



HAL
open science

High-Field NMR Metabolomics : Phenotyping the Metabolic Complexity from Humans to Cells

Clément Pontoizeau

► **To cite this version:**

Clément Pontoizeau. High-Field NMR Metabolomics : Phenotyping the Metabolic Complexity from Humans to Cells. Other. Ecole normale supérieure de lyon - ENS LYON, 2012. English. NNT : 2012ENSL0786 . tel-01124044

HAL Id: tel-01124044

<https://theses.hal.science/tel-01124044>

Submitted on 6 Mar 2015

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

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

THÈSE

en vue de l'obtention du grade de

Docteur de l'École Normale Supérieure de Lyon - Université de Lyon

Discipline : Chimie

Laboratoire : Centre de RMN à très hauts champs
Institut des Sciences Analytiques
Ecole Normale Supérieure de Lyon

École Doctorale de Chimie de Lyon

présentée et soutenue publiquement le 12 Décembre 2012

par Mr. Clément PONTOIZEAU

High-Field NMR Metabolomics: Phenotyping the Metabolic Complexity from Humans to Cells

Directeur de thèse : Prof. Pierre Toulhoat

Après l'avis de : Prof. Jean-Charles PORTAIS

Dr. Julian L. Griffin

Devant la commission d'examen formée de :

Dr. Bénédicte ELENA-HERRMANN, Université de Lyon, Membre

Dr. Julian L. GRIFFIN, Université de Cambridge, Rapporteur

Dr. Bruno MOUGIN, Centre Biomérieux, Membre

Prof. Jean-Charles PORTAIS, Université Paul Sabatier, Toulouse, Rapporteur

Prof. Pierre TOULHOAT, Université de Lyon, Directeur

Prof. Paolo VINEIS, Imperial College, Londres, Membre

Acknowledgements

Je souhaiterais tout d'abord remercier le Professeur Jean-Charles Portais et le Docteur Julian Griffin pour avoir accepté d'être les rapporteurs de ce manuscrit de thèse ainsi que le Professeur Paolo Vineis et le Docteur Bruno Mougin pour avoir accepté d'être membres du jury.

Je remercie les professeurs Pierre Toulhoat et Lyndon Emsley pour m'avoir offert l'opportunité de mener ce doctorat au CRMN et de m'avoir fait confiance dans la conduite du double cursus thèse-médecine. Je les remercie pour leurs conseils, leurs critiques et leurs idées au sujet des projets menés et pour la liberté laissée dans leur réalisation.

Je souhaiterais également remercier les doyens des facultés de médecine Lannec et Lyon Est, les professeurs Cochat et Etienne pour avoir accepté et soutenu, sur le versant médical, l'idée de ce double cursus médecine-sciences et d'en avoir favorisé la réalisation.

Je tiens tout particulièrement à remercier le Docteur Bénédicte Elena-Herrmann pour son encadrement au cours de cette thèse, ses conseils, sa disponibilité et pour la relecture de ce manuscrit. Merci Bénédicte.

Je remercie le Docteur Marc-Emmanuel Dumas pour avoir encadré mes premiers pas dans le domaine.

Je remercie Benjamin pour m'avoir montré la voie de ce double cursus, mais également pour l'aide apportée au début de ce doctorat, pour avoir accepté de travailler ensemble sur certains de ses projets et pour continuer à me faire partager son expérience.

Je remercie le Docteur Torsten Herrmann pour son aide précieuse en programmation, en particulier pour le programme UNIO et ses conseils avisés.

Je remercie l'ensemble de nos collaborateurs qui ont permis à tous ces projets de voir le jour et nous ont fait bénéficier de leurs expertises : Docteur Olivier Tredan, Docteur Thomas Bachelot, Docteur Mazda Jenab, Docteur Pierre Hainaut, Docteur Andre Da Costa, Yayun Dai, Professeur Dominique Gauguier, Professeur Robert Barouki, Docteur Xavier Coumoul, Docteur Linh-Chi Bui, Docteur Florence Solari, Docteur Laurent Mouchiroud, Nicolas Dallièrre et le Docteur Laurent Mollin.

Je remercie l'Institut National du Cancer pour le financement de la quatrième année de thèse.

Je souhaite remercier l'ensemble des membres du CRMN pour leur aide et les discussions et la bonne ambiance générale, même si nous n'avons pas eu l'occasion de travailler ensemble : Claire, James, Michele, Emeline, Stephan, Megan, Aaron, Maria, Andy, Michael, Alexandre, Corey, Emmanuel, Audrey, Laetitia, Moreno, Guido et Anne.

Je remercie également les anciens membres du CRMN que j'ai pu côtoyer au cours de ce doctorat: Gwendal, Robin, Julien, Elodie, Ségolène, Paul, Joseph, Marc, Amy et Mathilde.

Je tiens à remercier Vincent, Laetitia, Céline pour les travaux que nous avons pu mener ensemble pendant les premières années de ce doctorat et pour m'avoir fait partager leurs expertises sur les différents domaines de la métabolomique et de l'interprétation des données.

Je remercie Alessandro, qui a partagé le même carré de bureaux pendant ces quatre ans, pour ses conseils, ses connaissances historiques toujours très précises ainsi que sa disponibilité le soir et le week-end pour venir travailler au laboratoire.

Je tiens à remercier tout particulièrement Elodie, Anne et Andrei avec lesquels j'ai été amené à travailler sur différents projets, de manière très fructueuse et très agréable. Je vous remercie également pour votre disponibilité et votre aide précieuse.

Je tiens également à remercier Jean-Nicolas pour les discussions que nous avons pu avoir au cours de ces années, qui ont pu éclairer ou questionner notre quotidien.

Je remercie mes amis de l'ENS, en particulier Antoine et Gabriel ainsi que mes amis de Michelet pour leur présence et leur soutien indéfectible, malgré les périodes de silence radio.

Je remercie Gaëlle, Nicolas, Amélie et Jean-Nicolas qui ont partagé le quotidien de ces années avant de s'envoler vers d'autres horizons.

Je remercie ma famille pour son soutien total au cours de ces différentes années, mes grands parents, Maryvonne et Jacques, Raymonde et Pierre, mes parents Jacqueline et Philippe, mon oncle Jean-Paul et Véronique, ma sœur Lucile et mon frère Gabriel.

Je tiens enfin à remercier Mathilde pour sa patience et son soutien inestimable.

Résumé

La métabolomique a été introduite comme la dernière des approches “omics” dont l’objet est de caractériser l’ensemble des métabolites d’un système biologique et leurs variations en réponse à un facteur d’intérêt, comme une maladie, une perturbation génétique ou environnementale, d’un point de vue global. Cette approche exploite les derniers développements de méthodes analytiques telles que la spectroscopie par Résonance Magnétique Nucléaire (RMN) ou la spectrométrie de masse. Elle a été utilisée dans un grand nombre de domaines scientifiques, de la biologie moléculaire à la recherche de marqueurs biologiques de pathologies humaines, en tant que nouvel outil de caractérisation du métabolisme.

Cette thèse est dédiée aux développements méthodologiques et applications de la métabolomique exploitant la RMN à très hauts champs. Nous caractérisons une grande variété d’échantillons biologiques, des fluides biologiques humains aux organismes modèles comme le vers *Caenorhaditis elegans* (*C. elegans*) ou la lignée cellulaire HepG2, dans le cadre de problématiques très différentes. Nous soulignons les points communs et les spécificités de l’utilisation de la métabolomique dans ces différents contextes.

La première partie de ce manuscrit est dédiée à une présentation générale de la métabolomique. Nous décrivons les principes de cette approche, introduisons l’usage de la

RMN dans ce contexte et présentons l'analyse des données. Nous décrivons également les résultats obtenus concernant l'introduction d'une technique à dimensionnalité réduite pour la caractérisation des mélanges complexes, dénommée spectroscopie RMN par projections ciblées, dont l'objectif est de réduire le recouvrement entre les signaux des métabolites afin de permettre une identification non ambiguë. Nous introduisons cette approche dans le cadre d'un mélange complexe modèle en enregistrant une projection optimisée d'un spectre à 3 dimensions ^1H - ^1H - ^{13}C TOCSY-HSQC.

La seconde partie de ce manuscrit décrit les résultats de trois études métabolomiques portant sur des populations humaines. La première analyse démontre que les échantillons de sérum collectés dans le cadre de la cohorte européenne prospective internationale EPIC sont appropriés pour une étude métabolomique, par la bonne qualité des spectres RMN obtenus, l'identification réussie de 10 paires de doublons en aveugle et l'évaluation de différences entre pays de collecte. La seconde étude recherche une signature métabolique dans le sérum du cancer du sein métastatique par comparaison des profils sériques de patientes souffrant de cancers du sein localisés et métastatiques. La dernière analyse établit une signature plasmatique potentielle pour différentes pathologies hépatiques comme le carcinome hépatocellulaire (CHC) et la maladie chronique du foie, à partir d'échantillons collectés en Thaïlande. Nous montrons que la fonction hépatique, évaluée à partir de la concentration d'albumine, a un impact très important sur les profils métaboliques plasmatiques mais ne suffit pas à elle seule à expliquer la signature métabolique du CHC. Nous étudions également la corrélation entre profils métaboliques et infection par les virus de l'hépatite B ou C, et la corrélation avec les niveaux de LTBP2, nouveau marqueur potentiel de CHC.

La troisième partie de cette thèse est dédiée à la caractérisation d'organismes modèles variés dans le cadre de différentes questions biologiques. La première étude présente la caractérisation des différences métaboliques plasmatiques et urinaires entre quatre souches de rats, Fisher, Lewis, Wistar Kyoto and Brown Norway couramment utilisées comme contrôles en génétique, par RMN du liquide. Dans la seconde analyse, nous étudions les effets du vieillissement physiologique chez *C. elegans* et observons que le processus de restriction alimentaire, qui augmente la durée de vie d'un grand nombre d'organismes, tamponnent les modifications métaboliques associées au vieillissement, à travers la caractérisation de deux mutants de *C. elegans*, *slcf-1* et *eat-2*. Nous observons également que des perturbations du métabolisme de la phosphocholine corrélerent avec l'espérance de vie. La troisième étude caractérise un mutant de *C. elegans*, pour le gène *ahr-1* avec perte de fonction, codant pour un orthologue du récepteur AHR chez les mammifères, médiateur principal de la toxicité de la

dioxine. Nous observons de modifications métaboliques importantes chez ce mutant qui suggèrent une implication de ce gène dans le développement et le vieillissement de *C. elegans*. Enfin, nous étudions dans la dernière analyse les effets au niveau métabolique de l'interaction entre la protéine endogène E4F1 et la protéine virale HBx dans des cellules hépatiques infectées par le virus de l'hépatite B et suggérons que la protéine E4F1 équilibre les perturbations du métabolisme oxydatif induites par HBx, améliorant ainsi la tolérance des cellules hépatiques.

Abstract

Metabolomics has been introduced as the last of the “omics” approaches, aiming at delineating the ensemble of metabolites, low molecular weight molecules and changes in metabolite concentrations of biological systems in a holistic way, in response to specific stimuli such as diseases, genes or environmental perturbations. Relying on state-of-the-art analytical technologies, such as Nuclear Magnetic Resonance (NMR) spectroscopy and mass spectrometry (MS), this approach has been applied as a new readout casting light on metabolism, for a wide range of scientific fields, from cell biology to disease biomarker discovery in humans.

This thesis is dedicated to developments and applications of metabolomics, exploiting high field NMR spectroscopy. We characterized a wide range of biological samples, from human biofluids to *Caenorhaditis elegans* (*C. elegans*) and cell cultures, addressing various issues. We highlight the similarities and specificities in the use of metabolomics in these different contexts.

The first part is dedicated to a general presentation of metabolomics. We describe the principle of this approach, introduce the use of NMR spectroscopy in this context and present data analysis. We also report results about the introduction of reduced dimensionality

techniques for the characterization of complex mixtures, coined targeted projection NMR spectroscopy aiming at reducing metabolite signal overlaps and solve assignment ambiguities. We introduce this approach in a model complex mixture, recording an optimized projection of a 3D ^1H - ^1H - ^{13}C TOCSY-HSQC experiment.

The second part of this manuscript reports results about three different metabolomic studies carried out in human populations. The first analysis demonstrates the suitability for metabolomics of serum samples collected in the framework of the European Prospective Investigation into Cancer and Nutrition (EPIC) study by assessing the good quality of NMR spectra, identifying in a blind analysis 10 pairs of duplicates and evaluating metabolic differences between countries of sample origin. The second study investigates a serum metabolic signature of metastatic breast cancer by comparing serum metabolic profiles of patients suffering from localized and metastatic breast cancer. The last analysis establishes potential plasma metabolic signatures for different liver pathologies, hepatocellular carcinoma (HCC) and chronic liver disease (CLD), from samples collected in Thailand. We show that liver function through the measure of albumin levels has a strong impact on plasma metabolic profiles but can not alone explain the HCC metabolic signature. We investigate the correlation between plasma metabolic profiles and infection by hepatitis B (HBV) and C (HCV) viruses as well as LTPB2, a proteomic marker of HCC.

The third part of this thesis is dedicated to the characterization of various model organisms to address different types of biological questions. The first study presents a characterization of plasma and urine metabolic differences between four rat strains, Fisher, Lewis, Wistar Kyoto and Brown Norway, commonly used as controls in genetic studies, using liquid-state NMR spectroscopy. In the second study, we investigate the effects of physiological aging in *C. elegans* and observe that dietary restriction (DR), a process which increases lifespan buffers metabolic changes associated with aging through the characterization of two *C. elegans* DR mutants, *slcf-1* and *eat-2*. We further identify that perturbations in phosphocholine metabolism correlate with life expectancy. The third analysis of this part characterizes of *ahr-1* loss-of-function *C. elegans* mutant, a gene coding for an ortholog of the Aryl hydrocarbon receptor (AhR) in mammals, which mediates dioxin toxicity. We observe strong metabolic changes in *ahr-1* mutants, which suggest an involvement in development and aging processes, requiring further evaluation. We finally investigate in the last study the effects at the metabolic level of the interaction between an endogenous protein E4F1 and a viral protein HBx in liver cells infected by hepatitis B virus

and suggest that E4F1 balances cell oxidative metabolism perturbations induced by HBx, further enhancing tolerance of liver cells.

Table of Contents

| | |
|---|----|
| Acknowledgements | 3 |
| Résumé | 7 |
| Abstract | 11 |
| Table of Contents | 15 |
| Introduction | 19 |
| Part 1: The metabolomic approach | 23 |
| 1.1 Introduction to metabolomics | 23 |
| 1.2 NMR spectroscopy in metabolomics | 27 |
| 1.3 NMR of complex mixtures: towards targeted projection NMR spectroscopy | 33 |
| 1.3.1 Introduction | 33 |
| 1.3.2 Materials and methods | 34 |
| 1.3.3 Results and discussion..... | 37 |
| 1.3.4 Conclusion..... | 43 |
| 1.4 Data analysis in metabolomics | 45 |
| Part 2: Metabolomics for human population studies..... | 49 |
| 2.1 Assessment of the exposome: the EPIC cross-sectional pilot study | 51 |

| | | |
|---------|--|-----|
| 2.1.1 | Introduction | 51 |
| 2.1.2 | Material and methods (see Part 4 for ICL methods) | 52 |
| 2.1.3 | Results and discussion..... | 53 |
| 2.2 | Profiling systemic signatures of cancers: metastatic breast cancer serum signature | 57 |
| 2.2.1 | Introduction | 57 |
| 2.2.2 | Materials and methods | 58 |
| 2.2.3 | Results | 61 |
| 2.2.4 | Discussion | 64 |
| 2.2.5 | Conclusion..... | 67 |
| 2.3 | Plasma metabolic signature of hepatocellular carcinoma and chronic liver disease. | 69 |
| 2.3.1 | Introduction | 69 |
| 2.3.2 | Material and Methods | 70 |
| 2.3.3 | Results | 72 |
| 2.3.4 | Discussion | 81 |
| 2.3.5 | Conclusion..... | 84 |
| 2.4 | Conclusion..... | 85 |
| Part 3: | Probing model organisms metabolism | 87 |
| 3.1 | Natural metabotype variations in healthy control inbred rat strains | 90 |
| 3.1.1 | Introduction | 90 |
| 3.1.2 | Methods..... | 91 |
| 3.1.3 | Results | 95 |
| 3.1.4 | Discussion | 105 |
| 3.1.5 | Conclusion..... | 111 |
| 3.2 | Metabolic phenotyping of <i>Caenorhabditis elegans</i> | 112 |
| 3.2.1 | Dietary restriction buffers metabolic changes associated with aging in <i>Caenorhabditis elegans</i> | 114 |
| 3.2.2 | Metabolic phenotyping of <i>ahr-1(ia03)</i> <i>C. elegans</i> mutant..... | 129 |
| 3.3 | Metabolic phenotyping of HBV-infected liver cells | 137 |
| 3.3.1 | Introduction | 137 |
| 3.3.2 | Materials and methods | 138 |
| 3.3.3 | Results and discussion..... | 139 |
| 3.3.4 | Conclusion..... | 145 |

| | |
|--|-----|
| 3.4 Conclusion..... | 146 |
| Conclusion..... | 147 |
| Part 4: Appendixes | 151 |
| 4.1 On EPIC cross-sectional pilot study: Methods for UPLC-MS and 600 MHz NMR spectroscopy | 152 |
| 4.2 On metastatic breast cancer serum signature | 157 |
| 4.3 On plasma metabolic signature of hepatocellular carcinoma and chronic liver disease. | 163 |
| 4.4 Natural metabotype variations in healthy control inbred rat strains | 165 |
| 4.5 On metabolic phenotyping of HBV-infected liver cells | 170 |
| 4.6 Personal contribution to the different projects | 171 |
| References | 173 |
| Communications | 191 |
| Publications | 191 |
| Posters | 192 |
| Oral presentations..... | 192 |

Introduction

Over the past two decades, technological developments have driven the introduction of new tools and concepts for the study of biological systems, gathered under the generic term of “omics” approaches. These approaches aim at describing the whole content of a biological system in a global and unbiased manner, at a defined molecular level. Genomics intends to delineate the complete assembly of genes, transcriptomics and proteomics of transcripts or mRNA and proteins respectively. Their use allows the investigation of the global response of an organism at the system level to a particular event like a disease, deciphering complex interaction networks.

In this framework, metabolomics has been introduced as the study of the ensemble of metabolites, low molecular weight molecules, and of changes in metabolite concentrations of biological systems in a holistic way, in response to specific stimuli such as diseases, genes or environmental perturbations. Metabolomics relies on state-of-the-art analytical devices, such as Nuclear Magnetic Resonance (NMR) spectroscopy and mass spectrometry (MS) to profile metabolic compositions of biological samples. This approach has been applied as a new readout casting light on metabolism, from microbiology, toxicology, ecology, plant biology to biomedical studies for biomarker discovery or understanding of disease pathology.

This thesis is dedicated to developments and applications of metabolomics, exploiting high field NMR spectroscopy. Various types of biological samples have been characterized from human biofluids to *Caenorhabditis elegans* (*C. elegans*) worms and cell cultures to address different questions from investigating blood metabolic signature of human cancers to deciphering metabolic changes of *C. elegans* during aging. We highlight the similarities and specificities in the use of metabolomics in these different contexts.

The first part is dedicated to a general presentation of metabolomics and results about the introduction of targeted projection NMR spectroscopy for the study of complex mixtures. The first section of this part presents general principles of the metabolomic approach, highlights the fields of applications and gives a general picture of metabolomic studies. The second section focuses on the use of NMR spectroscopy for metabolomics. We introduce the different devices like High Resolution Magic Angle Spinning (HR-MAS) and NMR experiments necessary to obtain well-resolved metabolic profiles for liquid and semi-solid samples such as tissues or intact *C. elegans* worms and help metabolite identification. However, assignment of metabolites without ambiguities could be challenging. In this context, we report in the third section results about the introduction of reduced dimensionality techniques for the characterization of complex mixtures, coined targeted projection NMR spectroscopy. For complex mixtures with high signal overlaps, where metabolite assignment is ambiguous even in standard 2D experiments but possible in higher dimensional spectra, the principle is to record an optimized projection of a spectrum of high dimension to capture the relevant information necessary to solve the assignment issue in a small amount of time. We introduce this approach in a model complex mixture, recording a projection of a 3D ^1H - ^1H - ^{13}C TOCSY-HSQC experiment. Eventually, the fourth section of this first part describes the different steps of data analysis and focuses on the biostatistical tools used in this manuscript.

The second part of this manuscript reports results about a range of metabolomic studies carried out in human populations, from blood samples analyzed by liquid-state NMR spectroscopy for epidemiological or clinical purposes. The first section evaluates the suitability for metabolomics of serum samples collected in a large biobank between 1992 and 1998, in the framework of the European Prospective Investigation into Cancer and Nutrition (EPIC) study, designed to investigate the association between nutrition and cancer. In this pilot study, we analyze 50 samples including 10 pairs of duplicates, originating from three different countries. We assess the good quality of NMR spectra, identify in a blind analysis the 10 pairs of duplicates and evaluate differences between countries, demonstrating that these serum samples are appropriate for metabolomic analyses. Collaborators perform the

same experiments in parallel with mass spectrometry. The second section reports investigation about a serum metabolic signature of metastatic breast cancer. We compare the serum metabolic profiles of patients suffering from localized and metastatic breast cancer to characterize the metabolic differences between these two groups. We confirm these results with an independent validation set. The last section of this second part is dedicated to the analysis of plasma metabolic signatures of different liver pathologies like hepatocellular carcinoma (HCC) and chronic liver disease (CLD), from samples collected in Thailand, an area of high hepatitis B incidence. We investigate the effects of a range of biological and experimental factors on plasma metabolic profiles. We derive specific signatures for HCC and CLD, as well as common features by comparison to healthy subjects. We show that liver function through the measure of albumin levels has a strong impact on plasma metabolic profiles. However, we show that the signature of HCC cannot be reduced to this parameter. We investigate the correlation between plasma metabolic profiles and infection by hepatitis B (HBV) and C (HCV) viruses as well as LTPB2, a proteomic marker of HCC.

The third part of this thesis is dedicated to the characterization of various model organisms to address different types of biological questions. The first section presents a characterization of plasma and urine metabolic differences between four rat strains, Fisher, Lewis, Wistar Kyoto and Brown Norway, commonly used as controls in genetic studies, using liquid-state NMR spectroscopy. The following sections describe analyses carried out on small model organisms, *C. elegans* worms and HepG2 cell lines by HR-MAS NMR spectroscopy. In the second section, we investigate the effects of physiological aging in *C. elegans*. We then characterize the changes in worm metabolism during dietary restriction (DR), a process which increases lifespan, by the analysis of two long-lived worm mutants, *slcf-1* and *eat-2*, models of DR. We observe that DR buffers metabolic changes associated with aging. Double mutants carrying a mutation that suppresses the long-lived phenotype allow us to identify that perturbations in phosphocholine metabolism specifically correlate with life expectancy. The third section presents results about the characterization of *C. elegans* mutant for the gene *ahr-1*, undergoing a loss-of-function, by comparison to wild-type counterparts. The AHR-1 protein is an ortholog of the Aryl hydrocarbon receptor (AhR) in mammals, which mediates dioxin toxicity. The AHR-1 function in *C. elegans* is not completely understood, while an important endogenous role is expected. We observe strong metabolic changes in *ahr-1* mutants, which suggest an involvement in development and aging processes, requiring further evaluation. In the last section of this third part we investigate the effects at the metabolic level of the interaction between an endogenous protein E4F1 and a viral protein HBx in liver cells

infected by hepatitis B virus. It has been shown that this interaction enhances the tolerance of liver cells towards HBV infection. We suggest that E4F1 balances cell oxidative metabolism perturbations induced by HBx, further enhancing tolerance of liver cells.

Part 1: The metabolomic approach

1.1 Introduction to metabolomics

Since the beginning of the 21st century, metabolism has known a renewed interest through the development of metabolomics, with a reevaluation of its biological significance and potential applications.^[1] Metabolism gathers the relationships between metabolites, low-molecular weight (< 1.5 kDa) organic or inorganic compounds involved in enzyme-mediated biochemical reactions.^[2] The term metabolome was coined in 1998 as the quantitative complement of metabolites in a biological system.^[3] Different terms were introduced, referring to the study of metabolomes and changes in metabolite concentrations of biological systems in a holistic way, in response to specific stimuli such as diseases, genes or environmental perturbations. Metabolomics was originally defined as the comprehensive analysis of all metabolites present in a biological system,^[4] while metabonomics was introduced as “the quantitative measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification”.^[5] The difference between these two terms is essentially historical, originated respectively from microbial and plant studies by mass spectrometry and mammalian studies by Nuclear

Magnetic Resonance (NMR) spectroscopy and can now be considered as synonymous in practice.^[2,6] In this thesis we will only use the term metabolomics.

The idea of analyzing metabolism in a global manner to get insight into physiological processes dates back to the late 1960s, with the development of gas-liquid-chromatographic procedures to study biofluid metabolite compositions.^[7-10] In 1971, Pauling *et al.* conceived the idea that a quantitative analysis of metabolite body fluids composition could provide information-rich data reflecting individual status and permit disease diagnosis,^[10] while Horning *et al.* independently introduced the notion of metabolic profile.^[8]

In the late 1990s, metabolism studies knew a new enthusiasm, capitalizing on progresses in analytical technologies and the development of “omics” approaches. Indeed the complete sequencing of organism genetic information,^[11,12] coined genome, paved the way for “omics” approaches, which consist in delineating the whole content of biological systems for a given level of description, to depict global systemic behaviors. Genomics was introduced for the study of genetic information and genomes, as well as transcriptomics for the analysis of mRNA (or transcripts) and transcriptome^[13] and proteomics, focusing on the ensemble of proteins or proteome.^[14] In this context, metabolomics has emerged as the last and complementary “omics” approach, dedicating to the study of metabolites. Further integration of these different levels has become the core of systems biology, aiming at deciphering complex interactions between components of biological systems.

Metabolomics relies on state-of-the-art analytical techniques to profile metabolite composition of biological samples. Nuclear Magnetic Resonance (NMR) spectroscopy and mass spectrometry (MS) have been widely used to achieve this goal. Mass spectrometry is often coupled to chromatographic (gas-chromatography or liquid chromatography) or electrophoretic (capillary electrophoresis) separation devices prior to detection.^[2] The choice of the separation and detection devices defines the set of detectable metabolites. We can further distinguish untargeted approaches, without a priori selection of detected metabolites and targeted approaches, optimized to identify and accurately quantify a pre-defined set of metabolites.^[15] Up to now, none of these systems offers a complete coverage of the metabolome. NMR spectroscopy and the different MS systems could thus be used in a complementary manner to increase the number of detected compounds.^[16] Protocols have been developed to study a wide range of mammals body fluids such as plasma or serum,^[17, 18] urine,^[17, 19, 20] amniotic fluid,^[21] bile,^[22] cerebral spinal fluid,^[22] saliva^[23] or follicular fluids.^[24] NMR spectroscopy, through High-Resolution Magic Angle Spinning (HR-MAS) technology offers the possibility to study intact tissues,^[25] cell cultures^[26] or small model organisms.^[27]

Mass spectrometry and liquid-state NMR spectroscopy can also be used to study these biological samples following an extraction step.^[17, 28-30] The development of these analytical methods allows now the study of almost all types of biological compounds.

Metabolomics has now been applied in a wide range of scientific fields, as a new readout casting light on metabolism, from microbiology,^[31] toxicology,^[32] ecology,^[33] plant biology^[4] to biomedical studies for biomarker discovery or understanding of disease pathology. Historically, metabolomics has been successfully applied in clinic for the screening of inborn errors of metabolism, where metabolic perturbations constitute by definition the pathological process.^[15] Since that, the potential of metabolomics has been investigated in the other fields of medicine,^[34] with pilot studies launched in the clinic.^[35] The use of large cohorts of patients has been pioneered for biomarker discovery. Holmes *et al.* for instance showed correlations between high blood pressure, a set of urinary metabolites and geographical origins of patients,^[36] while Wang *et al.* observed an association between branched-chain and aromatic amino acid profiles and future development of diabetes.^[37] Beyond the study of pathological states, Suhre *et al.* investigated the links between genetics and serum or urine metabolic profiles.^[38, 39] Furthermore, Clayton *et al.* demonstrated that metabolic profiles determined before drug administration could predict metabolic response and liver toxicity of acetaminophen in rats, a concept termed pharmaco-metabonomics.^[40] Metabolomics identified the potential role of sarcosine and glycine in cancerous processes from cell lines.^[41, 42]

Metabolomic analyses are built as case-control studies, which consist in comparing two populations of samples coming from two relevant physio-pathological states. For instance one of the group gathers individuals suffering from a particular disease while the other one gathers healthy individuals used as controls. Comparing these two groups could give access to markers of the disease under study. Analytical devices derive a metabolic profile for each sample (**Figure 1.1.1a**). These metabolic profiles are then processed, normalized and scaled (**Figure 1.1.1b**). Different statistical tools such univariate tests or multivariate analyses, are then used to extract the significant differences in metabolite levels between groups defining a metabolic signature (**Figure 1.1.1c**).

The next section introduces the use of NMR spectroscopy in this context of metabolomic studies. In the third section of this part, we report results about the introduction of reduced dimensionality techniques for the characterization of complex mixtures, coined targeted projection NMR spectroscopy aiming at reducing metabolite signal overlaps and

solve assignment ambiguities. Eventually, the fourth section describes the different steps of data analysis and focuses on the biostatistical tools used in this manuscript.

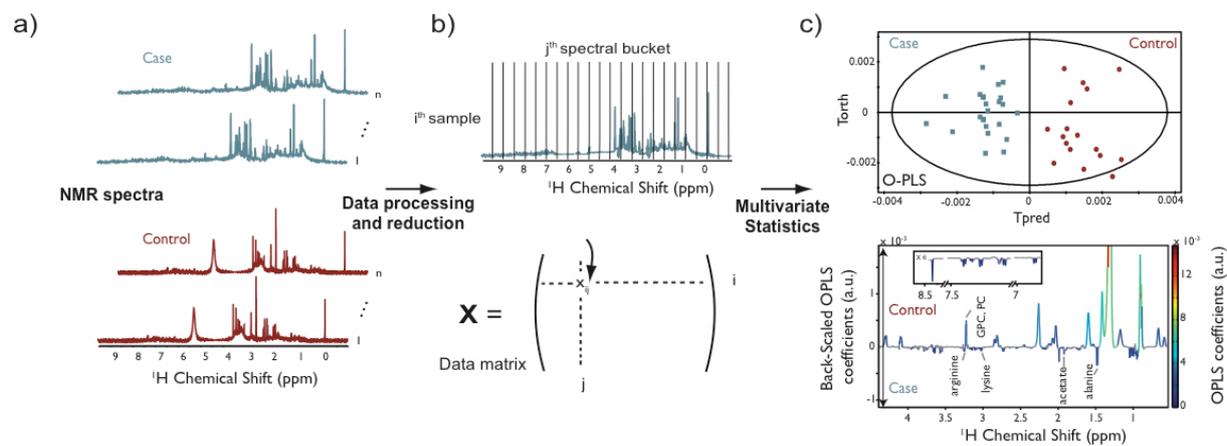


Figure 1.1.1. Principle of the NMR metabolomic approach: a) metabolic profiles are recorded; b) processed and reduced; and c) analyzed with multivariate statistical methods.

1.2 NMR spectroscopy in metabolomics

Nuclear Magnetic Resonance (NMR) spectroscopy is a quantitative and non-destructive technique widely used in chemistry, which provides detailed information on the molecular structure of compounds and on complex mixture compositions. It has become a technique of reference with mass spectrometry for metabolomics studies. NMR spectroscopy can detect various nuclei such as ^1H , ^{13}C , ^{31}P or ^{15}N according to the type of NMR experiment. In the context of biological sample analysis, ^1H detection is essentially used due to both the high abundance of this nucleus in organic molecules and to the high ^1H sensitivity for NMR. Each ^1H of a molecule creates a specific ^1H NMR signal which shape (or multiplicity) or coordinates along an axis of energy termed ^1H chemical shift are defined by the chemical environment of the ^1H nucleus. The ^1H NMR spectrum for a chemical compound is thus the sum of the signals of its different ^1H nuclei and for a compound mixture, the sum of the ^1H NMR spectra of the different compounds. In a biological sample, ^1H NMR spectroscopy detects thus without a priori all the small molecules containing ^1H nuclei reaching the detection limit, in an untargeted manner. The ^1H NMR spectrum of a urine sample is thus the superimposition of the contributions of different metabolites such as amino-acids or organic acids creating a metabolic fingerprint (**Figure 1.2.1**).

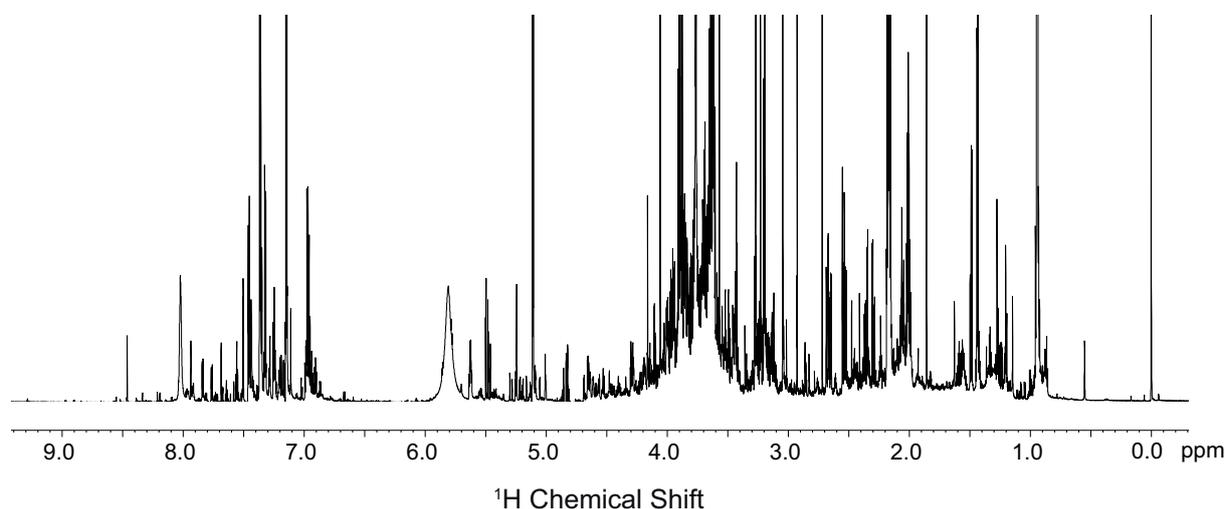


Figure 1.2.1. First 1GHz ^1H NOESY NMR spectrum of urine sample, recorded on our 1GHz.

NMR spectroscopy allows the characterization of liquid samples as well as semi-solid biological specimens, such as tissue biopsies, culture cells or small model organisms like the worm *Caenorhabditis elegans* (*C. elegans*).^[26, 43, 44] However, liquid-state NMR spectroscopy is not directly applicable for these samples due to the presence of anisotropic interactions

causing unresolved and useless spectra (**Figure 1.2.2a**). Magic Angle Spinning technique, first introduced by Andrew *et al.*^[45] and Lowe^[46] for the study of solids, can be applied for these samples. Samples are spun at an angle of 54.7° with respect to the static magnetic field, coined the magic angle, in order to average the main anisotropic interactions such as dipolar interactions, variation in magnetic susceptibility and chemical shift anisotropy, and obtain well-resolved spectra (**Figure 1.2.2b**). Cheng *et al.* introduced the first application of MAS for the study of biological tissues, termed High Resolution Magic Angle Spinning (HRMAS).^[47] Low spinning speeds are typically applied between 3.5 kHz and 5 kHz to obtain well-resolved spectra and minimize sample disruption.

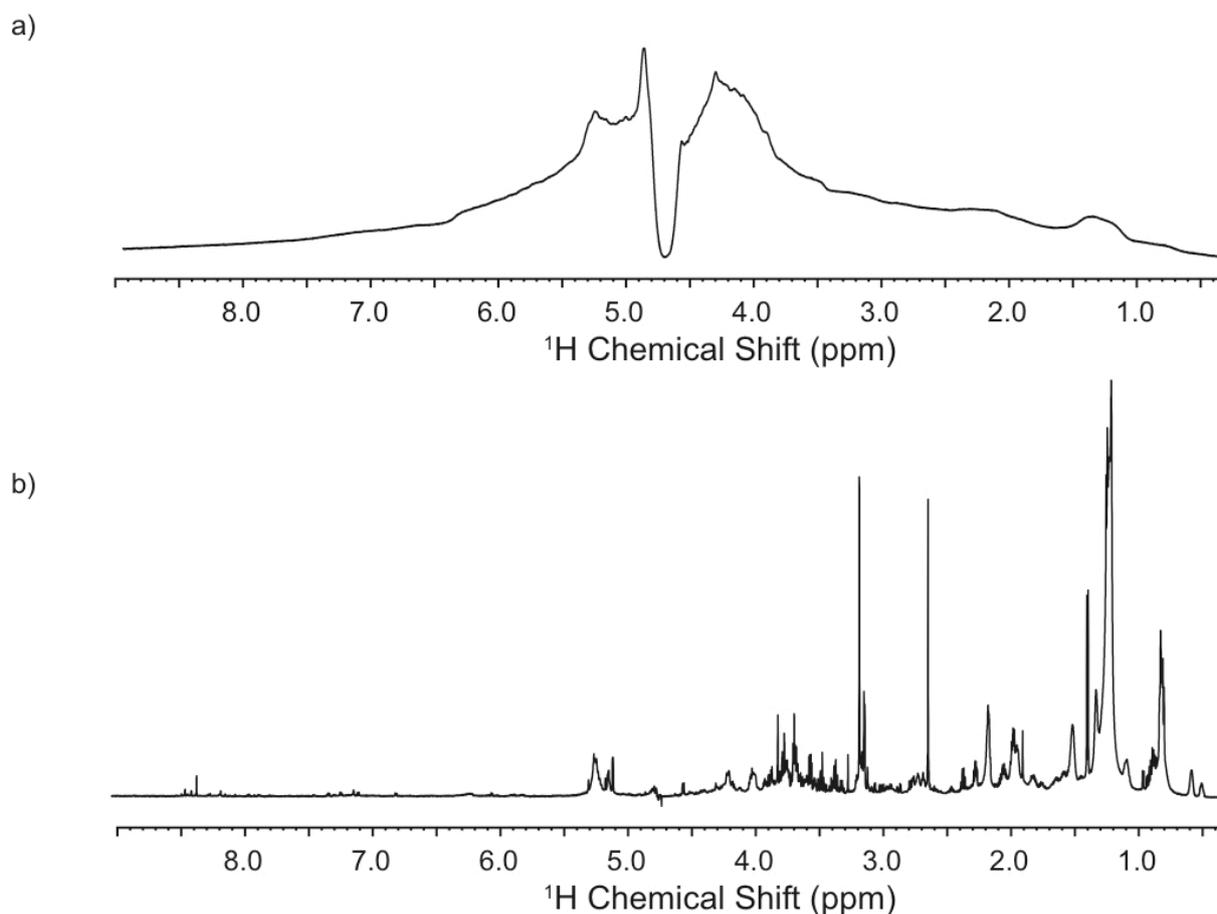


Figure 1.2.2. 700 MHz ^1H 1D NOESY spectrum of *C. elegans* with water presaturation in a) static conditions and b) under rotation at the magic angle (spinning speed: 3.5 kHz), recorded at the CRMN.

Sample preparation for NMR spectroscopy is quite straightforward. For liquid-state NMR spectroscopy, biofluids are mixed with an aqueous solution containing D_2O to lock the deuterium signal during acquisition and reduce signal linewidth. Serum or plasma samples are usually mixed with saline solution (NaCl 0.9% wt/vol, D_2O 10% vol/vol), while urine samples are mixed with a phosphate buffer (pH = 7.4) to control pH variations.^[17] For

HRMAS NMR spectroscopy, 30 μl disposable Kel-f inserts with sealing caps are filled with tissues, cells or small model organism and analyzed in 4 mm zirconium NMR rotors.

The same type of experiments can be performed for liquid-state and HRMAS NMR spectroscopy. The acquisition of metabolic profiles relies essentially on ^1H 1D NMR radio-frequency pulse sequences. Water signal suppression is achieved with solvent suppression methods such as water presaturation.^[48] The most common sequence is the 1D NOESY with water presaturation, which achieves a good water signal suppression and maintains a flat baseline (**Figure 1.2.1**).^[17, 48] Particular biological samples like serum or plasma contain proteins and lipids that generate broad signals, obscuring narrower signals from small molecules (**Figure 1.2.3a**). The contribution of proteins and lipids can be reduced through spectral editing. For instance, the Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence, through T2 editing, decreases the signals of large compounds or metabolites with reduced mobility (**Figure 1.2.3b**).^[17] On the contrary, editing through diffusion coefficient cancels out signals from small mobile molecules to focus on lipids and proteins.^[17]

After acquisition, NMR spectra are processed, with baseline and phase corrections, and calibrated. The calibration can be performed with an internal standard such as TSP or DSS introduced in samples, essentially for urine or directly on metabolite signals with stable chemical shifts such as the α -glucose anomeric proton signal ($\delta = 5.23$ ppm) in serum or plasma and the alanine doublet signal ($\delta = 1.48$ ppm) for *C. elegans*.

NMR spectroscopy can usually identify 30-50 metabolites in serum or plasma samples, 30-100 in urine samples and 20-40 in tissue or *C. elegans* samples.^[2] Indeed, one of the drawbacks of NMR is its poor sensitivity. The detection threshold, which varies between metabolites, is around 10 $\mu\text{mol/L}$ in routine conditions (600 MHz NMR spectrometer, equipped with a cryoprobe, 15 min of acquisition). Sensitivity can be improved with development in hardware, such as the increase in spectrometer magnetic field or introduction of cryoprobes. Resolution and the ability to discriminate metabolites is also an issue for NMR spectroscopy. Resolution can be increased by different methods such as increase in magnetic fields, the use of coupling separation methods as for MS or the use of high dimensional NMR experiments, which reduce signal overlap.

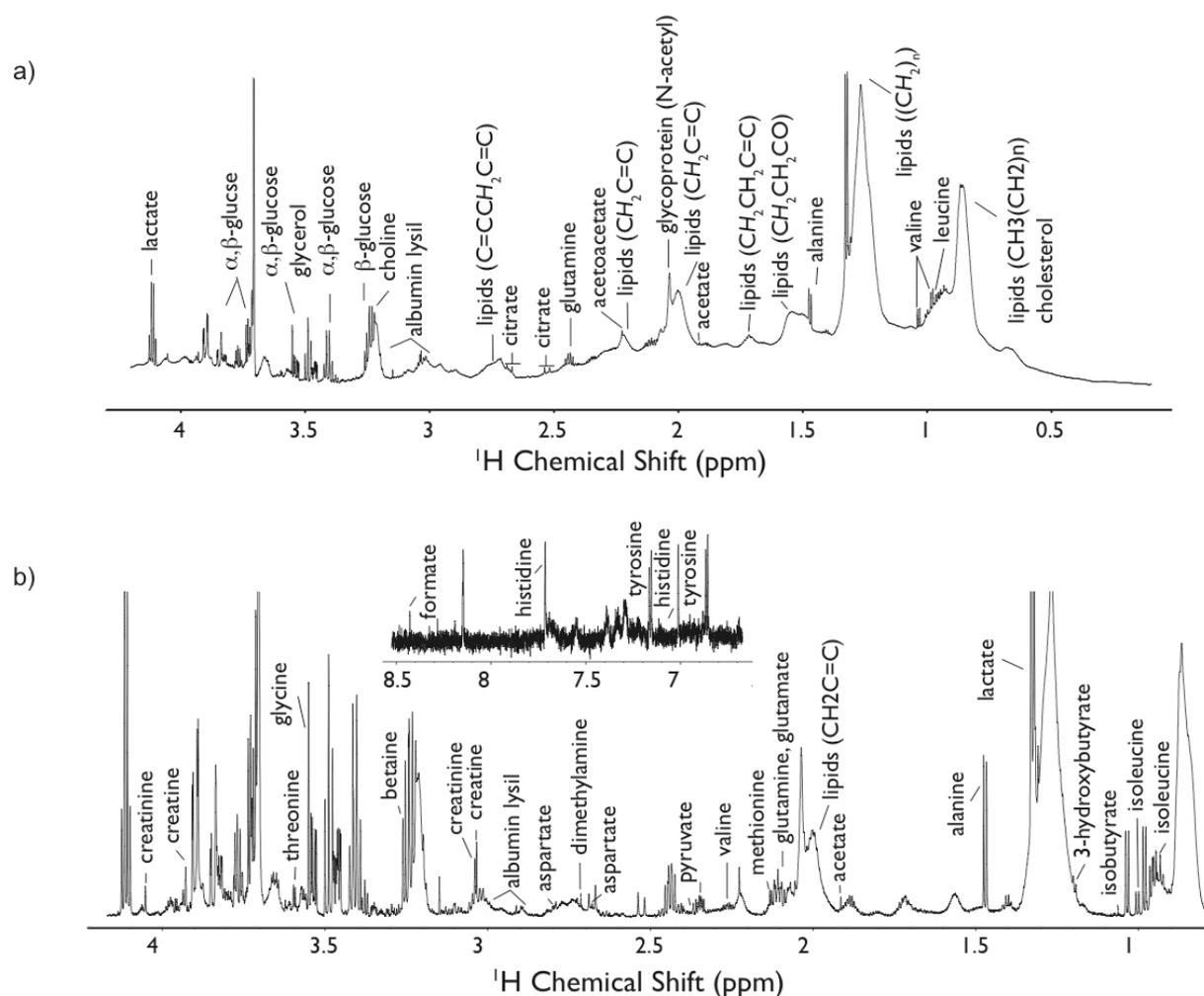


Figure 1.2.3. 800 MHz 1D NMR spectra of a typical serum sample: a) NOESY pulse sequence and b) CPMG pulse sequence, recorded at the CRMN.

Metabolite assignment is performed by comparison to reference spectra using academic spectral databases such as MMCD,^[49] HMDB^[50] and BMRB^[51] as well as proprietary databases (Chenomx NMR Suite, Chenomx Inc, Edmonton, Canada; AMIX SpectraBase, Bruker GmbH, Rheinstetten, Germany). Complementary NMR experiments are usually recorded on a subset of representative samples to provide further structural information, reduce ambiguities and overlaps between signals and eventually ease assignment. These NMR experiments correlate information in two dimensions. For instance, the 2D ^1H - ^{13}C HSQC experiment identifies adjacent proton and carbon nuclei (**Figure 1.2.4**),^[52] while the 2D ^1H - ^1H TOCSY experiment correlate all the ^1H nuclei belonging to a spin-system, usually corresponding to a molecule (**Figure 1.2.5**).^[53] The J-resolved experiment is also largely used in metabolomics and correlate chemical shift and signal multiplicity.^[54]

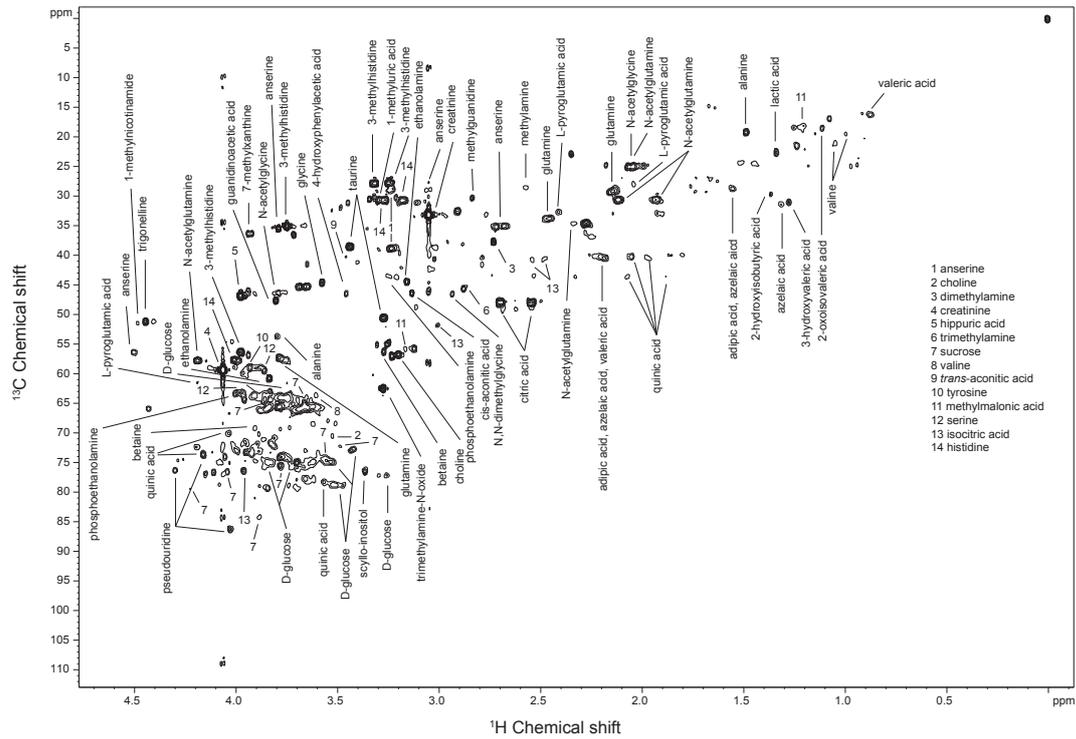


Figure 1.2.4. First 1GHz ^1H - ^{13}C HSQC NMR spectrum of urine (aliphatic region), obtained on our 1GHz (01/2010).

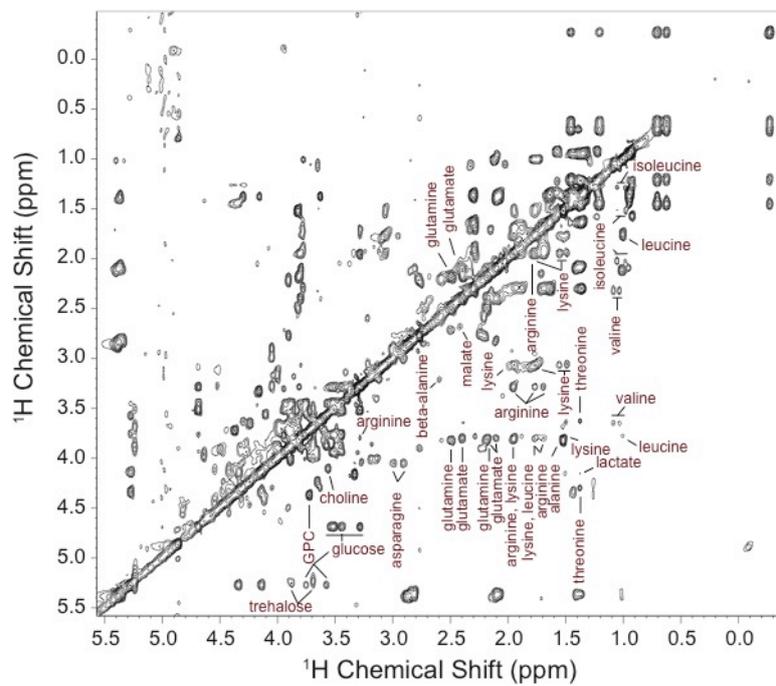


Figure 1.2.5 700 MHz HRMAS ^1H - ^1H TOCSY NMR spectrum of *C. elegans* (aliphatic region), at the CRMN.

However, assignment of metabolites without ambiguities could still be challenging. In this context, we report in the next section results about the introduction of reduced dimensionality techniques for the characterization of complex mixtures, coined targeted projection NMR spectroscopy. For complex mixtures with high signal overlaps, where metabolite assignment is ambiguous even in standard 2D experiments but possible in higher dimensional spectra, the principle is to record an optimized projection of a spectrum of high dimension to capture the relevant information necessary to solve the assignment issue in a small amount of time.

