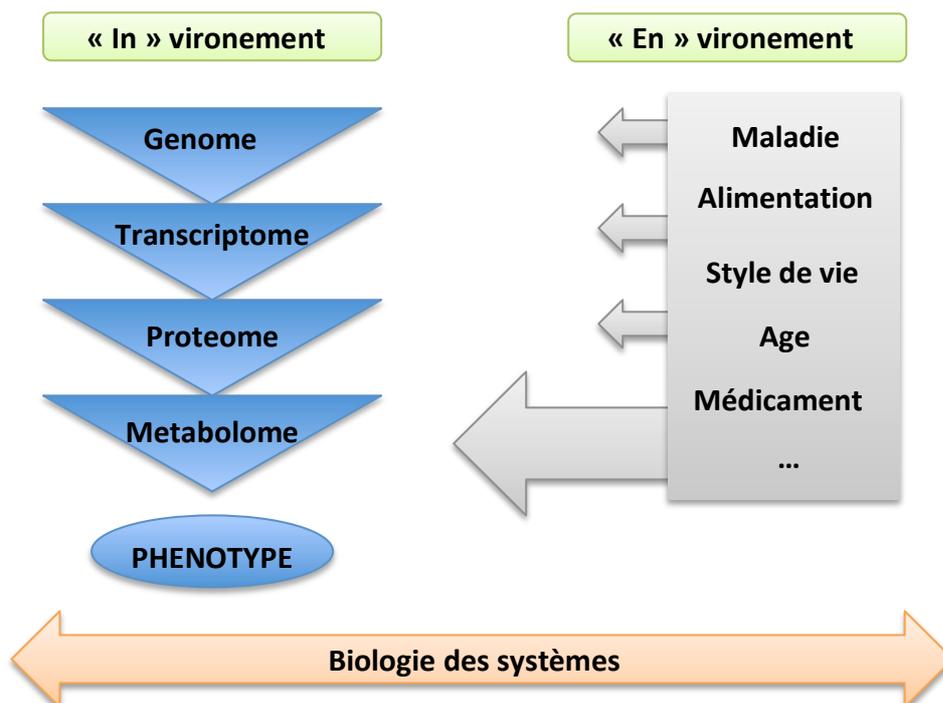
l'état nutritionnel des individus et faciliter la découverte de nouveaux biomarqueurs associés à divers dysfonctionnements métaboliques [42-46].

#### IV Vers la biologie des systèmes

##### ***IV.1 Croisée des sciences « omiques »***

De nos jours, il semble évident que la complexité d'un organisme ne peut être interprétée en se limitant à un seul domaine d'étude (l'analyse des gènes, des transcrits, des protéines ou des métabolites). La combinaison des approches appelée de manière plus générale, biologie des systèmes, représente une stratégie de recherche intégrée qui combine des approches expérimentales et computationnelles pour améliorer la compréhension des systèmes biologiques complexes [5,15]. La biologie des systèmes cherche, de cette manière, à étudier la façon dont les relations entre les composants de l'organisme vivant, gènes, messagers protéiniques, cellules et organes s'établissent au sein d'un système global. Un de ses objectifs est d'exploiter simultanément les données issues de ces différentes approches nécessitant par conséquent des outils bioinformatiques puissants [2]. L'intégration des sciences « omiques » ainsi que le développement de nouvelles méthodes d'analyses de données permettraient de cette façon une meilleure compréhension des processus biologiques d'organismes complexes. Ainsi, l'analyse combinée des sciences « omiques » constitue aujourd'hui un domaine émergent qui s'avère être prometteur pour aider à appréhender la complexité du vivant dans son ensemble (Fig.3) [47]. L'objectif est de combiner ces différentes plateformes « omiques » afin d'améliorer la prédiction des modèles pour des applications menant à la contribution de la conception de nouvelles cibles et traitements thérapeutiques, à la découverte de signatures typiques de maladie ainsi qu'à une meilleure compréhension des principes basiques de biochimie [48-50].

Néanmoins, certaines limitations sur le plan explicatif, sur la validité ainsi que l'interprétation biologique des biomarqueurs sont encore à améliorer nécessitant une collaboration plus étendue entre différents domaines de la recherche (biologie moléculaire, chimie, physique, bioinformatique) [6].



**Fig. 3 :** Illustration du concept de biologie des systèmes par une vision intégrée des différents niveaux des sciences « omiques » et de l'interaction des facteurs environnementaux [47].

#### **IV.2 Reconstruction des réseaux métaboliques**

La reconstruction des réseaux métaboliques à partir des données de métabolomique ou de toute autre approche « omique » constitue une étape importante pour l'interprétation biologique des perturbations observées. Cette reconstruction permet de retrouver les voies biochimiques qui sont affectées par la perturbation biologique étudiée. Différentes bases de données existent, parmi les plus utilisées nous pouvons citer : KEGG (Kyoto Encyclopedia of Genes and Genomes), ExPASy (Expert Protein Analysis System) ou encore Cytoscape [51-54].

### **V Techniques de spectrométrie utilisées en métabolomique**

La grande diversité des propriétés physicochimiques des métabolites rend compliquée l'analyse de l'ensemble des métabolites par un outil d'analyse privilégié.

Deux techniques sont couramment utilisées en métabolomique : la Spectrométrie de Masse (MS) et la Résonance Magnétique Nucléaire (RMN). Chacune d'entre elles présente des avantages et des limitations (Annexe 1). Comparée à la spectrométrie de masse, la RMN est quantitative et ne requiert pas d'étapes préliminaires de préparation d'échantillons telles

qu'une technique de séparation ou de dérivatisation. Par conséquent, elle présente l'avantage de pouvoir détecter une plus grande variété de métabolites (chargés, neutres, hydrophobes, hydrophiles) ainsi que d'être plus flexible concernant la nature de l'échantillon. Néanmoins, malgré les améliorations constantes dans ce domaine, la sensibilité de la RMN reste un point faible comparée à celle de la spectrométrie de masse. L'utilisation de la spectrométrie de masse requiert quant à elle à une technique de pré-séparation des métabolites utilisant la chromatographie gazeuse (GC) après dérivatisation, la chromatographie liquide (LC, HPLC, UPLC) ou encore l'électrophorèse capillaire (CE) [55-57].

### **V.1 La spectrométrie de masse**

La spectrométrie de masse est une technique d'analyse extrêmement sensible qui permet de détecter et de déterminer des structures moléculaires en les fragmentant. Son principe réside dans la séparation en phase gazeuse de molécules chargées en fonction de leur rapport masse/charge ( $m/z$ ). Il existe différents types de spectromètres, néanmoins ils ont tous en commun trois éléments qui peuvent différer suivant le type : une source d'ionisation, un analyseur et un détecteur. L'ionisation par champ électrique (*electrospray*) est une des méthodes de choix en métabolomique car l'une des plus polyvalentes. Au niveau des analyseurs, la trappe ionique et le temps de vol haute résolution (« *time of flight* », TOF) sont également les plus utilisés pour le profilage métabolique.

L'ionisation des métabolites est intrinsèquement dépendante de la source et loin d'être reproductible d'une famille de composés à l'autre. La spectrométrie de masse permet en revanche une détection moléculaire de l'ordre du pico au femto molaire [58,59].

### **V.2 La Résonance Magnétique Nucléaire**

La Résonance Magnétique Nucléaire (RMN), principal outil spectroscopique utilisé en chimie, permet de réaliser une analyse chimique et structurale non destructive.

La RMN est une technique robuste qui permet d'analyser des centaines d'échantillons en un temps restreint. La RMN permet l'analyse de liquides biologiques et de tissus intacts avec un minimum de manipulations par l'expérimentateur. Elle présente l'avantage d'être non destructive, néanmoins, en raison de sa faible sensibilité elle ne permet pas la détection de métabolites de concentration inférieure au  $\mu$ molaire [13,60]. Cette technique constitue la

technique de choix utilisée dans les différentes études présentées dans cette thèse et est pour cette raison décrite plus en détails dans le point suivant.

## Chapitre II La RMN en métabolomique

### I Principe

Le principe de la RMN est basé sur les propriétés magnétiques de certains noyaux ayant un spin non nul. Le phénomène de RMN a lieu lorsqu'un noyau atomique placé dans un champ magnétique est soumis à une onde électromagnétique de fréquence spécifique au noyau. On assiste alors à une concordance de fréquences entre le noyau et l'onde par un transfert d'énergie.

L'énergie de l'interaction des moments magnétiques des noyaux interagissant avec le champ magnétique extérieur est constituée de niveaux discrets caractéristiques ( $E_i$ ) (issus de la mécanique quantique). L'émission de radiofréquences à la fréquence des noyaux étudiés induit des transitions entre ces niveaux d'énergie. L'absorption d'énergie qui en résulte est enregistrée sous forme de signal de résonance, représenté par un ensemble de pics sur le spectre RMN.

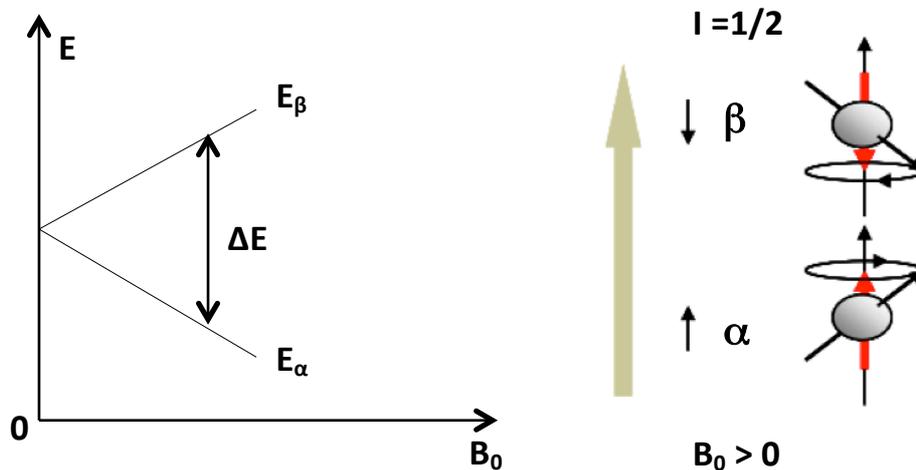
#### ***1.1 Propriétés magnétiques du noyau atomique***

Toutes les particules possèdent un spin qui caractérise le magnétisme du noyau. La RMN s'applique aux noyaux possédant un moment magnétique non nul. La parité du nombre de masse ainsi que du nombre atomique entraînent un moment magnétique nucléaire nul. Ces noyaux sont alors dits non-sensibles (comme par exemple :  $^{12}\text{C}$  ou  $^{16}\text{O}$ ). Dans les autres cas, deux conditions de résonance peuvent être observées. Dans le premier cas, les noyaux possédant un spin nucléaire  $I = 1/2$  tel que le  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{31}\text{P}$ ,  $^{15}\text{N}$  ou le  $^{19}\text{F}$  vont avoir une circulation de charge sphérique et posséder deux moments magnétiques. Dans le second cas, les noyaux possèdent un spin nucléaire  $I > 1/2$  avec une répartition non sphérique de la charge nucléaire. De manière à avoir une sensibilité et une résolution optimale, les noyaux les plus utilisés en RMN sont les noyaux de spin  $\frac{1}{2}$  [61].

#### ***1.2 Les états énergétiques***

Soumis à un champ magnétique  $B_0$ , un noyau de spin  $I$  non nul peut prendre  $2I+1$  orientations par rapport à la direction du champ magnétique (effet Zeeman). Dans le cas des noyaux de spin  $I = 1/2$  (tel que l'atome d'hydrogène,  $^1\text{H}$ ), ils s'orientent suivant deux directions : parallèle et anti-parallèle au champ magnétique. Autrement dit, les spins

nucléaires peuvent occuper deux niveaux d'énergie bien distincts : un état énergétique favorable,  $\alpha$ , qui, en situation d'équilibre est légèrement plus peuplé (distribution de Boltzmann des spins) et un état énergétique défavorable  $\beta$  (Fig. 4) [62,63].



**Fig. 4 :** Représentation de l'effet Zeeman pour un noyau  $I=1/2$ , avec  $B_0$ , le champ magnétique initial et  $\Delta E$ , la différence d'énergie entre les 2 états  $\alpha$  et  $\beta$ .

### 1.3 L'équation de Larmor

Le noyau peut être représenté par un vecteur de moment magnétique  $\vec{\mu}$  dont la composante  $\mu_z$  peut s'orienter parallèlement ou antiparallèlement au champ.  $\mu_z$  est dépendante du rapport gyromagnétique spécifique du noyau  $\gamma$  et du moment de spin  $I_z$  :

$$\mu_z = \gamma I_z$$

$$\text{avec } I_z = m_l \frac{h}{2\pi} \quad (m_l = \text{nombre quantique magnétique, } \pm \frac{1}{2} \text{ pour le } ^1\text{H})$$

On peut donc considérer le proton comme un dipôle magnétique qui tourne autour de l'axe du champ appliqué  $B_0$  (axe z) à la vitesse :

$$\omega_0 = 2\pi\nu_0$$

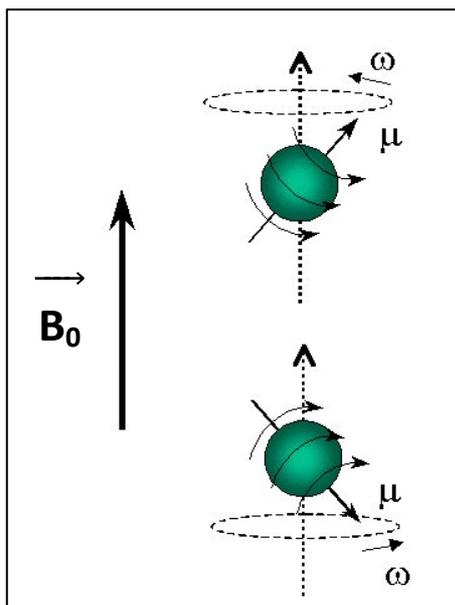
L'énergie d'un dipôle magnétique dans un champ magnétique est donnée par :

$$E = -\mu_z B_0$$

Sachant qu'une transition entre l'état  $E_\alpha$  et l'état  $E_\beta$  se produit quand un rayonnement électromagnétique d'énergie  $\Delta E = h\nu$  est absorbé,

alors, la condition de résonance est :  $h\nu_0 = \gamma \frac{h}{2\pi} B_0$

$$\omega_0 = \gamma B_0 \quad \text{Equation de Larmor}$$

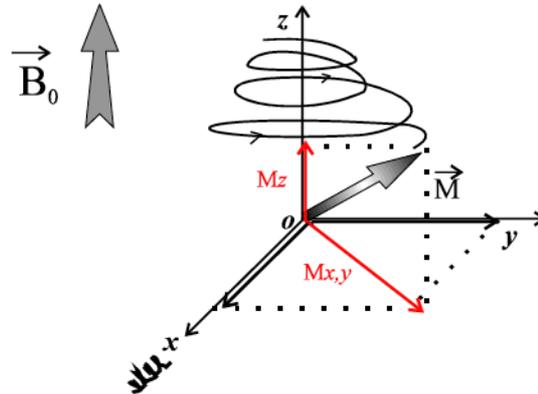


**Fig. 5 :** Représentation du mouvement de précession autour de  $B_0$  (Pr. R. Fort).

La résonance consiste donc à appliquer une radiofréquence ponctuelle (RF) perpendiculairement au champ  $B_0$  égale à la fréquence de précession du noyau.

La résultante vectorielle des spins occupant les deux orientations quantiques discutées au point précédent représente l'aimantation macroscopique  $\vec{M}$  alignée sur  $B_0$ . Lors de l'application de l'onde RF, on assiste à un basculement de l'aimantation  $\vec{M}$  par rapport à sa position initiale qui peut alors être scindée en deux composantes : l'une parallèle à la direction de  $B_0$  ( $M_z$ ) et l'autre dans un plan perpendiculaire à  $B_0$  ( $M_{xy}$ ) (Fig. 6). L'arrêt de cette RF va permettre le retour à l'équilibre des noyaux. L'énergie absorbée est restituée s'accompagnant du retour de l'aimantation à sa position d'équilibre en oscillant autour de l'axe z à la fréquence de précession des noyaux, c'est le phénomène de relaxation. Cette

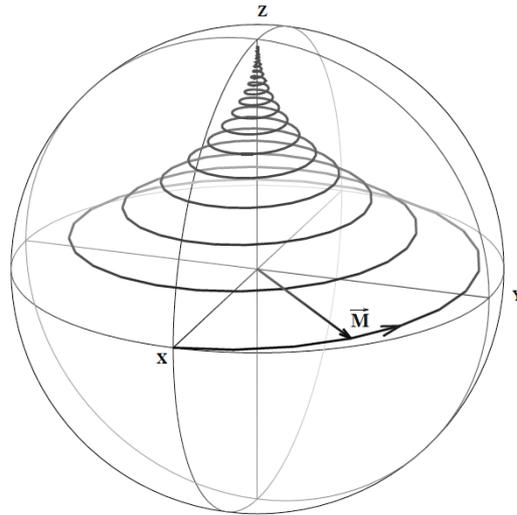
fréquence est spécifique à chaque espèce de noyaux et dépend de l'environnement électronique (densité électronique) [61,63].



**Fig. 6 :** Basculement de l'aimantation macroscopique  $M_0$  sous l'effet d'une impulsion de radiofréquence (Pr. E. Maes).

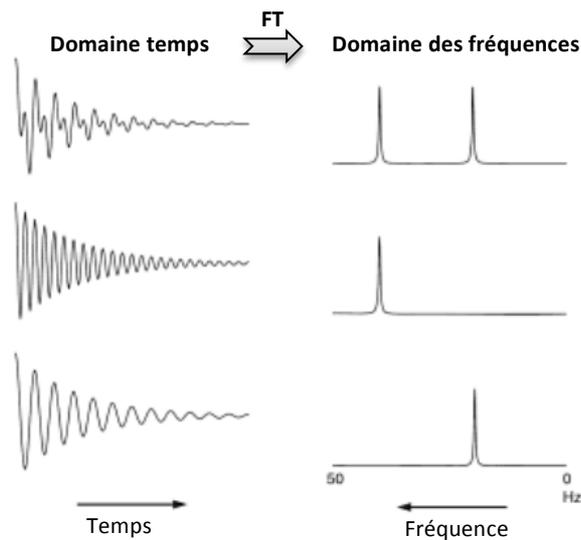
#### 1.4 Détection du signal

L'arrêt de l'application de la RF s'accompagne du retour de l'aimantation à sa position initiale le long de l'axe z avec une fréquence de précession spécifique et pendant un temps T (Fig. 7). Deux mécanismes sont responsables du retour de l'aimantation et sont caractérisés par deux constantes de temps :  $T_1$ , le temps de relaxation longitudinale et  $T_2$ , le temps de relaxation transverse. La relaxation longitudinale (appelée également relaxation spin-réseau) correspond au retour à l'équilibre de la composante  $M_z$  de l'aimantation macroscopique au cours du temps. Ce retour à l'équilibre des spins nucléaires n'est possible que si ceux-ci peuvent échanger leur énergie avec leur environnement extérieur. La relaxation transversale (ou relaxation spin-spin) provient du déphasage des composantes microscopiques de l'aimantation  $M_{xy}$  provoqué par l'inhomogénéité du champ et les interactions spin-spin [64].



**Fig. 7 :** Retour à l'équilibre de l'aimantation macroscopique (Pr. J-M. Nuzillard).

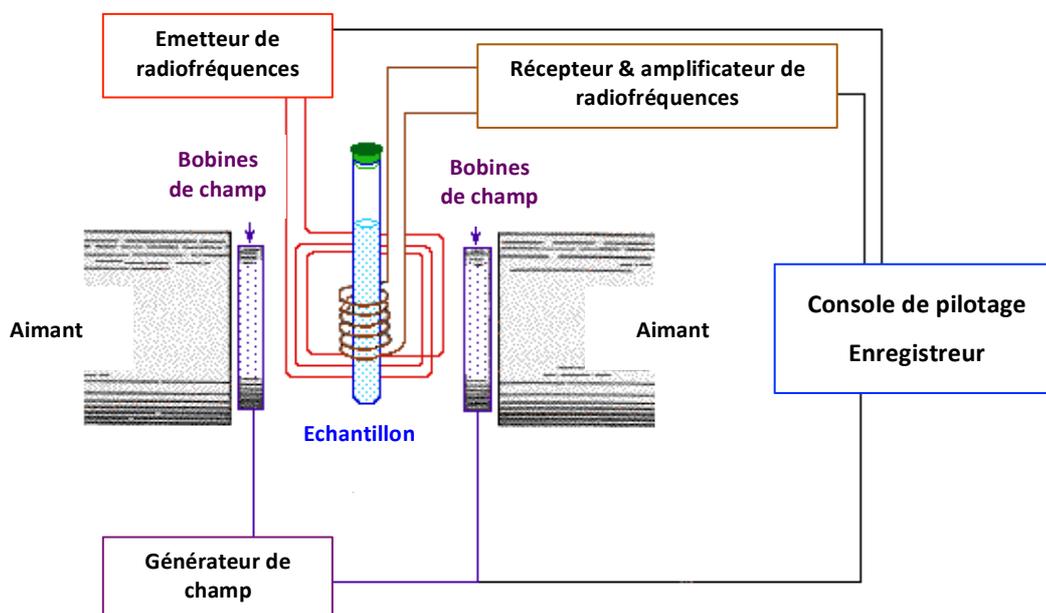
L'oscillation des noyaux lors du retour à l'aimantation provoque un champ électromagnétique qui est alors enregistré par un récepteur. Le signal détecté est appelé signal d'induction libre, ou « *Free induction decay* » (FID). Chaque sinusoïde amortie constituant le FID correspond à un type de noyau. Une transformée de Fourier (FT) est ensuite appliquée afin de passer du domaine temps (FID) au domaine des fréquences (spectre) (Fig.8).



**Fig. 8 :** Application de la transformée de Fourier permettant de transposer le FID en signal de fréquence [65].

### 1.5 Instrumentation

Le spectromètre est essentiellement constitué de l'aimant composé de matériaux supraconducteurs (niobium titane), d'un émetteur et d'un récepteur haute fréquence (Fig. 9). L'échantillon à analyser est introduit dans la cellule de mesure appelée sonde, elle-même placée dans le champ magnétique  $B_0$ . La sonde excite les noyaux grâce à un rayonnement de haute fréquence et réceptionne également le signal des noyaux relaxés. Le spectre est alors enregistré après amplification.



**Fig. 9 :** Représentation schématique de l'appareillage RMN (Pr. W. Reusch).

### 1.6 Interprétation des spectres

L'interprétation d'un spectre RMN se base sur trois éléments :

- Le déplacement chimique ( $\delta$ ) des noyaux (exprimé en ppm). Il provient d'un effet d'écrantage du champ. Cette valeur permet de définir la position du signal sur l'axe des fréquences. Il est mesuré par rapport à un signal de référence dépendant de la fréquence du spectromètre par la relation suivante [61] :

$$\delta = \frac{\nu - \nu_{\text{ref}}}{\nu_{\text{ref}}} \times 10^6$$

où  $\nu$  correspond à la fréquence de résonance du signal observé et  $\nu_{\text{ref}}$  à la fréquence de résonance du signal référent.

- La multiplicité du signal résultant des couplages spin-spin. Les couplages spin-spin ou couplages scalaires sont issus des interactions entre noyaux. Ces couplages provoquent un éclatement des raies RMN d'intensités relatives et d'espacements caractéristiques déterminés par une constante  $J$  spécifique.

- L'aire de chaque pic. Les protons chimiquement et magnétiquement identiques ont la même fréquence de précession et apparaissent au même déplacement chimique. L'aire du signal considéré est alors directement proportionnelle au nombre de protons impliqués dans ce signal. L'intégration du signal par rapport à un signal de référence permet par conséquent de déterminer le nombre de noyaux qui résonnent à une fréquence donnée.

### ***1.7 Optimisation des paramètres d'acquisition***

Pour pouvoir analyser les différents spectres, il faut que les conditions expérimentales soient comparables et par conséquent qu'elles soient optimisées pour l'analyse de l'ensemble des échantillons. Le réglage de ces paramètres va déterminer la nature et la qualité du spectre. Les principaux paramètres à prendre en compte lors de l'acquisition d'un spectre sont [61,63,66,67] :

- la fenêtre spectrale ( $SW$ , *spectral width*). Elle définit la gamme de fréquences observées.

- le temps d'acquisition du FID ( $AQ$ ). Il doit être suffisamment long afin de maximiser la quantité de signal qui est contenue dans le FID et garantir une bonne résolution digitale.

- le nombre de points acquis sur le FID ( $TD$ , *time domain data size*) autrement dit, la taille de mémoire utilisée.

Le temps d'acquisition ( $AQ$ ), la fenêtre spectrale ( $SW$ ) et le nombre de points ( $TD$ ) sont reliés par la relation suivante :

$$AQ = \frac{TD}{2 SW}$$

En général, plus le nombre de points définissant le FID est grand, meilleure est la résolution spectrale.

- le délai de relaxation ( $d_1$ ) est introduit afin de permettre le retour à l'équilibre du vecteur aimantation avant chaque nouvelle séquence d'impulsions. Le FID décroissant exponentiellement avec les temps de relaxation, il est préconisé d'attendre 5 fois la valeur du  $T_1$  le plus élevé afin d'obtenir le retour complet du vecteur aimantation après une impulsion de  $90^\circ$ .

- le nombre d'accumulations ( $NS$ , *number of scans*). Afin d'améliorer la sensibilité, une série de  $n$  impulsions peut être appliquée immédiatement après l'enregistrement du premier signal permettant l'enregistrement de  $n$  signaux qu'on accumulera avant d'obtenir le spectre. Le rapport signal sur bruit ( $S/B$ ) étant proportionnel à  $\sqrt{n}$ , plus le nombre d'accumulations sera grand, meilleur sera le rapport signal sur bruit.

- le gain ( $RG$  : *receiver gain*). Le gain de réception contrôle l'amplitude du FID qui dépend lui-même de la concentration de l'échantillon. Le gain optimum déterminé correspond à la valeur maximale obtenue pour l'intensité du signal le plus fort. Afin de pouvoir comparer les spectres, il est recommandé d'avoir la même valeur de gain et de fixer cette valeur un peu en dessous de celle déterminée de manière optimale pour éviter de saturer l'un des spectres, surtout si les échantillons à analyser diffèrent entre eux.

- La durée de l'impulsion de radiofréquence envoyée est de l'ordre de quelques  $\mu s$ .  $P_1$  définit la durée requise pour basculer entièrement le vecteur aimantation de l'axe  $Z$  dans le plan  $XY$  pour une impulsion à  $90^\circ$ . Cette durée est dépendante de l'intensité de l'irradiation.

- La puissance ( $PW$ ).  $PL$  définit le paramètre d'atténuation sur cette intensité d'irradiation.

- la fréquence d'irradiation,  $SFO_1$ , correspond à la fréquence envoyée pour exciter le noyau désiré. Elle est composée de deux termes :  $BF_1$ , qui correspond à la fréquence de base enregistrée pour le noyau choisi (500 MHz pour  $^1H$  correspondant à un  $B_0$  de 11.7 tesla) et de  $O_1$  (« offset value »), qui permet d'ajuster la valeur exacte de la fréquence de référence. Le réglage de  $O_1$  permet de se centrer sur la région spectrale d'acquisition.

## II La technique « High Resolution Magic Angle Spinning » (HRMAS)

Les analyses métabolomiques peuvent également être réalisées sur des échantillons inhomogènes comme des extraits tissulaires végétaux, animaux ou humains ou encore des cellules [68]. L'analyse de tels échantillons peut être réalisée grâce à l'utilisation d'une technique particulière utilisée en RMN appelée sonde « *High Resolution Magic Angle Spinning* », HRMAS. Cette technique permet d'obtenir des spectres RMN de haute résolution pour des échantillons donnant généralement des spectres mal résolus en RMN liquide. L'élargissement des raies est provoquée par les différents paramètres qui sont l'inhomogénéité du champ, la relaxation transversale, les interactions dipolaires, les interactions quadripolaires ainsi que l'anisotropie du déplacement chimique. Dans l'expression de l'interaction dipolaire et d'anisotropie de déplacement chimique, on retrouve le facteur  $(3 \cos^2 \theta - 1)$ . Pour un angle  $\theta = 54.7^\circ$ , l'interaction s'annule. Cette technique de RMN consiste, par conséquent, à faire tourner rapidement l'échantillon sur lui-même à un angle  $\theta$ , appelé angle magique, de  $54.7^\circ$  par rapport au champ magnétique  $B_0$  permettant de cette façon de réduire l'effet des inhomogénéités de susceptibilité magnétique [62,69].

## III Application à la métabolomique

### ***III.1 Principales séquences utilisées en métabolomique***

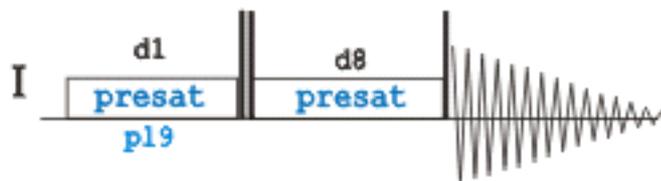
Il existe un grand nombre de séquences d'impulsions pouvant être utilisées en RMN. Elles ont pour but de faire basculer le vecteur aimantation sous différentes orientations et de suivre les différentes évolutions possibles en fonction des périodes de relaxation appliquées. Parmi le grand nombre de séquences existantes en RMN, les études métabolomiques utilisent plus particulièrement certaines séquences de RMN 1D du proton.

- Séquence « NOESY1D-PRESAT » [70,71]

Les échantillons biologiques sont des échantillons constitués de métabolites de faibles concentrations de l'ordre du mM en milieux aqueux. Les atomes d'hydrogène de ces métabolites sont minoritaires face aux hydrogènes de l'eau. La suppression du signal de l'eau doit, par conséquent être la plus efficace possible pour, non seulement récupérer le maximum de sensibilité mais surtout pour avoir une comparaison idéale des données RMN. Plusieurs techniques de suppression du signal de l'eau existent incorporées dans des

séquences RMN (*presat*, *watergate*, *jump return...*). La séquence la plus utilisée en métabolomique est la séquence « *noesy1D-presat* » [17]. Cette séquence, simple à appliquer, est la combinaison d'une pré-saturation du signal de l'eau et d'une expérience 1D NOESY. Lors de la pré-saturation, on applique une irradiation à la fréquence des protons de l'eau qui aura pour conséquence d'égaliser la population des noyaux entre les deux états énergétiques. Dans ce cas, il n'y a pas de variation d'énergie lors de l'application du champ B1 et donc pas de signal détectable pour ces protons là. La séquence 1D NOESY contribue également à la suppression du signal de l'eau par un choix judicieux du temps de mélange ( $d_8$ ), en figure 10, appliqué dans la séquence.

La séquence *noesy1D-presat* consiste en une première pré-saturation durant la relaxation ( $d_1$ ) suivie de deux impulsions de  $90^\circ$  successives. Le vecteur aimantation se retrouve alors renversé suivant l'axe  $-Z$ . Une deuxième pré-saturation suivant l'axe  $-Z$  est ensuite appliquée pendant le temps de mélange suivie d'une dernière impulsion à  $90^\circ$  avant l'enregistrement du FID. Il est important de noter que d'autres paramètres externes, tels que la température, le pH et la salinité doivent être pris en considération pour obtenir une bonne suppression du signal résiduel.

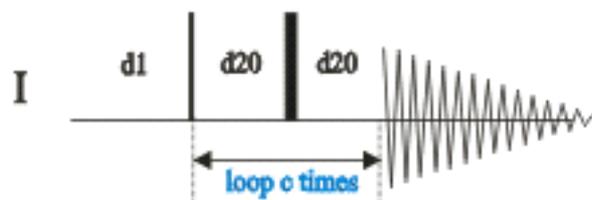


**Fig. 10** : Séquence *noesypr1d* telle que décrite par le constructeur Bruker.

- Séquence « CPMG » [72,73]

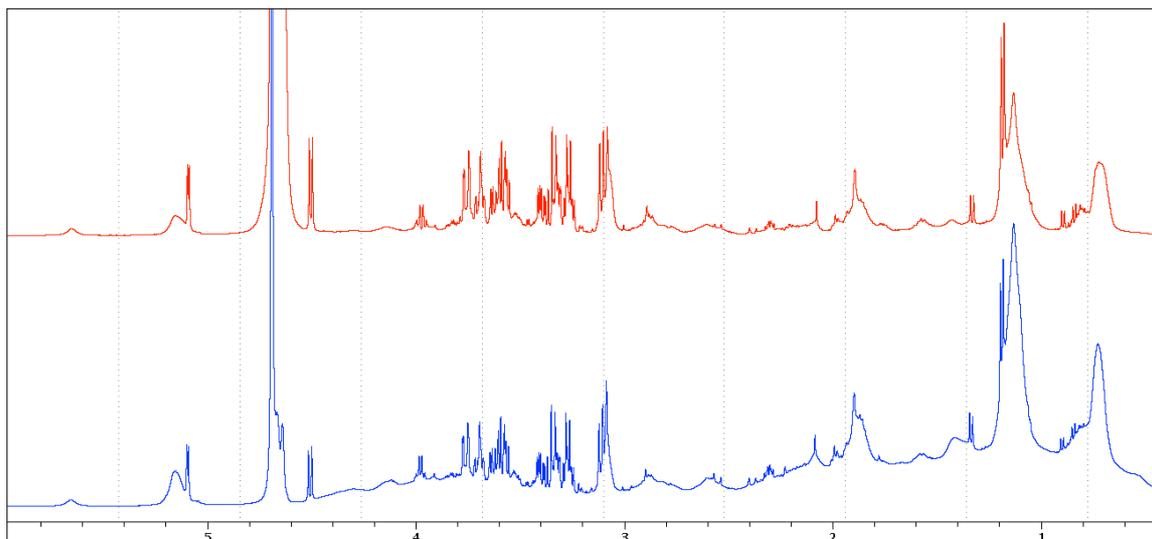
Un autre problème qui peut apparaître lors de l'acquisition des spectres est un élargissement des signaux. Certains liquides biologiques, tels que le plasma ou le sérum contiennent des composés d'une masse moléculaire de plusieurs dizaines de kDa, c'est par exemple le cas pour les protéines. Les temps de relaxation  $T_2$  étant fortement différents de ceux des métabolites inférieurs à 1000 Da, on assiste à une dispersion de l'énergie (déphasage) provoquant une perte de signal et un élargissement des signaux (le temps de relaxation des protéines étant bien inférieur à celui des métabolites). Les signaux des petits

métabolites sont superposés sur le spectre aux signaux larges des grosses molécules organisées (protéines). Afin de pallier à ce problème, une séquence également très utilisée en métabolomique utilise  $T_2$  pour différencier protéines et petits métabolites, c'est la séquence « *Carr-Purcell-Meiboom-Gill* » (CPMG). Cette séquence est basée sur l'utilisation de la technique de l'écho de spin permettant de refocaliser des signaux qui précessent à différentes fréquences et de s'affranchir de l'inhomogénéité de  $B_0$  (Fig. 11). Lorsque l'on répète  $n$  fois la séquence d'écho de spin, on obtient un train d'échos dont l'intensité décroît en fonction de  $T_2$ . Il devient alors possible d'éliminer le signal des molécules possédant un plus petit  $T_2$ . Couplée à une présaturation de l'eau, la séquence CPMG permet donc de réduire très fortement les signaux liés aux protéines mais pas sans épargner d'autres signaux qui présenteraient des  $T_2$  intermédiaires comme ceux des lipides.



**Fig. 11** : Séquence CPMG telle que décrite par le constructeur Bruker.

La figure 12 compare les spectres RMN de sérum obtenus avec les deux séquences noesygppr1D (en bleu) et CPMG (en rouge).

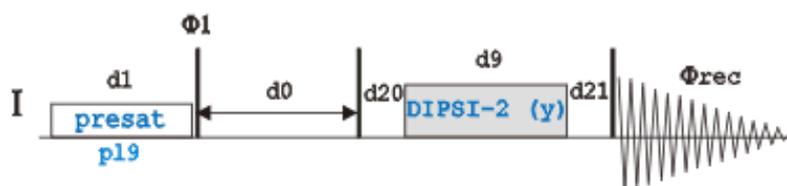


**Fig. 12 :** Spectres RMN de sérum obtenus, en rouge, avec application de la séquence CPMG et en bleu, avec application de la séquence nosygp1D.

### III.2 RMN multidimensionnelle [63,65]

Des expériences de RMN multidimensionnelles peuvent s'avérer très utiles pour l'amélioration de la résolution et pour l'identification des métabolites. Elles comprennent typiquement des expériences 2D de corrélation  $^1\text{H}$ - $^1\text{H}$  ou  $^1\text{H}$ - $^{13}\text{C}$ . Une des expériences 2D homonucléaires les plus utilisées que nous avons appliquée au sein du laboratoire est la séquence TOCSY (*Total Correlation Spectroscopy*). Elle permet d'indiquer la corrélation de tous les protons d'un même système de spins de manière directe ou indirecte par un couplage scalaire J à travers les liaisons chimiques. La TOCSY est caractérisée par un temps de mélange pendant lequel l'aimantation d'un spin A, soumise à un champ de radiofréquence est transférée à un spin relié à A par un couplage scalaire.

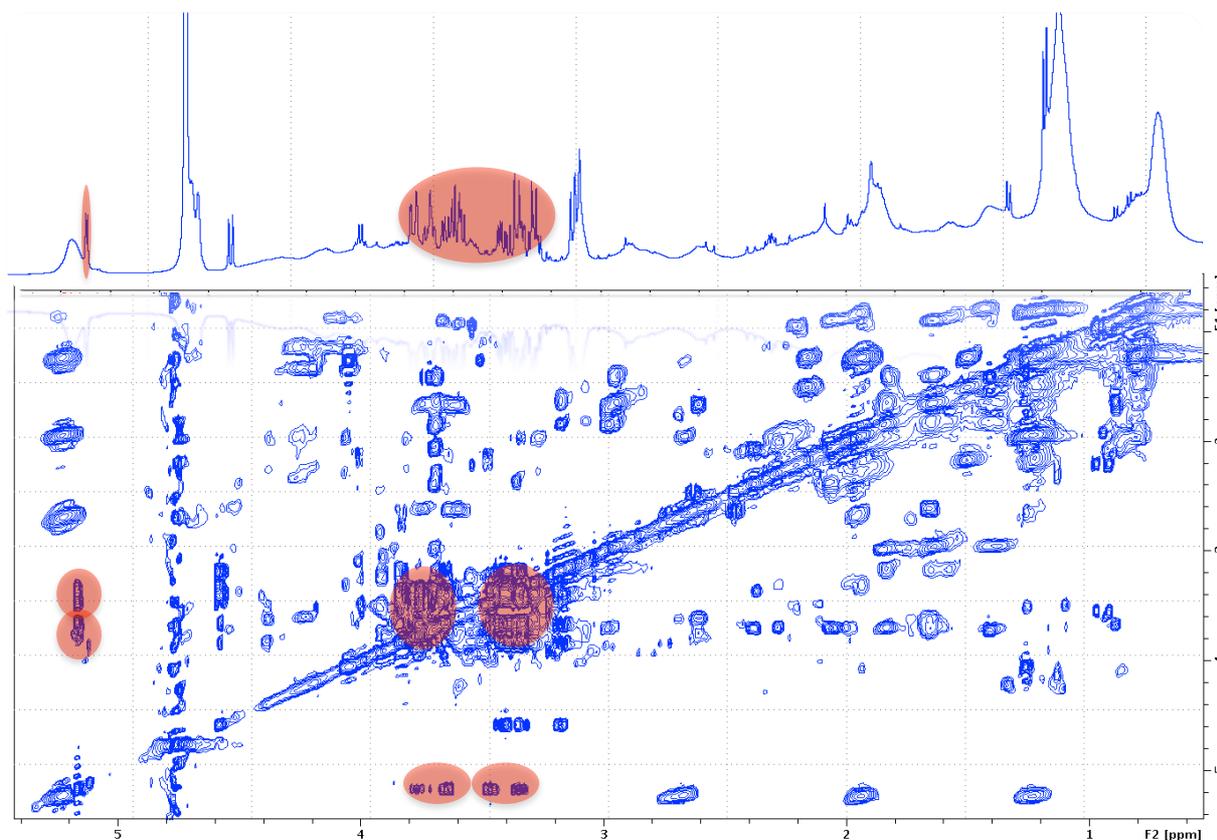
La séquence d'impulsion utilisée au laboratoire est schématisée en figure 13. Elle consiste en une séquence TOCSY reprenant un module dit de « *spin-lock* », c'est-à-dire un train d'impulsions appliqué lors de la période de mélange, qui est, dans ce cas, un module « dipsi ». Durant cette période, un transfert de polarisation s'opère vers les protons appartenant au même système de spins.



**Fig. 13 :** Séquence *dipsi2phpr* telle que décrite par le constructeur Bruker.

La recherche de noyaux couplés peut s'avérer être une étape très utile pour l'interprétation des spectres de molécules et de mélanges complexes. L'existence d'un couplage scalaire fournit une information sur la connectivité des atomes d'une molécule, c'est-à-dire sur sa formule développée.

La figure 14 représente un spectre TOCSY repris entre 0 et 5.5 ppm d'un échantillon de sérum sur lequel est illustré un exemple de connectivité entre les signaux des protons correspondant à l' $\alpha$ -glucose.



**Fig. 14 :** 2D-TOCSY (*dipsi2phpr*) comprise entre 0 et 5.5 ppm d'un échantillon de sérum illustrant la connectivité des protons correspondant à l' $\alpha$ -glucose.

### III.3 Méthodes de traitement des données RMN

Le traitement des données RMN est une étape intermédiaire entre l'obtention des données spectrales brutes et l'analyse des données [74].

Suite à l'imperfection de l'acquisition (acquisition de bruit), au traitement du signal ainsi qu'au caractère intrinsèque des échantillons biologiques (tel qu'un effet de dilution entre les échantillons), on est très souvent amené à appliquer quelques corrections aux spectres [75]. Différentes méthodes de traitement peuvent être utilisées et chacune d'entre elles possède ses avantages et ses inconvénients. Le choix d'une méthode dépend de la question biologique à laquelle on souhaite répondre, des propriétés du jeu de données ainsi que des méthodes d'analyses de données choisies. Les méthodes exposées ci-après ne constituent pas une liste exhaustive de toutes les méthodes existantes mais plutôt celles utilisées fréquemment au sein du laboratoire. Différents logiciels fournissent un nombre de fonctions permettant le traitement spectral et parfois même l'identification des métabolites et leur quantification (AMIX (*Bruker*), Chenomx NMR Suite ([www.chenomx.com](http://www.chenomx.com)), MestReNova [76], NMRpipe [77],...).

- Apodisation : la multiplication du signal par une fonction exponentielle

Avant d'appliquer la transformée de Fourier et afin d'améliorer le rapport signal/bruit, le signal FID peut être préalablement multiplié par une fonction exponentielle décroissante de type :  $e^{-t\pi LB}$ , dans laquelle LB est le facteur d'élargissement Lorentzien (*line-broadening*) et est exprimé en Hz. Cette fonction permet de diminuer progressivement l'importance du bruit sans introduire d'artéfacts [78].

- Phasage

L'acquisition du FID ne pouvant être enregistrée dès son commencement, on assiste à un déphasage des signaux sur le spectre. Deux types de correction de phase peuvent être nécessaires. La première est une correction à l'ordre 0. Elle applique la même correction de phase à l'ensemble du spectre et vise à rendre compte de tout déphasage pouvant se produire indépendamment de la fréquence du signal. La seconde est une correction à l'ordre 1. Elle applique, cette fois, une correction de phase proportionnelle à la fréquence du signal.

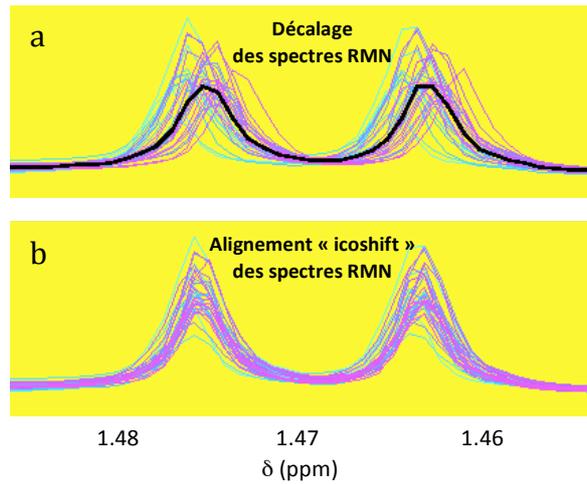
- Correction de la ligne de base

Un autre artefact qui apparaît dans le spectre RMN est la distorsion de la ligne de base due aux imperfections et à la non-linéarité du processus de détection électronique. Cette distorsion de la ligne de base peut alors être corrigée par soustraction d'une fonction polynomiale.

- Calibration

La calibration des spectres peut s'effectuer grâce à l'ajout, dans l'échantillon biologique, d'un composé référent résonant à une fréquence bien déterminée. Cependant, les références telles que l'acide TrimethylSilylPropionique (TSP) ou l'acide DimethylSilylpentaneSulfonique (DSS) sont connues pour avoir une certaine affinité avec des protéines comme l'albumine sérique humaine et par conséquent provoquer une variation dans le déplacement chimique des signaux [79]. L'absence de l'ajout d'une référence peut être pour cette raison privilégiée. Dans ce cas, le signal d'un métabolite dont le déplacement chimique est connu et peu sensible aux conditions expérimentales est utilisé pour la calibration (par exemple le doublet du signal de l' $\alpha$ -glucose à 5.23 ppm) [14].

Un autre problème, apparaissant lors du traitement statistique des données, est un possible décalage des signaux entre les différents profils RMN obtenus. Plusieurs paramètres ne provoquant pas toujours le même décalage entre les spectres peuvent influencer ce décalage des pics. Il peut s'agir de modifications du champ magnétique, d'un changement de pH, de température, d'une concentration saline différente ou encore de différentes concentrations relatives d'ions spécifiques [80]. Il s'avère donc crucial de procéder à un réalignement de ces signaux pour que le même métabolite apparaisse toujours au même endroit du spectre. Ce réalignement peut être obtenu à l'aide d'un algorithme utilisé sous Matlab : Icoshift [81]. La figure 15 illustre le réalignement par la fonction « Icoshift » d'une zone spectrale correspondant au signal de l'alanine.



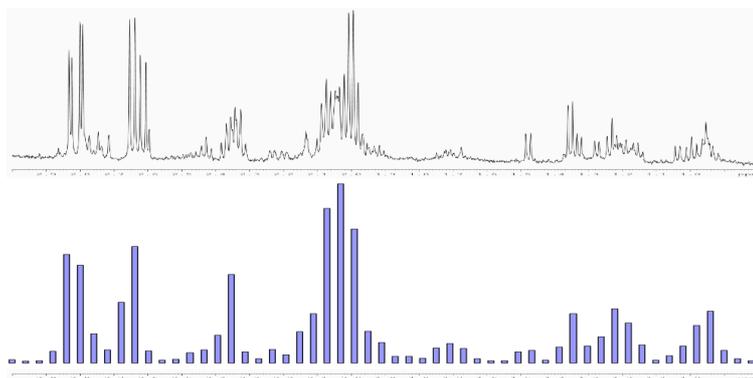
**Fig. 15 :** Effet du réalignement sur le signal du doublet de l'alanine : avant (a) et après traitement (b).

#### III.4 Méthodes de prétraitement statistique

- Découpage du spectre (« *binning* » ou « *bucketing* »)

Cette étape consiste à réduire la résolution spectrale, autrement dit le nombre de variables constituant le spectre. Chaque spectre est découpé en intervalles réguliers ou non, définis par l'utilisateur. Chaque intervalle appelé « *bucket* » correspond à une aire sous la courbe du spectre (Fig. 16) [82]. Dans le cadre des travaux de cette thèse, nous avons privilégié un découpage équidistant du spectre en régions d'une taille de  $10^{-3}$  ppm. Néanmoins il est également possible d'appliquer d'autres techniques dites « *intelligent bucketing* » permettant un découpage du spectre non équidistant [83].

Les valeurs des intégrales des buckets sont alors reportées dans une matrice  $\mathbf{X}$  à  $n$  lignes et  $m$  colonnes où les colonnes correspondent aux différents buckets (variables) et les lignes à chacun des spectres enregistrés (observations).



**Fig. 16 :** Découpage équidistant du spectre en régions appelées « buckets ».

- Exclusion de régions spectrales

Les régions du spectre qui ne contiennent pas d'informations sur les métabolites et qui sont susceptibles d'introduire des artéfacts pour l'analyse statistique sont de préférence exclues. Ainsi, les régions spectrales en dehors de la fenêtre de 0 à 10 ppm sont généralement retirées. Une autre partie du spectre correspondant aux résonances des signaux résiduels de solvant comme ceux de l'eau compris entre 4.6 ppm et 5 ppm est également très souvent exclue. Pour certaines études, nous sommes parfois amenés à retirer d'autres régions du spectre qui apportent une trop grande variabilité non liée à l'objectif recherché. C'est le cas, par exemple, pour les signaux correspondant à l'administration de médicaments.

- Normalisation des spectres

Le but de la normalisation des spectres ou encore des variables (voir point suivant) est de maximiser l'information à extraire des données et d'éliminer le maximum de bruit. Cette opération a lieu sur les lignes de la matrice reprenant les différentes données spectrales. La normalisation est appliquée aux données de chaque spectre (ou échantillon) et comprend différentes méthodes qui permettent une comparaison directe de ceux-ci. Cette normalisation permet de tenir compte des variations des conditions d'acquisition entre les spectres [13,84]. Elle est requise car il n'existe pas de standard interne fiable qui permettrait la mesure des concentrations absolues dans les échantillons. Cette normalisation permet également de minimiser les effets de dilution des échantillons qui apparaissent plus particulièrement lors de l'analyse d'échantillons d'urine. Parmi les différentes méthodes de

normalisation existantes, trois d'entre elles sont plus particulièrement appliquées : « *integral normalisation* » ; « *quotient normalisation* » et « *quantile normalisation* ».