1.3 NMR of complex mixtures: towards targeted projection NMR spectroscopy

1.3.1 Introduction

While Nuclear Magnetic Resonance (NMR) spectroscopy is established as a key technique for metabolomic approaches, providing unique and rich information content with high reproducibility,^[17, 55] the identification of individual metabolites present in biological samples still constitutes a primary obstacle for metabolomics. Although basic 1D ^1H NMR spectra are conventionally recorded for metabolic profiling in high throughput studies due to time constraints, these NMR experiments are not adequate for distinct metabolite identification by NMR which is based on unambiguous and comprehensive spin system assignment. Two-dimensional natural abundance ^1H and ^{13}C NMR spectroscopy, such as implemented in ^1H - ^1H TOCSY^[53, 56] or ^1H - ^{13}C HSQC^[52] experiments, are commonly used as key NMR experiments in metabolic profiling of complex mixtures. They rely on the identification of the distinct metabolite spin systems.^[57-59] However, severe NMR signal overlap, an inherent feature of two-dimensional NMR correlation spectra for complex systems or mixtures in practice often prevents unambiguous detection of metabolites. Various methods have been proposed to address the task of unambiguous metabolite identification, either based on semi-selective or selective excitation, such as selective 1D TOCSY,^[60, 61] or on the use of statistical correlation with the STOCSY methods.^[62, 63] Diffusion-Ordered Spectroscopy (DOSY) has also been proposed as a three-dimensional experiment to resolve analysis of complex mixtures,^[64] while isotope tagging methods can allow to detect classes of derivatized metabolites in biological samples.^[65, 66] Meanwhile, efforts have been made to speed up 2D acquisition, exploiting maximum entropy and non-linear sampling methods,^[67] ultrafast 2D NMR^[68] and Hadamard encoding^[69] or to automate information recovery.^[70, 71]

Multidimensional NMR data are usually acquired by sampling the time domain in all dimensions equidistantly at an operator-selected resolution. In the context of metabolites at natural abundance, sensitivity and resolution considerations of acquisition of ^1H and ^{13}C NMR data pose strong limitations on the use of high-dimensional experiments (3D, 4D...) that would clearly provide unambiguous spectra but would require unrealistically long acquisition times.

With recent advances in sensitivity, due to the use of increasing magnetic field strength and cryogenic NMR probes, the concept of reduced-dimensionality NMR spectroscopy^[72-75] has seen a strong revival^[76-78] which addresses the sampling limit issue of higher-dimensional NMR experiments by combining the acquisition of two or more indirect dimensions for multidimensional (4D, 5D...) NMR experiments in order to capture the desired high-dimensional chemical shift information in a lower dimensional experiment. Projection spectroscopy^[74] consists in recording discrete sets of projection spectra from higher-dimensional NMR experiments.^[78] Major applications, mainly focused on sequence-specific resonance assignment in the field of protein NMR spectroscopy, have recently illustrated this approach by speeding up the acquisition of high dimensional correlation spectra.^[79-85]

In the present study, we recruit projection NMR techniques for unambiguous metabolite identification in complex mixtures by capturing essential information of high-dimensional correlations in order to resolve NMR signal overlap present in standard 2D NMR experiments. We demonstrate this approach on a model metabolite mixture composed of ornithine, putrescine and arginine for which the resonances of putrescine overlap with the other metabolite signals in both classical 2D ^1H - ^1H TOCSY and ^1H - ^{13}C HSQC spectra, making this compound therefore undetectable. Acquisition of a single 2D projection of a 3D ^1H - ^1H - ^{13}C TOCSY-HSQC spectrum at an optimal projection angle selected by a fit-for-purpose algorithm, allowed the unambiguous assignment of this metabolite mixture, by providing well-resolved NMR signals of the putrescine spin system in the selected 2D projection plane.

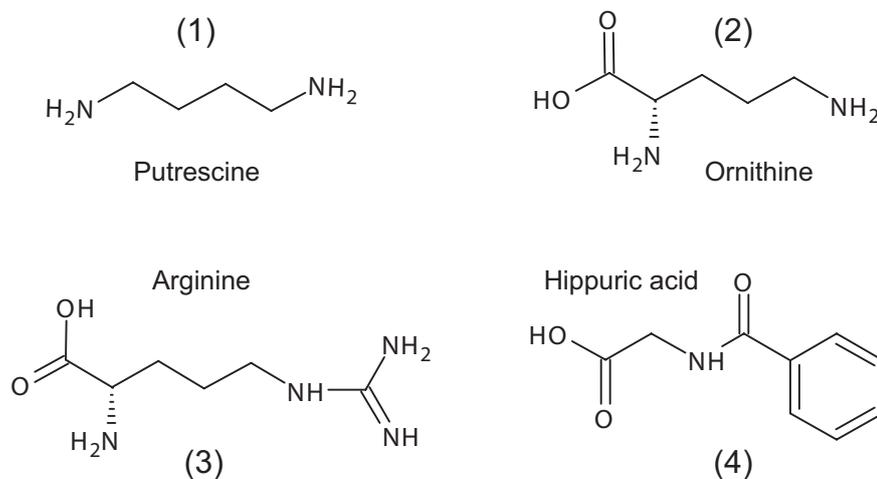
1.3.2 Materials and methods

Samples. Putrescine, hippurate, ornithine and arginine were purchased from Acros Organics, deuterated water (D_2O) from Euriso-top, the chemical shift reference, 3-(trimethylsilyl)propionic- 2,2,3,3- d_4 acid sodium salt (TSP) from Sigma-Aldrich and Sodium azide, Na_2HPO_4 and NaH_2PO_4 from Fluka Biochemika.

A phosphate buffer ($\text{pH} = 7.4$) was prepared by weighing 0.577 g of Na_2HPO_4 , 0.121 g of NaH_2PO_4 , 0.02 mM of TSP and 0.06 mM of NaN_3 into a 20 mL volumetric flask, adding 4 mL of D_2O and filling up to 20 mL with H_2O .

The model mixture was an aqueous solution of hippuric acid (10.9 mmol.L^{-1}), ornithine (58.2 mmol.L^{-1}), putrescine (3.6 mmol.L^{-1}) and arginine ($174.54 \text{ mmol.L}^{-1}$). The

phosphate buffer represented one third of total volume and thus D₂O 6.6% of sample in volume. A simpler model mixture was made similar to the previous one without arginine.



Scheme 1.3.1. Structures of model mixture compounds.

NMR spectroscopy. All NMR experiments were carried out at 300K, on a Bruker Avance II spectrometer, operating at a ¹H frequency of 700 MHz, using a standard triple resonance (¹H-¹³C-¹⁵N) 5 mm TXI probe. A standard 3D ¹H-¹H-¹³C TOCSY- HSQC pulse sequence using MLEV-17 isotropic mixing, States modulation in F₁ (¹H dimension) and gradient Echo/Anti-echo modulation in F₂ (¹³C dimension) was used. High power ¹H and ¹³C 90° pulses were measured at 12 μs and 13.5 μs respectively. The mixing time was set at 60 ms for all experiments. The spectral width in the direct dimension F₃ (¹H) was set to 9803 Hz with a maximum *t*₃ acquisition time of 52.2 ms and a relaxation delay between scans of 1 s.

The ¹H 2D TOCSY spectrum was recorded as the F₁-F₃ plane of the 3D ¹H-¹H-¹³C TOCSY- HSQC experiment, acquired by incrementing *t*₁ and keeping *t*₂ null. The F₁ spectral width SW₁ was set to 9803 Hz and 1024 Free Induction Decays (FID) were acquired corresponding to a *t*₁^{max} of 52 ms. The ¹H-¹³C 2D HSQC spectrum was recorded as the F₂-F₃ plane of the 3D ¹H-¹H-¹³C TOCSY- HSQC experiment, acquired by incrementing *t*₂ and keeping *t*₁ null. The F₂ spectral width SW₂ was set to 13203 Hz and 1024 FIDs were acquired corresponding to a *t*₂^{max} of 38.8 ms.

Projection planes were acquired by using the same 3D pulse sequence and incrementing *t*₁ and *t*₂ simultaneously in a constant ratio defining the projection angle α:

$$\begin{cases} t_1 = t \cos \alpha \\ t_2 = t \sin \alpha \\ t_2/t_1 = \tan \alpha \end{cases} \quad (1)$$

The value Δ of the increment of t depends on the spectral width SW of the tilted dimension as follows:

$$\Delta = 1/SW \quad (2)$$

with

$$SW = SW_1 \cos \alpha + SW_2 \sin \alpha \quad (3)$$

1024 FIDs were acquired with 16 scans each corresponding to a total experiment time of less than 5 hours for each set ($\pm\alpha$) of 2D projections.

Projection plane processing. As the experiment was acquired with States modulation in the F_1 dimension and gradient Echo/Anti-echo schemes in F_2 , recombination of the FIDs was performed to separate projections on $+\alpha$ and $-\alpha$ planes. For each indirect dimension time increment, we obtained 4 FIDs modulated as follows:

$$\cos(\Omega_1 t \cos \alpha) \exp(i\Omega_2 t \sin \alpha) \quad (S_1)$$

$$\sin(\Omega_1 t \cos \alpha) \exp(i\Omega_2 t \sin \alpha) \quad (S_2)$$

$$\cos(\Omega_1 t \cos \alpha) \exp(-i\Omega_2 t \sin \alpha) \quad (S_3)$$

$$\sin(\Omega_1 t \cos \alpha) \exp(-i\Omega_2 t \sin \alpha) \quad (S_4)$$

Recombination of these FIDs leads to:

$$\cos(\Omega_1 t \cos \alpha + \Omega_2 t \sin \alpha) \quad (S_1 - S_2/i + S_3 + S_4/i)$$

$$\sin(\Omega_1 t \cos \alpha + \Omega_2 t \sin \alpha) \quad (S_1/i + S_2 - S_3/i + S_4)$$

$$\cos(\Omega_1 t \cos \alpha - \Omega_2 t \sin \alpha) \quad (S_1 + S_2/i + S_3 - S_4/i)$$

$$\sin(\Omega_1 t \cos \alpha - \Omega_2 t \sin \alpha) \quad (S_1/i - S_2 - S_3/i - S_4)$$

Consequently, these combinations generate the sum and difference frequencies

$$\begin{cases} \Omega_{+\alpha} = \Omega_1 \cos \alpha + \Omega_2 \sin \alpha \\ \Omega_{-\alpha} = \Omega_1 \cos \alpha - \Omega_2 \sin \alpha \end{cases} \quad (4)$$

corresponding respectively to projection on planes tilted from F_1 by $+\alpha$ and by $-\alpha$. These FID combinations were done using MATLAB (The Mathworks, Inc). Each plane, $+\alpha$ and $-\alpha$, was then processed with NMRPipe^[86] as a classical 2D experiment acquired with the States method. Zero filling was performed in both dimensions to double the number of data points. Sine-bell apodization was used in both dimensions prior to two-dimensional Fourier transformation.

Projection angle optimization. In order to find the optimal projection angle α , to resolve each correlation signal of the spin system of a given metabolite with respect to all the other correlations originating from the other metabolites present in the mixture, signal positions

were calculated as a function of the projection angle α for a given NMR experiment using the correspondence between projected chemical shifts $\Omega_{\pm\alpha}$ in the $+\alpha$ or $-\alpha$ planes and chemical shifts from indirect original dimensions, as defined in equations 4.

The search routine to find the optimal angle and described in the following was written in FORTRAN 90 as an autonomous module and has been implemented into the framework and data structure of the UNIO application platform^[87-89] for automated NMR data analysis.

1.3.3 Results and discussion

Elucidation of the composition of a metabolite mixture by NMR spectroscopy requires unambiguous NMR signal assignment. This can be achieved when each metabolite displays at least one well-resolved characteristic NMR signal. Here, a characteristic NMR signal for a given metabolite is defined as a correlation peak of the expected metabolite spin system that cannot be explained by any other expected correlation peaks of any of the other metabolites present in the mixture composition. The specific question we address is then “is a given metabolite present in a mixture?”. To answer this question we make use of the fact that all metabolites are known, and their NMR spectra are available. As a model example, in this study we demonstrate the method on a model mixture made of four common metabolites (**Scheme 1.3.1**) found in biological samples,^[58] hippuric acid (10.9 mmol.L⁻¹), ornithine (58.2 mmol.L⁻¹), putrescine (3.6 mmol.L⁻¹) and arginine (174.54 mmol.L⁻¹) but for which unambiguous metabolite identification is not straightforward. Indeed, putrescine, cannot be unambiguously identified from either 2D ¹H-¹H TOCSY or ¹H-¹³C HSQC: TOCSY patterns for putrescine and ornithine are superimposed, while the putrescine HSQC correlations overlap with either the ornithine or arginine signals, as illustrated in the spectra of **Figure 1.3.1**.

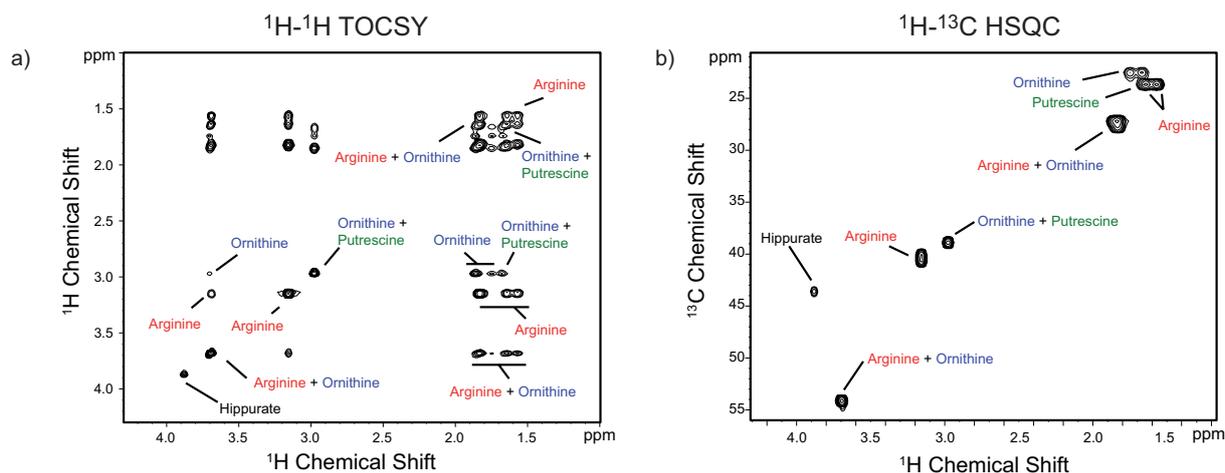


Figure 1.3.1. Contour plots of a) a 2D ^1H - ^1H TOCSY spectrum and b) a 2D ^1H - ^{13}C HSQC spectrum of a mixture composed of putrescine (3.6 mmol.L⁻¹), ornithine (58.2 mmol.L⁻¹), arginine (174.54 mmol.L⁻¹) and hippurate (10.9 mmol.L⁻¹).

Rather than recording a whole 3D ^1H - ^1H - ^{13}C TOCSY-HSQC experiment to resolve the assignment ambiguity, our strategy is to record a single appropriate 2D projection plane of the 3D ^1H - ^1H - ^{13}C TOCSY-HSQC, which allows identification of a characteristic putrescine NMR signal.

For a given 3D NMR experiment, combined exploration of the indirect time domains allows the acquisition of a 2D projection spectrum experiment defined by a projection angle α with respect to first indirect dimension F_1 as illustrated in **Figure 1.3.2**. The dimensions of the 2D projection plane (also termed tilted plane) are the direct dimension of acquisition F_3 and a tilted dimension, representing a linear combination of the two indirect dimensions F_1 and F_2 . Recording a projection plane is a 2D experiment, built from the original 3D experiment, where the two indirect time parameters t_1 and t_2 are simultaneously incremented in a constant ratio (Eq. 1).

Due to the use of quadrature detection schemes in both indirect dimensions, for each 2D experiment corresponding to a given value of α , we simultaneously obtain the projection of the 3D spectrum on the $+\alpha$ plane and $-\alpha$ plane, as detailed in the materials and methods. Chemical shifts in the tilted dimension, $\Omega_{\pm\alpha}$ for the projection on a $\pm\alpha$ plane, are defined by a linear combination of the chemical shifts Ω_1 and Ω_2 in the indirect dimensions and the projection angle α (Eq. 4).

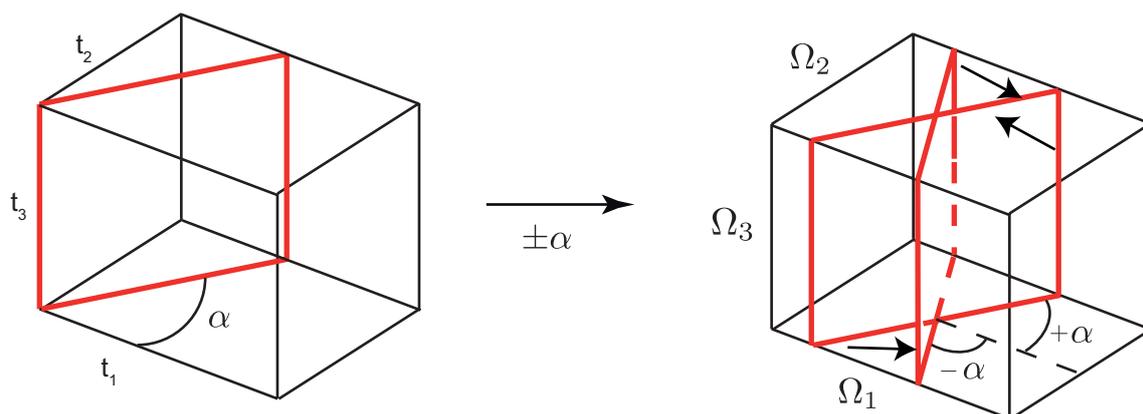


Figure 1.3.2. Schematic principle of the projection technique. In our case, the indirect dimension F_1 and F_2 , associated respectively with the time parameters t_1 and t_2 , and to the chemical shift Ω_1 and Ω_2 , correspond respectively to the ^1H and ^{13}C acquisition dimensions. F_3 is the direct acquisition dimension, associated with the chemical shift Ω_3 .

In our 3D ^1H - ^1H - ^{13}C TOCSY-HSQC experiment, the first indirect dimension F_1 corresponds to the ^1H dimension and F_2 to the ^{13}C dimension. As a consequence, for a projection angle of $\alpha = 0^\circ$ and $\alpha = 90^\circ$, the 2D projection spectrum is equivalent to a 2D ^1H - ^1H TOCSY spectrum and a 2D ^1H - ^{13}C HSQC spectrum, respectively. We recorded a set of 2D projection planes from the 3D ^1H - ^1H - ^{13}C TOCSY-HSQC experiment for the projection angles $\alpha = \pm 30^\circ, \pm 53^\circ, \pm 70^\circ$ and $\pm 83^\circ$ in order to assess the feasibility of the method in terms of sensitivity and resolution obtained. As illustrated **Figure 1.3.3**, this yields a progression of projected signal patterns from 2D TOCSY patterns to 2D HSQC patterns as α increases.

For a simpler mixture made of hippuric acid (10.9 mmol.L^{-1}), ornithine (58.2 mmol.L^{-1}) and putrescine (3.6 mmol.L^{-1}), we compared the experimentally measured values of the chemical shifts of the projection correlations with the values calculated with equation (3) from the values determined with 2D TOCSY and 2D HSQC, for the projected plane $\alpha = +30^\circ$, and shown in **Figure 1.3.4a**. Good agreement was observed between expected and observed projected correlations, confirming the reliability of the technique.

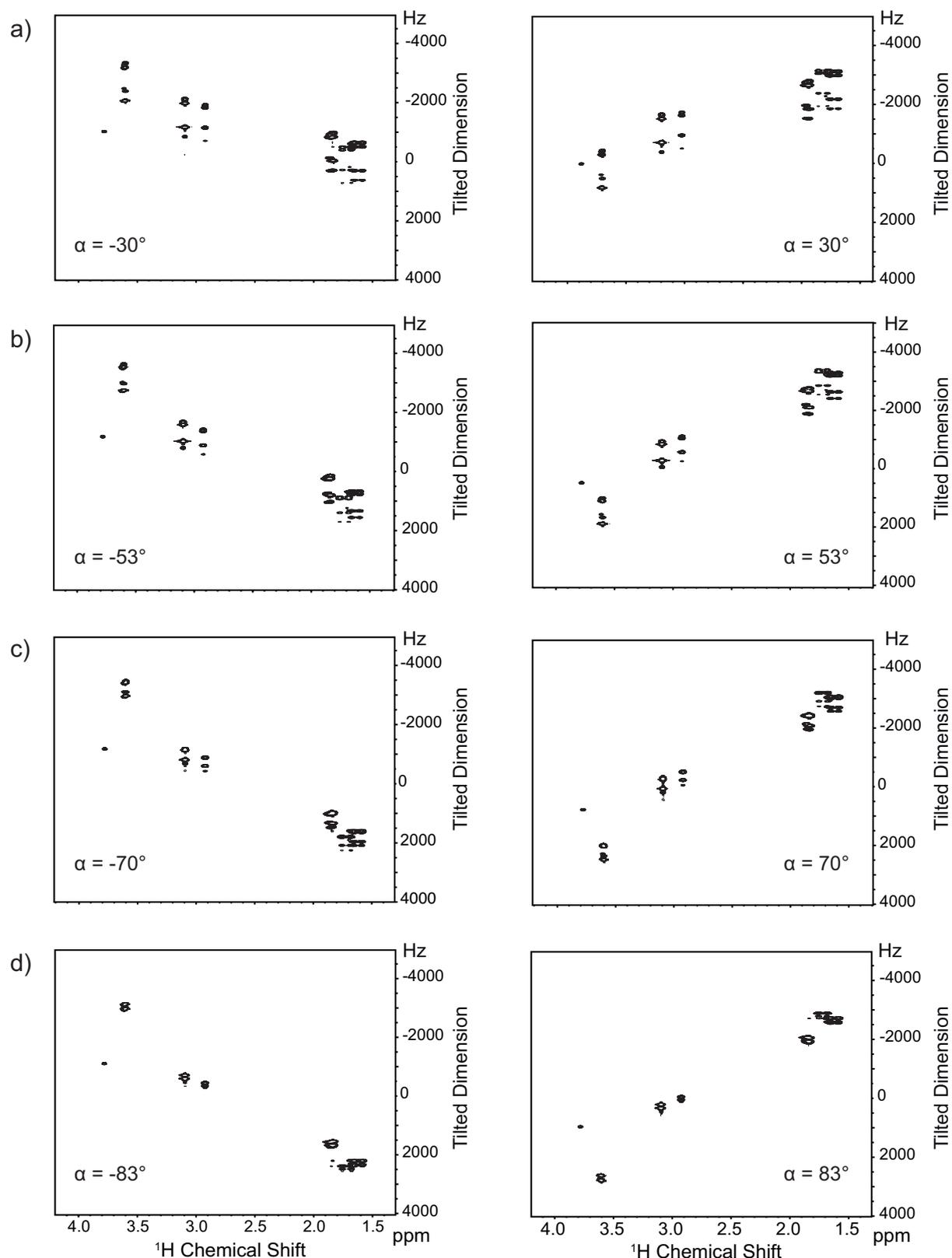


Figure 1.3.3. A series of four tilted 2D planes from a 3D ^1H - ^1H - ^{13}C TOCSY-HSQC experiment, recorded in distinct experiments, corresponding to (a) $\alpha = \pm 30^\circ$, (b) $\alpha = \pm 53^\circ$, (c) $\alpha = \pm 70^\circ$ and (d) $\alpha = \pm 83^\circ$. Note how the peaks move in characteristic, and perfectly predictable manners in the vertical (mixed) dimension. In particular peaks that are not resolved at $\alpha = \pm 83^\circ$ are clearly resolved at the other angles.

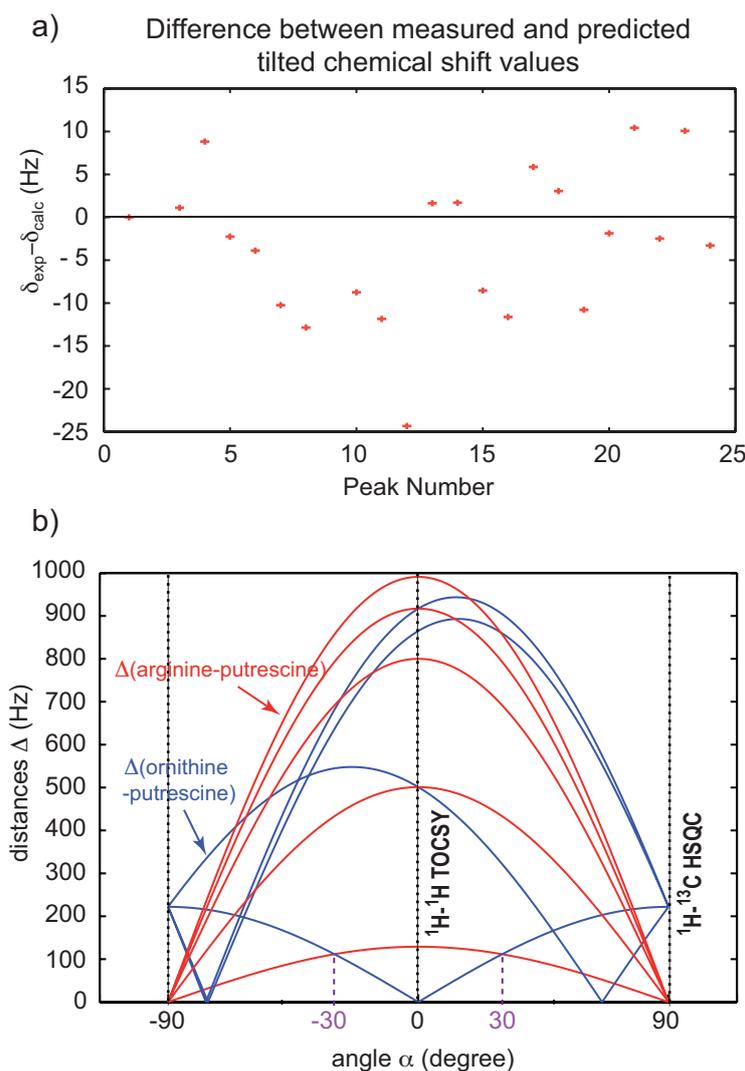


Figure 1.3.4. a) Comparison of experimental and calculated peak positions in tilted plane $\alpha = +30^\circ$ for a mixture composed of putrescine (3.6 mmol.L^{-1}), ornithine (58.2 mmol.L^{-1}) and hippurate (10.9 mmol.L^{-1}). b) Optimization of projection angle: the lines show the distances in Hz between the putrescine correlation that is potentially characteristic ($\Omega_{1\text{H}}: 2.97 \text{ ppm}$, $\Omega_{13\text{C}}: 23.7 \text{ ppm}$, $\Omega_{1\text{H}}: 1.67 \text{ ppm}$) and all the ornithine (blue) and arginine (red) correlations at 1.67 ppm in F_3 with respect to α . The optimum projection angle is determined automatically by selecting the angle with the largest gap between the baseline and the lowest difference, as illustrated in the figure for $\alpha = \pm 30^\circ$. At this point we predict that the putrescine resonance will be separated in the vertical dimension from all the other resonances by at least 110 Hz .

Selection of an optimal projection angle α . Here, we are interested in acquiring a single targeted 2D projection plane which contains a well-resolved characteristic NMR signal for putrescine, which is here the minor component of the mixture (and more generally for any given metabolite in a mixture). (As ornithine, arginine and hippurate have specific NMR

signals in both TOCSY and HSQC experiments, conserved in projection planes, we focused in this study on the ambiguous putrescine correlations for the optimization of the projection angle α .) To this end for each correlation in the spin system of putrescine, the predicted signal separation measured in Hertz with respect to all other correlation peaks of all the other metabolites present in the mixture are calculated as a function of the projection angle. This automatically performed analysis revealed the 3D putrescine correlation at $\Omega_{1H} = 2.97$ ppm, $\Omega_{13C} = 23.7$ ppm, and $\Omega_{1H} = 1.67$ ppm in the 3D 1H - 1H - ^{13}C TOCSY-HSQC as a candidate for a characteristic unambiguous putrescine NMR signal. The calculation of the signal separation between this correlation and all other ornithine and arginine signals correlating at $\delta = 1.67$ ppm in the direct dimension as a function of the projection angle α using Eq. 4 identified the projection angle $\alpha = \pm 30^\circ$ as the optimal 2D projection plane with maximum achievable resolution for the potential characteristic putrescine signal, as illustrated in **Figure 1.3.4b**. At this point we predict that the putrescine resonance will be separated in the vertical dimension from all the other resonances by at least 110 Hz. As expected, projection planes shown in **Figure 1.3.5** recorded for the optimized angle $\alpha = \pm 30^\circ$ display a well-resolved characteristic NMR signal for putrescine, making its identification unambiguous. The acquisition of a single targeted 2D projection plane allowed us to resolve signal overlap in a tailored manner and to identify unambiguously all the metabolites of the mixture. Though the main objective here is to resolve ambiguities arising from 2D data for known metabolites, the proposed approach is not limited to the studies of mixtures for which all metabolites are known. In the presence of unknown metabolites whose signals are fully overlapped with signals assigned to known spin systems, which would thus be undetectable at the start of the procedure and would not be included in the calculation of optimized projection angles, recording of a set of projections is likely to reveal the presence of these potential additional compounds in the mixture.

Note finally that the objective of our approach is not to reconstruct a 3D spectrum but to exploit a minimum and targeted set of artifact-free 2D projections in order to resolve potential ambiguities present in the conventional 2D spectra.

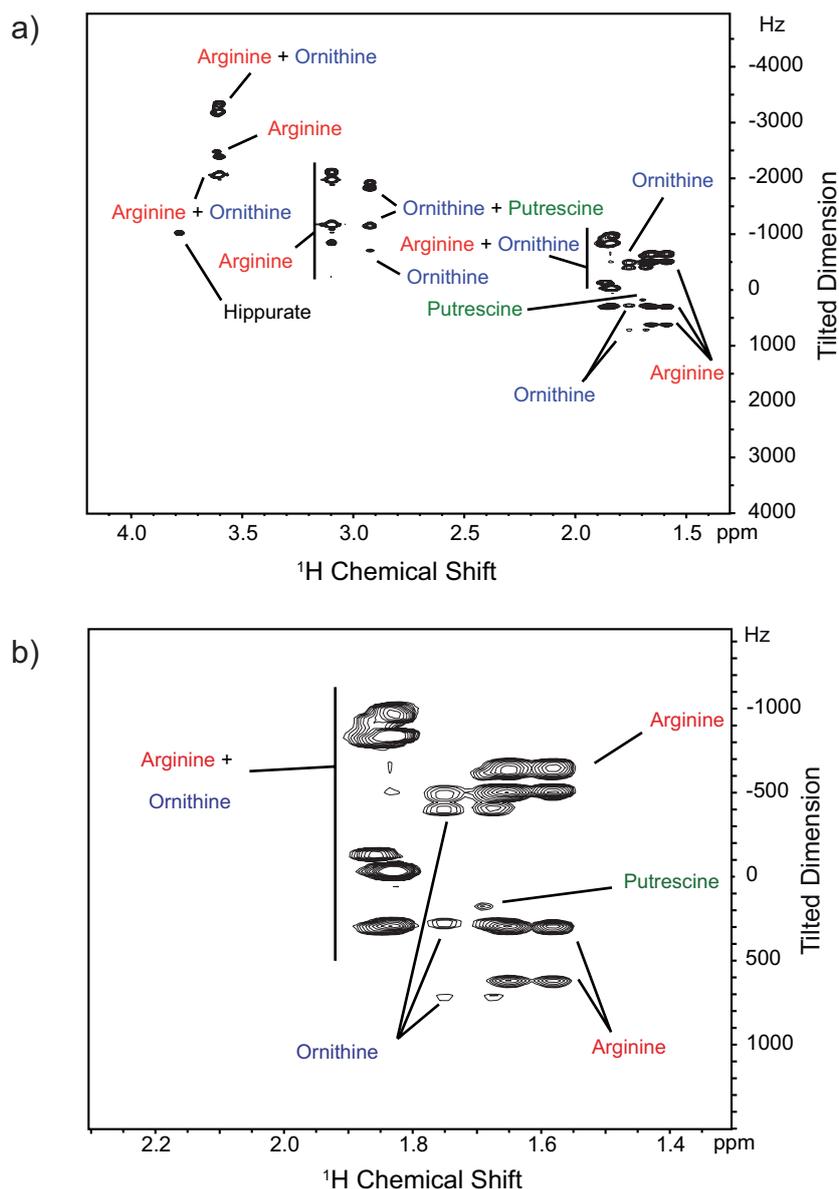


Figure 1.3.5. a) Projection plane for $\alpha = -30^\circ$ and b) zoom on the region [^1H : $\delta = 1.5 - 2.5$ ppm; tilted dimension: $-1500 - 1500$ Hz]. We see clearly that overlaps between ornithine, arginine and putrescine are now resolved, and the peak at $\delta(^1\text{H}) = 1.67$ ppm and at 200 Hz in the tilted dimension is an unambiguous indicator of the presence of putrescine in the mixture.

1.3.4 Conclusion

We have shown how targeted projection NMR spectroscopy appears to be a powerful and promising method for unambiguous elucidation of metabolite-mixture composition by giving efficient access to high dimensional correlations in a realistic amount of time, enabling resolution of spectral overlap and unambiguous metabolite identification. When ambiguities appear in signal identification with different plausible hypotheses, *a priori* knowledge of

potential mixture composition allows the selection of a discriminating optimized projection experiment in an automatic fashion.

Note that to illustrate the approach, we have here used a model mixture, and identified the minor component. Of course, since the NMR spectra of essentially all metabolites are known (and are even increasingly present in databases) or can easily be determined, it is perfectly possible to imagine extending this approach to the identification of minor metabolites in biological complex mixture samples such as plasma or urine, recovering more information from these thousand compound mixtures of medical interest.

1.4 Data analysis in metabolomics

After acquisition, NMR datasets are transformed to become suitable input for statistical analyses. Metabolite signal chemical shifts can change with the pH or ionic strength of the medium. If these parameters are not completely controlled with buffering, like in urine datasets, it is sometimes necessary to realign some metabolite signals for reliable comparisons.^[90, 91] NMR spectra are converted in a set of points termed NMR variables. This is usually performed with a regular binning (or bucketing), using typically buckets width of 0.001 to 0.01 ppm, the signals being then integrated for each spectral bin.^[92] With a resolution of 0.001 ppm for bucketing, each NMR spectrum is described by about 10000 variables. Different algorithms have been developed to perform an automatic binning, delineating meaningful variables such as peaks or multiplets.^[93, 94] The statistical recoupling of variables (SRV) method, widely used in this thesis, is an automatic binning procedure developed in our laboratory, which defines bins or clusters according to the covariance/correlation ratio profile of consecutive variables along the chemical shift axis.^[94] It is also possible to quantify metabolite concentrations in a targeted manner through deconvolution, prior to analysis.^[95] Spectra are normalized to cancel out effects of dilution between samples and focus on biologically relevant differences by dividing each variable by the total intensity of NMR signals or with more sophisticated approaches such as the probabilistic quotient normalization (PQN) procedure.^[96, 97] The PQN method evaluates the most probable coefficient of dilution between each spectrum and a reference spectrum and normalizes each spectrum by this coefficient.^[97] The dataset are finally mean-centered and possibly scaled with the Pareto or autoscaling methods.^[98] At the end of this step, the dataset corresponds to a matrix \mathbf{X} , on which statistical analyses are performed.

Unsupervised and supervised multivariate statistical methods are used to build models for data visualisation, sample classification and extract metabolic signatures between sample groups.^[34] Principal component analysis (PCA) is first performed to derive the main sources of variance within the dataset, check population homogeneity and eventually identify technical or biological outliers. PCA defines a new orthonormal basis set, which vectors termed principal components, new linear combinations of initial variables, correspond to the eigenvectors of the covariance matrix $\mathbf{X}'\mathbf{X}$ and are associated with maximum variance. PCA reduces the space of variables, initially represented by a few orthogonal principal components, which eases data visualisation. Data are represented as score plots, where each

point stands for the projection of a single sample on principal components and as loading plots, which represent the contributions of metabolic variables to principal components. Supervised methods such as Partial Least Square (PLS) or OPLS multivariate regression, are performed to build a sample classification model and derive group-specific metabolic phenotypes.^[99, 100] These methods are run to discriminate groups by regressing a supplementary data matrix **Y**, containing information about the group-class on the **X** NMR dataset matrix. A new basis set is defined where the correlation between the sample projection on the first component and the **Y** matrix is maximized. As for PCA, results are visualized through score and loading plots (**Figure 1.4.1**).

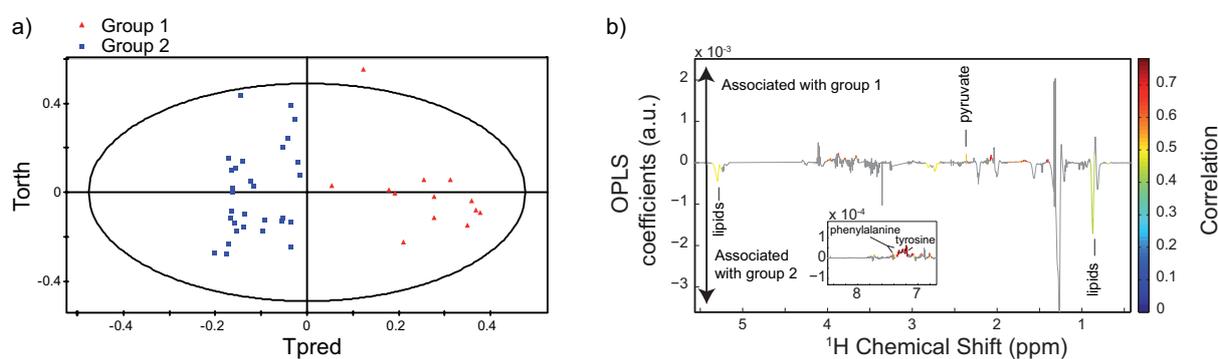


Figure 1.4.1. Visualization of results from an OPLS analysis. a) score plot: each point corresponds to one sample. A discrimination is observed between the group 1 in red and the group 2 in blue along the horizontal predictive component. b) loading plot: contributions of metabolite involved in this discrimination. Pyruvate here is associated with the group 1 while lipids are associated with the group 2.

Model performances are assessed by goodness-of-fit parameters R^2 and Q^2 calculated through a cross-validation procedure, related respectively to the explained and predicted variance by the model. We perform model validation by resampling the model 1000 times under the null hypothesis. The decrease of goodness-of-fit R^2 and Q^2 parameters when correlation between original model and random models decreases indicates the good quality of the model by rejecting the null hypothesis (**Figure 1.4.2**). However, the better test to assess model robustness is to evaluate its prediction ability on completely independent samples.

Univariate analyses are also performed to identify significant changes in metabolite concentrations.^[101] However, performing a large number of statistical tests simultaneously as encountered in metabolomics raises multiple testing issues. Indeed, the number of type I errors, i.e. false rejections of the null hypothesis, increases with the number of tests performed. Different error rates were defined in the framework of multiple testing with procedures for their control or estimation.

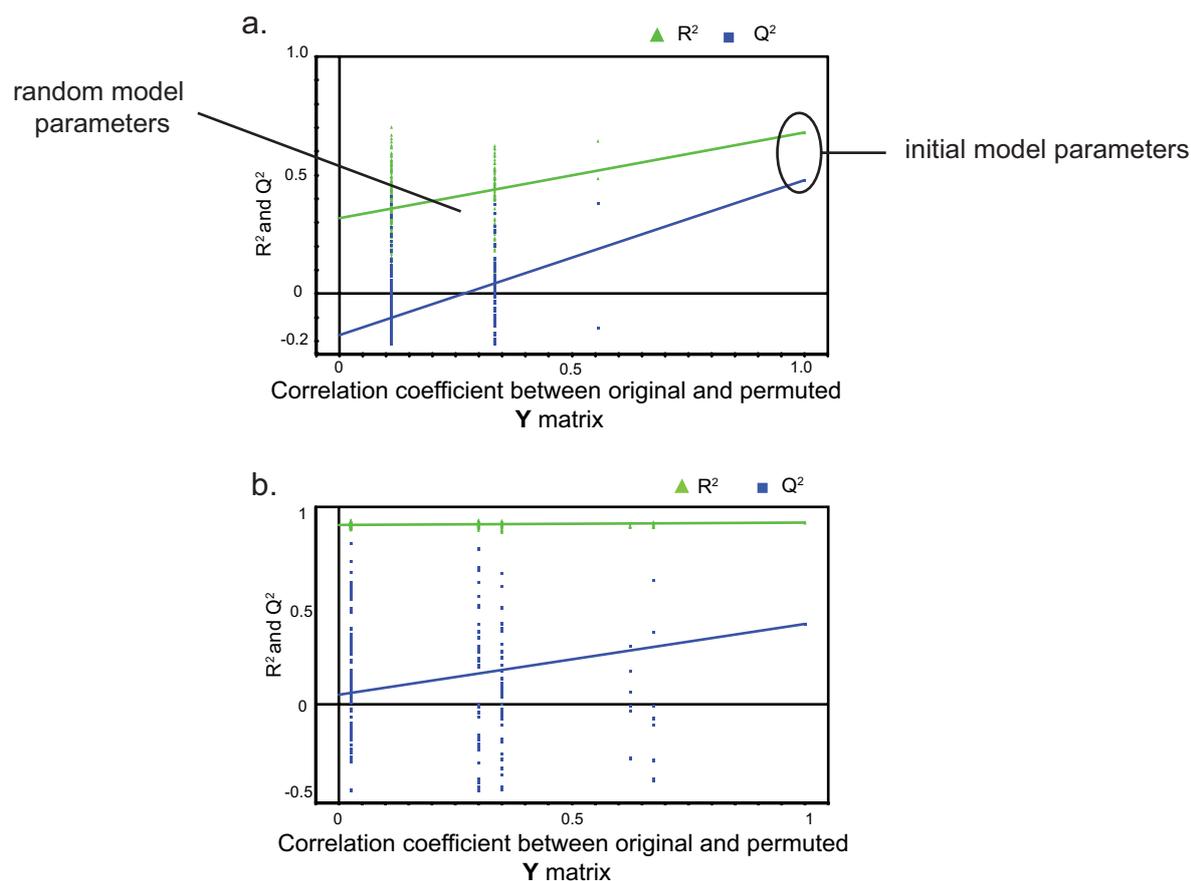


Figure 1.4.2. Typical results of a resampling validation procedure. a) R^2 and Q^2 goodness-of-fit parameters for random models are weaker than the initial model ones, which validates the initial model. b) R^2 and Q^2 goodness-of-fit parameters for random models are larger than the initial model ones. The performances observed for the initial model can thus be due to chance. The model is not validated.

The Family-Wise Error Rate (FWER) is defined as the probability to perform at least one type I error for the test family. The FWER is conservatively and strongly controlled by the Bonferroni procedure.^[101] The False Discovery Rate (FDR), defined as the expected proportion of type I errors among the rejected hypotheses, introduced by Benjamini and Hochberg is largely used in microarray experiments. Its control is less conservative than for the FWER.^[102] The procedure published by Benjamini and Hochberg controls the FDR for independent tests and for mild or limited correlation structure.^[102, 103] Benjamini and Yekutieli introduced a procedure, which controls the FDR under any dependence but with a lower power.^[104] Many procedures were since published for optimizing the FDR control.^[103]

The choice of the error rate control depends on the aims of the study. The FDR control is more relevant for exploratory analyses (discovery phase...), while the FWER control should be used for assessing more robust statistical significance (validation phase).^[103]

Different approaches have been developed to rationally understand the biological significance of variations in metabolite profiles. Some of them describe changes at the level of metabolites and chemical reactions, either by mapping changes on existing metabolic networks^[105] or by reconstructing metabolic networks based on correlation values^[63] while others focus on changes at the level of canonical pathways.^[106-108]

Part 2: Metabolomics for human population studies

Developing new tools for public health has been a driving force for metabolomics, to better characterize human populations, diagnose diseases or predict treatment outcomes. In this second part, we present different metabolomic applications aiming at discovering metabolic biomarkers in the context of molecular epidemiology and clinics.

Research for biomarker discovery relies mainly on observational case-control studies. The choice of the case and control groups defines the objectives of the study. The definition of control individuals is a crucial issue, on which the reliability of the study depends. Indeed, during this step, potential confounding factors are cancelled out through matching between cases and controls in order to focus on relevant differences.

Differences between groups are often small, in particular for epidemiological studies, and inter-individual variations important, due to the lack of control of external factors in this type of population, requiring thus large sample numbers to obtain statistically significant results.

In this part, we present three different studies, conducted on human biofluids, plasma and serum, using liquid-state NMR spectroscopy.

In the framework of molecular epidemiology, the first study aims at evaluating the feasibility of using the serum biobank collected from the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort for metabolomics.

The two following studies investigate systemic metabolic fingerprints of cancer in a clinical context. One aims at identifying metabolic differences between patients suffering from localized or metastatic breast cancers. The last study distinguishes a potential plasma metabolic signature for hepatocarcinoma, from healthy patients or patients suffering from chronic liver diseases.

2.1 Assessment of the exposome: the EPIC cross-sectional pilot study

2.1.1 Introduction

Genetics and environmental exposures, like radiation, infection, lifestyle or diet, play intricate key roles in the development of chronic diseases, such as cancer, diabetes or obesity. The purpose of epidemiology is to evaluate the individual contributions of these factors to the pathological processes. However, as well as the high complexity of the underlying corresponding mechanisms, epidemiology has to cope with the inherent difficulty to precisely evaluate the intensity of the exposures at the individual level.^[109, 110] For instance, the assessment of lifestyle or dietary exposures rely on questionnaires, associated with measurement errors, which might obscure disease risk associations. Molecular epidemiology was introduced to overcome these limitations, through the determination of unambiguous biomarkers of exposure, aiming at improving exposure assessment, detecting early changes preceding diseases or defining subgroups at risks.^[111]

In this context, the development of high-throughput screening methods like metabolomics could be highly valuable for molecular epidemiology, assessing simultaneously multiple biological markers. Recent works have evaluated the applicability of metabolomics for the study of large prospective multicentric human cohorts and have determined new biomarkers of disease risks.^[36, 112] A few studies have already compared the different analytical technologies, like NMR spectroscopy, GC-MS or LC-MS in a context of molecular epidemiology.^[113, 114] However such work has to be carried out on every new sample biobank, to assess the specimen quality and suitability for the metabolomic approach.

In this work, we have evaluated the applicability of metabolomics for a large serum biobank, collected in the framework of the European Prospective Investigation into Cancer and Nutrition (EPIC) study, a large multicentric prospective epidemiological study designed to investigate the association between nutrition and cancer, where 520,000 participants (healthy volunteers) were enrolled in 23 centers from 10 European countries (Denmark, France, Greece, Germany, Italy, Netherlands, Norway, Spain, Sweden and United Kingdom) between 1992 and 1998. The strength of this study was in the high quality of the recordings of individual lifestyle habits, diets, anthropometric data and blood samples, obtained at the inclusion before disease onset, designed for the identification of prospective markers of

pathologies.^[115] However, blood samples were collected around 11 and 19 years ago without collection and storage protocols specifically designed for metabolomic studies. It is thus mandatory to evaluate serum samples for NMR and MS-based metabolomics, prior to further analyses for discovery of new dietary and disease risk biomarkers.

This first pilot study was thus built to assess the suitability of serum samples from the EPIC biobank for high-field Nuclear Magnetic Resonance (NMR) spectroscopy and Ultra-Performance Liquid Chromatography-Mass Spectrometry (UPLC-MS) metabolomic analyses. This work was a collaboration between the NMR Center for High-Magnetic Field (CRMN) in Lyon, where NMR experiments using a high-field 800 MHz spectrometer were carried out and Imperial College, London (ICL), where our collaborators performed UPLC-MS experiments and 600 MHz NMR experiments.

2.1.2 Material and methods (see Part 4 for ICL methods)

Population. Sera were collected from volunteers between 1992 and 1998 and stored in liquid nitrogen, according to harmonized protocols.^[116] Within the EPIC cohort serum biobank, a set of samples, associated to incomplete individual data recording, termed “orphan” samples, is dedicated to feasibility and validation studies. We included in this study 40 orphan specimens from 3 different European countries (United Kingdom, n=16; France, n=8; Italy, n=16). Furthermore, 10 samples were duplicated, bringing the total number of analyzed samples to 50. Note that the 3 analytical platforms profiled aliquots of the same 50 samples. We conducted a blind analysis as the sample origins and the duplicate pair identities were only disclosed to all participants at the end of the analysis.

Sample Preparation. Sample preparation for ¹H-NMR spectroscopy was conducted in a similar manner as previously described by Beckonert *et al.*^[17] Briefly, samples were thawed at room temperature and 200 μ L of each mixed with 400 μ L saline solution (NaCl 0.9% wt/vol, D₂O 10% vol/vol), centrifuged for 5 min at 4°C at 12000g, 550 μ L transferred into 5 mm NMR tubes and kept at 4°C until analysis.

Profile Generation. All NMR experiments were carried out on a Bruker Avance III spectrometer operating at 800.14 MHz (proton resonance frequency), equipped with a 5 mm TXI probe and an automatic sample changer (SampleJet, Bruker) cooled at 4°C. Temperature was controlled at 300 K throughout the experiments. For each sample, a set of ¹H 1D NMR experiments with water presaturation including ¹H 1D NOESY with gradients and ¹H CPMG were recorded. 128 transient free induction decays (FID), with a 20 ppm spectral width and an

acquisition time of 1.36 s, corresponding to 43588 data-points, were recorded with a relaxation delay of 2 s. The NOESY mixing time was set to 10 ms and the CPMG spin-echo delay adjusted to 300 μ s for each of the 128 spin-echo loops (76.8 ms total echo time).

NMR data processing. NMR data were processed at CRMN. All FIDs were multiplied by an exponential function corresponding to a 0.3 Hz line-broadening factor prior to Fourier transform. Phasing and baseline correction were automatically performed in Topspin 2.1 (Bruker GmbH, Rheinstetten, Germany). Spectra were automatically calibrated on the α -glucose doublet at $\delta = 5.23$ ppm or on the lactate doublet at $\delta = 1.32$ ppm when glucose was lacking. Spectra were then divided into 10k or 1k bins of respectively 0.001 ppm width or 0.01 ppm width in AMIX software (Bruker GmbH). Water signal was excluded (4.5-5 ppm) and spectra normalized to total intensity. Key metabolite assignment was achieved exploiting published literature,^[55] the HMDB^[50] and bbiorefcode-2-0-0 (Bruker, GmbH, Rheinstetten, Germany) spectral databases.

Multivariate Analysis – PCA and HCA. Hierarchical cluster analysis (HCA) was performed in Matlab using mean-centred NOESY spectra with a bucket size of 0.01 ppm for NMR data and a single linkage algorithm based on Euclidean distances to identify sample clusters. ¹H CPMG NMR auto-scaled or mean-centred data with a bucket size of 0.001 ppm were imported into Simca P+ v12.0.1 (Umetrics AB, Sweden) for principal component analysis.

2.1.3 Results and discussion

Duplicate identification. The first step of analysis was to determine the 10 duplicates in the dataset in a blinded manner. We performed a hierarchical clustering using the single linkage algorithm based on Euclidean distances and recognized the duplicates as the sample pairs with the highest similarity. Using the ¹H NOESY 800 MHz NMR spectra, we were able to identify 10/10 duplicates (**Figure 2.1.1b**), while our collaborators determined 10/10 duplicates with UPLC-MS (**Figure 2.1.1a**) and 9/10 duplicates with ¹H NOESY 600 MHz NMR spectra (**Figure 4.1.1**).