La première représente la méthode standard de normalisation où l'intensité spectrale de chaque bucket est divisée par l'intensité totale de chaque spectre. Cette méthode est vulnérable lorsque un ou plusieurs signaux intenses varient considérablement entre les spectres [85].

La deuxième méthode, « quotient normalisation » consiste à diviser chaque variable (ou bucket) d'un spectre par l'intensité de cette même variable d'un spectre référent (médian). Puis de diviser le spectre total par la médiane de ces quotients. Cette méthode a l'avantage de ne plus dépendre des variations extrêmes d'intensités des métabolites entre les spectres [86].

La troisième méthode, la « quantile normalisation » est une technique qui va permettre une normalisation des différents spectres grâce à un classement des variables suivant leurs intensités par rapport au classement de celles d'un spectre référent (médian). Le classement des variables pour chacun des spectres permet alors une nouvelle attribution des valeurs de ces variables qui est établie d'après celles de la classification du spectre médian. On impose de cette manière que le premier bucket correspondant à l'intensité maximale soit égale pour chaque spectre, que le deuxième bucket correspondant à une intensité moindre soit égale pour chaque spectre et ainsi de suite de manière décroissante [84,87].

- Normalisation des variables (*scaling*)

La normalisation des variables appelée « *scaling* » est une opération qui porte, cette fois, sur les colonnes des données spectrales. Généralement, les variables associées aux métabolites les plus concentrés en moyenne ont une variance plus grande que les autres variables. De ce fait, elles vont jouer un rôle plus important que les autres dans les analyses de données multivariées s'appuyant sur la variance. Pour éviter ce biais, il s'avère essentiel de procéder à un rééquilibrage dans le poids des différentes variables. Plusieurs méthodes de normalisation des variables sont fréquemment utilisées. Néanmoins, en premier lieu, il convient de d'abord réaliser une opération qui consiste à translater le système de variables à l'origine appelée « centrage » [88]. Différentes méthodes de normalisation peuvent être appliquées grâce à la division de chaque variable par un facteur (« *scaling factor* »), qui dépend de la dispersion de la variable. Deux méthodes sont les plus couramment appliquées au

laboratoire, elles se basent toutes deux sur la mesure de la dispersion des données. La première est la méthode « *Unit Variance* » (UV scaling). Elle utilise l'écart type ( $s$ ) comme facteur de normalisation [85].

La seconde est une petite variante de l'UV scaling et se rapproche plus des mesures réelles puisqu'elle divise chaque variable par la racine carrée de l'écart-type ( $\sqrt{s}$ ), c'est la méthode dite « *Pareto scaling* ». Néanmoins cette méthode reste plus sensible aux larges variations des données [14].

<b>Méthode</b>	<b>Formule</b>
« <i>centering</i> »	$\tilde{x}_{ij} = x_{ij} - \bar{x}_i$
« <i>UV scaling</i> »	$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_i}{s_i}$
« <i>Pareto scaling</i> »	$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_i}{\sqrt{s_i}}$

**Tableau 1.** Résumé des méthodes de normalisation des variables qui sont principalement utilisées au laboratoire,  $\bar{x}_i$  correspond à la moyenne des variables ;  $s_i$  à l'écart-type des variables et  $\tilde{x}_{ij}$  représente la valeur des données après normalisation.

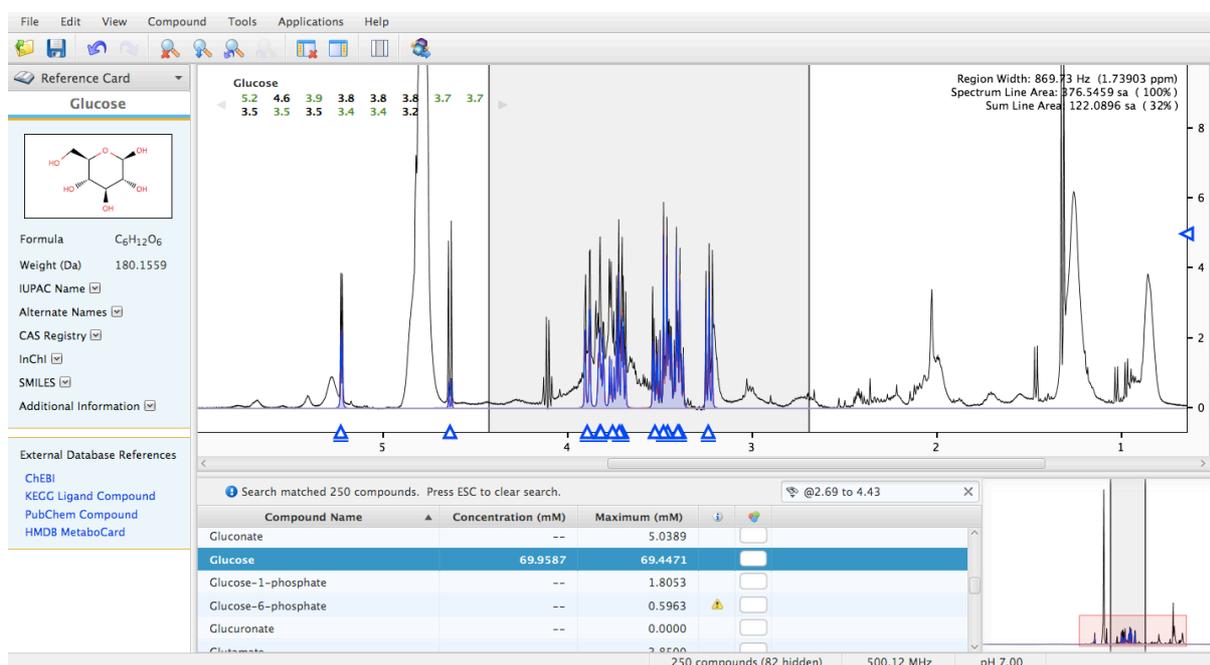
### III.5 Interprétations des spectres en métabolomique

La manipulation, l'intégration et l'annotation d'un nombre important de données telles qu'obtenues en métabolomique représentent un véritable challenge non seulement pour les chimistes, les biologistes mais également pour les biostatisticiens. Afin de pouvoir analyser et interpréter toutes ces données, la métabolomique fait appel à l'utilisation de méthodes d'analyses statistiques spécifiques que nous détaillerons dans le chapitre suivant.

L'attribution des métabolites d'un spectre RMN est une étape des plus délicates et importantes pour une interprétation biologique correcte des données métabolomiques.

Elle peut être réalisée grâce aux différents déplacements chimiques de métabolites connus dans la littérature pour un liquide biologique donné et grâce à l'apport d'informations de la

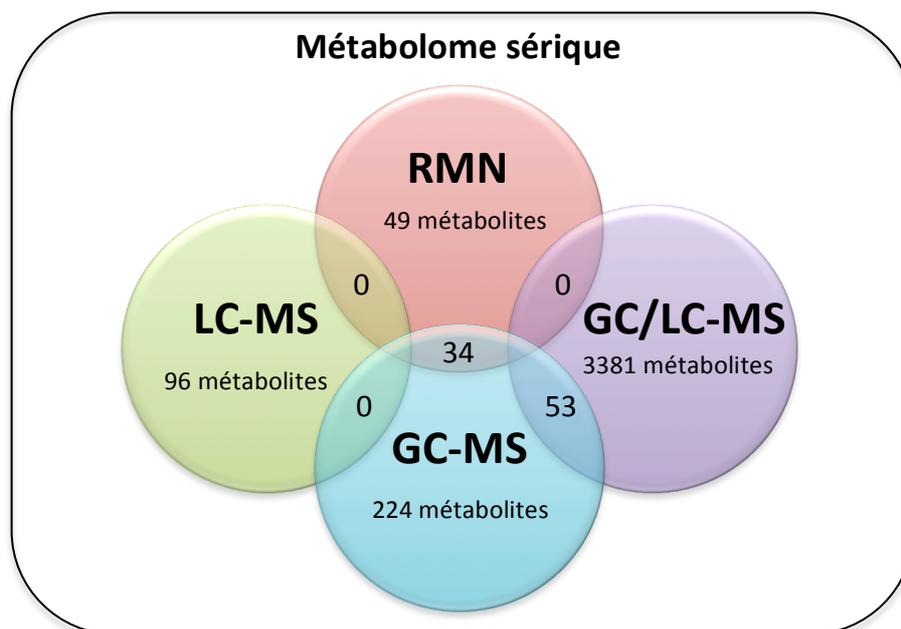
RMN 2D telle que décrite au point III.2. Néanmoins, en raison de la complexité que représente l'analyse non pas d'un composé mais d'un mélange de différents métabolites présents dans un liquide biologique, une stratégie d'attribution s'avère indispensable pour l'interprétation des spectres. Quelques bases de données de métabolites ont été développées grâce à l'utilisation de spectres référents reprenant les principaux métabolites recensés dans les liquides biologiques. Un des logiciels les plus utilisés au sein de notre laboratoire et permettant l'accès à un grand nombre de métabolites détectés en RMN est « Chenomx NMR Suite » (Fig. 17) ([www.chenomx.com](http://www.chenomx.com)). Il permet d'identifier et d'attribuer la plupart des signaux d'un spectre RMN d'un liquide biologique donné en prenant en compte les possibles fluctuations des déplacements chimiques causées par différents pH. En ligne, la « *Human Metabolomic Data Basis* » (HMDB) est une des bibliothèques les plus complètes et se place comme référence pour l'analyse du métabolome humain. Elle regroupe un ensemble de composés endogènes et exogènes atteignant les 40 000 métabolites expérimentalement détectés [89].



**Fig. 17 :** Exemple d'attribution du métabolite du glucose sur un spectre CPMG obtenu à partir de sérum à l'aide du logiciel « Chenomx NMR Suite ». En noir est représenté le spectre expérimental tandis que les signaux en bleu correspondent aux signaux du glucose pur issus de la banque de données.

### III.6 Métabolites présents dans le sérum

Depuis les grandes avancées technologiques réalisées en métabolomique et notamment grâce à la mise en place de l'*Human Metabolome Project* fondé en janvier 2005, on recense plus de 4 000 métabolites potentiellement présents dans le sang humain. Ils sont issus des différentes activités du métabolisme et la plupart d'entre eux ont été identifiés à l'aide des différentes techniques instrumentales : RMN, LC-MS, GC-MS, GC/LC-MS (figure 18) [90]. Les principales familles de métabolites endogènes détectées en RMN sont les acides gras, les acides aminés, les sucres, les intermédiaires du cycle de Krebs, les corps cétoniques ainsi que les nucléotides. L'ensemble des déplacements chimiques et multiplicités des métabolites du plasma d'individus sains identifiés en RMN  $^1\text{H}$  et  $^1\text{H}-^{13}\text{C}$ , à un pH=7.4, à une température de 292 K et à une fréquence de 750 MHz, a été repris dans la publication de J. Nicholson et son équipe [91].



**Fig. 18 :** Représentation schématique de l'analyse du métabolome sérique obtenue par les différentes techniques de RMN, LC-MS, GC-MS et par la combinaison des deux techniques GC et LC-MS. Adapté d'après [90].



## Chapitre III Traitement statistique des données RMN

Le grand nombre d'échantillons analysés mais aussi le nombre important de variables générées simultanément rendent difficile la visualisation et l'interprétation des variations métaboliques par simple comparaison des spectres. Le recours aux méthodes statistiques s'avère indispensable pour permettre une extraction des informations pertinentes et une interprétation plus aisée. Il existe de nombreuses méthodes d'analyse statistique. Néanmoins, les méthodes classiques ne peuvent être appliquées en métabolomique car le nombre de variables dépasse largement le nombre d'échantillons et de fortes corrélations entre les variables peuvent exister [92].

Les analyses métabolomiques privilégient les approches d'analyse de statistique multivariée. Elles ont l'avantage de considérer simultanément l'ensemble de tous les individus et de toutes les variables tout en permettant la réduction de la dimensionnalité des données.

Les méthodes univariées sont cependant utilisées en parallèle des méthodes multivariées. Les techniques présentées ici sont les plus utilisées au laboratoire et sont loin de constituer une liste exhaustive des techniques existantes.

### I Notions de statistique et méthodes univariées

Afin de mieux comprendre les approches et les tests statistiques appliqués lors d'une étude métabolomique, quelques notions sont rappelées ci-dessous [93].

#### ***1.1 Ecart-type et variance d'échantillons***

L'écart type d'échantillons ( $s$ ) représente la dispersion autour de la moyenne d'un ensemble de données. C'est la racine carrée de la variance. L'écart-type se calcule par l'équation suivante :

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Avec :  $n$  = nombre d'individus et  $\bar{x}$  = la moyenne de la variable  $x$ .

## 1.2 Paramètres de relation entre deux variables

- Covariance

Elle mesure la relation linéaire qui peut exister entre un couple de variables ainsi que la dispersion de ces variables. Si les deux variables évoluent dans le même sens, la covariance est de signe positif. Si elles évoluent dans le sens contraire, la covariance est négative. La covariance de variables  $x$  et  $y$  est donnée par l'expression suivante :

$$\text{Cov}(x, y) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

- Corrélation

Elle mesure l'intensité de la liaison linéaire qui peut exister entre un couple de variables. Elle est déterminée par le calcul du coefficient de corrélation linéaire,  $r$ . Ce coefficient est égal au rapport de leur covariance et du produit non nul de leurs écarts types, sa valeur est comprise entre -1 et 1. La corrélation entre deux variables  $X$  et  $Y$  est donnée par la relation suivante :

$$\text{Cor}(x, y) = \frac{\text{Cov}(x, y)}{s_X s_Y}$$

D'un point de vue géométrique, les deux séries de valeurs centrées  $(x_1^c, \dots, x_n^c)$  et  $(y_1^c, \dots, y_n^c)$ , où  $x_i^c$  et  $y_j^c$  représentent les valeurs prises par les variables centrées pour les individus  $i$  et  $j$ , peuvent être considérées comme les coordonnées de vecteurs dans un espace à  $n$  dimensions. Le cosinus de l'angle  $\alpha$  formé par ces vecteurs centrés représente le coefficient de corrélation.

Si  $r=1$ ,  $\alpha=0$ , les deux vecteurs sont colinéaires.

Si  $r=0$ ,  $\alpha=90^\circ$ , les deux vecteurs sont orthogonaux.

Si  $r=-1$ ,  $\alpha=180^\circ$ , les deux vecteurs sont colinéaires de sens opposé.

## 1.3 Test de Student

Le test de Student ou test  $t$ , consiste à comparer les valeurs prises par une variable pour deux groupes d'individus. Le calcul du test s'appuie sur une hypothèse appelée hypothèse nulle ( $H_0$ ) qui, dans le cas du test de Student postule le fait que les moyennes des

populations dont sont issus les groupes sont identiques. La différence des moyennes entre les groupes est d'abord calculée permettant ainsi de déterminer la probabilité  $\alpha$  d'obtenir une telle différence sous  $H_0$ . Si la valeur de  $\alpha$  est inférieure à une valeur seuil  $\alpha_c$  (seuil de confiance), on rejette  $H_0$ . Dans le cas de son rejet, l'hypothèse alternative ( $H_1$ ), il y a une différence significative entre les moyennes des échantillons, doit être acceptée.

Ce test est dit paramétrique parce qu'il teste des variables quantitatives suivant une loi normale.

Le seuil de confiance est généralement fixé à 0.05 dans les études biologiques ( $p\text{-value} < 0.05$ ). C'est-à-dire qu'il y a moins de 5% de chance que la différence soit observée alors qu'elle n'existe pas.

Lorsqu'il y a plus de deux ensembles de données, il devient nécessaire d'utiliser une ANOVA adaptée.

#### ***1.4 ANOVA (ANalysis Of VAriance)***

L'ANOVA est une analyse de variances. De la même façon que pour le test de Student, l'ANOVA est un test statistique vérifiant que plusieurs échantillons sont issus d'une même population. Il permet d'étudier le comportement d'une variable dépendante continue à expliquer en fonction d'un ou plusieurs facteurs catégoriels ou variables indépendantes.

## **II Méthodes multivariées**

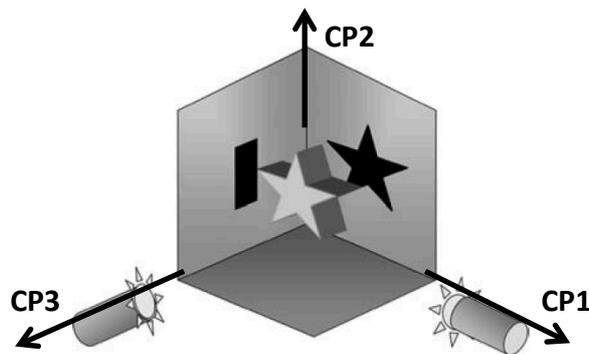
Les méthodes d'analyses multivariées permettent de synthétiser l'information par la réduction de la dimensionnalité des données afin d'en simplifier leur représentation et leur interprétation. Les résultats des analyses multivariées peuvent être représentés dans deux espaces distincts : l'espace des observations, qui établit la relation entre les individus (*scores plot*) et l'espace des variables, qui met en évidence les signatures métaboliques correspondantes, c'est-à-dire les différences caractéristiques des concentrations des métabolites détectés entre les classes d'individus (*loading plot*). Deux familles d'analyses multivariées se distinguent par leur caractère non supervisé ou supervisé.

### ***II.1 Méthodes « Non supervisées »***

La plus courante est la méthode d'analyse en composantes principales (ACP) [94,95]. Elle consiste à rechercher les directions de l'espace qui contiennent la plus grande quantité

d'informations, c'est-à-dire qui représentent le mieux la dispersion des individus. Elle permet de remplacer les variables d'origine, fortement redondantes, par des variables synthétiques (variables latentes), les composantes principales. L'ACP permet de cette manière de réduire les  $N$  variables en quelques composantes les plus représentatives. Ces composantes sont calculées par ordre d'importance et présentent l'avantage d'être orthogonales entre elles. Chaque composante principale correspond à une combinaison linéaire des variables d'origine, les poids représentant la contribution des variables à ces composantes.

Géométriquement, l'ACP peut être interprétée comme une technique de projection où la matrice  $X$  est projetée dans un sous-espace de dimensions réduites (Fig. 19).



**Fig. 19 :** Représentation imagée du principe de l'analyse en composantes principales où un objet tridimensionnel est projeté dans un espace bidimensionnel [96].

L'un des intérêts de cette méthode est de repérer, sans a priori, d'éventuels regroupements d'individus et/ou de variables. Elle permet aussi la détection d'individus au comportement atypique (« *outliers* ») par rapport au reste de la population. Ces échantillons aberrants peuvent avoir une grande influence sur la composante principale et sont, pour cette raison préférablement retirés.

Cependant, il est rare que les principales sources de variabilité observées au sein d'une cohorte d'individus soient directement reliées à l'effet étudié. L'analyse ACP ne conduit, par conséquent, généralement pas à la séparation des classes d'individus souhaitée. Les

méthodes d'analyses dites « supervisées » sont alors préférées pour identifier les variations des métabolites corrélées aux paramètres d'intérêt de l'étude.

## II.2 Méthodes « Supervisées » [97-99]

La plus courante en métabolomique est la méthode de régression des moindres carrés partiels, PLS (« *Partial Least Squares* » ou encore « *Projection to Latent Structure* »). Cette méthode a pour objectif de prédire une variable réponse  $Y$  à partir d'une matrice explicative  $X$ . Les composantes de la PLS sont calculées de manière à prendre en compte le maximum de variance des données ( $X$ ) tout en étant le plus corrélé possible à  $Y$ . Cette méthode est un compromis entre les méthodes d'ACP et de régression linéaire multiple. Son expression est donnée par l'équation suivante :

$$Y = XB + E$$

Avec  $Y$ , la matrice des variables réponses ;  $X$ , la matrice dont les colonnes représentent les variables explicatives et les lignes correspondent aux échantillons ;  $B$ , la matrice des coefficients de régression et  $E$ , la matrice des résidus.

La PLS permet l'identification des variables réponses qui sont les plus importantes dans la prédiction des variables explicatives.

La PLS-DA (« *Partial Least Squares-Discriminant Analysis* ») est une variante de la PLS pour laquelle la variable réponse est une variable discrète qui prend une valeur qui dépend uniquement de la classe catégorielle associée à l'échantillon. Le principe et l'interprétation restent inchangés par rapport à la PLS et permettent de mettre en évidence les variables discriminantes entre les groupes.

Une autre extension de l'analyse PLS, également très appliquée dans les analyses métabolomiques est l'OPLS (*Orthogonal Projection to Latent Structures*). Initialement proposée par Trygg et Wold en 2002 [100] , l'OPLS est une combinaison de la PLS avec une technique préalable de filtrage des données appelée « *Orthogonal Signal Correlation* » (OSC). Elle a pour but de retirer la variation systématique de  $X$  qui n'est pas corrélée à  $Y$  (c'est-à-dire orthogonale à  $Y$ ) rendant ainsi le modèle plus facilement interprétable.

De la même manière que pour la PLS-DA, l'OPLS-DA est une extension de l'OPLS où la variable réponse est déterminée par une classe. Ainsi avec l'OPLS-DA, une seule composante est utilisée pour prédire la classe tandis que les autres composantes décrivent la variation orthogonale par rapport à cette 1<sup>ère</sup> composante prédictive.

### III Validation

La validation est une étape essentielle en analyse statistique. En effet, la PLS ou l'OPLS appliquées aux données spectrales peuvent amener à des modèles biaisés et entraîner une surestimation des capacités de prédiction du modèle [14,101]. En effet le grand nombre de données engendre un espace ayant un très grand nombre de dimensions dans lequel il est presque toujours possible de trouver une direction de séparation parfaite entre les échantillons (qui sont moins nombreux face au nombre de variables). Par conséquent, il est indispensable de s'assurer des bonnes prédictibilité et reproductibilité des modèles établis à l'aide de différentes méthodes de validation.

D'autre part, il arrive également d'obtenir des résultats sans que le lien entre l'effet observé sur le spectre et le métabolisme n'existe réellement. Les résultats doivent par conséquent être interprétés de manière précautionneuse.

La meilleure technique pour l'évaluation de la prédiction d'un modèle reste l'utilisation d'une série indépendante de données permettant de tester le pouvoir prédictif du jeu de données originales. Néanmoins, cette série de validation est rarement disponible (nombre d'échantillons limité) obligeant le recours, par exemple, aux techniques de validation internes décrites ci-après [12].

#### **III.1 Evaluation de la performance du modèle**

Bien souvent, il n'est pas avantageux de chercher à expliquer toute la variabilité contenue dans X. En effet, le calcul d'un nombre trop important de variables latentes mène à la prise en compte de la variabilité associée au bruit. Il existe par conséquent différentes méthodes permettant de déterminer le nombre de variables latentes à inclure dans un modèle. La plus courante en métabolomique, la méthode dite de validation croisée (CV) définit deux paramètres dans l'évaluation quantitative de la performance du modèle [102] :

$R^2$  (X et Y) représente la proportion de variance expliquée de la matrice de variables X et de Y [95]. En travaillant sur les données centrées,  $R^2$  cumulé est calculé suivant l'équation :

$$R_{y,A}^2 = 1 - \frac{\sum_{i=1}^n (y_{i,A} - \widehat{y}_{i,A})^2}{\sum_{i=1}^n (y_i)^2} = 1 - \frac{SS_{RES,A}}{SS_{tot}} = \frac{\text{variance expliquée}}{\text{variance totale}}$$

où, A correspond à la dernière variable latente retenue,  $y_i$  est la valeur de y pour l'échantillon i,  $y_{i,A}$  est la valeur résiduelle de  $y_i$  une fois pris en compte les A-1 premières composantes,  $\widehat{y}_{i,A}$  est la valeur de  $y_{i,A}$  expliquée par la A<sup>ième</sup> composante.  $SS_{RES,A}$  est la somme des carrés des résidus de la prévision des observations après le calcul des A premières composantes et  $SS_{tot}$  est la somme des carrés des  $y_i$ .

$Q^2_Y$  cumulé représente la qualité de prédiction du modèle [103,104]. Elle peut être interprétée comme l'estimation de  $R^2$  pour de nouvelles données.  $Q^2_Y$  est calculée pour des données centrées par la relation suivante :

$$Q^2_{Y,A} = 1 - \frac{\sum_{i=1}^n (y_{i,A} - \widehat{y}_{i,A})^2}{\sum_{i=1}^n (y_{i,A})^2} = 1 - \frac{PRESS_A}{SS_A}$$

où, cette fois,  $\widehat{y}_{i,A}$  est la valeur de  $y_{i,A}$  prédite au cours de la validation croisée, le  $PRESS_A$  correspond à la somme des carrés des résidus de la prévision au cours de la validation croisée et  $SS_A$  est la somme des carrés des  $y_{i,A}$ .

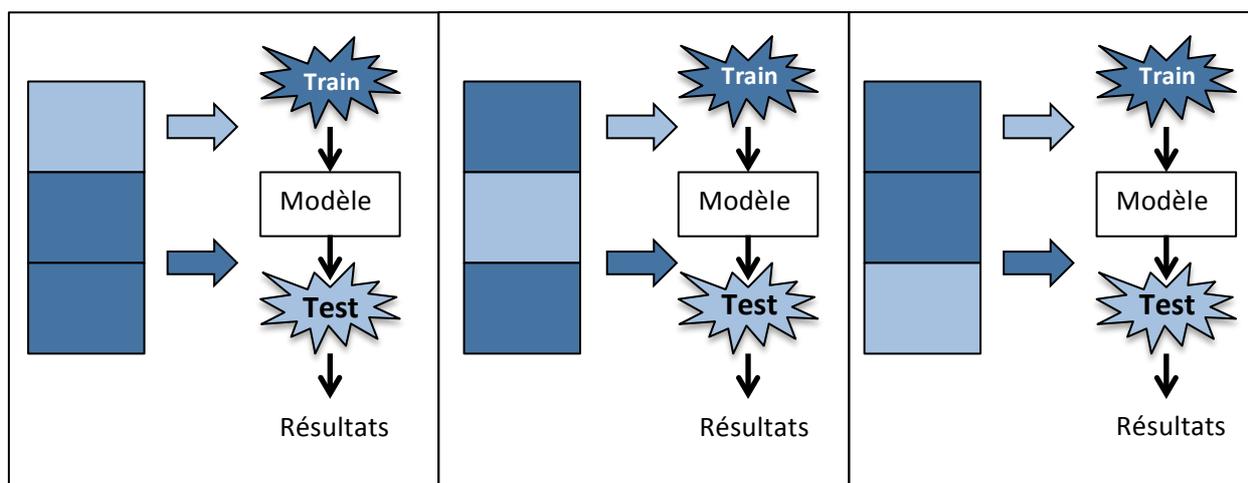
Ainsi,  $Q^2_Y$  cumulé est donné par :

$$Q^2_{Cum.Y} = 1 - \prod_1^A \frac{PRESS_A}{SS_A}$$

Au plus les valeurs de ces paramètres se rapprochent de 1, au plus le modèle est considéré comme prédictif et les résultats de séparation comme significatifs.

La méthode la plus couramment utilisée pour évaluer le nombre optimal de composantes principales ainsi que le  $Q^2_Y$  est la méthode de validation croisée, k-CV [105]. Son principe consiste à diviser le groupe d'échantillons originel en k groupes. On constitue un groupe de validation avec l'un des groupes (*test set*) puis on construit un modèle (*training set*) avec les autres groupes. L'échantillon test est alors prédit par le modèle avec estimation de l'erreur de prédiction (PRESS). L'opération se répète ainsi k-fois pour qu'au final chaque individu ait été utilisé pour la validation du modèle. Les valeurs partielles du PRESS sont sommées pour

donner le PRESS global. Si une nouvelle composante améliore le pouvoir prédictif comparé à la précédente valeur du PRESS alors cette nouvelle composante est retenue. La figure 20 illustre un exemple de validation croisée où  $k=3$ . Dans le cas de validation croisée où  $k=n$ , on parle de validation croisée « *leave-one-out* » [106]. Dans ce cas particulier de validation croisée, on réalise un modèle avec les  $(n-1)$  observations puis on le valide avec la  $n$ ème observation et on répète cette opération  $n$  fois. Ce type de validation est régulièrement appliqué lorsque peu d'échantillons sont disponibles pour l'étude.



**Fig. 20 :** Illustration du principe de validation croisée pour  $k=3$

### III.2 Autres techniques de validation

Malgré l'évaluation de la prédictibilité du modèle ( $Q^2$ ) obtenue grâce à la validation croisée, la détermination d'une valeur seuil significative du  $Q^2$  reste difficile à établir. Il arrive, de plus, que certains modèles présentant une valeur de  $Q^2$  optimiste soient non pertinents voire même biaisés. D'autres techniques complémentaires de validation doivent par conséquent être envisagées [103].

- Test des permutations

L'objectif de ce test est de confirmer que le modèle initial soit supérieur à d'autres modèles réalisés en permutant de manière aléatoire les éléments de la matrice  $Y$ . La permutation est réalisée plusieurs fois dans le but de déterminer un graphique des différentes valeurs de  $R^2$  et  $Q^2$  obtenues pour chaque permutation [107].

- Courbe CV-AUROC

L'aire sous la courbe ROC (*receiver operator characteristic*), AUROC, permet d'évaluer la qualité discriminante d'un modèle. Une courbe de ROC est un graphique reprenant le taux de vrais positifs en fonction du taux de faux positifs. L'aire sous la courbe de ROC indique la qualité discriminante d'un modèle : si la valeur est de 1, une séparation parfaite entre les deux classes est observée, si il n'y a pas de séparation entre les classes la valeur est égale à 0.5 [108].

Dans le cas du CV-AUROC, on utilise les valeurs de Y prédites par chaque sous modèle de la validation croisée pour construire la courbe de ROC.

- CV-ANOVA [109]

Ce test consiste à effectuer une analyse de la variance sur les résidus obtenus au cours de la validation croisée. Il permet de vérifier que l'erreur associée à la prédiction de la régression linéaire PLS ou OPLS est significativement plus petite que la variation de Y autour de sa moyenne.

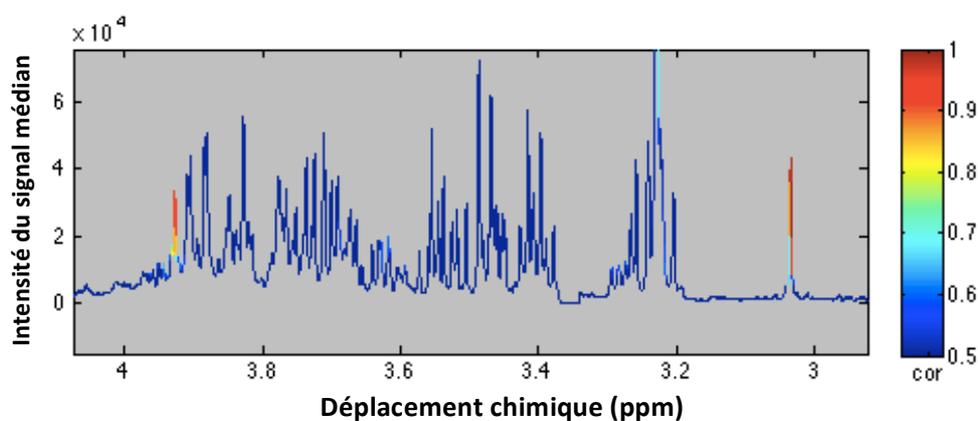
#### IV Intérêt des modèles appariés

Une trop grande variabilité entre individus peut potentiellement masquer des effets subtils reliés à une maladie ou à un traitement et par conséquent venir gêner les analyses métabolomiques. De plus, la réponse ou l'impact d'un traitement peut également être différente entre les individus.

Une analyse par appariement (*paired analysis*) peut permettre de résoudre ce problème en se débarrassant de la variabilité inter-individuelle. La variation inter-individuelle, estimée par les variations entre les moyennes individuelles peut être scindée de la variabilité intra-individuelle. Cet appariement présente l'avantage de se focaliser sur les changements métaboliques directement reliés à l'intervention et permet également une évaluation et une analyse de l'étendue de la variabilité inter-individus [110].

## V Technique complémentaire pour l'interprétation des données : la STOCSY

Introduite comme méthode complémentaire pour l'analyse des données métabolomiques, l'utilisation de la spectroscopie par corrélation totale statistique ou encore « *Statistical Total Correlation Spectroscopy* » (STOCSY) permet d'identifier les signaux de métabolites corrélés. En plus d'aider à l'identification de protons appartenant à la même molécule (exemple Fig. 21), elle présente l'avantage d'établir des connectivités de métabolites impliqués dans les mêmes voies biochimiques de manière corrélées ou anti-corrélées [17,111].



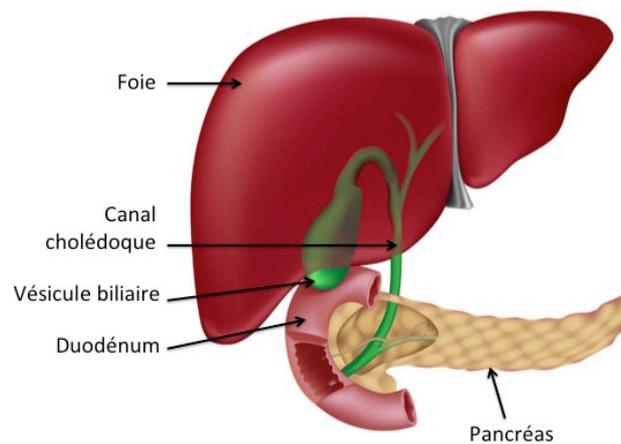
**Fig. 21** : Exemple de corrélation STOCSY des signaux de la créatine (2 singulets à 3.02 ppm et 3.92 ppm) réalisé à partir d'un spectre NOESY-1D d'une extraction de métabolites polaires de tissu de foie humain.

# Métabolomique et atteintes hépatiques

## Chapitre I Le Foie

### I Fonction du foie [112]

Le foie est un des organes les plus volumineux de l'organisme. Il est situé dans la partie droite de la cavité abdominale, sous le diaphragme. Il est divisé en deux parties par le ligament falciforme, le lobe droit et le lobe gauche. La vésicule biliaire est située au niveau du lobe droit. Le foie et la vésicule sont reliés à l'intestin par le canal cholédoque (Fig. 22). Le foie est très vascularisé lui donnant cet aspect rouge brunâtre. L'apport sanguin est réalisé par l'artère hépatique, amenant le sang oxygéné du coeur, et par la veine porte qui apporte le sang riche en nutriments provenant du tube digestif. Les cellules du foie sont majoritairement des hépatocytes mais il existe d'autres types cellulaires (endothéliales, de Küppfer, stellaires, cholangiocytes,...). L'unité fonctionnelle du foie est le lobule hépatique constitué de vaisseaux sanguins, de canaux et de cordons intercalés d'hépatocytes qui agissent comme des zones d'échanges.



**Fig. 22 :** Représentation simplifiée du foie (*anticancerfund.org*)

Le foie assure plusieurs fonctions vitales qui sont principalement des fonctions d'épuration, de synthèse et de stockage. Il joue un rôle de premier plan concernant le métabolisme des glucides et des lipides, notamment dans la production du glucose par glycogénolyse et

néoglucogenèse ainsi que son stockage sous forme de glycogène (glycogénogenèse) et dans la dégradation et synthèse du cholestérol. Il aide l'organisme à digérer les graisses en sécrétant de la bile (stockée dans la vésicule biliaire) qui se déverse dans le duodénum. Il assure le stockage d'une multitude d'autres substances dont la vitamine B12, le fer et le cuivre. Le foie est également le lieu de synthèse de certains acides aminés et de nombreuses protéines comme l'albumine, le fibrinogène ou les facteurs de coagulation ainsi que des enzymes digestives. Enfin, le foie possède des capacités antitoxiques. Il synthétise l'urée afin d'excréter les déchets azotés. Il détruit les globules rouges et les globules blancs vieillissants, ainsi que certaines bactéries présentes dans le sang. C'est également lui qui va capturer, transformer et rendre inoffensifs les toxiques auxquels nous pouvons être exposés en mangeant, buvant ou en respirant.

## II Les atteintes hépatiques [113]

### **II.1 Symptômes**

Les maladies hépatiques sont très souvent asymptomatiques pendant les débuts de leur apparition et se manifestent bien souvent à un stade déjà bien avancé de l'atteinte hépatique. Les principaux signes et complications qui peuvent apparaître et suggérer un dysfonctionnement hépatique sont [114,115] :

- L'ictère, plus couramment appelé jaunisse. Il apparaît des suites d'une augmentation de la bilirubine dans le sang.
- L'ascite. Il traduit la présence de liquide dans l'abdomen.
- Le prurit. Les dysfonctionnements hépatiques ou d'évacuation de la bile peuvent s'accompagner de démangeaisons de la peau probablement liées à la présence de sels biliaires irritant les terminaisons nerveuses.
- L'hypertension portale. Elle se définit par l'augmentation de la pression du sang dans la veine porte et peut mener aux hémorragies digestives.
- L'encéphalopathie hépatique. Elle est caractérisée par des manifestations neuropsychiatriques et des altérations de la neurotransmission. Elle est causée par la libération de substances neurotoxiques d'origine intestinale telles que l'ammoniaque et les mercaptans.

## II.2 Gradation des atteintes hépatiques

Le degré d'atteinte hépatique est très difficile à déterminer car ces maladies hépatiques peuvent rester asymptomatiques pendant plusieurs années et progresser différemment avec le temps. De plus, on distingue les atteintes dites « aiguës » des atteintes dites « chroniques » du foie. Dans le premier cas, on assiste à une destruction massive du parenchyme hépatique se déroulant sur environ deux mois et qui concerne généralement des patients ne présentant aucun antécédent de maladie hépatique. Dans les atteintes chroniques, on assiste à une évolution plus ou moins rapide vers différentes complications qui peuvent être l'apparition d'une cirrhose, d'une aggravation brutale de la fonction hépatique appelée « *acute on chronic liver failure* » (ACLF) [116,117] ou encore l'apparition du carcinome hépatocellulaire (CHC) [118]. Si les moyens sont encore limités pour prévenir et définir l'évolution des différentes atteintes hépatiques, il est possible d'établir par l'analyse du tissu hépatique différents stades de ces atteintes (Fig. 23).

- La stéatose

La stéatose est caractérisée par la présence de vacuoles graisseuses dans le cytoplasme des hépatocytes. En fonction du taux des vésicules graisseuses, quatre gradations de la stéatose peuvent être déterminées par un examen histologique [119]. La stéatose est asymptomatique et réversible.

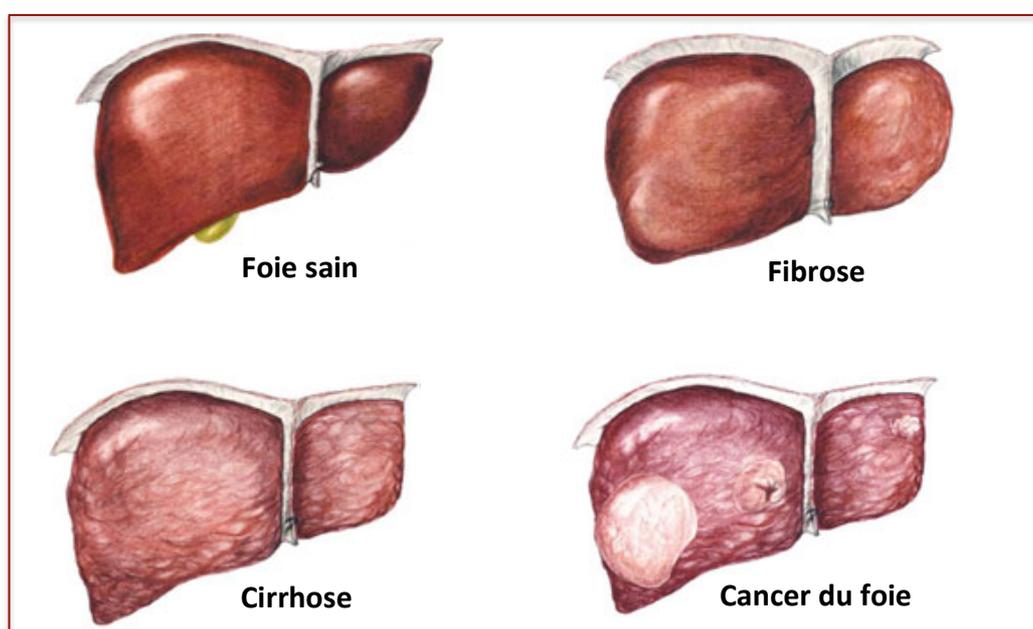
- De la fibrose à la cirrhose

L'agression chronique du foie provoque une destruction des cellules hépatiques entraînant la formation d'un tissu cicatriciel qui remplace les cellules hépatiques endommagées : c'est la fibrose hépatique. Le tissu cicatriciel entoure des amas de cellules hépatiques qui se régénèrent formant de cette façon des nodules de régénération. En fonction de l'étendue du tissu cicatriciel, on distingue différents stades de fibrose. Les stades F0 et F1 correspondent à un foie normal, le stade F2 traduit une fibrose légère et le stade F3 une fibrose sévère. On parle de cirrhose à partir du stade F4, lorsque tout le foie contient du tissu cicatriciel.

La fibrose étant dynamique et réversible, il est impératif de la quantifier précisément et précocement afin d'améliorer le diagnostic et l'évaluation d'un traitement thérapeutique.

La fibrose est donc responsable d'une désorganisation de l'architecture du foie sur le plan anatomique mais également fonctionnel pouvant mener au stade le plus avancé de fibrose

hépatique : la cirrhose [1]. La cirrhose est définie par une fibrose évolutive et diffuse, annulaire bouleversant l'architecture du lobule hépatique et vasculaire. Elle présente des nodules de régénération de petite taille (micronodules) entourés de fibrose. On distingue deux grandes étapes dans la cirrhose : la cirrhose compensée, au début totalement asymptomatique et qui peut être découverte de façon fortuite lors d'examens systématiques et la cirrhose décompensée, manifestée par les différents symptômes de complications hépatiques. La cirrhose prédispose à la survenue d'un cancer primitif du foie tel que le carcinome hépatocellulaire [120].



**Fig. 23** : Aspect macroscopique d'un foie : sain, avec fibrose, avec cirrhose et cancéreux  
(Pr. Lavanchy, OMS)

En plus des conséquences histologiques, ces différentes atteintes hépatiques mènent également à des dysfonctionnements des voies métaboliques hépatiques. L'organisation métabolique du foie a été étudiée sur des biopsies de foie sain, de foie avec fibrose et de foie cirrhotique. Les résultats ont démontré une organisation métabolique particulièrement perturbée dans le cas de la cirrhose [121]. Un grand nombre d'études ont été menées sur les différentes anomalies métaboliques détectées lors d'atteintes hépatiques en particulier la cirrhose. Ces anomalies touchent les principales fonctions métaboliques du foie notamment

manifestées dans la cirrhose par une augmentation de la gluconéogenèse [122], une augmentation du métabolisme des acides aromatiques [123], une diminution de la synthèse du HDL-cholestérol [124], ainsi que des perturbations dans le métabolisme de l'urée menant à une diminution de la détoxification de l'ammoniac [125].

### **II.3 Principales causes**

Plusieurs causes peuvent entraîner une fibrose progressive du foie et évoluer vers une cirrhose [112]. Les principales, en France, sont une consommation excessive d'alcool (>30 g par jour pendant plusieurs années) responsable pour environ 50% de l'apparition de fibrose ; la contamination des virus de l'hépatite B et C évoluant en une forme chronique ; ainsi que l'apparition, depuis ces deux dernières décennies, de stéatoses hépatiques non alcooliques (*non-alcoholic fatty liver disease : NAFLD*), dont l'obésité semble être la principale cause [126,127]. D'autres maladies, plus rares, responsables également de l'apparition de fibrose hépatique sont les maladies génétiques telles que l'hémochromatose (défaut de régulation de l'absorption intestinale du fer provoquant son accumulation dans le foie) et la maladie de Wilson (accumulation de cuivre dans le foie) ainsi que les maladies auto-immunes et les hépatopathies médicamenteuses.

### **II.4 Stéatose hépatique non alcoolique (NAFLD) et stéato-hépatite non alcoolique (NASH) [126,128,129].**

La stéatose hépatique non alcoolique (NAFLD) est une affection caractérisée par une accumulation excessive de graisses dans le foie sous forme de triglycérides (stéatose) qui survient chez un patient n'ayant pas manifesté d'autres maladies hépatiques d'origine virale, auto-immune, génétique, toxique ou encore qui soit reliée à une maladie alcoolique du foie. Ces atteintes NAFLD peuvent alors évoluer vers la stéato-hépatite non alcoolique (NASH) qui présente, en plus d'un excès de graisses, des lésions des cellules hépatiques (ballonisation, nécrose) ainsi qu'un phénomène d'inflammation. L'apparition des NAFLD semble pour la majorité des cas être associée à un syndrome métabolique lié aux effets d'obésité, d'hyperlipidémie, d'hypertension, de résistance à l'insuline ou encore de diabète de type 2. Depuis ces dernières années, alors que le surpoids et l'obésité touchent de plus en plus de personnes, la prévalence de ces atteintes hépatiques a doublé par rapport aux autres maladies hépatiques chroniques dans les pays occidentaux.

### ***II.5 Moyens diagnostiques et biomarqueurs cliniques [112,130].***

En cas de suspicion d'atteinte hépatique, l'évaluation de certains marqueurs sériques est quasiment systématique. Le syndrome d'insuffisance hépatocyttaire est défini par l'ensemble des manifestations en lien avec une diminution ou un arrêt des fonctions de l'hépatocyte. Les signes biochimiques recherchés reflètent par conséquent une diminution des fonctions de synthèse et une altération des fonctions d'épuration et de la fonction biliaire. Ces marqueurs n'étant à eux seuls pas suffisants pour poser un diagnostic, les analyses biologiques reposent principalement sur un contrôle des variations d'activités d'un ensemble de paramètres sanguins dont les principaux sont :

- les enzymes : les transaminases, l'aspartate aminotransférase (ASAT) et l'alanine aminotransférase (ALAT); les phosphatases alcalines; la 5'nucléotidase; l'alanine aminopeptidase; les gamma-glutamyltransférases (GGT)
- les plaquettes
- la bilirubine totale
- le taux de prothrombine
- les acides biliaires
- l'alphafoetoprotéine
- les différents anticorps et antigènes traduisant une infection virale ou une hépatite auto-immune
- le cholestérol total
- les phospholipides
- l'ammonium

Cependant, ces analyses étant peu sensibles et peu spécifiques pour l'évaluation de l'atteinte hépatique, elles nécessitent bien souvent d'être complétées par différents outils de diagnostic tels que l'histologie et les examens d'imagerie non invasifs comme par exemple l'échographie et l'élastographie impulsionnelle.

Les examens radiologiques tels que l'échographie abdominale permettent d'apprécier l'homogénéité du parenchyme hépatique, d'évaluer la grosseur du foie ainsi que de repérer des signes d'hypertension portale (dilatation veineuse, hypertrophie de la rate). L'élastographie impulsionnelle (Fibroscan), quant à elle, permet d'évaluer l'élasticité du foie,

marqueur de la qualité du parenchyme hépatique. Une onde élastique de basse fréquence est envoyée dans le foie. La propagation de l'onde est alors suivie par l'intermédiaire des ultrasons. Plus la vitesse de propagation de l'onde est rapide, plus le milieu est dur. La mesure de l'élasticité détermine de cette manière l'importance de la fibrose hépatique. Cette technique présente l'avantage d'être non invasive, indolore et reproductible pour le dépistage de fibrose extensive ou de cirrhose [131,132].

La biopsie du foie peut être utilisée pour confirmer le diagnostic des tests sanguins et pour déterminer l'ampleur des lésions du foie. L'analyse anatomopathologique du foie reste l'examen le plus spécifique et le plus fiable pour évaluer la nature et le degré de l'atteinte hépatique [115]. Néanmoins, par son caractère invasif, cette pratique présente quelques inconvénients. Elle peut engendrer des risques hémorragiques, infectieux et de douleurs qui ne sont pas sans danger pour le patient et ne peut par conséquent être fréquemment répétée. De plus, la biopsie requiert des praticiens expérimentés et peut être sujette à des risques d'erreur d'échantillonnage. Cette pratique s'inscrit par conséquent dans une démarche diagnostique bien précise [133].

### ***II.6 Utilisation de scores pronostiques***

Deux scores sont couramment utilisés pour prédire la survie des patients cirrhotiques : le score de Child-Pugh [134] et le score MELD (*Model for End-Stage Liver Disease*) [135]. Le premier comprend l'évaluation de cinq paramètres qui sont le taux de bilirubine, d'albumine, de prothrombine ainsi que la présence d'encéphalopathie et d'ascite. Le grade de la cirrhose est alors déterminé suivant la valeur du score et est repris sous trois classes : A, B ou C. La classe A correspond aux malades présentant généralement une cirrhose compensée alors que la classe B ou C correspond aux cirrroses décompensées. Les trois variables prises en compte pour le calcul du score MELD sont la bilirubine, l'INR (*International Normalized Ratio*) qui est un indicateur de la coagulation sanguine et la créatininémie. A l'inverse du score de Child-Pugh, le score MELD est un score continu variant de 6 à 40 points. Le MELD permet d'évaluer le degré d'insuffisance hépatocellulaire et est notamment utilisé pour une attribution équitable des greffons aux patients en attente d'une transplantation hépatique. Les tableaux des scores Child-Pugh et MELD sont donnés en annexes 2 et 3. Ces scores ne prennent pas en compte certaines complications de la cirrhose, telles que l'hémorragie digestive ou le carcinome hépatocellulaire. De plus, il est

important de noter que la normalité des tests hépatiques n'exclut pas l'existence d'une cirrhose.