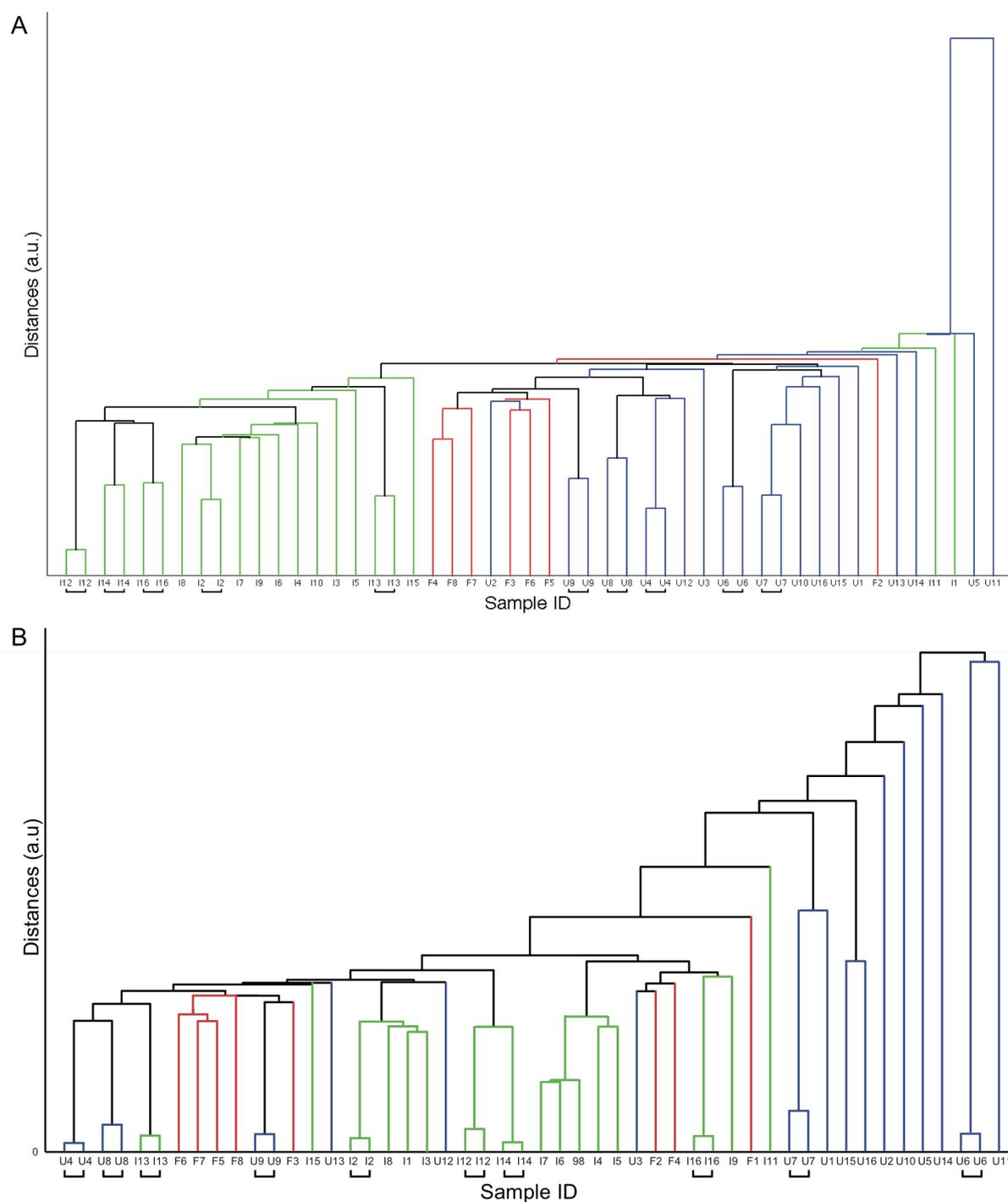


Figure 2.1.1. Hierarchical clustering of UPLC-MS (A) and 800 MHz ¹H NMR metabolic profiles (B). Duplicate samples are indicated by similar sample IDs. The dendrogram is colored according to country of origin (green = Italy, red = France, blue = UK). All species detected by UPLC-MS with RSD<5% across pooled QC samples were used for analysis. Peak intensities for each species were corrected individually for analytical batch differences and data were autoscaled prior to analysis. ¹H NOESY NMR data were mean-centred prior to analysis. Dendrograms were generated using single linkage and the Euclidean distance metric.

These results confirmed the analytical reproducibility of current protocols to delineate sample metabolic profiles, allowing robust characterization of inter-individual metabolic variations.

Metabolic profiles clustering according to the country of origin. Principal component analysis of the ^1H NOESY 800 MHz dataset indicated a clustering according to the country of origin as a major source of variance (**Figure 2.1.2a**). UK and France samples displayed more similar metabolic profiles by comparison to their Italian counterparts, based on an increase in lactate concentration and a decrease in glucose level (**Figure 2.1.2b**).

Similar clustering patterns were observed with UPLC-MS (data not shown). NMR and MS proved thus to be sensitive to subtle metabolic heterogeneity between the pools of samples from the different countries. Although we could not formally exclude at this stage a contribution of pre-analytical variations in sample collection, the observation of metabolic clustering according to the countries of origin, characterized by different diet habits, seem very promising towards the discovery of dietary or disease risk biomarkers. However, we could not further investigate these hypotheses, as for these orphan samples, we did not have any clinical or biological associated data, except the country of origin.

Complementarity in NMR and MS serum metabolite coverage. Metabolites identified by the 2 analytical platforms NMR spectroscopy and UPLC-MS were compared, in terms of redundancy and complementarity (**Table 4.1.2**). NMR provided a good coverage of highly polar molecules such as sugars, amino acids or organic acids, whereas UPLC-MS captured metabolites of lower polarity (phosphocholines, carnitines...) (**Table 4.1.2**). These differences were mainly due to the choice of the chromatographic column (C_{18}), chosen to provide a “one-shot” analysis, providing a good compromise between polar and non-polar metabolites. However, it may be valuable to optimize global metabolome coverage by improving NMR and MS complementarities, through the targeting with adequate chromatographic conditions of molecule classes poorly reported by NMR.

Our findings complemented the evaluation of reproducibility in metabolomics procedures, mainly conducted in urine in the context of large-scale studies^[112] and demonstrated that both NMR and UPLC-MS can produce accurate and convergent metabolic profiles from samples stored in a biobank for about 15 years.

Further large-scale studies are now being conducted in our laboratory, in the framework of the EPIC cohort, to compare metabolic profiles throughout European countries

and correlate them to individual nutritional data (24h recall dietary assessment data) or later onset for pancreatic or liver cancers, offering a unique opportunity to derive individual profiles associated with nutritional statuses and with risk factors for cancers.

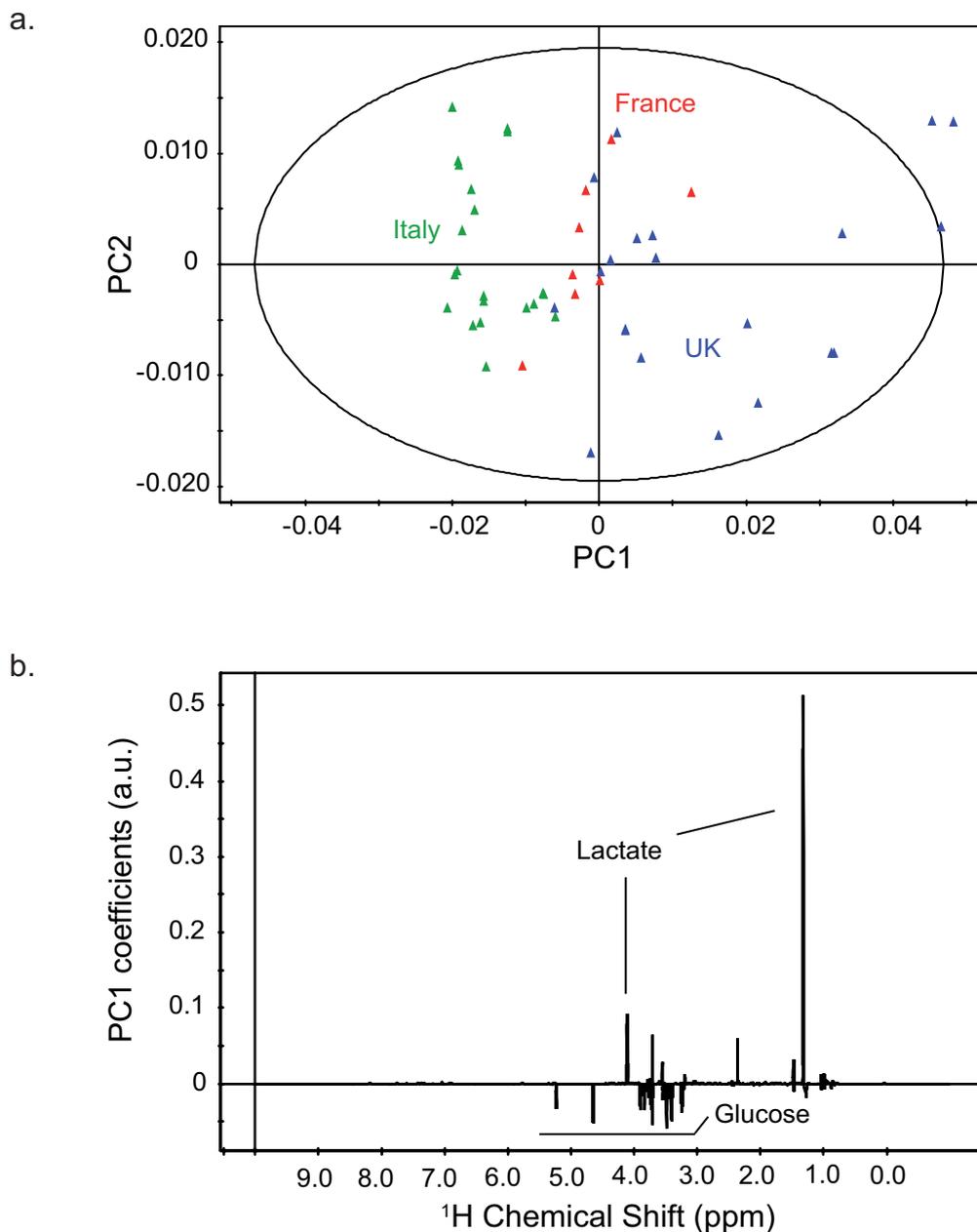


Figure 2.1.2. Principal component analysis (PCA) model of 800 MHz ^1H CPMG NMR metabolic profiles: a) scores plot showing clustering according to the countries of origin and b) loading plot giving the metabolite signature associated to PC1, responsible for the major variance within the dataset. Samples are colored according to the country of origin (green = Italy, red = France, blue = UK). Components PC1 and PC2 describe 74% and 13% of data variation respectively for the mean-centered NMR data. Lactate levels were increased in France and UK samples whereas the glucose concentration was higher in Italian samples.

2.2 Profiling systemic signatures of cancers: metastatic breast cancer serum signature

2.2.1 Introduction

Breast cancer (BC) incidence has slightly but steadily increased over the past three decades while its mortality rate has declined over the same period, thanks to advances in screening methods, early diagnosis and treatments. BC is the most common cancer and the leading cause of cancer death in women worldwide.^[117] Breast cancer displays a high heterogeneity in terms of etiology and histology as well as prognosis, metastatic evolution and response to treatments. Deciphering the molecular basis of such heterogeneity is a major challenge, now achievable through new bio-molecular and analytical techniques, aiming at a comprehensive cancer characterization for risk stratification, therapeutic target identification and appropriate treatment selection.^[111]

Identification of BC tumor types prone to evolve towards metastatic disease and early detection of sub-clinical metastases could represent a major advance in BC management, by selecting patients for follow up intensification and more appropriate treatment administration. Major efforts have been made to characterize genetic alterations and modifications of gene expression profiles in the context of BC tumorigenesis through genomic and transcriptomic approaches,^[118-120] suggesting prognosis factors for developing distant metastases.^[17] Specific changes in metabolism during tumor progression have also early been reported by Warburg^[121] and appear as keystone for understanding cancer evolution or response to drug treatment.^[122, 123] Yet, detailed characterization of BC tumor metabolism as well as broader investigations of the global alterations to individual metabolism still represent a tremendous challenge.

Metabolic phenotyping studies that provide untargeted identification of all detectable low molecular-weight molecules by profiling without any *a priori* the metabolic signatures of biological samples in connection to patho/physiological events,^[4, 5] are prone to play a key role towards this objective. Versatile analytical techniques mainly based on Nuclear Magnetic Resonance spectroscopy (NMR) and Mass Spectrometry (MS) allow the analysis of various samples from bio-fluids such as serum or urine, to intact cells or tissues.^[124, 125] Meanwhile, metabolomic approaches have already found promising applications in different fields from toxicology^[32], functional genomics^[4, 126] to oncology.^[41, 127]

In BC research, metabolomics has been so far generally used for the direct characterization of tumor metabolism alterations, mainly through analysis of intact biopsies by high-resolution magic angle spinning (HR-MAS) NMR spectroscopy. Specific tumor metabolic profiles have been identified by comparison of control and tumor tissues,^[128, 129] and correlation between metabolic tumor profiles and histological grade, hormone grade and axillary lymphatic spread were also observed,^[130] highlighting the potential of tumor metabolic signatures as prognosis factors. Recently work also showed the possibility to refine tumor gene-expression-based classification from tumor metabolic profiles, emphasizing the synergic use of transcriptomic and metabolomic approaches.^[34, 35]

Changes in global metabolism of individuals have also been identified in BC from biofluids such as urine and serum, most efforts being dedicated to the research of early non-invasive diagnosis biomarkers.^[131-133] One recent attempt was made so far to detect micrometastatic disease in early breast cancer patients from metabolomic analysis of peripheral blood serum.^[134] We report here a ¹H-NMR-based metabolomic study aiming at deciphering metabolic serum changes associated with advanced metastatic breast cancer by comparison to the localized disease. This work was carried out together with Elodie Jobard, PhD student at the CRMN

2.2.2 Materials and methods

Populations. This observational study was conducted in the Centre Léon Bérard, (Lyon, France) from January 2009 to February 2010. A cohort of female patients constituted of two subgroups was recruited, including patients suffering from early breast cancer (EBC; Group A) or metastatic breast cancer (MBC; Group B). Inclusion criteria for group A were: patients older than 18, with histologically proven EBC, who underwent a negative CT scan of the chest, abdomen, and pelvis, a negative whole body bone scan and were scheduled for a surgical procedure at Centre Léon Bérard. Inclusion criteria for the group B were: patients older than 18 with histologically proven MBC, eligible for a first line of chemotherapy; previous hormonal treatment was allowed. Diabetic patients were excluded due to the high concentration of glucose in their blood. For each patient, recorded clinical data included age, menopausal status, hormonal and HER2 receptor status, medical history and treatments. Chronologically, a first cohort of 52 patients, named the *training cohort*, was recruited to derive the statistical model, followed by recruitment of a second independent cohort of 33 individuals, named the *validation cohort*, used for model validation. Written informed consent

was obtained from each patient. The institutional ethics committee approved the study protocol before implementation.

Data collection and storage. Blood collection was performed in fasting conditions (12 hours without food intake) for each patient, in the morning before breast surgery for the group A, or in the morning before first chemotherapy cure for the group B. Blood samples were recovered from dry tubes and centrifuged 30 min after collection at 800 *g* for 10 min. Sera were then transferred in plastic straws, and stored in liquid nitrogen for a period of 6 to 12 months before acquisition of NMR data for the whole cohort.

Sample Preparation. Samples were prepared as previously described by Beckonert *et al.*^[17] Serum samples were thawed at room temperature before use. 200 μL of each was diluted with 400 μL of a 0.9% saline solution (NaCl 0.9% wt/vol, D₂O 10% vol/vol) in a microtube, then centrifuged for 5 min at 4°C at 12 000 *g*. Finally, 550 μL of supernatant was transferred into 5 mm NMR tubes. Samples were kept at 4°C until analysis.

¹H-NMR spectroscopy of serum samples. All NMR experiments were carried out on a Bruker Avance III spectrometer operating at 800.14 MHz (proton resonance frequency) equipped with a 5 mm TXI probe, and high-throughput sample changer that maintained the samples temperature at 4°C until actual NMR acquisition. The temperature was then regulated at 300K throughout the NMR experiments. Standard ¹H 1D NMR pulse sequences, NOESY and CPMG with water presaturation, were applied on each sample to obtain corresponding metabolic profiles. 128 transient free induction decays (FID) were collected for each experiment with a spectral width of 20 ppm, corresponding to 43588 data-points for an acquisition time of 1.36 s. For both sequences, the relaxation delay was set to 2 s. The NOESY mixing time was set to 100 ms and the CPMG spin-echo delay to 300 μs allowing an efficient attenuation of the lipid NMR signals. The 90° pulse length was automatically calibrated for each sample at around 9.25 μs . In addition, 2D NMR experiments (¹H-¹³C HSQC, ¹H-¹H TOCSY and J-Resolved) were recorded on a subset of samples to achieve structural assignment of the metabolic signals.

Data processing. All FIDs were multiplied by an exponential function corresponding to a 0.3 Hz line-broadening factor, prior Fourier transformation. ¹H-NMR spectra were automatically phased and referenced to the α -glucose anomeric proton signal ($\delta = 5.23$ ppm) using Topspin 2.1 (Bruker GmbH, Rheinstetten, Germany). Residual water signal (4.66 to 5.11 ppm) was excluded. Spectra were divided into 0.001 ppm-wide buckets over the chemical shift range

[0.1; 10 ppm] using the AMIX software (Bruker GmbH). Spectra were normalized to their total intensity and mean-centered prior to analysis. Data were then exported to SIMCA-P 12 (Umetrics, Umea, Sweden) for statistical analysis.

Population characterization (Elodie Jobard, CRMN). Descriptive statistical analysis was performed to characterize the two populations, using the Student t-test and Chi² test for quantitative and qualitative data respectively. The significance threshold was set to 0.05 for both tests.

Multivariate analysis of serum metabolic profiles. Unsupervised and supervised statistical multivariate methods were used to build models for sample classification and extract group-specific metabolic signatures. Principal component analysis (PCA) was performed to derive the main sources of variance within the dataset, check population homogeneity and eventually identify technical or biological outliers. Data were visualized as score plots, where each point stands for the projection of a single sample on the main principal components and as loading plots, which represent the contribution of the metabolic variables to principal components. Supervised regression methods such as Orthogonal Partial Least-Squares (O-PLS)^[100] were performed to build a robust sample classification model and derive group-specific metabolic phenotypes. The O-PLS analysis was run to discriminate populations by regressing a supplementary data matrix **Y**, containing information about the disease severity (EBC or MBC status) on the **X** NMR dataset matrix. Model performances were assessed by goodness-of-fit parameters R² and Q², related respectively to the explained and predicted variance, calculated through a cross-validation procedure. We performed model validation by resampling the model 1000 times under the null hypothesis in Matlab (The Mathworks, Inc). The decrease of goodness-of-fit R² and Q² parameters, when correlation between original model and random models decreased, indicated the good quality of our model.

Evaluation of classification performance (Elodie Jobard, CRMN). Model classification performance was characterized by the calculation of the receiver operating characteristic (ROC) curve and its area under the curve (AUC) from cross-validated results. Finally, evaluation of the overall model predictability was assessed on a completely independent dataset by predicting in blind condition the EBC or MBC status for the new individuals.

Identification of statistically significant biomarkers discriminating EBC and MBC. To derive statistically significant discriminating biomarkers, we used an univariate methodology previously described that couples an automatic binning procedure named statistical recoupling

of variables (SRV) to subsequent ANOVA analysis and multiple testing correction of the p -values,^[94] here the Benjamini-Hochberg correction that controlled the false-discovery rate below a 0.05 threshold,^[102] implemented with MATLAB (The MathWorks Inc., Natick, MA) homemade routines.

Annotation of significant metabolites was achieved through the identification of full spin systems from analysis of 2D NMR experiments (^1H - ^{13}C HSQC, ^1H - ^1H TOCSY and J-resolved experiments) as well as STOCSY analysis, which provides statistical correlations between NMR variables suggesting structural or biological connectivities.^[62] Metabolite assignment procedure exploited knowledge from academic spectral databases such as MMCD,^[49] HMDB^[50] and BMRB^[51] as well as proprietary databases (Chenomx NMR Suite, Chenomx Inc, Edmonton, Canada; AMIX SpectraBase, Bruker GmbH, Rheinstetten, Germany).

2.2.3 Results

Population and clinico-pathological characterization. In this study, we collected 52 serum samples from 28 female patients suffering from early breast cancer (EBC; Group A) and 24 female patients suffering from metastatic breast cancer (MBC; Group B), defining a training cohort. We then recruited a validation cohort, made of 33 additional independent individuals, 18 with EBC and 15 with MBC. For each cohort, we excluded one MBC sample, due to improper collection procedure as detailed herein. We compared clinicopathological data between group A and B for the training and validation cohorts to exclude biases related to patient selection, as summarized in **Table 4.2.1** and **Table 4.2.2**. No significant differences have been detected for age, menopausal or HER2 gene status, medical history and treatments (data not shown). The tumor hormonal status was significantly different between EBC and MBC groups ($p=0.007$).

NMR spectroscopy of serum samples. A typical high-field (800 MHz) ^1H NMR spectrum, as recorded for each of the serum samples to derive well-resolved metabolic profiles, is illustrated in **Figure 2.2.1**. Sharp lines typical of the signals from small metabolites are superimposed on broader signals from larger lipids or proteins. 40 metabolites were identified from these spectra, supplemented by information from 2D ^1H - ^1H and ^1H - ^{13}C NMR experiments. Assignments are provided in **Table 4.2.3**.

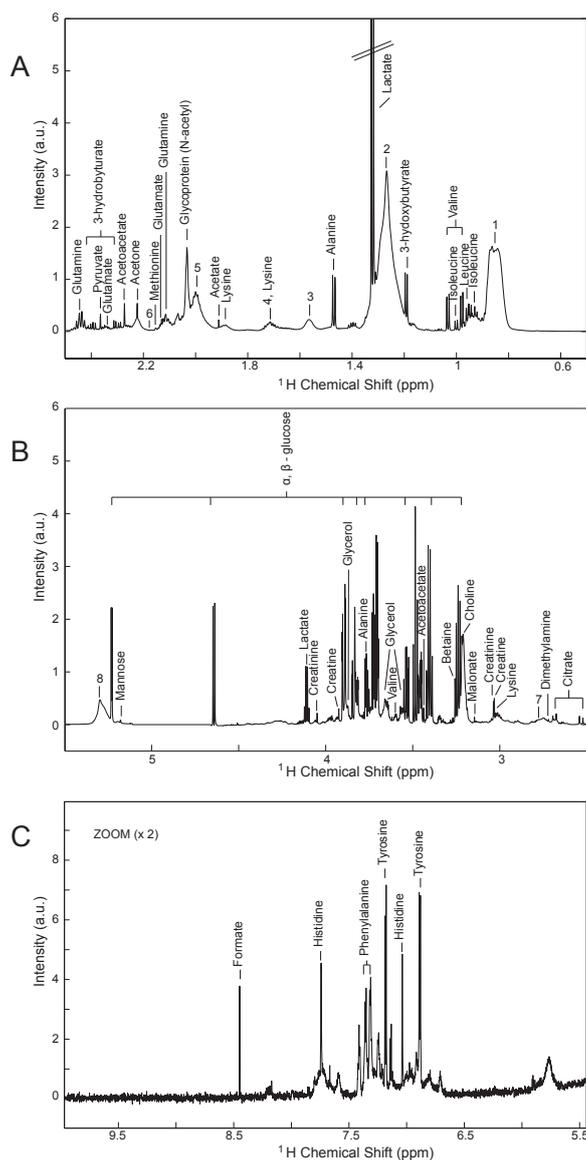


Figure 2.2.1. Overview of the ¹H NMR CPMG mean spectrum (800 MHz) for patients with metastatic breast cancer. (A) 0.5–2.5 ppm region; (B) 2.5–5.5 ppm region; (C) 5.4–10 ppm region. 1: Lipid (LDL) $\text{CH}_3(\text{CH}_2)_n$, Cholesterol (C26, C27), Cholesterol (C21); 2: Lipid $\text{CH}_3\text{CH}_2\text{CH}_2(\text{CH}_2)_n$; 3: Lipid (VLDL) $\text{CH}_2\text{CH}_2\text{CO}$; 4: Lipid $\text{CH}_2\text{CH}_2\text{C}=\text{C}$; 5: Lipid $\text{CH}_2\text{C}=\text{C}$; 6: Lipid CH_2CO ; 7: Lipid $\text{C}=\text{CCH}_2\text{C}=\text{C}$; 8: Unsaturated lipid ($\text{CH}=\text{CHCH}_2\text{CH}=\text{CH}$), Unsaturated lipid ($=\text{CHCH}_2\text{CH}_2$).

Variance and outliers within 1D ¹H NMR dataset. To derive the main sources of variance and identify potential analytical or biological outliers we performed a principal component analysis (PCA) on the 1D ¹H NMR datasets from the training or validation cohorts. We identified in each cohort one severe outlier, with a high concentration of citrate for the first one, probably due to mis-collection of the blood in a citrated tube, and a high concentration of lactate for the second outlier. These two samples were excluded before further analysis.

Discrimination between EBC and MBC from metabolic serum profiles. A supervised analysis by O-PLS^[100] analysis was carried out on the training cohort (51 spectra) to derive a robust statistical model based on discrimination between EBC and MBC metabolic profiles. We observed a clear discrimination between EBC and MBC individuals, as illustrated in **Figure 2.2.2a**, assessed by high values of goodness-of-fit model parameters R^2 and Q^2 related respectively to the explained and predicted variance in the model ($R^2 = 0.804$; $Q^2 = 0.447$). The discrimination robustness was further validated by re-sampling 1000 times the model under the null hypothesis showing a clear decrease of R^2 and Q^2 with the correlation between the original and permuted class information Y matrices (**Figure 2.2.2b**). In addition, we characterized the model classification performances with a ROC curve from cross-validated results (**Figure 2.2.2c**). We obtained a corresponding area under the curve of 0.92, assessing the powerful classification ability of the model.

Data from the validation cohort were analyzed independently, in order to test the classification model generated from the training cohort. The disease status of these individuals, either EBC or MBC, was predicted *in a blind analysis* from the NMR metabolic profiles by projecting the data onto the O-PLS multivariate model. From the 32 predicted subjects, 26 samples were correctly classified (10 MBC and 16 EBC) while 4 were misclassified as EBC instead of MBC and 2 misclassified as MBC instead of EBC, *corresponding to a sensitivity of 71% and a specificity of 89%* (**Figure 2.2.2d**). These good prediction parameters are highlighted visually with a ROC curve from independent validation (**Figure 2.2.2c**).

As an additional test to assess the robustness and reproducibility of the technique, 4 additional samples from patients belonging to the training cohort, were analyzed together with the validation cohort. The disease status for these four patients was correctly predicted from our classification model (data not shown).

Determination of a metabolic signature associated with breast cancer severity. Individual statistically significant metabolites discriminating EBC and MBC were identified with univariate analysis of the NMR metabolic profiles. Histidine was found in significantly higher concentration in the serum of EBC patients, while end-products of β -oxidation (acetoacetate and 3-hydroxybutyrate) and lipid degradation (glycerol), N-acetylglycoproteins (NAC 1 & 2), or pyruvate and phenylalanine concentrations in the serum increased for MBC individuals as compared to EBC (**Figure 2.2.3** and **Table 4.2.4**). Three NMR signals could not be identified from the literature or currently available databases. Additional ^1H NMR data was recorded

after microfiltration of serum samples (3 kDa filter), to ease the identification of possibly low-concentrated metabolites. The significant doublets observed at 1.38 and 1.39 ppm on unfiltered samples were absent from the spectra recorded for the filtered sera, which sustains hypothesis that these specific signals are linked to serum (glyco)proteins.

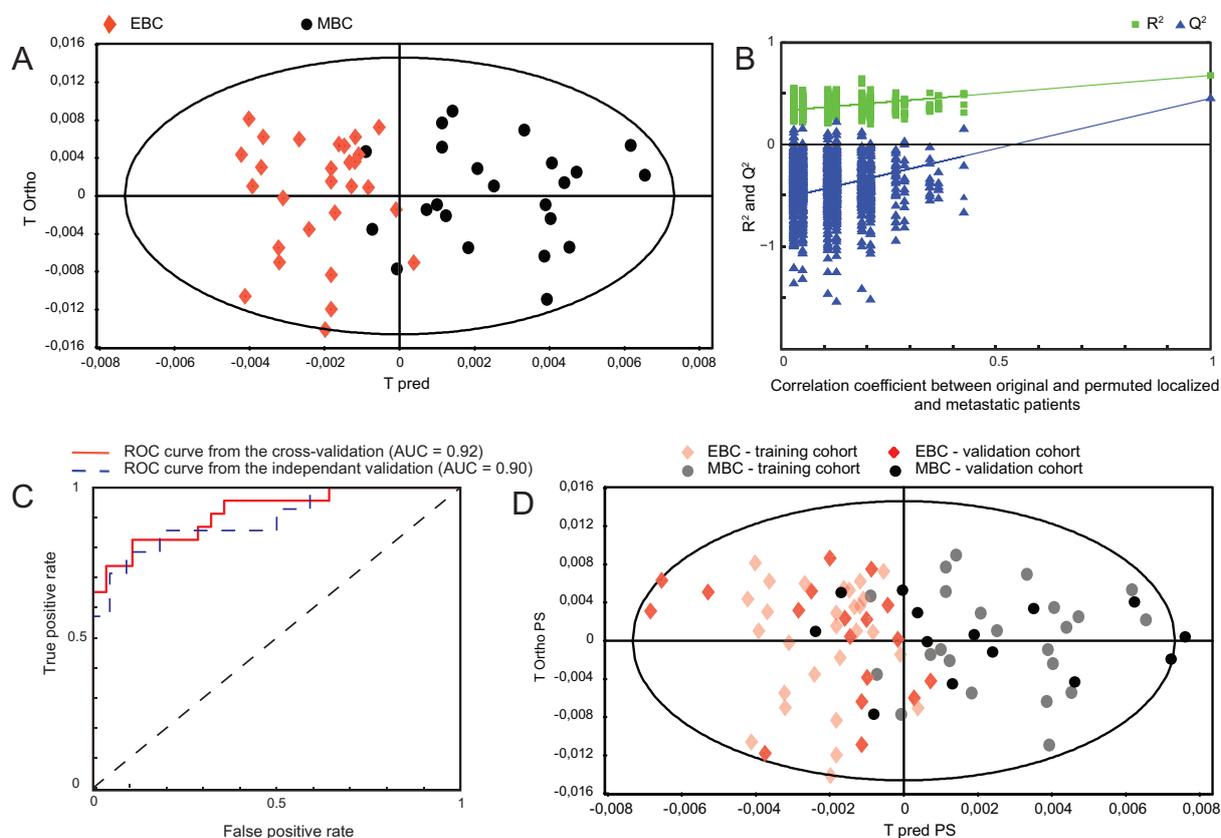


Figure 2.2.2. Serum metabolite profiles derived from metastatic breast cancer patients are different from localized breast cancer patients. (A) A 1+3 O-PLS model from the training cohort discriminating 28 localized breast cancer patients (red diamonds) and 23 metastatic breast cancer patients (black dots); $R^2 = 0.804$ and $Q^2 = 0.447$. (B) The O-PLS model was validated by re-sampling under the null hypothesis. (C) Receiver operating characteristic curve showing the good prediction from cross-validation and from independent validation. (D) O-PLS prediction of 32 additional subjects: 18 localized breast cancer patients (red diamonds) and 14 metastatic breast cancer (black dots) projected onto the training model (light colors).

2.2.4 Discussion

This study demonstrates that NMR metabolomic analysis from peripheral blood can robustly classify patients according to the severity of breast cancer: localized versus metastatic diseases. A major challenge for metabolic phenotyping approaches lies in the intrinsic variability of the tumors and hosts. While the present study on cancer patient sera is designed to circumvent this former aspect, a number of metadata may directly impact the metabolic

status of individuals (such as sex, age, dietary, lifestyle, medication, diseases, etc...),^[135] the related physiological variations hindering associations of metabolic fingerprints to specific pathological status or stimuli. All EBC and MBC patients in our cohort were thoroughly characterized from their metadata to verify the absence of such bias in our metabolomic study (**Table 4.2.1** and **Table 4.2.2**). While no significant differences were observed for age, menopause, HER2 status, medical history or drug treatment, the hormonal receptor status (HR) was however significantly different ($p = 0.007$) between EBC and MBC populations in the training cohort. This status reflects the intrinsic characteristics of hormone-sensitive tumorigenesis, associated to better prognosis from generally slower tumor growing rates and higher response to hormone-suppression treatments. As a result, HR- patients are consistently overrepresented in the MBC group.

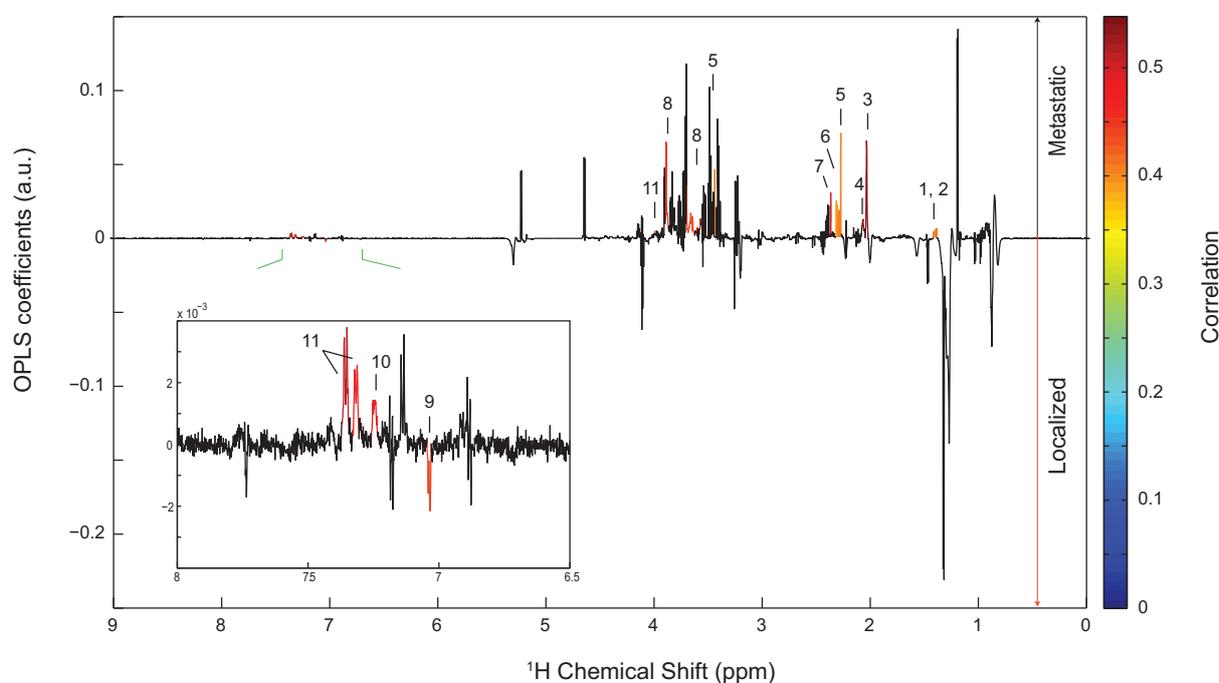


Figure 2.2.3. O-PLS loading plot after SRV analysis and Benjamini-Hochberg multiple testing correction. Statistically significant signals correspond to the colored spectral regions. Highlighted candidate biomarkers are 1) Unassigned doublets at 1.38 ppm and 1.39 ppm; 3) N-AcetylGlycoprotein (NAC1); 4) NAC2 ; 5) Acetoacetate; 6) Pyruvate; 7) 3-Hydroxybutyrate; 8) Glycerol; 9) Histidine; 10) Unassigned multiplet at 7.24 ppm; 11) Phenylalanine.

Statistical comparison of the NMR metabolic profiles of sera from breast cancer patients in our cohort provides a robust and predictive classification according to the localized or metastatic nature of the disease, as illustrated in **Figure 2.2.2**, with a sensitivity of 71% and a specificity of 89%. A range of investigations, using NMR or MS based metabolomic approach, have focused on establishing breast cancer biomarkers, from breast tumors or cell

lines.^[128, 129] Major metabolites identified as correlated with breast cancer in these studies are glucose, lactate, lipids, choline, and amino acids.^[136] Our findings highlight more specifically 11 significant metabolic markers associated with a metastatic status of the disease, from which 8 could be unambiguously identified. MBC patients, as compared to EBC cases, display higher serum concentrations of acetoacetate, 3-hydroxybutyrate (beta-hydroxybutyrate), glycerol, pyruvate, N-acetylglycoproteins and phenylalanine, and lower concentration of histidine. A subset from these significant metabolites has already been identified from patient sera as associated with either recurrent (histidine and 3-hydroxybutyrate) or metastatic (phenylalanine) breast cancer in two recent studies.^[133, 134] Oakman *et al.* compared, in a monocentric study design analogous to ours, the metabolic profiles of 44 patients with localized breast cancer versus 51 patients with metastatic breast cancer,^[134] with a sensitivity of 75% and specificity of 69%, slightly lower than the result of statistical analysis from our cohort. In their study, metastatic subjects were characterized by high values of phenylalanine, glucose, proline, lysine and N-acetylcysteine and low values of lipids. It may be noted that although some of these metabolites do not appear as significant biomarkers in our model, our multivariate metabolic fingerprints confirm the same trends of variation associated to MBC vs. EBC status (*e.g.* higher concentration of glucose, lower levels of lipids) (**Figure 2.2.3**).

An increased serum level of phenylalanine for patients with MBC was consistently highlighted in both studies, and therefore may constitute a potential robust marker of malignancy in breast cancer. Yet, phenylalanine was also previously shown from various serum or tissue metabolomic analyses, as associated with diverse cancers, regardless of their metastatic status.^[137-139] In those cases, high concentrations of phenylalanine was always correlated to cancer (vs. healthy subjects), with one notable exception of a study on serum from human colorectal cancer patients that showed a decreased level of phenylalanine with respect to controls.^[140] More generally, many studies to date have reported changes in amino acid metabolic profiles of patients with cancer, as reviewed by Lai and coworkers.^[141] While some aberrations in the metabolic profile of amino acids are expected in cancer patients, no consistent cancer-specific amino acids has yet emerged. The metabolic profile of free amino acids can vary between the early and late stage of cancer. Histidine has also been identified as a potential marker in the recent literature on breast cancer^[133] as well as for others types of cancers.^[142, 143] The fact that L-histidine is lower in MBC patients compared with EBC patients may be explained by the higher activity of histidine decarboxylase. This enzyme may

accelerate the decarboxylation of histidine in the MBC patients, as observed already in the case of colorectal cancer.^[143, 144]

The significant increase, for MBC patients, of serum concentrations of the two ketone bodies acetoacetate and 3-hydroxybutyrate, which are by-products of fatty acid catabolism, may arise from altered lipolysis linked to high energy demands in the cells. When carbohydrates are not sufficient as primary energy source, beta-oxidation of fatty acids can produce large amounts of acetyl-CoA, which is then primarily used by the Krebs cycle (mitochondrial respiration). When saturated the surplus undergoes ketogenesis, and ketone bodies in excess are excreted from the cells, accumulating in the blood.^[145] The presence of high ketones concentrations could therefore illustrate an increased fatty acids catabolism, consistent with the observed high levels of glycerol, and the incapacity of the Krebs cycle to fully metabolize acetyl-CoA by depletion of other TCA cycle intermediates. Meanwhile, recent investigations proposed a “reversed Warburg effect” mechanism that link metastatic dissemination to an increased mitochondrial metabolism in epithelial cancer cells, while ketone bodies produced by glycolytic stromal cells would feed the oxidative mitochondrial metabolism, to stimulate tumor growth and metastasis.^[146] N-acetyl glycoproteins are acute-phase glycoproteins, whose increased levels in circulating blood associated to human cancer is well established^[147] and that were observed in several studies in the presence of high levels of ketone bodies and glycerol. Finally, the increased concentration of pyruvate can be associated to an exalted glycolytic activity. Under these mechanistic hypotheses, our observations in blood sera therefore fully converge towards high-energy needs due to aggressive tumor growth in MBC patients.

2.2.5 Conclusion

In conclusion, our investigation proposes a robust serum metabolic signature of metastatic breast cancer as compared to patients with localized breast tumors. These results are promising as a non-invasive method to improve the diagnosis, prognosis and management of patients. They confirm the outcome of an earlier comparable monocentric investigation, while our statistical model discriminating MBC from EBC sera shows slightly higher specificity, and extend the range of observed significant metabolic markers, notably including several amino-acids and ketone bodies. The observed biomarker pattern is possibly not unique to MBC but may result of more general characteristics of malignant disease. In the future, larger cohorts, if possible recruited in a multicenter context, should be exploited to further validate

these results. In addition, follow-up of EBC patients with possible occurrence of metastatic disease will enrich our present model by the determination of early (predictive) markers of MBC occurrence.

2.3 Plasma metabolic signature of hepatocellular carcinoma and chronic liver disease.

2.3.1 Introduction

Primary liver cancer (PLC) represents a major public health issue worldwide. With 700.000 to 800.000 new cases each year, PLC is the 7th most common cancer and the 3rd cause of death by cancer.^[117] The two main histological types are hepatocellular carcinoma (HCC), which corresponds to 80% of cases, and cholangiocarcinoma (CHCA). Many risk factors have been identified for HCC, such as hepatitis B/C (HBV/HCV) virus infection, dietary exposure to aflatoxin B1 and all other potential sources of chronic liver disease like chronic alcohol consumption or non-alcoholic steatohepatitis (NASH). Geographical HCC incidence is thus superimposed on risk factor distribution around the globe. HCC represents a major burden for areas of endemic HBV or HCV carriage such as Sub-Saharan Africa and Eastern Asia.

Early detection of liver cancers is thus one of the main public health challenges for these areas of high incidence. The emergence of new technologies offers the opportunity to explore new fields of biomarker discovery. Metabolic profiling, relying on state-of-the-art high-throughput analytical techniques such as Nuclear Magnetic Resonance (NMR) spectroscopy or mass spectrometry appears as a key approach to probe metabolic perturbations induced by HCC and other liver diseases at the systemic level (plasma, serum...) and derive diagnostic biomarkers.

Different studies have observed changes in plasma or serum metabolic profiles between healthy individuals and patients suffering from chronic liver disease (CLD) or HCC, essentially in Chinese populations.^[138, 148-151] Similar studies have focused on the metabolic fingerprints of CLD and HCC in urine.^[148, 152-156] Ye *et al.* have monitored the urinary metabolic signatures of early HCC recurrence after surgical removal.^[157] Liver cirrhosis severity has also been evaluated from serum or urine metabolic profiles,^[158-160] as well as differences between cirrhosis induced by HBV infection or by alcohol consumption.^[161, 162]

Our objective was to characterize the specific plasma metabolic signatures of individuals suffering from HCC by comparison to healthy controls, chronic liver disease or cholangiocarcinoma carriers, in a case-control study conducted in Thailand, an area of high HBV-related HCC incidence, using NMR spectroscopy. We investigated correlations between plasma metabolic profiles and HBV/HCV infection status, LTBP2 protein levels, a potential

marker of HCC identified by proteomic analysis carried out on the same cohort, and occurrence of R249S mutation of the *TP53* gene, particularly associated with HCC induced by HBV infection.

2.3.2 Material and Methods

Sample collection. This study is part of the International Liver Cancer Study (ILCS), an international project aiming at contributing to intervention, prevention, early diagnosis and control of liver cancer. This case-control study was carried out in the National Cancer Institute, Thailand from April 2008 to December 2009. Diagnosis of hepatocellular carcinoma and cholangiocarcinoma was based on concordant clinical examination and abdominal imaging. Individuals assigned in the control group presented no clinical evidence of liver disease and were selected among individuals who came to the Institute for their annual check-up. Written consent was obtained from all participants. The Institutional Review Boards of the Thailand National Cancer Institute and the International Agency for Research on Cancer approved this study. Samples were collected in EDTA tubes and kept at -80°C until analysis.^[163]

Sample preparation. Samples were prepared as previously described by Beckonert *et al.*^[17] Plasma samples were thawed at room temperature before used. 200 μL of each was diluted with 400 μL of a 0.9% saline solution (NaCl 0.9% wt/vol, D_2O 10% vol/vol) and centrifuged at 12,000g for 5 min at 4°C . Finally, 550 μL of supernatant was transferred into a 5 mm NMR tube. Samples were kept at 4°C until analysis.

NMR spectroscopy of plasma samples. All NMR experiments were carried out on a Bruker Avance III spectrometer operating at 600 MHz (proton resonance frequency), equipped with a 5 mm CPI cryoprobe and high throughput sample changer (SampleJet, Bruker) that maintained the sample temperature at 4°C until NMR acquisition. Temperature was controlled at 310 K throughout the experiments. For each sample, standard ^1H 1D NMR pulse sequences, NOESY and CPMG with water signal presaturation were recorded for each samples to obtain corresponding metabolic profiles. 128 transient free induction decays (FID) were collected for each experiment with a spectral width of 20 ppm, corresponding to 32690 data-points for an acquisition time of 1.36 s. For both sequences, the relaxation delay was set to 2 s. The NOESY mixing time was set to 100 ms and the CPMG spin-echo delay to 400 μs for a total echo time = 64 ms, allowing an efficient attenuation of the protein and lipid NMR signals. The 90° pulse length was automatically calibrated for each sample at around 13 μs .

Serum quality controls (QC) were analyzed at the beginning and at the end of each batch of NMR experiments to assess analytical reproducibility. In addition, 2D NMR experiments, (^1H - ^1H TOCSY, ^1H - ^{13}C HSQC and J-resolved experiments) were carried out on a subset of selected samples to characterize structural connectivities between nuclei and refine metabolite identification.

Data processing. All FIDs were multiplied by an exponential function corresponding to a 0.3 Hz line-broadening factor prior to Fourier transform. ^1H -NMR phasing and baseline correction were automatically performed in Topspin 2.1 (Bruker GmbH, Rheinstetten, Germany). Spectra were automatically referenced to the α -glucose anomeric proton signal ($\delta = 5.23$ ppm). Residual water signal ($\delta = 4.3$ - 5.15 ppm) and EDTA signals ($\delta = 2.54$ - 2.57 , 2.68 - 2.71 , 3.06 - 3.24 , 3.57 - 3.63 ppm) were excluded. Spectra were then divided into 10k bins of 0.001 ppm width using the AMIX software (Bruker GmbH). Spectra were normalized with the Probabilistic Quotient Normalization (PQN) method considering as reference spectrum the median spectrum from the whole dataset^[97] and scaled with the Pareto method prior to analysis.

Statistical Analysis. Unsupervised and supervised statistical multivariate methods were used to build models for sample classification and extract group-specific metabolic signatures. Principal component analysis (PCA) was performed to derive the main sources of variance within the dataset, check population homogeneity and eventually identify technical or biological outliers. Data were visualized as score plots, where each point stands for the projection of a single sample on the main principal components and as loading plots, which represent the contribution of the metabolic variables to principal components. Supervised regression methods such as Orthogonal Partial Least-Squares (O-PLS)^[100] were performed to build a robust sample classification model and derive group-specific metabolic phenotypes.

To derive statistically significant discriminating biomarkers, we used an univariate methodology previously described that couples an automatic binning procedure named statistical recoupling of variables (SRV)^[94] to non-parametric Wilcoxon rank sum tests and p -value corrections for multiple testing, in our case the Benjamini-Yekutieli correction that controlled the false-discovery rate below a 0.05 threshold,^[104] implemented with MATLAB (The MathWorks Inc., Natick, MA) homemade routines. We generated loading plots representing back-scaled OPLS coefficients obtained from the Pareto-scaled dataset where each NMR variable coefficient was scaled back by the square root of its standard deviation to

ease metabolite identification. Only statistically significant metabolites were colored, according to their correlation with sample statuses.

Identification of statistically significant metabolites. Metabolite assignment exploited academic spectral databases such as HMDB,^[50] MMCD^[49] as well as proprietary databases (AMIX SpectraBase, Bruker, GmbH, Rheinstetten, Germany; Chenomx NMR Suite 7.0, Chenomx Inc, Edmonton, Canada).

2.3.3 Results

Patient cohort. We analyzed 312 plasma samples by 600 MHz ¹H NMR spectroscopy. The dataset was divided in 5 groups according to sample status: Control (n = 130), Chronic Liver Disease (CLD without HCC, n = 51), HCC (patients suffering from HCC, n = 86) further divided in HCC without cirrhosis (patients suffering from HCC without cirrhosis, n = 50) and HCC with cirrhosis (patients suffering from HCC with cirrhosis, n = 36), CHCA (patients suffering from cholangiocarcinoma, n = 37) and others (n = 8). Results concerning the CHCA group are not reported in this thesis. The group “others” gathered individuals suffering from two cancers simultaneously and was excluded during the analysis.

Clinical and biological parameters were recorded: age, gender, body mass index (BMI), year of sample collection, hepatitis B virus (HBV) and hepatitis C virus status (HCV) and liver function biological parameters such as albumin level. We also had measurements of LTBP2 protein concentrations, identified in a proteomic study within the same sample-set as a potential marker of HCC and R249S copy numbers, a mutation of the *TP53* gene particularly associated with HCC induced by HBV infection, measured from circulating free DNA (CFDNA) in plasma.^[163]

Principal component analysis. The analysis was carried out on CPMG ¹H NMR spectra. We first performed a principal component analysis (PCA) on the whole dataset to identify the main sources of variance and trends between the samples. The quality control (QC) sample dispersion was very limited (**Figure 2.3.1a**) compared to the whole dataset dispersion, assessing analytical reproducibility. The first principal component was dominated by variations in lipids and the second by variations in glucose. The third component (PC3) showed a discrimination between Control samples and a large number of pathological samples (**Figure 2.3.1a**). We performed a second PCA model focusing only on Control, CLD and HCC (pooled with and without cirrhosis) data. Again, we observed an interesting trend

along the PC3, suggesting a discrimination between Control and HCC samples, CLD samples being superimposed on both HCC and Control individuals (**Figure 2.3.1b**).

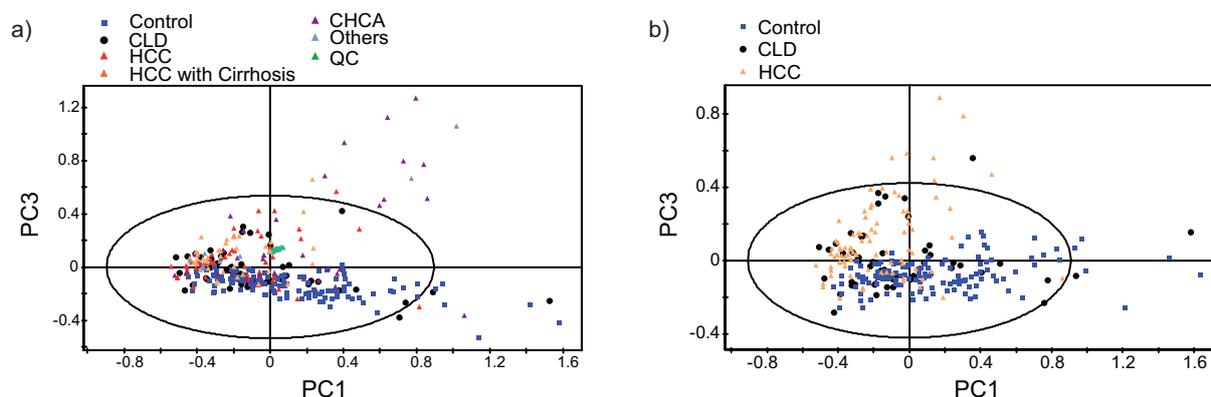


Figure 2.3.1. Principal component analysis for a) the whole dataset and b) control, CLD and HCC (with and without cirrhosis) groups.

Discrimination between Control, CLD and HCC (with and without cirrhosis). We first compared the plasma metabolic signatures of Control, CLD and total HCC samples (corresponding to individuals suffering from HCC with and without liver cirrhosis).

We observed a robust discrimination between Control and HCC samples, in an OPLS model, built with 5 orthogonal components ($R^2X = 0.769$, $R^2Y = 0.809$, $Q^2 = 0.732$, **Figure 2.3.2a**, **Figure 4.3.1a**, and **Table 2.3.1**). HCC samples were characterized by high levels of citrate, ethanol, formate, phenylalanine, pyruvate, tyrosine and N-acetylglycoproteins and low levels alanine and valine, compared to control samples. Lipid metabolism also contributed to this signature, with a global decrease in unsaturated and oxidized lipids in HCC individuals (**Figure 2.3.2b**).

To further evaluate whether this signature was specific of HCC, we compared HCC and CLD groups. We were able to build an OPLS model discriminating the two groups (1 orthogonal component, $R^2X = 0.444$, $R^2Y = 0.372$, $Q^2 = 0.305$, **Figure 2.3.2c**, **Figure 4.3.1b**, and **Table 2.3.1**). However, this discrimination was less important than the previous observed between HCC and Control samples. HCC samples displayed high concentrations of ethanol and N-acetylglycoproteins and low concentrations of alanine, glucose and valine. Global changes in the lipid profile also contributed to this signature (**Figure 2.3.2d**).

We finally compared the metabolic profiles of Control and CLD patients and observed a significant increase in acetate, citrate, ethanol, formate, glucose, phenylalanine, pyruvate and tyrosine for CLD individuals associated with a decrease in N-acetylglycoproteins, unsaturated and oxidized lipids (OPLS-DA, 3 orthogonal components, $R^2X = 0.766$, $R^2Y = 0.512$, $Q^2 = 0.392$, **Figure 2.3.2e-f**, **Figure 4.3.1c**, and **Table 2.3.1**).

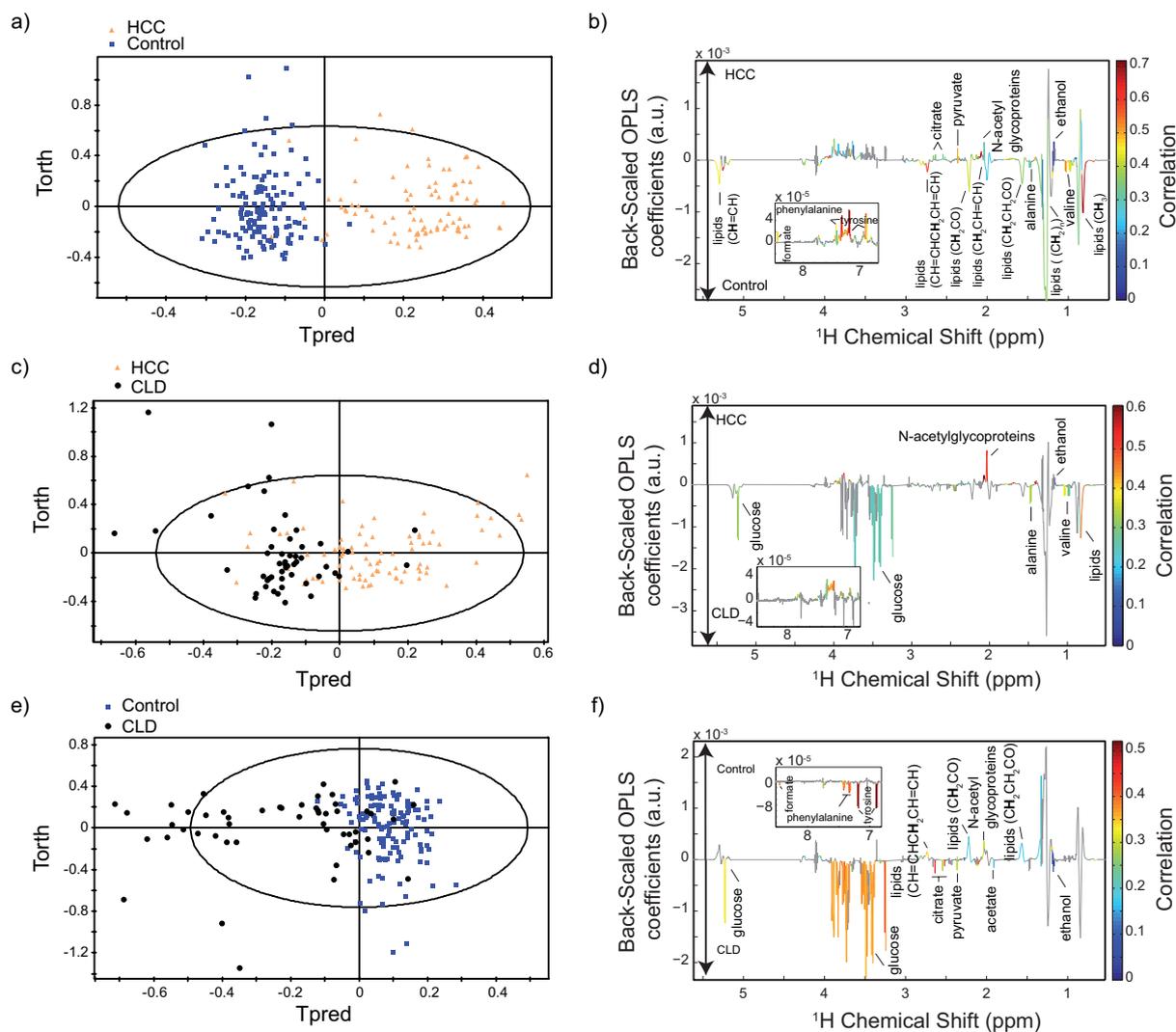


Figure 2.3.2. Discrimination between Control, CLD and HCC. OPLS models discriminating a) HCC and Control (5 orthogonal components, $R^2X = 0.769$, $R^2Y = 0.809$, $Q^2 = 0.732$), b) HCC and CLD (1 orthogonal component, $R^2X = 0.444$, $R^2Y = 0.372$, $Q^2 = 0.305$) and c) Control and CLD (3 orthogonal components, $R^2X = 0.766$, $R^2Y = 0.512$, $Q^2 = 0.392$).

These different analyses showed that HCC and CLD samples shared a common metabolic pattern by comparison to Control samples, involving an increase in citrate, ethanol, formate, phenylalanine, pyruvate and tyrosine and a decrease in unsaturated and oxidized lipids. This non-specific metabolic signature could be a fingerprint of liver disease. CLD individuals were specifically associated with high glucose and acetate levels, while HCC samples display high N-acetylglycoprotein levels and low alanine and valine levels. It seems thus that HCC could induce a specific plasma metabolic signature. However, in these comparisons, the HCC group gathered patients with and without cirrhosis, which increased the total individual number and statistical power but created an inhomogeneous population.

We thus performed a new set of analysis, distinguishing HCC individuals with or without cirrhosis, decreasing statistical power but increasing group homogeneity.

Discrimination between Control, CLD and HCC without cirrhosis. A strong discrimination was observed between Control and HCC without cirrhosis individuals (OPLS-DA, 3 orthogonal components, $R^2X = 0.723$, $R^2Y = 0.763$, $Q^2 = 0.714$, **Figure 2.3.3a**, **Figure 4.3.1d**, and **Table 2.3.1**). The HCC without cirrhosis group was associated with higher levels of formate, N-acetylglycoproteins, phenylalanine, pyruvate and tyrosine and lower levels of alanine, valine, oxidized and unsaturated lipids (**Figure 2.3.3b**). Here, citrate and ethanol did not significantly change between the two groups. CLD and HCC without cirrhosis groups were significantly discriminated (OPLS-DA, 2 orthogonal components, $R^2X = 0.691$, $R^2Y = 0.532$, $Q^2 = 0.457$, **Figure 2.3.3c**, **Figure 4.3.1e**, and **Table 2.3.1**) in a model displaying better performance than for the discrimination between total HCC and CLD groups. The HCC without cirrhosis group was associated with an increase in N-acetylglycoproteins, a decrease in acetate, alanine, citrate, glutamine, glucose, valine and global changes in the lipid profile (**Figure 2.3.3d**).

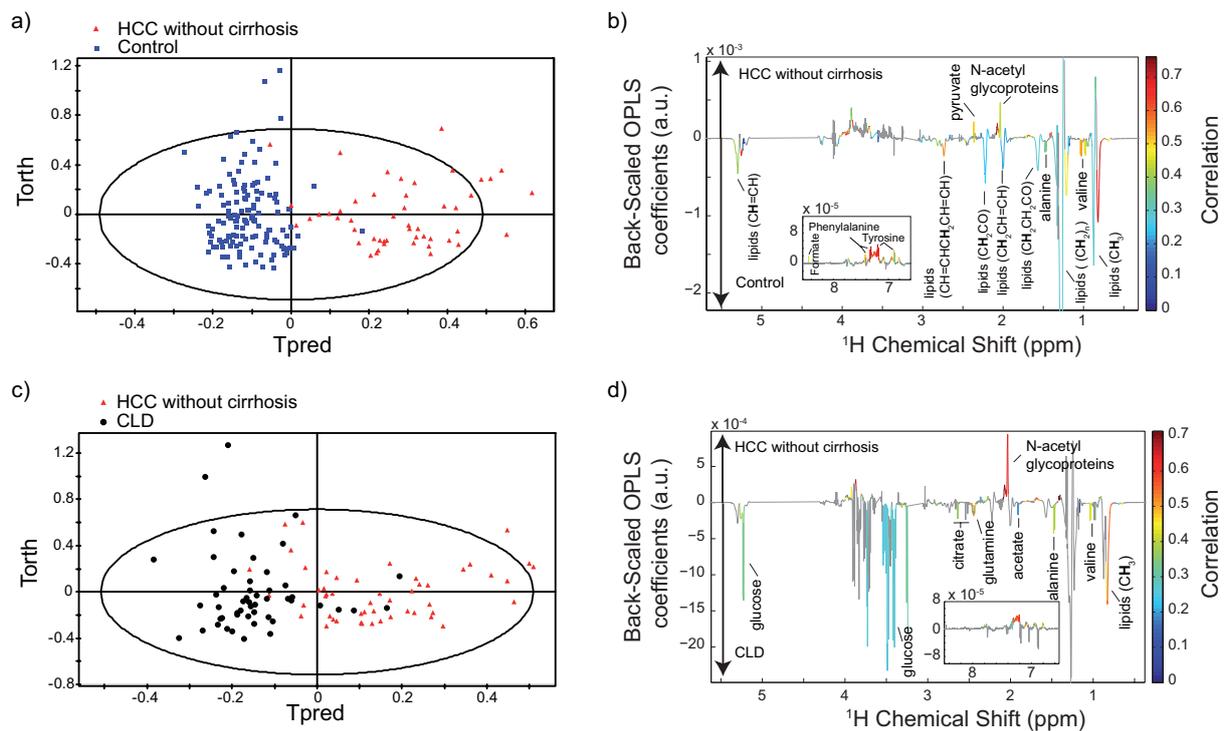


Figure 2.3.3. Discrimination between Control, CLD and HCC without cirrhosis. OPLS models discriminating a) HCC without cirrhosis and Control (3 orthogonal components, $R^2X = 0.723$, $R^2Y = 0.763$, $Q^2 = 0.714$), b) HCC without cirrhosis and CLD (2 orthogonal components, $R^2X = 0.691$, $R^2Y = 0.532$, $Q^2 = 0.457$).

Discrimination between Control, CLD and HCC with cirrhosis. We finally completed the study with the analysis of the HCC with cirrhosis group. These patients were significantly different from their Control counterparts (OPLS-DA, 2 orthogonal components, $R^2X = 0.645$, $R^2Y = 0.681$, $Q^2 = 0.609$, **Figure 2.3.4a**, **Figure 4.3.1f**, and **Table 2.3.1**), with higher levels of citrate, ethanol, formate, phenylalanine, pyruvate and tyrosine and lower levels of valine, unsaturated and oxidized lipids (**Figure 2.3.4b**). We compared the HCC with cirrhosis group from the CLD group. A small discrimination was observed, involving increased levels of N-acetylglycoproteins for HCC and changes in lipid profiles (OPLS-DA, 2 orthogonal components, $R^2X = 0.431$, $R^2Y = 0.365$, $Q^2 = 0.216$, **Figure 2.3.4c-d**, **Figure 4.3.1g**, and **Table 2.3.1**).

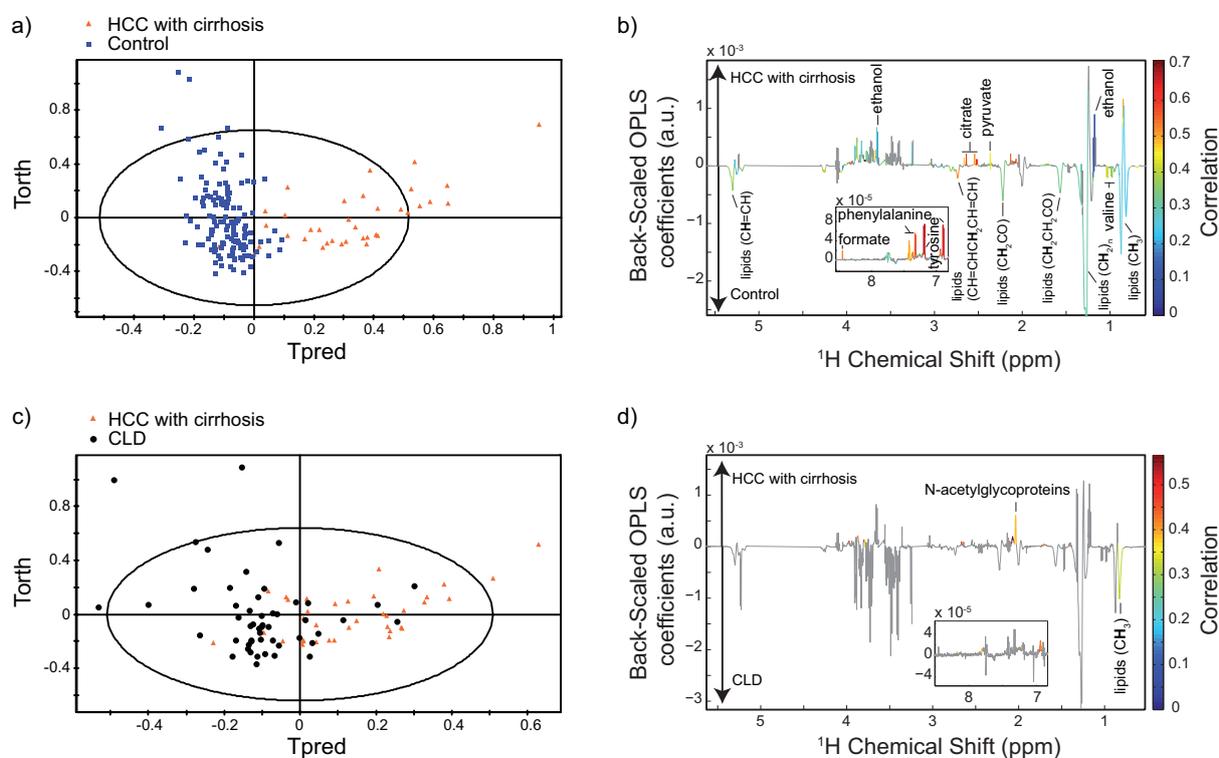


Figure 2.3.4. Discrimination between Control, CLD and HCC with cirrhosis. OPLS models discriminating a) HCC with cirrhosis and Control (2 orthogonal components, $R^2X = 0.645$, $R^2Y = 0.681$, $Q^2 = 0.609$), b) HCC with cirrhosis and CLD (2 orthogonal components, $R^2X = 0.431$, $R^2Y = 0.365$, $Q^2 = 0.216$).

Toward a plasma metabolic signature of CLD or HCC? These different analyses showed consistent results and allow us to distinguish metabolic signatures associated with CLD and HCC. High levels of N-acetylglycoproteins and low levels of alanine and valine characterized HCC. CLD samples displayed specifically high levels of acetate, citrate, glucose and ethanol. Furthermore, we identified a common metabolic signature for HCC and CLD, which could be

interpreted as a global marker of liver disease, involving an increase in formate, phenylalanine, pyruvate and tyrosine as well as changes in lipid metabolism.

We then investigated the influence of potential confounding factors on the plasma metabolic profiles in order to evaluate the reliability of these signatures and get insight into their biological meanings.

Table 2.3.1. Goodness-of-fit model parameters for OPLS models discriminating Control, CLD, HCC without cirrhosis, HCC with cirrhosis and total HCC.

| Model | Orthogonal component number | Sample number | R ² X | R ² Y | Q ² |
|----------------------------------|-----------------------------|---------------|------------------|------------------|----------------|
| Control vs HCC without Cirrhosis | 3 | 180 | 0.723 | 0.763 | 0.714 |
| Control vs CLD | 3 | 181 | 0.766 | 0.512 | 0.392 |
| Control vs HCC with Cirrhosis | 2 | 166 | 0.645 | 0.681 | 0.609 |
| CLD vs HCC without Cirrhosis | 2 | 101 | 0.691 | 0.532 | 0.457 |
| CLD vs HCC with Cirrhosis | 1 | 87 | 0.431 | 0.365 | 0.216 |
| Control vs total HCC | 5 | 216 | 0.769 | 0.809 | 0.732 |
| CLD vs total HCC | 1 | 137 | 0.444 | 0.372 | 0.305 |

Influence of gender, BMI, age, year of collection and albumin levels on plasma metabolic profiles. We investigated the influence of different parameters such as gender, body mass index (BMI), age and year of collection on the metabolic profiles. Gender was correlated with metabolic profiles in the Control group (OPLS-DA, 4 orthogonal components, R²X = 0.772, R²Y = 0.594, Q² = 0.316, **Figure 2.3.5a-b**, **Figure 4.3.2a**), involving changes in lipids and increase in branched-chain amino-acid concentrations (valine, leucine, isoleucine) for male individuals. To investigate the influence of BMI on the plasma metabolic profiles, we stratified the whole population in four groups: BMI < 20, 20 < BMI < 25, 25 < BMI < 30 and BMI > 30. We observed only a significant discrimination between the two extreme BMI groups (BMI < 20 and BMI > 30) associated with a global increase in lipids for patients with the highest BMI (OPLS-DA, 1 orthogonal components, R²X = 0.546, R²Y = 0.472, Q² = 0.273, **Figure 2.3.5c-d**, **Figure 4.3.2b**). Age was not associated with any metabolic signature, even when comparing extreme age classes. We compared samples obtained during the first year of collection, 2008, and the second year of collection, 2009. Samples collected during 2009 displayed higher lactate levels and lower glucose levels (OPLS-DA, 6 orthogonal components, R²X = 0.79, R²Y = 0.63, Q² = 0.473, **Figure 2.3.5e-f**, **Figure 4.3.2c**).

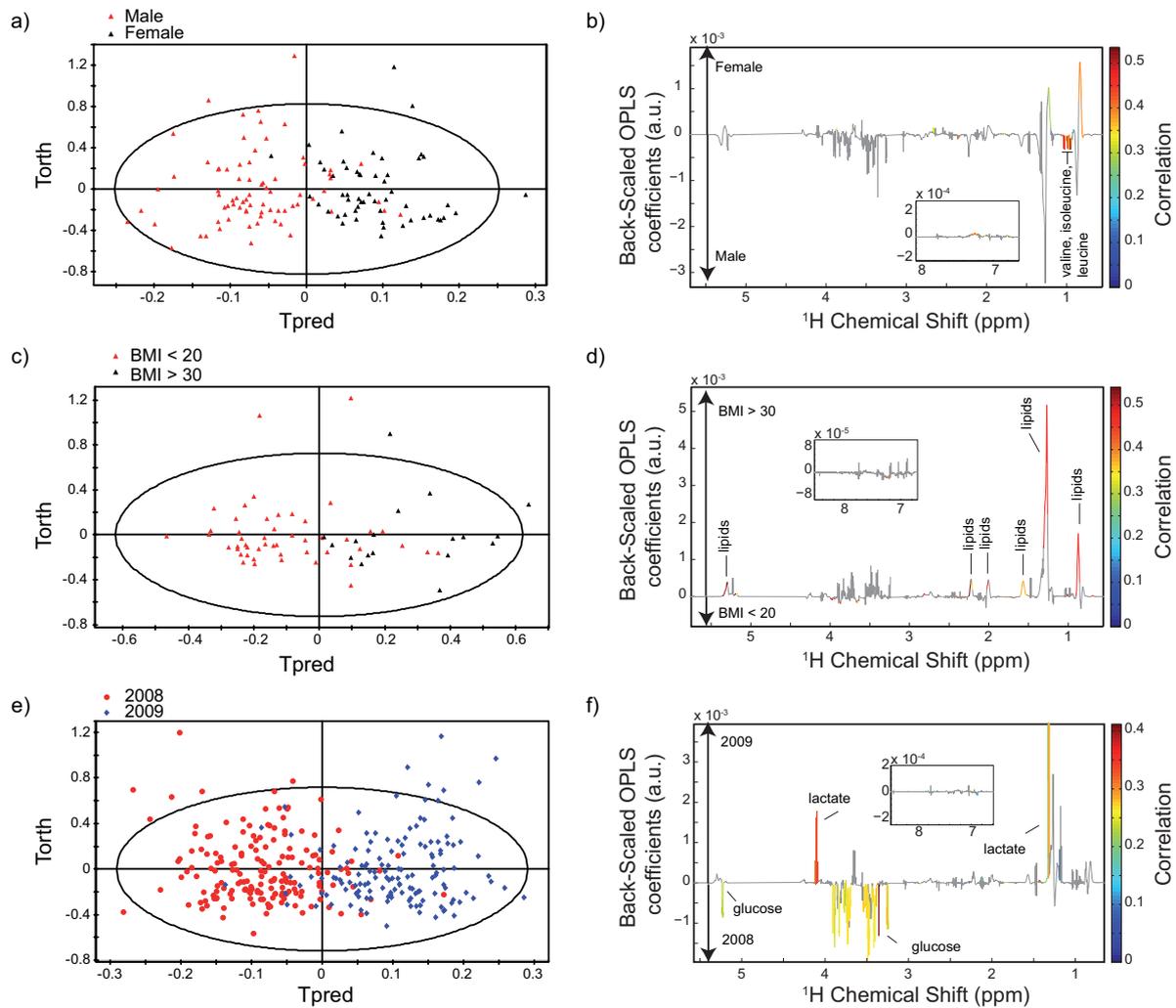


Figure 2.3.5. Effects of gender, BMI and year of collection on the metabolic signature. OPLS models discriminating a) gender (4 orthogonal components, $R^2X = 0.772$, $R^2Y = 0.594$, $Q^2 = 0.316$), b) BMI (1 orthogonal components, $R^2X = 0.546$, $R^2Y = 0.472$, $Q^2 = 0.273$) and years of collection. (6 orthogonal components, $R^2X = 0.79$, $R^2Y = 0.63$, $Q^2 = 0.473$).

Finally, albumin levels, potential markers of liver function for this population,^[164] considered as a continuous variable displayed a very strong correlation with the metabolic profiles (OPLS-DA, 5 orthogonal components, $R^2X = 0.797$, $R^2Y = 0.874$, $Q^2 = 0.786$, **Figure 2.3.6**, **Figure 4.3.2d**) where high levels of citrate, formate, glucose, phenylalanine, pyruvate and tyrosine and low levels of valine and isoleucine were associated with low albumin levels.

Evaluating disease metabolic signature reliability. We then re-evaluated the metabolic signatures observed for HCC or CLD groups in the light of these results, by matching groups on the different parameters to cancel out their effects for sample class discriminations.

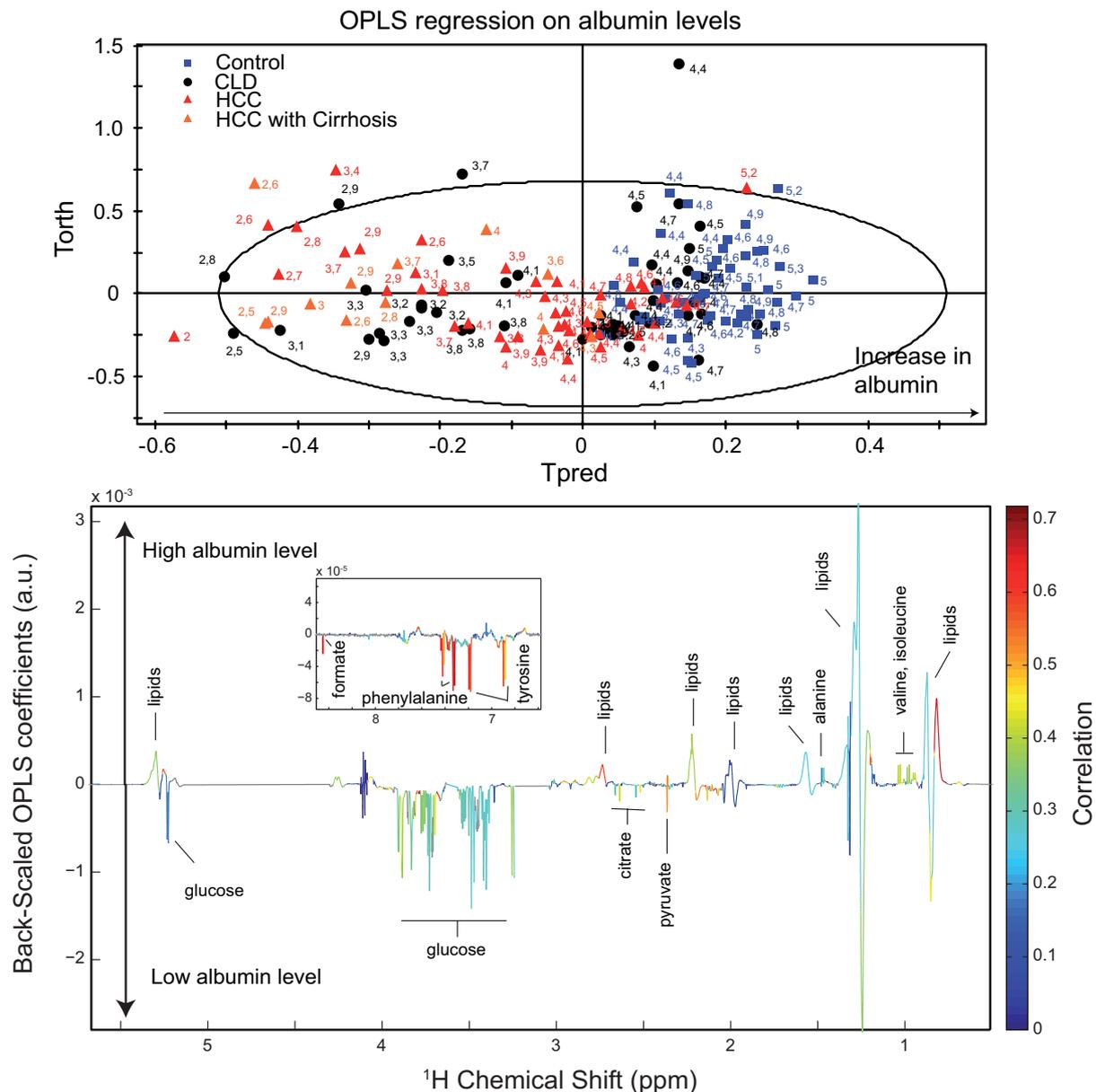


Figure 2.3.6. Correlation between albumin levels and metabolic profiles (5 orthogonal components, $R^2X = 0.797$, $R^2Y = 0.874$, $Q^2 = 0.786$).

After matching on gender, BMI, year of collection and albumin levels (keeping only high albumin levels (albumin (alb) > 40 g/L) corresponding to a normal liver function), we were still able to discriminate Control from HCC without cirrhosis individuals (OPLS-DA, 2 orthogonal components, $R^2X = 0.67$, $R^2Y = 0.847$, $Q^2 = 0.712$, **Figure 2.3.7a-b**, **Figure 4.3.2e**), involving an increase in phenylalanine, pyruvate and tyrosine for HCC individuals and global changes in lipid profiles. Only a few metabolites were involved in this discrimination but note that the statistical power was quite low due to the low number of samples kept for this analysis.

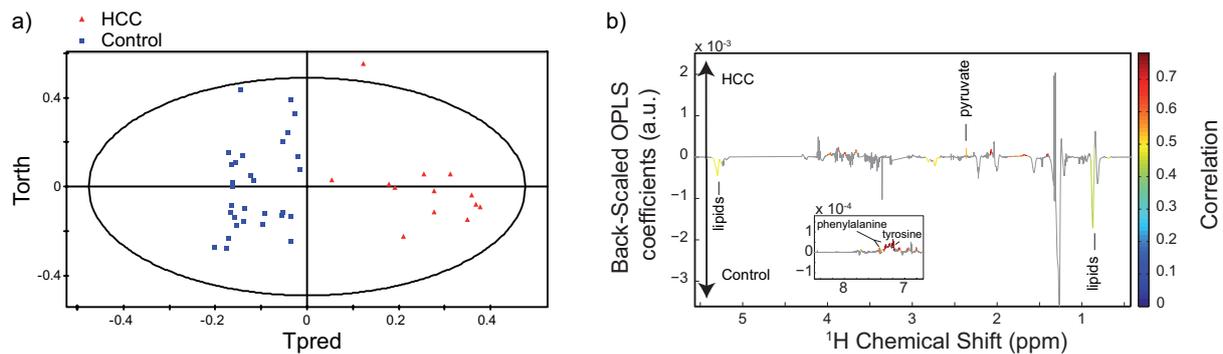


Figure 2.3.7. Discrimination between Control and HCC without cirrhosis individuals, after matching on gender, BMI, year of collection and liver function (albumin > 40g/L): OPLS-DA model (2 orthogonal components, $R^2X = 0.67$, $R^2Y = 0.847$, $Q^2 = 0.712$) a) score plot and b) loading plot.

We evaluated the effects of albumin level matching in the discrimination between Control and CLD individuals. We first compared Control and CLD samples with high albumin level, determined by standard clinical tests, (albumin > 40 g/L). We observed a very weak discrimination between the 2 groups after this step ($Q^2 = 0.261$), just above the significance level. We thus further matched the samples on the BMI or gender or year of collection and combinations of these parameters. In each case, we did not observe any discrimination between Control and CLD, probably due to the decrease in statistical power with matching steps. Only two Control samples had low albumin levels, making the comparison between Control and CLD for low albumin levels meaningless. We finally compared Control samples with high albumin levels (alb > 40 g/L) and CLD with low albumin levels (alb < 40 g/L). We observed a strong discrimination between the 2 groups ($Q^2 = 0.892$), still robust after further matching steps on BMI, gender or year of collection. Overall this analysis showed that there were 2 groups in the CLD samples, low and high albumin levels, displaying different behaviors. Those with low albumins levels were clearly different from Control samples whereas those with high albumin level were very similar to Control samples.

We were not able to perform a matching on albumin levels for the comparison of HCC with cirrhosis and Control groups as the former had only 2 samples with high albumin level (alb > 40 g/L) whereas the latter was associated with high albumin levels. After matching on albumin levels (keeping only samples with low albumin levels), we still observed discrimination between CLD and HCC with cirrhosis ($Q^2 = 0.538$).

Correlation with HBV and HCV status, LTBP2 levels and R249S copy numbers. We compared the metabolic profiles of individuals with positive and negative HBsAg, a marker of HBV infection. We did not observe any differences between positive and negative HBsAg

individuals, considering the whole dataset or each subgroup (Control, HCC without cirrhosis, HCC with cirrhosis, CLD or CHCA) individually.

We then compared the metabolic profiles of individuals with positive and negative HCVAg, a marker of HCV infection. We did not observe any differences between positive and negative HCVAg, considering the following subgroups: HCC without cirrhosis, HCC with cirrhosis, CLD and CHCA. This analysis was not performed in the control group, as this group only gathered HCVAg negative individuals.

LTBP2 protein was identified as a potential marker of HCC by comparison to healthy subjects. To further investigate whether LTBP2 levels could be linked to metabolic changes we correlated metabolic profiles and LTBP2 considered as a continuous variable. We did not observe any association in the subgroups individually, whereas we found a small association considering healthy and HCC samples as expected.

TP53 R249S mutation copy numbers, determined by circulating-free DNA are associated with HCC cases developed on HBV infection.

We considered either R249S copy numbers as a continuous variable or a discrete variable (values 1 and 2, defined with the threshold of 150 copies). In both cases, we did not observe any significant association between the metabolic profiles and the R249S copy numbers considering the HCC different subgroups individually. Note that in the HCC with cirrhosis group, we obtained a strong but not significant trend between individuals with low (< 150) and high (> 150) R249S copy numbers.

2.3.4 Discussion

We studied by 600 MHz ¹H NMR spectroscopy the influence of different clinical and biological parameters on plasma metabolic profiles and derived potential markers for patients suffering from hepatocarcinoma or chronic liver disease. We evaluated correlations between the plasma metabolic data and a proteomic marker of HCC to further build a multi-marker approach.

Influence of clinical and biological parameters on the plasma metabolic profiles. In this study gender was associated with branched-chain amino acid levels, through higher concentrations among males. Sexual dimorphism in serum metabolite concentrations, including similar variations in branched-chain amino acids, has already been observed with mass spectrometry in a large epidemiological study and correlated with genetic variants.^[165]

Global lipid levels discriminated individuals with low and high body mass indexes (BMI < 20

against BMI > 30), as already shown through characterization of effects of obesity on metabolism.^[166] Plasma metabolic profiles did not correlate with age in this study, although a recent study demonstrated an age dependence for human serum metabolic profiles.^[167] However, this last study was performed on 1038 females and 1124 males, thus with a larger statistical power allowing the detection of subtle variations.

Samples collected in 2008 and 2009 were discriminated according to glucose and lactate concentrations. These changes could suggest uncontrolled modification in sampling procedure. Changes in time between sample collection and first centrifugation could be involved as this parameter is known to correlate with plasma glucose decrease and lactate increase due to anaerobic metabolism of blood cells.^[168]

The strongest correlation in this study was between albumin levels, determined with standard clinical tests and plasma metabolic profiles. This analysis was performed on CMPG NMR spectra, where protein signals in the background, which include a strong albumin contribution, are removed through NMR signal editing. The metabolic signature associated with this correlation involved a wide-range of metabolites instead of changes in the background, confirming the removal of albumin signals. Decrease in albumin levels can be observed in different pathological conditions, such as liver failure, undernutrition or nephrotic syndrome for instance. In this context, albumin levels were a good marker of liver function. Control sample distribution over albumin level was homogenous with high albumin concentration (alb > 40g/L), whereas pathological samples displayed an important spread for this parameter, reflecting different stages of liver dysfunction. This analysis thus suggested that chronic alteration of liver function induced important plasma metabolic changes. Similar results have been observed among patients suffering from cirrhosis of different gravity. Most of the studies referred to impairment of liver function but did not try to evaluate the contribution of this parameter in the metabolic signatures associated with pathological states.^[138, 148-151]

Plasma metabolic signatures of hepatocellular carcinoma and chronic liver disease. Our analysis derived specific metabolic signatures for HCC involving high N-acetylglycoproteins and low levels of alanine and valine and for CLD with high concentrations of acetate, citrate, glucose and ethanol. These both groups displayed an increase in formate, phenylalanine, pyruvate and tyrosine by comparison to healthy subjects. Of particular clinical interest, we observed a discrimination between HCC with cirrhosis and CLD mainly based on increase in

N-acetylglycoprotein levels and changes in lipid profiles. N-acetylglycoproteins have been largely associated to human cancer.^[147]

Gao *et al.* investigated in a pilot study differences between healthy individuals and patients suffering from HCC or liver cirrhosis by NMR spectroscopy and observed metabolic signatures sharing some similarities with ours.^[138] High levels of phenylalanine, tyrosine, pyruvate and N-acetylglycoteins and low levels of unsaturated lipids and valine discriminated Control from HCC as in our study. Liver cirrhosis, as CLD here, was characterized by an increase in phenylalanine, tyrosine, pyruvate and acetate and decrease in unsaturated lipids. N-acetylglycoprotein levels, however, were lower for the liver cirrhosis group. Different studies observed metabolic changes between similar groups by mass spectrometry. Most of the metabolites involved in discriminations were not detectable by NMR spectroscopy, highlighting technique complementarities.^[148, 149, 151] Interestingly, Chen *et al.* observed high pyruvate levels in HCC patients by comparison to healthy subjects but low levels of phenylalanine contrary to Wang *et al.*, Zhou *et al.* and our study.

We highlighted the influence of the liver function on plasma metabolic profiles through evaluation of the correlation between albumin levels and metabolite concentrations. However, we still observed a discrimination between healthy subjects and patients with HCC without cirrhosis and normal liver function, suggesting metabolic differences between these two groups uncorrelated with liver function. On the contrary, for the CLD groups, individuals with normal liver function were poorly discriminated from healthy subjects, suggesting that metabolic changes in the CLD group were mainly due to liver dysfunction.

Influence of HBV or HCV infection. In this study, we did not monitor any effects of HBV or HCV infection on metabolic profiles. Zhou *et al.* did not observe also by UPLC-MS differences between patients with HCC infected by HBV or HCV.^[151] Furthermore, Shariff *et al.* detected similar changes in urine for patients with HCC infected by HCV or HBV by comparison to healthy subjects.^[153, 169] These different studies consistently suggest that HBV or HCV infection do not strongly influence biofluid metabolite composition.

Correlation with LTPB2 protein levels and TP53 R249S copy numbers. The LTBP2 protein was identified in a proteomic analysis, within the same sample set as a potential marker of HCC. We observed a correlation between LTBP2 concentrations and metabolic profiles considering healthy and HCC subjects, which was expected as both correlated with the HCC status. We did not detect any correlation considering each subgroup separately.

TP53 R249S copy numbers are increased in HCC with HBV infection. However, this parameter did not correlate with metabolic profiles in the HCC groups, which was expected, as we did not observe any correlation between HBV infection and metabolic profiles.

It thus would be interesting to integrate in a model these different parameters in order to improve classification between control and HCC and the different types of HCC case in a global multimarker approach.

2.3.5 Conclusion

In conclusion, our investigation has identified plasma metabolic signatures for hepatocellular carcinoma and chronic liver disease by ¹H NMR spectroscopy. Our collaborators are currently complementing this analysis with mass spectrometry measurements to crosscheck these results and extend metabolome coverage. These findings are promising as they potentially pave the way towards a non-invasive diagnostic tool for HCC. A global multimarker model will be built from genetics, proteomics and metabolomics data and should be validated with an independent patient cohort.

2.4 Conclusion

In this second part, we have presented different applications of metabolomics focusing on biomarker discovery and delineating metabolic signature of potential interest for physicians.

However, different issues have to be addressed, before planning a routine use of these results in hospitals. First of all, these metabolic signatures were derived from a relatively small number of patients coming from one geographical area. Validation of these results is mandatory on a larger number of samples collected in different places.

Furthermore, the question of the specificity of these metabolic signatures can be partially assessed with these examples. For instance, N-acetylglycoprotein levels correlate with both metastatic status of breast cancer and hepatocarcinomas, showing that this marker cannot be associated with a unique biological state. Such biological tests should thus be limited to well-defined clinical situations in order to reach enough sensitivity and specificity to be useful.

Extensive characterization of metabolic physiological changes in normal populations as well as effects of pathological states such as kidney or liver failures could help to delineate more precisely relevant clinical situations. This knowledge could also help to define more robust sampling protocols aiming at reducing inter-individual variations.

To fully validate a biomarker, it is necessary to understand the underlying biological mechanism responsible for its properties. The combination of different types of analyses, from biofluids to tissues or model animals is often required to achieve this goal.

Part 3: Probing model organisms metabolism

As widely used for the study of human populations in the framework of biomarker discovery, metabolomics has also been developed for the characterization of a wide range of model organisms, from yeasts to mammals, in order to address biological questions. In this third part, we shift our object of study from humans to model organisms, from epidemiological and clinical issues to biological ones. Describing the interest of model organisms for biology is beyond the scope of this introduction and we highlight here the specificities of model organisms metabolomic studies in terms of experimental design, analytical and statistical issues.

First of all, metabolomics is a new read-out for biological studies in monitoring metabolic variations in response to a wide-range of perturbations. This approach can thus complement any experimental schemes. These studies are not restricted to biomarker discovery as they can get further insight into molecular mechanisms. Contrary to investigations in human populations, essentially observational, studies on model organisms are experimental, through control of parameters, design and tools to test hypotheses. The perturbation under study is induced by the experiment and its effect can be isolated and recorded.

From a technological point of view, we can distinguish two situations in the study of model organisms according to their size. For “macroscopic” animals, such as mammals, the metabolomic approach is similar to human studies. Each animal is analyzed individually and many samples can be collected, from biofluids to tissues, opening different windows on global metabolic homeostasis. On the contrary, for “microscopic” or small organisms, such as bacteria, cells or *Caenorhabditis elegans* worms, we monitor global whole-organism metabolism without details about compartmentalization, in samples gathering from thousand individuals for worms to millions for cells.

Demonstrating statistically significant changes in model organisms requires less samples than for clinical or epidemiological studies, as metabolic perturbations induced in these systems are often larger and inter-individual variations lower through adequate control of experimental factors. Furthermore, in the case of whole-organism metabolomic studies, where each sample gathers thousand to million organisms, we monitor average behaviors without getting access to inter-individual variations. Metabolic variable distributions are thus tighter, reinforcing differences between conditions under study.

In this third part, we present four metabolomic analyses performed from rats to cells, which illustrate the different aspects relative to model organism studies.

We first investigate physiological metabolic differences in plasma and urine between four rat strains used as controls in genetic studies. Indeed, as for epidemiological or clinical studies, the choice of appropriate controls is a major issue in metabolomics and thus requires their extensive characterization. This descriptive analysis relies on liquid NMR spectroscopy of biofluids, as introduces in the previous part about human population studies.

The three following studies focus on two small model organisms: the worm *Caenorhabditis elegans* (*C. elegans*) and the HepG2 cell line, model of hepatocarcinoma and liver cells, using whole-organism HR-MAS NMR spectroscopy and aiming at characterizing biological processes, gene function or protein interaction at the metabolic level.

We investigate the effects of aging and dietary restriction on *C. elegans* metabolism by characterizing wild-type worms and different *C. elegans* genetic mutants, displaying different profiles of aging.

In the framework of functional genomics, we then delineate the impact on *C. elegans* metabolism of the protein receptor AHR-1 loss-of-function, ortholog of the Aryl Hydrocarbon Receptor (AHR), involved in dioxin toxicity in mammals.

Finally, we investigate the effects of the interaction between an endogenous transcription factor E4F1 and a viral protein HBx in a model of liver cells, HepG2, infected by hepatitis B virus (HBV).

These different projects were built as collaborations with different biological groups, experts in their area, which prepared all the samples and helped us to interpret the data.

3.1 Natural metabotype variations in healthy control inbred rat strains

3.1.1 Introduction

Maintenance of homeostatic biological functions in mammalian species is controlled by genetic and environmental factors that influence individual's risks to develop increasingly prevalent pathological elements of the cardiometabolic syndrome (type 2 diabetes mellitus, hypertension, dyslipidemia and obesity).^[170] Even though knowledge of their genetic basis keeps progressing with genome-wide association studies,^[171] the multiple interacting mechanisms at play involving numerous tissues still hamper investigations into the cause of these diseases. Among models that can assist the elucidation of disease aetiology, the rat has attracted the attention of geneticists interested in dissecting out and mapping complex traits in experimental cohorts (backcross, F2 cross, recombinant inbred, heterogeneous stock).^[172, 173]

Identifying a strain that exhibits normal or disease resistant phenotypes is a crucial step in the design of genetic studies in animal models to derive an optimal experimental cohort of hybrids between affected and control strains that will be used for genotype and phenotype analyses, and maximise quantitative trait locus (QTL) detection. The mapping of intermediate phenotypes in rodents can be extended to molecular phenotypes derived by functional genomic technologies, which provide quantitative information on gene expression regulation.^[174-176] In particular, metabolomics,^[5, 177] a hypothesis-free metabolic systems biology approach based on ¹H NMR spectroscopy and mass spectrometry, provides a powerful high-throughput molecular phenotyping system for acquiring repeated high-density multivariate phenotypic signatures of biological samples and for linking variations in metabolic abundance and genetic polymorphisms.^[178, 179]

Physiological and metabolomic studies in inbred control mouse strains have demonstrated that important interstrain genetic differences result in broad variability of biological functions that can nevertheless maintain metabolic and hormonal phenotypes within the normal range.^[180] These data contradict claims based on investigations in Sprague-Dawley *outbred* colonies, that all control laboratory rodents fed *ad libitum* display pathophysiological patterns associated to cancer, inflammation and metabolic syndrome.^[181] This view is challenged by studies in inbred mice and rats showing that, even though natural phenotype variability exists,^[182] metabolic and body weight variables in animals fed *ad*

libitum remain within a normal range.^[183, 184] In support of this idea, metabolic variability in biofluids is used to define metabolic entropy,^[185] for instance during a toxicological insult. Also, the establishment of a theory of biological robustness explains biological variability through essential systems robustness, *i.e.*, pathway redundancy, adaptation, parameter insensitivity.^[186]

In the present work, we investigate the quantitative metabolic phenotypes relevant to the cardiometabolic syndrome in genetically distinct inbred rat strains predominantly used as healthy control models in experimental cohorts designed to map the genetic basis of complex phenotypes. Analysis of adiposity and glucose and lipid homeostasis combined with NMR-based plasma and urine metabolomic profiles^[5, 177] highlighted unexpectedly high, albeit non-pathological, natural metabolomic variation between control strains, even when genetic polymorphism is reduced. These results underlie the involvement of distinct regulatory mechanisms of key biological functions, including transgenomic influences, which collectively contribute to maintain homeostasis within the normal, non-pathological range.

3.1.2 Methods

Animals (Jane F. Fearnside, University of Oxford). Male rats of four inbred rat strains (Fisher F344/NHsd [F344], n=5, Lewis Lew/SsNHsd [Lew], n = 5, Wistar Kyoto WKY/NHsd [WKY], n = 5, Brown Norway BN/Ox [BN], strain = 7-68) were used for this study. They were bred locally (BN) or purchased from a commercial supplier (F344, Lew, WKY) (Charles River Laboratories, Margate, Kent, UK). A group of aged matched Goto-Kakizaki (GK, n = 12) rats bred locally was also used for the preliminary study. All rats were housed in groups of 5, fed with standard laboratory chow pellets (ERB, Whitam, UK) and water *ad libitum*, and kept on 12 hours light/dark cycle. All experiments were performed in 13±1 week old rats. All experiments were carried out in accordance with UK national (Home Office) and institutional guidelines.

A summary of the animal experimental procedures, physiological and analytical assays is given in **Figure 4.4.1**.

Physiological procedures and sample collection (Jane F. Fearnside, University of Oxford). Intraperitoneal glucose tolerance tests (IPGTT) were carried out after 4 hours of fasting (from 9am until 1pm). Body weight and body mass index (BMI) were determined. Rats were anesthetized with ketamine hydrochloride (Ketalar, Parke-Davies, UK) (95 mg/kg body weight). After baseline blood sample had been taken from the tail vein, a solution of

glucose was injected intraperitoneally (2g/kg body weight). Subsequent blood samples were collected from the tail vein at 15, 30, 75 and 120 minutes post glucose injection to determine blood glucose concentration. The overall glucose tolerance of the animals was assessed by both the cumulative glycaemia (the total increment of plasma glucose during the IPGTT) and the ΔG (defined as the total increment of plasma glucose during the test over baseline). The K parameter, which measures glucose clearance rate following a glucose challenge, was calculated as the slope from the maximum glucose response 15 mins after glucose injection until the end of the IPGTT.

To avoid possible metabolic effects of anesthesia and glucose challenge, rats were individually housed one week later in metabolic cages for one overnight for collection of urine and plasma samples for lipid assays and metabonomic profiling. Food consumption values were recorded. The rats were then fasted overnight in the metabolic cages to minimize metabolic effects of interindividual variations of food intake. The following morning (9-10am) a second urine sample was obtained and a blood sample collected from the tail vein. Blood was centrifuged and plasma was taken and stored at -80 until lipid assays and $^1\text{H-NMR}$ metabonomic analyses.

One week later, rats were killed by CO_2 asphyxiation following an overnight fast. Retroperitoneal fat pad (RFP) was collected and weighed. Adiposity index (AI) was determined as the ratio between RFP weight and body weight.

Analytical assays (Jane F. Fearnside, University of Oxford). Blood glucose concentration was determined with a glucose meter (Accucheck, Roche Diagnostics). Total cholesterol (TC), cholesterol contained in high-density lipoproteins (HDL-C), in low-density lipoproteins (LDL-C) and triglycerides (TG) were determined using diagnostic enzymatic/colorimetric kits (ABX, Shefford, UK) on a Cobas Mira Plus automatic analyser (ABX, Shefford, UK).

Metabolic Profiling by $^1\text{H-NMR}$ Spectroscopy. Urine samples ($n = 20$, 5 per strain) were prepared according standard protocols^[17] by mixing 400 μL of urine with 200 μL of a phosphate buffer ($\text{pH} = 7.4$; 20% $\text{D}_2\text{O}/\text{H}_2\text{O}$ v/v) containing 1 $\text{mmol}\cdot\text{L}^{-1}$ of 3-(trimethylsilyl)propionic-2,2,3,3- d_4 acid (TSP). Plasma samples ($n = 16$, 4 per strain) were prepared^[17] by mixing 300 μL of plasma with 150 μL of a 0.9 $\text{g}\cdot\text{L}^{-1}$ saline solution (20% $\text{D}_2\text{O}/\text{H}_2\text{O}$ v/v). $^1\text{H-NMR}$ spectra were acquired on Bruker Avance spectrometers operating at 600 MHz and 700 MHz, using standard 5 mm TXI probes at the Rhône-Alpes Large Scale Facility for NMR and Imperial College London. Sample temperature was set to 300 K and controlled throughout the experiments, using low-power water presaturation pulse sequence

for water signal suppression during the relaxation time of 2 s and the mixing time of 10 ms. For each experiment, 128 transient free induction decays (FID) were collected with 49036 data points and a spectral width of 20 ppm. All FID were multiplied by an exponential function, equivalent to a 0.3 Hz line-broadening factor before Fourier transformation. Spectra were phased and ^1H chemical shifts were referenced to the TSP singlet ($\delta 0$) in urine and to the α -glucose doublet ($\delta 5.23$) in plasma. Each spectrum was reduced in 11k bins of 0.001 ppm width over the chemical shift range of -1 to 10 ppm using AMIX (Bruker). For urine samples, residual water signal ($\delta 4.60$ - 5.00) and urea signal ($\delta 5.30$ - 6.15) were excluded to remove effects of variation in water suppression. For plasma samples, residual water signal ($\delta 4.66$ - 5.00) was discarded. Spectra were scaled to the sum of intensities and mean-centered or scaled to unit variance (UV) prior to analysis.

Statistical analysis of phenotypic data. All phenotypic data collected from the inbred rat strains were analysed using SPSS version 14.0 statistical package. Covariates that account for variance that is not due to the dependent variable were selected using the univariate General Linear Model (GLM). Fisher's LSD and Tamhane's T2 post hoc tests according to Levene's test for equality of variance were carried out to identify any significant differences between strains for each of the phenotypes analysed. Principal Component Analysis (PCA)^[187] was performed on the UV-scaled physiological dataset and results were visualized as a biplot.^[188]

Multivariate analysis of metabolic profiles. OPLS analyses were performed to compare the different groups. The model coefficients were back-scaled according to Cloarec *et al.* to enhance interpretability.^[92] Model validations were performed by resampling the model 999 times under the null hypothesis, i.e. by generating models by random permutation of Y matrix, not related to the X matrix anymore. The analysis was performed in SIMCA P12 and in house Matlab script.

Hierarchical clustering analysis. Hierarchical clustering was used to derive a phylo-genetic tree from single nucleotide polymorphism (SNP) data (<http://gscan.well.ox.ac.uk/gsBleadingEdge/rat.snp.selector.cgi>) and phylo-metabonomic trees from plasma and urine metabolic profiles. Similarity between profiles was measured from Euclidian distance. Dendrograms were built from maximum similarity between classes. SNP data were coded as Strain Distribution Pattern (SDP) prior to clustering. Plasma and urine ^1H -NMR spectra were averaged for each strain to obtain plasma and urine phylometabonomic trees respectively. The pairwise comparison of dendrogram topology was assessed by computing the Pearson correlation coefficient between their cophenetic distances.^[189]

Hierarchical clustering and cophenetic analysis was carried out using R software (<http://www.R-project.org>).

Metabolite-strain correlation networks. The metabolite-strain correlation network was implemented as a bipartite graph, $G_{MS}=(M,S,C)$ where M is the set of nodes corresponding to discriminant metabolites, S the set of nodes corresponding to strains and C the set of edges corresponding to correlation between metabolites and strains as obtained from OPLS loadings analysis. The Graph Exploration System (GUESS) software (<http://graphexploration.cond.org>) was used to visualize graphically the metabolite-strain correlation network.

Metabolite-Set Enrichment Analysis (MSEA) (in collaboration with Vincent Navratil, University of Lyon). Metabolite-Set Enrichment Analysis (MSEA)^[106] is an extension of Gene-Set Enrichment Analysis (GSEA)^[190, 191] approach to test metabolic pathways enrichment from metabolomic data. The aim is to assess whether strain-associated metabolites correspond preferentially to particular metabolic pathways by comparing the proportions of pathway-associated metabolites in strain-associated metabolites and their respective proportions in global metabolism (**Figure 4.4.2**). The KEGG database (<http://genome.jp/kegg>) was first used to annotate discriminant strain-associated metabolites onto metabolic pathways using KEGG compound identification numbers.

The comparison between the list of metabolites from KEGG pathways and the list of significantly affected metabolites leads to the construction of a series of 2x2 contingency tables, for each strain and each pathway. For a given pathway α , significant metabolites belonging to pathway α are accounted for as “observed” (n11) whereas the other non-significant metabolites from pathway α are classified as “unobserved” (n21). n12 corresponds to the other significant metabolites that do not belong to pathway α according to KEGG pathway annotations, and finally, n22 corresponds to all the other metabolites listed in KEGG that are not significant and do not belong to pathway α . Lipids and glycoproteins moieties were not considered in this analysis, as they did not unequivocally match specific metabolites and therefore unique KEGG entries. To test for a significant overrepresentation in metabolites from pathway α , an exact Fisher test is then computed on this contingency table (**Figure 4.4.2**).

To control the false discovery rate associated to multiple testing, the exact Fisher test P-value was finally adjusted using Benjamini and Hochberg procedure. A pathway-strain association network was finally represented as a bipartite graph, $G_{PS}=(P,S,E)$ where P is the

set of nodes corresponding to enriched KEGG pathways with adjusted P-Values ≤ 0.05), S the set of nodes corresponding to strains and E the set of edges corresponding enriched pathways. The Graph Exploration System (GUESS) software (<http://graphexploration.cond.org>) was used to visualize graphically the pathway-strain correlation network.