### III Le carcinome hépatocellulaire (CHC)

Le carcinome hépatocellulaire (CHC), le plus fréquent des cancers primitifs du foie, apparaît dans la grande majorité des cas sur une cirrhose compensée. Il représente la cinquième cause de cancer dans le monde et la troisième cause de mortalité par cancer [136]. Le taux d'incidence augmente régulièrement depuis plusieurs décennies en France. Avec plus de 8000 nouveaux cas, il s'élevait à 12,1/100 000 chez l'homme et à 2,4/100 000 chez la femme en 2012 en France [137]. La survie globale des patients atteints de carcinome hépatocellulaire est de l'ordre de 20% à 2 ans et de 5 % à 5 ans. La survie globale des patients qui ont fait l'objet d'un traitement curatif dépasse 50% à 5 ans [118].

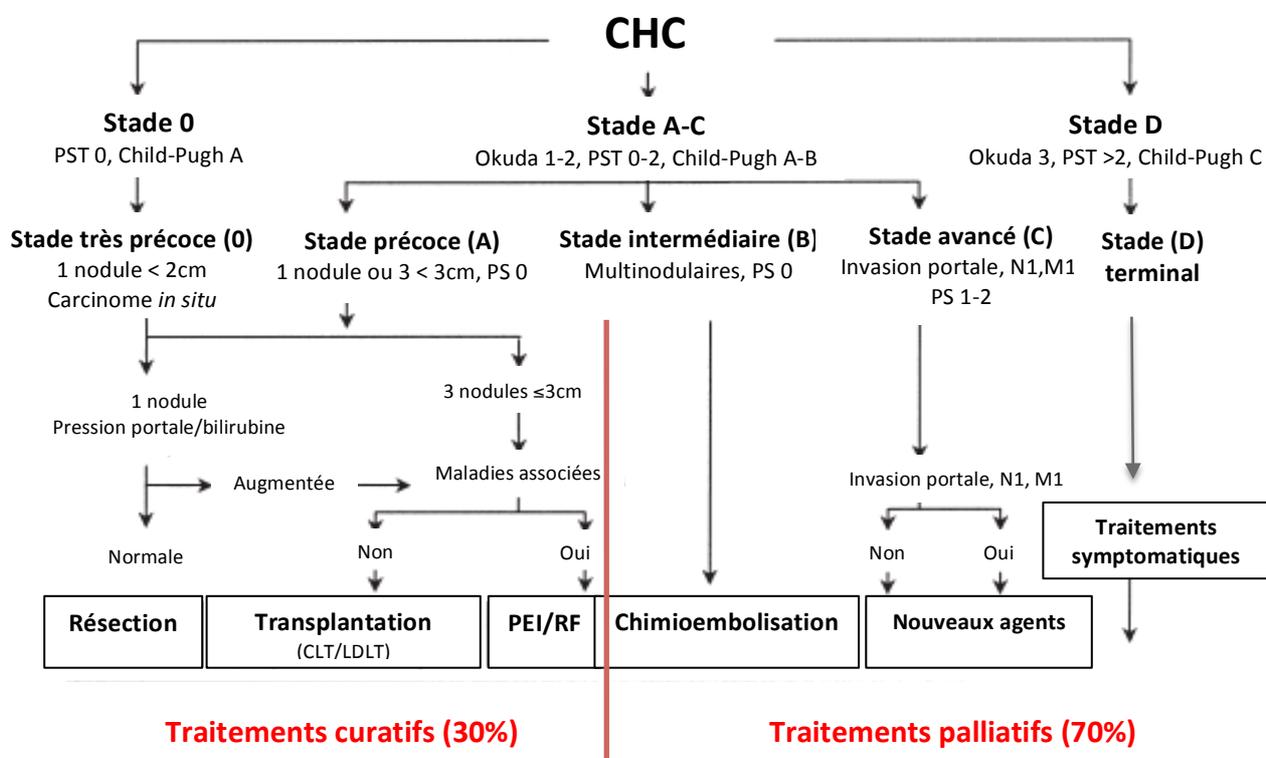
#### **III.1 Dépistage**

Le dépistage et la surveillance des patients atteints d'hépatopathie chronique permettent le diagnostic précoce d'un CHC. La détection repose essentiellement sur des examens échographiques et dosages réguliers de l'alpha-foetoprotéine (AFP) dans le sérum. Il n'existe cependant pas de marqueur biologique fiable, c'est-à-dire suffisamment sensible et spécifique pour affirmer le diagnostic du CHC [138]. D'autres techniques complémentaires telles que l'IRM et la tomodensitométrie peuvent également être entreprises afin de confirmer ce diagnostic. Les patients qui ont subi une surveillance accrue et qui ont été dépistés à un stade précoce du CHC peuvent bénéficier de traitements curatifs.

#### **III.2 Traitements [136,139-141]**

Grâce à l'apparition de nouvelles méthodes de traitement, aux améliorations du suivi des patients et à l'élaboration de recommandations, le traitement et la prise en charge du CHC se sont améliorés au cours de ces dernières années. Néanmoins, seul un diagnostic au stade initial permet un traitement potentiellement curatif. L'évaluation de la prise en charge thérapeutique est guidée selon le système de classification *Barcelona Clinic Liver Cancer* (BCLC) qui tient compte de l'état général du patient, de la fonction hépatique (Child-Pugh), du nombre de lésions, du diamètre tumoral ainsi que de la présence d'une obstruction vasculaire ou de métastases. Le tableau 2 reprend la classification BCLC. En fonction de celle-

ci, deux types de traitement peuvent être appliqués : un traitement curatif (tumeurs de petite taille) et un traitement palliatif (tumeurs évoluées) [138].

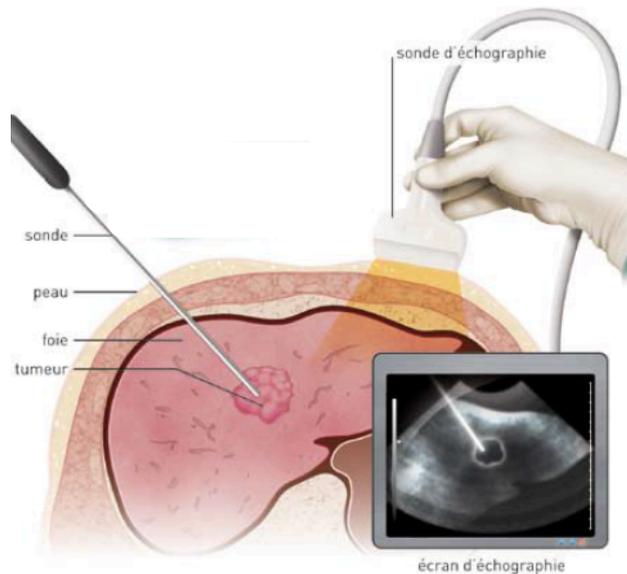


**Tableau 2.** Classification BCLC pour l'évaluation de la prise en charge du CHC [138].

- Traitements curatifs

Les traitements curatifs regroupent les techniques de résection chirurgicale (hépatectomie), de transplantation hépatique ainsi que de destruction percutanée telles que l'injection d'éthanol (PEI) et l'ablation par radiofréquence (RFA). Ils s'appliquent aux patients ayant un CHC au stade initial ou encore appelé « petit CHC » répondant aux critères de Milan (nodule unique de diamètre < 5 cm ou 2 à 3 nodules de diamètre < 3 cm) et possédant une fonction hépatique préservée. Si la résection a été pendant longtemps le traitement de première intention, les nouvelles techniques d'ablation par radiofréquence tendent à supplanter cette technique grâce à leurs applications dans des cas moins restrictifs de patients et par leur efficacité semblable à la résection. La RFA consiste en une nécrose coagulatrice utilisant un courant alternatif de haute fréquence qui est délivré par une électrode placée au centre de

la tumeur (Fig. 24). La RFA opère à des températures proches de 100°C provoquant la mort des cellules tumorales et parenchymateuses [142]. Néanmoins, le risque de récurrence reste très élevé. Plus de 50% des patients manifesteront une récurrence, locale ou à distance, dans les trois ans après la résection ou la RFA. La réduction de l'incidence de ces récurrences demeure un objectif majeur.



**Fig. 24 :** Représentation de la destruction tumorale par application de radiofréquences (Institut National du Cancer)

La transplantation hépatique est le traitement de référence pour les cancers qui se développent sur un foie atteint de cirrhose. Il permet de traiter à la fois la tumeur et la maladie chronique du foie associée. Néanmoins en raison du manque d'organes disponibles et avec les possibilités de traitements alternatifs, la transplantation n'est proposée qu'à un nombre restreint de patients.

- Traitements palliatifs

Les traitements palliatifs concernent les patients atteints d'une tumeur évoluée, hors des critères de Milan et/ou accompagnée d'une obstruction vasculaire. Ces patients peuvent profiter de traitements efficaces ralentissant la progression du CHC mais ne peuvent généralement plus être définitivement guéris. Les traitements les plus appliqués regroupent la chimio-embolisation transartérielle (TACE) et l'administration par voie orale de *sorafénib*.

La TACE concerne les patients ayant un CHC au stade intermédiaire et consiste en l'administration intra-artérielle d'un agent antitumoral, suivie d'une embolisation. Le traitement par le *sorafénib*, un inhibiteur multikinase, concerne principalement les patients présentant un CHC au stade avancé.



## **Chapitre II Etat de l'art des études métabolomiques dans les atteintes hépatiques**

Le foie est une véritable usine métabolique. La métabolomique est, par conséquent, un outil de choix pour l'exploration des fonctions métaboliques hépatiques.

Si les facteurs de risque des atteintes hépatiques sont bien connus, l'apparition de la pathologie, son évolution ainsi que les risques de complications sont encore impossibles à prédire. Le caractère asymptomatique de ces atteintes ainsi que l'absence de biomarqueurs sensibles et spécifiques rendent la classification et le diagnostic de ces hépatopathies difficiles à déterminer. De plus, malgré un dépistage précoce de l'apparition de complications telles que le carcinome hépatocellulaire permettant un traitement potentiellement curatif, une récurrence de la tumeur est observée pour plus de 50% des patients dans les 3 ans [143]. L'amélioration du diagnostic mais aussi du pronostic représente par conséquent un véritable challenge pour les cliniciens et requiert aujourd'hui le recours à des méthodes d'explorations globales du foie telle que la métabolomique.

Un grand nombre d'études métabolomiques, témoignant d'un intérêt grandissant des chercheurs et cliniciens pour cette approche, ont été entreprises dans les atteintes hépatiques que ce soit par MS ou par RMN et par l'analyse de différents liquides biologiques et tissus. Elles ont pour objectifs principaux :

- D'aider à une meilleure compréhension des perturbations métaboliques lors d'hépatopathies
- De permettre l'évaluation de la gravité de la dysfonction hépatique
- De décrire une approche diagnostique par la comparaison de profils et la recherche de biomarqueurs spécifiques de différentes hépatopathies
- De permettre la surveillance de l'évolution de l'atteinte
- D'étudier l'effet de la toxicité hépatique de xénobiotiques, par exemple médicamenteuse
- D'être prédictive d'une réponse à une thérapeutique
- D'évaluer la valeur pronostique des éventuels biomarqueurs

Nous avons choisi de faire une revue de la littérature des récentes avancées de la métabolomique dans les atteintes hépatiques en fonction des deux principales techniques

analytiques utilisées : la spectrométrie de masse et la résonance magnétique nucléaire. Comme énoncé précédemment, ces deux techniques présentent un aspect complémentaire en métabolomique notamment dans l'identification des métabolites et par conséquent des voies métaboliques.

## I Approches métabolomiques des atteintes hépatiques par spectrométrie de masse

### ***1.1 Infection par le VHB/VHC***

Si les facteurs de risque des atteintes hépatiques tels qu'une infection par le virus de l'hépatite B ou C sont bien connus, la progression et les effets de l'infection sur le métabolisme global restent encore imprécis. L'impact de l'infection du VHC sur le métabolome des hépatocytes a démontré une augmentation des métabolites impliqués dans les voies de synthèse des nucléotides dont l'ARN ainsi que des altérations du métabolisme lipidique et phospholipidique avec une augmentation notamment du niveau de la phosphatidylcholine (PC), du cholestérol et des sphingolipides. Des effets antioxydants impliquant la 5'methylthioadenosine ont également été observés lors de la réplication du VHC [144]. Une autre approche métabolomique a permis d'identifier un profil métabolique urinaire de patients atteints du VHB. Cette étude a montré une modification dans le métabolisme des xénobiotiques, de la phénylalanine ainsi que dans le métabolisme des sucres aminés et nucléotidiques [145].

### ***1.2 Insuffisance hépatique***

Une des complications des hépatopathies chroniques et de la cirrhose est l'aggravation brutale de la fonction hépatique (ACLF) [116]. Différentes études ont caractérisé une signature métabolique sérique d'atteinte chronique du foie induite par le VHB évoluant vers une insuffisance hépatique. Elles ont montré que les données métabolomiques pouvaient être capables de prédire la sévérité de l'atteinte hépatique et permettre une meilleure évaluation de la condition physiopathologique [146,147].

### ***1.3 Evaluation des atteintes « NAFLD »***

Les atteintes NAFLD regroupent un large spectre d'atteintes chroniques du foie allant de la stéatose hépatique à la stéato-hépatite (NASH), la fibrose, la cirrhose et pouvant même

engendrer l'apparition du CHC. Bien qu'une majorité des patients atteints de stéatose verra une évolution bénigne de leur maladie, une proportion significative de NASH est associée à un risque accru de cirrhose et de complications. Les critères de diagnostic du NASH par rapport à la stéatose sont par conséquent déterminants pour une prise en charge optimale du patient. C'est pour ces raisons que les études métabolomiques suscitent un intérêt de plus en plus marqué dans les atteintes NAFLD. Ces différentes études ont notamment démontré des perturbations dans l'homéostasie des lipides, des acides biliaires et de la fonction mitochondriale [148,149]. Il est également ressorti que les atteintes NAFLD s'accompagnent de la formation d'acides gras libres possédant des propriétés lipotoxiques, de la synthèse de lipides inflammatoires de la famille des eicosanoïdes et s'accompagnent également de stress oxydatif [150-152].

Il a été montré à partir d'analyses de plasma ou d'analyses de biopsies hépatiques que les acides biliaires étaient augmentés lors de la stéatose [148] et lors des stéato-hépatites [149]. Les concentrations plasmatiques totales des acides gras de longues chaînes des carnitines étaient aussi plus élevées lors de la stéatose et du NASH indiquant une altération de la  $\beta$ -oxydation mitochondriale lors d'une atteinte NAFLD [149,151]. D'autre part, une perturbation du métabolisme des phospholipides a également été observée lors des atteintes NAFLD [148,153].

Un métabolite évaluant la réponse au stress oxydatif, le dipeptide  $\gamma$ -glutamyl a souvent été identifié comme un biomarqueur sérique important des atteintes NAFLD [148,149,154].

Différentes études métabolomiques ont également indiqué que l'indice de masse corporel (IMC) constituait un paramètre important à considérer dans l'évaluation de l'atteinte NAFLD. Des différences entre trois sous groupes d'individus NASH classés suivant leur IMC ont été constatées sur le profil métabolique. Elles concernent une modification dans la concentration sérique des espèces lipotoxiques telles que les céramides et les sphingomyélines [151].

Une étude métabolomique entreprise sur du sérum en CE-MS et LC-MS a permis de déterminer un certain nombre de biomarqueurs utiles pour le diagnostic et l'évaluation de la sévérité de la fibrose. Les résultats de cette étude ont notamment montré un dysfonctionnement du métabolisme hormonal [155].

D'autre part, différentes études métabolomiques ont été entreprises sur des modèles expérimentaux d'atteintes NAFLD induites par administration de toxiques ou par le suivi

d'un régime particulier sur des animaux de laboratoire ou des cultures cellulaires. Ces études ont permis d'identifier un certain nombre de voies métaboliques altérées visant à aider à une meilleure compréhension mécanistique de cette atteinte [156,157].

#### ***1.4 Atteintes chroniques du foie menant au CHC***

La grande majorité des études des maladies du foie entreprises en métabolomique par MS concernent l'évolution des atteintes chroniques du foie en carcinome hépatocellulaire. Elles portent sur les dérégulations métaboliques, l'étude de la pathogenèse et/ou sur la mise en évidence de différents métabolites potentiellement candidats pour de nouveaux biomarqueurs de diagnostic ou pronostic du CHC. Pour la plupart, les profils métaboliques sériques ou urinaires de patients atteints du CHC sont comparés avec un groupe contrôle, ou avec un groupe de patients cirrhotiques par spectrométrie de masse de type HPLC- ou LC-MS et GC-MS. L'ensemble des études métabolomiques décrites dans ce paragraphe sont résumées dans le tableau 3.

Certaines des voies métaboliques altérées décrites précédemment lors des atteintes NAFLD se retrouvent également modifiées lors des atteintes chroniques du foie d'origine virale et/ou alcoolique et lors de l'apparition du CHC.

Le foie étant considéré comme une véritable usine du métabolisme des lipides, il n'est pas surprenant que la majorité des études rapportent des aberrations dans cette voie métabolique. Un certain nombre d'acides gras libres ont été identifiés comme métabolites sériques discriminants les patients CHC et un groupe d'individus contrôles [158-162]. Les principaux concernés sont les acides gras C16 :0, C18 :0, C20 :4, C24 :1. Il semble que l'oléamide, l'amide de l'acide oléique (C18 :1 cis-9) pourrait être un bon biomarqueur spécifique du CHC puisqu'il n'a jamais été proposé dans les autres atteintes hépatiques [163-165]. Une étude métabolomique entreprise sur le tissu hépatique tumoral a aussi déterminé une augmentation des acides gras libres saturés et monoinsaturés et une diminution des acides gras polyinsaturés par rapport au tissu non-tumoral [166].

D'autre part, une classe associée aux acides gras, la carnitine et les acylcarnitines ont également été identifiées comme voies altérées lors des atteintes hépatiques et du CHC [158,162,166-169].

Le métabolisme des phospholipides semble subir des modifications notamment dans les lysophosphatidylcholines (LPC). Plusieurs études ont montré une diminution de certaines de

ces LPC dans le sérum des patients CHC comparés à des individus sains [159,162,164,168,170,171] ou une augmentation comparés à des individus cirrhotiques [164,172]. Une altération dans ces LPC a également été observée dans le tissu hépatique tumoral [166,173].

Une des classes de composés largement citée lors des atteintes hépatiques est celle des acides biliaires. Une augmentation des acides biliaires conjugués est souvent observée lors d'une hépatopathie. La plupart des études métabolomiques ont montré une augmentation de certains acides biliaires conjugués dans le sérum de patients CHC comparés à celui d'individus sains [159,162,172]. De manière intéressante, une augmentation des acides glycochenodeoxycholique, glycocholique et taurocholique semble être plutôt associée à une signature métabolique liée à l'étiologie cirrhotique du CHC plutôt qu'à une signature du CHC [158,168,169,174].

Le métabolisme énergétique subit lui aussi des modifications liées à la prolifération des cellules tumorales, souvent décrites comme l'effet Warburg [175]. Cette hypothèse a par ailleurs été appuyée par un niveau de glycolyse plus élevé dans le tissu tumoral [166] ainsi que par une augmentation du lactate et du pyruvate dans le sérum des patients CHC [165,176].

Le glycérol et certains acides aminés ont montré une discrimination du sérum ou des urines de patients CHC et de sujets contrôles [161,165,170,177]. Une augmentation des acides aminés branchés et des acides aminés aromatiques a très souvent été observée chez les patients CHC [162,174,176].

D'autres altérations dans le métabolisme des purines et des nucléotides telles qu'une modification des concentrations de xantine, hypoxantine et acide urique sont susceptibles de se manifester lors du CHC [168,171,174,177,178]. La 1-méthyladenosine a également été rapportée comme biomarqueur sérique potentiel de l'apparition du CHC [165].

Un ensemble d'autres modifications notables dans la concentration des métabolites ont été relevées au cours de ces différentes études et sont reprises dans le tableau 3.

Associé par exemple au dosage de l' $\alpha$ -fétoprotéine, l'ensemble de ces altérations peut mener à l'établissement de nouveaux biomarqueurs permettant un meilleur diagnostic et une meilleure stratification des patients atteints du CHC [158,172].

Plateforme	Echantillon	Taille de la cohorte	Etiologie	« Biomarqueurs » identifiés	Reference
HPLC-MS/MS	Sérum	CIR : 22 CHC : 30	VHC	5-hydroxymethyl-2'-deoxyuridine, N2,N2-dimethylguanosine, Met, acide urique	Baniasadi et al, 2013 [171]
GC-MS	Tissu hépatique tumoral et non-tumoral	CHC : 31	NC	Glucose, glycérol, malate, Ala, myo-inositol, acide linoléique	Beyoglu et al, 2013 [179]
HPLC-MS	Sérum	VHC : 21 CHC : 37	VHC	Acide urique, xanthine, cholyglycine, acide leucique, l'acide 3-hydroxycaprique	Bowers et al, 2014 [178]
GC-TOF-MS UPLC-qTOF-MS	Sérum Urine	Sains : 71 Tumeur bénigne : 24 CHC : 82	VHB principalement	Inosine, His, Ala, Cys, Tyr, Phe, Thr, Met, Arg, intermédiaires du cycle de l'urée, créatinine, GCDCA, GCA, CDCA, TCA, acides gras, oléamide	Chen et al, 2011 [158]
UPLC-MS/MS	Sérum	Sains : 38 CHC : 41	VHB	1-methyladenosine	Chen et al, 2011 [165]
UFLC-MS (lipidomique)	Sérum	Sains : 30 VHB : 30 CIR : 30 CHC : 30	VHB	LPC, PC sphingomyélines, triglycerides, cholesterol	Chen et al, 2013 [170]
GC-MS UPLS-MS/MS	Sérum	Sains : 30 CIR : 27 CHC : 30	VHC	Eicosanoïdes (12-HETE et 15-HETE), sphingosine, dipeptide $\gamma$ -glutamyl, xanthine, LPC, Ser, Gly, Asp, acylcarnitines	Fitian et al, 2014 [174]
UHPLC-MS CE-MS	Tissu hépatique tumoral, adjacent et distant non-tumoral  Sérum	CHC <sub>tissu</sub> : 50  VHB <sub>serum</sub> : 81 CIR <sub>serum</sub> : 78 CHC <sub>serum</sub> : 139	VHB principalement	Intermédiaires de la glycolyse et du cycle de Krebs, Ser, Thr, Gln, BCAA, Trp, Phe, Tyr, acides gras, carnitines et acylcarnitines, LPC, PC, PE Bétaine, propionylcarnitine, Pro, Met, isoleucyl-L-proline	Huang et al, 2013 [166]

UPLC-MS	Tissu hépatique tumoral, adjacent et distant non-tumoral	CHC : 10	VHB	Phe, GPC, LPC, CDCA, betasitosterol, acide quinaldique, arachidyl carnitine, tetradecanal, oléamide	Liu et al, 2013 [173]
RPLC-MS HILIC-MS	Sérum	Sains : 26 CIR : 68 CHC : 71	VHB principalement	Acides gras, glycérol, LPC, PC, phytosphingosine, carnitine et acylcarnitines, acides biliaires, Val, Tyr, Phe	Liu et al, 2013 [162]
UPLC-qTOF-MS GC-MS	Plasma	Sains : 6 AML : 22 CIR : 7 CHC : 20	Mixte	LPC, acides gras, bilirubine, biliverdine, GCDCA	Patterson et al, 2011 [159]
GC-MS	Plasma	CIR : 49 CHC : 40	VHC	Glu, BCAA, citrate, lactate, $\alpha$ -tocopherol, cholestérol, sorbose	Ranjbar et al, 2015 [176]
UPLC-MS	Sérum	CIR : 184 CHC : 78	Mixte	Sphingosine, LPC, 3-sulfo-GCDCA, GCA, GDCA, TCA, TCDCA	Ressom et al, 2012 [172]
LC-MS	Sérum	VHB : 74 CIR : 76 CHC : 262	NC	TCA, LPC, LPE	Tan et al, 2011 [180]
UPLC-MS	Sérum	Sains : 90 CIR : 48 CHC : 82	VHB	Canavaninosuccinate, LPC, PC, GCDCA, oléamide	Wang et al, 2011 [164]
GC-MS	Urine	Sains : 20 CHC : 20	NC	Acides gras, Gly, Thr, Tyr, xylitol, urée, acides organiques, primidine, hypoxanthine	Wu et al, 2009 [177]
UPLC-qTOF-MS	Sérum	CIR : 49 CHC : 40	VHC	GCA, GDCA, LPC, PE, carnitines, peptides	Xiao et al, 2012 [169]
GC-MS	Sérum	Sains : 20 CHC : 20	NC	Acides gras, glycérol, BCAA, D-erythrose	Xue et al, 2008 [161]
RPLC-qTOF-MS HILIC-qTOF-MS	Sérum	Sains : 25 CIR : 25 CHC : 24	VHB	GCA, GCDCA, TCA, TCDCA, LPC, phytosphingosine, acides gras, acylcarnitines, bilirubine, taurine, hypoxanthine, inosine	Yin et al, 2009 [168]
CE-TOF-MS	Sérum	Sains : 61 CIR : 50 CHC : 72	VHB principalement	Trp, Gln, acide 2-hydroxybutyrique	Zeng et al, 2014 [181]

UPLC-qTOF-HDMS	Urine	Sains : 12 CHC : 25	NC	GCA	Zhang et al, 2013 [182]
UPLC-qTOF-MS	Sérum	Sains : 31 CIR : 28 CHC : 69	VHB et VHC	acides gras, LPC, anandamides	Zhou et al, 2012 [160]
RRLC-qTOF-MS	Sérum	VHB : 30 CIR : 30 CHC : 30	NC	Trp, GCA, GCDCA, LPC, acylcarnitines	Zhou et al, 2012 [167]

Abréviations : Ala= alanine ; AML= leucémie aigue myéloblastique ; Arg= arginine ; Asp= aspartate ; BCAA= acides aminés branchés ; CDCA= acide chenodeoxycholique ; CE= électrophorèse capillaire ; CHC= carcinome hépatocellulaire ; CIR= cirrhose ; Cys= cystéine ; GC= chromatographie gazeuse ; GCA= acide glycocholique ; GCDCA= acide glycochenodeoxycholique ; GDCA= acide glycodéoxycholique ; Gln= glutamine ; Glu= glutamate ; Gly= glycine ; GPC= glycérophosphocholine ; HETE= acide hydroxyicosatetraénoïque ; HDMS= spectrométrie de masse de haute résolution ; HILIC= chromatographie liquide à interaction hydrophile ; His= histidine ; HPLC= chromatographie liquide à haute performance ; LC= chromatographie liquide ; LPC= lysophosphatidylcholine ; LPE= lysophosphoéthanolamine ; Met= méthionine ; MS= spectrométrie de masse ; NC= non communiqué ; PC= phosphatidylcholine ; PE= phosphoéthanolamine ; Phe= phénylalanine ; Pro= proline ; qTOF= quadrupole temps de vol ; RPLC= chromatographie liquide en phase inversée ; RRLC= chromatographie liquide à résolution rapide ; Ser= sérine ; TCA= acide taurocholique ; TCDCA= acide taurochenodeoxycholique ; Thr= thréonine ; TOF= temps de vol ; Trp= tryptophane ; Tyr= tyrosine ; UFLC= chromatographie liquide rapide ; U(H)PLC= chromatographie liquide à ultra haute performance ; Val= valine ; VHB= virus hépatite B ; VHC= virus hépatite C.

Couleurs : Métabolite des acides biliaires, métabolite des acides gras, métabolite des nucléotides, métabolite des phospholipides, métabolite des purines

**Tableau 3.** Résumé des études métabolomiques par MS des atteintes chroniques du foie menant au CHC.

## II Approches métabolomiques des atteintes hépatiques par <sup>1</sup>H RMN

Un récapitulatif des récentes études métabolomiques d'atteintes hépatiques et leurs complications entreprises en RMN du proton a fait l'objet d'une revue acceptée dans « World Journal of Gastroenterology ». Elle retrace les principales applications des approches métabolomiques tout en intégrant la vision et les potentielles applications cliniques. Les différentes études métabolomiques sont regroupées sous trois sections qui sont les atteintes aigues du foie, les atteintes chroniques du foie ainsi que les applications en transplantation hépatique.

**Name of journal:** World Journal of Gastroenterology

**ESPS Manuscript NO:** 20058

**Columns:** MINIREVIEWS

**NMR based metabolomics and liver diseases: recent advances and future clinical applications**

Amathieu R *et al.* Metabolomics approaches in Liver Diseases

---

Roland Amathieu, Mohamed Nawfal Triba, Corentine Goossens, Nadia Bouchemal, Pierre Nahon, Philippe Savarin, Laurence Le Moyec

**Roland Amathieu**, Université Paris 13, Sorbonne Paris Cité, CSPBAT, UMR 7244, CNRS, Bobigny, France and Assistance Publique des Hôpitaux de Paris (AP-HP), Réanimation polyvalente, Hôpital Universitaire Jean Verdier, Bondy

**Mohamed Nawfal Triba, Corentine Goossens, Nadia Bouchemal, Philippe Savarin**, Université Paris 13, Sorbonne Paris Cité, CSPBAT, UMR 7244, CNRS, Bobigny

**Pierre Nahon**, Assistance Publique des Hôpitaux de Paris (AP-HP), Service d'Hépatologie, GHU PSSD, Hôpital Jean Verdier, Bondy, France and INSERM U1162, Génomique Fonctionnelle des Tumeurs solides, Université Paris 5, Paris, France

**Laurence Le Moyec**, Université d'Evry Val d'Essonne, UBIAE, INSERM U902, Evry, France

**Author contributions:** all authors solely contributed to this paper.

**Conflict-of-interest statement:** There is no conflict of interest associated with any of the senior author or other coauthors contributed their efforts in this manuscript.

**Open-Access:** This article is an open-access article which was selected by an inhouse editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

**Correspondence to:** Roland Amathieu, MD, PhD, Service d'Anesthésie et de Réanimation. Hopital Jean Verdier. Allée du 14 Juillet. 93140 Bondy. France. roland.amathieu@aphp.fr

**Telephone:** +33 1 48 02 66 80

**Fax:** +33 1 48 02 66 75

**Received:**

**Peer-review started:**

**First decision:**

**Revised:**

**Accepted:**

### **Abstract**

Metabolomics is defined as the quantitative measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification. It is an 'omics' technique that is situated downstream of genomics, transcriptomics and proteomics. Metabolomics is recognized as a promising technique in the field of systems biology for the evaluation of global metabolic changes. During the last decade, metabolomics approaches have become widely used in the study of liver diseases for the detection of early biomarkers and altered metabolic pathways. It is a powerful technique to improve our pathophysiological knowledge of various liver diseases. It can be a useful tool to help clinicians in the diagnostic process especially to distinguish malignant and non-

malignant liver disease as well as to determine the etiology or severity of the liver disease. It can also assess therapeutic response or predict drug induced liver injury. Nevertheless, the usefulness of metabolomics is often not understood by clinicians, especially the concept of metabolomics profiling or fingerprinting. In the present work, after a concise description of the different techniques and processes used in metabolomics, we will review the main research on this subject by focusing specifically on in vitro proton nuclear magnetic resonance spectroscopy based metabolomics approaches in human studies. We will first consider the clinical point of view enlighten physicians on this new approach and emphasis its future use in clinical "routine".

**Keywords:** Metabolomics, in vitro NMR spectroscopy, Liver Diseases, Cirrhosis

© **The Author(s) 2015.** Published by Baishideng Publishing Group Inc. All rights reserved.

**Core tip:** Metabolomics is a powerful technique to improve our pathophysiological knowledge of various liver diseases, to help clinicians in the diagnostic process as well as in the prognosis or therapeutic response assessment. Nevertheless, the usefulness of metabolomics is often not understood by clinicians. In the present work, after a concise description of the different techniques and processes used in metabolomics, we will review the main research on this subject by focusing specifically on  $^1\text{H}$  NMR based metabolomics in human studies. Three major themes will be enlightened: acute liver disease, chronic liver disease and liver transplantation.

## Introduction

Metabolomics is an “omics” technique that is situated downstream of genomics, transcriptomics and proteomics (Figure 1) [1]. Metabolomics is the study (and by the way the monitoring) of metabolic changes in an integrated biological system related to (patho-) physiological stimuli or genetic modification, using multiparametric analyses.

The fundamental of metabolomics is that disease (or therapeutic response) can be reflected by changes in metabolite concentrations in biological fluids or tissues. This concept is not really new. More than 500 years ago, the « urine wheel » was published by Ullrich Pinder in his book (the *Epiphanie Medicorum*, 1506). The wheel described color, taste and smell of urine to help physicians in making-diagnoses. In other words, the phenotype of the urine was used to help the clinician. Now analytical tools to measure small-molecule metabolites called biomarkers have become tools in hospital laboratories to assist physicians when diagnosing disease.

Proton nuclear magnetic resonance ( $^1\text{H}$  NMR) spectroscopy and mass spectroscopy (MS) based metabolomics approaches are the two most applied experimental methods in this area [2]. They allow identifying and quantifying metabolites within a biological fluid (serum, plasma, urine, ascites, bile...) in a single experiment. Metabolites are small molecules (less than 1 kDa) that participate in general metabolic reactions or pathways in biofluid or tissue. The term metabolome, derived from the word genome, refers to the complete set of metabolites in a biofluid, cell, tissue or organism [3]. For example, the human serum metabolome is composed of around 4,200 metabolites, half of which are phospholipids and over a thousand glycerolipids (triglycerides, diglycerides, and monoacylglycerols). Amino acids, peptides, carbohydrates, amines, and carboxylic acids are other metabolites being part of serum metabolome [4].

Data analysis in metabolomics and the process between, patients biofluid or tissue collection and the final interpretation have been well standardized [5]. The principal steps to achieve this approach are: collection of fluid or tissues, sample conservation,

sample preparation before signal acquisition, signal acquisition in NMR (or MS) platform, data processing and analysis (Figure 2).

Technically, each approach is grounded in different theoretical principles. MS spectroscopy is based on the separation of the metabolites using chemical and physical properties and NMR is based on the magnetic properties of some atoms. In MS spectroscopy, after specific preparation in function of the kind of metabolite analyzed (lipids or amino-acid or other type of metabolite), the sample is ionized and different metabolites are separated by their charges and mass.

NMR spectroscopy is often easier to understand for the physician because using exactly the same principle than magnetic resonance imaging (MRI) that is largely used daily in medicine. The principle behind NMR is that many nuclei have a nuclear spin value (for example: Hydrogen,  $^{31}$ -Phosphorus or  $^{13}$ -Carbon) and all nuclei are electrically charged. If an external magnetic field is applied, transitions are possible between the ground and excited spin states (i.e. energy transfert between the baseline to a higher energy level). The energy transfer takes place at a wavelength that corresponds to radio frequencies and when the spin returns to its base level, energy is emitted at the same frequency. The signal that matches this transfer is measured in many ways and processed in order to yield an NMR spectrum. Then instead to obtain an image like in MRI, in  $^1\text{H}$  NMR spectroscopy the signal is transformed in one or two dimensions spectra (Figure 3).

In  $^1\text{H}$  NMR spectroscopy, metabolites are identified by peaks shape, peaks multiplicity and peaks correlations. Large databases as the Human Metabolome Database (HMDB) are available to find out all the metabolite characteristics [6, 7].

Those techniques have advantages and disadvantages but are often complementary (Table 1).

Basically,  $^1\text{H}$  NMR metabolomics approach has the advantage of efficiently obtaining information on large numbers of metabolites in biofluids *in vitro* as well as in various tissues *ex vivo* (using HR-MAS: high-resolution magic angle spinning) and *in vivo* (using Medical Magnetic Resonance Imaging).  $^1\text{H}$  NMR is a highly reproducible,

rapid and non-destructive technique. Analysis can be performed with a minimal of sample preparation. Its sensitivity is far lower (microgram *vs* picogram) than mass spectrometry but sufficient to detect most of changes in biological fluid. MS needs more sample preparation and is a destructive technique. Well-conducted reviews provide larges understanding details on the different platforms especially concerning the technical aspect, physical based-mechanism and analytical method use in metabolomics [1, 3, 8].

Thus these approaches offer high-throughput analyses at relatively low cost per sample. Multiparameter datasets contain huge quantity of information concerning metabolites in the form of complex spectrum data. Many computer-based processing and statistical tools have been developed to facilitate analysis and interpretation of the data. Statistical models applied to metabolomics data can determine biomarkers or metabolomics profiles [9]. They may help identify biomarkers or the metabolic profile that characterize a disease, and/or evaluate metabolic modifications after treatment has been initiated [3]. The metabolite analysis provides information on the “metabolic phenotype” of the patient in response to various endogenous or exogenous stimuli. Then that may contribute to a better understanding of pathophysiology, to aid diagnosis, in decision-making process in choosing therapy or predicting outcome of therapeutic intervention. In the other hand, in the near future, these tools permit us to give our patients a more personalized approach in medicine [10].

Metabolomics approach showed sometime better diagnosis performance than usual biomarkers. Nevertheless, at the beginning, metabolomics approach should be considered more than complementary tool, helping clinician to diagnose disease, evaluate disease stage or follow treatment that substitution technic.

This review focuses on the recent human metabolomics data in the field of liver disease and their complications using *in vitro*  $^1\text{H}$  NMR metabolomics approaches. MS based metabolomics approaches and *in vivo* MR studies are excluded from this review.

### **Applications of $^1\text{H}$ NMR Metabolomics approaches in Liver Diseases**

The liver is a major organ with several intensive metabolic activities. The metabolic activities are distributed in different zones of the liver parenchyma. Hepatocytes metabolism varies according to its position in the liver: for example, oxidative phosphorylation, glucose output, urea synthesis, and bile acid synthesis is higher in the periportal area, whereas glucose uptake, glutamine formation, and xenobiotic metabolism are greater in the perivenous area [11]. Acute, chronic, and acute-on-chronic conditions perturb regulation of liver metabolism. From this point of view, the blood or urine metabolome should represent the final outcome of liver cellular regulation at many different levels, and for this reason represent the phenotype of a disease or a therapeutic response.

We divided the following parts into three domains: acute liver disease, chronic liver disease and liver transplantation. Major studies and biomarkers examples for each part are shown in table 2.

#### ***Acute Liver Disease***

##### ***Drug Induced Liver Injury***

Metabolomics have been used since decennia in toxicological studies principally in animal models [12]. Some recent studies have shown that metabolomics can be a powerful tool in human toxicology. Clayton *et al.* have used NMR metabolomics to study the secondary metabolism of acetaminophen in urine. The goal was to identify urine biomarkers that might be predictive of the manner in which that individual metabolizes the acetaminophen [13]. In this study, results suggested a mechanism whereby *p*-cresol and acetaminophen would compete for sulfation. Then higher *p*-cresol levels would lead to lower levels of sulfated acetaminophen and increase toxicity. More interestingly metabolome analysis have been in evidence the interplay between gut bacteria and secondary drug metabolism because *p*-cresol is a product of bacterial metabolism of tyrosine in the gut. Winnike *et al.* used a pharmacometabolomics approach to predict acetaminophen induced liver injury with a therapeutic dose (4g per day for 7 days). In human healthy volunteers and earlier

after the beginning of the treatment, they were able to identify patterns of urinary metabolites that distinguished subjects susceptible to acetaminophen-hepatotoxicity from those who were not susceptible [14].

Those examples represent a major advance in personalized medicine. A NMR based metabolomics approach could be a practical method for identifying susceptible patients shortly after starting drug treatment and with high risk of developing DILI.

#### *Liver injury and liver failure*

Acute hepatic liver failure is a very severe condition with a high risk of mortality. Predictability of survival without liver transplantation is often difficult. Using  $^1\text{H}$  NMR on serum and urine, Saxena *et al.* has shown the potential to predict an unfavorable outcome from a metabolomic profile-evaluation in patients with fulminant hepatic failure [15]. In this study, metabolomics profiles of patients with Fulminant Hepatic failure and favorable (n=20) or unfavorable (n=10) outcomes were compared. The results of this pilot study showed that one single biomarker (Glutamine) could permit to predict unfavorable outcome with a high sensitivity.

In another area of acute liver injury, Ranjan *et al.* examined serum in patients of liver injury to explore the possible clinical application of new metabolites as a biomarker to detect traumatic liver damage in blunt abdominal trauma. In this study,  $^1\text{H}$  NMR spectroscopy revealed only two amino acids, phenylalanine and tyrosine, in the sera of the subjects with liver injury, irrespective of the extent and type of injury gauged by radiology or laparotomy. The author concluded that these biomarkers detected by NMR spectroscopy could get a relevant clinical importance in establishing liver injury in patients with blunt abdominal trauma [16].

#### *AutoImmune Hepatitis*

Diagnosis of autoimmune hepatitis (AIH) is often difficult because of lack of specificity of biological and clinical signs. Moreover, diagnosis of AIH might be confounded with other acute liver conditions like DILI or primary biliary cirrhosis or overload from other diseases. Wang *et al.* assessed the utility of metabolomics in the diagnosis of the disease [17]. Plasma metabolomics profiles of patients with AIH, PBC, DILI, PBC and healthy subjects were compared. Nine biomarkers showed great

sensitivity in discriminating AIH from other disease were identified. The utility for the diagnosis of the biomarker panel was assessed and they achieved good sensitivity and specificity in distinguishing AIH from other diseases.

### ***Chronic Liver Disease***

#### *Chronic Viral Hepatitis*

Different viruses cause hepatitis infection. Hepatitis C virus is one of the most important and has infected about 150 million people worldwide [18]. Expensive biological methods to diagnosis HVC infection are problematic, especially in developing countries.

Several studies using metabolomics approaches in the field of viral hepatitis have been done. One of the most relevant is the study ran by Godoy *et al.* In a metabolomics model based on <sup>1</sup>H NMR spectra of urine they discriminated patients with HCV infection with high sensitivity (94%), specificity (97%) and accuracy (95%)<sup>[19]</sup>. These findings reveal the potential of metabolomics as a low cost, noninvasive diagnosis of HCV related hepatitis using urine samples.

#### *Assessment of Cirrhosis and its complications*

Studies have shown a close relationship between metabolic abnormalities and the severity of the disease in sera and tissues [20-23]. In those studies, a metabolomics approach was a powerful tool in assessing the severity of chronic liver failure in alcohol-induced cirrhosis within a cohort of patients without acute decompensation. For instance, in our previous study, the severity of chronic liver failure was evaluated using the MELD score, and correlated well with impairment of lipid, glucose, and amino acid metabolism [20]. Other studies show that metabolomics can fingerprint the differences between compensated and decompensated cirrhosis, and between cirrhosis caused by alcohol or viruses [24, 25].

Assessments of complication in patients with chronic liver disease, especially with cirrhosis, were performed using metabolomics approaches.

Hepatic encephalopathy (HE) is a well-known complication of liver cirrhosis. Its diagnosis is currently performed at the bedside using clinical examination. Unfortunately low grade or “sub-clinical” HE (also called minimal hepatic encephalopathy) is more difficult to diagnose and request using more complex psychometric tests that are time consuming and not realized in routine. In this case, biomarkers could be easier to use and helpful. Jimenez *et al.* have shown that using  $^1\text{H-NMR}$  to assess the metabolomics of sera from patients with cirrhosis could discriminate between patients with minimal hepatic encephalopathy and those with no encephalopathy [26].

Recently, we described a serum metabolite fingerprint for acute-on-chronic liver failure, obtained with  $^1\text{H-NMR}$  [27]. The hypothesis in this study was that cirrhotic patients with acute event (Acute-on-chronic liver failure or ACLF group) have had a specific metabolic response as compared to cirrhotic patients with stable cirrhosis (Chronic Liver Failure or CLF group). Both groups were distinguished using multivariable statistical methods and specific metabolomics fingerprint of ACLF patients in intensive care unit was identified. Several metabolites were identified and reflected major change in liver function such as energy metabolism, urea metabolism or amino-acid metabolism but also major extra-liver function change such as renal impairment or related to inflammation and/or necrosis.

#### *Liver Cancer*

Recent studies have attempted to describe the metabolic phenotype of liver cancer in heterogeneous populations of patients with hepatocellular carcinoma (HCC) using  $^1\text{H NMR}$ . This has led to the identification of various metabolically impaired pathways in serum and urine [28-31]. Goa *et al.* have shown that metabolite profiles obtained from  $^1\text{H NMR}$ -based metabolomics analysis of blood serum may be different in healthy volunteers compared to patients with liver cirrhosis or hepatocellular carcinoma [28]. Major changes to metabolites within the sera include lipids, ketone bodies, and amino-acid metabolism. Soper *et al.* have performed an *in vitro*  $^1\text{H NMR}$  spectroscopy study to characterize liver biopsy samples into normal, cirrhotic or hepatocellular carcinoma on the basis of a computer-based statistical classification strategy with changes in lipids, choline and creatine identified [32].

Ninety-eight percent of hepatocellular carcinomas in this series were distinguished from non-malignant tissue on the basis of reduced lipid and increased choline content. We used this approach to assess the metabolomic profiles of serum from alcoholic cirrhotic patients with and without hepatocellular carcinoma [29]. This study included 154 consecutive patients with compensated biopsy-proven cirrhosis. Among these, 93 had cirrhosis without HCC, 28 had biopsy-proven HCC eligible for curative treatment ('small' HCC) and 33 had HCC outside the curative treatment criteria ('large' HCC). The first step in this study was to create a diagnostic model with 'large' HCC population and validate it (figure 4 A and B). In this model, discriminant metabolites that increased with large HCC were, glutamate, acetate and N-acetyl glycoproteins, whereas metabolites that correlated with cirrhosis were lipids and glutamine (Figure 4 C). The second step was to assess the diagnostic performance of the model using the 'Small HCC' population. Unfortunately projection of small HCC samples into the first model showed a heterogeneous distribution between large HCC and cirrhotic samples. Nevertheless, small HCC patients with metabolomic profiles similar to those of large HCC group had higher incidences of recurrence or death during follow-up (63% vs 47%). Serum NMR-based metabolomics identified metabolic fingerprints that could be specific to large HCC in cirrhotic livers. From a metabolomic standpoint, some patients with small HCC, who are eligible for curative treatments, seem to behave as patients with advanced cancerous disease. This finding indicates the usefulness of the monitoring by this approach during the follow-up of those patients before and after treatment.

Biliary tract cancer is an uncommon type of liver cancer with high mortality and which is difficult to diagnosis. Wen *et al.* used metabolomics profile of bile to distinguish patients with biliary cancer from patients with benign biliary duct diseases<sup>[33]</sup>. This approach provides a good performance to discriminate cancer from benign diseases. Moreover, metabolomics approach has shown higher diagnostic performance (sensitivity of 88% and specificity of 81%) than conventional tests (CA19-9; CEA and bile cytology).

### *Liver transplantation*

Clinical course after liver transplantation is sometimes chaotic and life threatening. Unfortunately, bad outcomes are unpredictable, especially immediately after the transplant. One of the major concerns for the transplant team immediately after the surgery is to know whether the graft works. Failure of graft function remains an important cause of mortality. Some serial reports using metabolomics approaches have shown interesting results concerning the follow-up of patients. Serkova *et al.* described the blood metabolomics profile, which permits early detection of graft failure. In this case report, they were able to identify several metabolites in case of graft failure using sequential approach with multiple samples in the same patient: lactate, uric acid, glutamine, methionine increase and total fatty acids, citrate decrease. Thus, they have shown that metabolomics profiling can be an additional tool in clinical decision making [34]. Unfortunately, validation on this interesting work is not available.

To assess the quality of the liver before transplantation, Melendez *et al.* analyzed metabolomic profiles of hepatic bile in vitro  $^1\text{H}$  NMR spectroscopy. They included bile samples from donors and recipients. The main result showed a greater phosphatidylcholine signal in bile from steatotic graft compared with normal grafts [35].

**Current limitation of metabolomics and future clinical application**

In this review we have highlighted studies that have advanced our understanding of various liver diseases. The examples above demonstrate that the integration of metabolomics approaches into basic and biomedical research is already improving our understanding of biological mechanism, with important applications in the study of disease and/or therapeutic response. Technological advances have brought metabolomics to the point where these techniques can find general application in medicine.

However, there are several challenges to be overcome before metabolomics approaches can become a valuable clinical tool. It will be necessary to translate the technic in hospital lab, improve, simplify and automatized bioinformatics strategies, automatic recon of biomarker or metabolomics profile, validated the results in large prospective observational and interventional studies with meaningful clinical-end point.

Only after that and probably in a near future, results from metabolomics approach (with analysis and interpretation) will be available to the clinician in less than 1 day.

Metabolomics profiling could be useful in personalized medicine for diagnosis, prognosis and to follow patients-response before and after treatment. Metabolomics has the potential of providing new criteria to risk-stratify patients and develop novel approaches for individualized treatment.

### **Conclusion**

Metabolomics is a powerful method that can be used to quantitatively assess the differences in metabolite abundance or to determine metabolic fingerprint discriminating different disease states, the severity of disease extension, drug treatment metabolic impairment or pathophysiological mechanism investigation. The ability to perform such studies in a large range of biological samples, especially urine samples, which are easy to collect and non-invasive, makes it an attractive platform for translation to clinical use.

### **Acknowledgments**

The authors thank Ms. Virginia Eskridge (Pittsburgh, Pennsylvania, USA) for her invaluable assistance in preparing this manuscript.