3.1.3 Results

Metabolic variability in disease and control rat strains. A crucial step in the experimental design of control/case and subsequent QTL studies lies in the choice of a healthy control strain bred to the disease strain to derive the cohort of hybrids. To test the power of metabolomics to separate a disease model from controls, we carried out a preliminary study designed to test the segregation of plasma metabonomic data of the Goto-Kakizaki (GK) strain, a model the cardiometabolic syndrome, and normoglycemic (BN, WKY) rats, using an OPLS-DA model. The distribution of the three strains along the OPLS score clearly shows that the WKY strain is closer to GK than BN (**Figure 4.4.3**), indicating the strong metabolic impact of genetic divergence between BN and Wistar-derived strains (WKY, GK). In fact, the WKY strain appears at the centre of the OPLS score, whereas BN and GK strains lie at either ends of the distribution. This result shows that although BN and WKY are both considered as healthy negative control strains, the BN strain is more metabolically distant from GK strain, and therefore a better control to map the genetic control of metabolic traits. This result is also suggestive of an unexpected level of metabonomic variability in inbred rat strains, otherwise used indifferently as neutral controls. To investigate whether this natural metabolomic variation in control strains is metabolically morbid, we characterized the genomic, physiological and metabolomic variability of several standard control rat strains (Fisher – F344, Lew, BN, WKY) used in cardiometabolic syndrome research. To characterize natural physiological and metabolic variation in healthy control strains, we applied a series of physiological procedures and metabolic assays outlined in **Figure 4.4.1**.

Definition of strain-specific physiological plasticity in inbred rats. To identify strain-specific phenotypic patterns, we performed a principal component analysis (PCA) model on average physiological data, visualized as a Gabriel's biplot (**Figure 3.1.1a-b**). This representation illustrates the clear separation of the four control strains and the physiological variables that contribute to strain divergence (**Figure 3.1.1a**). Interestingly, variability between control strains remains important when physiological data of the spontaneously

diabetic GK strain were included in the model (**Figure 3.1.1b**). The physiological variation between BN and GK strains was of the same order of magnitude as between BN and WKY.

When compared to the four control strains, GK rats showed marked increased adiposity (**Table 4.4.1**), hyperglycaemia (**Figure 3.1.1c**) and glucose intolerance (**Figure 3.1.1c-d**), which is the single pathophysiological selection criterion used to derive this strain over many successive generations of breeding outbred Wistar rats. Impaired glucose homeostasis, altered lipid metabolism (reduced plasma concentrations of total, HDL and LDL cholesterol and elevated triglycerides) (**Figure 3.1.1e**), and systematically elevated body weight, BMI, adipose tissue weight and adiposity index (**Table 4.4.1**) in GK were significant when compared to WKY, which also derives from an outbred Wistar stock and is therefore the genetically closest control to GK. This pattern of pathological features provides confirmatory evidence that the GK is a model of the cardiometabolic syndrome and that the other four strains used in the study exhibit relative normal relevant phenotypic patterns.

All four control inbred rat strains exhibited close to normoglycemia (5.5mM). Consistent with data shown in the Gabriel's biplot (**Figure 3.1.1a-b**), WKY show relative basal hyperglycemia, reduced glucose tolerance as indicated by increased glycemic response to glucose during the intra peritoneal glucose tolerance test (IPGTT, see Methods) and elevated cumulative glycemia and ΔG (**Figure 3.1.1c-d**), increased plasma levels of total and HDL cholesterol (**Figure 3.1.1e**), elevated body weight (**Table 4.4.1**), increased retroperitoneal fat pad weight and adiposity index (**Table 4.4.1**) when compared to BN, F344 and Lew rats. Rats of the BN strain are characterized by enhanced glucose tolerance and high plasma level of LDL cholesterol (**Figure 3.1.1c-e**). The BN strain showed significantly lower body weight, body mass index, RFP weight and adiposity index than Lew and WKY (**Table S1**). F344 rats showed specifically lower plasma LDL cholesterol and higher triglyceride levels than the other three strains (**Figure 3.1.1e**). BN and F344 shared identical low plasma concentrations of total and HDL cholesterol. Lew rats generally showed intermediate phenotype values. Altogether, these physiological results show that inbred control rat strains fed *ad libitum* maintain non-pathological body weight and glucose and lipid regulations when compared to outbred rat strains^[181] and rat models of diabetes (GK) and obesity (Zucker *falfa*).

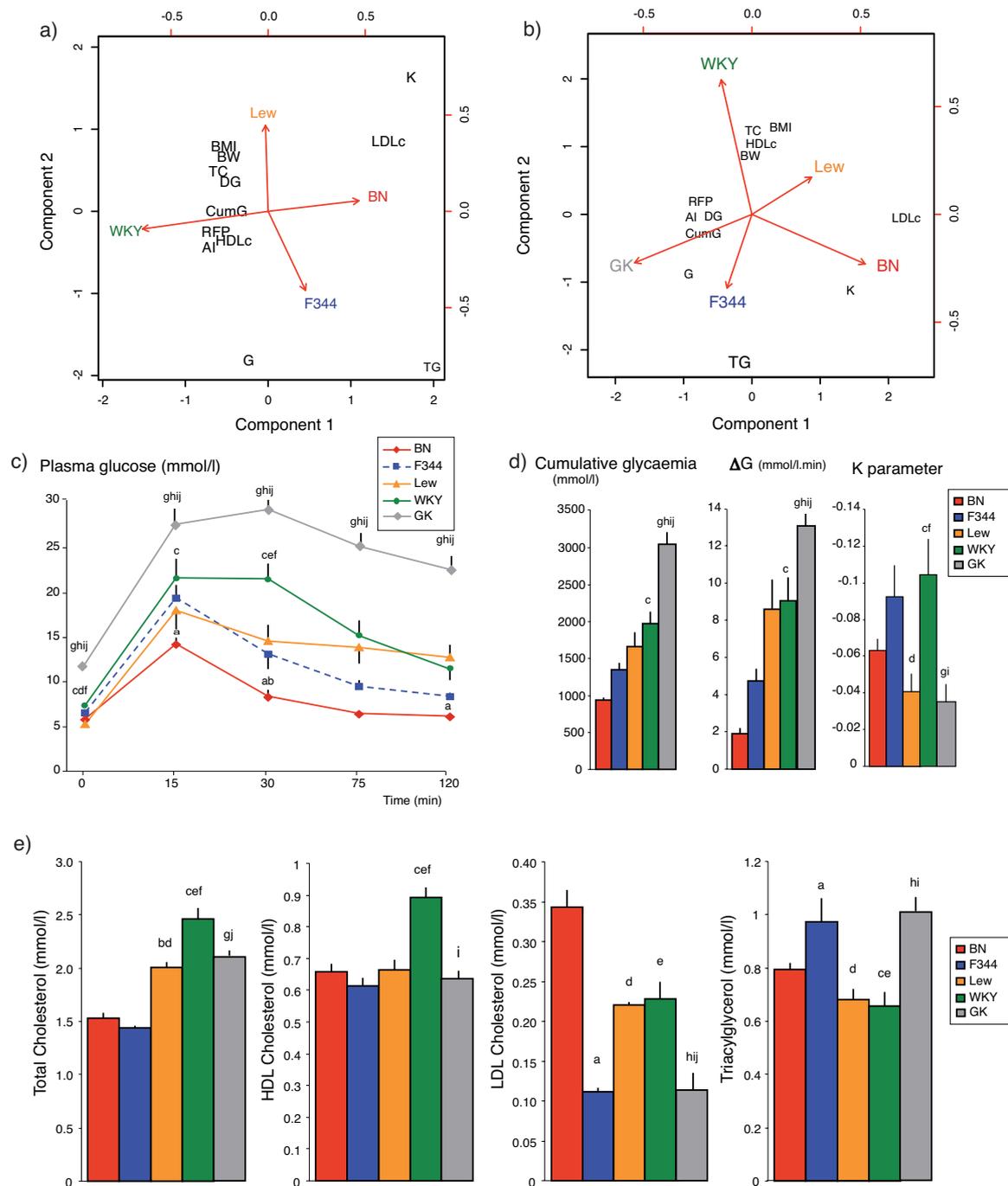


Figure 3.1.1 Evidence for physiological heterogeneity in inbred rat strains. Biplot graph of PCA model built with physiological data obtained in strains BN, Lew, F344 and WKY a) and in BN, Lew, F344, WKY and GK b), glucose tolerance illustrated by glycemic profiles during the IPGTT c), cumulative glycaemia, ΔG and K parameter during the test d), and plasma concentrations of total, HDL and LDL cholesterol and triacylglycerol e) in the five inbred rat strains are shown. Data shown are mean \pm SEM. Differences between strains were assessed by Fisher's LSD and Tamhane's T2 post hoc tests. Significant differences ($P < 0.05$) between strains are shown as: a, BN vs F344; b, BN vs Lew; c, BN vs WKY; d, F344 vs Lew; e, F344 vs WKY; f, Lew vs WKY; g, GK vs F344; h, GK vs Lew; i, GK vs WKY; j, GK vs BN. BW: body weight; BMI: body mass index; RFP: retroperitoneal fat pad weight; AI: adiposity index; TC: total cholesterol; HDLc: HDL cholesterol; LDLc: LDL cholesterol; TG:

triglycerides; G: glycemia; CumG: cumulative glycemia, K and ΔG were calculated with data from the IPGTTs (see Methods). For BW, BMI, RFP and AI sample numbers are the same as stated in **Table 4.4.1**. For BN: LDLc, HDLc, TC, TG (n = 32), G, CumG, DG, K (n = 15). For Lewis n = 4, for GK, F344 and WKY, n = 5.

Metabotyping of inbred control rats. To identify strain-specific metabolites underlying such striking physiological plasticity in healthy control strains, we acquired high-resolution 700 MHz $^1\text{H-NMR}$ spectra from urine and plasma collected from BN, Lew, F344 and WKY rats (**Figure 4.4.4**). An O-PLS-DA model was derived for each biofluid. Robustness of these models was assessed by high goodness-of-fit parameter values, $R^2X = 0.780$ and $Q^2Y = 0.590$ for the model based on plasma data and $R^2X = 0.753$ and $Q^2Y = 0.637$ for the model based on urinary data. For validation purposes, models were randomly re-sampled 999 times following the null hypothesis (*i.e.* assuming there are no differences between strains). We observed a decrease in associated goodness-of-fit parameters (**Figure 4.4.5**), reinforcing the validity of our initial models. A significant discrimination between the four strains was observed for both O-PLS-DA scores plots (**Figure 3.1.2a-b**), which define obvious clusters (**Figure 3.1.2c-d**), suggestive of strain-specific urinary and plasma metabolites, as derived from O-PLS-DA loadings (**Figure 3.1.3**). Only metabolites displaying correlation values superior to 0.5 were assigned in model coefficient plots.

Natural variation in metabolites. Structural assignment of the biofluid-specific O-PLS-DA loading plots reveals broad-ranging metabolic variations in plasma (**Table 3.1.1** and **Figure 3.1.3**), involving lipids (CH_3 ; CH_2 ; $\text{C}=\text{CCH}_2\text{C}=\text{C}$; $\text{CH}_2=\text{C}$; $\text{CH}_2\text{CH}_2\text{CO}$, CH_2CO ; VLDL; LDL) and amino acids (alanine, glycine, valine and isoleucine). In urine, variation in the choline oxidation pathway (choline, betaine, *N,N*-dimethylglycine, creatinine) but also tricarboxylic acid (TCA) intermediates and energy metabolites (lactate, citrate, succinate, 2-oxoglutarate), short-chain fatty acids (valerate) or gut microbiota related metabolites (phenylacetyl-glycine and hippurate) amongst others is observed (**Table 3.1.1**).

Relationships between natural variations at genetic, physiology and metabolic levels. To investigate the relationship between strain-specific genetic, physiological and metabolic patterns, we compared phylogenetic trees obtained from published single nucleotide polymorphism (SNP) data in the rat strains (<http://gscan.well.ox.ac.uk/gscanBleedingEdge/rat.snp.selector.cgi>),^[192] physiological data and $^1\text{H-NMR}$ based metabolic profiles (**Figure 3.1.4**). Variance within the SNP, plasma and urine datasets was similar. Pairwise dendrogram comparisons were performed by computing cophenetic

correlations (cophenetic correlations measure the similarity between dissimilarity matrices obtained from the different types of data: genotypes, physiology, plasma and urine metabolomes). Physiological data and plasma metabolic profiles presented positive cophenetic correlations, but no cophenetic correlation was observed between urine metabolic profile and SNP clustering trees (**Table 3.1.2** and **Figure 3.1.4**). Interestingly these cophenetic tendencies reflect the influence of genetics over homeostasis, including physiological phenotypes and plasma metabolotypes, whereas urinary metabolotypes exhibit phenotypic hypervariability unmatched by SNP variability in the rat genome.

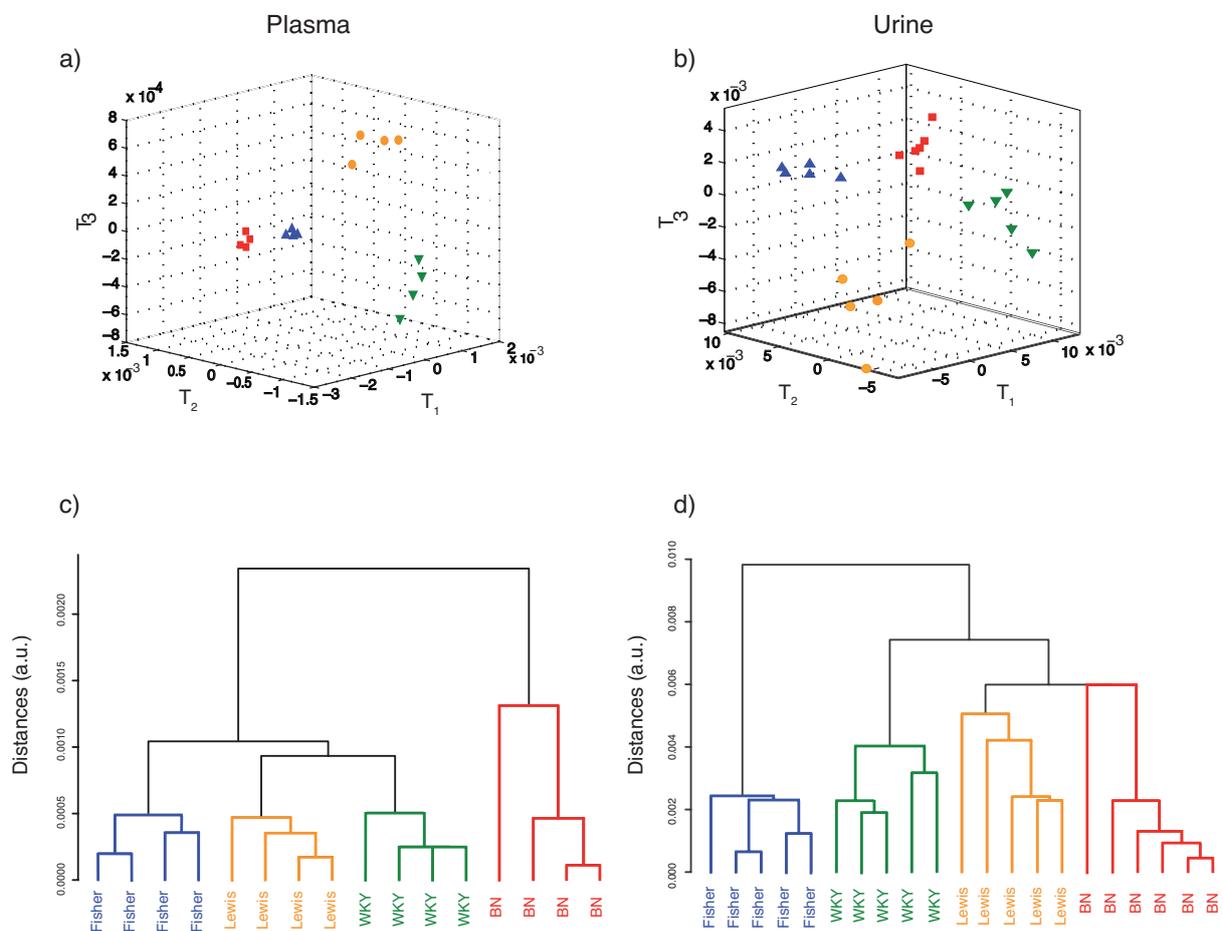


Figure 3.1.2. Urine and plasma metabolic variability in control strains. 3D score plots of OPLS-DA models a), b) and subsequent hierarchical clustering trees c), d) for plasma samples ($n = 16$, $Q^2Y = 0.590$), a) c), urine samples ($n = 20$, $Q^2Y = 0.753$), b), d). ■ BN, ● Lewis, ▲ Fisher, ▼ WKY. For a) and b), the different axis T1, T2 and T3 correspond to the 3 predictive components of each OPLS-DA model. For c) and d), vertical axis correspond to distances between individuals calculated with Euclidean metrics (in the OPLS predictive score subspace), expressed in arbitrary units.

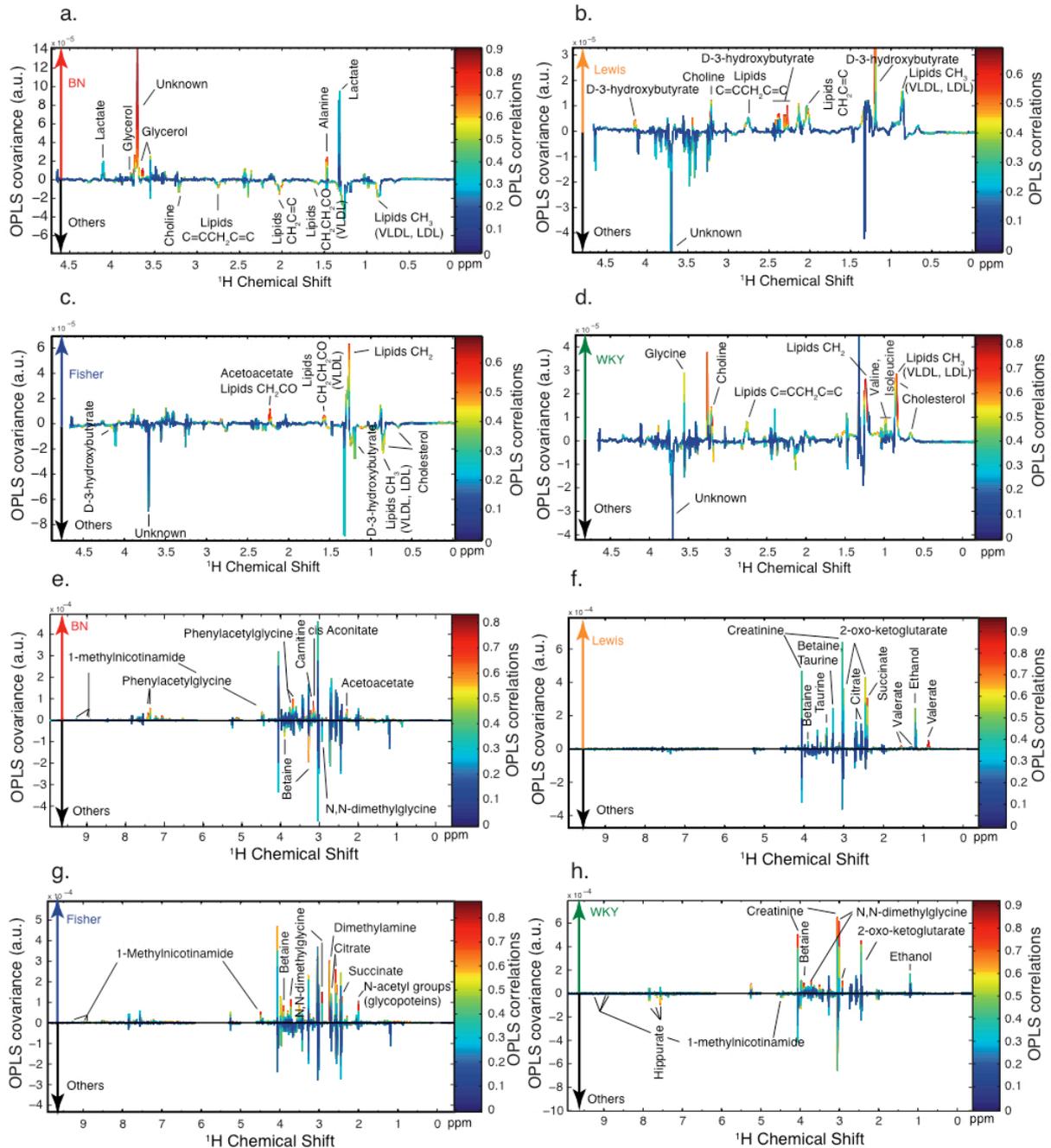


Figure 3.1.3. Metabolic signatures. Loadings plot from plasma O-PLS-DA models a) BN vs others, b) Lewis vs others, c) Fisher vs others and d) WKY vs others and urine O-PLS-DA models e) BN vs others, f) Lewis vs others, g) Fisher vs others and h) WKY vs others.

Natural metabotype variation networks. We used correlation networks (**Figure 3.1.5**) to simultaneously visualize the metabotype of the different strains. Metabolites and strains were connected in a bipartite graph according to their correlation values (see Methods). Metabolites located at the network centre, connecting a large number of strains, were suggestive of common differentially active pathways. On the contrary, strain-specific mono-associated

metabolites located at the network periphery, reflected strain-specific metabolic pathway activity.

Table 3.1.1. Summary of significant metabolites derived from OPLS-DA models of urine and plasma ^1H NMR spectra^a.

| Metabolite | δ (ppm) & multiplicity | BN (r) | Lewis (r) | F344 (r) | WKY (r) |
|--|-------------------------------|--------|-----------|----------|---------|
| Plasma | | | | | |
| acetoacetate | 2.23 (s) | | | 0.79 | |
| alanine | 1.48 (d) | 0.83 | | | |
| D-3-hydroxybutyrate | 2.29 (m), 1.19 (d) | | 0.8 | -0.6 | |
| cholesterol (C18 in HDL) | 0.67 (m) | | | -0.57 | 0.69 |
| choline | 3.21 (s) | -0.65 | 0.7 | | 0.85 |
| glycerol | 3.64 (dd) | 0.92 | | | |
| glycine | 3.55 (s) | | | | 0.68 |
| isoleucine | 1.00 (d) | | | | 0.73 |
| lactate | 4.11 (q) | 0.63 | | | |
| lipoprotein (C=CCH ₂ C=C) | 2.74 (m) | -0.82 | 0.69 | | 0.68 |
| lipoprotein (CH ₂) | 1.26 (m) | | | 0.71 | 0.9 |
| lipoprotein (CH ₂ =C) | 2.03 (m) | -0.82 | 0.72 | | |
| lipoprotein (CH ₂ CH ₂ CO, VLDL) | 1.59 (m) | -0.8 | | 0.71 | |
| lipoprotein (CH ₂ CO) | 2.23(m) | | | 0.79 | |
| lipoprotein (CH ₃ , LDL, VLDL) | 0.87(m) | -0.73 | 0.65 | -0.63 | 0.85 |
| valine | 0.97 (d) | | | | 0.72 |
| unknown | 3.7 (s) | 0.92 | | | |
| Urine | | | | | |
| 1-methylnicotinamide | 4.47 (s) | 0.73 | | 0.84 | -0.76 |
| 2-oxoglutarate | 3.01 (t), 2.45 (t) | | 0.72 | | 0.89 |
| acetoacetate | 2.3 (s) | 0.8 | | | |
| betaine | 3.89 (s), 3.27 (s) | -0.73 | 0.6 | 0.84 | 0.92 |
| carnitine | 3.23 (s) | 0.8 | | | |
| citrate | 2.68 (d), 2.57 (d) | | 0.54 | 0.87 | |
| creatinine | 4.05 (s), 3.04 (s) | | 0.58 | | 0.85 |
| dimethylamine | 2.73 (s) | | | 0.76 | |
| ethanol | 1.19 (t) | | 0.65 | | 0.53 |
| hippurate | 7.55 (t) | | | | -0.79 |
| N,N-dimethylglycine | 2.93 (s) | -0.56 | | 0.91 | 0.83 |
| phenylacetyl glycine | 7.35 (m) | 0.81 | | | |
| succinate | 2.41 (s) | | 0.83 | 0.83 | |
| taurine | 3.43 (t) | | 0.54 | | |
| cis-aconitate | 3.16 (s) | 0.88 | | | |
| glycoprotein (N-acetyl) | 2 (s) | | | 0.92 | |
| valerate | 0.87 (t) | | 0.97 | | |

^a Correlation coefficients (r) between metabolites and strains were displayed when $|r| > 0.5$. Note that other resonances from a given metabolites may not pass the threshold because of signal overlap.

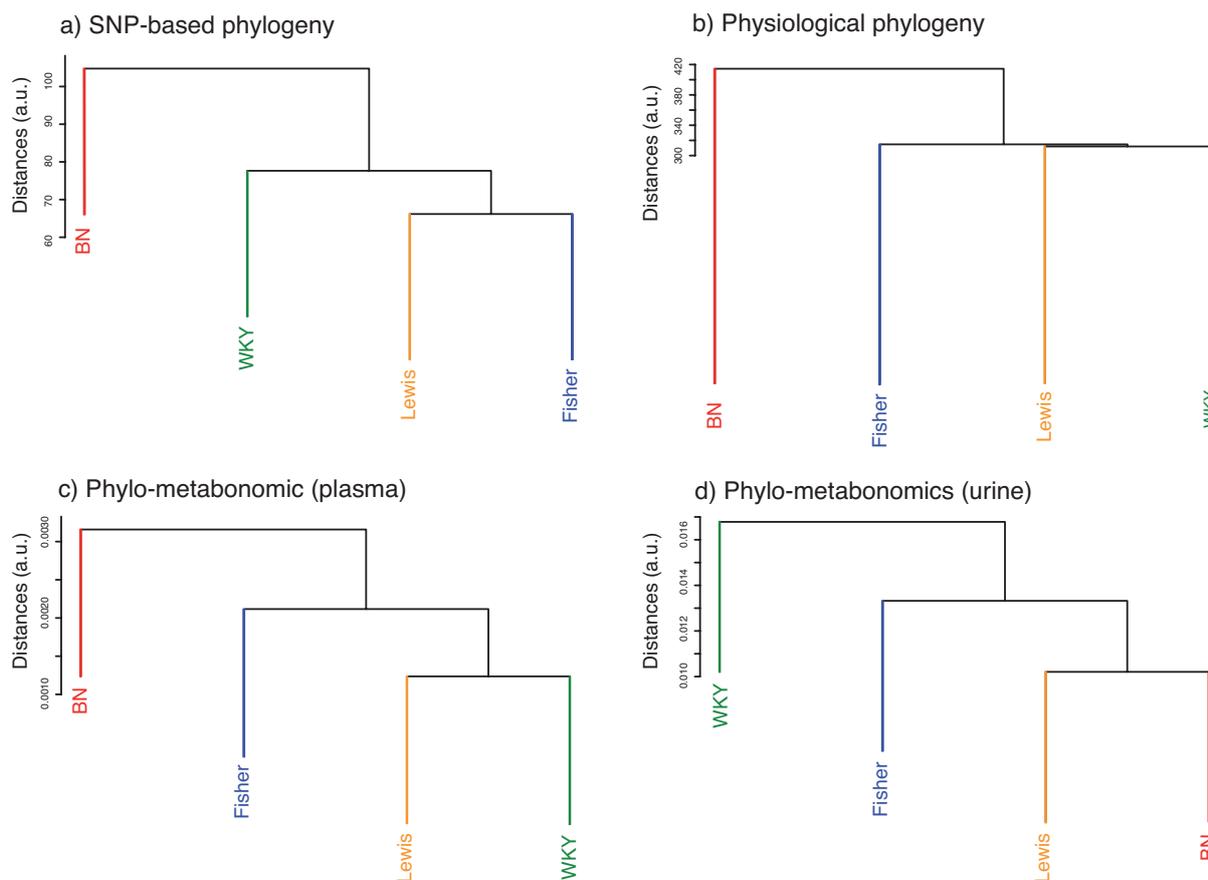


Figure 3.1.4. Urine and plasma metabolic variability in control strains and relationship to phylogeny. Clustering analysis of the BN, Lewis, Fisher and WKY strains, derived from a) SNP data (16,829 variables, $n = 1$ per isogenic strain), b) physiological data (12 variables, $n > 8$ per strain), c) $^1\text{H-NMR}$ plasma metabonomic data (10,999 variables, $n = 4$ per strain) and d) $^1\text{H-NMR}$ urine metabonomic data (10,999 variables, $n = 5$ per strain). The vertical axis corresponds to distances between strains calculated with an Euclidean metrics in arbitrary units (derived from the initial variables). Total variance for each dataset was calculated by computing the sum of eigenvalues from the UV-scaled matrix of strain average spectra: $\text{Var}(\text{SNP dataset}) = 380$; $\text{Var}(\text{physiological dataset}) = 13$; $\text{Var}(\text{plasma dataset}) = 275$; $\text{Var}(\text{urine dataset}) = 293$.

The correlation networks based on plasma data (**Figure 3.1.5a**) reveals that saturated lipids (CH_3 ; VLDL, LDL) were the most connected metabolites (*i.e.* hubs), associated with the four strains, whereas unsaturated lipids ($\text{C}=\text{CCH}_2\text{C}=\text{C}$) and choline were associated with three strains. Glycine, valine and isoleucine were specific to WKY, CH_2CO and acetoacetate to F344 and glycerol, alanine, lactate and an unknown signal at 3.70 ppm to BN. Lew rats did not display any specific mono-association pattern.

According to the correlation network derived from urine data (**Figure 3.1.5b**), BN rats displayed the largest number of specific, mono-associated metabolites: *cis*-aconitate,

carnitine, acetoacetate, and phenylacetyl glycine. Taurine and valerate were specifically mono-associated to Lew, dimethylamine and N-acetyl groups of glycoproteins to F344 and hippurate to WKY. Betaine and 1-methylnicotinamide were highly connected to respectively 3 or 4 strains, showing pluri-associations.

Table 3.1.2. Pairwise comparison of hierarchical clustering trees using cophenetic Pearson correlation coefficient.

| | SNP | Physiology | Plasma | Urine |
|------------|-----|-------------|-------------|-------|
| SNP | 1 | 0,96 | 0,84 | -0,32 |
| Physiology | | 1 | 0,92 | -0,44 |
| Plasma | | | 1 | -0,5 |
| Urine | | | | 1 |

Finally, in order to evaluate the systems-wide metabolic variation in both biofluids, we merged plasma and urine data in a unique correlation network (**Figure 3.1.5c**) and we observed that saturated lipids (CH_3 ; VLDL, LDL), unsaturated lipids ($\text{C}=\text{CCH}_2\text{C}=\text{C}$), betaine, 1-methylnicotinamide and choline were highly connected to three or four strains. Valine, isoleucine, glycine, hippurate and cholesterol were WKY-specific, whereas *cis*-aconitate, glycerol, carnitine, lactate, alanine, phenylacetyl glycine and an unknown metabolite corresponding to a NMR signal at 3.70 ppm were BN-specific. The other two strains displayed a relatively small number of specific mono-associated metabolites. Only valerate and taurine are characteristically associated to Lewis and N-acetyl groups of glycoproteins, dimethylamine and oxidized lipids (CH_2CO) to F344. These results illustrate the power of metabolomics to identifying genetically determined metabolic signatures that contribute to phenotypic variability irrespective of disease susceptibility.

Mapping broad-ranging metabotype variation onto known metabolic pathways.

Metabolic network reconstruction is now part of the post-genomic effort.^[193] Such metabolic pathway databases can be used to reveal higher-order systemic operation of the cell and the organism. We developed a metabolite-set enrichment analysis (MSEA) strategy to identify over-represented metabolic pathways^[106] and map strain-specific metabolites (or metabolite-set), onto the KEGG database,^[193] (see Methods, **Figure 4.4.2** and **Figure 3.1.6a**). We then visualized significant strain-pathway associations (assessed by Fisher's exact test) under the form of a correlation network (**Figure 3.1.6b**). Natural variation in glycine, serine and threonine metabolism was associated with all four strains and propanoate metabolism with three strains. WKY displayed the largest number of specifically active pathways with bile

Synthesis and degradation of ketone bodies and tyrosine metabolism were F344-specific. F344 and Lew were both associated with butanoate metabolism, tricarboxylic acid (TCA) cycle and glyoxylate and dicarboxylate metabolism. **Figure 3.1.5** summarizes the *metabolites* associated with each strain for plasma (**Figure 3.1.5a**), urine (**Figure 3.1.5b**) or both (**Figure 3.1.5c**), whereas **Figure 3.1.6b** summarizes the *metabolic pathways* significantly enriched in each strain, as derived by a MSEA performed on the metabolites from **Figure 3.1.5c**.

Note that in BN and WKY strains, a compound involved in butanoate metabolism was significantly affected. However, the MSEA revealed that this pathway was not significantly enriched, as the number of observed metabolites for these strains was larger than that of LEW F344, decreasing the relative representation for this pathway. Interestingly, although BN had the largest number of specific metabolites (**Figure 3.1.5c**), this strain did not display any specific pathway. In fact, these specific metabolites belonged to several metabolic pathways and thus none of them was significantly enriched.

The application of NMR-based MSEA in genetically diverse individuals is a powerful approach to generate a global and accurate overview of altered metabolic pathways caused by genetic polymorphisms.

3.1.4 Discussion

We report an unexpectedly high level of non-morbid phenotype variability in normal control inbred rat strains at both metabolite and metabolic pathway levels, using a purpose-built metabolite-set enrichment analysis (MSEA)^[106] based on ¹H NMR metabolic profiles. These knowledge-based over-representation approaches, popular in transcriptomic studies,^[190, 191] enhance the interpretation of complex hypothesis-free metabolic signatures. Natural metabolic variation patterns between rat strains revolve around key compounds and pathways (**Figure 3.1.7**) and appear as subtle modulations of main hubs of mammalian metabolism (e.g. lipid, amino acid, energy metabolism) in interaction with the gut microbiome, which collectively reflect distinct physiological regulatory mechanisms contributing to the maintenance of homeostasis.

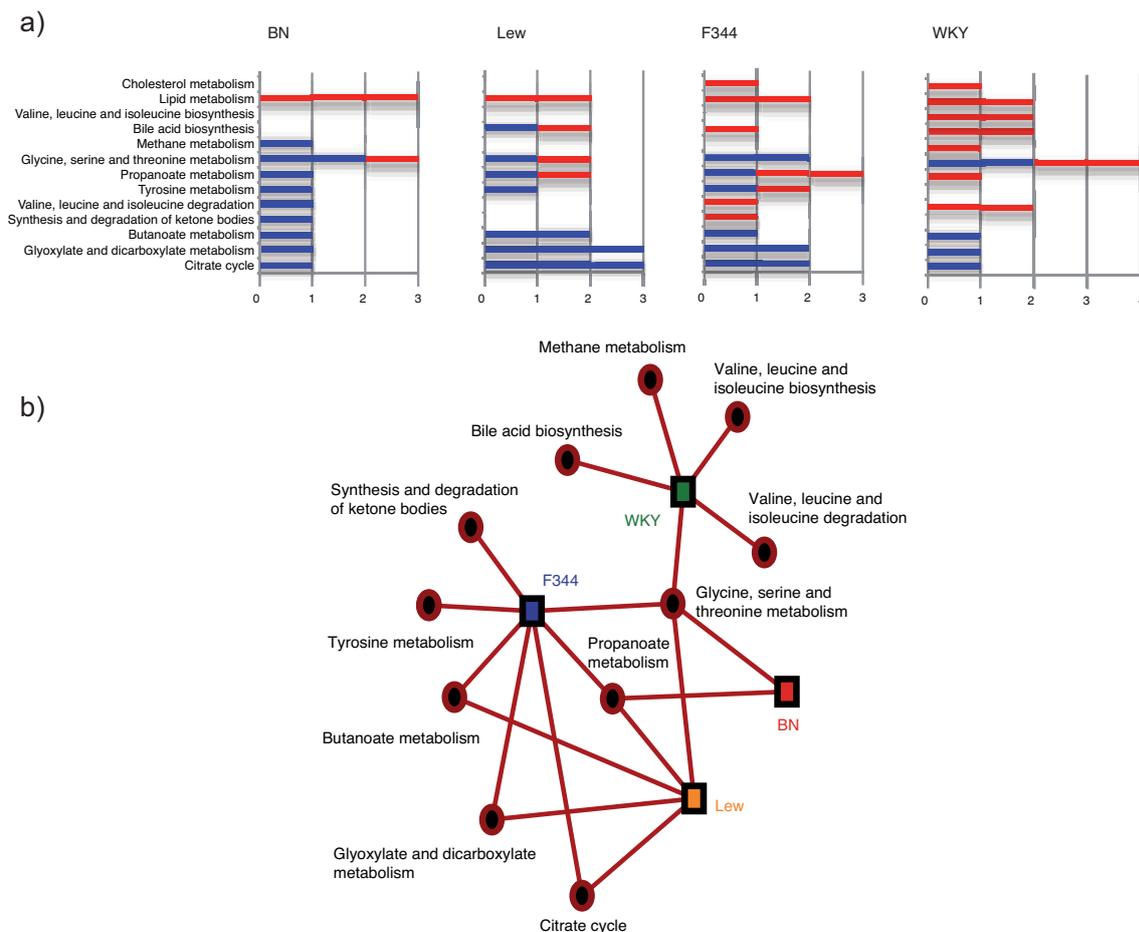


Figure 3.1.6. Strain-specific Metabolic Set Enrichment Analysis (MSEA). a) List of pathways significantly enriched for each strain. Bars correspond to number of involved metabolites in each pathway in blue for urine and red for plasma. b) Correlation networks visualizing association between strains (colored squares) and metabolic pathways (black circles), derived from OPLS-DA models listed in **Figure 3.1.2**, **Figure 3.1.3** and **Table 3.1.1**.

The maintenance of healthy metabolic phenotypes that we demonstrate in inbred control strains strongly argues against recently published opinions that rodent models fed *ad libitum* systematically exhibit morbid phenotypes.^[181] The definition of both healthy phenotypes in inbred control strains and covariates that affect phenotype expression are central considerations in the experimental design of rodent genetic studies, in order to maximise the detection of genetic polymorphisms in hybrid cohorts accounting for phenotypic differences between disease susceptible and resistant models (www.jax.org/phenome).^[183, 184, 194] The vast majority of these studies in all disease areas have used phenotypic data from animals fed *ad libitum* and led to the identification of disease-related QTLs. Results from genetic studies in cohorts derived from inbred rat models of hypertension and diabetes bred to different inbred controls have demonstrated chromosomal

clustering of linkages to cardiometabolic traits,^[195-197] providing evidence of QTL replication which is a gold standard criterion to assess rodent QTL robustness^[198] Our results provide a comprehensive phenotype screen that can be used for selecting the most appropriate healthy control strain in genetic studies of metabolic phenotypes. Phenotype and metabonomic divergences between control strains also support the possibility to map the genetic basis of complex phenotypes in the rat heterogeneous stock which derives from eight progenitor strains, including the WKY, BN and F344 strains tested here.^[183,184]

Our results illustrate the system-level biological robustness achieved through distributed control of cellular processes leading to an overall stabilization of the organism.^[186] According to the metabolic entropy framework, homeostatic normalization, a physiological process involved in correcting any biological process diverging in physiologically-controlled parameters, leads to a metabolic penalization.^[185] We find an illustration of this metabolic penalization in the higher cophenetic correlation between SNP, physiological and plasma metabolites, these parameter being controlled at the genetic level, whilst urinary metabolites present a stronger variation, presumably related to metabolic compensation at the cellular and systemic level, as well as to active transport phenomena in kidney physiology.^[199] The identification of the genes influencing physiological data and plasma or urinary metabolites would require analyzing genome polymorphisms and metabolites in a segregating population (mQTL studies^[178,179]) and is clearly outside the scope of this work.

Physiological plasticity and metabolic variability in control strains. Physiological data and metabolic profiles highlighted broad differences between the four strains tested, which, for all phenotypes, remained within a normal range. Rats of the WKY strain displayed low glucose tolerance, high plasma cholesterol, elevated body weight and adiposity when compared to BN, F344 and Lew rats. F344 rats showed divergence in plasma LDL cholesterol and triglycerides, whereas Lew rats generally had intermediate phenotype values. Interestingly, such heterogeneity in the four strains could be replicated using urine and plasma metabolic profiling data, suggesting the existence of strain-specific metabolic signatures. The development and application of MSEA allowed mapping spectroscopically-derived metabolic signatures onto known metabolic pathways. For a given set of metabolites, several different affected pathways were identified due to interconnection and involvement of a given metabolite in several pathways. For instance, metabolites associated with Lew include citrate, succinate and 2-oxoglutarate, which indicated perturbations in TCA cycle and glyoxylate and dicarboxylate metabolism. These pathways were also altered in F344, whereas valine,

leucine, isoleucine biosynthesis and degradation were found altered in WKY, compared to the other strains. Although BN displayed the largest number of mono-associated metabolites, MSEA only identified two metabolic pathways. This observation highlights the necessity to derive appropriate knowledge-based bioinformatic tools, such as MSEA, to enhance the systems-wide interpretation of broad-range metabolomic variation derived from experimental metabolic datasets. In the present case, broad-range natural metabotypic variation supports physiological plasticity, *i.e.*, homeostatic, non-pathological variation, in primary phenotypes, *i.e.*, BMI, blood lipids, IPGTT, and can be summarized as outlined hereafter.

Differential affinity for energy metabolism substrates and products. Energy metabolites contributed to strain discrimination, in particular for F344 and Lew, which showed excretion of metabolites involved in TCA cycle (**Figure 3.1.7**). TCA intermediates excretion is dominated by renal tubular pH and renal transporter function.^[200] The TCA cycle is the common final pathway for the oxidation of fuel molecules, such as carbohydrates, fatty acids and amino acids. Associated with oxidative phosphorylation, in aerobic conditions, it allows the release of the maximum amount of energy from fuel molecules.^[201] Thus, it is a central hub role in intermediate metabolism and as a consequence is connected to a large number of pathways. For instance, 2-oxoglutarate, citrate and succinate are all involved in alanine, glutamate and aspartate degradation, and in glyoxylate and dicarboxylate pathways as well as in reductive carboxylate pathways. Amino acid carbon backbones enter the TCA cycle through 2-oxoglutarate for arginine, glutamate, glutamine, histidine, proline and through succinyl-CoA for isoleucine, methionine, threonine and valine.^[201] These TCA cycle intermediates can also be converted into phosphoenolpyruvate and then into glucose. The level of excretion should depend on different complex interactions between these previous pathways, including regulation by substrate availability and product inhibition. Furthermore F344 was associated with higher acetoacetate and D-3-hydroxybutyrate, *i.e.*, ketone bodies, suggestive of ketosis. These molecules are mainly produced in the liver from fatty acids or certain amino acids and are released in blood to be transported in peripheral tissues where they can be used as sources of energy instead of glucose (**Figure 3.1.7**). These data thus suggested differential mobilisation of fatty acid to produce energy.^[201]

Consequences for lipid metabolism. Lipid parameters quantified by classical assays provide complementary data to those measured by ¹H NMR spectroscopy, which provides data on functionality of lipids (unsaturation patterns, oxidized functions) rather than quantifying the different lipoproteins. For this reason, lipid metabolism was not part of MSEA via KEGG

obesity-associated insulin resistance, in the context of high fat diet through the interplay of mTor,^[205] which seems consistent with increased adiposity and low glucose tolerance in WKY.

Influence of gut microbiota metabolism. We investigated the different transgenomic interactions between host and gut microbiota in the four strains. Gut microbiota metabolism influences mammalian phenotypes, through a complex molecular crosstalk between bacterial and mammalian genes.^[206-208] For instance, we observed a specific reduction of hippurate excretion in WKY rats. Hippurate is a mammalian-microbial co-metabolite formed by glycine conjugation of benzoate, which is produced from bacterial degradation of plant polyaromatic compounds.^[179, 209] Consistently, WKY displayed also specific high plasma level of glycine. These data suggest that WKY metabolism exhibits a specific transgenomic interaction with gut microbiota, leading to reduced benzoate formation and thus reduced glycine conjugation and hippurate excretion, explaining the observed metabotype. Along the same line, changes in dimethylamine excretion were found specifically in F344. Methylamine formation is related to microbial processing of dietary choline in the gut.^[203, 210] Interestingly, when compared to WKY and Lew, F344 had lower plasma choline level. We identified BN-specific changes in phenylacetyl-glycine excretion, which is directly linked to gut microbiota activity.^[211] This observation indicates BN-specific microbiota transgenomic interaction. As all strains used in this study were fed the same diet and maintained in identical environmental conditions, these different results indicate a complex modulation from the host strain on the composition or activity of its gut microbiota, highlighting the complexity of mammalian organisms through transgenomic interactions and fully justifying an integrative systems biology approach.

Relationships between natural metabolic variation, host genetic polymorphisms and microbiome metabolic functions. We observed that the four strains, often used as healthy or disease resistant controls in genetic studies, were markedly different at the physiological and metabolic levels. Our results show that, quantitatively, genetic variations investigated through SNP-based hierarchical clustering, did not match metabolic changes in our experimental conditions used here for physiological studies and sample collection. However, plasma metabotypes correlate strikingly more with genetics than urinary metabotypes. Plasma composition and their associated metabolic processes are tightly controlled by homeostatic regulation to maintain organism integrity, involving complex protein interaction networks, determined to a certain extent by genetics. As for urine, it contains end products of metabolism, the excretion of which is a reflection of the balance between the dietary inputs

and the metabolic requirements, so that the relationship with genetic determinants is weaker. These observations are made at the genome-wide and metabolome-wide level. The identification of gene variants influencing specific metabolite levels is possible using an mQTL approach, which is out of the scope of the current study.^[178, 179] These results support the concept that metabolic variations cannot be entirely reduced to their genetic component, and should be understood through an integrative systems biology process associating different parameters, such as genetics, but also environmental influences such as exercise, diet and host-gut microbial transgenomic interactions.^[212]

3.1.5 Conclusion

Altogether, these results introduce the concept of natural metabolomic variation driving physiologically stable, albeit diverse, phenotypic outputs within the range of normality, through dynamic readjustment of the fluxes among the metabolic network,^[213] as well as differential excretion rates, both phenomena being suggestive of systems-wide robustness. Ongoing efforts in genome resequencing in inbred rat strains^[214] may shed light on naturally occurring polymorphisms that cause non-morbid phenotype variability. Our results bring further support to the complexity of mammalian metabolism involving strain-specific transgenomic (host/symbiont) interactions. By highlighting physiological and metabolic differences between strains, our data contribute to improving design of future genome-wide metabolome-wide studies, such as mQTLs discovery, in a truly integrative genomic and systems biology context. The possibility to identify metabolotypes separating groups of control individuals demonstrates the power and sensitivity of metabolomics, which may directly impact human clinical and genetic studies through the detection of metabolic biomarkers allowing fine stratification of patients and controls.

3.2 Metabolic phenotyping of *Caenorhabditis elegans*

Originally introduced in the early 1970s by Sydney Brenner as a multicellular genetic model for development and organogenesis,^[215-217] the nematode *Caenorhabditis elegans* (*C. elegans*) became a key model organism in modern biology, being complex enough to address high-order questions but still, easy to work with.

This 1 mm long transparent roundworm belongs to the invertebrate multicellular eukaryotes. The worms are sexually dimorphic with males and self-fertilizing hermaphrodites. The later possess 959 somatic cells, fully differentiated in organs, organized in systems (epithelial system, reproductive system, nervous system, excretory system and muscle system).^[218] At 20°C, *C. elegans* has a short generation time of approximately 3 days, corresponding to the development of fertilized eggs into sexually mature adults, through 4 larval stages termed L1, L2, L3 and L4. *C. elegans* has an average lifespan of two to three weeks. After the L2 stage, if the environmental conditions are not favorable for growth, worms may enter an alternative larval stage termed *dauer* in which they develop a thick cuticle and stop feeding, surviving for several months, until experiencing more favorable conditions to proceed to L4 stage.

C. elegans attractiveness as a model organism was reinforced by its experimental tractable management and by major biological characterizations. Indeed, *C. elegans* is inexpensive and easy to grow on a diet of bacteria in Petri dishes. Important developmental and anatomical descriptions were achieved as for instance, the identification of the complete lineage of every cell^[219] or the entire nervous system reconstruction.^[220] The discovery of the usefulness of RNA interference to manipulate gene expression^[221] and the complete *C. elegans* genome sequencing were also two major steps in the field of genetics.^[222]

In combination, all these features made *C. elegans* a key model organism for the study of genomics, neuroscience, cell biology, aging or human diseases, such as neurodegenerative diseases.^[217]

Intermediary metabolism appears highly conserved between *C. elegans* and humans, as for most key eukaryotic enzymes involved in these metabolic pathways, corresponding genetic patterns, termed orthologs have been found in the worm genome.^[217, 223] The nematode has thus all the metabolic machinery for the fundamental pathways of cellular respiration or anaerobic glycolysis.^[223] Furthermore, unlike humans, *C. elegans* can perform also in anaerobic ethanolic fermentation^[224] and malate dismutation, which produces acetate,

propionate and succinate as waste products while using fumarate as electron acceptor instead of oxygen.^[217, 225]

With the development of metabolic profiling strategies, the study of *C. elegans* metabolism is of a growing interest recently. Blaise *et al.* developed a whole-organism NMR-based approach, relying on the use of the HR-MAS NMR spectroscopy to derive metabolic signatures directly from whole organisms.^[27, 44] In the meantime, different approaches were also developed, coupling a first extraction step with solution NMR spectroscopy or mass spectrometry.^[30] As the metabolic phenotype can be considered as a functional readout, many studies used it in a functional genomics context, aiming at deciphering gene functions, in the fields of aging,^[226-229] oxidative stress^[27, 230, 231] or fatty acid metabolism.^[232, 233] Metabolic profiling was also used to characterize worm metabolic perturbations in response to external stimuli such as exposure to toxic agents^[234, 235] or diet variations.^[236] Metabolomics has been also used to identify small molecules signals that regulate behavior and development in *C. elegans*.^[237-239]

In this chapter, we first investigate the effects of aging and dietary restriction on *C. elegans* metabolism by characterizing wild-type worms and different *C. elegans* genetic mutants, displaying different profiles of aging. We then delineate the impact on *C. elegans* metabolism of the protein receptor AHR-1 loss-of-function, ortholog of the Aryl Hydrocarbon Receptor (AHR), involved in dioxin toxicity in mammals, in the framework of functional genomics.

3.2.1 Dietary restriction buffers metabolic changes associated with aging in *Caenorhabditis elegans*

3.2.1.1 Introduction

Metabolism and lifespan are tightly linked. It has been known for decades that modulating metabolism *via* genetic interventions such as mutation in genes encoding the insulin/IGF-1 receptor^[240] or *via* environmental changes such as dietary restriction (DR), significantly extend longevity and delays aging in many species.^[241] Yet, whether physiological aging involves a metabolic shift as an early or late event in the process of aging, and how interventions such as DR specifically affect the metabolic changes associated to aging has not been extensively studied. Metabolomic investigation provides a comprehensive and unbiased strategy to address this question at the whole organism level. In mammals, this approach is restricted to the description of metabolite concentrations in biofluids or specific tissues, which provides complementary but partial information on the homeostatic network of the whole body.^[242-245] We therefore address this question by using the nematode *C. elegans* as a model system, as metabolic phenotypes can be investigated by Nuclear Magnetic Resonance (NMR) spectroscopy to detect global metabolic changes.^[27]

Our study identified a list of metabolites, which concentrations vary during early aging. We show that those changes are affected by DR and can be used to discriminate, among worms of the same chronological age, those that are healthier and which will live longer from those that have a short life expectancy.

3.2.1.2 Material and Methods

Nematode strains and culture conditions (Laurent Mouchiroud *et al.* Université de Lyon). *C. elegans* strains were cultured at 20°C on nematode growth media (NGM)^[215] agar plates freshly poured and seeded with *E. coli* strain OP50 culture. Wild-type Bristol N2, *eat-2(ad465)* II and *daf-18(e1375)* IV strains were provided by the *Caenorhabditis* Genetics Center (University of Minnesota). *slcf-1(tm2258)* mutants were obtained from the *C. elegans* knockout consortium directed by Pr Mitani and outcrossed five times in our wild-type strain. *slcf-1(tm2258);daf-18(e1375)* double mutants were obtained by genetic crosses.