**REFERENCES**

- 1 Lindon JC, Nicholson JK. Spectroscopic and statistical techniques for information recovery in metabonomics and metabolomics. *Annu Rev Anal Chem (Palo Alto Calif)* 2008; **1**: 45-69 [PMID: 20636074 DOI: 10.1146/annurev.anchem.1.031207.113026]
- 2 Fernie AR, Trethewey RN, Krotzky AJ, Willmitzer L. Metabolite profiling: from diagnostics to systems biology. *Nature reviews Molecular cell biology* 2004; **5**(9): 763-769 [PMID: 15340383 DOI: 10.1038/nrm1451]
- 3 Dunn WB, Broadhurst DI, Atherton HJ, Goodacre R, Griffin JL. Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chem Soc Rev* 2011; **40**(1): 387-426 [PMID: 20717559 DOI: 10.1039/b906712b]
- 4 Psychogios N, Hau DD, Peng J, Guo AC, Mandal R, Bouatra S, Sinelnikov I, Krishnamurthy R, Eisner R, Gautam B, Young N, Xia J, Knox C, Dong E, Huang P, Hollander Z, Pedersen TL, Smith SR, Bamforth F, Greiner R, McManus B, Newman JW, Goodfriend T, Wishart DS. The human serum metabolome. *PLoS one* 2011; **6**(2): e16957 [PMID: 21359215]
- 5 Goodacre R, Broadhurst D, Smilde AK, Kristal BS, Baker D, Beger RD, Bessant C, Connor S, Capuani G, Craig A, Ebbels T, Kell DB, Manetti C, Newton J, Paternostro G, Somorjai RL, Sjöström M, Trygg J, Wulfert F. Proposed minimum reporting standards for data analysis in metabolomics. *Metabolomics* 2007(3): 231-241
- 6 <http://www.hmdb.ca/>.
- 7 Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, Cheng D, Jewell K, Arndt D, Sawhney S, Fung C, Nikolai L, Lewis M, Coutouly MA, Forsythe I, Tang P, Shrivastava S, Jeroncic K, Stothard P, Amegbey G, Block D, Hau DD, Wagner J, Miniaci J, Clements M, Gebremedhin M, Guo N, Zhang Y, Duggan GE, Macinnis GD, Weljie AM, Dowlatabadi R, Bamforth F, Clive D, Greiner R, Li L, Marrie T, Sykes BD, Vogel HJ, Querengesser L. HMDB: the Human Metabolome Database. *Nucleic Acids Res* 2007; **35**(Database issue): D521-526 [PMID: 17202168]
- 8 Trygg J, Holmes E, Lundstedt T. Chemometrics in metabonomics. *J Proteome Res* 2007; **6**(2): 469-479 [PMID: 17269704 DOI: 10.1021/pr060594q]

- 9 Dunn WB, Bailey NJ, Johnson HE. Measuring the metabolome: current analytical technologies. *Analyst* 2005; **130**(5): 606-625 [PMID: 15852128]
- 10 Holmes E, Wilson ID, Nicholson JK. Metabolic phenotyping in health and disease. *Cell* 2008; **134**(5): 714-717 [PMID: 18775301 DOI: 10.1016/j.cell.2008.08.026]
- 11 Jungermann K, Kietzmann T. Oxygen: modulator of metabolic zonation and disease of the liver. *Hepatology* 2000; **31**(2): 255-260 [PMID: 10655244 DOI: 10.1002/hep.510310201]
- 12 Lindon JC, Keun HC, Ebbels TM, Pearce JM, Holmes E, Nicholson JK. The Consortium for Metabonomic Toxicology (COMET): aims, activities and achievements. *Pharmacogenomics* 2005; **6**(7): 691-699 [PMID: 16207146 DOI: 10.2217/14622416.6.7.691]
- 13 Clayton TA, Baker D, Lindon JC, Everett JR, Nicholson JK. Pharmacometabonomic identification of a significant host-microbiome metabolic interaction affecting human drug metabolism. *Proc Natl Acad Sci U S A* 2009; **106**(34): 14728-14733 [PMID: 19667173 PMCID: 2731842 DOI: 0904489106 [pii] 10.1073/pnas.0904489106]
- 14 Winnike JH, Li Z, Wright FA, Macdonald JM, O'Connell TM, Watkins PB. Use of pharmaco-metabonomics for early prediction of acetaminophen-induced hepatotoxicity in humans. *Clin Pharmacol Ther* 2010; **88**(1): 45-51 [PMID: 20182423 DOI: clpt2009240 [pii] 10.1038/clpt.2009.240]
- 15 Saxena V, Gupta A, Nagana Gowda GA, Saxena R, Yachha SK, Khetrpal CL. <sup>1</sup>H NMR spectroscopy for the prediction of therapeutic outcome in patients with fulminant hepatic failure. *NMR Biomed* 2006; **19**(5): 521-526 [PMID: 16598697 DOI: 10.1002/nbm.1034]
- 16 Ranjan P, Gupta A, Kumar S, Gowda GA, Ranjan A, Sonker AA, Chandra A, Ramakant. Detection of new amino acid markers of liver trauma by proton nuclear magnetic resonance spectroscopy. *Liver Int* 2006; **26**(6): 703-707 [PMID: 16842327 DOI: LIV1283 [pii] 10.1111/j.1478-3231.2006.01283.x]
- 17 Wang JB, Pu SB, Sun Y, Li ZF, Niu M, Yan XZ, Zhao YL, Wang LF, Qin XM, Ma ZJ, Zhang YM, Li BS, Luo SQ, Gong M, Sun YQ, Zou ZS, Xiao XH. Metabolomic

- Profiling of Autoimmune Hepatitis: The Diagnostic Utility of Nuclear Magnetic Resonance Spectroscopy. *J Proteome Res* 2014 [PMID: 24940827 DOI: 10.1021/pr500462f]
- 18 Lavanchy D. The global burden of hepatitis C. *Liver Int* 2009; **29 Suppl 1**: 74-81 [PMID: 19207969 DOI: 10.1111/j.1478-3231.2008.01934.x]
- 19 Godoy MM, Lopes EP, Silva RO, Hallwass F, Koury LC, Moura IM, Goncalves SM, Simas AM. Hepatitis C virus infection diagnosis using metabonomics. *J Viral Hepat* 2010; **17**(12): 854-858 [PMID: 20070502 DOI: 10.1111/j.1365-2893.2009.01252.x JVH1252 [pii]]
- 20 Amathieu R, Nahon P, Triba M, Bouchemal N, Trinchet JC, Beaugrand M, Dhonneur G, Le Moyec L. Metabolomic approach by <sup>1</sup>H NMR spectroscopy of serum for the assessment of chronic liver failure in patients with cirrhosis. *J Proteome Res* 2011; **10**(7): 3239-3245 [PMID: 21568267 DOI: 10.1021/pr200265z]
- 21 Martinez-Granados B, Monleon D, Martinez-Bisbal MC, Rodrigo JM, del Olmo J, Lluch P, Ferrandez A, Marti-Bonmati L, Celda B. Metabolite identification in human liver needle biopsies by high-resolution magic angle spinning <sup>1</sup>H NMR spectroscopy. *NMR Biomed* 2006; **19**(1): 90-100 [PMID: 16411169 DOI: 10.1002/nbm.1005]
- 22 Martinez-Granados B, Morales JM, Rodrigo JM, Del Olmo J, Serra MA, Ferrandez A, Celda B, Monleon D. Metabolic profile of chronic liver disease by NMR spectroscopy of human biopsies. *Int J Mol Med* 2011; **27**(1): 111-117 [PMID: 21072494 DOI: 10.3892/ijmm.2010.563]
- 23 Yu K, Sheng G, Sheng J, Chen Y, Xu W, Liu X, Cao H, Qu H, Cheng Y, Li L. A metabonomic investigation on the biochemical perturbation in liver failure patients caused by hepatitis B virus. *J Proteome Res* 2007; **6**(7): 2413-2419 [PMID: 17539670 DOI: 10.1021/pr060591d]
- 24 Qi S, Tu Z, Ouyang X, Wang L, Peng W, Cai A, Dai Y. Comparison of the metabolic profiling of hepatitis B virus-infected cirrhosis and alcoholic cirrhosis patients by using (1) H NMR-based metabonomics. *Hepatol Res* 2012 [PMID: 22404306 DOI: 10.1111/j.1872-034X.2011.00964.x]
- 25 Qi SW, Tu ZG, Peng WJ, Wang LX, Ou-Yang X, Cai AJ, Dai Y. (1)H NMR-based serum metabolic profiling in compensated and decompensated cirrhosis.

*World J Gastroenterol* 2012; **18**(3): 285-290 [PMID: 22294833 PMCID: 3261547 DOI: 10.3748/wjg.v18.i3.285]

26 Jimenez B, Montoliu C, MacIntyre DA, Serra MA, Wassel A, Jover M, Romero-Gomez M, Rodrigo JM, Pineda-Lucena A, Felipe V. Serum metabolic signature of minimal hepatic encephalopathy by (1)H-nuclear magnetic resonance. *J Proteome Res* 2010; **9**(10): 5180-5187 [PMID: 20690770 DOI: 10.1021/pr100486e]

27 Amathieu R, Triba MN, Nahon P, Bouchemal N, Kamoun W, Haouache H, Trinchet JC, Savarin P, Le Moyec L, Dhonneur G. Serum 1H-NMR metabolomic fingerprints of acute-on-chronic liver failure in intensive care unit patients with alcoholic cirrhosis. *PLoS one* 2014; **9**(2): e89230 [PMID: 24586615 PMCID: 3929651 DOI: 10.1371/journal.pone.0089230]

28 Gao H, Lu Q, Liu X, Cong H, Zhao L, Wang H, Lin D. Application of 1H NMR-based metabolomics in the study of metabolic profiling of human hepatocellular carcinoma and liver cirrhosis. *Cancer Sci* 2009; **100**(4): 782-785 [PMID: 19469021]

29 Nahon P, Amathieu R, Triba MN, Bouchemal N, Nault JC, Ziol M, Seror O, Dhonneur G, Trinchet JC, Beaugrand M, Le Moyec L. Identification of serum proton NMR metabolomic fingerprints associated with hepatocellular carcinoma in patients with alcoholic cirrhosis. *Clin Cancer Res* 2012; **18**(24): 6714-6722 [PMID: 23136190 DOI: 10.1158/1078-0432.CCR-12-1099]

30 Shariff MI, Gomaa AI, Cox IJ, Patel M, Williams HR, Crossey MM, Thillainayagam AV, Thomas HC, Waked I, Khan SA, Taylor-Robinson SD. Urinary metabolic biomarkers of hepatocellular carcinoma in an Egyptian population: a validation study. *J Proteome Res* 2011; **10**(4): 1828-1836 [PMID: 21275434 DOI: 10.1021/pr101096f]

31 Shariff MI, Ladep NG, Cox IJ, Williams HR, Okeke E, Malu A, Thillainayagam AV, Crossey MM, Khan SA, Thomas HC, Taylor-Robinson SD. Characterization of urinary biomarkers of hepatocellular carcinoma using magnetic resonance spectroscopy in a Nigerian population. *J Proteome Res* 2010; **9**(2): 1096-1103 [PMID: 19968328 DOI: 10.1021/pr901058t]

32 Soper R, Himmelreich U, Painter D, Somorjai RL, Lean CL, Dolenko B, Mountford CE, Russell P. Pathology of hepatocellular carcinoma and its precursors

using proton magnetic resonance spectroscopy and a statistical classification strategy. *Pathology* 2002; **34**(5): 417-422 [PMID: 12408339]

33 Wen H, Yoo SS, Kang J, Kim HG, Park JS, Jeong S, Lee JL, Kwon HN, Kang S, Lee DH, Park S. A new NMR-based metabolomics approach for the diagnosis of biliary tract cancer. *J Hepatol* 2010; **52**(2): 228-233 [PMID: 20036026 DOI: S0168-8278(09)00728-4 [pii] 10.1016/j.jhep.2009.11.002]

34 Serkova NJ, Zhang Y, Coatney JL, Hunter L, Wachs ME, Niemann CU, Mandell MS. Early detection of graft failure using the blood metabolic profile of a liver recipient. *Transplantation* 2007; **83**(4): 517-521 [PMID: 17318087 PMCID: 2709529 DOI: 10.1097/01.tp.0000251649.01148.f8 00007890-200702270-00022 [pii]]

35 Melendez HV, Ahmadi D, Parkes HG, Rela M, Murphy G, Heaton N. Proton nuclear magnetic resonance analysis of hepatic bile from donors and recipients in human liver transplantation. *Transplantation* 2001; **72**(5): 855-860 [PMID: 11571450]

36 Munshi SU, Taneja S, Bhavesh NS, Shastri J, Aggarwal R, Jameel S. Metabonomic analysis of hepatitis E patients shows deregulated metabolic cycles and abnormalities in amino acid metabolism. *J Viral Hepat* 2011; **18**(10): e591-602 [PMID: 21914081 DOI: 10.1111/j.1365-2893.2011.01488.x]

37 Andersson U, Lindberg J, Wang S, Balasubramanian R, Marcusson-Stahl M, Hannula M, Zeng C, Juhasz PJ, Kolmert J, Backstrom J, Nord L, Nilsson K, Martin S, Glinghammar B, Cederbrant K, Schuppe-Koistinen I. A systems biology approach to understanding elevated serum alanine transaminase levels in a clinical trial with ximelagatran. *Biomarkers* 2009; **14**(8): 572-586 [PMID: 19780643 DOI: 10.3109/13547500903261354]

38 Sharif AW, Williams HR, Lampejo T, Khan SA, Bansal DS, Westaby D, Thillainayagam AV, Thomas HC, Cox IJ, Taylor-Robinson SD. Metabolic profiling of bile in cholangiocarcinoma using in vitro magnetic resonance spectroscopy. *HPB (Oxford)* 2010; **12**(6): 396-402 [PMID: 20662790 PMCID: 3028580 DOI: HPB185 [pii] 10.1111/j.1477-2574.2010.00185.x]

39 Yang Y, Li C, Nie X, Feng X, Chen W, Yue Y, Tang H, Deng F. Metabonomic studies of human hepatocellular carcinoma using high-resolution magic-angle

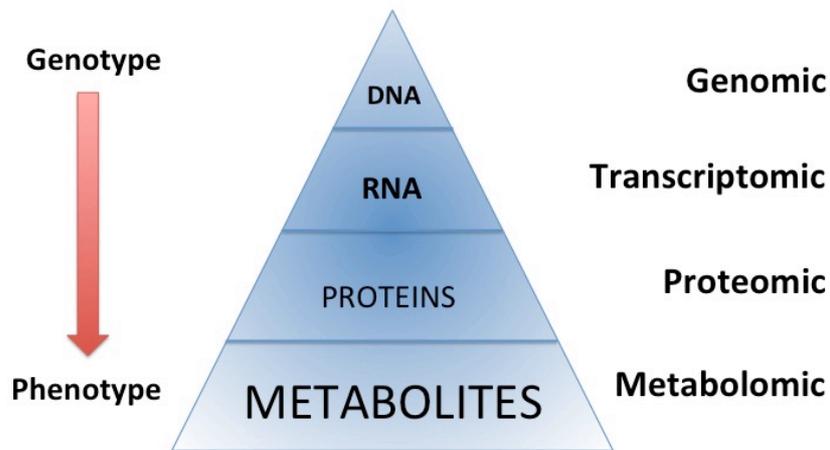
spinning  $^1\text{H}$  NMR spectroscopy in conjunction with multivariate data analysis. *J Proteome Res* 2007; **6**(7): 2605-2614 [PMID: 17564425]

40 Cobbold JF, Patel JH, Goldin RD, North BV, Crossey MM, Fitzpatrick J, Wylezinska M, Thomas HC, Cox IJ, Taylor-Robinson SD. Hepatic lipid profiling in chronic hepatitis C: an in vitro and in vivo proton magnetic resonance spectroscopy study. *J Hepatol* 2010; **52**(1): 16-24 [PMID: 19913320 DOI: S0168-8278(09)00652-7 [pii] 10.1016/j.jhep.2009.10.006]

**Figure 1: "Omic" techniques.** Schematic representation in biological system : Each functional level from the DNA, RNAs, Proteins and metabolites who constitute respectively the genome, transcriptome, proteome and metabolome, have bidirectional flow of information and complex interactions together and with the environment (diseases, drug, lifestyle, genre, habit, diet...). Those interactions produce the phenotype that represents the final output of the system measured in metabolomics.

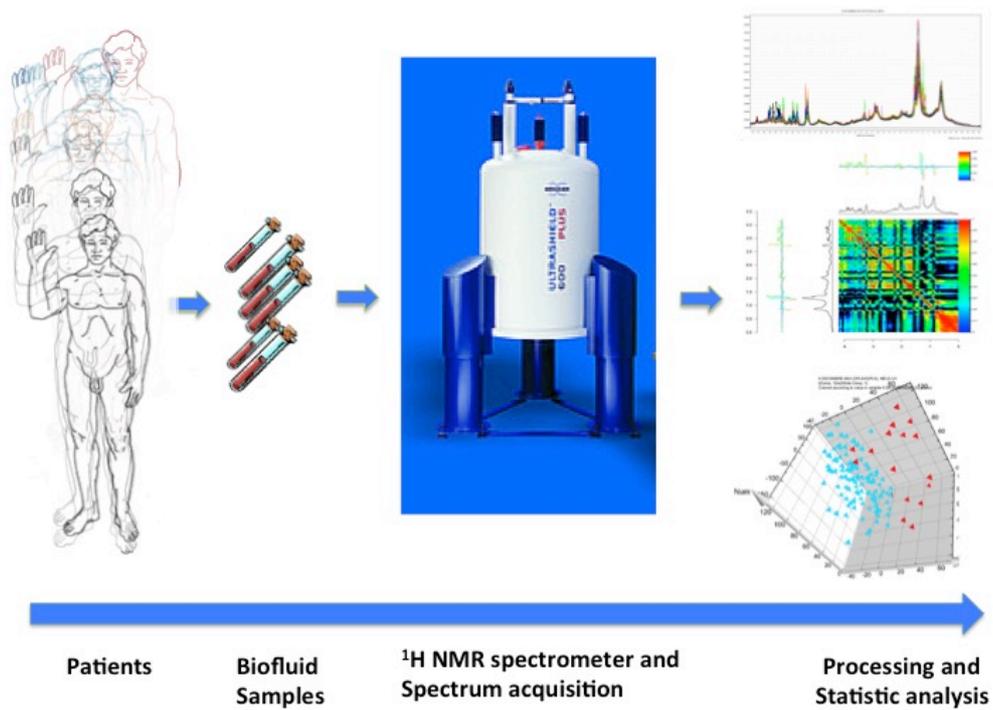
**Figure**

**2:**



**Schematic view of workflow for metabolomic studies: from bedside to bench.**

Proposed standards for metabolomic approach are presented in this schematic view. Clinical question, selection of the population, standardized biofluid collection and conservation, biofluid preparation and spectra acquisition, pre-processing to clean the data for data processing, pre-treatment (i.e. scaling, centering...) to transform the clean data to make them ready for data processing, data analysis (multivariate analysis, unsupervised and supervised analysis), metabolite identification and interpretation.

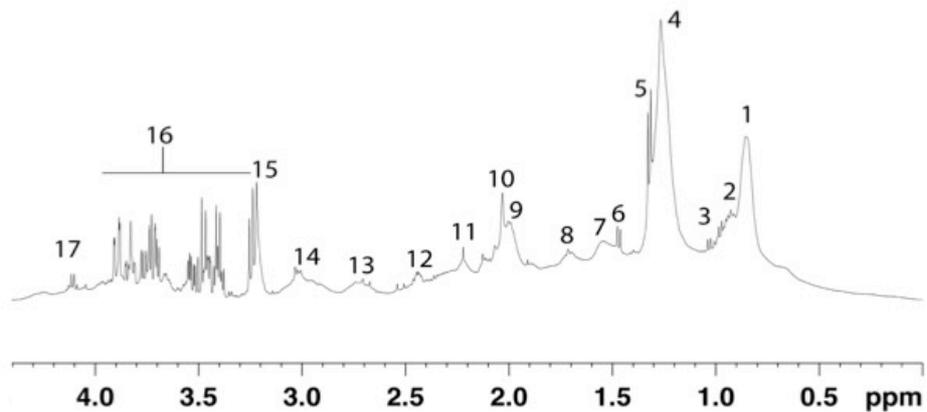


**Figure 3: Typical  $^1\text{H}$  NMR (500 MHz) spectra of the region between 0 and 6 ppm from cirrhotic patient.**

Region between 4.5 and 5.0 ppm corresponding to the water and urea was suppressed.

Peak assignment: 1: Fatty acids ( $-\text{CH}_2-\text{CH}_2-\text{CH}_2-\text{CH}_3$ ); 2: Isoleucine; 3: Valine; 4: Fatty acids ( $-\text{CH}_2-\text{CH}_2-\text{CH}_2-$ ); 5: Lactate; 6: Alanine; 7: Fatty acids ( $-\text{CH}_2-\text{CH}_2-\text{CO}-$ ); 8: Fatty acids ( $-\text{CH}_2-\text{CH}_2-\text{CH}=\text{}$ ); 9: Fatty acids ( $=\text{CH}-\text{CH}_2-\text{CH}_2-$ ); 10: Acetyl signals from a1-acid glycoprotein; 11: Fatty acids ( $-\text{CH}_2-\text{CO}-$ ); 12: Glutamine; 13: Fatty acids ( $=\text{CH}-\text{CH}_2-\text{CH}=\text{}$ ); 14: Albumin lysyl; 15: Choline; 16: Glucose; 17: Lactate; 18: Fatty acids ( $-\text{CH}=\text{CH}-$ ).

*Adapted from [29]*



**Figure 4 (Panel A, B, C): Example of metabolomic study using a training and test set to validate the model.** *Adapted from [29]*

**Creation of the model (A)**

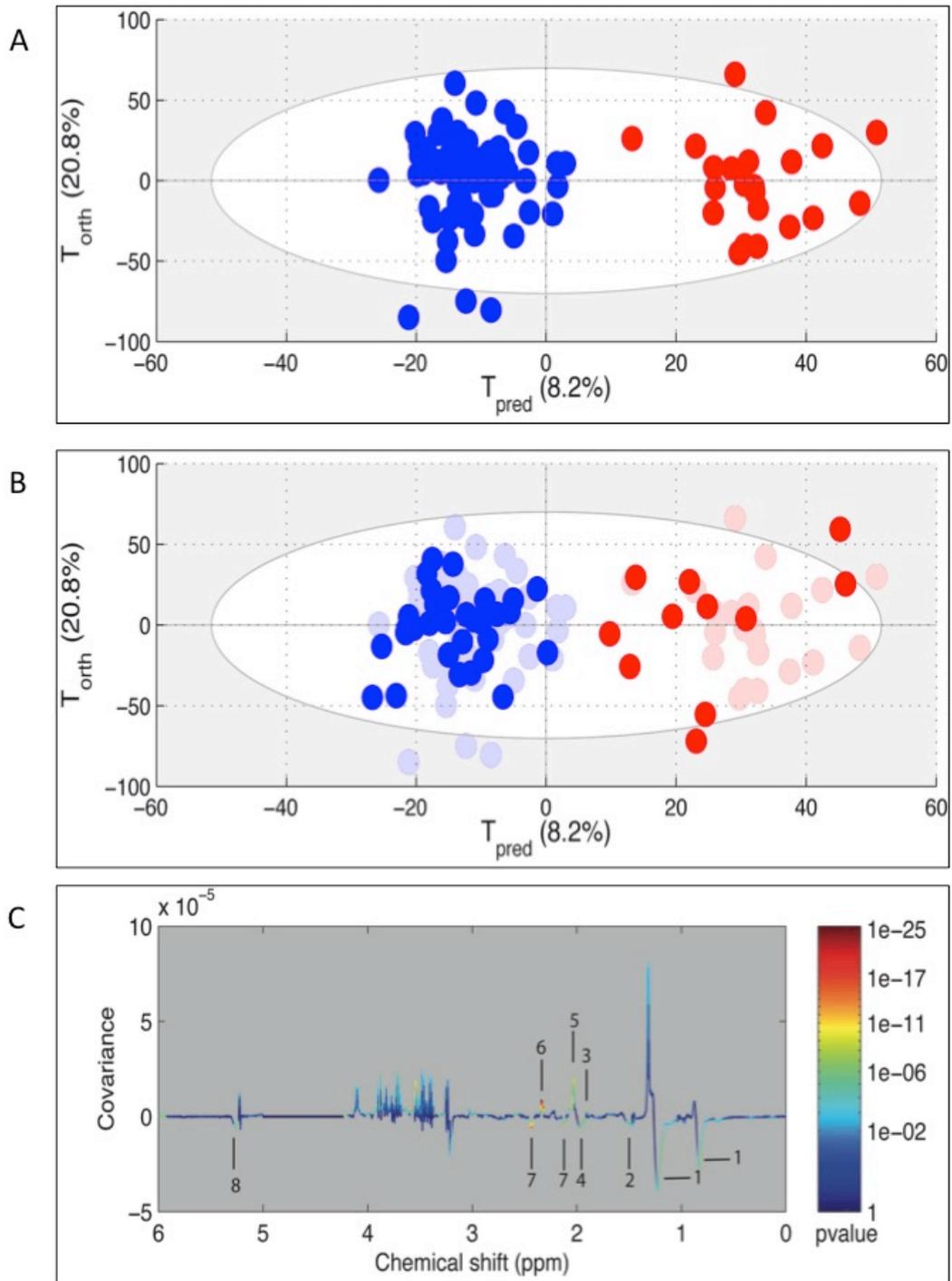
On this figure called score plot, each point represented the projection of an NMR spectrum (and thus one patient is sample) on both axes of the model. On this score plot, each dot corresponds to a spectrum colored according to the absence (blue) or the presence (red) of hepatocellular carcinoma (HCC). The constructed model provides a good distinction between the spectrum of cirrhotic patients without HCC and those with HCC.

**Validation of the model (B)**

Each new spectrum was projected in the score plot using the previously constructed model to enable prediction of the presence or absence of hepatocellular carcinoma (HCC). Each dot corresponds to a spectrum coloured depending on the absence (blue) or presence (red) of HCC.

**Discriminant metabolites (C)**

On this figure called loading plot, variations of bucket intensities are represented using a line plot between 0 to 6 ppm. Positive signals correspond to the metabolites present at increased concentrations in patients with large HCC. Conversely, negative signals correspond to the metabolites present at increased concentrations in patients without HCC. 1 HDL; 2 Fatty Acids; 3 Acetate; 4 Fatty Acids; 5 N-Acetyl-Glycoprotein; 6 Glutamate; 7 Glutamine; 8 Fatty Acids



**Table 1: Advantages and Disadvantages of NMR and Mass spectroscopy.**

	<b>NMR</b>	<b>MS</b>
<b>Sensitivity (detection limit)</b>	Usually micromolar (nanomolar with cryosonde)	picomolar
<b>Reproducibility</b>	High	Low
<b>Detected Metabolite</b>	Non targeted approach Detect metabolite Only if contain proton on the molecule	Targeted approach Need specific preparation to well detected some metabolites (Lipids...)
<b>Metabolite identification</b>	Easy, using 1D and/or 2D spectra and databases	More difficult, need sometime complementary analysis
<b>Number of know identifiable metabolites</b>	More than 200	More than 4000
<b>Sample</b>	Simple preparation (minimal add of D2O, Buffer and sometime reference) Non destructive method Need 400 microliters (less than 10 microliters with microprobe)	Preparation more complex (protein extraction...) Destructive method Need few microliters
<b>Type of sample</b>	Liquid (urine, whole blood, serum, plasma...) and intact tissue	Liquid
<b>Cost of machine</b>	Very High	High
<b>Cost of sample analysis</b>	Lower	Higher
<b>Signal acquisition time</b>	5 to 15 min for 1D spectra More longer for 2D spectra (few hours)	Around 10 minutes

NOTES: NMR, Nuclear Magnetic Resonance; MS, Mass spectroscopy

**Table 2: Examples of Metabolite changes involved in Liver Diseases in NMR based metabolomics approaches.**

Metabolites	Variation	Model/pathology	sample	Reference
2OH butyrate	-	HBV vs HEV	plasma	[36]
3OH butyrate	+	ximelagatran toxicity	plasma	[37]
3OH butyrate	+	Cirrhosis(HBV/ Alcohol) vs controls	serum	[24]
acetate	+	acetaminophen toxicity	urine	[14]
acetate	+	HCC vs Cirrhosis	serum	[29]
acetate	-	Cirrhosis severity	serum	[20]
acetoacetate	-	Cirrhosis and encephalopathy	serum	[26]
acetoacetate	+	Cirrhosis(HBV/ Alcohol) vs controls	serum	[24]
acetoacetate	-	HBV vs Alcohol cirrhosis	serum	[24]
acetoacetate	-	HBV vs control	urine	[36]
acetoacetate	+	AIH vs healthy, PBC, OS and DILI	plasma	[17]
acetone	-	HEV vs control	plasma	[36]
acetone	+	Cirrhosis(HBV/ Alcohol) vs controls	serum	[24]
acetone	-	decompensated vs compensated cirrhosis	serum	[25]
acetone	-	HCC vs Cirrhosis vs controls	urine	[31]
Acetone	+	AIH vs healthy, PBC, OS and DILI	plasma	[17]
bile acids	+	Cholangiocarcinoma vs other causes	bile	[38]
bile acids	+	HCC vs adjacent tissue	liver tissue	[39]
carnitine	-	HBV vs control	plasma	[36]
citrate	+	AIH vs healthy, PBC, OS and DILI	Plasma	[17]
choline	-	fibrosis vs cirrhosis	liver tissue	[21]
choline	+	HCC vs adjacent tissue	liver tissue	[39]
choline, P-choline	-	Cirrhosis severity	serum	[20]
creatine	+	AIH vs healthy, PBC, OS and DILI	plasma	[17]
creatine	+	high grade HCC vs low grade HCC	liver tissue	[39]
Dimethylamine	+	AIH vs healthy, PBC, OS and DILI	plasma	[17]
fatty acids	-	biliary tract cancer	bile	[33]

fatty acids	-	non fonctionnal vs fonctionnal graft after liver transplantation	blood (extraction)	[34]
fatty acids	-	Cirrhosis and encephalopathy	serum	[26]
fatty acids (HDL)	-	Cirrhosis severity	serum	[20]
fatty acids (HDL)	-	HCC vs Cirrhosis	serum	[29]
Glutamine	+	AIH vs healthy, PBC, OS and DILI	plasma	[17]
glycerol	+	HEV vs control	plasma	[36]
glycerol	+	Cirrhosis and encephalopathy	serum	[26]
GPC	-	mild vs moderate fibrosis vs cirrhosis	liver tissue	[40]
GPC	+	HCC vs adjacent tissue	liver tissue	[39]
GPC/P-choline	-	Cirrhosis(HBV/ Alcohol) vs controls	serum	[24]
Histidine	+	AIH vs healthy, PBC, OS and DILI	plasma	[17]
isobutyrate	+	Cirrhosis(HBV/ Alcohol) vs controls	serum	[24]
isobutyrate	+	HBV vs Alcohol cirrhosis	serum	[24]
LDL	-	Cirrhosis(HBV/ Alcohol) vs controls	serum	[24]
LDL	-	decompensated vs compensated cirrhosis	serum	[25]
lipids	-	HCC vs adjacent tissue	liver tissue	[39]
lipids	-	high grade HCC vs low grade HCC	liver tissue	[39]
OH-butyrate	-	Cirrhosis severity	serum	[20]
P-choline	+	fibrosis vs cirrhosis	liver tissue	[21]
P-choline	-	mild vs moderate fibrosis vs cirrhosis	liver tissue	[40]
P-choline	+	HCC vs adjacent tissue	liver tissue	[39]
P-choline	+	Cirrhosis and encephalopathy	serum	[26]
P-choline/GPC	-	high grade HCC vs low grade HCC	liver tissue	[39]
Pdt-choline	+	Cholangiocarcinoma vs non cancer	bile	[38]
Pdt-choline	-	mild vs moderate fibrosis vs cirrhosis	liver tissue	[40]
P-ethanolamine	+	HCC vs adjacent tissue	liver tissue	[39]
P-ethanolamine	+	high grade HCC vs low grade HCC	liver tissue	[39]

P-ethanolamine	+	fibrosis vs cirrhosis	liver tissue	[21]
PUFA	-	mild vs moderate fibrosis vs cirrhosis	liver tissue	[40]
Pyruvate	+	AIH vs healthy, PBC, OS and DILI	Plasma	[17]
saturation index	+	mild vs moderate fibrosis vs cirrhosis	liver tissue	[40]
total lipids	+, -	mild vs moderate fibrosis vs cirrhosis	liver tissue	[40]
Total choline/lipids	-	mild vs moderate fibrosis vs cirrhosis	liver tissue	[40]
unsaturated FA	+	fibrose vs cirrhose	liver tissue	[21]
unsaturated FA	+	Cirrhosis(HBV/Alcohol) vs controls	serum	[24]
VLDL	-	Cirrhosis(HBV/Alcohol) vs controls	serum	[24]
VLDL	-	decompensated vs compensated cirrhosis	serum	[25]

NOTES: HBV: hepatitis B virus; HBE: hepatitis E virus; HBC: hepatitis C virus; HCC: hepatocellular carcinoma; AIH: autoimmune hepatitis; OS: overleap syndrome; PBC: primary biliary cirrhosis; DILI: drug induced liver injury; FA: Fatty Acid; HDL: high density lipoproteins; LDL: low density lipoproteins; VLDL: very low density lipoproteins; P-choline: phospho-choline; GPC: glycerophospho-choline; P-ethanolamine: phospho-ethanolamine; PUFA: polyunsaturated fatty acids; Pdt-choline: phosphatidylcholine;



# Travaux de recherche

---

## **Chapitre I Stéatose hépatique non alcoolique (NAFLD) et stéato-hépatite non alcoolique (NASH) : approche diagnostique par étude métabolomique RMN à partir d'extractions de tissus de foies humains.**

### I Contexte introductif

Au cours de ces dernières années, la stéatose hépatique non alcoolique « NAFLD » est devenue, dans les pays industrialisés, la cause la plus fréquente de maladie hépatique chronique. La prévalence de cette atteinte hépatique, estimée à 20-30% dans les pays occidentaux, progresse de manière constante depuis plusieurs années en raison de l'épidémie mondiale de surcharge pondérale et d'obésité [183]. Les facteurs de risque de la survenue des NAFLD sont associés au syndrome métabolique regroupant le diabète de type 2, l'insulino-résistance, l'obésité, l'hypertension artérielle, l'hypercholestérolémie et l'hypertriglycéridémie. La NAFLD est caractérisée par une accumulation graisseuse au sein des hépatocytes en l'absence de consommation excessive d'alcool ou autre cause secondaire. Souvent asymptomatique, cette accumulation peut entraîner une réponse inflammatoire appelée stéato-hépatite non alcoolique (NASH) qui favorise l'apparition d'une cirrhose voire même le développement d'un carcinome hépatocellulaire. La prévalence des patients souffrants de stéatose progressant en NASH est estimée de 10 à 25% et de 2 à 3% pour l'apparition du CHC [184,185]. Le diagnostic du NASH relève par conséquent d'une grande importance pour le pronostic. Les diagnostics du NAFLD et du NASH sont très souvent difficiles à poser en raison du caractère asymptomatique de ces hépatopathies. La biopsie hépatique reste l'outil diagnostique le plus approprié mais cette technique invasive et risquée ne doit se pratiquer qu'en cas de doute persistant. Aucun marqueur clinique ou biochimique ne dispose aujourd'hui de sensibilité ni de spécificité suffisante pour établir le diagnostic du NAFLD ou du NASH [127].

Actuellement, il est admis que l'accumulation de lipides au niveau des hépatocytes est due à un déséquilibre entre un apport important d'acides gras par l'alimentation, la lipolyse des tissus graisseux et la lipogenèse hépatique par rapport à la diminution de l'élimination et de l'oxydation des acides gras. Les voies de signalisation intracellulaire se retrouvent alors

perturbées provoquant une résistance à l'action de l'insuline. Un autre mécanisme participant à une insulino-résistance est la dérégulation de la biogenèse et du fonctionnement des mitochondries causant une baisse de leur capacité oxydative [186].

Cependant, l'apparition d'une atteinte NAFLD n'est pas toujours expliquée par un syndrome métabolique rendant la pathogenèse du NAFLD et du NASH complexe et loin d'être complètement élucidée. En raison de sa prévalence croissante, la compréhension des mécanismes impliqués dans le développement des NAFLD constitue un enjeu majeur de santé publique [127,187].

La biopsie hépatique étant l'outil diagnostique le plus fiable pour la détermination des atteintes NAFLD et en particulier le NASH, l'analyse du tissu hépatique constitue actuellement le meilleur moyen d'évaluer l'étendue de l'atteinte hépatique.

D'autre part, la métabolomique entreprise sur les échantillons tissulaires peut constituer un outil utile dans l'étude du métabolisme anormal des maladies et être susceptible de fournir une information des modifications métaboliques en amont des mécanismes de régulation de ces maladies [188].

Différentes études métabolomiques ont été appliquées pour l'étude du NAFLD et de sa progression en stéato-hépatite ou en CHC. Néanmoins, la majorité d'entre elles ont été entreprises par spectrométrie de masse et se sont portées sur l'étude des altérations métaboliques dans les liquides biologiques tels que l'urine et le plasma [149-151,155]. Très peu d'entre elles ont étudié les modifications métaboliques dans le tissu hépatique humain [148].

Dans le but d'aider à une meilleure compréhension des mécanismes des atteintes NAFLD dans le métabolome du foie mais également dans l'optique d'améliorer la classification de ces hépatopathies, une étude métabolomique en RMN a été entreprise sur l'analyse d'extraits tissulaires d'un ensemble de patients présentant un foie normal, une stéatose ou un NASH.

## II Matériel et Méthode

### **II.1 Constitution de la cohorte**

Entre juin 2011 et mai 2015, un total de 72 patients obèses a été recruté au sein du service de chirurgie digestive et métabolique du CHU de Jean Verdier de Bondy. Ces patients étaient candidats à une procédure bariatrique (gastroplastie par pose d'anneau ajustable, gastrectomie longitudinale en manchon, By-Pass gastrique roux en Y). Des biopsies hépatiques ont été réalisées en peropératoire par laparotomie lors des différentes chirurgies bariatriques. Tous les patients inclus dans la cohorte prospective ont signé un consentement couvrant la biopsie hépatique peropératoire ainsi que l'utilisation ultérieure des échantillons congelés.

Les critères d'inclusion étaient les suivants : (1) âge de 18 ans et plus avec une obésité morbide :  $IMC \geq 40$  ou  $\geq 35$  avec au moins une comorbidité associée (hypertension artérielle, diabète de type 2, dyslipidémie, syndrome d'apnées obstructives du sommeil, arthrose et/ou NAFLD) ; (2) absence d'autre infection hépatique (critères d'exclusion) : maladie hépatique auto-immune ou infectieuse (hépatite virale chronique), hémochromatose, antécédents de chimiothérapie ou prise de médicaments hépatotoxiques, consommation alcoolique supérieure à 20 g/jour pour les femmes et supérieure à 30g/jour pour les hommes ; (3) Eligibilité physique et psychologique d'une intervention chirurgicale bariatrique (décidé lors de la réunion de concertation pluridisciplinaire en chirurgie bariatrique). Les données démographiques et cliniques des patients ont été enregistrées au moment de l'inclusion et sont reprises dans le tableau 4.

La biopsie hépatique a été réalisée sur le bord externe du lobe gauche du foie d'une profondeur de 2 cm minimum dans le parenchyme. Cette biopsie a immédiatement été divisée en trois portions : 1) une partie destinée pour les analyses des flux métaboliques ; 2) une partie pour l'analyse histologique ; 3) une partie destinée à la congélation directe à -80 °C pour les futures analyses métaboliques. L'analyse histologique basée sur un algorithme décrit par Bedossa *et al.* [189] a permis la classification des biopsies hépatiques des patients en trois catégories : **foie normal** (stéatose, inflammation et ballonisation = 0) ; **stéatose** (stéatose 1-3 + soit inflammation 1-3, soit ballonisation 1-2) et **NASH** (stéatose 1-3 avec inflammation 1-3 et ballonisation 1-2 + fibrose 1-4).

	<b>Foie normal (n=16)</b>	<b>Stéatose (n=30)</b>	<b>NASH (n=26)</b>	<b>p-value*</b>
<b>Age (années)</b>	33.5 ± 8.6	37.7 ± 11.5	39.3 ± 10.1	-
<b>Sexe féminin (%)</b>	94	83	65	-
<b>IMC (kg/m<sup>2</sup>)</b>	43.4 ± 5.1	44.9 ± 5.7	46 ± 5.9	0.41
<b>ASAT (IU/L)</b>	21.7 ± 6.2	24.8 ± 11.8	31.3 ± 14.2	<b>0.03</b>
<b>ALAT (IU/L)</b>	19.3 ± 7.8	29.3 ± 19.5	41.6 ± 20.7	<b>0.0009</b>
<b>GGT (IU/L)</b>	21.7 ± 9.5	33.1 ± 31.4	42.2 ± 30.5	0.075
<b>Cholestérol (mmol/L)</b>	4.7 ± 0.7	4.8 ± 1.4	5 ± 1.1	0.64
<b>TG (mmol/L)</b>	0.9 ± 0.4	1.3 ± 0.7	2.2 ± 1.3	<b>7.1 10<sup>-5</sup></b>
<b>Glycémie à jeun (mmol/L)</b>	4.8 ± 0.5	5.4 ± 1.7	6.2 ± 2.3	<b>0.034</b>
<b>Insuline à jeun (pmol/L)</b>	75.2 ± 41.7	158.3 ± 117.6	163.4 ± 148	0.066

La classification des patients a été établie à partir des biopsies hépatiques. IMC= Indice de masse corporel ; ASAT= aspartate aminotransférase ; ALAT= alanine aminotransférase ; GGT= gamma-glutamyl transférases ; TG= triglycérides. Les valeurs sont données par la moyenne ± l'écart-type par rapport à la moyenne. \*Les p-values ont été calculées d'après une ANOVA et sont significatives lorsque  $P < 0.05$ .

**Tableau 4.** Caractéristiques clinicopathologiques des patients inclus dans l'étude

## II.2 Méthodes d'extraction des métabolites

Les biopsies hépatiques congelées ont été réduites en poudre et 100 mg d'échantillon ont été soumis à un protocole d'extraction par un mélange de solvants : eau/méthanol/chloroforme dans des proportions volumiques de 1/1.4/1.4 afin de séparer simultanément les métabolites hydrosolubles des métabolites liposolubles. Le protocole utilisé a été adapté suivant les recommandations des procédures métabolomiques de Beckonert *et al.* [17] et de Beltran *et al.* [55]. Afin de limiter les processus enzymatiques lors de la décongélation des tissus, les solvants ont été maintenus dans un bain de glace et la méthode d'extraction a également été réalisée dans un bain de glace. Les 100 mg de poudre de foie ont été homogénéisés pendant une minute avec 100 µl d'eau et 400 µl de méthanol

suivi de l'ajout de 400  $\mu\text{l}$  de chloroforme et 185  $\mu\text{l}$  d'eau. L'homogénat a ensuite été stocké à 4°C pendant 15 min avant d'être centrifugé 15 min à 4°C et à 1000g. Après centrifugation, les deux phases formées ont alors été séparées et lyophilisées et séchées pendant une nuit. Pour certaines biopsies, les quantités de foie (supérieures à 200 mg) ont permis de réaliser deux extractions. La réalisation de ces doublons a pour but de préciser la reproductibilité de la méthode d'extraction.

### ***II.3 Analyses RMN***

Les échantillons lyophilisés ont été redissous dans 580  $\mu\text{l}$  de PBS-D<sub>2</sub>O pour la fraction aqueuse et dans 580  $\mu\text{l}$  de CDCl<sub>3</sub>/CD<sub>3</sub>OD (2 :1) pour la fraction organique. 550  $\mu\text{l}$  ont alors été placés dans chacun des tubes RMN de 5 mm afin de procéder à l'acquisition des spectres. Les spectres proton 1D ont été acquis pour chaque échantillon à 500 MHz sur un spectromètre Bruker Avance III à 23°C. Les fractions aqueuses ont été enregistrées avec une séquence noesygppr1D tandis qu'une séquence noesy1d a été appliquée pour les fractions organiques. Pour chaque échantillon, 32 accumulations ont été réalisées avec 64 k points pour une fenêtre spectrale de 20 ppm. Le temps de relaxation était de 2 s.

### ***II.4 Traitement des spectres***

Les spectres ont été traités comme décrit au point III.3 du chapitre 2 de la section Méthodologies à l'aide du logiciel NMRPipe [77]. La région spectrale comprise entre 0 et 10.5 ppm a été divisée en 10 500 buckets de 0.001 ppm. Les régions correspondantes aux solvants résiduels (eau, méthanol et chloroforme) ont été exclues de l'analyse. Avant l'analyse statistique les données ont été centrées-réduites (UV scaling) et normalisées (quotient normalisation).

### ***II.5 Analyses statistiques***

Une analyse non supervisée de type ACP ainsi que des analyses supervisées de type OPLS-DA ont été réalisées comme décrites à la section II du chapitre 3 de la section Méthodologies.

### III Résultats & Discussion

#### ***Caractéristiques des données clinico-pathologiques des patients.***

Les données cliniques et biochimiques des patients inclus dans cette étude sont résumées dans le tableau 4. L'étude inclut des patients d'un âge similaire ( $37.3 \pm 10.5$  ans) et une majorité de sujets féminins (79%). Nous pouvons constater une augmentation significative de quatre paramètres clinico-pathologiques qui sont les enzymes hépatiques ASAT et ALAT, les triglycérides et la glycémie à jeun. L'indice de masse corporel (IMC) n'a pas montré d'augmentation qui soit significative chez les patients avec une stéatose et un NASH comparés aux patients « foie normal ». Néanmoins, différentes études menées sur les facteurs de risque du NAFLD ont montré une augmentation linéaire de la prévalence du NAFLD avec l'IMC [190,191] et ont démontré que l'IMC était un paramètre à considérer dans l'évaluation d'une atteinte NAFLD [151].

#### ***Reproductibilité de la méthode d'extraction.***

La manipulation des échantillons lors des étapes pré-analytiques peut introduire des variabilités supplémentaires liées à cette manipulation. Il a été montré, par exemple, que le temps de décongélation des tissus pouvait influencer l'interprétation des résultats à cause des processus de dégradation et activité enzymatique [17]. Il s'avère par conséquent crucial, pour une meilleure interprétation des résultats, d'appliquer un protocole bien établi pour l'ensemble des échantillons. Il convient alors de vérifier que la variation introduite lors de cette étape pré-analytique soit plus faible que la variabilité biologique intrinsèque.

Le tableau 5 résume l'ensemble des échantillons disponibles pour l'extraction. Pour les 72 échantillons disponibles, 53 étaient en quantité suffisante pour être dupliqués. Un total de 125 échantillons ont subi une extraction.

	<i>Nombre de patients</i>	<i>Nombre de replica disponibles</i>	<i>Nombre d'échantillons</i>
<i>Foie normal</i>	16	11	27
<i>Stéatose</i>	30	23	53
<i>NASH</i>	26	19	45
<i>TOTAL</i>	72	53	125

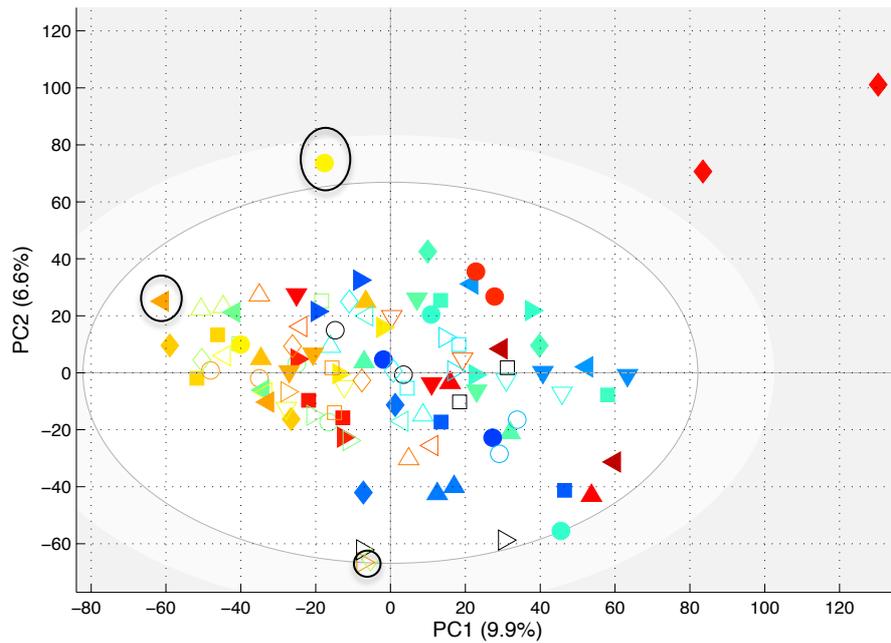
**Tableau 5.** Echantillonnage pour l'extraction.

Afin d'évaluer l'ampleur de la variabilité engendrée par les manipulations d'extraction des échantillons, une ACP a été réalisée avec l'ensemble des échantillons possédant un duplicate, en premier lieu, issus de la phase aqueuse et ensuite avec les échantillons issus de la phase organique.

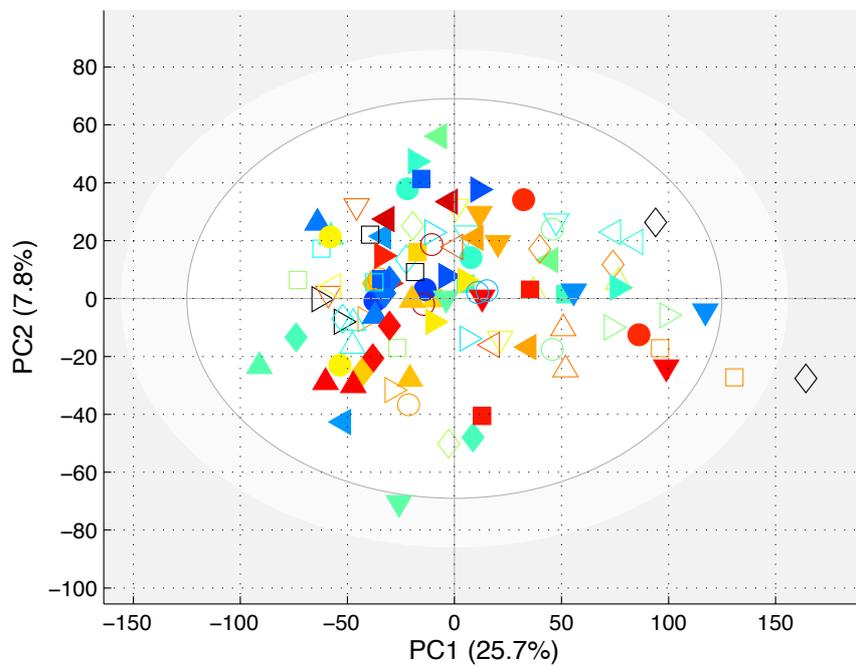
Le score plot de l'ACP des échantillons issus de la phase aqueuse est représenté en figure 25. De manière générale, les duplicates se retrouvent assez proche l'un de l'autre (représentés par une même forme et une même couleur). Certains échantillons plus éloignés de leurs semblables, représentés par les 3 points entourés sur le score plot, trouvent une justification de mauvaise manipulation relevée lors de l'extraction (prélèvement d'un peu de phase lipophile et temps d'attente à 4°C plus long).

Le score plot de l'ACP des échantillons issus de la phase organique est représenté en figure 26. De la même façon que pour l'ACP des fractions aqueuses, nous avons constaté que les duplicates étaient relativement proches entre eux suggérant une faible variabilité liée à la manipulation comparée à la variabilité biologique intrinsèque des échantillons.

Nous avons choisi, par conséquent, pour chacune des fractions, de moyenner les données centrées réduites des échantillons duplicates pour les futurs modèles réalisés excepté pour les trois échantillons qui ont subi une erreur de manipulation pour la fraction aqueuse.



**Fig. 25 :** Score plot de l'ACP obtenue à partir des extractions aqueuses des 106 échantillons dupliques correspondant à 53 patients. Les 3 points entourés correspondent à une mauvaise manipulation lors de l'extraction.



**Fig. 26 :** Score plot de l'ACP obtenue à partir des extractions organiques des 106 échantillons dupliques correspondant à 53 patients.

***Effet décongélation-recongélation d'une partie des échantillons.***

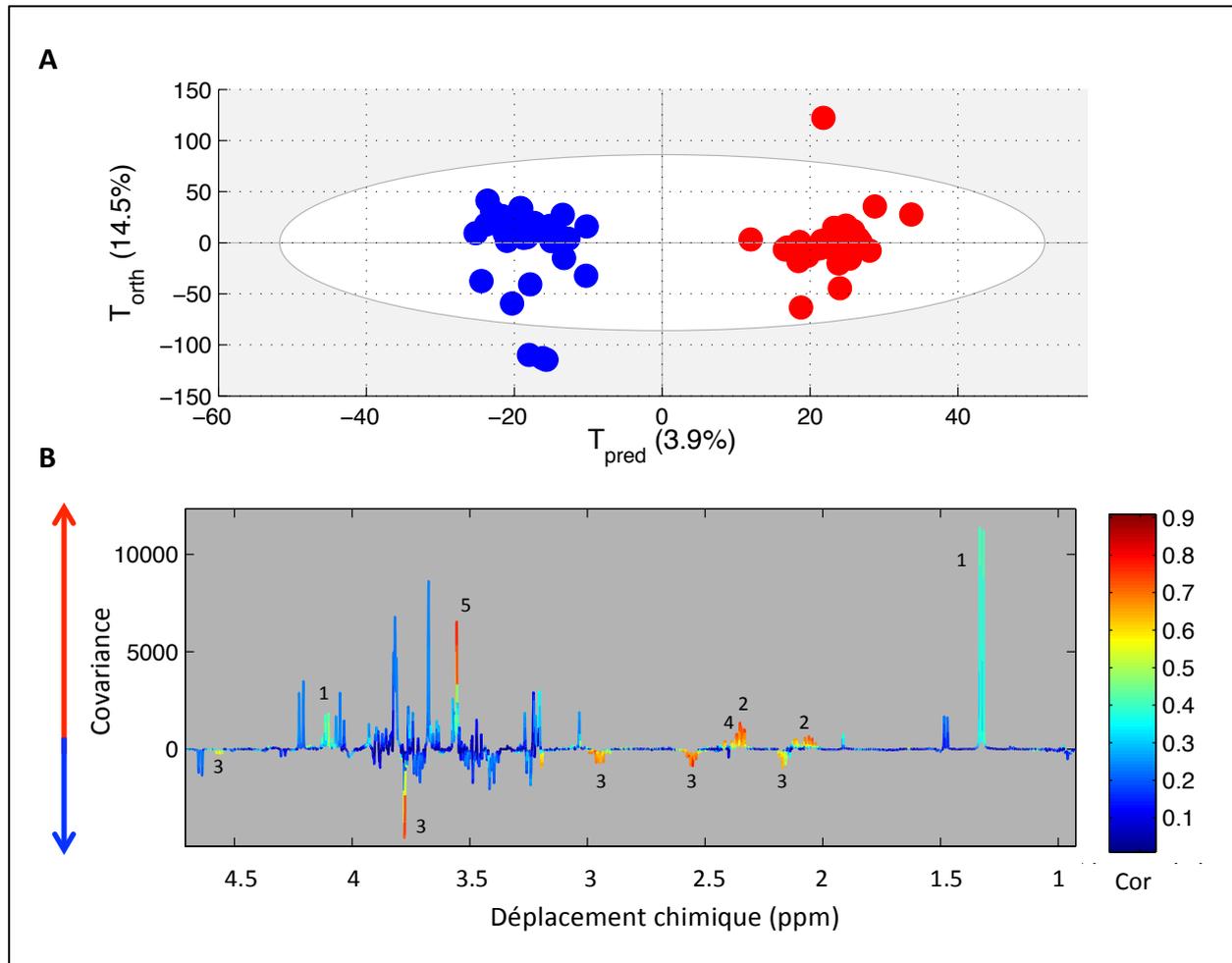
Une partie des échantillons était susceptible d'avoir subi un cycle de décongélation-recongélation qui, de manière prévisible, ne serait pas sans conséquences sur l'interprétation des résultats.

En effet, sur les échantillons des 72 patients, 32 étaient potentiellement victimes d'un cycle de décongélation-recongélation. Une analyse OPLS-DA a, par conséquent, été réalisée sur l'ensemble des fractions aqueuses afin d'évaluer l'existence et l'impact d'une possible décongélation/recongélation sur ces 32 échantillons. A noter qu'aucune séparation entre les deux groupes d'échantillons ayant probablement ou non subi ce cycle de décongélation-recongélation au niveau de l'ACP, n'a pu être observée préalablement. Les résultats de l'analyse discriminante réalisée avec 4 composantes sont présentés en figure 27. Les paramètres de qualité du modèle sont donnés par une valeur de  $R^2$  de 0.96, de  $Q^2$  de 0.8 et de CV-AUROC égale à 1, indiquant une discrimination significative entre les deux groupes d'échantillons. Ainsi, il apparaît maintenant très probable que ces 32 échantillons aient subi ce cycle de décongélation-recongélation.

Les métabolites impactés par l'effet de la décongélation/recongélation sont mis en évidence sur le loading plot (Fig. 27B) et correspondent à une augmentation des concentrations du lactate (1), du glutamate (2), du pyruvate (4) et de la glycine (5), et une diminution de celle du glutathion (3) observées dans les échantillons victimes de la décongélation/recongélation.

Différentes études avaient déjà démontré que l'intégrité des métabolites dans le sérum et l'urine était compromise lors des cycles de décongélation/recongélation [192-194]. Les métabolites concernés par ces facteurs sont les métabolites qui sont intrinsèquement instables tels que ceux impliqués dans le métabolisme énergétique intervenant dans la glycolyse, le cycle de Krebs ainsi que dans la voie des pentoses-phosphates [16]. Les cycles de décongélation/recongélation sont ainsi évités le plus souvent possible lors des analyses.

Concernant l'analyse ACP et OPLS-DA de la fraction organique, le cycle de décongélation/recongélation n'a pas eu d'impact sur l'analyse des métabolites lipophiles puisque aucun modèle n'a permis de discriminer les échantillons ayant subi ce cycle de décongélation/recongélation des autres.



**Fig. 27** : Score (A) et loading plot (B) de l'analyse OPLS-DA réalisée sur les fractions aqueuses des 32 échantillons ayant subi un cycle de décongélation/recongélation (en rouge) et des 40 autres échantillons préservés de cet événement (en bleu). Les variations significatives des métabolites sont représentées entre 1 et 4.5 ppm sur le loading plot. Un signal positif correspond aux métabolites présents à des concentrations plus élevées pour les échantillons ayant subi le cycle de décongélation/recongélation. A l'inverse, les signaux négatifs correspondent aux métabolites présents à des concentrations plus élevées pour les échantillons préservés du cycle décongélation/recongélation. Annotations : 1= lactate ; 2= glutamate ; 3= glutathion ; 4= pyruvate ; 5= glycine ; cor= corrélation.

***Evaluation des atteintes NAFLD par profilage métabolique.***

Une analyse OPLS-DA a été entreprise dans le but de déterminer si un profil métabolique se différencie des extraits tissulaires obtenus à partir de foies normaux, stéatoses et NASH.

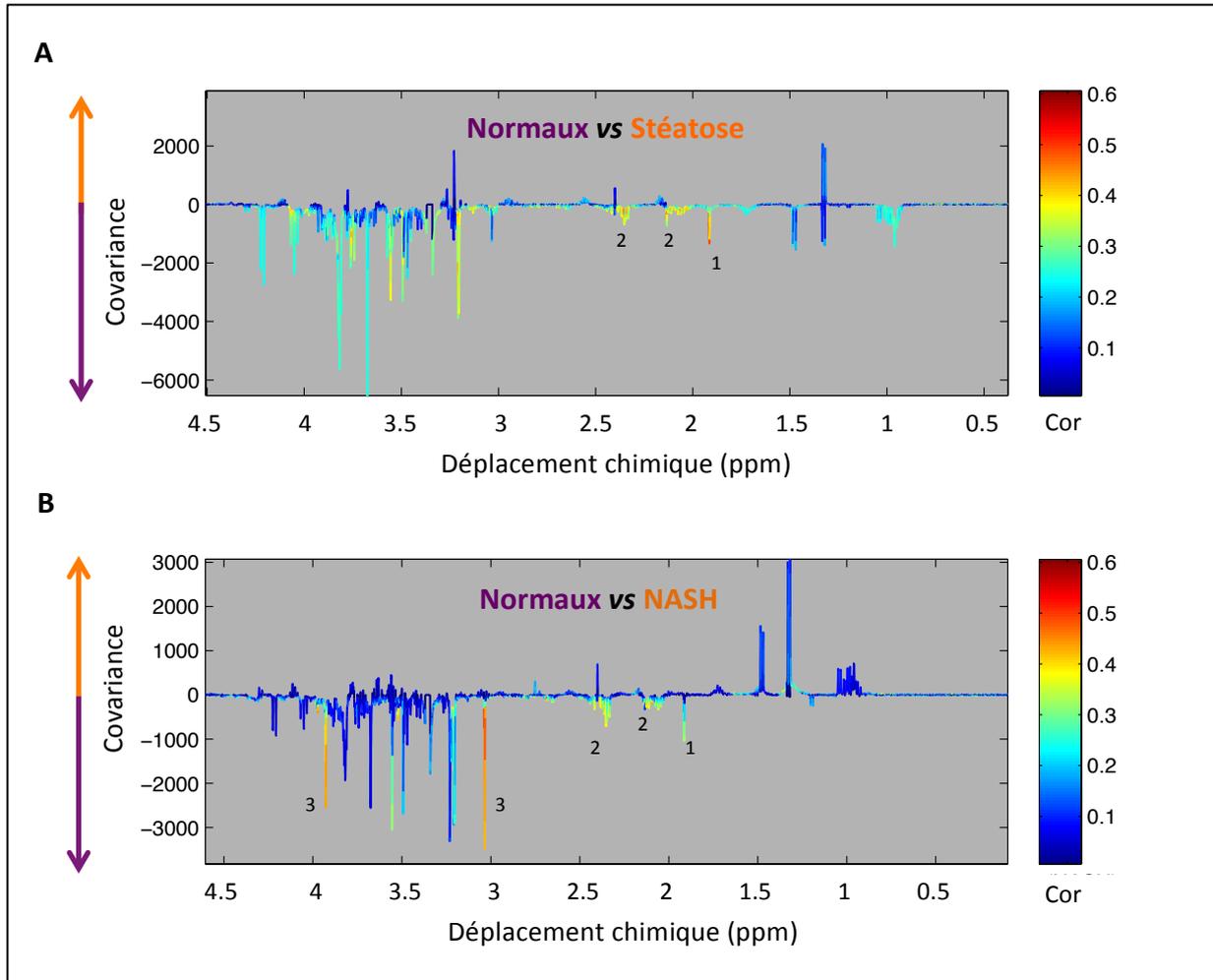
- Analyse de la fraction aqueuse

L'exclusion des échantillons ayant subi le cycle de décongélation/recongélation implique une trop petite cohorte d'échantillons restante et ne permet, par conséquent pas de réaliser des modèles statistiques qui soient significatifs en métabolomique.

L'analyse OPLS-DA de la fraction aqueuse de l'ensemble des échantillons n'a pas permis d'obtenir de modèles significatifs permettant la différenciation des extraits tissulaires des foies contrôles, stéatoses et NASH. Néanmoins, certains métabolites semblent être corrélés en univarié tels que le N-acétyl des glycoprotéines (1), le glutamate (2) et la créatine (3) comme représentés sur les loadings de la figure 28.

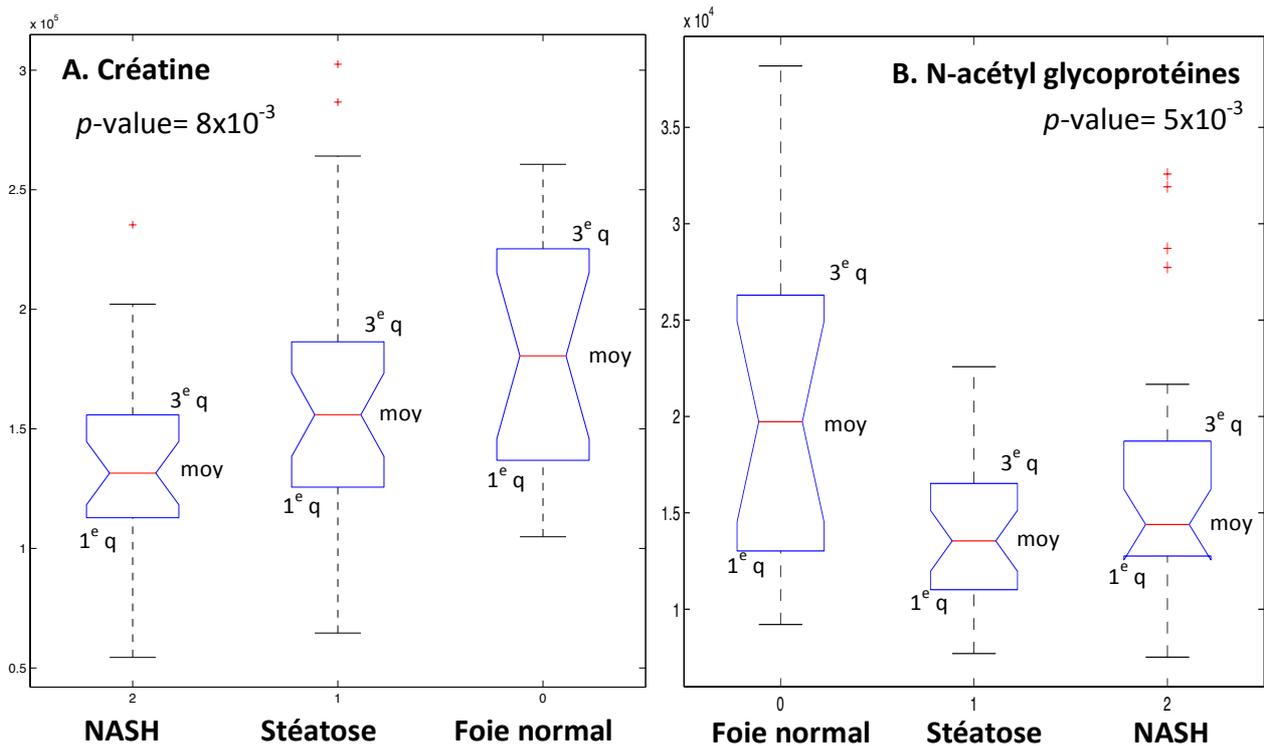
Il est à noter que ces métabolites, excepté pour le glutamate ne sont pas les métabolites qui avaient été observés précédemment lors du cycle de décongélation/recongélation.

Le signal de la créatine semble être plus particulièrement corrélé et sa concentration diminuée chez les NASH comparée à celle du groupe contrôle (foie normal). Le signal N-acétyl des glycoprotéines semble quant à lui être plus particulièrement corrélé et sa concentration diminuée lors de la stéatose comparée au groupe contrôle.



**Fig. 28 :** Loading plots de l'analyse OPLS-DA obtenue à partir des fractions aqueuses des échantillons (A) de foies normaux et stéatoses ; (B) de foies normaux et NASH. Les signaux des N-acétyl des glycoprotéines (1), du glutamate (2) et de la créatine (3) semblent être corrélés en univarié et augmentés chez les foies normaux comparés aux stéatoses et aux NASH.

Afin de vérifier la significativité des observations de l'augmentation de la créatine, du glutamate et du groupement N-acétyl des glycoprotéines lors de la stéatose et du NASH, un test ANOVA a été réalisé sur les trois groupes : contrôles, stéatoses, NASH. Le test ANOVA réalisé sur les valeurs de la créatine a montré une différence significative ( $p\text{-value} = 8 \times 10^{-3}$ ) du taux de créatine observé entre les 3 groupes avec un niveau de créatine plus élevé chez les contrôles suivi des stéatoses et des NASH (Fig. 29A). L'ANOVA du signal N-acétyl des glycoprotéines a également montré une différence significative ( $p\text{-value} = 5 \times 10^{-3}$ ) entre les 3 groupes (Fig. 29B). En revanche, la concentration du glutamate n'a pas montré de différence significative entre les 3 groupes.



**Fig. 29 :** Représentation des valeurs en « boîte à moustache » de la créatine (A) et du groupement N-acétyl des glycoprotéines (B) pour les 3 groupes correspondant aux extraits aqueux des tissus de foies normaux, stéatoses et NASH. Moy= moyenne ; 1<sup>e</sup> q= 1<sup>er</sup> quartile ; 3<sup>e</sup> q= 3<sup>e</sup> quartile. Les croix rouges correspondent aux échantillons aberrants.

Très peu d'études métabolomiques ont été entreprises sur l'évaluation des atteintes NAFLD en RMN. Une étude s'est intéressée à un modèle de NAFLD induit chez la souris par l'administration d'un régime déficitaire en méthionine et choline [195]. Il en est ressorti une augmentation de la concentration d'alanine, d'arginine/lysine, glutamate/glutamine et de créatine et une diminution de VLDL/LDL, leucine, N-acétyl des glycoprotéines, du TMAO/bétaine et de glucose dans le sérum des souris atteintes de divers grades de NAFLD comparés aux souris contrôles. Parmi ces métabolites, la combinaison de 4 d'entre eux (glucose, lactate, glutamate/glutamine, taurine) a été sélectionnée par des méthodes d'analyse de cluster hiérarchique et de réseau de neurones artificiels comme potentiels biomarqueurs pour le diagnostic des atteintes NAFLD.

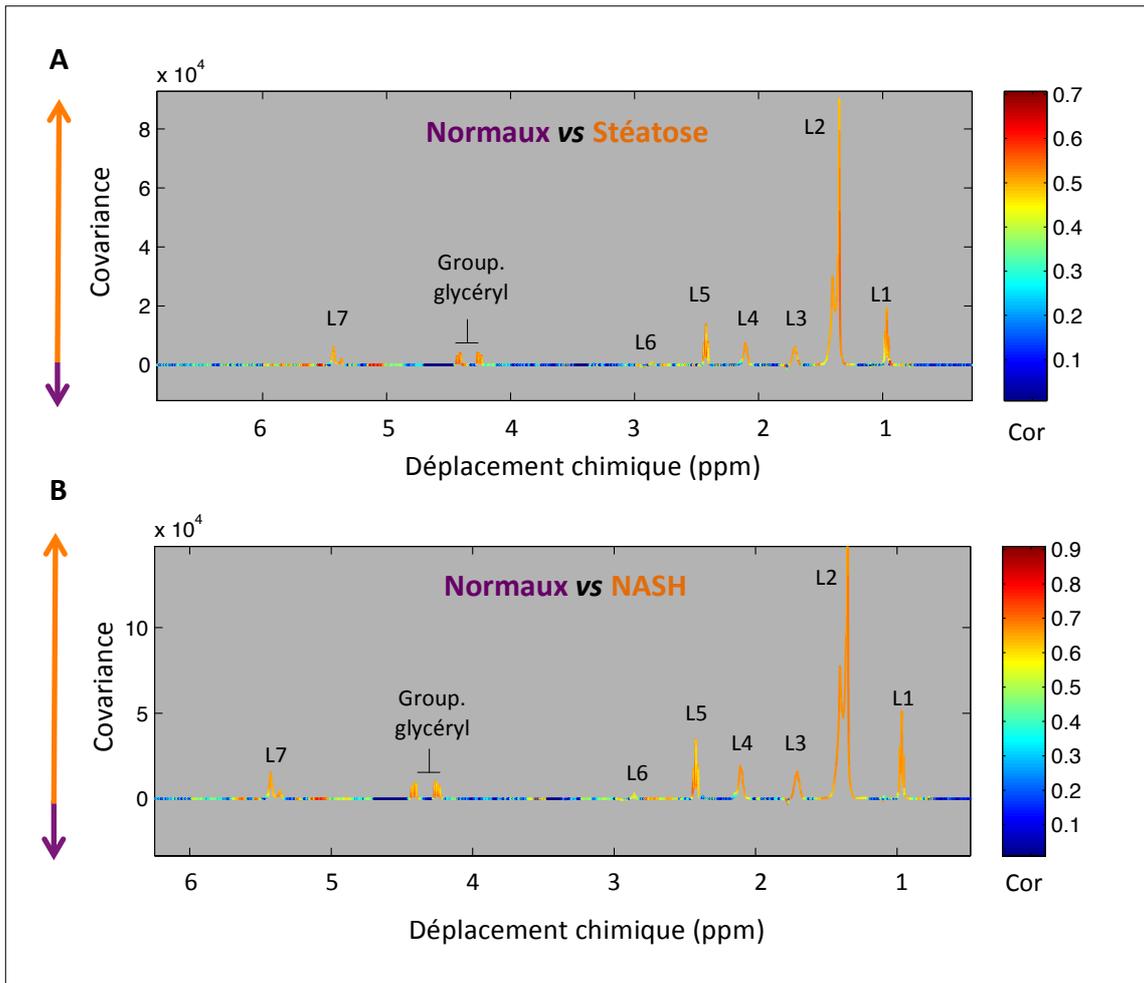
En spectrométrie de masse, Garcia Canaveras *et al.* [148] ont également constaté une diminution du niveau de la créatine lors de l'atteinte NAFLD au cours d'une étude métabolomique réalisée sur le tissu de foie de patients avec et sans stéatose.

Par ailleurs, il a été montré qu'une supplémentation de créatine entraîne une réduction de l'accumulation de triglycérides hépatiques grâce à une stimulation de l'oxydation des acides gras et aiderait de cette manière à prévenir l'apparition de stéatose hépatique [196,197]. En équilibre avec l'apport alimentaire, le foie est également capable de synthétiser la créatine. Une diminution du niveau de la créatine lors d'une atteinte NAFLD et NASH telle qu'observée dans notre étude pourrait être révélatrice d'un dysfonctionnement dans l'équilibre de biosynthèse de ce composé dans le foie.

D'autres études métabolomiques réalisées en spectrométrie de masse ont également mis en évidence une modification du taux de glutamate dans le plasma ou le sérum de sujets atteints de NAFLD [149,151,155].

- Analyse de la fraction organique

Concernant la fraction organique, les modèles réalisés en OPLS-DA ont permis de discriminer les trois groupes correspondant au foie normal, stéatose et NASH. Après exclusion de deux échantillons de patients outliers, les paramètres de qualité du modèle OPLS-DA réalisé avec 2 composantes pour la discrimination foie normal (n=15) et stéatose (n=29) sont donnés par une variance expliquée,  $R^2$  de 0.75 et pour une prédictibilité du modèle,  $Q^2$  de 0.48 ainsi qu'une CV-AUROC de 0.93. Le modèle OPLS-DA discriminant foies normaux (n=15) et NASH (n=26) a été obtenu avec 4 composantes et avec des paramètres de qualité pour  $R^2$  de 0.96,  $Q^2$  de 0.7 et CV-AUROC de 1. On peut constater sur les loading plots des deux modèles représentés en figure 30 une augmentation générale des lipides pour le groupe stéatose (Fig. 30A) et pour le groupe NASH (Fig. 30B) comparé au groupe foie normal.



**Fig. 30 :** Loading plots de l'analyse OPLS-DA obtenue à partir des fractions organiques des échantillons (A) de foies normaux et stéatoses ; (B) de foies normaux et NASH. L1= groupement lipidique  $\text{CH}_3\text{-CH}_2\text{-}$  (méthyle) ; L2= groupement lipidique :  $\text{-(CH}_2\text{)}_n\text{-}$  (méthylène) ; L3= groupement lipidique  $\text{-CH}_2\text{-CH}_2\text{CO-}$  (méthylène des L5) ; L4= groupement lipidique  $\text{-CH=CH-CH}_2\text{-CH}_2\text{-}$  (allylique) ; L5= groupement lipidique  $\text{-CH}_2\text{-CH}_2\text{-CO-}$  ( $\alpha$ -méthylène du carbonyle) ; L6= groupement lipidique  $\text{-CH=CH-CH}_2\text{-CH=CH-}$  (diallylique – acides gras polyinsaturés « PUFA ») ; L7= groupement lipidique  $\text{-CH=CH-}$  (n-ene – acides gras monoinsaturés « MUFA ») ; group. glycéryl= groupement glycéryl  $\text{CH}_2\text{OR}$ .

Parmi les signaux des lipides, nous pouvons distinguer les signaux correspondant aux acides gras monoinsaturés (MUFA) et polyinsaturés (PUFA), tous deux en augmentation également lors de la stéatose et du NASH.

Dans la littérature, une diminution du niveau total des acides gras polyinsaturés a très souvent été rapportée chez des rats obèses [198], dans des modèles murins de NAFLD [199], ainsi que chez des patients obèses souffrant d'une stéatose [200]. Différentes études lipidomiques réalisées en spectrométrie de masse ont également montré des modifications dans la voie des acides gras polyinsaturés avec une diminution du taux total de ces PUFA après extraction du tissu hépatique chez des patients obèses, NAFLD et NASH comparés à des patients sains [201,202].

Johnson *et al.* [203] ont également rapporté par une étude *in vivo* en résonance magnétique une augmentation des acides gras saturés et une diminution des acides gras polyinsaturés chez des patients obèses avec ou sans stéatose comparés à des patients sains.

Une étude menée *in vitro* sur des biopsies de foie en  $^1\text{H}$  HR-MAS et *in vivo* en résonance magnétique sur des patients souffrant d'hépatopathie chronique du VHC a été entreprise pour évaluer la composition lipidique hépatique des différents grades de la fibrose [204]. Cette étude rapporte une augmentation du pourcentage total des lipides ainsi que des acides gras saturés et une diminution du pourcentage des acides gras polyinsaturés avec l'intensité de la fibrose hépatique.

Les acides gras polyinsaturés avec les phospholipides jouent un rôle important dans la construction et le maintien des membranes cellulaires [205]. Il est possible que le processus d'extraction, par l'intermédiaire des méthodes drastiques de précipitation des protéines, puisse jouer sur la composition des lipides membranaires et par conséquent influencer le taux de PUFA obtenu dans l'analyse métabolomique [206].

#### IV Conclusion

La prévalence grandissante des atteintes dites « NAFLD » et le manque de biomarqueurs sensibles et spécifiques pour les diagnostiquer constituent actuellement un véritable enjeu de santé publique. Allant de la simple stéatose à une condition inflammatoire appelée NASH, cette atteinte peut également progresser vers une forme plus sévère et irréversible d'hépatopathie telle que la cirrhose et le carcinome hépatocellulaire. C'est pourquoi la distinction précoce des différents stades de cette atteinte est devenue indispensable pour la prise en charge. La biopsie hépatique reste l'outil diagnostique le plus fiable. Cependant cette pratique invasive ne peut être utilisée en routine dans le dépistage de cette atteinte. Aujourd'hui il existe un réel besoin de déterminer de nouveaux biomarqueurs qui soient non invasifs et plus fiables pour le diagnostic des différentes atteintes NAFLD.

Au cours de cette étude, nous avons exploré les différentes voies métaboliques qui peuvent être altérées lors d'une atteinte NAFLD et NASH par une analyse complémentaire des extractions des métabolites hydrosolubles et liposolubles du foie. Les différentes atteintes n'ont pas pu être discriminées par une analyse OPLS-DA de la fraction aqueuse. Néanmoins certains biomarqueurs, tels que la créatine, se sont révélés être significativement diminués avec le degré de l'atteinte hépatique. D'autre part, l'analyse OPLS-DA de la fraction organique a montré une discrimination significative entre le foie normal, la stéatose et le NASH avec une augmentation générale des lipides lors de l'atteinte hépatique. Si cette augmentation des lipides lors du NAFLD est largement connue et décrite dans la littérature, la composition lipidique reste difficile à déterminer.

Différents algorithmes basés notamment sur les examens histologiques sont aujourd'hui utilisés pour la détermination du stade de l'atteinte NAFLD [119,129,189,207]. Néanmoins, le panel des différents algorithmes existants témoigne de la difficulté d'établir une classification de cette atteinte. Une exploration des différents paramètres cliniques et histologiques mériterait d'être investie plus en détails dans cette étude pour une stratification des patients peut être plus appropriée qui permettrait également une meilleure prédiction de l'atteinte NASH. Différents facteurs tels que les taux de TG, ASAT, ALAT, cholestérol, insuline et glycémie à jeun doivent par conséquent encore être testés dans des modèles de type OPLS-DA.

Une autre perspective de ce travail est la comparaison des données obtenues par cette méthode d'extraction avec la technique spectroscopique HR-MAS qui permet l'analyse du tissu intact. L'application de cette technique a déjà été étudiée en comparaison avec les méthodes d'extraction et a montré une similarité pour la plupart des métabolites observés avec ces deux techniques [208]. Elle a également été appliquée dans d'autres études de maladies chroniques du foie [206]. La spectroscopie HR-MAS offre l'avantage d'avoir une préparation quasi nulle des échantillons diminuant fortement le risque d'erreurs expérimentales et d'introduction de variabilité liée à la manipulation. De plus, l'analyse de tissu intact permet de conserver un certain degré d'intégrité du tissu et permettrait par conséquent de réaliser une étude relativement non perturbée des interactions moléculaires dynamiques. Par conséquent, la technique HR-MAS apparaît être très prometteuse pour une meilleure compréhension du métabolisme local et des voies biochimiques qui sont impliquées lors des atteintes NAFLD.

Ces deux approches d'analyse (HR-MAS et méthodes d'extractions) du tissu représentent une analyse plus ciblée du dysfonctionnement direct du métabolisme de l'organe atteint mais sont également complémentaires des analyses plus classiques réalisées sur les liquides biologiques. Ainsi, l'analyse métabolomique réalisée sur les échantillons de sérum de ces patients contrôles, NAFLD et NASH constituera également une approche complémentaire à cette étude.

## **Chapitre II Etude métabolomique du carcinome hépatocellulaire précoce : approche pronostique et suivi de traitement par ablation par radiofréquence (Article soumis).**

Le carcinome hépatocellulaire (CHC) est une des complications les plus fréquentes de la cirrhose dont les causes majeures d'apparition sont les maladies chroniques du foie telles que les infections chroniques par le VHB et VHC, la consommation excessive d'alcool ainsi que la stéato-hépatite (NASH). Le développement rapide du CHC ainsi que son dépistage tardif sont le résultat d'un pronostic peu optimiste avec un taux de survie à 5 ans ne dépassant pas 5% [9,209]. Afin de diminuer l'incidence et la mortalité liée au CHC, la recherche dans les mécanismes d'hépatocarcinogénèse est primordiale. Le métabolisme anormal des cellules tumorales représente une caractéristique importante qui pourrait éclaircir non seulement la pathogénèse mais également fournir des outils thérapeutiques potentiels pour de futurs traitements curatifs. Le dysfonctionnement du métabolisme énergétique des cellules cancéreuses peut aussi modifier les autres voies métaboliques qui influencent de nombreux processus biologiques [210]. La modification du métabolisme constitue, par conséquent, un sujet phare dans la recherche du cancer et en particulier en métabolomique [27]. Néanmoins, si de nombreuses études métabolomiques ont été entreprises pour aider à une meilleure compréhension de l'apparition du CHC, très peu d'entre elles ont investigué les effets d'un suivi thérapeutique ainsi que l'apparition d'une récurrence qui a pourtant lieu pour plus de 50% des cas après traitement curatif [118]. A notre connaissance, seules deux études métabolomiques réalisées en spectrométrie de masse ont cherché à déterminer le profil urinaire [211] et plasmatique [212] prédictif de la récurrence du CHC après résection hépatique. La première étude entreprise sur 19 patients a montré une discrimination des urines prélevées 7 jours après l'intervention des patients qui ont eu une récurrence et des patients sans récurrence suivis sur une période de 21 mois. Elle a montré une perturbation dans le métabolisme des carbohydrates et nucléotides, du cycle de Krebs, de la glycine, sérine et thréonine, des acides aminés ramifiés et aromatiques, des glyoxylates et de la flore intestinale. La seconde a comparé le profil métabolique plasmatique de 22 patients ayant eu une récurrence précoce (< 2 ans) et de 18 patients ayant eu une récurrence tardive (> 2 ans). Des différences dans le métabolisme des acides aminés, acides biliaires, acides gras,

phospholipides, carbohydrates ainsi que dans le métabolisme du cholestérol ont été relevées entre ces deux groupes. Néanmoins, ces deux études ont été menées sur un nombre de patients relativement restreint présentant un stade tumoral variable suivant les sujets.

D'autre part, les hépatopathies chroniques menant au CHC peuvent être de différentes origines. Les deux causes cirrhotiques : la consommation excessive d'alcool et les infections chroniques des virus des hépatites B et C sont les causes d'apparition du CHC les plus fréquemment observées. Il est souvent énoncé que le CHC est associé aux maladies hépatiques indépendamment de la cause spécifique de la maladie (World Gastroenterology Organisation Global Guideline de novembre 2009). De plus, de manière générale, l'étiologie de la maladie hépatique ne paraît pas influencer de façon importante la discussion thérapeutique et le résultat des traitements [213]. Néanmoins, mêmes si elles restent controversées, certaines études ont montré une influence de l'étiologie de la maladie du foie au niveau de la présentation clinique des patients au moment du diagnostic et de la rapidité de l'évolution de la tumeur [214,215]. Il a notamment été montré que la fonction hépatique était, dans l'ensemble, moins détériorée chez les patients atteints de cirrhose virale que chez ceux atteints de cirrhose alcoolique [216]. De plus, en cas d'infection chronique par le VHC ou de consommation excessive d'alcool, le CHC survient presque toujours après un stade cirrhotique. Ce qui n'est pas le cas chez les porteurs chroniques du VHB, principalement en Asie et en Afrique suggérant un rôle direct du VHB dans la carcinogenèse hépatique [217].

Ces différentes étiologies impliquent, par conséquent, des mécanismes de carcinogenèse hépatique probablement différents qui sont encore aujourd'hui loin d'être complètement élucidés [218,219].

L'étiologie cirrhotique du CHC constitue ainsi un sujet d'étude intéressant qui n'a été que très peu approché dans les études métabolomiques.

Le but de cette étude a été :

- d'évaluer la réponse métabolique suite au traitement par ablation de radiofréquence sur le profil sérique grâce au suivi longitudinal des patients ;
- d'évaluer l'impact de l'étiologie sur le profil sérique du CHC.

Cent vingt patients présentant un CHC d'origine cirrhotique correspondant aux critères de Milan (nodule unique de diamètre < 5 cm ou 2 à 3 nodules de diamètre < 3 cm) et éligibles

pour un traitement curatif de type ablation par radiofréquence (RFA) ont été prospectivement inclus dans l'étude. Trois échantillons de sérum ont été collectés : le premier avant la RFA (t0), le suivant un jour après l'intervention (t1) et le dernier entre 1-4 mois suivant l'intervention (t2).

La survie à 2 ans sans récurrence de cette cohorte s'est avérée être différente suivant l'origine de la cause de la cirrhose. De plus, les analyses OPLS-DA inter- et intra-individuelles ont permis d'identifier un profil métabolique sérique différent suivant l'origine virale ou non des patients atteints du CHC. Les effets de la RFA sont également visibles dans le sérum et se traduisent par deux profils différents pour les deux cohortes virale et non virale. Chez les patients viraux, nous avons pu différencier les profils métaboliques sériques du CHC (à t0) des profils métaboliques sériques de leur rémission à 1-4 mois (t2). Une augmentation des sucres et une diminution des lipides ont été remarquées dans le profil de la rémission ainsi qu'une modification du taux de certains métabolites tels que le glycérol, la  $\tau$ -methylhistidine, le 3-hydroxybutyrate et la choline.

La récurrence de la tumeur n'a quant à elle, pas pu être prédite par une signature métabolique spécifique du sérum dans cette étude. Néanmoins, cette étude métabolomique a permis de démontrer que les modifications biochimiques observées lors du CHC étaient en partie dépendantes de l'étiologie du CHC et suivaient une réponse métabolique différente à la suite d'un traitement RFA.



**Sequential serum metabolomic profiling after radiofrequency ablation of hepatocellular carcinoma reveals different response patterns according to aetiology**

Corentine Goossens<sup>1</sup>, Pierre Nahon<sup>2,3</sup>, Laurence Le Moyec<sup>4</sup>, Mohamed Nawfal Triba<sup>1</sup>, Nadia Bouchemal<sup>1</sup>, Roland Amathieu<sup>1,5</sup>, Nathalie Ganne-Carrié<sup>2,3</sup>, Marianne Ziol<sup>6,7</sup>, Jean-Claude Trinchet<sup>2,3,7</sup>, Nicolas Sellier<sup>8</sup>, Abou Diallo<sup>9</sup>, Olivier Seror<sup>3,8</sup>, Philippe Savarin<sup>1</sup>

**Departments and institutions:**

<sup>1</sup> Université Paris 13, Sorbonne Paris Cité, CSPBAT, UMR 7244, CNRS, Bobigny, France

<sup>2</sup> Groupe hospitalier Paris Seine-Saint-Denis, pôle d'activités oncologiques spécialisées, APHP, Hôpital Jean Verdier, Bondy et Université Paris 13, Sorbonne Paris Cité, UFR SMBH, Bobigny, France

<sup>3</sup> INSERM U1162, Génomique Fonctionnelle des Tumeurs solides, Université Paris 5, Paris, France

<sup>4</sup> Université d'Evry Val d'Essonne, UBIAE, INSERM U902, Evry, France

<sup>5</sup> Service d'Anesthésie-Réanimation, GHU PSSD, Hôpital Jean Verdier, Bondy et Université Paris 13, Sorbonne Paris Cité, UFR SMBH, Bobigny, France

<sup>6</sup> APHP, service d'Anatomie Pathologique, Hôpital Jean Verdier, Bondy et Université Paris 13, Sorbonne Paris Cité, UFR SMBH, Bobigny, France

<sup>7</sup> BB-0033-00027. Centre de Ressources Biologiques Maladies du foie, Groupe hospitalier Paris-Seine-Saint-Denis, Bondy, France

<sup>8</sup> APHP, service de Radiologie, Hôpital Jean Verdier, Bondy, France

<sup>9</sup> Service d'Information Médicale, GHU PSSD, Hôpital Jean Verdier, Bondy, France

**Running title.** Aetiology influence on the metabolic profiling of hepatocellular carcinoma

**Keywords:** Liver cancer; RFA; NMR; non-invasive biomarkers; OPLS

**Financial support:** Supported by grants from SESAME Ile de France (SESAME NMR Biomolécules) and from Association Française pour l'Etude du Foie (AFEF).

**Number of figures and tables :** 7

The authors disclose no potential conflicts of interest.

## Abstract

Radiofrequency ablation (RFA) is commonly performed as a curative approach in patients with hepatocellular carcinoma (HCC). However, the risk of tumor recurrence is difficult to predict due to a lack of reliable clinical and biological markers and identification of new biomarkers poses a major challenge for improving prognoses. Metabolomics is a promising technique that may lead to the identification and characterization of new disease fingerprints. The objective of the present study was to explore, preoperatively and at various time points post-RFA, the metabolic profile of serum samples from HCC patients in order to identify factors associated with treatment response and recurrence.

Sequential sera obtained before and after RFA procedures for 120 patients with HCC due to cirrhosis were investigated using Nuclear Magnetic Resonance (NMR) metabolomics. A multilevel orthogonal projection to latent structure (OPLS) analysis was used to discriminate intra-individual metabolic changes in response to RFA treatment.

Recurrence-free survival differed depending on the underlying cause of cirrhosis. Two distinct serum metabolomic profiles were observed in the sera of HCC patients depending on whether their liver disease had a viral or non-viral aetiology. These profiles were also associated with specific and distinct metabolic responses after RFA.

As demonstrated by NMR metabolomics, timing of recurrence and biochemical adaptations to treatment depend upon the underlying aetiology associated with the development of HCC. This discovery suggests that future studies of HCC treatment response should be more investigated in order to implement new personalized therapeutic management plans.

## Introduction

Hepatocellular carcinoma (HCC) is the most common type of primary liver cancer and is mainly due to cirrhosis in Western countries (1). With the advent of diagnostic imaging techniques and medical surveillance of patients at risk, diagnosis at an earlier stage can be achieved and curative procedures can be implemented (2). Among the latter, radiofrequency ablation (RFA) has gained popularity and has been shown to be highly effective in the management of primary hepatic tumors. However, despite current progress in the use of this technique, the rate of overall recurrence is about 70% at 5 years and this remains one of the most challenging aspects of this field (3). Prognostication for HCC after curative treatment is difficult, in part due to the lack of useful biomarkers that would allow for the selection of patients at higher risk of tumor recurrence or enable accurate assessment of treatment response (4).

Metabolomics is recognized as a promising technique in the field of systems biology that allows the evaluation of global metabolic changes (5). Proton nuclear magnetic resonance ( $^1\text{H}$  NMR) spectroscopy-based metabolomics can be used to identify and quantify metabolites within a biological fluid in a single experiment, thus defining a metabolic profile, also called a metabolomic fingerprint (6). NMR-based metabolomics has been recently used in the field of liver disease for the detection of early biomarkers and altered metabolic pathways (7,8). NMR-based metabolomic studies of HCC have also been performed in sera, urine, and tissues to elucidate significant changes in tumor progression (9-11). Changes in serum metabolomic profiles in patients with advanced HCC reflect the distinct activation or impairment of specific biological pathways, in particular, energetic metabolism (12). Even more interestingly, inter- and intra-individual analysis of serum changes during follow-up of treated patients may provide essential information related to therapeutic responses and/or patterns of recurrence. Such improvements might lead to better stratification of patients and provide the data needed for implementation of personalized treatment algorithms.

In the present study, serum samples from patients with stage 0 or A HCC (classified by Barcelona Clinic Liver Cancer (BCLC) criteria) were investigated using NMR-metabolomics. The objective was to explore, preoperatively and at various time points post-RFA, the

metabolic profiles of patients with HCC to identify factors associated with disease recurrence and treatment response.

## **Patients and Methods**

**Ethics Statement.** The Institutional Review Board of Jean Verdier University Hospital approved the protocol and the French Research Delegation Office accepted the creation of a dedicated bio-collection for the patients who have been included in the protocol. The CNIL (National Informatics and Liberty Commission) also approved the creation of both bio-collection and database. Written informed consent was obtained directly from each patient at inclusion.

**Patient selection, Follow-up, and Collection of Serum Samples.** Between January 2002 and December 2012, 711 cirrhotic patients referred to our institution for the management of HCC by percutaneous ablation were considered. Among them, 349 underwent monopolar or multipolar RFA and had frozen serum samples available before and after the procedure. Serum samples were collected, stored and provided by the “centre de ressources biologiques maladie du foie”, Bondy, France (BB-0033-00027). The following inclusion criteria were considered: 1) biopsy-proven cirrhosis, 2) tumor stage 0 (one nodule smaller than 2 cm) or A (one nodule smaller than 5 cm or a maximum of three nodules smaller than 3cm), classified according to BCLC criteria, 3) availability for regular follow-up. HCC was diagnosed according to BCLC criteria: histological evidence or demonstration of a focal lesion more than 1 cm in size with arterial hyper-vascularization and wash-out at the portal phase by imaging techniques. The decision to treat patients with RFA was made by a multidisciplinary team including hepatologists, oncologists, liver surgeons, and interventional radiologists according to international recommendations (13). Clinical, biological, and radiological data for patients were recorded on the day of the first blood collection which was considered to be the inclusion date. The cause of the underlying liver disease was recorded, patients were considered to be in the “viral-HCC” group if they had hepatitis B virus (HBV) or hepatitis C virus (HCV)-related cirrhosis, and in the “non-viral” group if they had alcohol- and/or dysmetabolic-related cirrhosis. Response was assessed at one month after RFA, every three months for the next two years, and then every six months after that. Assessments included a three-phase contrast

computed tomography (CT) scan or magnetic resonance imaging (MRI) examination, and alpha-fetoprotein (AFP) serum level measurement. A tumor was considered as completely ablated if no nodular or irregular enhancement of the liver was visible at the arterial phase. Local or distant tumor recurrences were assessed. Local and overall tumor progression-free survival was computed.

Serum samples were collected at different times before and after RFA intervention and were immediately stored at  $-80^{\circ}\text{C}$  until NMR spectroscopic analysis. Sequential blood samples were selected: 1) within 14 days before RFA procedure (t0), 2) the day after the RFA procedure (t1) and 3) at the time of treatment response evaluation, ranging from one to four months after RFA (t2). At this time, liver examination using abdominal CT scan showed complete liver necrosis at the location of the tumor without other visible signs of nodules.

**Statistical Analysis of Clinical Data.** Patient status was specified at the date of the most recent follow-up visit before September 2013. Endpoint of interest was recurrence-free survival. Patients who underwent liver transplantation after RFA treatment were censored from the study. Survival was estimated using the Kaplan-Meier method. Univariate and multivariate Cox proportional hazards regressions were performed to determine the significant clinical and biological parameters for prediction of tumor recurrence-free survival. All statistical analyses were conducted using SAS software 9.4.1.

**$^1\text{H-NMR Spectroscopy}$ .** For NMR analysis, serum samples were thawed at room temperature and 450  $\mu\text{L}$  aliquots were mixed with 50  $\mu\text{L}$  of  $\text{D}_2\text{O}$ . The proton NMR spectra were acquired at  $20^{\circ}\text{C}$  on a Bruker AVANCE III 500 spectrometer. For all samples,  $^1\text{H}$  NMR spectra were recorded using two pulse sequences (noesygppr1D and a Carr-Purcell-Meiboom-Gill (CPMG) sequence (14)). The signal was acquired on 64K data points for a spectral window of 6000 Hz. For each sample, noesygppr1D experiments were collected with 64 transients and mixing time was 100 ms. The CPMG sequence is frequently used to suppress broad macromolecular signals according to their short T2 relaxation times. The number of transients was 128 and the applied spin-spin relaxation delay was 18 ms.

For resonance assignment, 2D total correlation spectroscopy (TOCSY) spectra were acquired (mixing time 80ms, 4K data points, 32 transients). J-resolved spectra were also acquired with 4K data points and 32 transients. Resonance assignments were performed based on the HMDB databank (15). Both types of 1D spectra were used to compute the models and those with the best statistical performance are presented here. All presented models were

obtained with CPMG spectra except for the one obtained with t0 samples discriminating viral from non-viral patient sera before RFA.

**Data Processing.** The spectra obtained were processed with NMRPipe software (16). The spectra were manually phased and a linear baseline correction was applied. The chemical shifts were referenced using the  $\alpha$ -glucose signal at 5.23 ppm. The spectral region between -1 and 10.5 ppm was divided into 11500 spectral regions of 0.001 ppm width, called buckets. The spectral region containing residual water (4.5-5.1 ppm) was excluded from the analysis. The spectra normalization algorithm employed was the quantile normalization (17). Prior to analysis, all remaining spectral buckets were mean-centered and scaled (UV scaling) (18).

**Multivariate Data Analysis of clinical data.** A principal component analysis (PCA) was performed on quantitative clinical variables. These variables were centered and reduced for this analysis. Results of the PCA analysis were presented in a score plot.