Sample preparation for metabolomics analysis (Laurent Mouchiroud *et al.* Université de Lyon). In order to get rid of variations relative to sample preparation or analysis, the assays were performed on large number of worms (40 000 worms for each age in total, split into

1000 worms per analysed NMR sample) prepared in at least 3 independent experiments. For worm amplification and synchronization, ten adult worms were allowed to lay eggs, on *E. Coli* OP50 seeded 55mm NGM plates, for 2-3 hours at 20°C then removed. When F1 worms reached the preadult-L4 stage, 5-fluorouracil (5-FU, Sigma) was added on top of the plate at a final concentration of 1.30 mg/L (10µM). 5-FU inhibits the development of eggs laid by F1 progeny, allowing the maintenance of a synchronized F1 population until old age while avoiding to transfer worms every couple of days to separate them from their progeny, which represents stressful conditions. In order to standardize the experimental conditions, 5-FU was also added when worms were recovered at the young adult stage. Davis *et al.* have shown that the addition of FUdR to maintain synchronous *C. elegans* population had an impact on the metabolic signature of worms.^[246] However, FUdR, which acts in a concentration dependent manner, was introduced in high concentration (0.4mM, 100 mg.L⁻¹), and induces major biological perturbation in worms as revealed by complete sterility of worms. In our experimental conditions (we used 40 times less concentrated 5-FU) worms still lay eggs that fail to develop.

Synchronized worms were recovered 24 hours later (YA stage) or 7 days later (A7). Worms cultures synchronisation and recovery were set up in order to recover both young adult and 7 days old worms on the same day for all genotypes, and repeated at least 3 times. On the day of recovery, 50 plates for each condition (age/genotype) were washed 5 times in 50 ml of M9 buffer, separated by 5 min sedimentation steps in order to get rid of residual bacteria. Worms were then fixed for 45 minutes in 1% paraformaldehyde and then washed 5 times in distilled water followed by 5 washes in deuterium oxide. 30 µl disposable Kel-f inserts with sealing caps for 4 mm NMR rotors were filled with around 1000 whole worms and frozen at -80°C for storage until NMR analysis. Samples were thawed at room temperature 15 minutes before the NMR experiments.

Whole *C. elegans* HR-MAS NMR spectroscopy. *C. elegans* HR-MAS NMR spectroscopy was performed as previously described by Blaise *et al.*^[27, 44] All experiments were carried out on a Bruker Avance II spectrometer, operating at 700 MHz (proton resonance frequency), equipped with a 4 mm HR-MAS double resonance (¹H-¹³C) probe. Temperature was controlled at 295 K throughout the experiments and magic angle spinning speed set to 3.5 KHz. A set of 1D ¹H NMR experiments with water presaturation, including 1D ¹H NOESY and 1D ¹H CPMG were performed on each sample to derive metabolic profiles. 256 free induction decays (FIDs) were co-added, with a 12 ppm spectral width and an acquisition time

of 1.4 s, corresponding to 23568 data-points, with a relaxation delay of 1.7 s, for a total experimental time of 13 minutes per spectrum. The NOESY mixing time was set to 100 ms and the CPMG spin-echo delay adjusted to 300 μ s for each of the 80 spin-echo loops (48 ms total echo time). The ^1H 90° hard pulse length was calibrated at 7.5 μ s.

2D NMR experiments, including ^1H - ^1H TOCSY and ^1H - ^{13}C HSQC experiments, were carried out on a subset of selected samples to characterize structural connectivities between nuclei and refine metabolite identification.

NMR data processing. All FIDs were multiplied by an exponential function corresponding to a 0.3 Hz line-broadening factor prior to Fourier transform. Phasing and baseline correction were performed manually in Topspin 2.1 (Bruker, GmbH, Rheinstetten, Germany). Spectra were automatically calibrated on the CH_3 alanine doublet at $\delta = 1.48$ ppm. Spectra were reduced over the chemical range of 0.55-8.75 ppm to 8200 bins (10^{-3} ppm wide) with integration of signal intensity. Residual water signal ($\delta = 4.5$ -5 ppm), residual methanol signal originated from formaldehyde fixation step ($\delta = 3.32$ -3.39 ppm) and a noise area ($\delta = 5.5$ -6.5 ppm) were discarded prior to analysis. Spectra were normalized using the probabilistic quotient normalization (PQN) approach,^[97] with a median of all spectra as reference spectrum. We applied a Pareto scaling on the dataset for multivariate analysis only. Metabolite assignment was completed exploiting reference data from the literature,^[27, 226] the HMDB,^[50] MMCD,^[49] bbiorecode-2-0-0 (Bruker, GmbH, Rheinstetten, Germany) and Chenomx NMR Suite 7.0 (Chenomx Inc, Edmonton, Canada) spectral databases.

NMR data analysis. Principal component analysis (PCA)^[187] was first conducted in SIMCA P12+ (Umetrics, Umea, Sweden) to derive the main sources of variance within the dataset, assess sample homogeneity and exclude biological or technical outliers. Orthogonal projection to latent structure discriminant analysis (OPLS-DA) was then performed in MATLAB (The MathWorks Inc., Natick, MA) to derive pair-wise comparison between the different conditions (strains and ages).^[100]

Metabolites involved in class discrimination are then derived from an univariate approach based on the statistical recoupling of variables (SRV) analysis recently described.^[94] SRV corresponds to an automatic binning scheme based on the relationship of covariance and correlation between consecutive variables, which is followed by an univariate unpaired two-tailed t-test calculated for each variable under the Benjamini-Yekutieli correction to cope with multiple testing issue.^[104]

Statistically significant metabolites found in the previous analysis are finally quantified either by direct signal integration in case of non-overlapping signals or by computer assisted manual fitting (deconvolution) of overlapping NMR peaks using the Chenomx NMR Suite 7.0 (Chenomx Inc, Edmonton, Canada). Results are plotted as means and 95% confidence intervals and p-values were calculated for each pair-wise comparison from univariate unpaired two-tailed t-tests.

3.2.1.3 Results and discussion

Metabolic changes correlate with both chronological and physiological age in *C. elegans*.

Wild-type (WT) worms raised at 20°C have a median and maximal lifespan of respectively 17 and 30 days in average.^[228] Under these experimental conditions, obvious morphological changes and functional decline appear after a week and progressively increase until death.^[247] In order to investigate the metabolic variations that occur during early adulthood, we analysed the metabolome of worms staged at two different adult ages: as young adults (YA) before egg production starts, and at day 7 of adulthood (A7), just after egg production ceases,^[228] in order to target a time window preceding the onset of strong morphological alterations while minimising the impact of eggs production on the metabolism. Acquisition of ¹H NMR metabolic profiles (**Figure 3.2.1**) was performed on a pool of intact fixed animals following our recently described high-resolution magic angle-spinning (HRMAS) protocol.^[44] Unsupervised and supervised multivariate statistical modelling show that WT YA and A7 worms can clearly be distinguished by their metabolic fingerprints (**Figure 3.2.2a**, **Figure 3.2.3a-c**, **Table 3.2.1**). YA and A7 worms were essentially isogenic and maintained in a steady environment, while any bias linked to individual phenotype is precluded by our sampling conditions. These data therefore show that metabolic profiles correlate with the chronological age of adult worms and may constitute a fingerprint characteristic of physiological aging. In this case, one would expect that the metabolic profile of worms with extended longevity should harbour a “young fingerprint”, i.e similar to WT YA, at more advanced age. In order to test this hypothesis we analysed the metabolome of worms carrying a mutation in the *slcf-1* gene, which have been shown to increase worms average lifespan by 30% compared to WT animals.^[228] Similarly to WT animals, YA and A7 *slcf-1(tm2258)* mutants can still be separated according to their metabolic profiles from unsupervised or supervised analysis (**Figure 3.2.2a**, **Figure 3.2.4a-c**). Furthermore, WT and long-lived worms can also be discriminated at the same chronological age (YA or A7) (**Table 3.2.1**) Indeed, principal component analysis (PCA) reveals that the metabolic fingerprint of A7 *slcf-*

I mutants is closer to the profiles of young adults, either *slcf-1(tm2258)* or WT, than to the A7 WT fingerprint (**Figure 3.2.2a**).

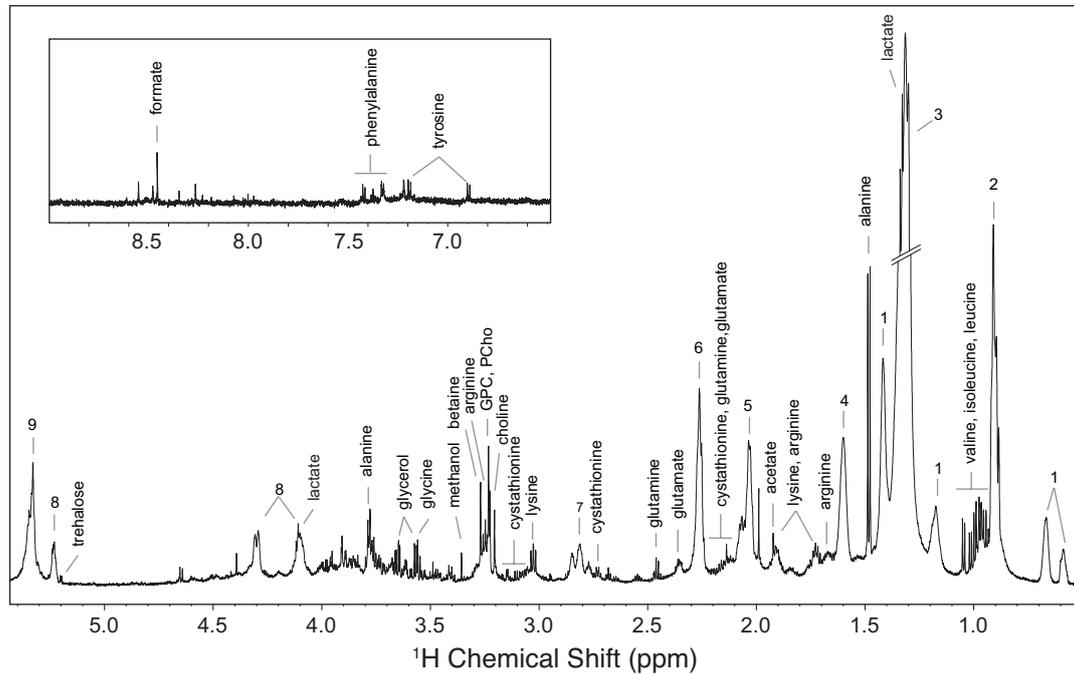


Figure 3.2.1. Typical 700 MHz ^1H HR-MAS NOESY NMR spectrum of whole *slcf-1(tm2258)* *C. elegans* worms for aliphatic ($\delta = 0.5\text{-}5.3$ ppm) and aromatic ($\delta = 6.5\text{-}9$ ppm, magnified 5 times) regions. The resolution of a ^1H HR-MAS NMR spectrum is typically of 1.3 Hz (measured as the width at half height for one of the alanine doublet peaks). Spectra were recorded with a signal-to-noise ratio of 300. Keys: 1. cyclic fatty acids, 2. lipids (CH_3), 3. lipids ($(\text{CH}_2)_n$), 4. lipids ($\text{CH}_2\text{CH}_2\text{CO}$), 5. unsaturated lipids ($\text{CH}_2\text{CH}=\text{CH}$), 6. lipids (CH_2CO), 7. unsaturated lipids ($\text{CH}=\text{CHCH}_2\text{CH}=\text{CH}$), 8. glyceryl of lipids, 9. unsaturated lipids ($\text{CH}=\text{CH}$), PCho: phosphocholine, GPC: glycerophosphocholine

Dietary restriction prevents metabolic changes associated with aging. The *slcf-1* gene encodes a putative monocarboxylates transporter expressed in the intestine of the worm and we have recently shown that its mutation increases longevity by mechanisms similar to DR.^[228] We thus asked whether the difference in the metabolic shift observed with age between WT and *slcf-1(tm2258)* mutants was specific for *slcf-1(tm2258)* mutants or may be a paradigm for metabolic changes that take place in response to DR. To this end we aimed to validate these results by using *eat-2(ad465)* mutants as a second genetic model of DR. The *eat-2* gene encodes a subunit of nicotinic acetylcholine receptors that regulates pharyngeal pumping, which is dramatically slowed down in *eat-2(ad465)* mutants, thus inducing a strong reduction in food intake. *eat-2(ad465)* mutants have been extensively used as a genetic model of dietary restriction since its identification^[248] as it was the only existing DR model before the identification of *slcf-1*. PCA shows a distinct cluster for *eat-2(ad465)* mutants and

discrimination between *eat-2(ad465)* YA and A7 (**Figure 3.2.2a**) confirmed by supervised analysis (**Figure 3.2.4d-f**, **Table 3.2.1**). PCA reveals a common axis for discrimination between YA and A7 in the three strains, but with less amplitude for the two long-lived mutants. In order to further evaluate how long-lived mutants behave along the metabolic coordinates of WT, we projected *slcf-1(tm2258)* and *eat-2(ad465)* individuals onto the OPLS model discriminating YA and A7 WT worms (**Figure 3.2.3d**). YA, for both *slcf-1(tm2258)* and *eat-2(ad465)* mutants, cluster with the WT YA worms, whereas long-lived A7 adults are projected at an intermediate position on the physiological aging axis, between YA and A7 WT worms.

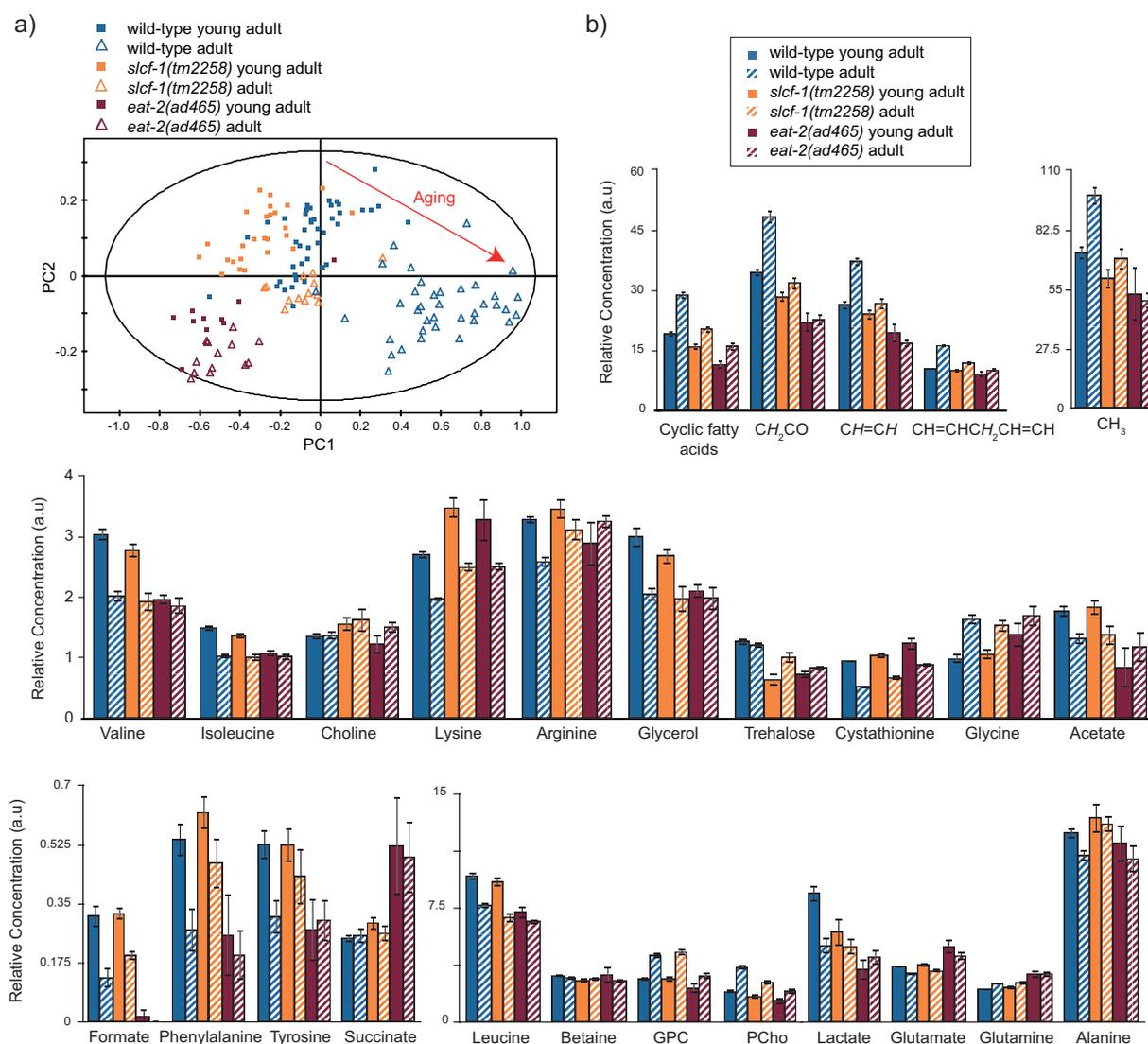


Figure 3.2.2. Metabolic variations in WT, *slcf-1(tm2258)* and *eat-2(ad465)* worms during aging. a) PCA including young adults and adults WT, *slcf-1(tm2258)* and *eat-2(ad465)*; b) Relative concentrations in arbitrary units of 22 metabolites and lipid signals corresponding to specific chemical functions. Results are reported with means and 95% confidence intervals.

Overall those results show that there are fewer differences between old and young long-lived worms for metabolic variations associated with physiological aging than between young and old WT worms and suggest that the metabolic reprogramming triggered by DR specifically prevents the age-associated metabolic variations.

To further investigate this hypothesis we sought to define metabolites that discriminate the different WT worm populations. We identified a set of metabolites which concentrations increase with age: saturated and unsaturated lipids, glycerophosphocholine (GPC), phosphocholine (PCho), glutamine and glycine, while a decrease in concentration is observed for 14 metabolites that includes a range of amino acids (alanine, arginine, isoleucine, leucine, lysine, phenylalanine, tyrosine, valine,), formate and cystathionine, which are both linked to folate metabolism, as well as tricarboxylic acid cycle (TCA) metabolites (glutamate, acetate and lactate) and glycerol (**Figure 3.2.2b and Table 3.2.2**). When considering specifically the metabolites which levels significantly vary with age for the WT worms, we measured for *slcf-1(tm2258)* mutants lower basal levels of lipids and PCho at the YA stage and only a moderate increase with age (**Figure 3.2.2b, Figure 3.2.4b, Table 3.2.2 and Table 3.2.3**). An attenuated decrease in the concentration of alanine, arginine, phenylalanine, tyrosine, cystathionine, and formate was also observed for *slcf-1(tm2258)* aging animals as compared to WT (**Figure 3.2.2b**). Furthermore, a set of common metabolic features clearly discriminated both *eat-2(ad465)* and *slcf-1(tm2258)* animals from the WT worms. These differences include lower levels of lipids, leucine, PCho, trehalose and higher levels of lysine and cystathionine (**Figure 3.2.2b, Table 3.2.3**). These metabolites may therefore constitute a common signature of the long life phenotype for DR mutants.

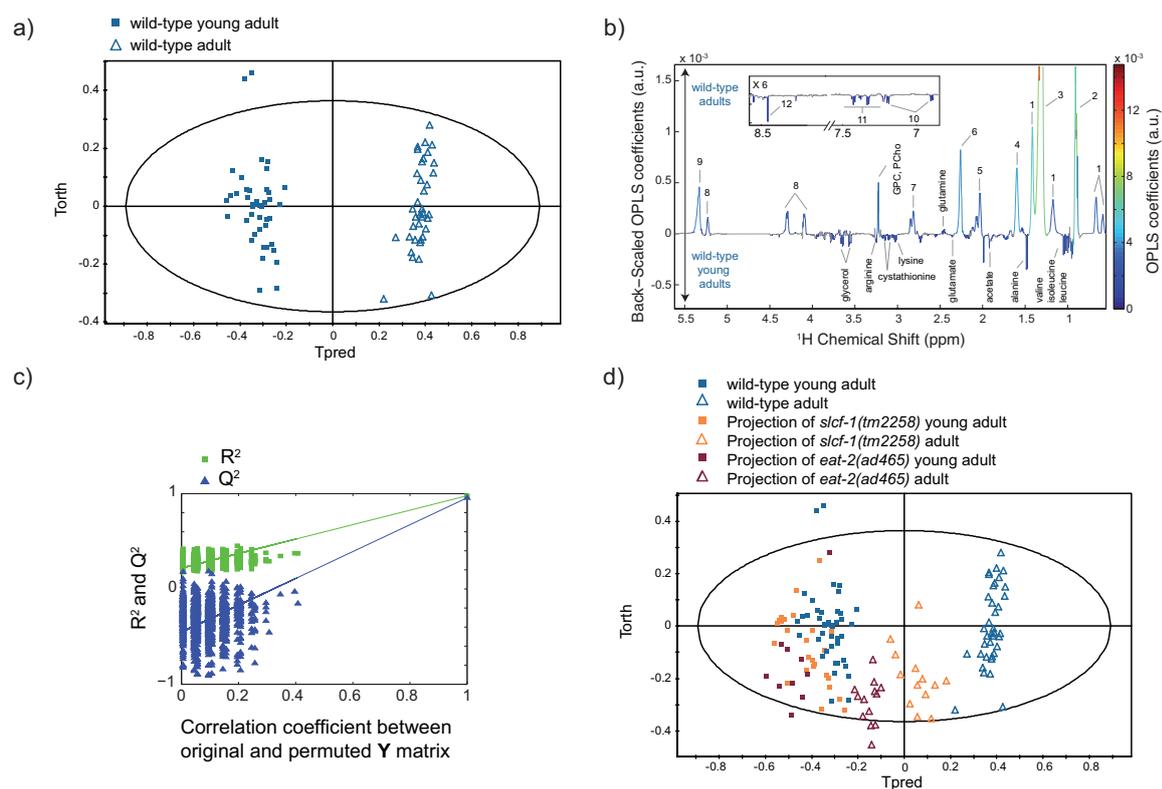


Figure 3.2.3. Metabolic signature of aging in wild-type *C. elegans* worms. OPLS model discriminating wild-type young adults and wild-type adults (1 predictive component and 3 orthogonal components; $R^2X = 0.846$, $R^2Y = 0.978$, $Q^2 = 0.956$) from Pareto-scaled dataset: a) score plot, b) loadings plot resulting from the SRV analysis, showing back-scaled OPLS coefficients values, colored from the original OPLS coefficients if variables were found statistically significant after a multiple testing univariate procedure (Benjaminin-Yekutieli correction) and c) model validation resulting from 1000 permutations, demonstrating the model robustness, as model R^2 and Q^2 values were significantly higher than random model ones. d) Score plot of the projections of *slcf-1(tm2258)* and *eat-2(ad465)* adults and young adults in the OPLS model (A), discriminating wild-type adults and young adults. Keys: 1. cyclic fatty acids, 2. lipids (CH_3), 3. lipids ($(CH_2)_n$), 4. lipids (CH_2CH_2CO), 5. unsaturated lipids ($CH_2CH=CH$), 6. lipids (CH_2CO), 7. unsaturated lipids ($CH=CHCH_2CH=CH$), 8. Glycerol of lipids, 9. unsaturated lipids ($CH=CH$), 10. tyrosine, 11. phenylalanine, 12. formate, PCho: phosphocholine, GPC: glycerophosphocholine.

Table 3.2.1. Goodness-of-parameter values for the different OPLS models built to discriminate WT, *slcf-1* and *eat-2* young adults and adults.

| | Number of orthogonal components | R^2X | R^2Y | Q^2 |
|--------------------------|---------------------------------|--------|--------|-------|
| WT: YA vs A7 | 3 | 0.846 | 0.978 | 0.956 |
| <i>slcf-1</i> : YA vs A7 | 3 | 0.794 | 0.97 | 0.934 |
| <i>eat-2</i> : YA vs A7 | 2 | 0.728 | 0.978 | 0.934 |
| YA: WT vs <i>slcf-1</i> | 5 | 0.827 | 0.987 | 0.96 |
| A7: WT vs <i>slcf-1</i> | 4 | 0.903 | 0.985 | 0.947 |
| YA: WT vs <i>eat-2</i> | 2 | 0.759 | 0.963 | 0.933 |
| A7: WT vs <i>eat-2</i> | 1 | 0.868 | 0.955 | 0.94 |

YA: 1-day old young adult; A7: 7-day old adult.

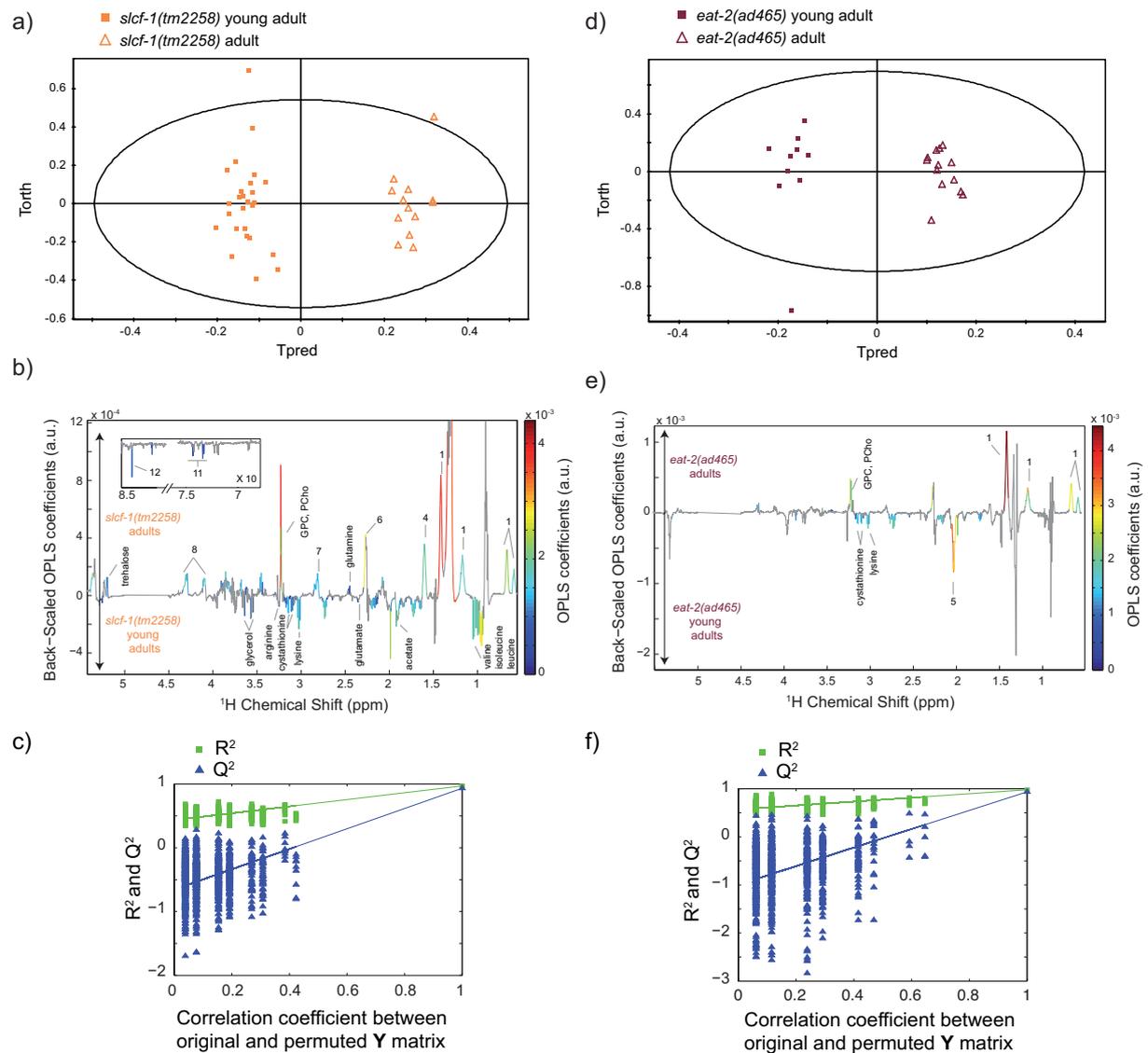


Figure 3.2.4. Metabolic signatures of aging in *slcf-1(tm2258)* and *eat-2(ad465)* *C. elegans* worms. OPLS model discriminating *slcf-1(tm2258)* young adults and *slcf-1(tm2258)* adults (1 predictive component and 3 orthogonal components; $R^2X = 0.794$, $R^2Y = 0.97$, $Q^2 = 0.934$) from Pareto-scaled dataset: a) scores plot, b) loadings plot resulting from the SRV analysis and c) model validation resulting from 1000 permutations, demonstrating the model robustness, as model R^2 and Q^2 values were significantly higher than random model ones. OPLS model discriminating *eat-2(ad465)* young adults and *eat-2(ad465)* adults (1 predictive component and 2 orthogonal components; $R^2X = 0.728$, $R^2Y = 0.978$, $Q^2 = 0.934$) from Pareto-scaled dataset: d) scores plot, e) corresponding loadings plot resulting from the SRV analysis and f) model validation resulting from 1000 permutations, demonstrating the model robustness. Keys: 1. cyclic fatty acids, 4. lipids (CH₂CH₂CO), 5. unsaturated lipids (CH₂CH=CH), 6. lipids (CH₂CO), 7. unsaturated lipids (CH=CHCH₂CH=CH), 8. glyceryl

Higher phosphocholine content is predictive of a short lifespan expectancy. These observations show that DR is associated with a metabolic reprogramming that involves an attenuation of the metabolic variations associated with the physiological aging observed in worms fed *ad libitum*, and that this effect could participate in beneficial effect of DR on lifespan. A mutation that suppresses the extended lifespan phenotype of DR worms should thus affect those metabolite levels in a converse manner. In order to test this hypothesis, we investigated the metabolic pattern of short-lived *daf-18(e1375)* mutants. *daf-18(e1375)* mutation shortens average lifespan by 30% compared to WT while it completely suppresses the extended longevity of *slcf-1(tm2258)* worms by reducing their lifespan by 60%.^[228, 249] Interestingly, phosphocholine levels which are both lower in A7 *eat-2(ad465)* and *slcf-1(tm2258)* mutants compared to WT (**Figure 3.2.2b**), are dramatically increased in *daf-18(e1375)* single mutants and *slcf-1(tm2258);daf-18(e1375)* double mutants (**Figure 3.2.5a**, **Table 3.2.4** and **Table 3.2.5**). It is noteworthy that this observation is not restricted to long-lived DR worms. Indeed Fuchs *et al.* recently reported that long-lived insulin/IGF-1/*daf-2* mutants also harbour lower level of PCho compared to WT, and this level is increased in short-lived FOXO/*daf-16* single or *daf-16; daf-2* double mutants.^[226] Overall, these results strongly suggest that PCho levels reflect the healthspan of worms: while low level is associated with youthfulness, high PCho level is predictive of short life expectancy in both WT animals and different long lived conditions including dietary restriction.

Table 3.2.2. Metabolite variations with age in WT, *slcf-1(tm2258)* and *eat-2(ad465)* .

| Metabolite | Chemical Shift (ppm) and multiplicity | WT: A7 vs YA ^a | P ^b | <i>slcf-1</i> : A7 vs YA ^a | P ^b | <i>eat -2</i> : A7 vs YA ^a | P ^b |
|----------------|---------------------------------------|---------------------------|----------------|---------------------------------------|----------------|---------------------------------------|----------------|
| Alanine | 1.48 (d) | ↓ | 5.82E-07 | ns | 1.00E+00 | ns | 8.42E-01 |
| Betaine | 3.27 (s) | ns | 2.29E-01 | ns | 1.00E+00 | ns | 6.57E-01 |
| Choline | 3.204 (s) | ns | 1.00E+00 | ns | 1.00E+00 | ↑ | 2.92E-02 |
| Cystathionine | 3.145 (m) | ↓ | 2.00E-40 | ↓ | 1.96E-10 | ↓ | 3.07E-06 |
| Formate | 8.46 (s) | ↓ | 1.82E-11 | ↓ | 1.96E-10 | ns | 8.42E-01 |
| GPC | 3.235 (s) | ↑ | 2.28E-28 | ↑ | 5.50E-14 | ↑ | 2.01E-03 |
| Glutamate | 2.355 (m) | ↓ | 3.87E-11 | ↓ | 3.38E-04 | ns | 9.01E-02 |
| Glutamine | 2.465 (m) | ↑ | 1.44E-07 | ↑ | 4.89E-03 | ns | 1.00E+00 |
| Isoleucine | 1.015 (d) | ↓ | 1.12E-24 | ↓ | 5.08E-11 | ns | 7.38E-01 |
| Leucine | 0.96 (d) | ↓ | 6.99E-18 | ↓ | 5.08E-11 | ↓ | 2.92E-02 |
| Lysine | 3.025 (t) | ↓ | 1.22E-30 | ↓ | 6.33E-09 | ↓ | 6.25E-04 |
| Phenylalanine | 7.43 (t), | ↓ | 5.83E-09 | ↓ | 5.41E-03 | ns | 1.00E+00 |
| Phosphocholine | 3.225 (s) | ↑ | 1.85E-28 | ↑ | 1.77E-08 | ↑ | 8.65E-04 |
| Succinate | 2.41 (s) | ns | 1.00E+00 | ns | 3.03E-01 | ns | 1.00E+00 |
| Tyrosine | 7.20 (d) | ↓ | 2.11E-08 | ns | 2.00E-01 | ns | 1.00E+00 |
| Trehalose | 5.20 (d) | ns | 4.13E-01 | ↑ | 8.20E-05 | ↑ | 2.92E-02 |
| Valine | 1.045 (d) | ↓ | 8.47E-23 | ↓ | 1.01E-09 | ns | 8.42E-01 |
| Arginine | 3.246 (t) | ↓ | 2.46E-20 | ↓ | 4.56E-02 | ns | 1.80E-01 |
| Acetate | 1.92 (s) | ↓ | 2.39E-09 | ↓ | 2.50E-04 | ns | 5.33E-01 |
| Lactate | 1.33 (d) | ↓ | 4.86E-13 | ns | 6.69E-01 | ns | 2.14E-01 |
| Glycerol | 3.56 (m) | ↓ | 1.47E-13 | ↓ | 2.25E-06 | ns | 1.00E+00 |
| Glycine | 3.56 (s) | ↑ | 6.39E-18 | ↑ | 2.96E-07 | ns | 1.80E-01 |

^a↑: increase in metabolite concentration with age; ↓: decrease in metabolite concentration with age; ns: non significant metabolite variation. ^b P-values obtained an unpaired two-tailed t-test, corrected with the Benjamini-Yekutieli method for multiple testing. P-values < 0.05 are significant. YA: young adult; A7: adult.

Table 3.2.3. Metabolite variations between WT and long-lived mutants (*slcf-1(tm2258)* or *eat-2(ad465)*) in young adults and adults.

| Metabolite | Chemical Shift (ppm) and multiplicity | YA: <i>slcf-1</i> vs WT ^a | P ^b | A7: <i>slcf-1</i> vs WT ^a | P ^b | YA: <i>eat-2</i> vs WT ^a | P ^b | A7: <i>eat-2</i> vs WT ^a | P ^b |
|----------------|---------------------------------------|--------------------------------------|----------------|--------------------------------------|----------------|-------------------------------------|----------------|-------------------------------------|----------------|
| Alanine | 1.48 (d) | ns | 1.68E-01 | ↑ | 2.27E-06 | ns | 4.34E-01 | ns | 1.00E+00 |
| Betaine | 3.27 (s) | ↓ | 5.05E-04 | ns | 1.00E+00 | ns | 1.00E+00 | ns | 4.67E-01 |
| Choline | 3.204 (s) | ↑ | 3.65E-03 | ↑ | 7.79E-03 | ns | 2.05E-01 | ns | 1.29E-01 |
| Cystathionine | 3.145 (m) | ↑ | 5.85E-03 | ↑ | 8.79E-07 | ↑ | 2.67E-11 | ↑ | 2.13E-20 |
| Formate | 8.46 (s) | ns | 1.00E+00 | ns | 6.98E-02 | ↓ | 4.01E-10 | ↓ | 4.59E-06 |
| GPC | 3.235 (s) | ns | 1.00E+00 | ns | 3.91E-01 | ↓ | 1.09E-04 | ↓ | 1.22E-12 |
| Glutamate | 2.355 (m) | ns | 1.00E+00 | ns | 8.72E-02 | ↑ | 5.55E-12 | ↑ | 3.78E-15 |
| Glutamine | 2.465 (m) | ns | 8.03E-01 | ns | 1.00E+00 | ↑ | 3.75E-11 | ↑ | 2.38E-08 |
| Isoleucine | 1.015 (d) | ↓ | 3.26E-03 | ns | 1.00E+00 | ↓ | 1.76E-10 | ns | 1.00E+00 |
| Leucine | 0.96 (d) | ns | 2.81E-01 | ↓ | 2.92E-04 | ↓ | 4.38E-10 | ↓ | 8.07E-08 |
| Lysine | 3.025 (t) | ↑ | 6.47E-14 | ↑ | 3.08E-19 | ↑ | 2.04E-06 | ↑ | 2.13E-20 |
| Phenylalanine | 7.43 (t), | ns | 1.68E-01 | ↑ | 7.79E-03 | ↓ | 3.07E-05 | ns | 9.41E-01 |
| Phosphocholine | 3.225 (s) | ↓ | 3.65E-03 | ↓ | 1.68E-07 | ↓ | 1.94E-05 | ↓ | 1.78E-14 |
| Succinate | 2.41 (s) | ↑ | 2.29E-03 | ns | 1.00E+00 | ↑ | 3.18E-10 | ↑ | 9.08E-08 |
| Tyrosine | 7.20 (d) | ns | 1.00E+00 | ns | 9.81E-02 | ↓ | 2.01E-05 | ns | 1.00E+00 |
| Trehalose | 5.20 (d) | ↓ | 4.74E-17 | ↓ | 2.17E-03 | ↓ | 3.55E-11 | ↓ | 5.29E-10 |
| Valine | 1.045 (d) | ↓ | 1.42E-02 | ns | 1.00E+00 | ↓ | 5.55E-12 | ns | 4.42E-01 |
| Arginine | 3.246 (t) | ns | 1.60E-01 | ↑ | 2.27E-06 | ↓ | 1.21E-03 | ↑ | 2.50E-10 |
| Acetate | 1.92 (s) | ns | 1.00E+00 | ns | 1.00E+00 | ↓ | 2.78E-10 | ns | 9.41E-01 |
| Lactate | 1.33 (d) | ↓ | 7.50E-05 | ns | 1.00E+00 | ↓ | 2.61E-10 | ns | 6.06E-01 |
| Glycerol | 3.56 (m) | ns | 6.74E-02 | ns | 1.00E+00 | ↓ | 2.32E-05 | ns | 1.00E+00 |
| Glycine | 3.56 (s) | ns | 8.80E-01 | ns | 1.00E+00 | ↑ | 8.21E-05 | ns | 1.00E+00 |

^a↑: increase in metabolite concentration in long-lived mutant (*slcf-1* or *eat-2*) by comparison to WT; ↓: decrease in metabolite concentration in long-lived mutant (*slcf-1* or *eat-2*) by comparison to WT; ns: non significant metabolite variation. ^bP-values obtained an unpaired two-tailed t-test, corrected with the Benjamini-Yekutieli method for multiple testing. P-values < 0.05 are significant. YA: young adult; A7: adult.

Choline kinase expression correlated with physiological age. In mammals, higher PCho content is also a hallmark of several cancers and its increase has been associated with choline kinase overexpression in tumour cells where the tumour suppressor PTEN pathway is inactivated.^[250] PTEN is the ortholog of *daf-18* and regulates longevity in worms and mammals.^[249, 251, 252] Interestingly, genes overexpressed in *daf-18(e1375)* mutants (data not shown) include *ckb-2* which encodes a choline kinase.^[253] In order to test the hypothesis that variations in PCho levels may be linked to transcriptional regulation of the *ckb-2* gene, we quantified *ckb-2* transcripts in WT and lifespan mutants at different ages. Indeed, *ckb-2* transcripts levels correlate with PCho content in worms for all genotypes and ages, further supporting the idea that choline pathway activation may reflect the physiological age of the worms (**Figure 3.2.5b**).

Table 3.2.4. Metabolite variations with age in *daf-18(e1375)* mutants and *slcf-1(tm2258);daf-18(e1375)* double mutants.

| Metabolite | Chemical Shift (ppm) and multiplicity | <i>daf-18</i> : A7 vs YA ^a | P ^b | <i>slcf-1;daf-18</i> : A7 vs YA ^a | P ^b |
|----------------|---------------------------------------|---------------------------------------|----------------|--|----------------|
| Alanine | 1.48 (d) | ↓ | 1.92E-08 | ↓ | 1.03E-06 |
| Betaine | 3.27 (s) | ↓ | 9.63E-03 | ↑ | 1.44E-04 |
| Choline | 3.204 (s) | ↑ | 6.74E-05 | ↑ | 1.82E-06 |
| Cystathionine | 3.145 (m) | ↓ | 9.63E-03 | ↓ | 4.37E-05 |
| Formate | 8.46 (s) | ↓ | 3.92E-10 | ↓ | 4.29E-08 |
| GPC | 3.235 (s) | ↑ | 2.66E-08 | ↑ | 1.12E-11 |
| Glutamate | 2.355 (m) | ns | 9.16E-01 | ns | 3.40E-01 |
| Glutamine | 2.465 (m) | ↑ | 1.48E-07 | ↑ | 2.33E-08 |
| Isoleucine | 1.015 (d) | ↓ | 1.33E-13 | ↓ | 3.54E-11 |
| Leucine | 0.96 (d) | ↓ | 8.53E-11 | ↓ | 3.10E-12 |
| Lysine | 3.025 (t) | ↑ | 3.34E-02 | ↓ | 1.82E-06 |
| Phenylalanine | 7.43 (t), | ↓ | 1.37E-07 | ↓ | 1.88E-07 |
| Phosphocholine | 3.225 (s) | ↑ | 4.24E-06 | ↑ | 4.50E-24 |
| Succinate | 2.41 (s) | ↓ | 1.69E-06 | ↑ | 1.19E-13 |
| Tyrosine | 7.20 (d) | ns | 6.12E-01 | ns | 1.00E+00 |
| Trehalose | 5.20 (d) | ↓ | 1.09E-04 | ↓ | 1.38E-06 |
| Valine | 1.045 (d) | ↓ | 1.01E-12 | ↓ | 1.09E-10 |
| Arginine | 3.246 (t) | ns | 1.00E+00 | ↓ | 5.51E-12 |
| Acetate | 1.92 (s) | ↓ | 2.06E-07 | ↓ | 8.64E-07 |
| Lactate | 1.33 (d) | ↓ | 3.22E-05 | ↓ | 4.88E-16 |
| Glycerol | 3.56 (m) | ↓ | 2.06E-07 | ↓ | 1.63E-08 |
| Glycine | 3.56 (s) | ns | 6.91E-01 | ns | 1.28E-01 |

^a↑: increase in metabolite concentration with age; ↓: decrease in metabolite concentration with age; ns: non significant metabolite variation. ^b P-values obtained with an unpaired two-tailed t-test, corrected with the Benjamini-Yekutieli method for multiple testing. P-values < 0.05 are significant. YA: young adult; A7: adult. Acetate, lactate, glycerol and glycine p-values are not reliable due to signal overlaps.

Table 3.2.5. Metabolite variations between WT and *daf-18(e1375)* and between *slcf-1(tm2258)* and *slcf-1(tm2258);daf-18(e1375)* in young adults and adults.

| Metabolite | YA: <i>daf-18</i> ^a vs WT ^a | P ^b | A7: <i>daf-18</i> vs WT ^a | P ^b | YA: <i>slcf-1</i> ; <i>daf-18</i> vs <i>slcf-1</i> ^a | P ^b | A7: <i>slcf-1</i> ; <i>daf-18</i> vs <i>slcf-1</i> ^a | P ^b |
|----------------|---|----------------|--|----------------|--|----------------|--|----------------|
| Alanine | ↑ | 2.21E-08 | ↓ | 4.27E-03 | ns | 1.00E+00 | ↓ | 1.61E-11 |
| Betaine | ns | 4.09E-01 | ns | 1.00E+00 | ns | 1.00E+00 | ↑ | 1.35E-02 |
| Choline | ns | 1.00E+00 | ↑ | 4.57E-08 | ns | 7.20E-01 | ↑ | 1.54E-06 |
| Cystathionine | ns | 1.00E+00 | ↑ | 7.29E-14 | ns | 1.00E+00 | ↑ | 2.93E-14 |
| Formate | ns | 1.00E+00 | ↓ | 3.30E-07 | ns | 2.03E-01 | ↓ | 9.55E-08 |
| GPC | ns | 1.00E+00 | ns | 4.95E-01 | ns | 5.19E-01 | ↓ | 3.03E-07 |
| Glutamate | ns | 1.00E+00 | ↑ | 2.00E-05 | ↓ | 3.48E-04 | ns | 5.72E-01 |
| Glutamine | ↓ | 2.20E-02 | ↑ | 1.25E-03 | ns | 5.19E-01 | ↑ | 4.06E-03 |
| Isoleucine | ↑ | 3.26E-02 | ↑ | 4.74E-02 | ns | 5.19E-01 | ns | 1.38E-01 |
| Leucine | ↑ | 1.72E-03 | ↑ | 2.56E-03 | ↑ | 3.66E-02 | ↑ | 6.65E-05 |
| Lysine | ↑ | 2.54E-03 | ↑ | 9.91E-11 | ns | 1.00E+00 | ns | 1.00E+00 |
| Phenylalanine | ↑ | 4.20E-03 | ↑ | 2.00E-05 | ns | 8.54E-01 | ns | 6.42E-01 |
| Phosphocholine | ns | 5.39E-01 | ↑ | 3.30E-07 | ns | 1.00E+00 | ↑ | 3.93E-26 |
| Succinate | ns | 5.11E-02 | ↑ | 1.22E-13 | ns | 1.21E-01 | ↑ | 3.11E-14 |
| Tyrosine | ns | 4.75E-01 | ↑ | 2.46E-08 | ns | 2.03E-01 | ↑ | 2.24E-06 |
| Trehalose | ↓ | 2.96E-04 | ↓ | 1.54E-09 | ns | 8.83E-01 | ↓ | 4.58E-05 |
| Valine | ↑ | 4.94E-03 | ns | 1.00E+00 | ns | 1.00E+00 | ns | 1.00E+00 |
| Arginine | ↑ | 1.53E-04 | ↑ | 4.01E-06 | ↑ | 3.66E-02 | ↓ | 1.21E-06 |
| Acetate | ↑ | 1.54E-04 | ↑ | 1.37E-07 | ns | 1.00E+00 | ns | 1.00E+00 |
| Lactate | ns | 1.00E+00 | ↓ | 3.07E-19 | ↑ | 2.19E-02 | ↓ | 3.51E-17 |
| Glycerol | ↓ | 1.96E-03 | ↓ | 4.52E-05 | ns | 5.19E-01 | ↓ | 2.34E-04 |
| Glycine | ↑ | 3.79E-02 | ↑ | 1.85E-02 | ns | 1.00E+00 | ns | 1.00E+00 |

^a↑: increase in metabolite concentration in *daf-18* or *slcf-1;daf-18* by comparison to WT or *slcf-1*; ↓: decrease in metabolite concentration in *daf-18* or *slcf-1;daf-18* by comparison to WT or *slcf-1*; ns: non significant metabolite variation. ^bP-values obtained with an unpaired two-tailed t-test, corrected with the Benjamini-Yekutieli method for multiple testing. P-values < 0.05 are significant. YA: young adult; A7: adult. Acetate, lactate, glycerol and glycine p-values are not reliable due to signal overlaps.

These results thus suggest that the regulation of PCho pathway by DAF-18/PTEN is conserved across species. Furthermore DR not only extends lifespan but also prevents cancer.^[241] Our data therefore underline a new mechanism by which DR may counteract tumour development by preventing the up-regulation of choline kinase expression.

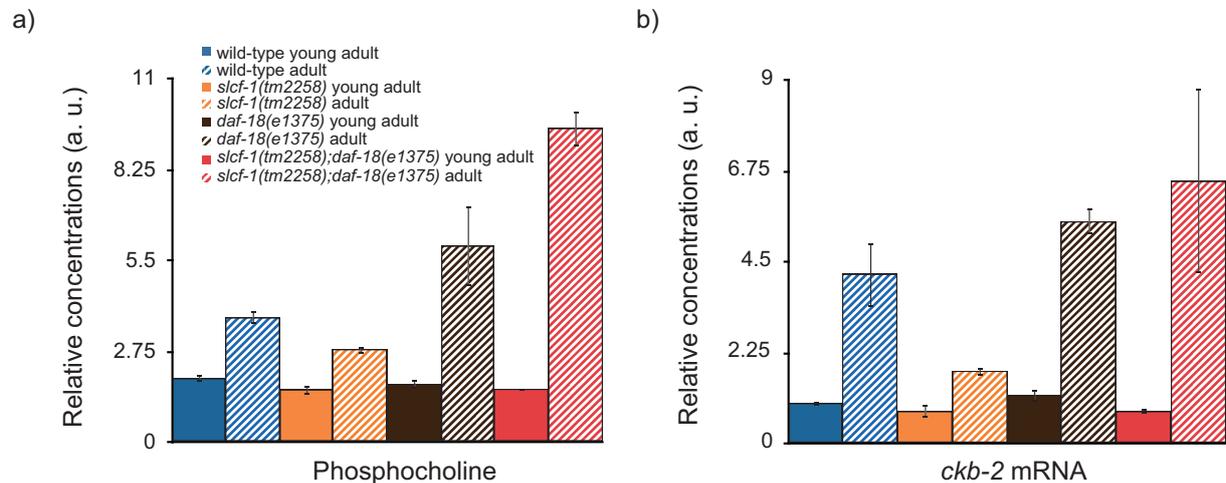


Figure 3.2.5. Activation of the phosphocholine pathway with aging; a) Relative concentrations in arbitrary units of phosphocholine in young and 7 day-old adults WT, *slcf-1(tm2258)*, *daf-18(e1375)* and *slcf-1(tm2258);daf-18(e1375)* double mutants. Results are reported with means and 95% confidence intervals. b) Relative concentrations in arbitrary units of *ckb-2* mRNA in young and 7 day-old adults WT, *slcf-1(tm2258)*, *daf-18(e1375)* and *slcf-1(tm2258);daf-18(e1375)* double mutants. Results are reported with means and standard deviations.

Recent studies also reported the modification with age of PCho levels and/ or other choline metabolites, among many other metabolites, in different species including human.^[167, 243-245, 254] Our data thus establish *C. elegans* as a relevant model to investigate the regulation of choline metabolism in aging and its relationship with age associated diseases.

3.2.2 Metabolic phenotyping of *ahr-1(ia03)* *C. elegans* mutant

3.2.2.1 Introduction

Toxicity and carcinogenic effects of a wide variety of environmental contaminants such as dioxin (TCDD; 2,3,7,8-tetrachlorodibenzo-[p]-dioxin) are mediated in a large extent by the Aryl hydrocarbon (dioxin) receptor (AhR), a ligand-activated transcription factor.^[255] Beyond activation with xenobiotics, studies have shown that AhR possesses an independent endogenous cellular function, involved in cell proliferation, cell differentiation and cell migration as well as carcinogenesis. However, the molecular bases of the AhR endogenous function have not been yet completely characterized and understood.^[255]

In the model organism *Caenorhabditis elegans*, an AhR ortholog termed AHR-1 has been identified.^[256] This receptor is not activated by dioxin or related chemical compounds.^[257] The endogenous AHR-1 function, evaluated with the characterization of loss-of-function in *C. elegans* mutants plays at least a role in regulation of neural development,^[257] or fatty acid metabolism.^[258] Furthermore, a transcriptomic analysis of AHR-1 mutants has characterized expression changes of genes involved in development, growth and metabolism.^[258]

Metabolomics approaches have focused mainly on the effects of dioxin-related compounds on metabolism in different systems such as HepG2 liver cells,^[259] mice^[260] or rats^[261, 262] in the context of toxicological studies. However, endogenous AhR function on global metabolism has never been investigated with untargeted metabolomic techniques.

In this study, we applied whole-organism HR-MAS NMR spectroscopy to characterize metabolic changes induced by *ahr-1* *C. elegans* mutants and get insight into the effects of the endogenous AHR-1 function on worm metabolism. We thus compared the metabolic profiles of control wild-type worms (N2) and *ahr-1(ia03)* *C. elegans* mutants where the mutation causes a loss of function for the AHR-1 receptor, with or without AHR-1 coupling with a GFP reporter (WT-GFP and KO-GFP).