**Multivariate Data Analysis of NMR data.** A PCA was first performed to detect any group separation based on NMR signal variability. This unsupervised method also enabled detection and exclusion of any outliers, defined as observations located outside the 95% confidence region of the model. A supervised method, the orthogonal projection to latent structure-discriminant analysis (OPLS-DA) was used. Multilevel OPLS-DA was also applied as a paired analysis through the follow-up of one patient, as proposed by Westerhuis et al.(19). Briefly, the *between* subject variation which is described by the average of two observations from each subject (for example before and after treatment), can be separated from the *within* subject variation which is the difference between the two observations. This pairing provides the advantage of particularly discriminating the metabolic changes within individuals caused by the intervention and allows for evaluation of the large variability between human subjects. For each model, the number of considered patients was dependent on the availability of samples t0, t1, or t2 for each patient. PCA, OPLS-DA, and Multilevel OPLS-DA analyses were performed using Simca-P12 (Umetrics, Umea) and in-house Matlab 2012b (Mathworks, Natick, Massachusetts, USA) code based on the Trygg and Wold method (20). The quality parameters of the models, such as the explained variance ( $R^2Y$ ) and the predictability of the model ( $Q^2Y$ ) were calculated.  $Q^2Y$  was calculated by a 7-fold cross validation procedure and confirmed by exploring the impact of permutations in the dataset rows (21). The discriminatory power of the model was evaluated by calculating

the area under the receiver operating curve during the cross validation (CV-AUROC). Models were validated by permutation tests. The results are presented as a score-plot and a loading plot. Each point in the score-plot represents the projection of an NMR spectrum (and thus a sample from one patient) on the predictive (horizontal axis) and the first orthogonal component of the model (vertical axis). The loading plot represents the covariance between the Y-response matrix and the signal intensity of the various spectral domains. Colors were also used in the loading plot depending on the correlation between the corresponding bucket intensity and Y variable.

## Results

**Baseline characteristics of patients.** To investigate the metabolic changes in serum between baseline and post-RFA follow-up, 120 patients fulfilling inclusion criteria were selected for this study. Patients characteristics are shown in table 1 for the whole cohort and stratified according to the cause of underlying cirrhosis (viral and non-viral). Among viral-HCC patients, 11/59 (18.6%) had negative viral replication at time of HCC diagnosis following or during the course on an antiviral treatment. There were significant differences in baseline parameters between the two subgroups for cholesterol and transaminases levels (aspartate aminotransferase (AST) and alanine aminotransferase (ALT)). Moreover, a PCA analysis showed a tendency in the differentiation of these two sub-groups on the first and third component (supplementary figure 1), demonstrating that a major part of the variance is related to aetiology. Sequential blood samples pre- and post-RFA were collected on the whole cohort and included a set of 273 sera. The flow chart of patients and available serum samples is presented in figure 1. The data were first analyzed by PCA, leading to the exclusion of 4 spectra because of poor recording. The PCA score is shown in supplementary figure 2. A representative  $^1\text{H}$  NMR spectrum (CPMG) from the serum of one HCC patient at (a) t2 (from one to four months post-ablation time), (b) t1 (one day post-ablation time) and (c) t0 (pre-ablation period) is presented in supplementary figure 3.

**HCC recurrence-free survival: univariate and multivariate analyses.** After a median follow-up time of 2.2 years, 61 (50.8%) patients experienced HCC recurrence (local=8, distant=53, figure 2A). When stratified according to the cause of cirrhosis (figure 2B),

patients with viral-HCC had a higher probability of HCC recurrence (HR=1.9, Log-Rank=0.02, supplementary table S1). According to multivariate Cox model, this feature remained an independent predictor of HCC recurrence when considering the whole population, as did higher gamma-glutamyl transferase (GGT) activity (HR=2.6, P=0.003).

**Metabolic profile comparisons related to HCC aetiology.** In order to provide a more in-depth comprehension analysis of the observed difference in the recurrence-free survival between the two cohorts, metabolomic profiles of non-viral and viral sera were compared. An OPLS-DA model was performed on the t0 sera of the whole cohort (118 t0 samples available). The score plot of the 118 spectra (figure 3A) revealed a clear separation between non-viral and viral HCC patients ( $Q^2Y=0.34$ ). Some metabolite variations that differed between the two HCC aetiology subgroups could be identified in the serum spectra according to the loading plot (figure 3B). This model specifically reflects a different metabolic profile in the serum of non-viral versus viral-related HCC patients.

**Metabolic profiles resulting from the RFA intervention.** To further explore and compare the metabolic changes that occurred as a result of RFA treatment for these two cohorts, metabolites in sera collected before and after RFA intervention were studied. The metabolic signature of the intervention was determined by comparison of the two OPLS-DA models for the two subgroups: one comparing serum at t0 and at t1, and the other comparing serum at t2 and t1. The comparison of these two OPLS-models allows to accurately assess the proper effects of the RFA intervention. Considering the high levels of inter-individual variability observed in the cohort, paired samples were used in multilevel-OPLS-DA models (19). Consequently, four OPLS-DA models were obtained and their loading plots are presented in figure 4 (score plots are shown in supplementary figure S4). The t0 vs. t1 OPLS-DA models in figure 4A and 4B showed good predictability ( $Q^2Y=0.38$  for the 27 patients in the non-viral cohort, and 0.62 for the 34 patients in the viral cohort). For t2 vs. t1 models, the predictability was 0.4 for the non-viral cohort (figure 4C, 17 patients) and 0.58 for the viral cohort (figure 4D, 27 patients). The four loading plots show that the RFA intervention is correlated with an increase of lactate (5), glutamine (10), and 3-phenylpropionate (23) and a decrease of isoleucine (2), proline (7), phosphatidylcholine (PC), and glycerophosphocholine (GPC) (18). The comparison of loading plots for both sub-groups (comparing column 1 and column 2 in figure 4) allowed visualization of a different profile for viral and non-viral patients and determination of some metabolites which vary differently

post-RFA intervention depending on the aetiology of cirrhosis (lipids (1), aspartate (14), choline (17), glucose (20)). However, the loading plot in figure 4D demonstrates different behavior for glucose and lipids when compared to the three other loading plots. These differences seem to indicate that the t2 metabolic state in the viral cohort could be different from the one observed at t0. In contrast, profiles for the non-viral cohort (loading plots 4A and 4C), appear similar. All statically relevant metabolites ( $p$  value < 0.05) are reported in table 2.

***Discrimination of t0 from t2 sera for viral-related HCC patients.*** Profiles of the pre-RFA sera at t0 and the sera collected after one to four months post-ablation (t2) were compared in order to evaluate the evolution of the metabolic pathways post RFA. Interestingly, no significant OPLS-DA model discriminating t0 and t2 could be computed in the non-viral cohort. However, for viral patients, the t0 and t2 serum spectra could be discriminated with OPLS-DA analysis (42 patients). Figures 5A and 5B present the paired samples score and loading plots. The model demonstrated good predictability of 0.37. Some discriminating metabolites could be identified and are summarized in figure 5C. The t2 state was mainly characterized by an increase of glucose (20), glycerol (22),  $\tau$ -methylhistidine (16), and a decrease of lipids (1), 3-hydroxybutyrate (4), and choline (17) when compared to the t0 state.

***Metabolomic prediction of HCC recurrence.*** As these patients were prospectively followed-up, the recurrence of the tumor after t2 was tested as a driving factor for OPLS-DA. Several models were tested according to the time or aetiology but none of them was statistically relevant.

## Discussion

Metabolomics technology offers the ability to associate changes in serum metabolic profiles with environmental, pathological, or treatment effects. Due to the fact that there is a pressing need for prognostic biomarkers that can stratify patients according to their risk of HCC recurrence, we used metabolomics technology to study potential differences in metabolic profiles in the sera of HCC patients before and after RFA treatment. In the present study, inter- and intra-individual analysis of serum changes during follow-up of RFA-treated

patients highlighted aetiological factors related to therapeutic response. Three strengths characterized the present study: i) selection of an homogeneous cohort of 120 patients with HCC developed in cirrhotic liver and classified according to the cause of liver disease, ii) longitudinal assessment of the serum metabolomic changes at various pivotal points of follow-up, iii) application of statistical analyses exploiting the paired data structure in cross-over multivariate data (19). It was essential to explore and apply the OPLS derivative methods such as paired data structures because of the large variability between individuals that could potentially mask small and subtle effects linked to disease aetiology or treatment.

The present study shows that two-year recurrence-free survival for this prospective cohort differed according to the cause of cirrhosis, a parameter that is usually not taken into account in clinical practice. The effect of viral infection on clinico-pathological features and long-term outcome in HCC-treated patients has been the subject of several studies (22). Zhou et al. reported that patients with HBV or HCV infection had a worse 5-year disease-free survival when compared to non-viral HCC patients. They also noted no difference in the 5-year survival between the HCV and HBV-positive HCC patients (23). In the present cohort, in both subgroups, recurrence was associated with a higher baseline level of serum GGT, independently of the cause of liver disease. As all patients had compensated cirrhosis with small HCC, this biological feature might reflect, at least partly, the degree of insulin resistance related to comorbidities known to be frequent in these populations. This simple and non-specific biological feature suggests that the impairment of glucose and lipid metabolism in liver cancer recurrence after curative treatment could be targeted by adjuvant therapies.

Based on these simple clinical observations, serum metabolomic profiling of patients was able to provide a more in-depth comprehensive analysis at the metabolic level. When stratified according to the cause of cirrhosis, patients presented specific metabolomic profiles at baseline, as already demonstrated by Qi et al. (24). High throughput analyses based on genomics and/or transcriptomics of liver tumors highlighted differences related to aetiology (25). Although metabolic pathways are greatly interlinked and interpretation of the meaning of changes in a particular metabolite is complicated, it appears that the distinct metabolites responsible for differences observed in this prospective non-viral- and viral-HCC cohort revealed significant alterations in lipid metabolism (increase of lipids, N-acetyl glycoprotein, choline, and glycerol, and a decrease in GPC in the viral cohort compared to

the non-viral cohort). This feature is consistent with previous metabolic studies that demonstrated that HBV and HCV infection perturbed lipid classes such as free fatty acids, phospholipids, and sphingolipids in cultured hepatoma cells (26,27). In these experiments, an increase of choline and phosphatidylcholine was observed in HCV-infection.

The aetiological subgroups also exhibited differences in the evolution of their metabolomic profiles post-RFA. First, comparison of sera from different time points allowed identification of a specific metabolic signature associated with RFA treatment. The main metabolites related to the RFA intervention were metabolites associated with energy metabolism alterations, such as lactate and acetate, but also with amino acid metabolism, such as isoleucine and glutamine. The application of electrical current during RFA treatment causes burns in the liver and produces coagulative necrosis which results in parenchymal and tumor cell death (28). Indeed, numerous studies have reported that lactate is a major biomarker for tissue hypoxia and necrosis (29) and that an enhancement of consumption of branched amino acids (BCAA), such as isoleucine, may characterize the inflammatory response in liver, which could lead to further increases in BCAA catabolism and glutamine production, consistent with the findings obtained in this study (30-32).

However, through comparison of the profiles of RFA intervention effects, we noticed a profile quite different for the viral cohort. The metabolic state at t2 appeared to differ from the one at t0, contrary to what was observed in the non-viral cohort. Moreover, an OPLS-DA model discriminating t0 and t2 could be only computed for the viral cohort. This model supports once again the idea that there is an evolution in the metabolic state at t2 compared to the one at t0 for the viral cohort unlike the non-viral cohort. Major changes between t0 and the t2 profiles involved metabolites implicated in the lipid and glucose metabolic pathways. It has been reported by Bollard et al. (33) that liver regeneration requires phospholipids for cell membrane synthesis and very-low-density lipoprotein (VLDL) synthesis requires the availability of phospholipids, particularly PC. A possible explanation could be that the process of regeneration is underway and could affect lipid metabolism. Modifications in glucose concentration post-RFA could be linked to hyperglycaemia after liver resection, as reported by Dungan et al. (34). Consequently, for viral-related HCC patients, metabolic processes in the serum appear to follow different biological pathways than those for the non-viral-related HCC patients post-RFA.

Discrepancies in metabolomic pathways between non-viral- and viral-HCC were furthermore characterized by different patterns of evolution after RFA procedure. Recurrence-free survival was not identical in the two groups despite similar therapeutic managements. This metabolomics investigation demonstrated that biochemical adaptations to liver injury depend, at least in part, on the aetiology of the disease and this results in divergent biochemical responses to RFA treatment. Further metabolomic investigations focused on HCC and response to curative treatment should be undertaken in order to implement new paradigms for personalized therapeutic management.

## References

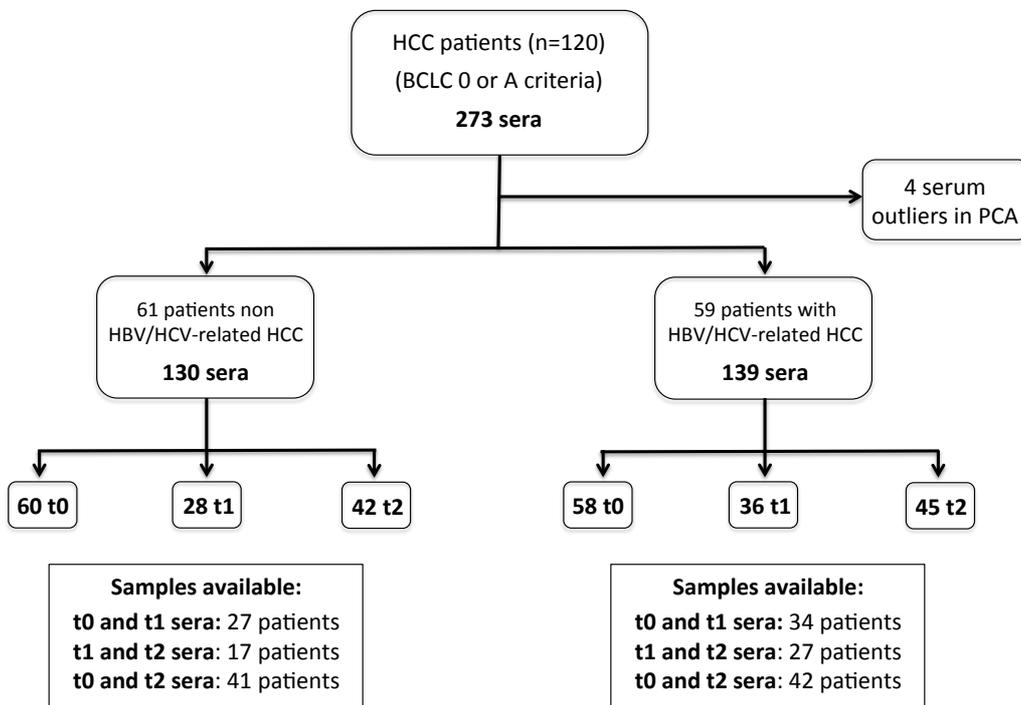
1. Daoudaki M, Fouzas I. Hepatocellular carcinoma. *Wien Med Wochenschr* 2014.
2. Cartier V, Aube C. Diagnosis of hepatocellular carcinoma. *Diagn Interv Imaging* 2014.
3. Brunello F, Cantamessa A, Gaia S, Carucci P, Rolle E, Castiglione A, et al. Radiofrequency ablation: technical and clinical long-term outcomes for single hepatocellular carcinoma up to 30 mm. *Eur J Gastroenterol Hepatol* 2013;25(7):842-9.
4. Breuhahn K, Gores G, Schirmacher P. Strategies for hepatocellular carcinoma therapy and diagnostics: lessons learned from high throughput and profiling approaches. *Hepatology* 2011;53(6):2112-21.
5. Nagrath D, Caneba C, Karedath T, Bellance N. Metabolomics for mitochondrial and cancer studies. *Biochim Biophys Acta* 2011;1807(6):650-63.
6. Dunn WB, Bailey NJ, Johnson HE. Measuring the metabolome: current analytical technologies. *Analyst* 2005;130(5):606-25.
7. Ranjan P, Gupta A, Kumar S, Gowda GA, Ranjan A, Sonker AA, et al. Detection of new amino acid markers of liver trauma by proton nuclear magnetic resonance spectroscopy. *Liver Int* 2006;26(6):703-7.
8. Amathieu R, Nahon P, Triba M, Bouchemal N, Trinchet JC, Beaugrand M, et al. Metabolomic approach by <sup>1</sup>H NMR spectroscopy of serum for the assessment of chronic liver failure in patients with cirrhosis. *J Proteome Res* 2011;10(7):3239-45.
9. Liu Y, Hong Z, Tan G, Dong X, Yang G, Zhao L, et al. NMR and LC/MS-based global metabolomics to identify serum biomarkers differentiating hepatocellular carcinoma from liver cirrhosis. *Int J Cancer* 2013.

10. Ladep NG, Dona AC, Lewis MR, Crossey MM, Lemoine M, Okeke E, et al. Discovery and validation of urinary metabotypes for the diagnosis of hepatocellular carcinoma in West Africans. *Hepatology* 2014;60(4):1291-301.
11. Yang Y, Li C, Nie X, Feng X, Chen W, Yue Y, et al. Metabonomic studies of human hepatocellular carcinoma using high-resolution magic-angle spinning <sup>1</sup>H NMR spectroscopy in conjunction with multivariate data analysis. *J Proteome Res* 2007;6(7):2605-14.
12. Nahon P, Amathieu R, Triba MN, Bouchemal N, Nault JC, Ziol M, et al. Identification of serum proton NMR metabolomic fingerprints associated with hepatocellular carcinoma in patients with alcoholic cirrhosis. *Clin Cancer Res* 2012;18(24):6714-22.
13. Bruix J, Sherman M. Management of hepatocellular carcinoma: an update. *Hepatology* 2011;53(3):1020-2.
14. Foxall PJ, Spraul M, Farrant RD, Lindon LC, Neild GH, Nicholson JK. 750 MHz <sup>1</sup>H-NMR spectroscopy of human blood plasma. *J Pharm Biomed Anal* 1993;11(4-5):267-76.
15. Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, et al. HMDB: the Human Metabolome Database. *Nucleic Acids Res* 2007;35(Database issue):D521-6.
16. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of biomolecular NMR* 1995;6(3):277-93.
17. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 2003;19(2):185-93.
18. Van den Berg RA, Hoefsloot HC, Westerhuis JA, Smilde AK, van der Werf MJ. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* 2006;7:142.
19. Westerhuis JA, van Velzen EJ, Hoefsloot HC, Smilde AK. Multivariate paired data analysis: multilevel PLSDA versus OPLSDA. *Metabolomics* 2009;6(1):119-28.
20. Trygg J, Wold S. Orthogonal projections to latent structures (O-PLS). *JChemometrics* 2002;16:119-28.
21. Triba MN, Le Moyec L, Amathieu R, Goossens C, Bouchemal N, Nahon P, et al. PLS/OPLS models in metabolomics: the impact of permutation of dataset rows on the K-fold cross-validation quality parameters. *Molecular bioSystems* 2014.
22. Tanase AM, Dumitrascu T, Dima S, Grigorie R, Marchio A, Pineau P, et al. Influence of hepatitis viruses on clinicopathological profiles and long-term outcome in patients undergoing surgery for hepatocellular carcinoma. *Hepatobiliary Pancreat Dis Int* 2014;13(2):162-72.

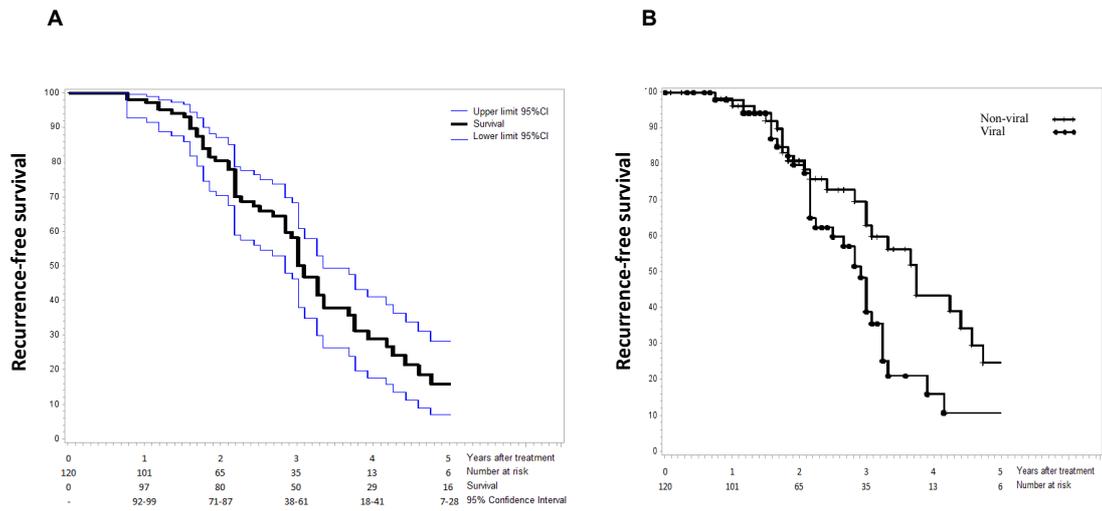
23. Zhou Y, Si X, Wu L, Su X, Li B, Zhang Z. Influence of viral hepatitis status on prognosis in patients undergoing hepatic resection for hepatocellular carcinoma: a meta-analysis of observational studies. *World J Surg Oncol* 2011;9:108.
24. Qi S, Tu Z, Ouyang X, Wang L, Peng W, Cai A, et al. Comparison of the metabolic profiling of hepatitis B virus-infected cirrhosis and alcoholic cirrhosis patients by using (1) H NMR-based metabonomics. *Hepatol Res*;42(7):677-85.
25. Nault JC. Pathogenesis of hepatocellular carcinoma according to aetiology. *Best Pract Res Clin Gastroenterol* 2014;28(5):937-47.
26. Negro F, Sanyal AJ. Hepatitis C virus, steatosis and lipid abnormalities: clinical and pathogenic data. *Liver Int* 2009;29 Suppl 2:26-37.
27. Roe B, Kensicki E, Mohny R, Hall WW. Metabolomic profile of hepatitis C virus-infected hepatocytes. *PLoS One* 2011;6(8):e23641.
28. Higgins H, Berger DL. RFA for liver tumors: does it really work? *Oncologist* 2006;11(7):801-8.
29. Bernal W, Donaldson N, Wyncoll D, Wendon J. Blood lactate as an early predictor of outcome in paracetamol-induced acute liver failure: a cohort study. *Lancet* 2002;359(9306):558-63.
30. Holecek M. Branched-chain amino acids and ammonia metabolism in liver disease: therapeutic implications. *Nutrition* 2013;29(10):1186-91.
31. Saxena V, Gupta A, Nagana Gowda GA, Saxena R, Yachha SK, Khetrpal CL. 1H NMR spectroscopy for the prediction of therapeutic outcome in patients with fulminant hepatic failure. *NMR Biomed* 2006;19(5):521-6.
32. Tripathi P, Bala L, Saxena R, Yachha SK, Roy R, Khetrpal CL. 1H NMR spectroscopic study of blood serum for the assessment of liver function in liver transplant patients. *J Gastrointest Liver Dis* 2009;18(3):329-36.
33. Bollard ME, Contel NR, Ebbels TM, Smith L, Beckonert O, Cantor GH, et al. NMR-based metabolic profiling identifies biomarkers of liver regeneration following partial hepatectomy in the rat. *J Proteome Res* 2010;9(1):59-69.
34. Dungan KM, Braithwaite SS, Preiser JC. Stress hyperglycaemia. *Lancet* 2009;373(9677):1798-807.

**Figures**

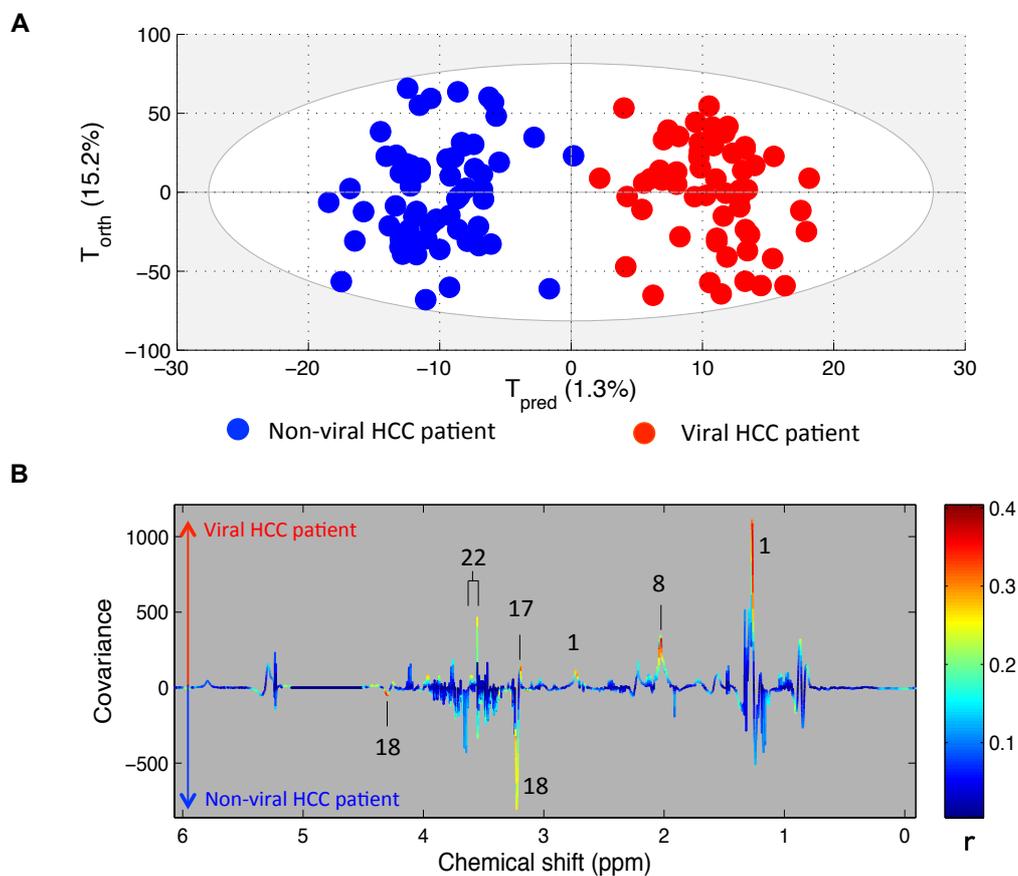
**Fig. 1. Flow chart of patients and serum samples.** t0 corresponds to serum pre-RFA, t1: serum taken one day post-RFA, t2: serum collected within a period from one to four months post-RFA. HCC, hepatocellular carcinoma; BCLC, Barcelona clinic liver cancer criteria; HCV, hepatitis C virus; HBV, hepatitis B virus ; PCA, principal component analysis.



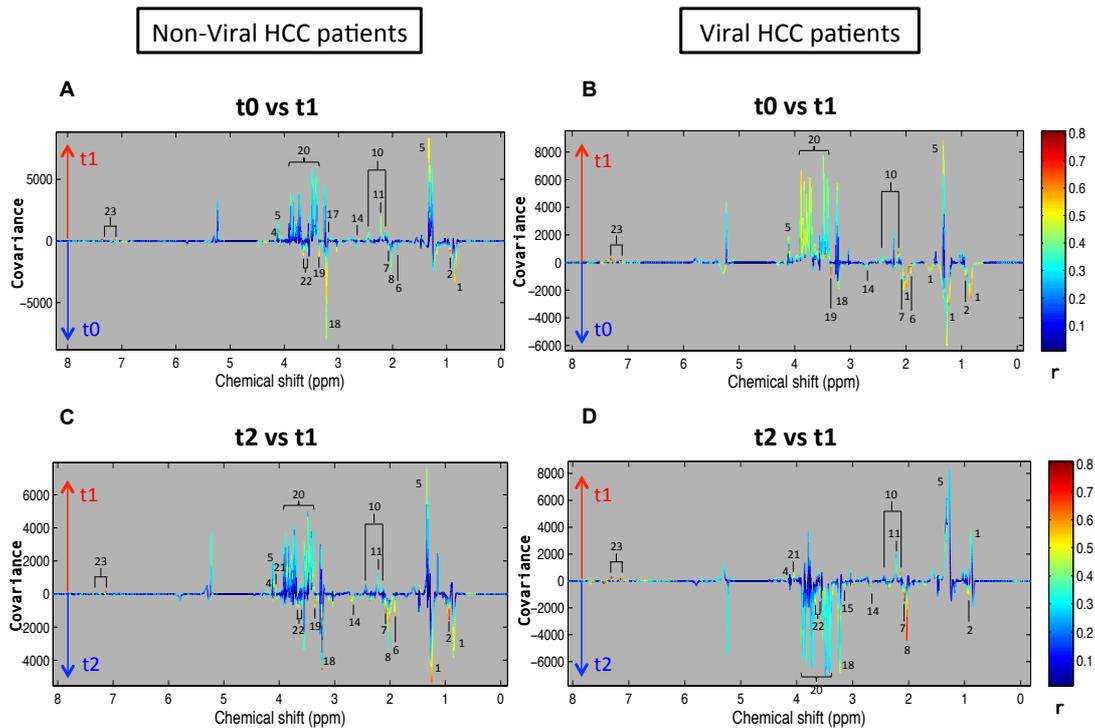
**Fig. 2. Recurrence-free survival after RFA in the whole cohort (A), and according to the cirrhosis aetiology (B).**



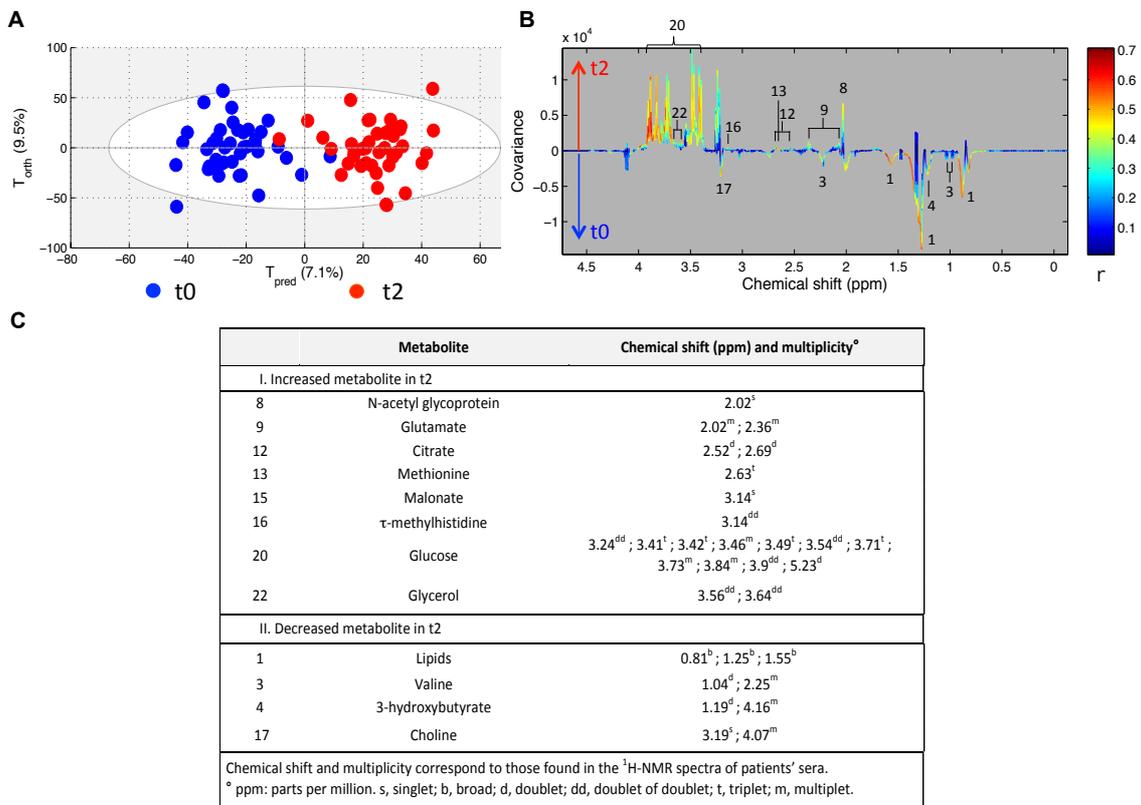
**Fig. 3. OPLS-DA score plot (A) and loading plot (B) obtained from viral HCC patients (red) and non-viral HCC patients (blue) at baseline (t0) (118 patients,  $Q^2Y=0.34$ ,  $R^2Y=0.89$ , 6 components, CV-AUROC=0.87).** Variations of metabolites are represented using a line plot between 0-6 ppm. Positive signals correspond to metabolites present at increased concentrations in the viral HCC sera. Conversely, negative signals correspond to metabolites present at increased concentrations in non-viral HCC sera. Metabolite 1= lipids, 8= N-acetyl glycoprotein, 17= choline, 18= glycerophosphocholine, 22= glycerol.



**Fig. 4. Multilevel OPLS-DA loading plots of the within-subject variation obtained from the sera of HCC patients pre-RFA (t0) and post-RFA intervention (t1) comparing non-viral HCC patients (A) (27 patients,  $Q^2Y=0.38$ ,  $R^2Y=0.99$ , 3 components, CV-AUROC=0.86) and viral HCC patients (B) (34 patients,  $Q^2Y=0.62$ ,  $R^2Y=0.99$ , 4 components, CV-AUROC=0.97). Panels C and D represent the within-subject variation obtained from the sera taken from one to four months after intervention (t2) and post-RFA intervention (t1) comparing non-viral HCC patients (C) (17 patients,  $Q^2Y=0.4$ ,  $R^2Y=0.98$ , 2 components, CV-AUROC=0.86) and viral HCC patients (D) (27 patients,  $Q^2Y=0.58$ ,  $R^2Y=0.9$ , 2 components, CV-AUROC=0.97). Variations of metabolites are identified in table 2.**



**Fig. 5. Multilevel OPLS-DA score plot (A) and loading plot (B) of the within-subject variation obtained from serum viral HCC patients before RFA, t0 (blue) and after 1-4 months post-RFA, t2 (red) (42 patients,  $Q^2Y=0.4$ ,  $R^2Y=0.6$ , 1 component, CV-AUROC=0.87). Variations of metabolites are represented using a line plot between 0-4.5 ppm and are summarized in the presented table (C).**



## Tables

**Table 1. Characteristics of HCC patients at the time of RFA procedure according to the cause of underlying cirrhosis.**

	Whole cohort (n=120)	Non viral-HCC (n=61, 50.9%)	Viral-HCC (n=59, 49.1%)	P-value#
<b>Age (years) *</b>	66 [59.5-73]	65 [60-73]	68 [57-75]	0.96
<b>Male gender †</b>	90 (75%)	46 (75.4%)	44 (74.6%)	0.92
<b>Uninodular HCC †</b>	87 (72.5%)	43 (70.5%)	44 (74.6%)	0.62
<b>Biopsy-proven HCC †</b>	71 (59.2%)	35 (57.4%)	36 (61%)	0.69
<b>Largest HCC nodule (mm)*</b>	24.5 [18-34]	20 [18-35]	25 [18-33]	0.64
<b>AFP (ng/mL) *</b>	6 [4-21]	6 [4-11]	6.5 [4-40]	0.42
<b>Triglycerides (g/L) *</b>	1.06 [0.7-1.6]	1.08 [0.7-1.4]	0.9 [0.7-1.4]	0.67
<b>Cholesterol (g/L) *</b>	4.3 [3.5-5]	4.4 [3.9-5.1]	3.6 [3.2-4.7]	0.005
<b>HDL (g/L) *</b>	1.3 [1.1-1.6]	1.4 [1.1-1.8]	1.3 [1.2-1.6]	0.4
<b>LDL (g/L) *</b>	2.3 [1.7-2.9]	2.3 [2-2.8]	1.8 [1.6-3.1]	0.47
<b>Glycaemia (mmol/L) *</b>	5.9 [5.1-8]	6.2 [5.2-8.2]	5.7 [5-7.6]	0.20
<b>CRP (g/L) *</b>	4 [4-7]	5 [4-8]	4 [4-5]	0.51
<b>AST (IU/L) *</b>	48 [36-69]	45 [33-61]	56 [37-101]	0.006
<b>ALT (IU/L) *</b>	36 [26-54]	32 [25-41]	44 [30-92]	<0.0001
<b>GGT (IU/L) *</b>	114 [63-182]	125 [68-204]	110 [47-168]	0.61

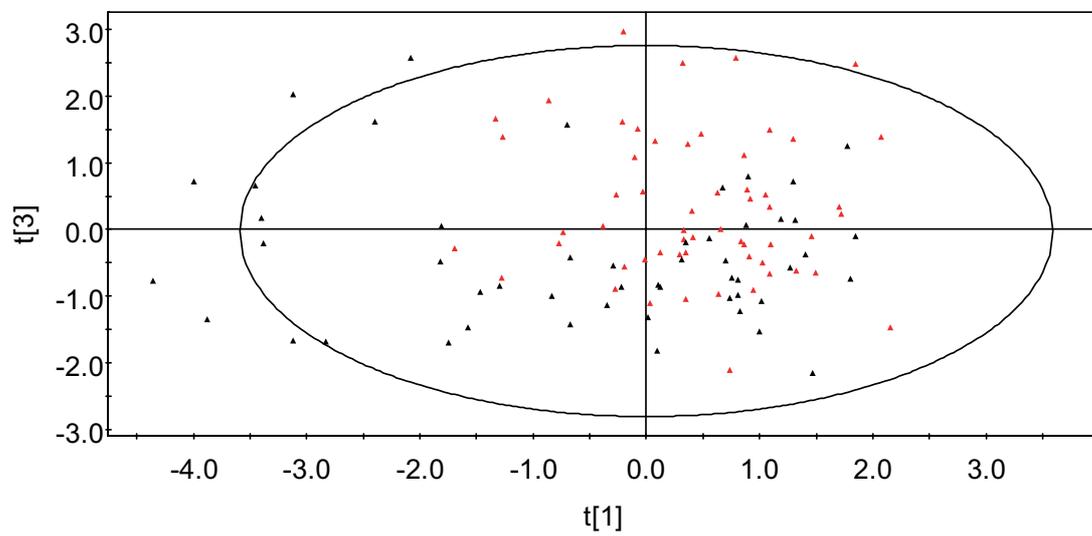
\* Data expressed as median [Q1-Q3]. † Number (percentage) of patients. AFP, alpha-fetoprotein; HDL, high density lipoprotein; LDL, low density lipoprotein; CRP, C-reactive protein; AST, aspartate aminotransferase; ALT, alanine aminotransferase; GGT, gamma-glutamyl transferase. # P-values were calculated by the student test considering the null hypothesis that there is no difference between the two groups of population.

**Table 2. Discriminant metabolites observed in sera of non-viral and viral HCC patients by  $^1\text{H}$  NMR spectroscopy, according to the loading plot in the pre-ablation period (t0) group and in the one-day post-ablation (t1) group and according to the loading plot in the one from four months post-ablation period (t2) group and in the one-day post-ablation (t1) group.** Metabolites affected by the RFA procedure correspond to peaks that simultaneously increased or simultaneously decreased for non-viral and viral cohorts.

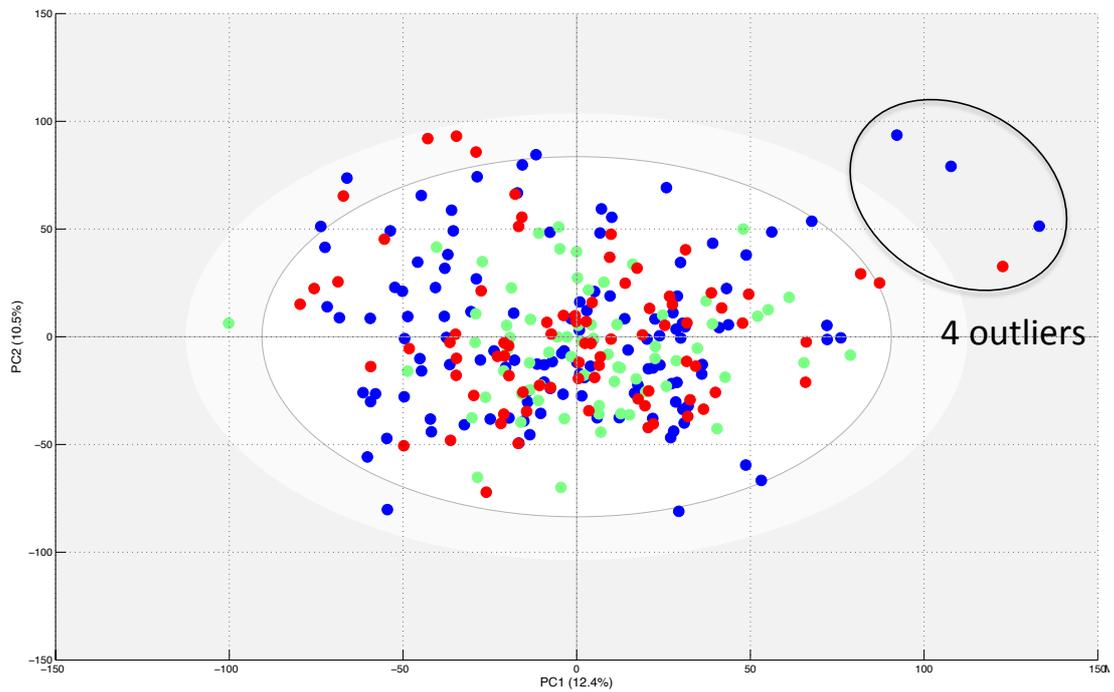
	Metabolite	Chemical shift (ppm) and multiplicity <sup>#</sup>	t0 vs t1		t2 vs t1	
			Non-Viral (A)	Viral (B)	Non-Viral (C)	Viral (D)
<b>I. Metabolites affected by the RFA intervention</b>						
2	Isoleucine	0.93 <sup>t</sup> ; 1.01 <sup>d</sup>	↓**	↓**	↓**	↓**
5	Lactate	1.32 <sup>d</sup> ; 4.11 <sup>q</sup>	↗**	↗**	↗**	↗**
7	Proline	1.95 <sup>m</sup> ; 2.07 <sup>m</sup> ; 3.35 <sup>m</sup>	↓*	↓**	↓**	↓**
10	Glutamine	2.15 <sup>m</sup> ; 2.44 <sup>m</sup>	↗**	↗**	↗*	↗**
18	PC/GPC	3.22 <sup>s</sup>	↓**	↓**	↓**	↓**
23	3-phenylpropionate	7.25 <sup>t</sup> ; 7.35 <sup>d</sup>	↗**	↗**	↗**	↗**
<b>II. Metabolites affected by the RFA procedure depending on cirrhosis aetiology</b>						
1	Lipids	0.83 <sup>b</sup> ; 1.23 <sup>b</sup> ; 1.57 <sup>b</sup>	↓**	↓**	↓**	↗*
14	Aspartate	2.69 <sup>dd</sup> ; 2.80 <sup>dd</sup>	↗*	↓**	↓**	↓**
17	Choline	3.19 <sup>s</sup>	↗**	↓*	-	-
20	Glucose	3.24 <sup>dd</sup> ; 3.41 <sup>t</sup> ; 3.42 <sup>t</sup> ; 3.46 <sup>m</sup> ; 3.49 <sup>t</sup> ; 3.54 <sup>dd</sup> ; 3.71 <sup>t</sup> ; 3.73 <sup>m</sup> ; 3.84 <sup>m</sup> ; 3.9 <sup>dd</sup> ; 5.23 <sup>d</sup>	↗**	↗**	↗**	↓*
<b>III. Metabolites affected by the RFA procedure and identified for one of the cohorts</b>						
4	3-hydroxybutyrate	1.19 <sup>d</sup> ; 4.16 <sup>m</sup>	↗**	-	↗*	↗**
6	Acetate	1.91 <sup>s</sup>	↓*	↓**	↓*	-
8	N-acetyl glycoprotein	2.02 <sup>s</sup>	↓*	-	↓**	↓**
11	Acetone	2.22 <sup>s</sup>	↗*	-	↗**	↗**
15	Malonate	3.14 <sup>s</sup>	-	-	-	↓**
19	Methanol	3.35 <sup>s</sup>	↓**	↓**	↓**	-
21	Myo-Inositol	3.27 <sup>t</sup> ; 4.06 <sup>t</sup>	-	-	↗**	↗**
22	Glycerol	3.56 <sup>dd</sup> ; 3.64 <sup>dd</sup>	↓**	-	↓**	↓**
<p>Chemical shift and multiplicity correspond to those found in the <sup>1</sup>H-NMR spectra of the patients' sera.  <sup>#</sup> ppm : parts per million. s, singlet; b, broad; d, doublet; dd, doublet of doublet; t, triplet; m, multiplet.                      PC/GPC : phosphatidylcholine/glycerophosphocholine                      An arrow pointing down corresponds to an increased metabolite concentration in the t0, pre-ablation group (for models A and B)/ in the t2, from two to four months post-radiofrequency ablation group (for models C and D) and an arrow pointing up to an increased metabolite concentration in the t1, one-day post-ablation group for all models.                      * P-value&lt;0.001                      ** P-value&lt;0.0001</p>						

## Supplementary data

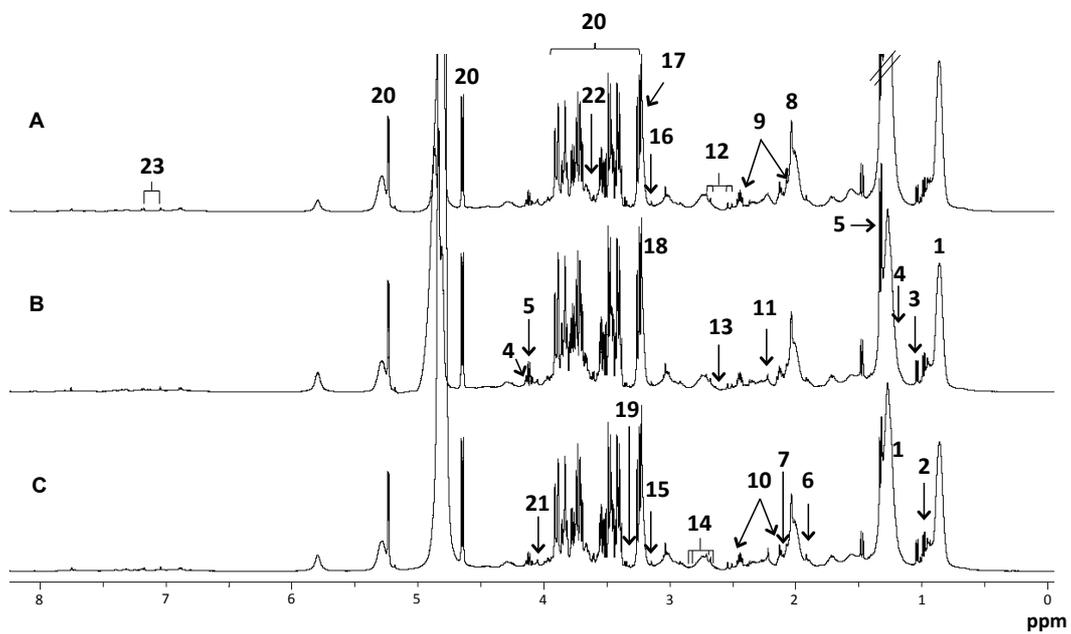
**Fig. S1. PCA score plot of 120 patients considering six biochemical parameters : glycaemia, gamma-glutamyl transferase, alpha-fetoprotein, cholesterol, aspartate aminotransferase and alanine aminotransferase. Black triangles correspond to viral HCC patients and red triangles to the non-viral HCC patients.**



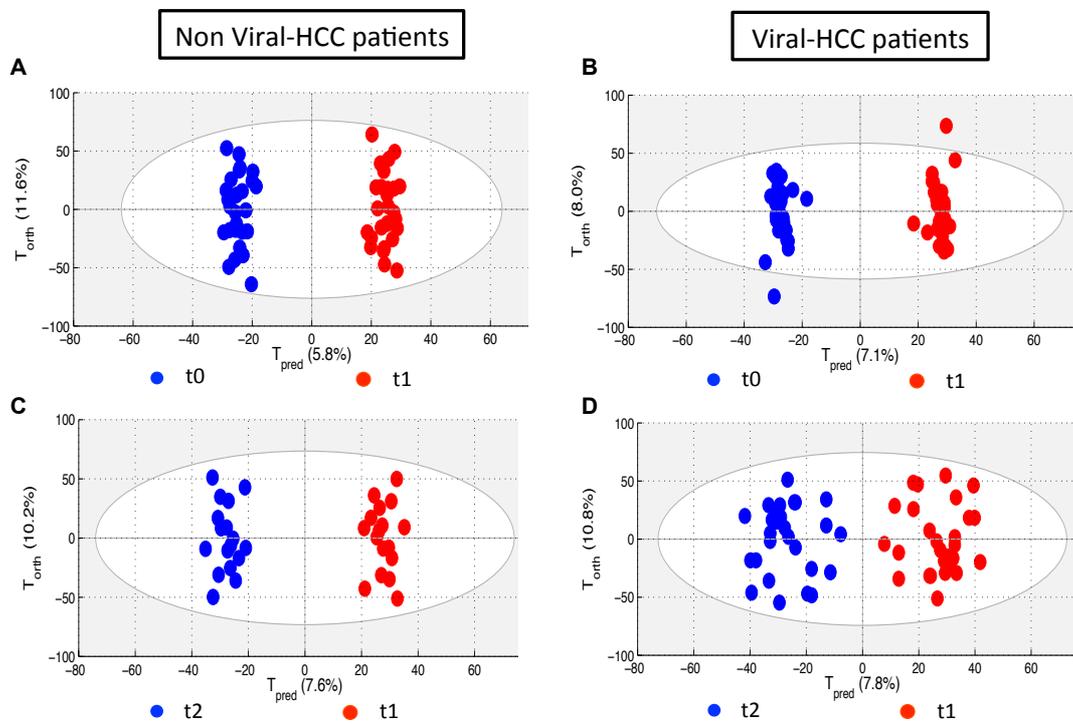
**Fig. S2. PCA score plot of the 273 samples obtained from the sera of HCC patients pre-RFA (blue), 1 day post-RFA intervention (green) and after 1-4 months post-RFA (red). Four samples were considering as outliers and excluded for the following analysis.**



**Fig. S3. Representative  $^1\text{H}$  CPMG spectra of serum obtained from one patient at t2 (A), t1 (B) and t0 (C).** 1: lipids, 2: isoleucine, 3: valine, 4: 3-hydroxybutyrate, 5: lactate, 6: acetate, 7: proline, 8: N-acetyl glycoprotein, 9: glutamate, 10: glutamine, 11: acetone, 12: citrate, 13: methionine, 14: aspartate, 15: malonate, 16:  $\tau$ -methylhistidine, 17: choline, 18: phosphatidylcholine/glycerophosphocholine, 19: methanol, 20: glucose, 21: myo-inositol, 22: glycerol, 23: 3-phenyl-propionate.



**Fig. S4. Multilevel OPLS-DA score plots of the within-subject variation obtained from the sera of HCC patients pre-RFA (t0) and post-RFA intervention (t1) comparing non-viral HCC patients (A) (27 patients) and viral HCC patients (B) (34 patients).** Panels C and D represent score plots of the within-subject variation obtained from the sera taken from one to four months after intervention (t2) and post-RFA intervention (t1) comparing non-viral HCC patients (C) (17 patients) and viral HCC patients (D) (27 patients).



**Table S1. Risk factors for HCC recurrence after RFA procedure.** HR, hazard ratio; AST, aspartate aminotransferase; ALT, alanine aminotransferase; GGT, gamma-glutamyl transferase.

	<b>Cox univariate analysis</b>	<b>Cox multivariate analysis</b>
<b>Viral-HCC</b>	HR=1.9 [1.1-3.4] <i>P</i> =0.02	HR=1.8 [1.01-3.4] <i>P</i> =0.04
<b>AST&gt;48 IU/mL</b>	HR=1.8 [1.03-3.3] <i>P</i> =0.03	—
<b>ALT&gt;36 IU/mL</b>	HR=2.1 [1.1-3.8] <i>P</i> =0.01	—
<b>GGT&gt;114 IU/mL</b>	HR=2.6 [1.4-4.9] <i>P</i> =0.002	HR=2.6 [1.4-4.8] <i>P</i> =0.003



### **Chapitre III Validation des modèles statistiques : impact d'une permutation des individus dans la matrice des données sur les paramètres de qualité du modèle (« Molecular BioSystems », octobre 2014 [220])**

L'étude précédente portant sur le carcinome hépatocellulaire précoce a présenté certaines caractéristiques particulièrement marquantes. Ainsi, la taille relativement importante de la cohorte de cette étude ainsi que les valeurs obtenues pour les paramètres de qualité de prédictibilité des modèles ( $Q^2$ ) en deçà du seuil limite (0.5) généralement admis dans ce type d'étude [107], nous ont mené à une réflexion plus approfondie sur la validation des résultats observés.

Par son large potentiel de découverte de nouveaux biomarqueurs, la métabolomique ouvre la porte à de nouvelles applications cliniques. Néanmoins en raison de la manipulation d'un grand jeu de données, il est essentiel de s'assurer de la validité et de la significativité des résultats obtenus afin de pouvoir envisager de les intégrer en milieu clinique.

Les approches d'analyses de données multivariées associées aux études métabolomiques offrent de nouvelles opportunités d'extraire des informations qui ne peuvent être obtenues lors d'analyses univariées classiques. Cependant ces techniques peuvent mener à de mauvaises interprétations et nécessitent par conséquent une validation rigoureuse des informations générées. La meilleure technique de validation reste incontestablement l'utilisation d'un jeu de données externe [105]. Néanmoins, cette série de validation est rarement disponible à cause d'un nombre limité d'échantillons nécessitant le recours à d'autres techniques de validations internes. La technique de validation croisée est une technique de validation largement utilisée, en particulier la k-fold validation [105]. Elle consiste à diviser le groupe d'échantillons en k groupes, à mettre l'un d'entre eux de côté, d'estimer le modèle sur les k-1 autres groupes restants et de calculer l'erreur sur chacune des observations qui n'ont pas participé à l'estimation. Les résultats de la validation croisée sont alors estimés par les valeurs des différents paramètres de qualité du modèle tels que  $R^2$  et  $Q^2$ .

Simca (v12.0, Umetrics, Umea, Suède) est un logiciel couramment utilisé pour réaliser les analyses multivariées en particulier en métabolomique. Il propose par défaut la technique

de k-fold validation. L'objectif de ce travail a été de tester la robustesse de cette technique de validation proposé par le logiciel Simca. De manière intéressante, nous avons pu montrer que les valeurs des paramètres de qualité d'un modèle pouvaient varier lors d'une permutation dans la classification des individus constituant les sous groupes de validation. Par un exemple concret, nous avons montré que cette simple permutation pouvait même aboutir à des conclusions contradictoires sur la validité du modèle. Une vigilance particulière doit par conséquent être accordée lors de la validation des résultats obtenus par cette technique de k-fold validation.

Ainsi, afin de s'assurer de la robustesse et de valider les valeurs définies pour ces paramètres, d'autres méthodes de validation croisée peuvent être utilisées telles que la double validation croisée (2CV) ou encore la méthode de Monte Carlo (MCCV).

CrossMark  
click for updatesCite this: *Mol. BioSyst.*, 2015,  
11, 13

## PLS/OPLS models in metabolomics: the impact of permutation of dataset rows on the K-fold cross-validation quality parameters†

Mohamed N. Triba,<sup>a</sup> Laurence Le Moyec,<sup>b</sup> Roland Amathieu,<sup>c</sup> Corentine Goossens,<sup>a</sup> Nadia Bouchemal,<sup>a</sup> Pierre Nahon,<sup>d</sup> Douglas N. Rutledge<sup>e</sup> and Philippe Savarin<sup>a</sup>

Among all the software packages available for discriminant analyses based on projection to latent structures (PLS-DA) or orthogonal projection to latent structures (OPLS-DA), SIMCA (Umetrics, Umeå Sweden) is the more widely used in the metabolomics field. SIMCA proposes many parameters or tests to assess the quality of the computed model (the number of significant components,  $R^2$ ,  $Q^2$ ,  $p_{CV-ANOVA}$ , and the permutation test). Significance thresholds for these parameters are strongly application-dependent. Concerning the  $Q^2$  parameter, a significance threshold of 0.5 is generally admitted. However, during the last few years, many PLS-DA/OPLS-DA models built using SIMCA have been published with  $Q^2$  values lower than 0.5. The purpose of this opinion note is to point out that, in some circumstances frequently encountered in metabolomics, the values of these parameters strongly depend on the individuals that constitute the validation subsets. As a result of the way in which the software selects members of the calibration and validation subsets, a simple permutation of dataset rows can, in several cases, lead to contradictory conclusions about the significance of the models when a K-fold cross-validation is used. We believe that, when  $Q^2$  values lower than 0.5 are obtained, SIMCA users should at least verify that the quality parameters are stable towards permutation of the rows in their dataset.

Received 16th July 2014,  
Accepted 23rd October 2014

DOI: 10.1039/c4mb00414k

www.rsc.org/molecularbiosystems

### Introduction

Projection to latent structures (PLS) and orthogonal projection to latent structures (OPLS) are popular methods for multivariate statistical analysis in metabolomics.<sup>1</sup> For classification or discrimination problems these methods are referred to as PLS-DA and OPLS-DA where DA stands for discriminant analysis. Many software packages are available for these types of analysis

in commercial or academic statistical packages: R Project for Statistical Computing (<http://www.r-project.org>),<sup>2</sup> MetaboAnalyst,<sup>3</sup> MVAPACK,<sup>4</sup> Multibase (Numerical Dynamics), IFRNOPLS-DA,<sup>5,6</sup> STATISTICA (StatSoft), Unscrambler (CAMO Software), SAS (SAS Institute Inc, Cary NC), the PLS-toolbox for Matlab (Eigenvector Research Inc, Wenatchee WA) and SIMCA (Umetrics, Umeå Sweden). However, in comparison to the other software packages, SIMCA seems to be much more often used in the metabolomics field (Table 1). Consequently, an appropriate use of this statistical package is necessary to assure the quality of the results published by the metabolomics community.

PLS/OPLS models try to find a linear relationship between a  $X$  predictor matrix (e.g. spectrometric data of biological samples) and a  $Y$  response matrix (e.g. clinical results, treatment...). In metabolomics, the  $X$  predictor matrix frequently has more columns (predictor variables) than rows (individuals). Because of this property of metabolomics data, PLS/OPLS models can easily be overfitted and their predictability overestimated.

The only way to reliably estimate the ability of the model to predict  $Y$  values of new individuals is to predict individuals from an independent dataset (*i.e.* that were not used to build this model). This can be achieved by splitting the dataset into a training set and a test set. The training set is used to build the model and the test set is used to estimate the predictability. However, the cost of this splitting is that the model is built with

<sup>a</sup> Université Paris 13, Sorbonne Paris Cité, Laboratoire Chimie, Structures, Propriétés de Biomatériaux et d'Agents Thérapeutiques (CSPBAT), Unité Mixte de Recherche (UMR) 7244, Centre National de Recherche Scientifique (CNRS), Equipe Spectroscopie des Biomolécules et des Milieux Biologiques (SBMB), 74 rue Marcel Cachin, 93037, Bobigny, France.  
E-mail: mohamed.triba@univ-paris13.fr

<sup>b</sup> Université d'Evry Val d'Essonne, Unité de Biologie Intégrative des Adaptations à l'Exercice (UBIAE), U902, INSERM, Bd François Mitterrand, 91025 Evry Cedex, France

<sup>c</sup> Service d'Anesthésie et des Réanimations Chirurgicales, Université Paris 12, Hôpital Henri Mondor, Assistance Publique des Hôpitaux de Paris (AP-HP), Créteil, France

<sup>d</sup> Service d'Hépatologie et Université Paris 13, Hôpital Jean Verdier, Assistance Publique des Hôpitaux de Paris (AP-HP), 93143 Bondy Cedex, France

<sup>e</sup> Laboratoire de Chimie Analytique, AgroParisTech, 16 rue Claude Bernard, 75231 Paris, France

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c4mb00414k

**Table 1** Estimation of the level of use of various software programs for OPLS/PLS analysis in metabolomic studies until June 2014

Bibliographic database	No. of articles containing metabolomics and PLS	Using Metabo analyst	Using SIMCA	Using R	Using Statistica	Using Unscrambler
Science direct	1117	51	464 (42%)	49	37	37
Royal Society of Chemistry	135	9	54 (40%)	5	3	2
PLOS One	245	13	108 (44%)	28	16	3
Springer link	654	32	274 (42%)	45	22	18
ACS	473	32	278 (59%)	30	14	22

only a fraction of the information that is present in the whole dataset. This may reduce the ability of this model to correctly predict a new dataset. Thus, splitting the dataset into a training set and a test set can be done only if enough individuals are available to build a reliable model. As in univariate statistics, the significance of the results of multivariate models depends on the sample size. However, the minimum number of individuals needed to attain a given significance threshold for the PLS models is very application-dependent and no easily applicable rules have been proposed to estimate this number.<sup>7</sup>

When no test set is available, the cross-validation method is the main strategy proposed by commercial or academic statistical packages to assess the quality of a model. Different cross-validation procedures exist. The default SIMCA cross-validation is the so-called K-fold cross-validation. Results of the cross-validation procedure are summarized by the value of different quality parameters. The most frequently mentioned in the metabolomics literature are  $R^2$  and  $Q^2$  parameters (also called cross-validated  $R^2$ ).  $R^2$  measures the goodness of fit while  $Q^2$  measures the predictive ability of the model.  $R^2 = 1$  indicates perfect description of the data by the model, whereas  $Q^2 = 1$  indicates perfect predictability.  $R^2$  increases monotonically with the number of components (NC) and will automatically approach 1 if NC approaches the rank of the  $X$  matrix.  $Q^2$  will not necessarily approach 1. At a certain value of NC,  $Q^2$  reaches a plateau and usually will finally decrease with addition of more components. This indicates that at a certain degree of complexity the predictive ability of the model decreases.<sup>8</sup> At this stage, it is very likely that the model is trying to fit dataset characteristics that are no longer representative of the studied population. A large discrepancy between  $R^2$  and  $Q^2$  indicates an overfitting of the model through the use of too many components. According to the SIMCA users' guide,  $Q^2 > 0.5$  is admitted for good predictability (SIMCA P12 users' guide, p. 514).<sup>9</sup> It has been shown that in practice it is difficult to give a general limit that corresponds to a good predictability since this strongly depends on the properties of the dataset.<sup>8,10</sup> For example, an acceptable  $Q^2$  threshold will strongly depend on the number of observations included. During the last few years, a large number of SIMCA PLS-DA/OPLS-DA models have been published with  $Q^2$  below 0.4 or even below 0.3 (for example, see ref. 11 and 12). These models with poor predictability are frequently validated by a permutation test that consists in comparing the  $Q^2$  obtained for the original dataset with the distribution of  $Q^2$  values calculated when original  $Y$  values are randomly assigned to the individuals.<sup>10</sup> The cross-validation procedure also provides the possibility to calculate a  $p$ -value to estimate the significance of PLS/OPLS models ( $p_{CV-ANOVA}$ ).<sup>13</sup>

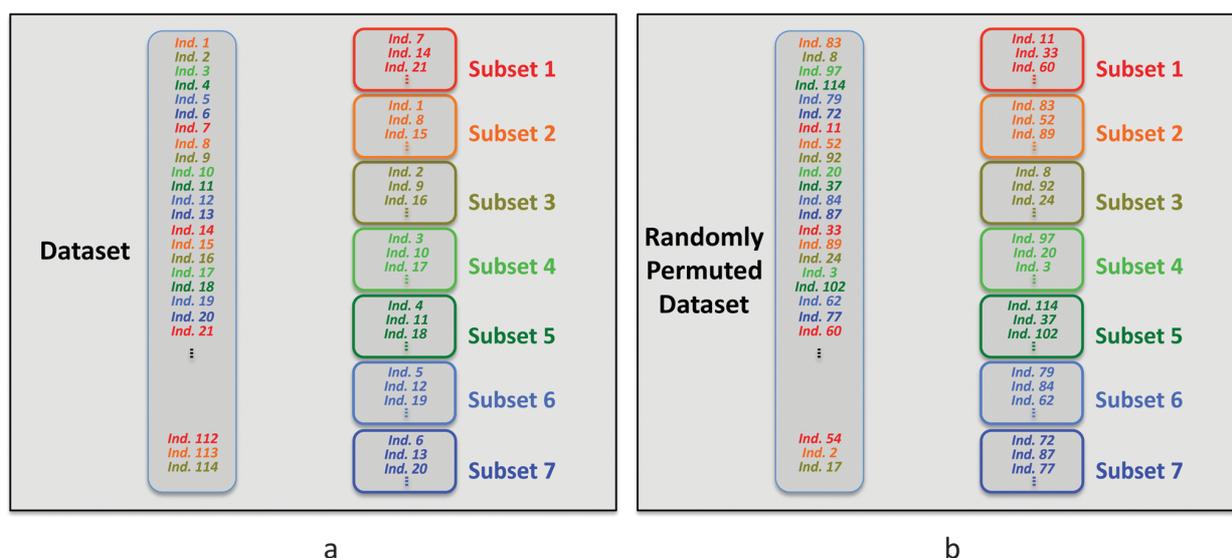
As recently published in this journal,<sup>14</sup> metabolomic results based on PLS/OPLS models should always give the values of the quality parameters of the multivariate models. The number of components used in the final model,  $Q^2$  and  $p_{CV-ANOVA}$  values should be presented to allow the reader to assess the quality of the model calculated by SIMCA. However, in this Opinion piece, we want to point out that in some cases, because of the way in which the default SIMCA cross-validation procedure selects members of the calibration and validation subsets, permutation of the rows of a dataset can result in variations in the values of the quality parameters. As a consequence, in these circumstances, different conclusions on the quality of the PLS/OPLS models may be drawn from the same dataset. In a first part, we will show that under some conditions a random permutation of rows in the dataset strongly affects the quality parameter values obtained when default SIMCA cross-validation settings are used. In a second part, we will discuss three different types of situations frequently encountered in metabolomics studies where the K-fold cross-validation procedure fails to calculate a  $Q^2$  that is not strongly dependent on the arbitrary order of the rows in a dataset.

## Default SIMCA cross-validation procedure

We give here a very basic description of the default SIMCA cross-validation procedure. Only the way the validation sets are built will be discussed in detail. For an exhaustive description of the procedure, the reader should refer to the Umetrics documentation.<sup>8</sup>

Cross-validation allows to estimate the ability of a model to correctly predict the  $Y$  response matrix of new individuals. In the SIMCA software, cross-validation is also used to avoid overfitting by estimating the number of significant components (NSCs) to use in the model. Many cross-validation procedures are used in the metabolomic community (K-fold, Leave One Out, Monte-Carlo, 2CV, *etc.*). The default SIMCA cross-validation procedure is a 7-fold cross-validation<sup>8</sup> where the dataset is split into 7 different subsets. For a fixed number of components (NC), the  $Y$  values of all individuals of each subset are predicted using a submodel built with the 6 other subsets (calibration subset). The differences between the predicted  $Y$  values and the observed  $Y$  values are used to calculate the  $Q_{NC}^2$  parameter for this number of components. The procedure starts at  $NC = 1$  and is repeated by incrementing NC as long as the increase of  $Q_{NC}^2$  is larger than a limit value fixed by various rules.<sup>9</sup>

Each subset is constituted by selecting one row every seven rows in the dataset. The first subset is built with the individuals



**Scheme 1** Selection of the individuals used to build the cross-validation subsets in SIMCA for the original dataset (a) and when the rows of this dataset are randomly permuted (b).

corresponding to rows 7, 14, 21 and so on. The second subset is constituted with the individuals corresponding to rows 1, 8, 15, ... The other subsets are built in the same way (Scheme 1a).

Considering the way the subsets are built, it is clear that a permutation in row order of the  $X$  and  $Y$  dataset changes the individual positions and modifies the composition of these subsets (Scheme 1b). Thus, submodels and predicted  $Y$  values calculated during the cross-validation procedure are also affected by a permutation of rows.

The major consequences of this are:

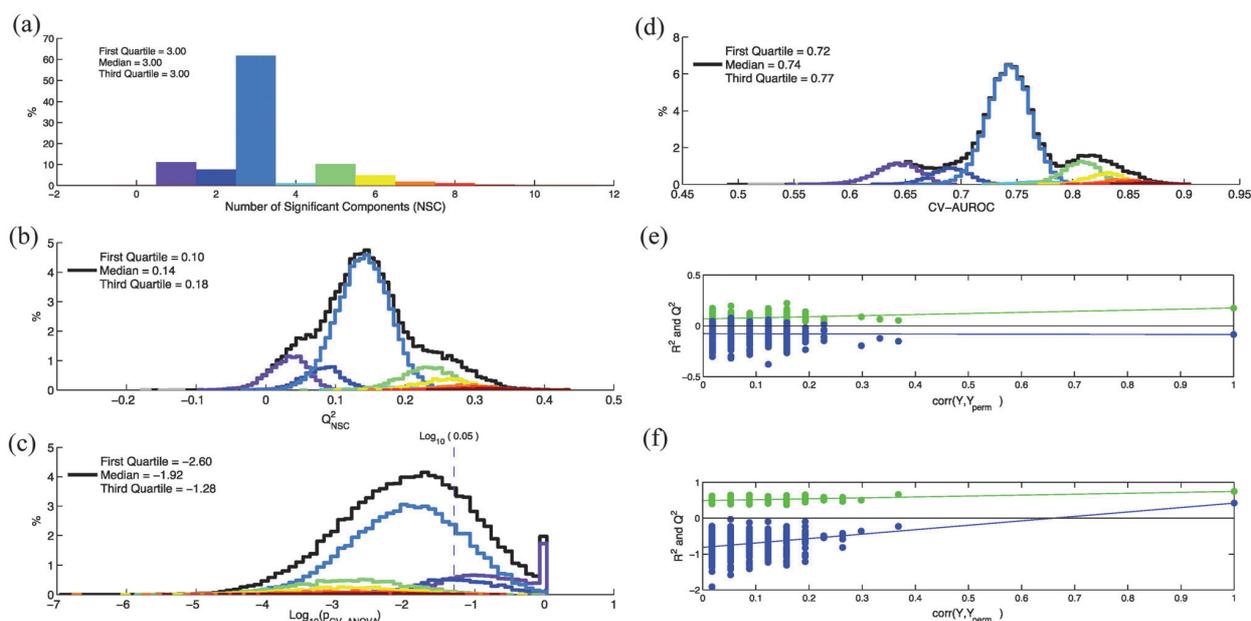
- Row permutations can potentially change the number of components considered as significant (NSC) by SIMCA.
- For the same number of significant components, row permutations will change the value of the  $Q_{NSC}^2$  parameter.
- The CV-ANOVA  $p$ -value, which depends on the cross-validation procedures, is also affected by row permutations in the dataset.
- The conclusion of the permutation test can be different when the order of rows is changed.

## Row permutation can strongly affect the K-fold cross-validation procedure

In this example we used serum NMR spectra of a metabolomics study on the influence of hepatitis viruses on patients with Hepatocellular Carcinoma (HCC).<sup>15</sup> OPLS-DA was used to discriminate 57 HCC patients with hepatitis infection from 57 HCC patients without viral infection. Spectra were normalized with the probabilistic quotient normalization method<sup>16</sup> to eliminate any dilution effect. They were divided into 230 domains of 0.05 ppm and the water signal region was suppressed. The resulting  $X$  predictor matrix was composed of 114 rows and 196 non-null columns.  $X$  and  $Y$  matrices are available as ESI† (Dataset1.xlsx). The  $X$  matrix was centered and scaled (UV scaling) prior to multivariate

analysis. OPLS-DA analyses were performed with an in-house Matlab 2012b (The Mathworks Inc., Natick, Massachusetts, USA) code based on the Trygg and Wold method.<sup>17</sup> The number of significance components is determined according to SIMCA rules (SIMCA P12 users' guide, p. 529).<sup>9</sup>  $X$  and  $Y$  dataset rows were randomly permuted in the same way and for each permutation various quality parameters (*i.e.* NSC,  $Q_{NSC}^2$ ,  $p_{CV-ANOVA}$ ) were calculated. In Fig. 1a, b and c are represented, respectively, the distributions of the NCS,  $Q_{NSC}^2$  and  $p_{CV-ANOVA}$  calculated for 50 000 random permutations. For each permutation, an area under ROC curve is calculated with  $Y$  values predicted during the cross-validation process (CV-AUROC).<sup>10</sup> The distribution of this parameter is shown in Fig. 1d. In Fig. 1b–d, the black lines represent the distributions of the quality parameters calculated for the 50 000 random permutations. In order to estimate the variability of the quality parameters for a given number of components, these distributions were decomposed according to the NSC value calculated for each permutation (colored lines). It clearly appears that an important part of the variability of these parameters is a consequence of NSC variability with row re-arrangements.

Thus, a better estimation of the number of components could help to reduce the variability of the quality parameters. According to Wheelock and Wheelock,<sup>14</sup> “the default automatic fitting in SIMCA extracts the maximal number of significant components, which in most cases results in an overfitted model”. These authors suggested that the optimal number of components (ONC) can be estimated by using the  $p_{CV-ANOVA}$  parameter: when this ONC is reached then the addition of another component would increase  $p_{CV-ANOVA}$ . As shown in Fig. 1c, values of  $p_{CV-ANOVA}$  can strongly depend on the arbitrary order of the rows in the dataset. This dependence is also observed for a given number of components (colored lines). As a consequence, an ONC based on  $p_{CV-ANOVA}$  may also strongly vary when the dataset lines are permuted. In order to



**Fig. 1** Variation of the OPLS quality parameters when 50 000 random permutations of the dataset lines were performed. For each permutation, values of NSC,  $Q_{\text{NSC}}^2$ ,  $p_{\text{CV-ANOVA}}$  and CV-AUROC were calculated (a): distribution of the number of significant components (NSC). Each color is associated to a specific value of NSC. (b)–(d): distributions of  $Q_{\text{NSC}}^2$ ,  $p_{\text{CV-ANOVA}}$  and CV-AUROC parameters calculated for the 50 000 random permutations (black lines). These distributions were decomposed according the NSC value calculated for each permutation (colored lines). (e) and (f): 500 random permutation tests (permutation of group membership) performed on the datasets that corresponded to the lowest and highest  $Q_{\text{NSC}}^2$  values among the 50 000 models. The vertical axis corresponds to  $R^2$  (green points) and  $Q^2$  (blue points) values of each model. The horizontal axis corresponds to the correlation coefficient between the original  $Y$  and the permuted  $Y$ .

estimate this variability, we performed 1000 permutations of the rows and, for each permutation, the value of ONC was determined by looking for the first local minimum of  $p_{\text{CV-ANOVA}}$  when NC is incremented. We found a large variability of the ONC with rows rearrangement (Fig. S1, Supplementary Data A, ESI†). Thus, an ONC determined by using  $p_{\text{CV-ANOVA}}$  can also strongly depend on the arbitrary order of the lines in the dataset if the K-fold cross-validation procedure is used. More generally, the number of components estimated by using parameters that depend on row order (such as  $Q^2$ ,  $p_{\text{CV-ANOVA}}$ , ...) can potentially exhibit a large variability with row permutations.

The datasets with the row arrangement corresponding to the lowest and the highest calculated values of  $Q_{\text{NSC}}^2$  (i.e.  $-0.09$  and  $0.42$ ) were compared by calculating the quality parameters of the OPLS models. These permuted datasets are available as ESI† (Dataset2.xlsx and Dataset3.xlsx). We observed that, for the same experimental result, quality parameters of the two models (Table 2) lead to contradictory conclusions on the significance of the

metabolic differences between the two classes. Contradictory conclusions are also obtained when permutation tests (random permutation of group affiliation) were performed on these two models (Fig. 1e and f). This particular dataset proves that, in some situations, quality parameter values calculated with the default SIMCA cross-validation procedure are strongly determined by chance. This result also suggests that performing row permutations allows an estimation of confidence intervals for the various quality parameters.

## In which situations?

Considering this result, an important question is now: in which situations do row permutations strongly affect the K-fold cross-validation procedure? We will not try to give an exhaustive or theoretical answer to this question but considering our practical experience in the PLS/OPLS analysis of metabolomic datasets, we will discuss some situations where row permutations can have a strong effect. According to Eriksson *et al.* “a necessary condition for PLS-DA to work reliably is that each class is tight and occupies a small and separate volume in  $X$ -space.”<sup>18</sup> This condition is fulfilled when interclass variability is large enough relative to the intraclass variability and is observed for example when classes can be discriminated by a simple principal component analysis (PCA). When this situation was observed in our experimental results, we noticed that row permutation in the dataset did not significantly affect the K-fold cross-validation procedure.

**Table 2** Quality parameters for the permuted datasets that corresponded to the lowest (Permuted dataset 1) and highest (Permuted dataset 2)  $Q^2$  values

	Permuted dataset 1	Permuted dataset 2
NSC	1	10
$R^2$	0.18	0.75
$Q^2$	$-0.09$	0.42
$p_{\text{CV-ANOVA}}$	1	0.00004
CV-AUROC	0.57	0.91

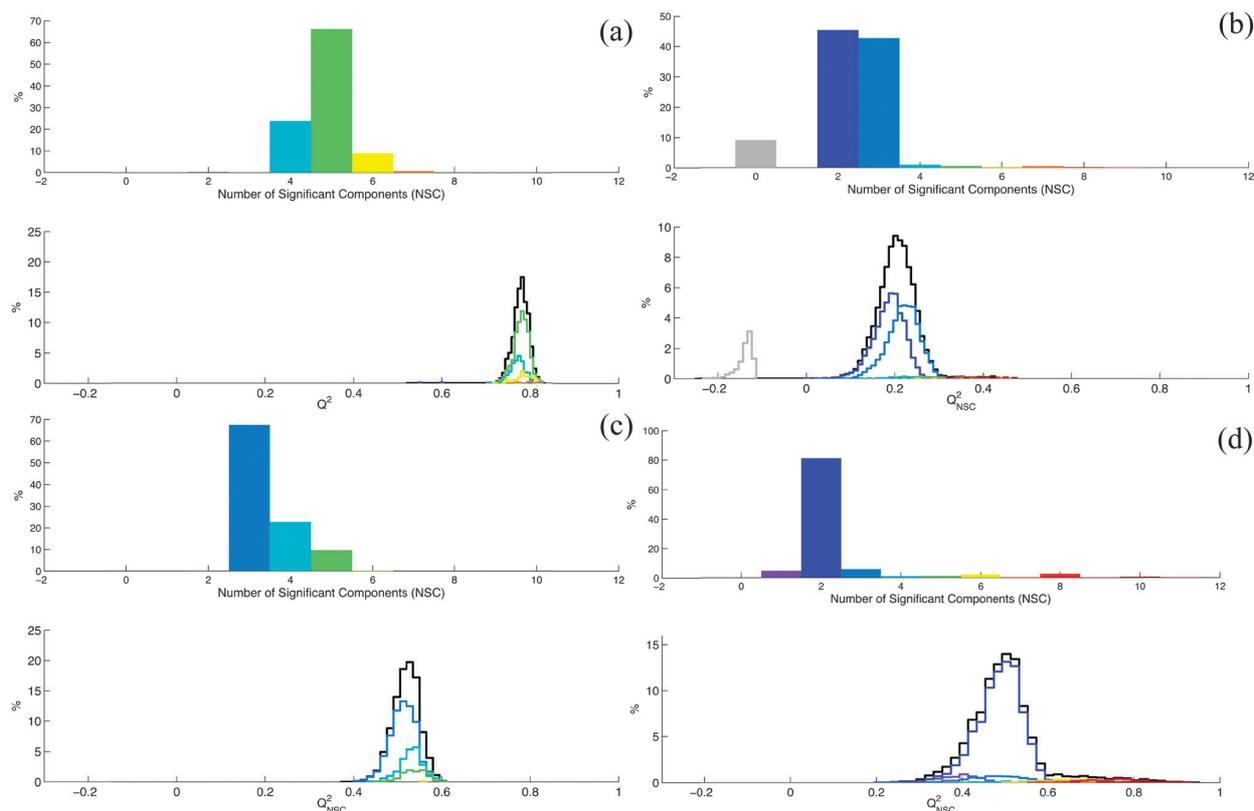
In particular, under these conditions no contradictory conclusions can be drawn from the same experimental dataset. However, Eriksson *et al.* also noticed that "... when some of the classes are not homogeneous and spread significantly in  $X$ -space, the discriminant analysis does not work".<sup>8</sup> According to our experience, when this situation is encountered, row permutations can have serious consequences on the conclusions of the K-fold cross-validation procedure. This is related to the fact that we can no longer neglect the probability to build by chance calibration and validation subsets that are not representative of the whole dataset in terms of difference between classes. These non-representative subsets can lead to an underestimation or an overestimation of the capability of the model to predict new data.

To illustrate this point, we modified the experimental results of a second metabolomics study where we evaluated the influence of HCC on the metabolism of cirrhotic patients.<sup>18</sup> OPLS-DA was used to discriminate 33 patients without HCC from 33 patients with large HCC. Spectra were normalized with the probabilistic quotient normalization method. They were divided into 230 domains of 0.05 ppm and the water signal region was suppressed. The resulting  $X$  and  $Y$  matrices are available as ESI† (Dataset4.xlsx). The properties of this dataset correspond to the first situation mentioned by Eriksson *et al.*<sup>8</sup> (*i.e.* interclass variability is large enough relative to

the intraclass variability) and no strong effect of row permutations on the  $Q^2$  parameter was observed for this dataset (Fig. 2a). We modified this dataset until we reached the second condition mentioned by Eriksson and coworkers<sup>8</sup> (*i.e.* non-homogeneous classes and large intraclass variability). The modifications introduced in the original dataset were chosen to simulate three types of circumstances frequently observed in metabolomic studies.

– The first situation is when the main source of variability in the dataset is uncorrelated with the  $Y$  response variable. This can be observed for example when incorrect sample normalization has been applied to correct for dilution effects. To simulate this situation we multiplied each line of the original dataset by a dilution factor randomly chosen between 1 and 50 (ESI† Dataset5.xlsx). We randomly permuted the rows of the resulting dataset and calculated the NSC and  $Q^2$  values for each permutation (Fig. 2b). We observed a larger distribution of  $Q^2$  compared to Fig. 2a. For some permutations, no significant component was obtained.

– A second situation corresponds to the inaccurate labeling of group membership of individuals. The situation is known as class noise.<sup>19</sup> It is frequently encountered in metabolomic studies applied to clinical problems especially when a reliable diagnostic tool is unavailable. To simulate this situation, 10% of the individuals of each group were incorrectly labeled in the



**Fig. 2** The first figures correspond to the original dataset (a) while the others correspond to the original dataset modified to simulate the effects of dilution (b), class noise (c) and small dataset size (d). Each figure illustrates the variation of the NSC and  $Q_{NSC}^2$  parameters when 10 000 random permutations of the dataset lines are performed. Each color is associated to a specific value of NSC. The  $Q_{NSC}^2$  distributions (black lines) correspond to the 10 000 random permutations. These distributions were decomposed according to the NSC value calculated for each permutation (colored lines).

## Opinion

original dataset (ESI, † Dataset6.xlsx). NSC and  $Q^2$  distributions after random permutations of rows were calculated (Fig. 2c). Here again, we observed a larger distribution of  $Q^2$  compared to Fig. 2a.

– Finally, a third situation is when the number of individuals used to build the model is too small. In this case, the probability to build by chance a sample with at least one non-homogeneous or non-representative class is not negligible even if classes are homogeneous in the population. We selected 8 individuals of each class from the original dataset and we randomly permuted the rows of the resulting 16 rows dataset (ESI, † Dataset7.xlsx). For each permutation we calculated the NSC and  $Q^2$  parameters (Fig. 2d). In this case,  $Q^2$  values spread from 0.23 to 0.92.

These results showed that when the situations mentioned above are encountered, quality parameters could be strongly affected by row permutations in the dataset if the K-fold cross-validation procedure is used. Moreover, many combinations of these three situations can be encountered in metabolomics studies.

## Suggestions

We agree with Wheelock *et al.*<sup>14</sup> that the number of components used in the final model,  $Q^2$  and  $p_{CV-ANOVA}$  values should be presented to allow the reader to assess the quality of the multivariate model published on metabolomics studies. Our opinion is that the cross-validation method used to build the model should also be described since it determines the value of the parameters mentioned above. This is particularly necessary if the K-fold procedure is used and the  $Q^2$  value found is smaller than 0.5. Indeed, we have shown in the first example (Fig. 1) that  $Q^2$  values of 0.4 and  $-0.1$  can be obtained for the same dataset. We also strongly recommend SIMCA users to permute the lines of their dataset to control that the  $Q^2$  value calculated is stable regarding this permutation. In the supplementary material, a simple procedure is proposed to estimate  $Q^2$  variability with row rearrangements (Supplementary Data B, ESI†).

The SIMCA software allows users to modify the cross-validation procedures by changing the number of cross-validation sets and/or selecting the individuals of each set. We believe that this possibility can also help users to estimate a confidence interval of the calculated quality parameters. The Leave One Out (LOO) procedure can also be tested on SIMCA by setting the number of subsets to the number of samples. This method does not depend on the order of the rows in the dataset, however, as pointed out by several authors,<sup>20,21</sup> the LOO procedure can lead to over-fitting and over estimation of  $Q^2$ .

Other cross-validation methods that (to our knowledge) are not yet implemented in SIMCA should be tested. We particularly recommend the double cross-validation (2CV) method.<sup>22,23</sup> As the K-fold method, the 2CV uses all the available individuals to build the models and to estimate their predictability. However, in 2CV, the estimation of NSC and  $Q^2$  are decoupled. This is a very important issue since, as illustrated by the Fig. 1b, an

overestimation/underestimation of NSC frequently leads to overestimation/underestimation of  $Q^2$ . Thus, even if the double validation loop process is time consuming compared to the simple validation loop performed in the K-fold procedure, the risk of overestimation or underestimation of the predictability is reduced with the 2CV procedure.

Another interesting method is the Monte Carlo Cross-validation (MCCV) procedure.<sup>24</sup> By randomly building many subsets with many combinations of individuals, this procedure averages the opposite effects of too optimistic and too pessimistic cross-validation submodels.

Finally, we want to remind readers that a truly reliable estimation of the predictability of a model is obtained with individuals that are independent of those used to build this model.<sup>25</sup>

## Conclusion

PLS-DA/OPLS-DA is a powerful method for multivariate problems and SIMCA is a very robust and well-adapted software for this type of analysis. However, as mentioned by Erikson *et al.*, the PLS method fails to build reliable discriminant models under some conditions (*i.e.* non-homogeneous classes and large intraclass variability). We have shown that when these conditions are encountered, the K-fold cross-validation procedure results can be strongly determined by the composition of the different subsets. As a consequence, a simple permutation of the row order in the dataset can strongly modify the value of the quality parameters calculated by the K-fold cross-validation procedures. In some cases, contradictory conclusions concerning the significance of the multivariate model can be drawn from the same dataset.

## Abbreviations

PLS	Projection to latent structures
OPLS	Orthogonal projection to latent structures
DA	Discriminant analysis
PCA	Principal components analysis
NC	Number of components
NSC	Number of significant components
ONC	Optimal number of components
LOO	Leave one out
MCCV	Monte Carlo cross-validation

## References

- 1 B. Worley and R. Powers, *Curr. Metabolomics*, 2013, **1**, 92–107.
- 2 B.-H. Mevik and R. Wehrens, *Journal of Statistical Software*, 2007, **18**, 1–24.
- 3 J. Xia, R. Mandal, I. V. Sinelnikov, D. Broadhurst and D. S. Wishart, *Nucleic Acids Res.*, 2012, **40**, 127–133.
- 4 B. Worley and R. Powers, *ACS Chem. Biol.*, 2014, **9**, 1138–1144.

- 5 H. S. Tapp and E. K. Kemsley, *TrAC, Trends Anal. Chem.*, 2009, **28**, 1322–1327.
- 6 E. K. Kemsley and H. S. Tapp, *J. Chemom.*, 2009, **23**, 263–264.
- 7 G. A. Marcoulides and C. Saunders, *Management Information Systems Quarterly*, 2006, **30**, iii–ix.
- 8 L. Eriksson, E. Johansson, N. Kettaneh-Wold, J. Trygg, C. Wikström and S. Wold, *Multi- and Megavariate Data Analysis, Part I: Basic Principles and Applications*, Umetrics AB, Umeå, Sweden, 2006.
- 9 Umetrics, *User Guide SIMCA+ 12*, Umetrics, 2008.
- 10 J. Westerhuis, H. J. Hoefsloot, S. Smit, D. Vis, A. Smilde, E. J. van Velzen, J. M. van Duijnhoven and F. van Dorsten, *Metabolomics*, 2008, **4**, 81–89.
- 11 H. L. Cai, H. D. Li, X. Z. Yan, B. Sun, Q. Zhang, M. Yan, W. Y. Zhang, P. Jiang, R. H. Zhu, Y. P. Liu, P. F. Fang, P. Xu, H. Y. Yuan, X. H. Zhang, L. Hu, W. Yang and H. S. Ye, *J. Proteome Res.*, 2012, **11**, 4338–4350.
- 12 B. Dong, J. Jia, W. Hu, Q. Chen, C. Jiang, J. Pan, Y. Huang, W. Xue and H. Gao, *Clin. Biochem.*, 2013, **46**, 346–353.
- 13 L. Eriksson, J. Trygg and S. Wold, *J. Chemom.*, 2008, **22**, 594–600.
- 14 A. M. Wheelock and C. E. Wheelock, *Mol. BioSyst.*, 2013, **9**, 2589–2596.
- 15 C. Goossens, M. N. Triba, L. Le Moyec, O. Seror, P. Nahon and P. Savarin, 2014, in preparation.
- 16 F. Dieterle, A. Ross, G. Schlotterbeck and H. Senn, *Anal. Chem.*, 2006, **78**, 4281–4290.
- 17 J. Trygg and S. Wold, *J. Chemom.*, 2002, **16**, 119–128.
- 18 P. Nahon, R. Amathieu, M. N. Triba, N. Bouchemal, J. C. Nault, M. Zioli, O. Seror, G. Dhonneur, J. C. Trinchet, M. Beaugrand and L. Le Moyec, *Clin. Cancer Res.*, 2012, **18**, 6714–6722.
- 19 D. B. Kell and R. D. King, *Trends Biotechnol.*, 2000, **18**, 93–98.
- 20 A. Golbraikh and A. Tropsha, *J. Mol. Graphics Modell.*, 2002, **20**, 269–276.
- 21 H. A. Martens and P. Dardenne, *Chemom. Intell. Lab. Syst.*, 1998, **44**, 99–121.
- 22 S. Smit, M. I. J. van Breemen, H. C. J. Hoefsloot, A. K. Smilde, J. M. F. G. Aerts and C. G. de Koster, *Anal. Chim. Acta*, 2007, **592**, 210–217.
- 23 M. Stone, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 1974, **36**, 111–147.
- 24 Q.-S. Xu, Y.-Z. Liang and Y.-P. Du, *J. Chemom.*, 2004, **18**, 112–120.
- 25 D. Broadhurst and D. Kell, *Metabolomics*, 2006, **2**, 171–196.



## Conclusions & Perspectives

---

Les atteintes hépatiques touchent de plus en plus de personnes en France. D'après les chiffres de l'INSERM, plus de 700 000 personnes sont atteintes de cirrhose en France entraînant de 10 000 à 15 000 décès par an et constituant aujourd'hui un véritable problème de santé publique. Etant asymptomatiques pour la plupart d'entre elles, leurs apparitions, leurs évolutions ainsi que les risques de complications sont encore loin d'être totalement compris et prévisibles. Ainsi, l'absence de marqueurs cliniques et biologiques fiables rendent la classification ainsi que le diagnostic de ces hépatopathies difficiles à déterminer.

Nous avons vu au travers des différents travaux exposés dans ce manuscrit que la métabolomique est une approche prometteuse pour l'identification des changements métaboliques notamment lors d'une atteinte hépatique. L'analyse métabolomique permet de mettre en évidence les métabolites impliqués dans la rupture de l'homéostasie induite par les changements physiopathologiques étudiés. Ainsi, il est maintenant largement reconnu que la détermination de l'état pathologique d'un individu ne dépend plus de l'identification d'un biomarqueur unique. Grâce aux méthodes analytiques actuelles, il est aujourd'hui possible d'établir une véritable cartographie de plusieurs centaines voire milliers de descripteurs caractérisant, par exemple l'état métabolique d'un individu. Une telle approche permet de constituer des profils d'évolution de ces descripteurs dans le but de préciser et même de prédire l'atteinte. D'autre part, l'analyse d'un profil métabolique complet permet d'avoir une vision plus globale des différents processus biochimiques pouvant également mener à une meilleure compréhension des mécanismes biochimiques et biologiques. Les études métabolomiques possèdent par conséquent le potentiel de fournir certains éléments de réponses à un grand nombre de questions biologiques.

Au cours de la première étude, nous avons exploré les changements métaboliques du foie observés lors d'une atteinte hépatique de type NAFLD à partir d'extractions des métabolites hydrosolubles et liposolubles de biopsies de foie humain.

Notre étude par une analyse multivariée n'a pas permis de déterminer des métabolites permettant de classer les foies (normaux, stéatoses et NASH) suite à l'extraction en phase

aqueuse. Cependant, une tendance univariée s'est avérée être significative pour deux métabolites qui sont la créatine et le signal des N-acétyl correspondant aux glycoprotéines. Les modifications dans la concentration de ces métabolites avaient déjà été rapportées dans différentes études en particulier une diminution du niveau de la créatine observée lors des atteintes de type NAFLD.

L'analyse de la phase liposoluble a quant à elle démontré une discrimination significative des 3 groupes correspondant aux extraits de foies normaux, stéatoses et NASH.

D'autre part nous avons également mis en évidence par cette étude que différentes conditions de stockage, en particulier liées à la comparaison d'échantillons ayant subi un cycle de décongélation-recongélation, ont un impact sur les analyses des métabolites hydrosolubles.

Ces résultats préliminaires ont permis d'avoir une estimation de la reproductibilité des techniques d'extractions utilisées ainsi que d'obtenir une première idée sur les modifications métaboliques observées entre le foie contrôle, stéatose et NASH. Néanmoins, l'analyse de ce jeu de données nécessite d'approfondir encore les investigations dans la recherche d'une meilleure classification d'après diverses combinaisons des paramètres clinico-biologiques relevés.

D'autre part, les résultats obtenus par cette méthode d'extraction seront prochainement, comparés et complétés avec l'analyse  $^1\text{H}$  RMN d'échantillons de sérum de ces mêmes patients. Par ailleurs, l'analyse complémentaire par la technique spectroscopique HR-MAS du tissu intact est également prévue. En plus d'avoir une préparation quasi nulle des échantillons, cette technique présente l'avantage d'une analyse directe de l'organe ciblé. Les signatures métaboliques des prélèvements tissulaires permettront alors de compléter et d'enrichir celles des échantillons obtenus par les méthodes d'extraction et du sérum.

Lors de la seconde étude, nous nous sommes intéressés à une des complications les plus fréquentes des hépatopathies qui est le carcinome hépatocellulaire. Les effets de la RFA sur les 120 patients présentant un CHC se sont traduits sur le profil sérique par des marqueurs de l'inflammation impliquant des modifications dans la concentration des métabolites intervenant principalement dans le métabolisme énergétique et le métabolisme des acides aminés. Les profils métaboliques obtenus à partir des différents sera de ces patients n'ont pas permis de mettre en évidence une signature métabolique caractéristique de la récurrence

de la tumeur. Néanmoins, cette étude a montré une différenciation des profils métaboliques en fonction de l'origine virale ou non du CHC. Cette différence de profils métaboliques s'est non seulement révélée dans le sérum avant l'ablation mais aussi après la RFA suggérant également différents mécanismes dans l'évolution métabolique post intervention.

Une collaboration avec l'équipe du Pr. Guowang Xu de l'Institut de Chimie Physique de Dalian en Chine, permettra également d'analyser ces échantillons de sérum en spectrométrie de masse et devrait fournir une information complémentaire aux différents modèles obtenus par RMN.

L'exploration *in vivo* grâce à la spectroscopie par résonance magnétique nucléaire (SRM) offre également de nombreuses possibilités de champ d'exploration non invasives. La SRM du proton est une technique d'imagerie moléculaire permettant la quantification relative de certains métabolites dans les tissus. L'efficacité potentielle de cette technique apporterait un élément supplémentaire de caractérisation des atteintes hépatiques telles que les tumeurs pour lesquelles la spécificité de l'imagerie morphologique standard est encore limitée [221]. Cette technique fournirait des informations supplémentaires pour distinguer les tumeurs des lésions non tumorales. Elle permettrait l'évaluation du grade de la tumeur de manière non invasive et également un suivi de l'évolution après la mise en place de différents traitements.

Ce type de projet est actuellement en train de voir le jour grâce aux collaborations actives entre l'équipe de recherche RMN du laboratoire et les services d'hépatologie de l'hôpital Jean Verdier de Bondy et de l'hôpital Avicenne de Bobigny.

Enfin, le dernier point de la thèse s'est articulé autour de l'importance de la validation des techniques utilisées en métabolomique et plus particulièrement concernant l'analyse des données multivariées.

De nombreuses applications cliniques en métabolomique sont actuellement possibles grâce, notamment, aux innovations techniques avec le développement d'outils d'analyse de plus en plus performants mais également grâce à l'utilisation des analyses de données multivariées. Si le but ultime pour beaucoup de chercheurs repose sur une application en routine de la métabolomique en milieu clinique, il réside cependant encore une certaine réserve notamment liée à la rigueur de la significativité des analyses statistiques. Ainsi, l'intérêt de s'assurer des bonnes méthodes de validation comme nous les avons également

proposées dans cette dernière publication est de rationaliser les résultats obtenus lors des différentes études métabolomiques et de permettre ainsi leurs potentielles applications en routine.

D'autre part, plus la taille de l'échantillon représentant la population étudiée sera grande, plus les modèles prédictifs seront fiables. Il est également impératif d'obtenir des prélèvements de liquides biologiques et tissulaires de qualité. Le développement des bases de données et des centres de ressources biologiques uniformisés en accord avec les registres épidémiologiques et les équipes de recherche fondamentale, représente un point fort pour mener à bien des études métabolomiques.

Cependant, les approches actuelles de métabolomique peuvent s'avérer longues et fastidieuses lors de la collecte, la préparation et l'analyse des échantillons, l'identification des métabolites, l'interprétation biologique ainsi que lors de la validation des analyses statistiques. Un des futurs challenges en métabolomique reste, par conséquent, le développement de méthodes automatisées d'analyse et d'identification des métabolites afin de réduire non seulement le temps passé à ce type d'étude mais également de pouvoir appliquer la métabolomique en routine. De plus, la complexité, essentiellement liée à la nature multivariée des analyses métabolomiques requiert encore des efforts particuliers afin d'extraire les informations biologiques les plus pertinentes [222].