3.2.2.2 Material and Methods

Sample preparation (Bui Linh-Chi, Université Paris Descartes). Worms were grown according to standard protocol except for a liquid step in worm amplification. Worms were fixed for 45 minutes in 1% paraformaldehyde and then washed 5 times in distilled water followed by 5 washes in deuterium oxide. 30 ml disposable Kel-f inserts with sealing caps for 4 mm NMR rotors were filled with around 1000 whole worms and frozen at -80°C for storage

until NMR analysis. Samples were thawed at room temperature 15 minutes before the NMR experiments.

Whole *C. elegans* HR-MAS NMR spectroscopy. *C. elegans* HR-MAS NMR spectroscopy was performed as previously described by Blaise *et al.*^[27, 44] All experiments were carried out on a Bruker Avance II spectrometer, operating at 700 MHz (proton resonance frequency), equipped with a 4 mm HR-MAS double resonance (¹H-¹³C) probe. Temperature was controlled at 295 K throughout the experiments and magic angle spinning speed set to 3.5 KHz. A set of 1D ¹H NMR experiments with water presaturation, including 1D ¹H NOESY and 1D ¹H CPMG were performed on each sample to derive metabolic profiles. 256 free induction decays (FIDs) were co-added, with a 12 ppm spectral width and an acquisition time of 1.4 s, corresponding to 23568 data-points, with a relaxation delay of 1.7 s, for a total experimental time of 13 minutes per spectrum. The NOESY mixing time was set to 100 ms and the CPMG spin-echo delay adjusted to 300 μs for each of the 80 spin-echo loops (48 ms total echo time). The ¹H 90° hard pulse length was calibrated at 6.2 μs. 2D NMR experiments, including ¹H-¹H TOCSY and ¹H-¹³C HSQC experiments, were carried out on a subset of selected samples to characterize structural connectivities between nuclei and refine metabolite identification.

NMR data processing. All FIDs were multiplied by an exponential function corresponding to a 0.3 Hz line-broadening factor prior to Fourier transform. Phasing and baseline correction were performed manually in Topspin 2.1 (Bruker, GmbH, Rheinstetten, Germany). Spectra were automatically calibrated on the CH₃ alanine doublet at δ = 1.48 ppm. Spectra were then imported in the AMIX software (Bruker, GmbH, Rheinstetten, Germany) and reduced over the chemical range of 0.55-8.75 ppm to 8200 10⁻³ ppm wide bins with integration of signal intensity. Residual water signal (δ = 4.45-5.1 ppm) as well as the DMSO signal (δ = 2.71-2.75 ppm) and the methanol signal (δ = 3.35-3.36 ppm) were discarded prior to analysis. Spectra were normalized using the probabilistic quotient normalization (PQN) approach,^[97] with a median of all spectra as reference spectrum and mean-centred. Metabolite assignment was completed exploiting reference data from the literature,^[27, 226] the HMDB,^[50] MMCD,^[49] bbiorefcodes-2-0-0 (Bruker, GmbH, Rheinstetten, Germany) and Chenomx NMR Suite 7.0 (Chenomx Inc, Edmonton, Canada) spectral databases.

NMR data analysis. Principal component analysis (PCA)^[187] was first conducted in SIMCA P12+ (Umetrics, Umea, Sweden) to derive the main sources of variance within the dataset, assess sample homogeneity and exclude biological or technical outliers. Orthogonal

projection to latent structure discriminant analysis (OPLS-DA) was then performed in MATLAB (The MathWorks Inc., Natick, MA) to derive pair-wise comparison between the different conditions (strains and ages).^[100] Metabolites involved in class discrimination are then derived from an univariate approach based on the statistical recoupling of variables (SRV).^[94] Univariate analysis is performed using the non-parametric two-sided rank sum test calculated for each recoupled clusters of variables. Variable statistical significance threshold was evaluated from the Benjamini-Hochberg correction for multiple testing.^[102]

3.2.2.3 Results and Discussion

The *ahr-1(ia03)* mutation carried by the *ahr-1(ia03)* *C. elegans* mutants corresponds to a 1517-bp deletion in the *ahr-1* gene inducing a strong loss-of-function for that gene.^[257] We recorded HR-MAS NMR metabolic signatures for wild-type (N2) worms, *ahr-1(ia03)* mutants and their counterparts which possessed a GFP reporter, WT-GFP and KO-GFP. A typical 700 MHz ¹H NOESY NMR *C. elegans* spectra displayed broad signals from lipids on which narrower signals from small mobile molecules such as amino acids, organic acids, choline derivatives or carbohydrates were superimposed. We detected in the different strains of this study some metabolites such as allantoin or β -alanine, which were not observed in the previous study on the effects of aging on *C. elegans* metabolism (see Part 1.2.1). These differences could come from the sample preparation protocols, which were slightly different with an amplification step in liquid phase for this study on AHR-1, but also from *C. elegans* strains themselves, which were probably not genetically identical, even for N2 wild-type worms, as their origins were different. Above all, this observation highlights the necessity to include adequate controls within each study to get the most reliable comparisons between strains.

Effects of AHR-1 GFP-reporter. We investigated the effects of the GFP reporter for AHR-1 on worm metabolism by comparing N2 and WT-GFP in an OPLS model built with 7 orthogonal components, and associated with high goodness-of-fit-parameter values ($R^2X = 0.933$, $Q^2 = 0.955$, **Figure 3.2.6**). We observed a robust discrimination between the two strains. Coupling with GFP is associated with an increase in phenylalanine, tyrosine, asparagine, β -alanine and glutamate and a decrease in phosphocholine.

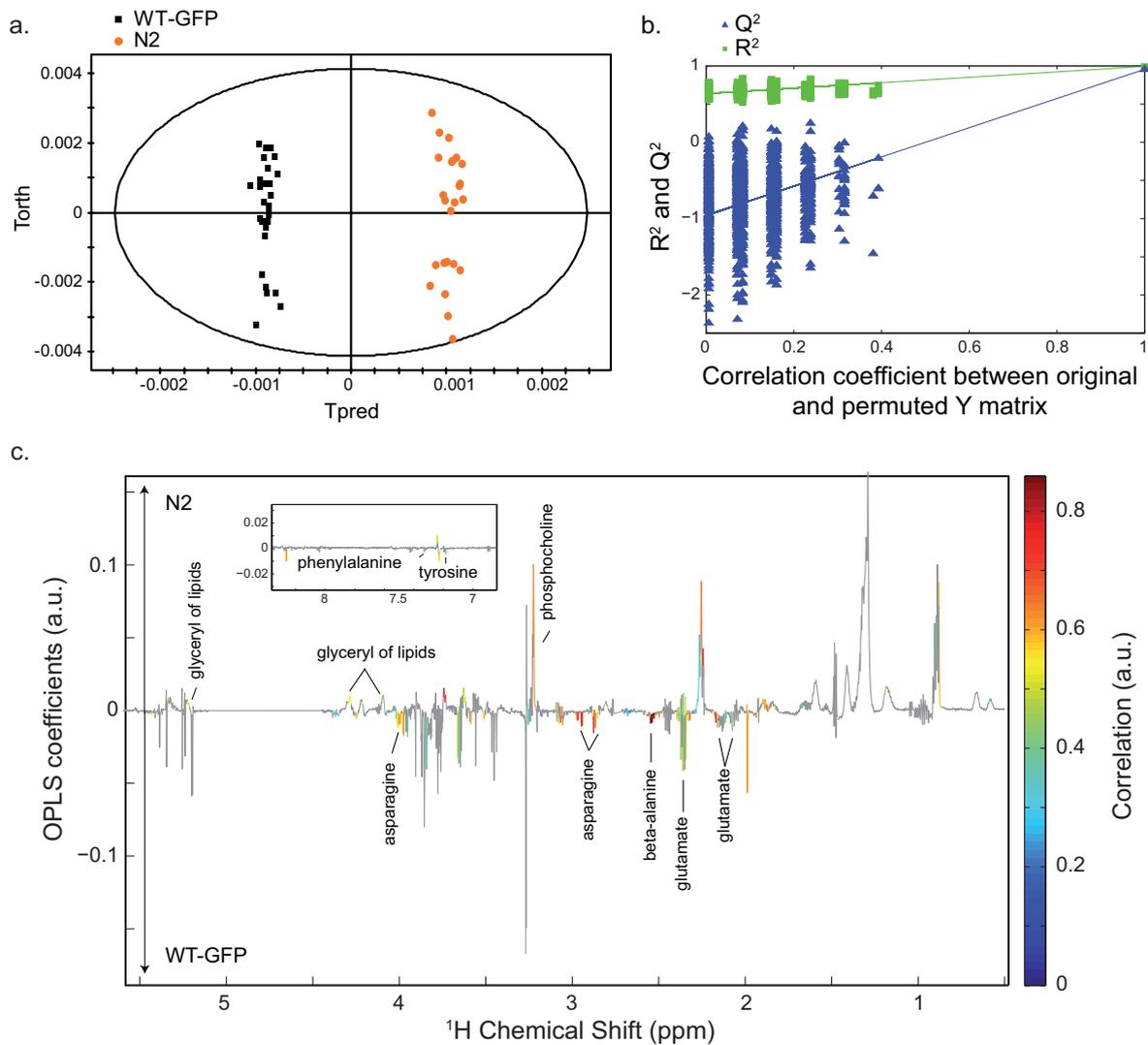


Figure 3.2.6. OPLS model discriminating N2 worms and WT-GFP worms (7 orthogonal components; $R^2X = 0.933$, $Q^2 = 0.955$): a) score plot, b) validation results with 1000 permutations and c) corresponding loadings plot resulting from the SRV analysis, showing OPLS coefficients values, colored from correlation values if variables were found statistically significant after an univariate procedure corrected for multiple testing (Benjaminin-Hochberg correction).

We performed the same analysis in the *ahr-1* mutants. An OPLS model was built with 3 orthogonal components, associated with high goodness-of-fit-parameters ($R^2X = 0.875$, $Q^2 = 0.853$, **Figure 3.2.7**). Worms with GFP reporter for AHR-1 are associated with an increase in phenylalanine, tyrosine, cystathionine, lysine, asparagine, β -alanine, glutamine and leucine and a decrease in phosphocholine and glycerophosphocholine.

These results showing that coupling AHR-1 with GFP has an effect on the worm metabolism are quite unexpected and suggest that AHR-1 function is probably altered when coupled with GFP. For *ahr-1* mutants, we could interpret this data as a modification of the residual AHR-1 function in response to coupling with GFP.

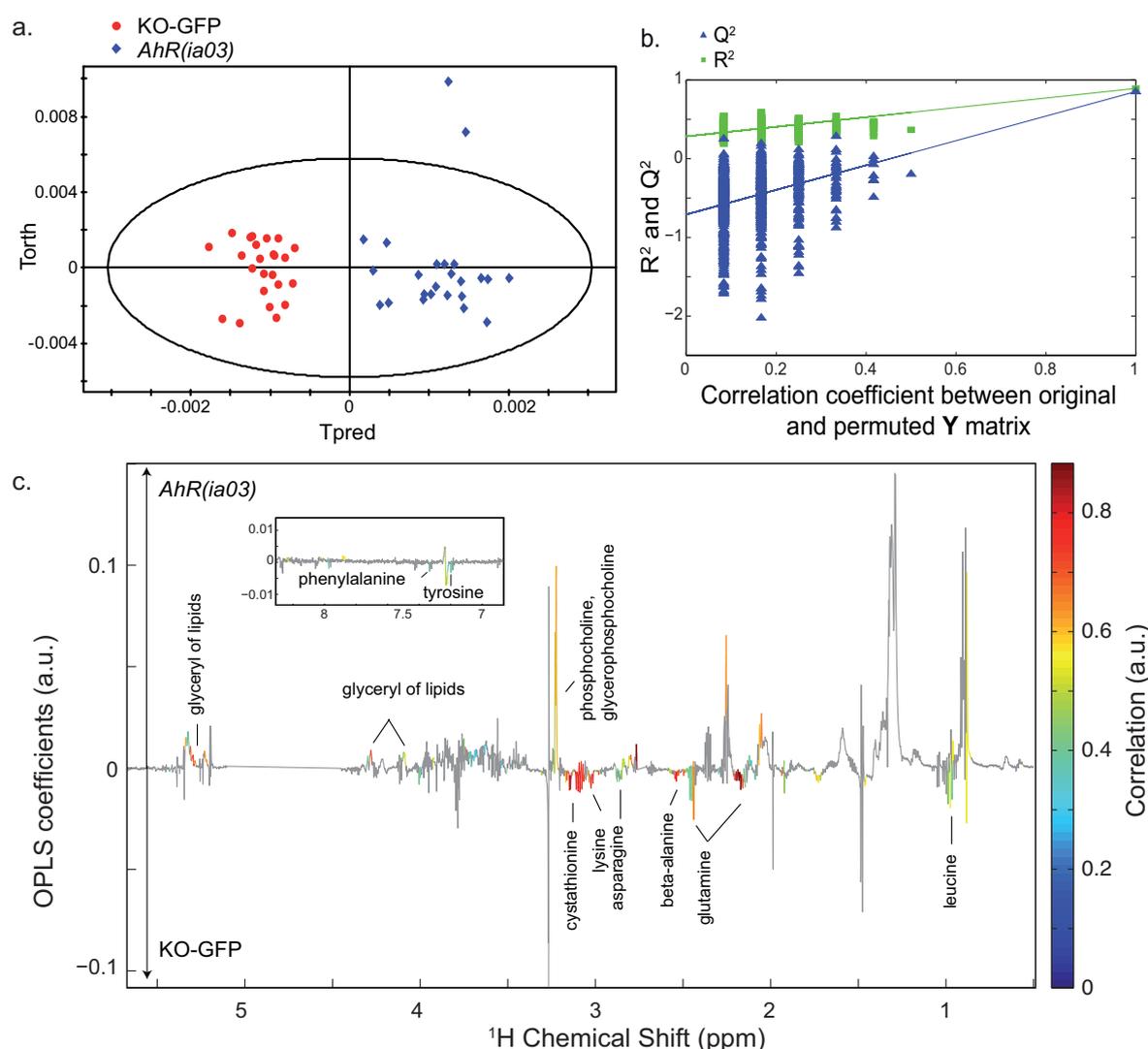


Figure 3.2.7. OPLS model discriminating KO-GFP mutants and *AhR(ia03)* mutants (3 orthogonal components; $R^2X = 0.875$, $Q^2 = 0.853$): a) score plot, b) validation results with 1000 permutations and c) corresponding loadings plot resulting from the SRV analysis, showing OPLS coefficients values, colored from correlation values if variables were found statistically significant after an univariate procedure corrected for multiple testing (Benjaminin-Hochberg correction).

The inactivation of the AHR-1 receptor has an impact on *C. elegans* metabolism. To evaluate the effects of the AHR-1 receptor inactivation on *C. elegans* metabolism, we compared the metabolic profiles of wild-type N2 control worms and *ahr-1(ia03)* mutants. We built an OPLS model with 3 orthogonal components, associated with high goodness-of-fit-parameter values, showing a robust discrimination between the two worm strains involving a wide range of metabolites ($R^2X = 0.903$, $Q^2 = 0.93$, **Figure 3.2.8**). Compared to N2 worms, *ahr(ia03)* mutants are associated with high levels of tyrosine, glycerol, cystathionine, lysine, asparagine, β -alanine, valine, leucine and isoleucine, and low levels of allantoin, trehalose, phosphocholine, glycerophosphocholine, cyclic and oxidized fatty acids.

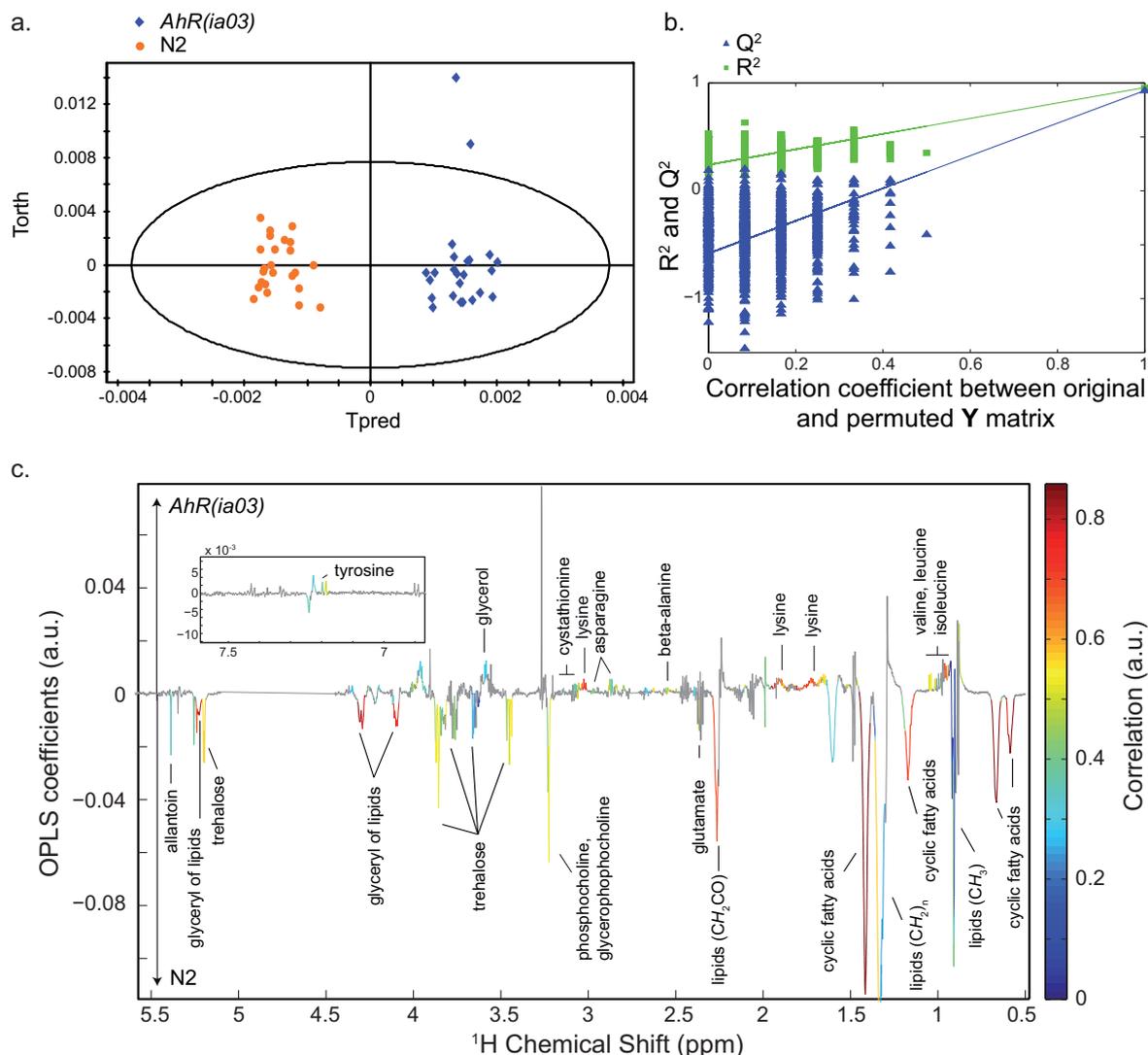


Figure 3.2.8. OPLS model discriminating N2 worms and AhR(ia03) mutants (3 orthogonal components; $R^2X = 0.903$, $Q^2 = 0.93$): a) score plot, b) validation results with 1000 permutations and c) corresponding loadings plot resulting from the SRV analysis, showing OPLS coefficients values, colored from correlation values if variables were found statistically significant after an univariate procedure corrected for multiple testing (Benjaminin-Hochberg correction).

KO-GFP and WT-GFP worms are also metabolically distinct. To confirm the previous metabolic signature of AHR-1 inactivation in *C. elegans*, we compared the metabolic profiles of WT-GFP and KO-GFP worms in an OPLS model, built with 2 orthogonal components and associated with high goodness-of-fit-parameter values ($R^2X = 0.783$, $Q^2 = 0.973$, **Figure 3.2.9**). We observed a robust discrimination between the two strains where KO-GFP worms are associated with high levels of phenylalanine, tyrosine, cystathionine, lysine, β -alanine, glutamine, valine, leucine and isoleucine, and low levels of allantoin, trehalose, phosphocholine and glycerophosphocholine, glutamate, cyclic and oxidized fatty acids by comparison to WT-GFP.

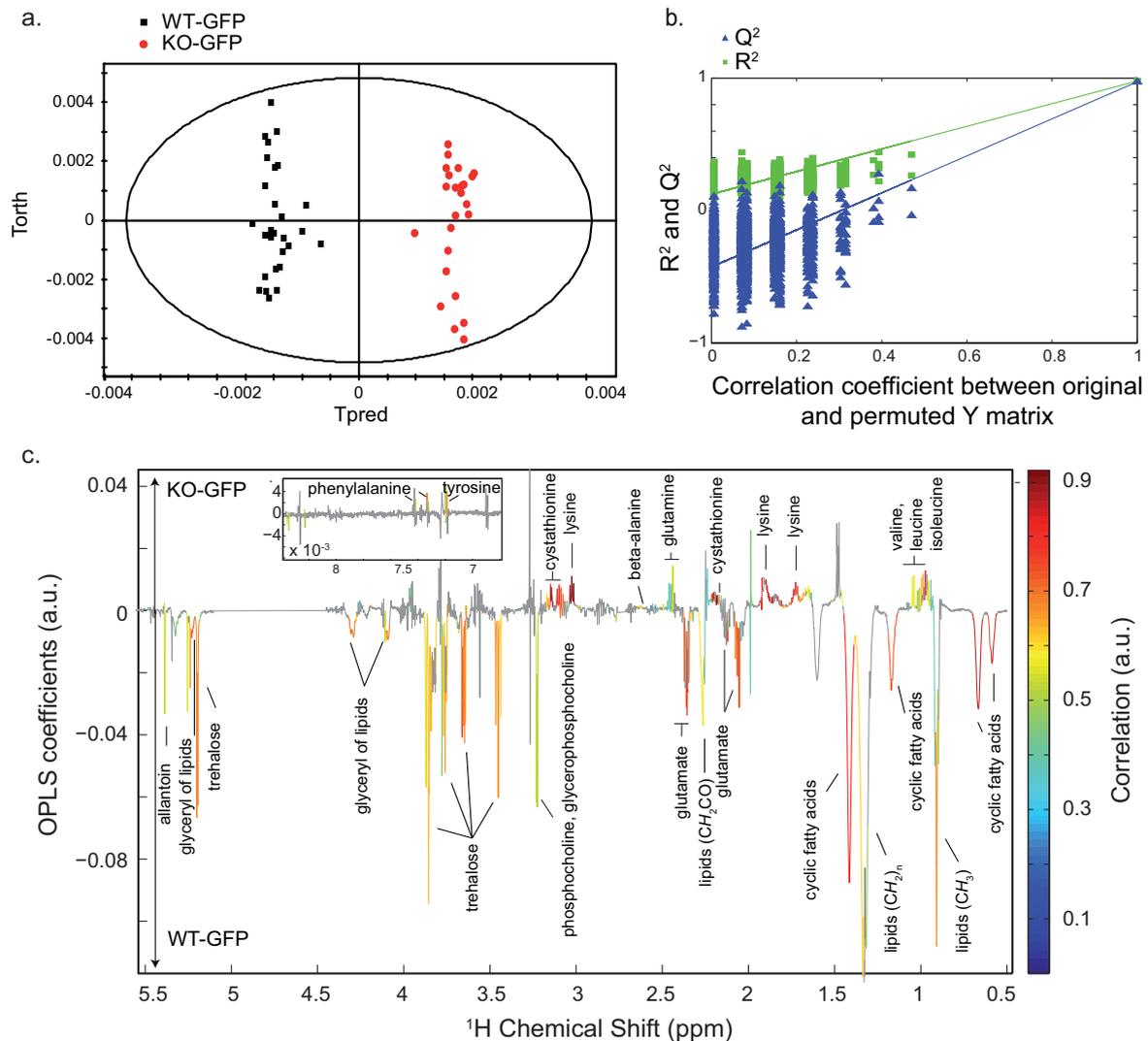


Figure 3.2.9. OPLS model discriminating WT-GFP worms and KO-GFP mutants (2 orthogonal components; $R^2X = 0.783$, $Q^2 = 0.973$): a) score plot, b) validation results with 1000 permutations and c) corresponding loadings plot resulting from the SRV analysis, showing OPLS coefficients values, colored from correlation values if variables were found statistically significant after an univariate procedure corrected for multiple testing (Benjaminin-Hochberg correction).

Specific metabolic signature of AHR-1 inactivation. The metabolic signatures observed in response to AHR-1 inactivation are very consistent between the conditions with or without coupling with GFP. We were thus able to define a robust specific metabolic signature of AHR-1 inactivation as the intersection of these two metabolic signatures. The AHR-1 inactivation in *C. elegans* thus induces a decrease in levels of allantoin, trehalose, phosphocholine, glycerophosphocholine, glutamate, cyclic and oxidized fatty acids and an increase in levels of tyrosine, cystathionine, lysine, β -alanine, valine, leucine and isoleucine.

The AHR-1 inactivation induces thus important changes in amino-acid and carbohydrate metabolism of *C. elegans*. This observation is consistent with a transcriptomic

analysis performed in *ahr-1 C. elegans* mutants, where many genes involved in these metabolic pathways were over or under-expressed.^[258] Furthermore, fatty acid metabolism appears regulated by AHR-1, as total fatty acid levels are decreased in *ahr-1* mutants. Links between AHR-1 and fatty acid biosynthesis have already been suggested, in particular through the changes of the length of fatty acid methyl esters in response to AHR-1 inactivation.^[258]

Curiously, the metabolic signature of *ahr-1* mutants shares large similarities with the metabolic signature of physiological aging determined in the previous study (see Part 1.2.1). Indeed, *ahr-1* mutants are characterized by high levels of valine, isoleucine, leucine, lysine, cystathionine and tyrosine and low levels of oxidized and cyclic fatty acids, glycerophosphocholine and phosphocholine as young adult wild-type worms compared to adult wild-type worms. Our results could thus suggest an effect of AHR-1 on the process of development and aging of *C. elegans*. To further support this hypothesis, a transcriptomic analysis has also shown the regulation of genes involved in larval development and normal growth.^[258]

3.2.2.4 Conclusion

In this study, we demonstrated that loss-of-function mutation in the *ahr-1* gene induces strong metabolic changes in the model organism *C. elegans* involving amino-acid metabolism, carbohydrate metabolism as well as fatty acid metabolism. These metabolic perturbations were consistently observed in *C. elegans* strains including or not a GFP reporter for AHR-1. Above all, our results suggest an effect of AHR-1 on the development and aging of *C. elegans*, which would require validation and characterization. Experiments are already planned to evaluate *ahr-1 C. elegans* longevity modification and further interpret this dataset.

3.3 Metabolic phenotyping of HBV-infected liver cells

3.3.1 Introduction

Understanding the molecular bases of the interaction between hepatitis B virus (HBV) and its host, the liver cell, leading to chronic hepatitis and liver cancer, is a rational approach to derive new drug targets to fight the HBV pandemic.

Dai *et al.* identified a physical and functional interaction between the HBV protein HBx and the endogenous multifunctional E4F1 protein in a cellular model of hepatocarcinoma, HepG2, infected by HBV.^[263] The E4F1 protein is known as a transcription factor involved in different biological processes like cellular proliferation control,^[264, 265] oxidative stress metabolism^[266] or survival of embryonic and somatic adult stem cells.^[267, 268] It was demonstrated that E4F1 protein neutralizes p53 protein-dependent detrimental activities of HBx on cytoplasmic vesicle processing and cell proliferation. This enhances the tolerance of liver cells towards HBV infection and thus contributes to the maintenance and survival of chronically infected cells. We aimed at complementing these observations by investigating the effects of the E4F1-Hbx interaction on the HBV-infected cell metabolism.

Different metabolomic approaches have been developed to probe mammalian cell metabolism, based on mass spectrometry, liquid-state NMR spectroscopy or high-resolution magic angle spinning (HR-MAS) NMR spectroscopy.^[269] In this context, the HepG2 cell line has been used as a cellular model of hepatocarcinoma for cancer research^[270] and also as a cellular model of liver to evaluate drug liver metabolism and hepatotoxicity in toxicology studies.^[259, 271-273] In our study, we investigated whether the interaction between E4F1 and HBx had an effect on the HBV-infected liver cell metabolism. We probed HBV-infected HepG2 cell metabolism from intact whole cells by HR-MAS NMR spectroscopy.^[26] We characterized these cells in three different conditions, with reduction of E4F1 levels through RNA interference (E4F1 silencing, siE4F1), reduction of both E4F1 and HBx levels (E4F1 and HBx double silencing, siE4F1-HBx) and control silencing (scr), looking for HBx-dependent metabolic perturbations of E4F1 silencing.

3.3.2 Materials and methods

Experiment model of HBV-infected hepatocarcinoma cells and targeted gene silencing.

To model HBV-infected liver cells, we used the cellular model of hepatocarcinoma, HepG2, modified to express cDNA encoding HBV genotype D and assemble viral particles *in vitro*, referred as HepG2/2.2.15.^[274]

Sample preparation (Yayun Dai, IARC). $4 \cdot 10^5$ cells were seeded in six-well plates and transfected for 48hrs with E4F1 or Scramble small interfering RNA (siRNA) as described in Appendix. Cells were then harvested by trypsinization and the cell pellet ($32 \cdot 10^6$) was washed in a freshly prepared 0.9% NaCl solution in D₂O (D215B - Deuterium oxide 100%, CortecNet, France). After gentle homogenization in 250 ml of NaCl/D₂O solution the cell suspension was distributed in 30 ml disposable NMR inserts (B4495, CortecNet, France) and immediately stored at -80°C until analyzed.

Whole HepG2 cells HR-MAS NMR spectroscopy. All experiments were carried out on a Bruker Avance II spectrometer, operating at 700 MHz (proton resonance frequency), equipped with a 4 mm HR-MAS double resonance (¹H-¹³C) probe. Temperature was controlled at 293 K throughout the experiments and magic angle spinning speed set to 3.5 KHz. 1D ¹H NOESY NMR experiments with water presaturation were performed on each sample to derive metabolic profiles. 512 free induction decays (FIDs) were co-added, with a 12 ppm spectral width and an acquisition time of 1.36 s, corresponding to 22856 data-points, with a relaxation delay of 2 s, for a total experimental time of 30 minutes per spectrum. The NOESY mixing time was set to 100 ms. The ¹H 90° hard pulse length was calibrated at 7.5 μs. 2D NMR experiments, including ¹H-¹H TOCSY and ¹H-¹³C HSQC experiments, were carried out on a subset of selected samples to characterize structural connectivities between nuclei and refine metabolite identification.

NMR data processing. All FIDs were multiplied by an exponential function corresponding to a 0.3 Hz line-broadening factor prior to Fourier transform. Phasing and baseline correction were performed manually in Topspin 2.1 (Bruker, GmbH, Rheinstetten, Germany). Spectra were automatically calibrated on the CH₃ alanine doublet at δ = 1.48 ppm. Spectra were then imported in the AMIX software (Bruker, GmbH, Rheinstetten, Germany) and reduced over the chemical range of 0.5-8.6 ppm to 8100 10⁻³ ppm-wide bins with integration of signal intensity. Residual water signal (δ = 4.5-5.18 ppm) and ethanol signals (δ = 1.17-1.197, 3.636-3.674 ppm) were excluded. Spectra were normalized using the probabilistic quotient

normalization (PQN) approach,^[97] with a median of all spectra as reference spectrum and mean-centred. Metabolite assignment was completed exploiting reference data from the literature,^[259] the HMDB,^[50] MMCD,^[49] bbiorefcode-2-0-0 (Bruker, GmbH, Rheinstetten, Germany) and Chenomx NMR Suite 7.0 (Chenomx Inc, Edmonton, Canada) spectral databases.

NMR data analysis. Principal component analysis (PCA)^[187] was first conducted in SIMCA P12+ (Umetrics, Umea, Sweden) to derive the main sources of variance within the dataset, assess sample homogeneity and exclude biological or technical outliers. Orthogonal projection to latent structure discriminant analysis (OPLS-DA) was then performed in MATLAB (The MathWorks Inc., Natick, MA) to derive pair-wise comparison between the different conditions (strains and ages).^[100] Metabolites involved in class discrimination are then derived from an univariate approach based on the statistical recoupling of variables (SRV) analysis.^[94] Univariate analysis is performed using the non-parametric two-sided rank sum test on each recoupled clusters of variables. Variable statistical significance was evaluated with different thresholds for p-values: 0.05, 0.01, and the threshold calculated from the Benjamini-Hochberg correction for multiple testing.^[102]

3.3.3 Results and discussion

We studied HepG2/2.2.15 cells with reduction in E4F1 levels only (siE4F1) or in both E4F1 and HBx levels (siE4F1-HBx), using small interfering RNA. This process of protein depletion corresponds to targeted post-transcriptional gene silencing based on RNA interference.^[275] We used HepG2/2.2.15 cells undergoing silencing with a non functional siRNA (scr) as controls to cancel out the effects of the silencing procedure in pair-wise comparisons. We performed HR-MAS NMR spectroscopy on intact whole HepG2/2.2.15 cells to derive a metabolic profile for each sample. A typical 700 MHz ¹H NOESY NMR spectrum of siE4F1-HBx cells displays signals from mobile low-molecular weight molecules, such as amino-acids, choline compounds or nucleoside derivatives, superimposed on broad signals from lipids (**Figure 3.3.1**).

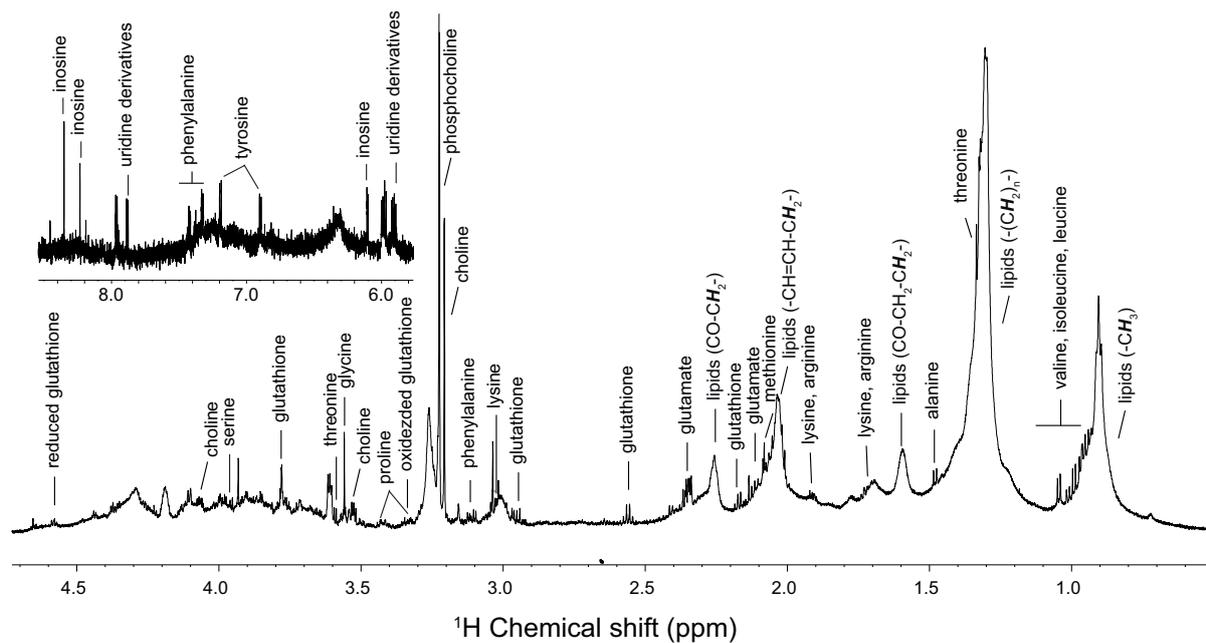


Figure 3.3.1 700 MHz 1D ^1H HR-MAS NMR spectrum of siE4F1-HBx cells for aliphatic ($\delta = 0.5\text{--}4.5$ ppm) and aromatic ($\delta = 5.5\text{--}8.5$ ppm, magnified 5 times) regions.

Effects of E4F1 silencing on HepG2/2.2.15 metabolism. We first probed the effects of E4F1 level reduction on HBV-infected hepatocarcinoma cell metabolism, by comparing the metabolic profiles of HepG2/2.2.15 cells undergoing E4F1 silencing (siE4F1) and control HepG2/2.2.15 cells undergoing silencing with a non functional siRNA (scr). We observe a significant discrimination between scr control cells and siE4F1 cells (**Figure 3.3.2a, b**, $R^2\text{X} = 0.965$, $R^2\text{Y} = 0.976$ and $Q^2 = 0.854$). Glutathione, acetate, leucine and tyrosine levels are lower in siE4F1 cells while inosine and phosphocholine levels are higher, reaching the level of significance ($q\text{-value} < 0.05$) after Benjamini-Hochberg correction for multiple testing. Without reaching this degree of significance but still associated with low p-values (without correction for multiple testing, $p\text{-value} < 0.05$), glutamate, succinate and uridine derivatives show lower levels and unsaturated lipids higher levels in siE4F1 (**Figure 3.3.2c, Table 3.3.1**). These metabolic variations may correspond to the cell endogenous response to E4F1 protein depletion as well as to perturbations of the effects on the cell metabolism of the viral HBx protein interacting with E4F1.

Effects of E4F1 and HBx double silencing on HepG2/2.2.15 metabolism. We characterized the metabolic profiles of HepG2/2.2.15 cells undergoing a double silencing, for E4F1 and HBx proteins (siE4F1-HBx).

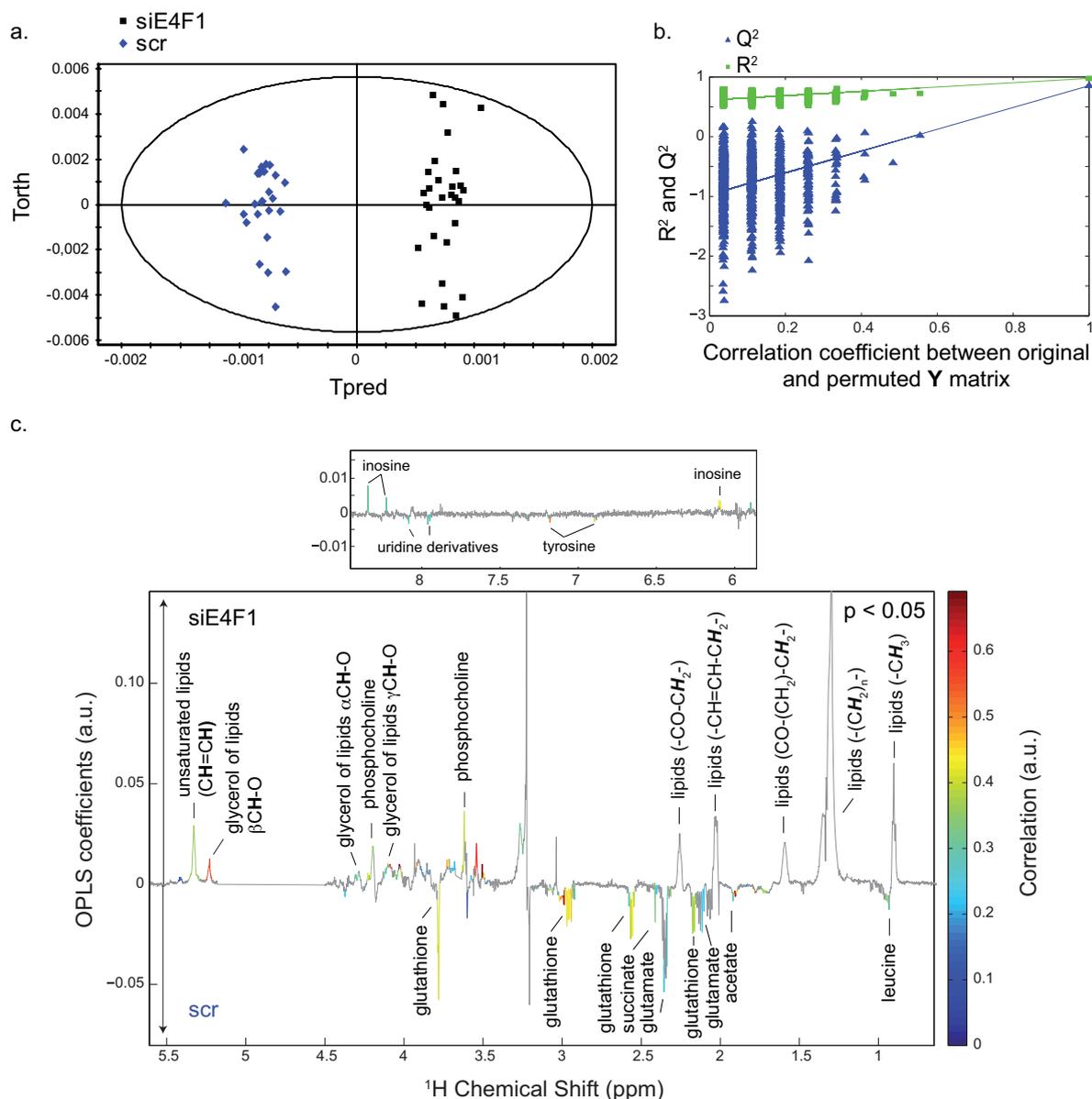


Figure 3.3.2 Discrimination between control scr cells and siE4F1 cells obtained with an OPLS model (7 orthogonal components, $R^2X = 0.965$, $R^2Y = 0.976$ and $Q^2 = 0.854$): a) score plot, b) validation results with 100 permutations and c) loading plot. NMR variable clusters associated with a p -value < 0.05 are colored according to their correlation values.

We first observe a smaller but significant discrimination between siE4F1 cells and siE4F1-HBx cells (**Figure 3.3.3a, b**, 2 orthogonal components, $R^2X = 0.857$, $R^2Y = 0.592$ and $Q^2 = 0.39$) involving higher levels of glutathione, uridine derivatives in siE4F1-HBx cells, reaching the level of significance (q -value < 0.05) after Benjamini-Hochberg correction for multiple testing (**Figure 3.3.3c, Table 3.3.1**). Phosphocholine, glutamate and tyrosine are also increased and inosine decreased in siE4F1-HBx cells without reaching this level of significance but with p -values below 0.05 (**Figure 3.3.3c, Table 3.3.1**).

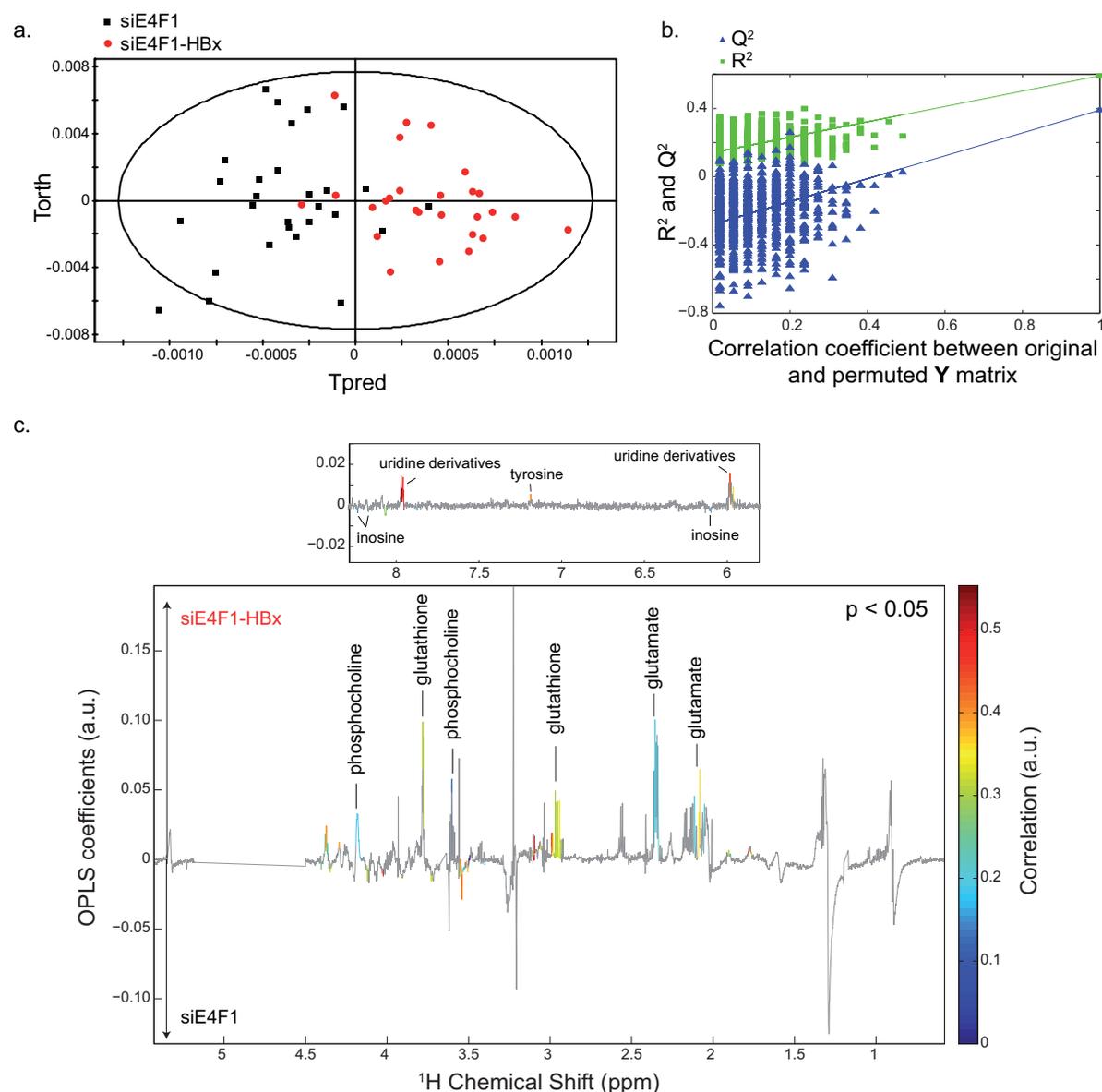


Figure 3.3.3 Discrimination between control siE4F1 cells and siE4F1-HBx cells obtained with an OPLS model (2 orthogonal components, $R^2X = 0.857$, $R^2Y = 0.592$ and $Q^2 = 0.39$): a) score plot, b) validation results with 1000 permutations and c) loading plot. NMR variable clusters associated with a p-value < 0.05 are colored according to their correlation values.

Furthermore, siE4F1-HBx and scr cells are significantly discriminated (**Figure 3.3.4a**, b, 4 orthogonal components, $R^2X = 0.908$, $R^2Y = 0.884$ and $Q^2 = 0.748$). The metabolic signature involves an increase in unsaturated lipids reaching the level of significance (q-value < 0.05) after Benjamini-Hochberg correction and with less significance but still with low p-values (p < 0.05) higher levels of creatine, inosine, uridine derivatives, phosphocholine, oxidized lipids and lower levels of acetate, alanine and choline in siE4F1-HBx cells (**Figure 3.3.4c**, **Table 3.3.1**). Glutathione, glutamate and tyrosine levels are not significantly different between siE4F1-HBx and scr cells (**Figure 3.3.4c**, **Table 3.3.1**).

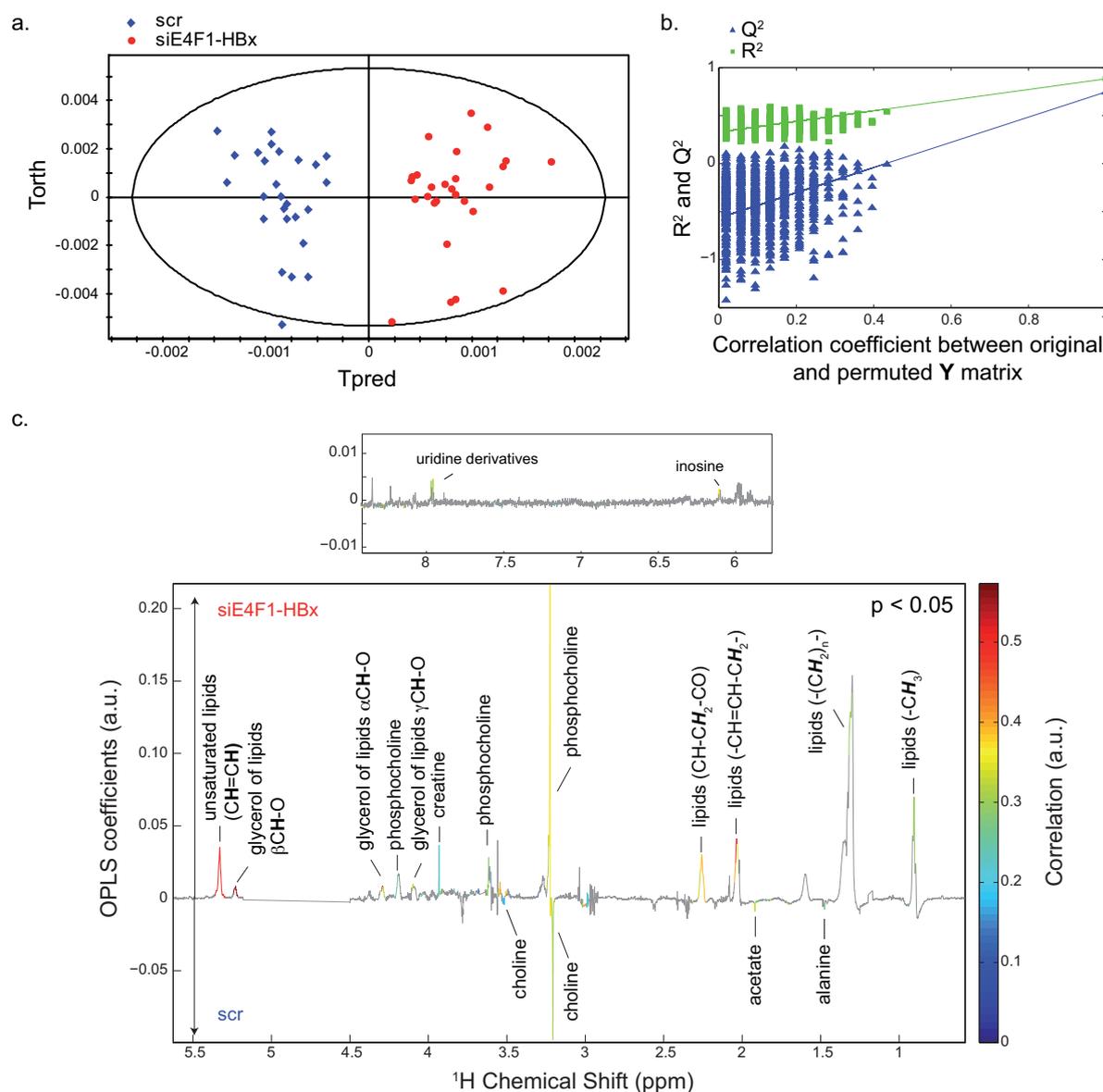


Figure 3.3.4 Discrimination between control scr cells and siE4F1-HBx cells obtained with an OPLS model (4 orthogonal components, $R^2X = 0.908$, $R^2Y = 0.884$ and $Q^2 = 0.748$): a) score plot, b) validation results with 1000 permutations and c) loading plot. NMR variable clusters associated with a p-value < 0.05 are colored according to their correlation values.

Phosphocholine and lipid metabolism perturbations in siE4F1 and siE4F1-HBx cells.