A cette fin, la combinaison des données de la métabolomique avec celles des autres sciences « omiques » peut s'avérer prometteuse pour fournir une information significative au sein de la biologie des systèmes. Les futures recherches et collaborations pour le développement de nouvelles méthodes d'intégration des différentes données « omiques » sont les prochaines barrières à franchir pour les futures avancées dans la recherche médicale.

# Annexes

## Annexe 1.

Avantages et limitations de la RMN et la SM comme outils analytiques utilisés en métabolomique [13].

	<b>NMR</b>	<b>Mass spectrometry</b>
Sensitivity	Low but can be improved with higher field strength, cryo- and microprobes, and dynamic nuclear polarization	High and detection limit reach nanomolar
Selectivity	Even though few selective experiments are available such as selective TOCSY, it is in general used for nonselective analysis	Can be used for both selective and nonselective (targeted and nontargeted) analyses
Sample measurement	All metabolites that have NMR concentration level can be detected in one measurement	Usually need different chromatography techniques for different classes of metabolites
Sample recovery	Nondestructive; sample can be recovered and stored for a long time; several analyses can be carried out on the same sample	Destructive technique but need a small amount of sample
Reproducibility	Very high	Moderate
Sample preparation	Need minimal sample preparation	More demanding; needs different columns and optimization of ionization conditions
Tissue samples	Yes, using HRMAS NMR tissue samples analyzed directly	No, requires tissue extraction
Number of detectable metabolites in urine sample	40–200 depending on spectral resolution	Could be more than 500 using different MS techniques
Target analysis	Not relevant for targeted analysis	Superior for targeted analysis
In vivo studies	Yes—widely used for $^1\text{H}$ magnetic resonance spectroscopy (and to a lesser degree $^{31}\text{P}$ and $^{13}\text{C}$ )	No—although suggestion that desorption electrospray ionization (DESI) may be a useful way to sample tissues minimally invasively during surgery

**Annexe 2.**

Score de Child Pugh [113].

	1 point	2 points	3 points
Encéphalopathie	Absente	Confusion	Coma
Ascite	Absente	Minime	Importante
Bilirubine totale ( $\mu\text{mol/l}$ )	< 35	35 à 50	> 50
Albumine (g/l)	> 35	28 à 35	< 28
Taux de prothrombine (%)	> 50	40 à 50	< 40

La gravité est croissante avec la valeur du score : entre 5 et 6 points (classe A) ; entre 7 et 9 points (classe B) ; entre 10 et 15 points (classe C)  
 En cas de cirrhose compensée, la plupart des malades sont en classe A. La cirrhose décompensée correspond à une classe B ou C. Ce score ne prend pas en compte certaines complications de la cirrhose telles que l'hémorragie digestive ou le carcinome hépatocellulaire.

**Annexe 3.**

Score MELD [135].

<b>MELD=</b> (3,8 x ln[bilirubine (mg/dL)])+(11,2 x ln[INR])+(9,6 x ln[créatinine (mg/dL)]+(6,4 x étiologie*))
<p>Dans ce calcul :</p> <ul style="list-style-type: none"> <li>- toute valeur inférieure à 1 reçoit la valeur de 1 (c'est-à-dire si la bilirubine est à 0,8, une valeur de 1 est utilisée)</li> <li>- si le patient a été dialysé à deux reprises au cours des 7 derniers jours, la valeur de la créatinine sérique doit être donnée à 4 mg/dL.</li> </ul> <p>* [0 si cholestatique ou alcoolique, 1 dans le cas contraire]</p>

# Bibliographie

---

1. Crespo Yanguas S, Cogliati B, Willebrords J, Maes M, Colle I, van den Bossche B, et al. Experimental models of liver fibrosis. *Archives of toxicology* 2015.
2. Loscalzo J, Kohane I, Barabasi AL. Human disease classification in the postgenomic era: a complex systems approach to human pathobiology. *Mol Syst Biol* 2007;3:124.
3. Holmes E, Wilson ID, Nicholson JK. Metabolic phenotyping in health and disease. *Cell* 2008;134(5):714-7.
4. German JB, Hammock BD, Watkins SM. Metabolomics: building on a century of biochemistry to guide human health. *Metabolomics* 2005;1(1):3-9.
5. Nicholson JK, Lindon JC. Systems biology: Metabonomics. *Nature* 2008;455(7216):1054-6.
6. Beisken S, Eiden M, Salek RM. Getting the right answers: understanding metabolomics challenges. *Expert review of molecular diagnostics* 2015;15(1):97-109.
7. Lindon JC, Holmes E, Bollard ME, Stanley EG, Nicholson JK. Metabonomics technologies and their applications in physiological monitoring, drug safety assessment and disease diagnosis. *Biomarkers : biochemical indicators of exposure, response, and susceptibility to chemicals* 2004;9(1):1-31.
8. 2020 visions. *Nature* 2010;463(7277):26-32.
9. Bruix J, Gores GJ, Mazzaferro V. Hepatocellular carcinoma: clinical frontiers and perspectives. *Gut* 2014.
10. Fiehn O. Metabolomics--the link between genotypes and phenotypes. *Plant molecular biology* 2002;48(1-2):155-71.
11. Nicholson JK, Lindon JC, Holmes E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* 1999;29(11):1181-9.
12. Smolinska A, Blanchet L, Buydens LM, Wijmenga SS. NMR and pattern recognition methods in metabolomics: from data acquisition to biomarker discovery: a review. *Anal Chim Acta*;750:82-97.
13. Emwas A-H, Salek R, Griffin J, Merzaban J. NMR-based metabolomics in human disease diagnosis: applications, limitations, and recommendations. *Metabolomics* 2013.
14. Roberts MJ, Schirra HJ, Lavin MF, Gardiner RA. NMR-based metabolomics: global analysis of metabolites to adress problems in prostate cancer. *Cervical, Brest and Prostate Cancer: IConcept Press*; 2014. p 1-43.
15. Weckwerth W, Morgenthal K. Metabolomics: from pattern recognition to biological interpretation. *Drug discovery today* 2005;10(22):1551-8.
16. Gil A, Siegel D, Permentier H, Reijngoud DJ, Dekker F, Bischoff R. Stability of energy metabolites - an often overlooked issue in metabolomics studies: A review. *Electrophoresis* 2015.
17. Beckonert O, Keun HC, Ebbels TM, Bundy J, Holmes E, Lindon JC, et al. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat Protoc* 2007;2(11):2692-703.
18. Lindon JC, Holmes E, Nicholson JK. Metabonomics in pharmaceutical R&D. *The FEBS*

- journal 2007;274(5):1140-51.
19. Averna TA, Kline EE, Smith AY, Sillerud LO. A decrease in <sup>1</sup>H nuclear magnetic resonance spectroscopically determined citrate in human seminal fluid accompanies the development of prostate adenocarcinoma. *J Urol* 2005;173(2):433-8.
  20. Deepinder F, Chowdary HT, Agarwal A. Role of metabolomic analysis of biomarkers in the management of male infertility. *Expert Rev Mol Diagn* 2007;7(4):351-8.
  21. Menon R, Jones J, Gunst PR, Kacerovsky M, Fortunato SJ, Saade GR, et al. Amniotic fluid metabolomic analysis in spontaneous preterm birth. *Reproductive sciences* 2014;21(6):791-803.
  22. Han X, Rozen S, Boyle SH, Hellegers C, Cheng H, Burke JR, et al. Metabolomics in early Alzheimer's disease: identification of altered plasma sphingolipidome using shotgun lipidomics. *PLoS One* 2011;6(7):e21643.
  23. Brown MV, McDunn JE, Gunst PR, Smith EM, Milburn MV, Troyer DA, et al. Cancer detection and biopsy classification using concurrent histopathological and metabolomic analysis of core biopsies. *Genome Med* 2012;4(4):33.
  24. Duarte IF, Stanley EG, Holmes E, Lindon JC, Gil AM, Tang H, et al. Metabolic assessment of human liver transplants from biopsy samples at the donor and recipient stages using high-resolution magic angle spinning <sup>1</sup>H NMR spectroscopy. *Anal Chem* 2005;77(17):5570-8.
  25. Serkova NJ, Hasebroock KM, Kraft SL. Magnetic resonance spectroscopy of living tissues. *Methods Mol Biol* 2009;520:315-27.
  26. Lindon JC, Nicholson JK, Holmes E, Antti H, Bollard ME, Keun H, et al. Contemporary issues in toxicology the role of metabonomics in toxicology and its evaluation by the COMET project. *Toxicol Appl Pharmacol* 2003;187(3):137-46.
  27. Spratlin JL, Serkova NJ, Eckhardt SG. Clinical applications of metabolomics in oncology: a review. *Clin Cancer Res* 2009;15(2):431-40.
  28. Vinayavekhin N, Homan EA, Saghatelian A. Exploring disease through metabolomics. *ACS chemical biology* 2010;5(1):91-103.
  29. Duarte IF, Diaz SO, Gil AM. NMR metabolomics of human blood and urine in disease research. *J Pharm Biomed Anal* 2014;93:17-26.
  30. Griffin JL. The potential of metabonomics in drug safety and toxicology. *Drug discovery today Technologies* 2004;1(3):285-93.
  31. Robertson DG, Watkins PB, Reily MD. Metabolomics in toxicology: preclinical and clinical applications. *Toxicological sciences : an official journal of the Society of Toxicology* 2011;120 Suppl 1:S146-70.
  32. Mantle PG, Nicholls AW, Shockcor JP. <sup>1</sup>H NMR spectroscopy-based metabolomic assessment of uremic toxicity, with toxicological outcomes, in male rats following an acute, mid-life insult from ochratoxin a. *Toxins (Basel)* 2011;3(6):504-19.
  33. Ebbels TM, Keun HC, Beckonert OP, Bollard ME, Lindon JC, Holmes E, et al. Prediction and classification of drug toxicity using probabilistic modeling of temporal metabolic data: the consortium on metabonomic toxicology screening approach. *J Proteome Res* 2007;6(11):4407-22.
  34. Cho SH, Choi MH, Kwon OS, Lee WY, Chung BC. Metabolic significance of bisphenol A-induced oxidative stress in rat urine measured by liquid chromatography-mass spectrometry. *J Appl Toxicol* 2009;29(2):110-7.
  35. Wolfender JL, Glauser G, Boccard J, Rudaz S. MS-based plant metabolomic approaches for biomarker discovery. *Nat Prod Commun* 2009;4(10):1417-30.

36. Asai T, Matsukawa T, Kajiyama S. Metabolic changes in Citrus leaf volatiles in response to environmental stress. *J Biosci Bioeng* 2015.
37. Tohge T, Fernie AR. Metabolomics-Inspired Insight into Developmental, Environmental and Genetic Aspects of Tomato Fruit Chemical Composition and Quality. *Plant Cell Physiol* 2015;56(9):1681-96.
38. Kusano M, Yang Z, Okazaki Y, Nakabayashi R, Fukushima A, Saito K. Using metabolomic approaches to explore chemical diversity in rice. *Mol Plant* 2015;8(1):58-67.
39. Weber RJ, Selander E, Sommer U, Viant MR. A stable-isotope mass spectrometry-based metabolic footprinting approach to analyze exudates from phytoplankton. *Marine drugs* 2013;11(11):4158-75.
40. Southam AD, Lange A, Al-Salhi R, Hill EM, Tyler CR, Viant MR. Distinguishing between the metabolome and xenobiotic exposome in environmental field samples analysed by direct-infusion mass spectrometry based metabolomics and lipidomics. *Metabolomics* 2014;10(6):1050-58.
41. Kuhlisch C, Pohnert G. Metabolomics in chemical ecology. *Natural product reports* 2015;32(7):937-55.
42. Van Dorsten FA, Daykin CA, Mulder TP, Van Duynhoven JP. Metabonomics approach to determine metabolic differences between green tea and black tea consumption. *J Agric Food Chem* 2006;54(18):6929-38.
43. Holmes E, Loo RL, Stampler J, Bictash M, Yap IK, Chan Q, et al. Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature* 2008;453(7193):396-400.
44. Ismail NA, Posma JM, Frost G, Holmes E, Garcia-Perez I. The role of metabonomics as a tool for augmenting nutritional information in epidemiological studies. *Electrophoresis* 2013;34(19):2776-86.
45. Mora-Cubillos X, Tulipani S, Garcia-Aloy M, Bullo M, Tinahones FJ, Andres-Lacueva C. Plasma metabolomic biomarkers of mixed nuts exposure inversely correlate with severity of metabolic syndrome. *Mol Nutr Food Res* 2015.
46. Kim S, Kim J, Yun EJ, Kim KH. Food metabolomics: from farm to human. *Curr Opin Biotechnol* 2015;37:16-23.
47. De Preter V. Metabonomics and systems biology. *Methods Mol Biol* 2015;1277:245-55.
48. Casado-Vela J, Cebrian A, Gomez del Pulgar MT, Lacal JC. Approaches for the study of cancer: towards the integration of genomics, proteomics and metabolomics. *Clin Transl Oncol* 2011;13(9):617-28.
49. Lin D, Hollander Z, Meredith A, Stadnick E, Sasaki M, Cohen Freue G, et al. Molecular signatures of end-stage heart failure. *Journal of cardiac failure* 2011;17(10):867-74.
50. Adamski J. Genome-wide association studies with metabolomics. *Genome Med* 2012;4(4):34.
51. Kind T, Scholz M, Fiehn O. How large is the metabolome? A critical analysis of data exchange practices in chemistry. *PLoS One* 2009;4(5):e5440.
52. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000;28(1):27-30.
53. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExpASY: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 2003;31(13):3784-8.

54. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13(11):2498-504.
55. Beltran A, Suarez M, Rodriguez MA, Vinaixa M, Samino S, Arola L, et al. Assessment of compatibility between extraction methods for NMR- and LC/MS-based metabolomics. *Anal Chem* 2012;84(14):5838-44.
56. Lindon JC, Nicholson JK. Spectroscopic and statistical techniques for information recovery in metabonomics and metabolomics. *Annu Rev Anal Chem (Palo Alto Calif)* 2008;1:45-69.
57. Emwas AH. The strengths and weaknesses of NMR spectroscopy and mass spectrometry with particular focus on metabolomics research. *Methods in molecular biology* 2015;1277:161-93.
58. Lindon JC, Nicholson JK. Spectroscopic and statistical techniques for information recovery in metabonomics and metabolomics. *Annu Rev Anal Chem (Palo Alto Calif)* 2008;1:45-69.
59. Bedair M, Sumner LW. Current and emerging mass-spectrometry technologies for metabolomics. *Trends Anal Chem* 2008;27(3):238-50.
60. Lenz EM, Wilson ID. Analytical strategies in metabonomics. *Journal of proteome research* 2007;6(2):443-58.
61. Canet D, Boubel J-C., Canet Soulas, E. La RMN. Concepts, méthodes et applications. 2002. 236 p.
62. Günther H. La spectroscopie de RMN. 1993.
63. Keeler J. Understanding NMR spectroscopy. 2010.
64. Levitt MH. Spin Dynamics. Basics of Nuclear Magnetic Resonance. edition S, editor2008.
65. Claridge T. High-Resolution NMR Techniques in Organic Chemistry. 2009.
66. Sanders JH, B. Modern NMR Spectroscopy. A guide for Chemists. 1993.
67. Hoch JC, Stern, A.S. NMR data processing. 1996.
68. Bollard ME, Garrod S, Holmes E, Lindon JC, Humpfer E, Spraul M, et al. High-resolution (1)H and (1)H-(13)C magic angle spinning NMR spectroscopy of rat liver. *Magn Reson Med* 2000;44(2):201-7.
69. Andrew ER, Brodbury A, Eades R. Removal of Dipolar Broadening of Nuclear Magnetic Resonance Spectra of Solids by Specimen Rotation. *Nature* 1959;183((4678)):1802-03.
70. Giraudeau P, Silvestre V, Akoka S. Optimizing water suppression for quantitative NMR-based metabolomics: a tutorial review. *Metabolomics* 2015;11(2).
71. McKay RT. How the 1D-NOESY Suppresses Solvent Signal in Metabonomics NMR Spectroscopy: An Examination of the Pulse Sequence Components and Evolution. *Concepts in Magnetic Resonance Part A* 2011;38A((5)):197-220.
72. Carr HY, Purcell EM. Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments. *Phys Rev* 1954;94:630-38.
73. Meiboom S, Gill D. Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. *Rev Sci Instrum* 1958;29:688-91.
74. Vettukattil R. Preprocessing of raw metabonomic data. *Methods Mol Biol* 2015;1277:123-36.
75. De Meyer T, Sinnaeve D, Van Gasse B, Rietzschel ER, De Buyzere ML, Langlois MR, et al. Evaluation of standard and advanced preprocessing methods for the univariate

- analysis of blood serum <sup>1</sup>H-NMR spectra. *Anal Bioanal Chem* 2010;398(4):1781-90.
76. Cobas JC, Sardina FJ. Nuclear magnetic resonance data processing. MestRe-C: A software package for desktop computers. *Concepts in Magnetic Resonance Part A* 2003;19A(2):80-96.
  77. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of biomolecular NMR* 1995;6(3):277-93.
  78. Rule GS, T.K. H. *Fundamentals of Protein NMR Spectroscopy*. 2006.
  79. Shimizu A, Ikeguchi M, Sugai S. Appropriateness of DSS and TSP as internal references for (<sup>1</sup>H) NMR studies of molten globule proteins in aqueous media. *J Biomol NMR* 1994;4(6):859-62.
  80. Defernez M, Colquhoun IJ. Factors affecting the robustness of metabolite fingerprinting using <sup>1</sup>H NMR spectra. *Phytochemistry* 2003;62(6):1009-17.
  81. Savorani F, Tomasi G, Engelsen SB. icoshift: A versatile tool for the rapid alignment of 1D NMR spectra. *J Magn Reson* 2010;202(2):190-202.
  82. Wishart DS. Quantitative metabolomics using NMR. *TrAC Trends in Analytical Chemistry* 2008;27(3):228-37.
  83. Sousa SAA, Magalhaes A, Ferreira MMC. Optimized bucketing for NMR spectra: Three case studies. *Chemometrics and Intelligent Laboratory Systems* 2013(122):93-102.
  84. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 2003;19(2):185-93.
  85. Craig A, Cloarec O, Holmes E, Nicholson JK, Lindon JC. Scaling and normalization effects in NMR spectroscopic metabolomic data sets. *Anal Chem* 2006;78(7):2262-7.
  86. Dieterle F, Ross A, Schlotterbeck G, Senn H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in <sup>1</sup>H NMR metabolomics. *Anal Chem* 2006;78(13):4281-90.
  87. Lee J, Park J, Lim MS, Seong SJ, Seo JJ, Park SM, et al. Quantile normalization approach for liquid chromatography-mass spectrometry-based metabolomic data from healthy human volunteers. *Anal Sci* 2012;28(8):801-5.
  88. Van den Berg RA, Hoefsloot HC, Westerhuis JA, Smilde AK, van der Werf MJ. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* 2006;7:142.
  89. Wishart DS. Human Metabolome Database: completing the 'human parts list'. *Pharmacogenomics* 2007;8(7):683-6.
  90. Psychogios N, Hau DD, Peng J, Guo AC, Mandal R, Bouatra S, et al. The human serum metabolome. *PLoS One*;6(2):e16957.
  91. Nicholson JK, Foxall PJ, Spraul M, Farrant RD, Lindon JC. 750 MHz <sup>1</sup>H and <sup>1</sup>H-<sup>13</sup>C NMR spectroscopy of human blood plasma. *Anal Chem* 1995;67(5):793-811.
  92. Hendriks MMWB, Van Eeuwijk FA, Jellema RH, Westerhuis JA, Reijmers TH, Hoefsloot HCJ, et al. Data-processing strategies for metabolomics studies. *Trends in Analytical Chemistry* 2011;30(10):1685-98.
  93. Wonnacott TH, Wonnacott RJ. *Statistique: Economie-Gestion-Sciences-Médecine (avec exercices d'application)*. 1999.
  94. Lindon JC, Holmes E, Nicholson JK. Metabonomics techniques and applications to pharmaceutical research & development. *Pharmaceutical research* 2006;23(6):1075-88.

95. Jolliffe IT. Principal Component Analysis. 2002.
96. Wishart DS. Computational approaches to metabolomics. *Methods Mol Biol*;593:283-313.
97. Barker M, Rayens W. Partial least squares for discrimination. *Journal of Chemometrics* 2003;17(3):166-73.
98. Trygg J, Holmes E, Lundstedt T. Chemometrics in metabonomics. *J Proteome Res* 2007;6(2):469-79.
99. Tenenhaus M. La regression PLS: Théorie et pratique. 1998.
100. Trygg J, Wold S. Orthogonal projections to latent structures (O-PLS). *J Chemometrics* 2002;16:119-28.
101. van Velzen EJ, Westerhuis JA, van Duynhoven JP, van Dorsten FA, Hoefsloot HC, Jacobs DM, et al. Multilevel data analysis of a crossover designed human nutritional intervention study. *J Proteome Res* 2008;7(10):4483-91.
102. Madsen R, Lundstedt T, Trygg J. Chemometrics in metabolomics--a review in human disease diagnosis. *Anal Chim Acta* 2010;659(1-2):23-33.
103. Szymanska E, Saccenti E, Smilde AK, Westerhuis JA. Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics* 2012;8(Suppl 1):3-16.
104. Westerhuis JA, Van Velzen EJJ, Hoefsloot HCJ, Smilde AK. Discriminant Q2 (DQ2) for improved discrimination in PLS-DA models. *Metabolomics* 2008(4):293-96.
105. Broadhurst DI, Kell DB. Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics* 2006;2(4):171-96.
106. Eliasson M, Rannar S, Trygg J. From data processing to multivariate validation--essential steps in extracting interpretable information from metabolomics data. *Curr Pharm Biotechnol* 2011;12(7):996-1004.
107. Westerhuis J, Hoefsloot H, Smit S, Vis D, Smilde A, van Velzen E, et al. Assessment of PLS-DA cross validation. *Metabolomics* 2008;4(1):81-89.
108. Fawcett T. An introduction to ROC analysis. *Pattern Recognition Letters* 2006;27:861-74.
109. Eriksson L, Trygg J, Wold S. CV-ANOVA for significance testing of PLS and OPLS® models. *Journal of Chemometrics* 2008;22(11-12):594-600.
110. Westerhuis JA, van Velzen EJ, Hoefsloot HC, Smilde AK. Multivariate paired data analysis: multilevel PLS-DA versus OPLS-DA. *Metabolomics* 2009;6(1):119-28.
111. Cloarec O, Dumas ME, Craig A, Barton RH, Trygg J, Hudson J, et al. Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic 1H NMR data sets. *Anal Chem* 2005;77(5):1282-9.
112. Beaudeau J-L, Durand G. Biochimie médicale. Marqueurs actuels et perspectives. 2011.
113. Benhamou J-P, Erlinger S. Maladies du foie et des voies biliaires. 2007.
114. James J, Liou IW. Comprehensive Care of Patients with Chronic Liver Disease. *The Medical clinics of North America* 2015;99(5):913-33.
115. Keane MG, Lai C, Pereira SP. Detecting patients with cirrhosis in primary care. *Practitioner* 2014;258(1773):15-20, 2-3.
116. Amathieu R, Triba MN, Nahon P, Bouchemal N, Kamoun W, Haouache H, et al. Serum 1H-NMR Metabolomic Fingerprints of Acute-On-Chronic Liver Failure in Intensive Care Unit Patients with Alcoholic Cirrhosis. *PLoS One* 2014;9(2):e89230.
117. Bernal W, Jalan R, Quaglia A, Simpson K, Wendon J, Burroughs A. Acute-on-chronic

- liver failure. *Lancet* 2015.
118. Forner A, Llovet JM, Bruix J. Hepatocellular carcinoma. *Lancet* 2012;379(9822):1245-55.
  119. Kleiner DE, Brunt EM, Van Natta M, Behling C, Contos MJ, Cummings OW, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* 2005;41(6):1313-21.
  120. Nusrat S, Khan MS, Fazili J, Madhoun MF. Cirrhosis and its complications: Evidence based treatment. *World J Gastroenterol* 2014;20(18):5442-60.
  121. Racine-Samson L, Scoazec JY, D'Errico A, Fiorentino M, Christa L, Moreau A, et al. The metabolic organization of the adult human liver: a comparative study of normal, fibrotic, and cirrhotic liver tissue. *Hepatology* 1996;24(1):104-13.
  122. Petersen KF, Krssak M, Navarro V, Chandramouli V, Hundal R, Schumann WC, et al. Contributions of net hepatic glycogenolysis and gluconeogenesis to glucose production in cirrhosis. *The American journal of physiology* 1999;276(3 Pt 1):E529-35.
  123. Tessari P, Kiwanuka E, Vettore M, Barazzoni R, Zanetti M, Cecchet D, et al. Phenylalanine and tyrosine kinetics in compensated liver cirrhosis: effects of meal ingestion. *Am J Physiol Gastrointest Liver Physiol* 2008;295(3):G598-604.
  124. Habib A, Mihas AA, Abou-Assi SG, Williams LM, Gavis E, Pandak WM, et al. High-density lipoprotein cholesterol as an indicator of liver function and prognosis in noncholestatic cirrhotics. *Clin Gastroenterol Hepatol* 2005;3(3):286-91.
  125. Romero-Gomez M, Jover M, Galan JJ, Ruiz A. Gut ammonia production and its modulation. *Metabolic brain disease* 2009;24(1):147-57.
  126. Morgan B. Fatty liver disease: the liver labyrinth. *Nature* 2014;516(7529):S8-9.
  127. Hashimoto E, Taniai M, Tokushige K. Characteristics and diagnosis of NAFLD/NASH. *J Gastroenterol Hepatol* 2013;28 Suppl 4:64-70.
  128. Loomba R, Sanyal AJ. The global NAFLD epidemic. *Nat Rev Gastroenterol Hepatol* 2013;10(11):686-90.
  129. Ekstedt M, Hagstrom H, Nasr P, Fredrikson M, Stal P, Kechagias S, et al. Fibrosis stage is the strongest predictor for disease-specific mortality in NAFLD after up to 33 years of follow-up. *Hepatology* 2015;61(5):1547-54.
  130. Vergniol J, de Ledinghen V. [Non-invasive diagnosis of liver fibrosis: guidelines for the use of biomarkers and FibroScan]. *Gastroenterologie clinique et biologique* 2009;33(4):334-44.
  131. Castera L, Forns X, Alberti A. Non-invasive evaluation of liver fibrosis using transient elastography. *J Hepatol* 2008;48(5):835-47.
  132. Foucher J, Chanteloup E, Vergniol J, Castera L, Le Bail B, Adhoute X, et al. Diagnosis of cirrhosis by transient elastography (FibroScan): a prospective study. *Gut* 2006;55(3):403-8.
  133. Bedossa P, Dargere D, Paradis V. Sampling variability of liver fibrosis in chronic hepatitis C. *Hepatology* 2003;38(6):1449-57.
  134. Pugh RN, Murray-Lyon IM, Dawson JL, Pietroni MC, Williams R. Transection of the oesophagus for bleeding oesophageal varices. *Br J Surg* 1973;60(8):646-9.
  135. Kamath PS, Kim WR, Advanced Liver Disease Study G. The model for end-stage liver disease (MELD). *Hepatology* 2007;45(3):797-805.
  136. Daoudaki M, Fouzas I. Hepatocellular carcinoma. *Wien Med Wochenschr* 2014.
  137. Binder-Foucard F, Belot A, Delafosse P, Remontet L, Woronoff A, Bossard N. Estimation nationale de l'incidence et de la mortalité par cancer en France entre 1980

- et 2012. Rapport Francim-HCL-InVS-INCa 2013:32-5.
138. Bruix J, Sherman M. Management of hepatocellular carcinoma: an update. *Hepatology* 2011;53(3):1020-2.
  139. Llovet JM, Beaugrand M. Hepatocellular carcinoma: present status and future prospects. *J Hepatol* 2003;38 Suppl 1:S136-49.
  140. Raoul JL. Natural history of hepatocellular carcinoma and current treatment options. *Semin Nucl Med* 2008;38(2):S13-8.
  141. Crissien AM, Frenette C. Current management of hepatocellular carcinoma. *Gastroenterol Hepatol (N Y)* 2014;10(3):153-61.
  142. Higgins H, Berger DL. RFA for liver tumors: does it really work? *Oncologist* 2006;11(7):801-8.
  143. Hatzaras I, Bischof DA, Fahy B, Cosgrove D, Pawlik TM. Treatment Options and Surveillance Strategies After Therapy for Hepatocellular Carcinoma. *Ann Surg Oncol* 2013.
  144. Roe B, Kensicki E, Mohney R, Hall WW. Metabolomic profile of hepatitis C virus-infected hepatocytes. *PLoS One* 2011;6(8):e23641.
  145. Zhang A, Sun H, Han Y, Yan G, Wang X. Urinary metabolic biomarker and pathway study of hepatitis B virus infected patients based on UPLC-MS system. *PLoS One*;8(5):e64381.
  146. Yu K, Sheng G, Sheng J, Chen Y, Xu W, Liu X, et al. A metabonomic investigation on the biochemical perturbation in liver failure patients caused by hepatitis B virus. *J Proteome Res* 2007;6(7):2413-9.
  147. Nie CY, Han T, Zhang L, Li Y, Liu H, Xiao SX, et al. Cross-sectional and dynamic change of serum metabolite profiling for Hepatitis B-related acute-on-chronic liver failure by UPLC/MS. *J Viral Hepat*;21(1):53-63.
  148. Garcia-Canaveras JC, Donato MT, Castell JV, Lahoz A. A comprehensive untargeted metabonomic analysis of human steatotic liver tissue by RP and HILIC chromatography coupled to mass spectrometry reveals important metabolic alterations. *J Proteome Res* 2011;10(10):4825-34.
  149. Kalhan SC, Guo L, Edmison J, Dasarathy S, McCullough AJ, Hanson RW, et al. Plasma metabolomic profile in nonalcoholic fatty liver disease. *Metabolism*;60(3):404-13.
  150. Puri P, Wiest MM, Cheung O, Mirshahi F, Sargeant C, Min HK, et al. The plasma lipidomic signature of nonalcoholic steatohepatitis. *Hepatology* 2009;50(6):1827-38.
  151. Barr J, Caballeria J, Martinez-Arranz I, Dominguez-Diez A, Alonso C, Muntane J, et al. Obesity-dependent metabolic signatures associated with nonalcoholic fatty liver disease progression. *J Proteome Res* 2012;11(4):2521-32.
  152. Dumas ME, Kinross J, Nicholson JK. Metabolic phenotyping and systems biology approaches to understanding metabolic syndrome and fatty liver disease. *Gastroenterology* 2014;146(1):46-62.
  153. Barr J, Vazquez-Chantada M, Alonso C, Perez-Cormenzana M, Mayo R, Galan A, et al. Liquid chromatography-mass spectrometry-based parallel metabolic profiling of human and mouse model serum reveals putative biomarkers associated with the progression of nonalcoholic fatty liver disease. *J Proteome Res* 2010;9(9):4501-12.
  154. Soga T, Sugimoto M, Honma M, Mori M, Igarashi K, Kashikura K, et al. Serum metabolomics reveals gamma-glutamyl dipeptides as biomarkers for discrimination among different forms of liver disease. *J Hepatol* 2011;55(4):896-905.
  155. Tokushige K, Hashimoto E, Kodama K, Tobari M, Matsushita N, Kogiso T, et al. Serum

- metabolomic profile and potential biomarkers for severity of fibrosis in nonalcoholic fatty liver disease. *J Gastroenterol* 2013.
156. Willebrords J, Pereira IV, Maes M, Yanguas SC, Colle I, Bossche BV, et al. Strategies, models and biomarkers in experimental non-alcoholic fatty liver disease research. *Prog Lipid Res* 2015.
  157. Peng B, Liu F, Han R, Luo G, Cathopoulos T, Lu K, et al. Dynamic metabolic change is indicative of inflammation-induced transformation of hepatic cells. *Int J Biochem Cell Biol* 2015;66:45-58.
  158. Chen T, Xie G, Wang X, Fan J, Qiu Y, Zheng X, et al. Serum and urine metabolite profiling reveals potential biomarkers of human hepatocellular carcinoma. *Mol Cell Proteomics*;10(7):M110 004945.
  159. Patterson AD, Maurhofer O, Beyoglu D, Lanz C, Krausz KW, Pabst T, et al. Aberrant lipid metabolism in hepatocellular carcinoma revealed by plasma metabolomics and lipid profiling. *Cancer Res*;71(21):6590-600.
  160. Zhou L, Ding L, Yin P, Lu X, Wang X, Niu J, et al. Serum metabolic profiling study of hepatocellular carcinoma infected with hepatitis B or hepatitis C virus by using liquid chromatography-mass spectrometry. *J Proteome Res* 2012;11(11):5433-42.
  161. Xue R, Lin Z, Deng C, Dong L, Liu T, Wang J, et al. A serum metabolomic investigation on hepatocellular carcinoma patients by chemical derivatization followed by gas chromatography/mass spectrometry. *Rapid Commun Mass Spectrom* 2008;22(19):3061-8.
  162. Liu Y, Hong Z, Tan G, Dong X, Yang G, Zhao L, et al. NMR and LC/MS-based global metabolomics to identify serum biomarkers differentiating hepatocellular carcinoma from liver cirrhosis. *Int J Cancer* 2013.
  163. Kimhofer T, Fye H, Taylor-Robinson S, Thursz M, Holmes E. Proteomic and metabolomic biomarkers for hepatocellular carcinoma: a comprehensive review. *British journal of cancer* 2015;112(7):1141-56.
  164. Wang B, Chen D, Chen Y, Hu Z, Cao M, Xie Q, et al. Metabolomic profiles discriminate hepatocellular carcinoma from liver cirrhosis by ultraperformance liquid chromatography-mass spectrometry. *J Proteome Res*;11(2):1217-27.
  165. Chen F, Xue J, Zhou L, Wu S, Chen Z. Identification of serum biomarkers of hepatocarcinoma through liquid chromatography/mass spectrometry-based metabolomic method. *Anal Bioanal Chem*;401(6):1899-904.
  166. Huang Q, Tan Y, Yin P, Ye G, Gao P, Lu X, et al. Metabolic characterization of hepatocellular carcinoma using nontargeted tissue metabolomics. *Cancer Res* 2013;73(16):4992-5002.
  167. Zhou L, Wang Q, Yin P, Xing W, Wu Z, Chen S, et al. Serum metabolomics reveals the deregulation of fatty acids metabolism in hepatocellular carcinoma and chronic liver diseases. *Anal Bioanal Chem*;403(1):203-13.
  168. Yin P, Wan D, Zhao C, Chen J, Zhao X, Wang W, et al. A metabolomic study of hepatitis B-induced liver cirrhosis and hepatocellular carcinoma by using RP-LC and HILIC coupled with mass spectrometry. *Mol Biosyst* 2009;5(8):868-76.
  169. Xiao JF, Varghese RS, Zhou B, Nezami Ranjbar MR, Zhao Y, Tsai TH, et al. LC-MS based serum metabolomics for identification of hepatocellular carcinoma biomarkers in Egyptian cohort. *J Proteome Res* 2012;11(12):5914-23.
  170. Chen S, Yin P, Zhao X, Xing W, Hu C, Zhou L, et al. Serum lipid profiling of patients with chronic hepatitis B, cirrhosis, and hepatocellular carcinoma by ultra fast LC/IT-TOF

- MS. Electrophoresis 2013;34(19):2848-56.
171. Baniasadi H, Nagana Gowda GA, Gu H, Zeng A, Zhuang S, Skill N, et al. Targeted Metabolic Profiling of Hepatocellular Carcinoma and Hepatitis C using LC-MS/MS. *Electrophoresis* 2013.
  172. Resson HW, Xiao JF, Tuli L, Varghese RS, Zhou B, Tsai TH, et al. Utilization of metabolomics to identify serum biomarkers for hepatocellular carcinoma in patients with liver cirrhosis. *Analytica chimica acta* 2012;743:90-100.
  173. Liu SY, Zhang RL, Kang H, Fan ZJ, Du Z. Human liver tissue metabolic profiling research on hepatitis B virus-related hepatocellular carcinoma. *World J Gastroenterol*;19(22):3423-32.
  174. Fitian AI, Nelson DR, Liu C, Xu Y, Ararat M, Cabrera R. Integrated metabolomic profiling of hepatocellular carcinoma in hepatitis C cirrhosis through GC/MS and UPLC/MS-MS. *Liver international : official journal of the International Association for the Study of the Liver* 2014;34(9):1428-44.
  175. Warburg O, Wind F, Negelein E. The Metabolism of Tumors in the Body. *The Journal of general physiology* 1927;8(6):519-30.
  176. Nezami Ranjbar MR, Luo Y, Di Poto C, Varghese RS, Ferrarini A, Zhang C, et al. GC-MS Based Plasma Metabolomics for Identification of Candidate Biomarkers for Hepatocellular Carcinoma in Egyptian Cohort. *PLoS One* 2015;10(6):e0127299.
  177. Wu H, Xue R, Dong L, Liu T, Deng C, Zeng H, et al. Metabolomic profiling of human urine in hepatocellular carcinoma patients using gas chromatography/mass spectrometry. *Anal Chim Acta* 2009;648(1):98-104.
  178. Bowers J, Hughes E, Skill N, Maluccio M, Raftery D. Detection of hepatocellular carcinoma in hepatitis C patients: Biomarker discovery by LC-MS. *J Chromatogr B Analyt Technol Biomed Life Sci* 2014.
  179. Beyoglu D, Imbeaud S, Maurhofer O, Bioulac-Sage P, Zucman-Rossi J, Dufour JF, et al. Tissue metabolomics of hepatocellular carcinoma: tumor energy metabolism and the role of transcriptomic classification. *Hepatology* 2013;58(1):229-38.
  180. Tan Y, Yin P, Tang L, Xing W, Huang Q, Cao D, et al. Metabolomics study of stepwise hepatocarcinogenesis from the model rats to patients: potential biomarkers effective for small hepatocellular carcinoma diagnosis. *Mol Cell Proteomics*;11(2):M111 010694.
  181. Zeng J, Yin P, Tan Y, Dong L, Hu C, Huang Q, et al. Metabolomics Study of Hepatocellular Carcinoma: Discovery and Validation of Serum Potential Biomarkers by Using Capillary Electrophoresis-Mass Spectrometry. *J Proteome Res* 2014.
  182. Zhang A, Sun H, Yan G, Han Y, Ye Y, Wang X. Urinary metabolic profiling identifies a key role for glycocholic acid in human liver cancer by ultra-performance liquid-chromatography coupled with high-definition mass spectrometry. *Clin Chim Acta* 2013;418:86-90.
  183. Milic S, Stimac D. Nonalcoholic fatty liver disease/steatohepatitis: epidemiology, pathogenesis, clinical presentation and treatment. *Dig Dis* 2012;30(2):158-62.
  184. Masarone M, Federico A, Abenavoli L, Loguercio C, Persico M. Non Alcoholic Fatty Liver. *Epidemiology and Natural history. Rev Recent Clin Trials* 2014.
  185. Schuppan D, Schattenberg JM. Non-alcoholic steatohepatitis: pathogenesis and novel therapeutic approaches. *J Gastroenterol Hepatol* 2013;28 Suppl 1:68-76.
  186. Ahmed M. Non-alcoholic fatty liver disease in 2015. *World J Hepatol* 2015;7(11):1450-9.

187. Basaranoglu M, Ormeci N. Nonalcoholic fatty liver disease: Diagnosis, pathogenesis, and management. *Turk J Gastroenterol* 2014;25(2):127-32.
188. Matheus N, Hansen S, Rozet E, Peixoto P, Maquoi E, Lambert V, et al. An Easy, Convenient Cell and Tissue Extraction Protocol for Nuclear Magnetic Resonance Metabolomics. *Phytochem Anal* 2014.
189. Bedossa P, Poitou C, Veyrie N, Bouillot JL, Basdevant A, Paradis V, et al. Histopathological algorithm and scoring system for evaluation of liver lesions in morbidly obese patients. *Hepatology* 2012;56(5):1751-9.
190. Eguchi Y, Hyogo H, Ono M, Mizuta T, Ono N, Fujimoto K, et al. Prevalence and associated metabolic factors of nonalcoholic fatty liver disease in the general population from 2009 to 2010 in Japan: a multicenter large retrospective study. *J Gastroenterol* 2012;47(5):586-95.
191. Miyake T, Kumagi T, Furukawa S, Tokumoto Y, Hirooka M, Abe M, et al. Non-alcoholic fatty liver disease: Factors associated with its presence and onset. *J Gastroenterol Hepatol* 2013;28 Suppl 4:71-8.
192. Yin P, Peter A, Franken H, Zhao X, Neukamm SS, Rosenbaum L, et al. Preanalytical aspects and sample quality assessment in metabolomics studies of human blood. *Clin Chem* 2013;59(5):833-45.
193. Rosenling T, Slim CL, Christin C, Coulier L, Shi S, Stoop MP, et al. The effect of preanalytical factors on stability of the proteome and selected metabolites in cerebrospinal fluid (CSF). *J Proteome Res* 2009;8(12):5511-22.
194. Gika HG, Theodoridis GA, Wilson ID. Liquid chromatography and ultra-performance liquid chromatography-mass spectrometry fingerprinting of human urine: sample stability under different handling and storage conditions for metabolomics studies. *J Chromatogr A* 2008;1189(1-2):314-22.
195. Li H, Wang L, Yan X, Liu Q, Yu C, Wei H, et al. A proton nuclear magnetic resonance metabolomics approach for biomarker discovery in nonalcoholic fatty liver disease. *J Proteome Res*;10(6):2797-806.
196. da Silva RP, Kelly KB, Leonard KA, Jacobs RL. Creatine reduces hepatic TG accumulation in hepatocytes by stimulating fatty acid oxidation. *Biochim Biophys Acta* 2014;1841(11):1639-46.
197. Barcelos RP, Stefanello ST, Mauriz JL, Gonzalez-Gallego J, Soares FA. Creatine and the Liver: metabolism and possible interactions. *Mini Rev Med Chem* 2015.
198. Serkova NJ, Jackman M, Brown JL, Liu T, Hirose R, Roberts JP, et al. Metabolic profiling of livers and blood from obese Zucker rats. *J Hepatol* 2006;44(5):956-62.
199. Cobbold JF, Anstee QM, Goldin RD, Williams HR, Matthews HC, North BV, et al. Phenotyping murine models of non-alcoholic fatty liver disease through metabolic profiling of intact liver tissue. *Clin Sci (Lond)* 2009;116(5):403-13.
200. Araya J, Rodrigo R, Videla LA, Thielemann L, Orellana M, Pettinelli P, et al. Increase in long-chain polyunsaturated fatty acid n - 6/n - 3 ratio in relation to hepatic steatosis in patients with non-alcoholic fatty liver disease. *Clin Sci (Lond)* 2004;106(6):635-43.
201. Elizondo A, Araya J, Rodrigo R, Poniachik J, Csendes A, Maluenda F, et al. Polyunsaturated fatty acid pattern in liver and erythrocyte phospholipids from obese patients. *Obesity (Silver Spring)* 2007;15(1):24-31.
202. Puri P, Baillie RA, Wiest MM, Mirshahi F, Choudhury J, Cheung O, et al. A lipidomic analysis of nonalcoholic fatty liver disease. *Hepatology* 2007;46(4):1081-90.
203. Johnson NA, Walton DW, Sachinwalla T, Thompson CH, Smith K, Ruell PA, et al.

- Noninvasive assessment of hepatic lipid composition: Advancing understanding and management of fatty liver disorders. *Hepatology* 2008;47(5):1513-23.
204. Cobbold JF, Patel JH, Goldin RD, North BV, Crossey MM, Fitzpatrick J, et al. Hepatic lipid profiling in chronic hepatitis C: an in vitro and in vivo proton magnetic resonance spectroscopy study. *J Hepatol*;52(1):16-24.
205. Cabre E, Gassull MA. Polyunsaturated fatty acid deficiency in liver diseases: pathophysiological and clinical significance. *Nutrition* 1996;12(7-8):542-8.
206. Martinez-Granados B, Morales JM, Rodrigo JM, Del Olmo J, Serra MA, Ferrandez A, et al. Metabolic profile of chronic liver disease by NMR spectroscopy of human biopsies. *Int J Mol Med*;27(1):111-7.
207. Takahashi Y, Fukusato T. Histopathology of nonalcoholic fatty liver disease/nonalcoholic steatohepatitis. *World journal of gastroenterology : WJG* 2014;20(42):15539-48.
208. Waters NJ, Holmes E, Waterfield CJ, Farrant RD, Nicholson JK. NMR and pattern recognition studies on liver extracts and intact livers from rats treated with alpha-naphthylisothiocyanate. *Biochem Pharmacol* 2002;64(1):67-77.
209. Cartier V, Aube C. Diagnosis of hepatocellular carcinoma. *Diagn Interv Imaging* 2014.
210. Vander Heiden MG, Cantley LC, Thompson CB. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* 2009;324(5930):1029-33.
211. Ye G, Zhu B, Yao Z, Yin P, Lu X, Kong H, et al. Analysis of urinary metabolic signatures of early hepatocellular carcinoma recurrence after surgical removal using gas chromatography-mass spectrometry. *J Proteome Res* 2012;11(8):4361-72.
212. Zhou L, Liao Y, Yin P, Zeng Z, Li J, Lu X, et al. Metabolic profiling study of early and late recurrence of hepatocellular carcinoma based on liquid chromatography-mass spectrometry. *Journal of chromatography B, Analytical technologies in the biomedical and life sciences* 2014;966:163-70.
213. Bruix J, Sherman M, Llovet JM, Beaugrand M, Lencioni R, Burroughs AK, et al. Clinical management of hepatocellular carcinoma. Conclusions of the Barcelona-2000 EASL conference. European Association for the Study of the Liver. *J Hepatol* 2001;35(3):421-30.
214. Trevisani F, Magini G, Santi V, Morselli-Labate AM, Cantarini MC, Di Nolfo MA, et al. Impact of etiology of cirrhosis on the survival of patients diagnosed with hepatocellular carcinoma during surveillance. *Am J Gastroenterol* 2007;102(5):1022-31.
215. Fattovich G, Bortolotti F, Donato F. Natural history of chronic hepatitis B: special emphasis on disease progression and prognostic factors. *J Hepatol* 2008;48(2):335-52.
216. Benvegna L, Gios M, Boccato S, Alberti A. Natural history of compensated viral cirrhosis: a prospective study on the incidence and hierarchy of major complications. *Gut* 2004;53(5):744-9.
217. Kew MC. Interaction between hepatitis B and C viruses in hepatocellular carcinogenesis. *J Viral Hepat* 2006;13(3):145-9.
218. Shlomai A, de Jong YP, Rice CM. Virus associated malignancies: The role of viral hepatitis in hepatocellular carcinoma. *Semin Cancer Biol* 2014.
219. Nault JC. Pathogenesis of hepatocellular carcinoma according to aetiology. *Best Pract Res Clin Gastroenterol* 2014;28(5):937-47.
220. Triba MN, Le Moyec L, Amathieu R, Goossens C, Bouchemal N, Nahon P, et al.

- PLS/OPLS models in metabolomics: the impact of permutation of dataset rows on the K-fold cross-validation quality parameters. *Molecular bioSystems* 2014.
221. ter Voert EG, Heijmen L, van Laarhoven HW, Heerschap A. In vivo magnetic resonance spectroscopy of liver tumors and metastases. *World journal of gastroenterology : WJG* 2011;17(47):5133-49.
222. Wishart DS. Introduction to cheminformatics. *Curr Protoc Bioinformatics* 2007;Chapter 14:Unit 14 1.

# Résumé

---

Les atteintes hépatiques, asymptomatiques pour la plupart d'entre elles et pouvant évoluer vers des complications sévères telles que le carcinome hépatocellulaire (CHC) sont responsables de plus de 15 000 décès par an en France.

Le manque de marqueurs cliniques et biologiques fiables pour déterminer le degré de sévérité de l'hépatopathie ainsi que pour reconnaître les stades précoces du CHC constitue actuellement un obstacle majeur à une prise en charge optimale de la maladie.

Grâce aux approches de type métabolomique et aux techniques analytiques telles que la résonance magnétique nucléaire, il est désormais possible d'obtenir une véritable cartographie des métabolites d'un individu.

L'objectif de ce travail a été d'explorer, par une approche RMN métabolomique, les changements métaboliques dans le foie et dans le sérum causés par différentes pathologies hépatiques afin de proposer de nouvelles pistes dans l'amélioration du diagnostic et de la prise en charge de ces maladies.

Une attention particulière a également été donnée à l'étude de la validité des paramètres de qualité des modèles de discrimination réalisés lors des analyses statistiques des données multivariées.

**Mots-clés :** Métabolomique, RMN, analyses multivariées, carcinome hépatocellulaire, NAFLD, biomarqueurs

# Abstract

---

Most liver diseases nowadays remain symptomless and tend to lead to hepatocellular carcinoma responsible for more than 15.000 patient deaths per year in France. Liver diseases are therefore a major concern for public health.

Clinicians lack of non-invasive biomarkers allowing them to enhance identification of liver diseases stages in order to efficiently target the first HCC signs and accordingly improve clinical prognosis.

Identification of new biomarkers set new challenges in translational research in order to refine the prognosis and adapt therapeutic procedures.

Proton nuclear magnetic resonance spectroscopy-based metabolomics enable to identify and quantify such metabolites by defining individual metabolic fingerprints.

First part of this work was to explore the metabolic modifications of liver tissue to further establish diseases stages profiles.

Second part was focused on the assessment of metabolic variations in HCC patients, by analyzing sequential serums taking, before and after a radiofrequency ablation curative treatment.

Third and last part was centered on the validation of the quality parameters of the discriminant models used in multivariate statistical analysis.

**Keywords :** Metabolomics, NMR, multivariate analysis, hepatocellular carcinoma, NAFLD, biomarkers