Phosphocholine and unsaturated lipid levels are both increased in siE4F1 and siE4F1-HBx cells by comparison to scr cells. Dai *et al.* observed an increase in intra-cytoplasmic large vesicle turnover and accumulation after E4F1 silencing in HepG2/2.2.15 cells which is not cancelled out by further HBx silencing. Changes in phosphocholine and lipid metabolism could thus be linked to higher lipid membrane production and disruption in HepG2/2.2.15 cells undergoing E4F1 or E4F1-HBx silencing.^[276]

Table 3.3.1. Metabolites involved in the discrimination between scr, siE4F1 and siE4F1-HBx

| Metabolites | $\delta^1\text{H}$ (ppm) | siE4F1 vs scr ^a | p-value siE4F1 vs scr ^d | siE4F1- HBx vs scr ^b | p-value siE4F1- HBx vs scr ^d | siE4F1- HBx vs siE4F1 ^c | p-value siE4F1- HBx vs siE4F1 ^d |
|---|-----------------------------|-------------------------------|--|---------------------------------------|---|--|---|
| acetate | 1.919 | ↓*** | 5.89×10^{-4} | ↓** | 7.03×10^{-3} | - | - |
| alanine | 1.48 | - | - | ↓* | 2.33×10^{-3} | - | - |
| creatine | 3.93 | - | - | ↑** | 9.63×10^{-3} | - | - |
| glutamate | 2.355 | ↓* | 3.39×10^{-2} | - | - | ↑** | 4.42×10^{-2} |
| glutathione | 2.96 | ↓*** | 2×10^{-3} | - | - | ↑*** | 3.47×10^{-4} |
| inosine | 6.10 | ↑*** | 3.93×10^{-5} | ↑* | 1.12×10^{-2} | ↓* | 2.25×10^{-3} |
| leucine | 0.96 | ↓*** | 9.175×10^{-3} | - | - | - | - |
| succinate | 2.41 | ↓* | 1.43×10^{-2} | - | - | - | - |
| tyrosine | 7.187 | ↓*** | 1.91×10^{-4} | - | - | ↑** | 2.96×10^{-3} |
| uridine derivatives | 7.96 | ↓* | 2.5×10^{-2} | ↑* | 3.20×10^{-2} | ↑*** | 3.58×10^{-5} |
| choline | 3.20 | - | - | ↓* | 1.24×10^{-2} | - | - |
| glyceryl lipids | 5.228 | ↑*** | 2.90×10^{-5} | ↑*** | 6.99×10^{-5} | - | - |
| phosphocholine | 3.62, 3.225 | ↑*** | 1.68×10^{-3} | ↑** | 7.41×10^{-3} | ↑** | 8.01×10^{-3} |
| unsaturated lipids (CH=CH) | 5.33 | ↑* | 1.23×10^{-2} | ↑*** | 1.93×10^{-4} | - | - |
| oxydized lipids (CH ₂ -CH ₂ -CO) | 2.255 | - | - | ↑* | 1.12×10^{-2} | - | - |
| lipids (-(CH ₂) _n)- | 1.31 | - | - | ↑* | 3.66×10^{-2} | - | - |

^a ↑: increase in siE4F1 cells by comparison to scr cells, ↓: decrease in siE4F1 cells by comparison to scr cells; ^b ↑: increase in siE4F1-Hbx cells by comparison to scr cells, ↓: decrease in siE4F1-Hbx cells by comparison to scr cells; ^c ↑: increase in siE4F1-HBx cells by comparison to siE4F1 cells, ↓: decrease in siE4F1-HBx cells by comparison to siE4F1 cells; ^d P-values obtained from a two-sided rank-sum test; *: p-value < 0.05; **: p-value < 0.01; ***: q-value (defined with Benjamini-Hochberg correction) < 0.05.

Perturbations of oxidative metabolism. Glutamate and total glutathione, both involved in glutathione pathway, display low levels in siE4F1 cells by comparison to scr or siE4F1-HBx cells and no significant changes between scr and siE4F1-HBx cells. We did not monitor any variations in the ratio between oxidized and reduced glutathione. These metabolic variations are thus induced by siE4F1 silencing and cancelled out by siHBx silencing. We can thus hypothesize that these metabolic perturbations reflect the effects of the viral HBx protein on HepG2/2.2.15 cell metabolism, counteracted by E4F1. Glutathione is known to be an antioxidant specie associated with oxidative metabolism. Low glutathione levels observed in siE4F1 cells only could thus be interpreted as the reflect of a high oxidative cell state induced by HBx but balanced by E4F1 in normal conditions. It has already been shown that HBx is closely linked to oxidative stress, as HBx increases production of reactive oxygen species^[277] and conversely, reactive oxygen species increase HBx levels.^[278] Our results thus suggest that E4F1 balances HBx effects on cell oxidative metabolism.

3.3.4 Conclusion

Dai *et al.* identified E4F1 as partner for HBx in HCC cells and demonstrated that this interaction neutralizes detrimental activities of HBx on cytoplasmic vesicle processing and cell proliferation. Our study shows that these changes are reflected at the metabolic level through perturbations of lipid and phosphocholine metabolism. Furthermore, our data suggests that E4F1 balances cell oxidative metabolism perturbations induced by HBx, further enhancing liver cell tolerance towards HBV infection.

3.4 Conclusion

In this third part, we have introduced four metabolomic studies on model organisms, from rats to cells, highlighting how metabolomics can help to get insight into molecular mechanisms through characterization of metabolic perturbations. Indeed, this approach allowed us to make hypotheses about the function of *ahr-1* gene and link phosphocholine metabolism to aging in *Caenorhabditis elegans*, as well as monitoring the effects on HepG2 cell metabolism of the interaction between the endogenous E4F1 protein and the HBx viral protein.

However, these studies allow us only to generate hypotheses at the molecular level, as they relied on observations and statistical correlations. For instance, we have observed a correlation between phosphocholine metabolism variations and aging in *C. elegans* but we are not able to distinguish if these variations are directly involved in aging mechanisms or due to global homeostatic metabolic changes indirectly linked to aging. It is thus mandatory to complement these studies by biological evidences proving involvement of metabolism in biological mechanisms under investigation with for instance modulation of enzyme activities through RNA interference.

Metabolomic studies of small organisms should soon benefit from progress in analytical technologies towards higher sensitivity and spatial resolution, opening access to single-organism and single-cell characterization. These achievements would allow an increase in the throughput of analyses, improving biological condition coverage and result robustness. Furthermore, monitoring inter-organism or inter-cell variations would be particularly relevant, for instance, to characterize the different cell populations, which constitute tumors.

Conclusion

The improvements of analytical techniques such as Nuclear Magnetic Resonance spectroscopy and mass spectrometry, as well as the emergence of “omics” approaches have cast a new light on metabolism during this last decade, through the development of metabolomics. It has been discovered for instance that metabolism plays a central role in the biology of particular cancers, like glioblastoma. Metabolomics have been widely applied in the framework of clinical or epidemiological for biomarker discovery, with the hope to capture a global picture of diseases and organisms in their context as metabolism is located at the cross-road of different influences such as genetics, pathologies, diet, gut microbiota and more generally global environment.

In this thesis, we have presented different developments and applications of metabolomics by high field NMR spectroscopy, from human biofluids studies to model animal characterizations, highlighting differences and specificities of this approach in different contexts. We have introduced the notion of targeted projection NMR spectroscopy for the analysis of complex mixtures, aiming at reducing metabolite signal overlaps and solve assignment ambiguities. We have introduced this approach in a model complex mixture, recording an optimized projection of a 3D ^1H - ^1H - ^{13}C TOCSY-HSQC experiment. We have demonstrated that the serum samples from the EPIC cohort are suitable for NMR

metabolomics studies, by assessing the good quality of NMR spectra, identifying in a blind analysis 10 pairs of duplicates and evaluating metabolic differences between countries of sample origin, opening access to this rich biobank for the discovery of new dietary and disease risk biomarkers. Furthermore, we have shown that a blood metabolic signature can be associated with metastatic breast cancer. We have established potential plasma metabolic signatures for different liver pathologies, hepatocellular carcinoma (HCC) and chronic liver disease (CLD), from samples collected in Thailand. We have showed that liver function through the measure of albumin levels has a strong impact on plasma metabolic profiles but could not alone explain the HCC metabolic signature. Moving on to analyses of model organisms, we have characterized plasma and urine metabolic profiles of different rat strains, Fisher, Lewis, Wistar Kyoto and Brown Norway, commonly used as controls in genetic studies. We have investigated the effects of physiological aging in *C. elegans* and observed that dietary restriction (DR), a process which increase lifespan buffers metabolic changes associated with aging through the characterization of two *C. elegans* DR mutants, *slcf-1* and *eat-2*. We have further identified that perturbations in phosphocholine metabolism correlate with life expectancy. We have then characterized the effects on metabolism of the loss of function mutation of *ahr-1* gene, suggesting an involvement in development and aging processes, requiring further evaluation. We have finally investigated the metabolic influences of the interaction between the endogenous E4F1 protein and the HBx viral protein in a model of liver cells infected by HBV and suggested that E4F1 balances cell oxidative metabolism perturbations induced by HBx, further enhancing tolerance of liver cells.

Metabolomics have strongly benefited from the advances of analytical devices we can expect some new improvements in the following years. For NMR spectroscopy, development of hardware technologies such as cryoprobes, microcoils or increase in magnetic field could improve sensitivity and resolution for this technique. Furthermore, the use of Dynamic Nuclear Polarization (DNP) raises hopes to dramatically increase the sensitivity of NMR devices. However, a global metabolome coverage by a single analytical technique seems still utopian and combination of complementary devices would be necessary in the future. Getting access to single single-cell analysis could address new relevant questions, related for instance to the spatial distribution of metabolites in tissues and tumors, delineating different cellular behaviors.

In vivo magnetic resonance spectroscopy with the increase of magnetic fields in particular, could soon completely benefit from the results of in vitro metabolomics analytical strategies. Indeed, the chemical resolution of metabolic profiles, which can be obtained in

vivo, in small volumes, is increasing, capturing more and more information contained in ex-vivo metabolomic profiles. A reinforced crosstalk between the two fields could thus be expected.

Metabolomics offers now a new readout at the metabolic level for many biological systems, such as yeast, bacteria or cell cultures, small or large model organisms. We can imagine that biologists would exploit this tool to characterize their systems and further investigate the role of metabolism in biological processes.

Except for the characterization of inborn errors of metabolism, metabolomics has not yet reached real clinical applications and is still in a discovery phase. A growing number of articles report metabolic differences between individuals in various conditions, of potential interests. However these results need to be validated in large patient cohorts, which should be now prospectively recruited according to the standard and designs of metabolomics studies. The issue of the specificity of these results for their use in a daily clinical environment is also of major interest. Many metabolites are now associated with completely different physiological or pathological processes. It seems thus necessary to refine clinical questions to obtain enough specificity. A global picture of metabolic variations in normal physiological conditions seems thus mandatory as well as the characterization of the effects on metabolism of non-specific pathological states encountered in many diseases such as anemia, kidney or liver failure.

Metabolomics is a growing field which benefit from the late technological advances. From a discovery phase, the field is now moving to a validation phase, through the replication of studies with an increase in sample numbers and refinement of clinical questions towards real biomedical applications.

Part 4: Appendixes

4.1 On EPIC cross-sectional pilot study: Methods for UPLC-MS and 600 MHz NMR spectroscopy

In this part, we reported protocols used by our collaborators in Imperial College London for Ultra-Performance Liquid Chromatography-Mass Spectrometry (UPLC-MS) and Nuclear Magnetic Resonance (NMR) spectroscopy performed on a spectrometer operating at 600 MHz, corresponding to established procedures.^[17]

Chemicals. All chemicals (water, methanol, acetonitrile, formic acid, leucine encephalin acetate hydrate, fetal bovine serum) were obtained from Sigma with all solvents of LCMS grade.

Ultra Performance Liquid Chromatography – Quadrupole Time-of-Flight Mass Spectrometry (UPLC-QToFMS). All samples were thawed together on ice, vortex mixed (30 s) and kept on ice until prepared. Quality control (QC) samples were prepared using a linear combination of all samples. Column preconditioning samples were prepared using a surrogate FBS matrix, prepared in the same way as the biobank samples.

For each sample, an aliquot (80 μL) was mixed with water (80 μL) and methanol (480 μL , 4°C) and vortex mixed (30 s). Samples were kept at 4°C before a further vortex mix (30 s) and then centrifugation (10 min, 16000 g, 4°C) to effectively deproteinize the sample. The supernatant (480 μL) was removed and transferred to a clean sample tube. Samples were then dried (Eppendorf SpeedVac, 40°C, 2hrs), recapped and stored at -80°C until the analysis by UPLC-QToFMS. Samples were thawed (30 min, RT) before reconstitution in water (100 μL), aided by vortex mixing (30 s). To ensure no particulate matter remained in the sample, each was centrifuged (10 min, 16000g, 4°C) before the supernatant (75 μL) was transferred to a clear glass sample vial (Waters Certified LCMS Vials).

Analyses were performed on Acquity UPLC system (Waters Corporation, Milford, MA, USA) comprising a binary solvent manager, sample manager, column oven and photodiode array detector, hyphenated to a Water LCT Premier mass spectrometer (Waters MS Technologies, Manchester, UK) and controlled using MassLynx v4.1 software. An Acquity UPLC BEH C₁₈ 1.7 μm 2.1 x 100mm column was used for the chromatographic separation of serum metabolites. The column was maintained at 50°C and the sample manager at 4°C. Samples remained in the cooled sample manager (4°C) during the entire analysis.

A binary gradient (solvent A = water + 0.1 % formic acid (v/v); solvent B = methanol + 0.1 % formic acid (v/v)) was supplied to the chromatographic column at a flow rate of 0.4 ml min⁻¹. Gradient conditions are given in the **Table 4.1.1**.

Table 4.1.1. Gradient conditions for UPLC-ToF-MS analysis of EPIC samples. Solvent A = water + 0.1 % formic acid (v/v); Solvent B = methanol + 0.1 % formic acid (v/v)).

| Time (min) | Flow Rate (ml ,min-1) | %A | %B | Gradient Curve |
|------------|-----------------------|------|------|----------------|
| Initial | 0.4 | 99.9 | 0.1 | Linear |
| 1 | 0.4 | 99.9 | 0.1 | Linear |
| 16 | 0.4 | 0.1 | 99.9 | Linear |
| 18 | 0.4 | 0.1 | 99.9 | Linear |
| 19 | 0.4 | 99.9 | 0.1 | Linear |
| 20 | 0.4 | 99.9 | 0.1 | Linear |

The chromatographic column was preconditioned with injections of FBS surrogate sample matrix, and then pooled QC samples (n=3) prior to the start of the full analytical batch. Visual inspection of the retention times of the main spectral features across the chromatograms resulting from these injections was used to determine whether the system was adequately reproducible and therefore equilibrated. Sample injections (5 μ L) were made following an equilibration time (2 min) at the starting gradient conditions (100% A). Pooled QC samples were injected and analysed at the start, end, and after every 10 samples within the batch.

The chromatographic eluent was delivered to the inlet of the mass spectrometer with the ion optics operating in V mode (resolution \sim 8000). Data was collected in the m/z range 50-1000 in electrospray positive mode (ESI+). The source temperature was held at 120°C with a flow of nitrogen desolvation gas supplied at a rate of 900 L hr⁻¹ at 350°C. The nitrogen sample cone gas was supplied at 25 L hr⁻¹. The sample capillary voltage was set at 3.2 kV and the sampling cone voltage at 35 V. A reference solution (50:50 water:acetonitrile) containing 1% (v/v%) formic acid and leucine enkephalin (m/z 556.2771, 50 pg μ L⁻¹) was infused into the LockSpray inlet at 10 μ L min⁻¹ to provide a linear mass calibration.

Spectral Alignment, Feature Selection and Data Preprocessing. Mass spectral data was converted to NetCDF format using the DataBridge utility supplied with MassLynx v4.1 (Waters). The resulting profile data from the analyte channel was processed using XCMS v1.16.3 running in the R computing environment. Briefly, spectral features comprising m/z -retention time pairs were identified and extracted from the profile data using the *centWave* algorithm in XCMS. This uses a continuous wavelet transformation to identify regions of interest (ROIs) in the data and performs integration to generate a spectral feature intensity

matrix. The mass spectral data for one sample could not be processed at this stage (file error), and therefore this sample was omitted from related analysis. The relative standard deviation (RSD) of intensity data for mass-retention time spectral features selected by XCMS was calculated across the pooled QC (n=4) samples run throughout the batch for each mode (first and then after every 10 consecutive injections). “Well-behaved” features with RSD<5% across the QC samples were selected for use in multivariate analysis.

¹H Nuclear Magnetic Resonance Spectroscopy. Sample preparation and data processing were similar between ICL and CRMN. Concerning profile generation, ¹H All NMR experiments were carried out on a Bruker AVANCE 600 NMR operating at 600.13 MHz (proton resonance frequency), equipped with a 5mm BBI probe and an automatic sample changer (B-ACS, Bruker) at ambient room temperature, Temperature was controlled at 300 K throughout the experiments. ¹H 1D CPMG spectra were acquired into 32 K data-points as the sum of 128 transients, following 16 dummy scans. The water resonance was suppressed using resonance presaturation during the relaxation delay (2s).

Table 4.1.2. Major putative serum metabolites detected by ¹H NMR and UPLC-MS.

| Metabolite Class | Metabolite | NMR | UPLC-MS | Metabolite Class | Metabolite | NMR | UPLC-MS | |
|------------------------|----------------|--------------------|---------|------------------|-----------------------|-----------------------|---------|---|
| Lipids | Cholesterol | Y | N | Organic acids | 2-Hydroxyisobutyrate | Y | N | |
| | Choline | Y | Y | | 3-Hydroxybutyrate | Y | N | |
| | Oleamide | N | Y | | Acetate | Y | N | |
| | 16:0 LPC | N | Y | | Acetoacetate | Y | N | |
| | 18:1 LPC | N | Y | | Alanine | Y | N | |
| | 18:0 LPC | N | Y | | Arginine | Y | N | |
| | 17:0 LPC | N | Y | | Asparagine | N | Y | |
| | 20:0 LPC | N | Y | | Aspartate | Y | N | |
| | 18:2 LPC | N | Y | | Betaine | Y | N | |
| | 20:4 LPC | N | Y | | Citrate | Y | Y | |
| | 22:6 LPC | N | Y | | Creatine | Y | N | |
| | 14:0 LPC | N | Y | | Creatinine | Y | N | |
| | 16:1 LPC | N | Y | | Formate | Y | N | |
| | 22:5 LPC | N | Y | | Glutamate | Y | Y | |
| | 22:4 LPC | N | Y | | Glutamine | Y | Y | |
| | 34:1 PC | N | Y | | Glycine | Y | N | |
| | 36:1 PC | N | Y | | Hippuric acid | N | Y | |
| | 32:2 PC | N | Y | | Histidine | Y | N | |
| | 34:2 PC | N | Y | | Isoleucine | Y | Y | |
| | 36:2 PC | N | Y | | Lactate | Y | N | |
| | SM(d18:1/16:0) | N | Y | | Leucine | Y | Y | |
| | SM(d18:1/18:0) | N | Y | | Lysine | Y | N | |
| | SM(d18:1/20:0) | N | Y | | Malonate | Y | N | |
| | SM(d18:1/22:0) | N | Y | | Methionine | Y | Y | |
| | Carnitines | Carnitine | N | | Y | N-acetyl muramic acid | N | Y |
| | | Acetylcarnitine | N | | Y | Norleucine | N | Y |
| | | Propionylcarnitine | N | | Y | Phenylalanine | Y | Y |
| Butyrylcarnitine | | N | Y | Proline | Y | Y | | |
| Isovalerylcarnitine | | N | Y | Pyroglutamate | N | Y | | |
| Hexanoylcarnitine | | N | Y | Pyruvate | Y | N | | |
| Octanoylcarnitine | | N | Y | Serine | Y | N | | |
| Nonanoylcarnitine | | N | Y | Threonine | Y | Y | | |
| Acylcarnitine C10 | | N | Y | Tryptophan | Y | Y | | |
| Acylcarnitine C12 | | N | Y | Tyrosine | Y | Y | | |
| Acylcarnitine C14 | | N | Y | Uric acid | N | Y | | |
| Acylcarnitine C16 | | N | Y | Valine | Y | Y | | |
| Acylcarnitine C18 | | N | Y | Other | Acetone | Y | N | |
| Methylmalonylcarnitine | | | Y | | Dimethylamine | Y | N | |
| Octenoylcarnitine | | N | Y | | Inosine | N | Y | |
| Acylcarnitine C10:1 | | N | Y | | Bilirubin | N | Y | |
| Acylcarnitine C14:1 | | N | Y | | Hypoxanthine | Y | Y | |
| Acylcarnitine C16:1 | | N | Y | | Chenodeoxycholic acid | N | Y | |
| Acylcarnitine C18:1 | | N | Y | | Dihydrothymine | N | Y | |
| Acylcarnitine C18:2 | | N | Y | | Caffeine | N | Y | |
| | | | | | Glucose | Y | N | |
| | | | | | Glycerol | Y | N | |
| | | | | | myo-Inositol | Y | N | |

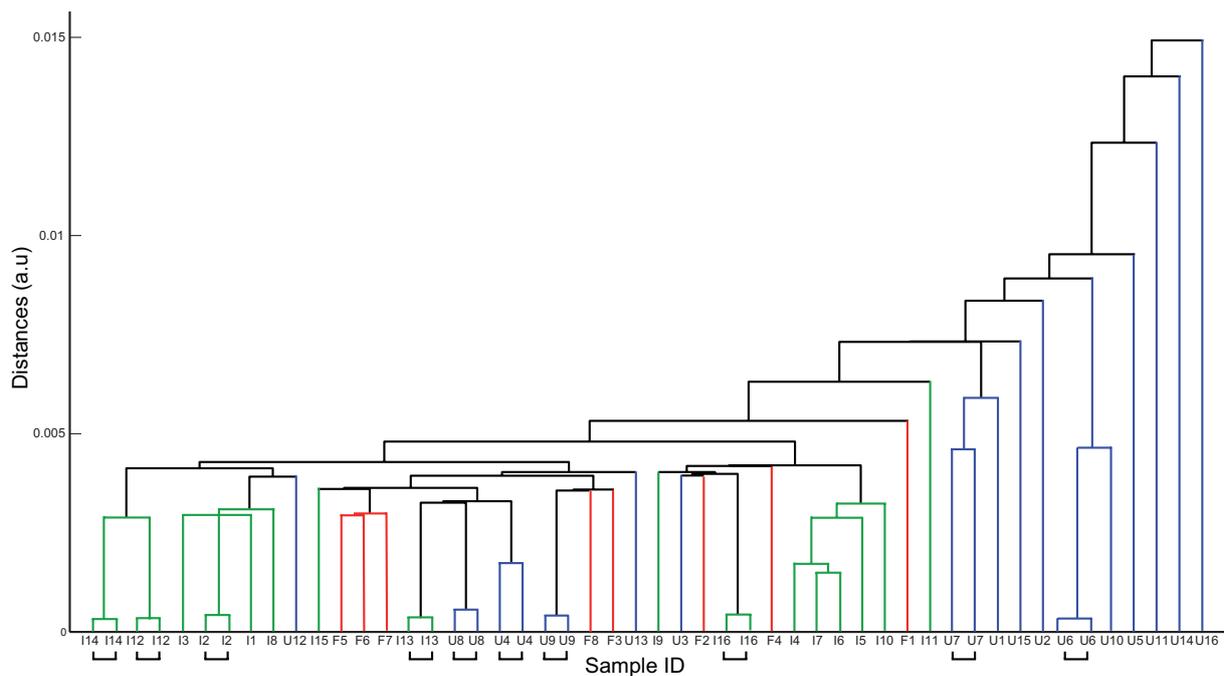


Figure 4.1.1. Hierarchical clustering of 600 MHz ^1H NMR metabolic profiles. Duplicate samples are indicated by similar sample IDs. The dendrogram is coloured according to country of origin (green = Italy, red = France, blue = UK). Peak intensities for each species were corrected individually for analytical batch differences and data were autoscaled prior to analysis. ^1H NOESY NMR data were mean-centred prior to analysis. Dendrograms were generated using single linkage and the Euclidean distance metric.

4.2 On metastatic breast cancer serum signature

Table 4.2.1. Summary of clinicopathological characteristics of breast cancer patients: a) Training cohort characteristics; b) Validation cohort characteristics (Elodie Jobard, CRMN)

| A- Model | | | |
|--------------------------------------|--------------------|---------------------|-----------------------------|
| Characteristics | Localized Patients | Metastatic Patients | <i>p</i> value ^a |
| No. of subjects | 28 | 23 | |
| Age (mean/SD) | 56.0 ± 9.2 | 57.5 ± 11.5 | 0.63 |
| Menopause | | | 0.51 |
| Post-menopausal women | 15(54%) | 16(70%) | |
| Pre-menopausal women | 11(39%) | 6(26%) | |
| Unknown | 2(7%) | 1(4%) | |
| HER2 status | | | 0.29 |
| Positive | 5(18%) | 7(30%) | |
| Negative | 23 (82%) | 16(70%) | |
| Hormone receptors^b | | | 0.007 |
| HR+ | 25(89%) | 13(57%) | |
| HR- | 3(11%) | 10(43%) | |

| B - Validation | | | |
|--------------------------------------|--------------------|---------------------|-----------------------------|
| Characteristic | Localized Patients | Metastatic Patients | <i>p</i> value ^a |
| No. of subjects | 18 | 14 | |
| Age (mean/SD) | 56.2 ± 11.0 | 56.7 ± 11.8 | 0.91 |
| Menopause | | | 0.65 |
| Post-menopausal women | 9(50%) | 9(64%) | |
| Pre-menopausal women | 8(44%) | 4(29%) | |
| Unknown | 1(6%) | 1(7%) | |
| HER2 status | | | 0.65 |
| Positive | 2(11%) | 2(14%) | |
| Negative | 15(83%) | 12(86%) | |
| Unknown | 1(6%) | 0(0%) | |
| Hormone receptors^b | | | 0.66 |
| HR+ | 11(61%) | 9(64%) | |
| HR- | 6(33%) | 5(36%) | |
| Unknown | 1(6%) | 0(0%) | |

The two outliers are not included in these tables.^a *p*-value calculated using either the Student or Chi² tests respectively for the mean and for proportions.^b Hormone receptors are receptors for estrogen and progesterone; HR-: at least one of the two receptors is negative; HR+: both receptors are positive.

Table 4.2.2. Summary of medical history of breast cancer patients (Elodie Jobard, CRMN)

A- Medical History - Training Cohort

| Characteristics | Localized Patients | Metastatic Patients | p value |
|----------------------|--------------------|---------------------|---------|
| Hypertension | 7 (25%) | 7 (30%) | 0.66 |
| Hypothyroidism | 3 (11%) | 1 (4%) | 0.40 |
| Deep vein thrombosis | 1 (4%) | 0 (0%) | 0.36 |
| Hypercholesterol | 3 (11%) | 3 (13%) | 0.79 |
| Bronchiectasis | 1 (4%) | 0 (0%) | 0.36 |
| Viral cirrhosis C | 1 (4%) | 0 (0%) | 0.36 |
| Glomerulonephritis | 1 (4%) | 0 (0%) | 0.36 |
| Of allergy | 1 (4%) | 0 (0%) | 0.36 |
| Dyslipidemia | 0 (0%) | 1 (4%) | 0.26 |
| Tyroidectomy | 0 (0%) | 1 (4%) | 0.26 |
| Stroke | 0 (0%) | 1 (4%) | 0.26 |
| Tuberculosis | 1 (4%) | 0 (0%) | 0.36 |
| Parkinson | 1 (4%) | 0 (0%) | 0.36 |

B- Medical History - Validation Cohort

| | Localized Patients | Metastatic Patients | p value |
|--------------------|--------------------|---------------------|---------|
| Hypertension | 1 (13%) | 2 (40%) | 0.4 |
| Hypothyroidism | 3 (38%) | 1 (20%) | 0.42 |
| Acute pancreatitis | 1 (13%) | 0 (0%) | 0.37 |
| Bouveret's disease | 1 (13%) | 0 (0%) | 0.37 |
| Osteoporosis | 1 (13%) | 0 (0%) | 0.37 |
| Depression | 0 (0%) | 1 (20%) | 0.25 |
| Epilepsy | 0 (0%) | 1 (20%) | 0.25 |
| Osteoarthritis | 1 (13%) | 0 (0%) | 0.37 |

Table 4.2.3. Metabolites identified from 1D and 2D NMR profiles of blood sera of patients with either metastatic or localized breast cancer (Elodie Jobard, CRMN).

| Metabolites | $\delta^1\text{H}$ ppm | $\delta^{13}\text{C}$ ppm | Multiplicity | Group | Observed |
|--------------------|---------------------------|------------------------------|--------------|------------------------------|--------------------------|
| 3-hydroxybutyrate | 1.19 | 24.3 | d | $\gamma\text{-CH}_3$ | CPMG, J-Res, HSQC, TOCSY |
| | 2.29 | | q | half $\beta\text{-CH}_2$ | CPMG, J-Res, TOCSY |
| | 2.39 | | q | half $\beta\text{-CH}_2$ | CPMG, J-Res, TOCSY |
| | 4.14 | | m | $\beta\text{-CH}$ | CPMG, J-Res, TOCSY |
| Acetate | 1.91 | | s | CH_3 | CPMG, J-Res, TOCSY |
| Acetoacetate | 2.27 | 32.2 | s | CH_3 | CPMG, J-Res, HSQC |
| | 3.43 | 56.1 | s | | CPMG, J-Res |
| Acetone | 2.22 | | s | CH_3 | CPMG, J-Res, TOCSY |
| Alanine | 1.46 | 18.9 | d | CH_3 | CPMG, J-Res, HSQC, TOCSY |
| | 3.76 | 53 | q | $\alpha\text{-CH}$ | CPMG, J-Res, HSQC, TOCSY |
| Albumin lysil | 2.89 | 42 | t | $\varepsilon\text{-CH}_2$ | CPMG, HSQC |
| | 2.96 | 42 | t | $\varepsilon\text{-CH}_2$ | HSQC |
| | 3.01 | 42 | t | $\varepsilon\text{-CH}_2$ | HSQC |
| Aspartate | 2.65 | | dd | half $\beta\text{-CH}_2$ | J-Res |
| | 2.80 | | dd | half $\beta\text{-CH}_2$ | J-Res |
| Betaine | 3.25 | 56.3 | s | CH_3 | CPMG, J-Res |
| | 3.90 | | s | | CPMG, J-Res, HSQC |
| Cholesterol | 0.65 | 14.2 | m | C18 | NOESY, HSQC |
| | 0.84 | 25.1 | m | C26, C27 | CPMG, HSQC |
| | 0.9 | 21.2 | m | C21 | HSQC |
| | 1.47 | 30.3 | m | C25 | HSQC |
| | 1.11 | 42 | m | C12 | CPMG, HSQC |
| Choline | 3.21 | 56.5 | s | $\text{N}(\text{CH}_3)_3$ | CPMG, J-Res, HSQC, TOCSY |
| Choline (in lipid) | 3.66 | | m | | CPMG, J-Res |
| Citrate | 2.52 | | d | half- CH_2 | CPMG, J-Res, TOCSY |
| | 2.67 | | d | half- CH_2 | CPMG, J-Res, TOCSY |
| Creatine | 3.02 | | s | CH_3 | CPMG, J-Res, TOCSY |
| | 3.91 | | s | CH_2 | CPMG, J-Res, TOCSY |
| Creatinine | 3.03 | | s | CH_3 | CPMG, J-Res, TOCSY |
| | 4.05 | | s | CH_2 | CPMG, J-Res, TOCSY |
| Dimethylamine | 2.71 | | s | CH_3 | CPMG, J-Res |
| Formate | 8.45 | | s | CH | CPMG, J-Res, TOCSY |
| b-glucose | 3.24 | 76.8 | dd | H2 | CPMG, J-Res, HSQC, TOCSY |
| | 3.40 | 72.2 | q | H4 | CPMG, J-Res, HSQC, TOCSY |
| | 3.46 | 78.4 | t | H5 | CPMG, J-Res, HSQC, TOCSY |
| | 3.48 | 78.4 | t | H3 | CPMG, J-Res, HSQC, TOCSY |
| | 3.89 | 63.4 | dd | half CH_2C_6 | CPMG, J-Res, HSQC, TOCSY |
| | 4.64 | 98.4 | d | H1 | CPMG, J-Res, HSQC, TOCSY |
| a-glucose | 3.40 | 72.2 | q | H4 | CPMG, J-Res, HSQC, TOCSY |
| | 3.53 | 74 | dd | H2 | CPMG, J-Res, HSQC, TOCSY |
| | 3.71 | 75.3 | t | H3 | CPMG, J-Res, HSQC, TOCSY |
| | 3.72 | 63.4 | dd | half CH_2C_6 | CPMG, J-Res, HSQC, TOCSY |

| | | | | | |
|--------------------------------------|------|------|-----|---|--------------------------|
| | 3.76 | 63.3 | m | half CH ₂ C ₆ | CPMG, J-Res, HSQC, TOCSY |
| | 3.82 | 74 | ddd | H5 | CPMG, J-Res, HSQC, TOCSY |
| | 3.84 | 63.2 | m | half CH ₂ C ₆ | CPMG, J-Res, HSQC, TOCSY |
| | 5.23 | 94.6 | d | H1 | CPMG, J-Res, HSQC, TOCSY |
| Glutamate | 2.04 | | m | half β-CH ₂ | CPMG, TOCSY |
| | 2.12 | | m | half β-CH ₂ | CPMG, J-Res, TOCSY |
| | 2.34 | | m | half γ-CH ₂ | CPMG, J-Res, TOCSY |
| Glutamine | 2.09 | 29.5 | m | half β-CH ₂ | CPMG, J-Res, HSQC, TOCSY |
| | 2.43 | 33.6 | m | half γ-CH ₂ | CPMG, J-Res, HSQC, TOCSY |
| | 3.76 | 57.1 | t | α-CH ₂ | CPMG, J-Res, HSQC, TOCSY |
| Glycerol | 3.56 | 65.5 | dd | half-CH ₂ | CPMG, J-Res, HSQC, TOCSY |
| | 3.64 | 65.3 | dd | half-CH ₂ | CPMG, J-Res, HSQC, TOCSY |
| | 3.88 | 74.3 | m | C2-H | CPMG, J-Res, TOCSY |
| Glycine | 3.54 | 44.4 | s | CH ₂ | CPMG, J-Res, HSQC, TOCSY |
| Glycoprotein (N-acetyl) NAC 1 | 2.03 | 24.7 | s | NHCOCH ₃ | CPMG, JRes, HSQC, TOCSY |
| Glycoprotein (N-acetyl) NAC 2 | 2.06 | 24.8 | s | NHCOCH ₃ | CPMG, JRes, HSQC, TOCSY |
| Histidine | 7.03 | | s | H4 | CPMG, TOCSY |
| | 7.73 | | s | H2 | CPMG, J-Res, TOCSY |
| Isobutyrate | 1.06 | | d | CH ₃ | CPMG, J-Res, TOCSY |
| Isoleucine | 0.92 | | t | δ-CH ₃ | CPMG, J-Res |
| | 1 | | d | β-CH ₃ | CPMG, J-Res |
| Lactate | 1.32 | 22.7 | d | CH ₃ | CPMG, J-Res, HSQC, TOCSY |
| | 4.10 | 71 | q | CH | CPMG, J-Res, HSQC, TOCSY |
| Leucine | 0.94 | | t | δ-CH ₃ | CPMG, J-Res, TOCSY |
| | 0.95 | | d | δ-CH ₃ | CPMG, J-Res |
| Lipid (LDL) | 0.84 | 16.5 | m | CH ₃ (CH ₂) n | CPMG, HSQC |
| Lipid (VLDL) | 0.86 | 23.4 | m | | CPMG,HSQC |
| | 1.24 | 34.2 | m | CH ₃ CH ₂ CH ₂ | CPMG, HSQC |
| | 1.26 | 32.1 | m | (CH ₂) n | CPMG, HSQC |
| | 1.55 | 27.3 | m | CH ₂ CH ₂ CO | CPMG, HSQC |
| Lipid | 0.93 | 27.3 | m | CH ₃ CH ₂ | |
| | 1.26 | 25.1 | m | CH ₃ CH ₂ CH ₂ (CH ₂) | CPMG, HSQC |
| | 1.7 | 29.1 | m | CH ₂ CH ₂ C=C | CPMG, HSQC |
| | 2 | 29.6 | m | CH ₂ C=C | CPMG, HSQC |
| | 2.22 | 36.1 | m | CH ₂ CO | CPMG, HSQC |
| | 2.71 | 28 | m | C=CCH ₂ C=C | CPMG, HSQC |
| Lysine | 1.47 | | m | | CPMG, TOCSY |
| | 1.71 | | m | | CPMG, TOCSY |
| | 1.89 | | m | | CPMG, TOCSY |
| | 3.02 | | t | | CPMG, TOCSY |
| Mannose | 5.17 | | d | | CPMG, J-Res, TOCSY |
| Malonate | 3.15 | | s | | CPMG, J-Res, TOCSY |
| Methanol | 3.35 | | s | CH ₃ | CPMG, J-Res, TOCSY |
| Methionine | 2.13 | | s | S-CH ₃ | CPMG, J-Res, TOCSY |
| Phenylalanine | 3.99 | | m | α -CH | CPMG |

| | | | | | |
|--------------------------|------|-------|----------|---------------------------------------|--------------------------|
| | 7.31 | | m | H2, H6 | CPMG |
| | 7.35 | | m | H4 | CPMG |
| | 7.43 | | m | H3,H5 | CPMG |
| Pyruvate | 2.36 | | s | CH ₂ | CPMG, J-Res |
| Tyrosine | 3.95 | | dd | α-CH | J-Res |
| | 6.87 | | m | H3, H5 | CPMG, J-Res, TOCSY |
| | 7.17 | | m | H2, H6 | CPMG, J-Res, TOCSY |
| Valine | 0.97 | 19.2 | d | CH ₃ | CPMG, J-Res, HSQC, TOCSY |
| | 1.02 | 20.6 | d | CH ₃ | CPMG, J-Res, HSQC, TOCSY |
| | 2.25 | 32 | spt of d | β-CH | CPMG, J-Res, HSQC, TOCSY |
| | 3.58 | 63.2 | d | α-CH | CPMG, J-Res, HSQC, TOCSY |
| Unsaturated lipid | 5.23 | | m | | CPMG |
| | 5.26 | 130.2 | m | (CH=CHCH ₂ CH =CH) | CPMG, HSQC |
| | 5.27 | | m | (=CHCH ₂ CH ₂) | CPMG |
| | 5.29 | 131.7 | m | (CH=CHCH ₂ CH =CH) | CPMG, HSQC |
| | 5.31 | | m | (=CHCH ₂ CH ₂) | CPMG |
| | 5.33 | | m | (=CHCH ₂ CH ₂) | CPMG |

Table 4.2.4. Significant metabolites differentiating patients with metastatic and localized breast cancer.

| ID | Metabolite | ¹ H ppm | ¹³ C ppm | Signal Multiplicity | Group | Observed by NMR | p-value | q-value BH ^b |
|----|--------------------|-----------------------|------------------------|------------------------|------------------------|-----------------------------|---------|----------------------------|
| 1 | | 1.38 | | d | | CPMG, J-Res, TOCSY | 0.003 | 0.03 |
| 2 | | 1.39 | | d | | CPMG, J-Res, TOCSY | 0.003 | 0.03 |
| 3 | NAC ^a 1 | 2.03 | 24.7 | s | NHCOCH ₃ | CPMG, J-Res, HSQC, TOCSY | 0.0001 | 0.005 |
| 4 | NAC ^a 2 | 2.06 | 24.8 | s | NHCOCH ₃ | CPMG, J-Res, HSQC, TOCSY | 0.0000 | 0.005 |
| 5 | Acetoacetate | 2.27 | 32.2 | s | CH ₃ | CPMG, J-Res, HSQC, TOCSY | 0.002 | 0.03 |
| | | 3.43 | 56.1 | s | | CPMG, J-Res, HSQC, TOCSY | 0.003 | 0.03 |
| 6 | Pyruvate | 2.36 | | s | CH ₂ | CPMG, J-Res | 0.0003 | 0.009 |
| 7 | 3-hydroxybutyrate | 2.29 | | q | half β-CH ₂ | CPMG, J-Res, TOCSY | 0.005 | 0.04 |
| 8 | Glycerol | 3.56 | 65.5 | dd | half β-CH ₂ | CPMG, J-Res, HSQC, TOCSY | 0.0001 | 0.006 |
| | | 3.64 | 65.3 | dd | half β-CH ₂ | CPMG, J-Res, HSQC, TOCSY | 0.001 | 0.01 |
| | | 3.88 | 74.3 | m | C ₂ H | CPMG, J-Res, TOCSY | 0.0009 | 0.01 |
| 9 | Histidine | 7.03 | | s | H4 | CPMG, TOCSY | 0.001 | 0.01 |
| 10 | | 7.24 | | m | | CPMG | 0.0005 | 0.01 |
| 11 | Phenylalanine | 3.99 | | | α-CH | CPMG | 0.001 | 0.02 |
| | | 7.31 | | m | H2, H6 | CPMG | 0.0006 | 0.01 |
| | | 7.35 | | m | H4 | CPMG | 0.0002 | 0.006 |

^a NAC = N-acetylglycoprotein, ^b q-value BH = *p*-value after Benjamini-Hochberg false discovery rate multiple testing correction

4.3 On plasma metabolic signature of hepatocellular carcinoma and chronic liver disease.

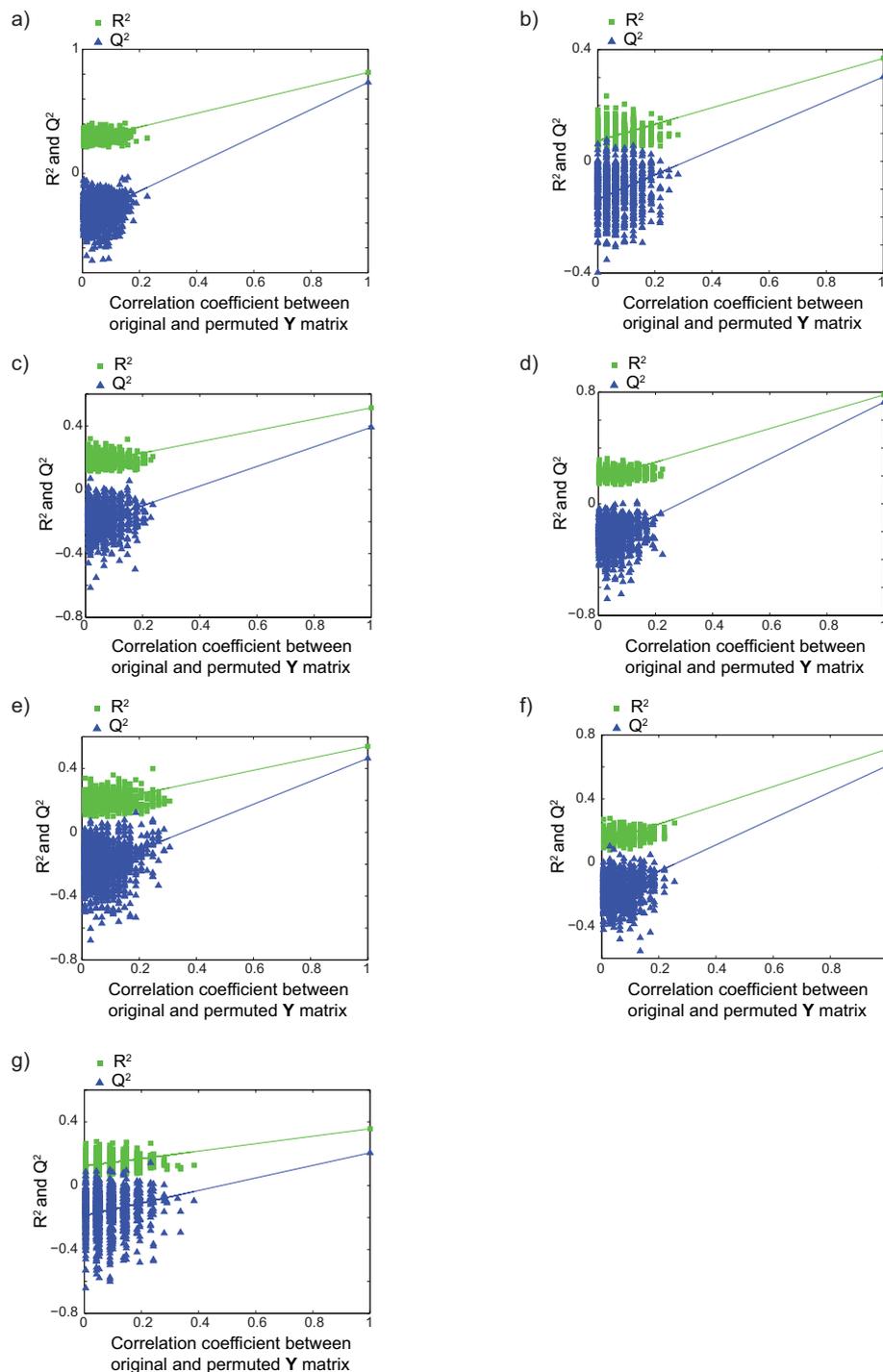


Figure 4.3.1. Model validations resulting from 1000 permutations, demonstrating model robustness for discrimination between a) HCC and Control, b) HCC and CLD, c) Control and CLD, d) HCC without cirrhosis and Control, e) HCC without cirrhosis and CLD, f) HCC with cirrhosis and Control and f) HCC with cirrhosis and CLD.

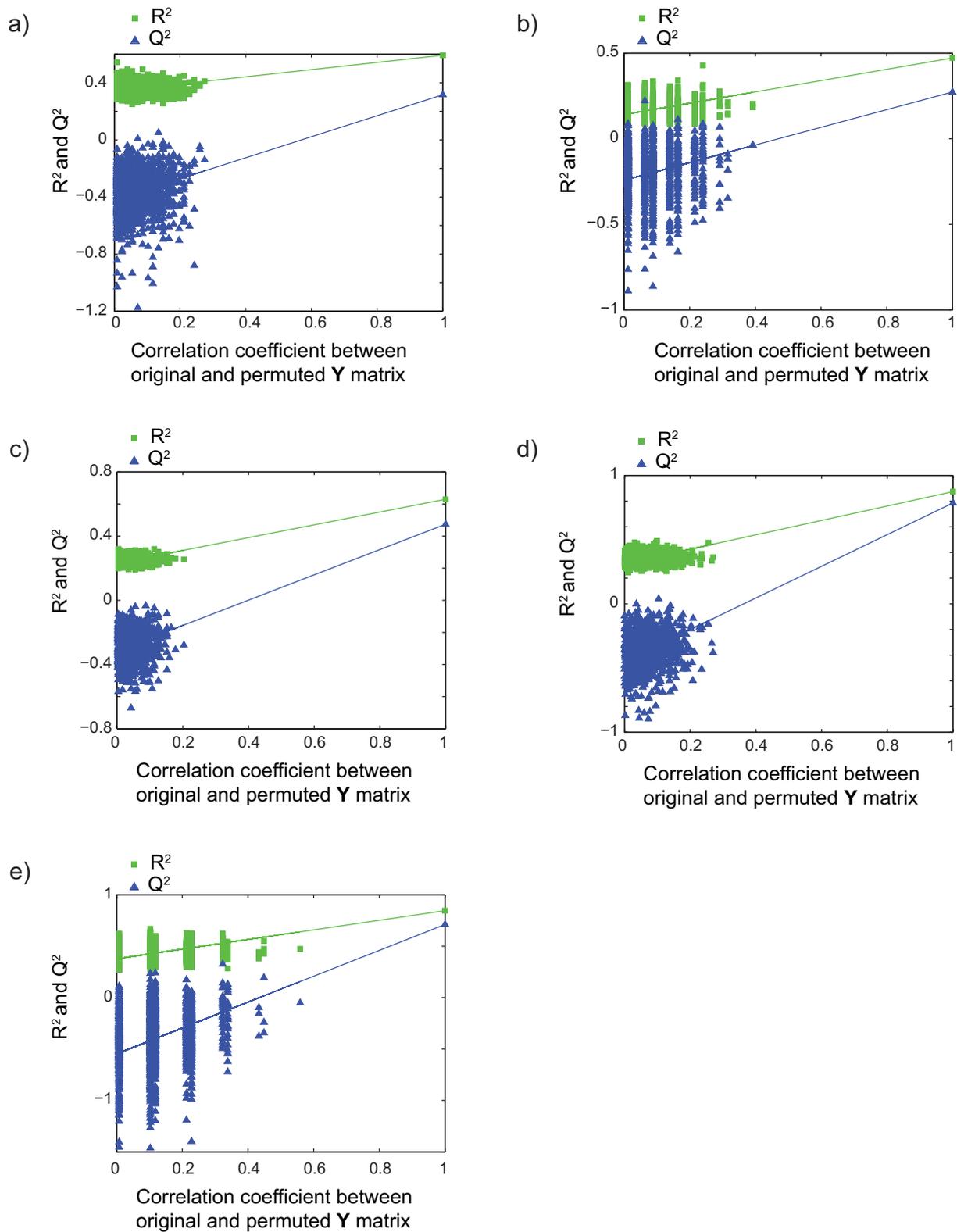


Figure 4.3.2. Model validations resulting from 1000 permutations, demonstrating model robustness for discrimination between a) males and females, b) high and low BMI, c) years of collection (2008 vs 2009), d) correlation between albumin levels and NMR dataset, e) discrimination between HCC and Control after matching on gender, BMI, years of collection and albumin levels.

4.4 Natural metabotype variations in healthy control inbred rat strains

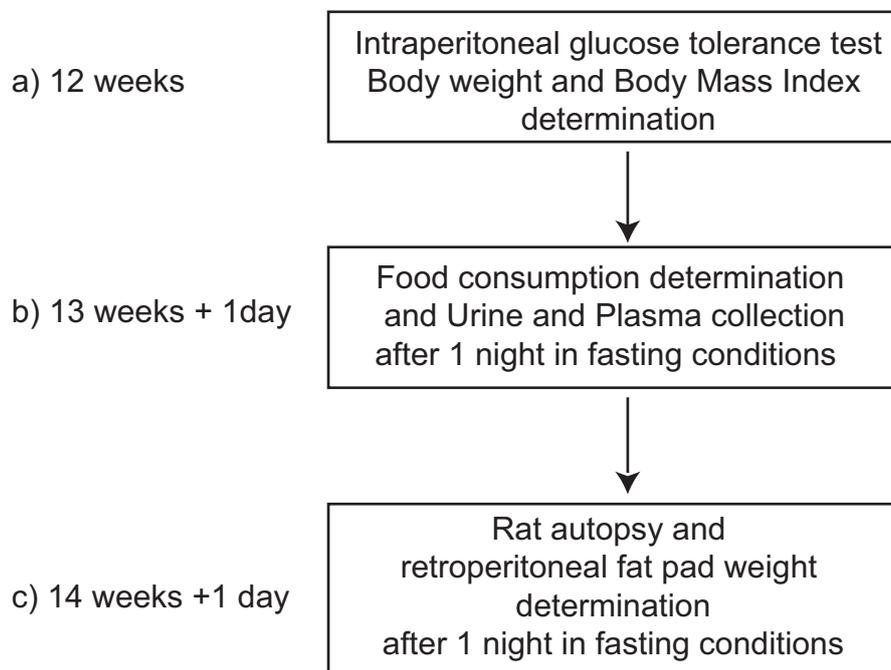


Figure 4.4.1. Summary of animal experimental procedures, physiological and analytical assays.

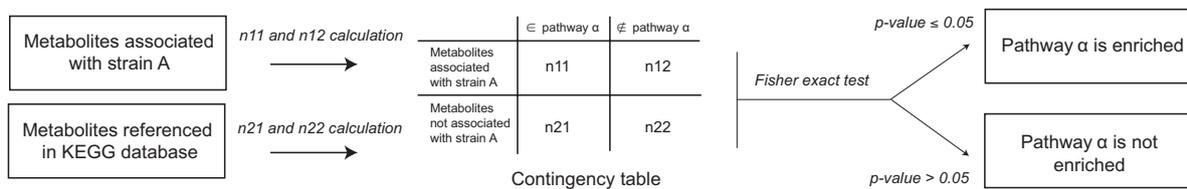


Figure 4.4.2. Flowchart for Metabolite-Set Enrichment Analysis.

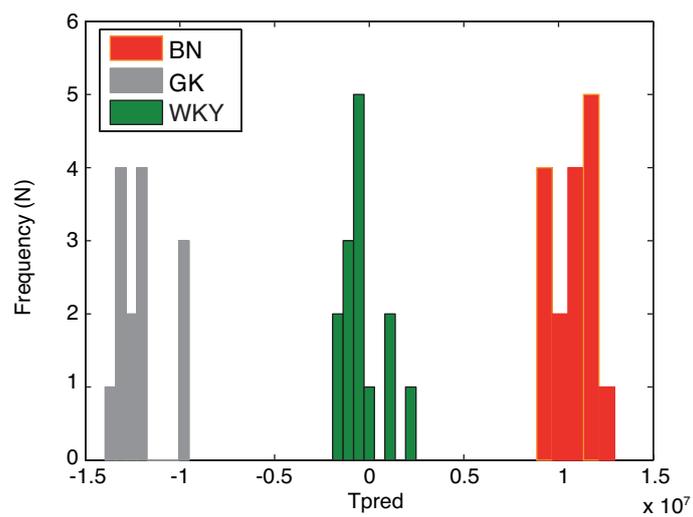


Figure 4.4.3. Evidence for metabonomic-based metabolic hypervariability in inbred rat strains. Projections are shown for BN ($n = 13$), WKY ($n = 14$), and GK ($n = 16$) rats on predictive component from an OPLS-DA model using plasma ¹H NMR spectra obtained at 600 MHz.

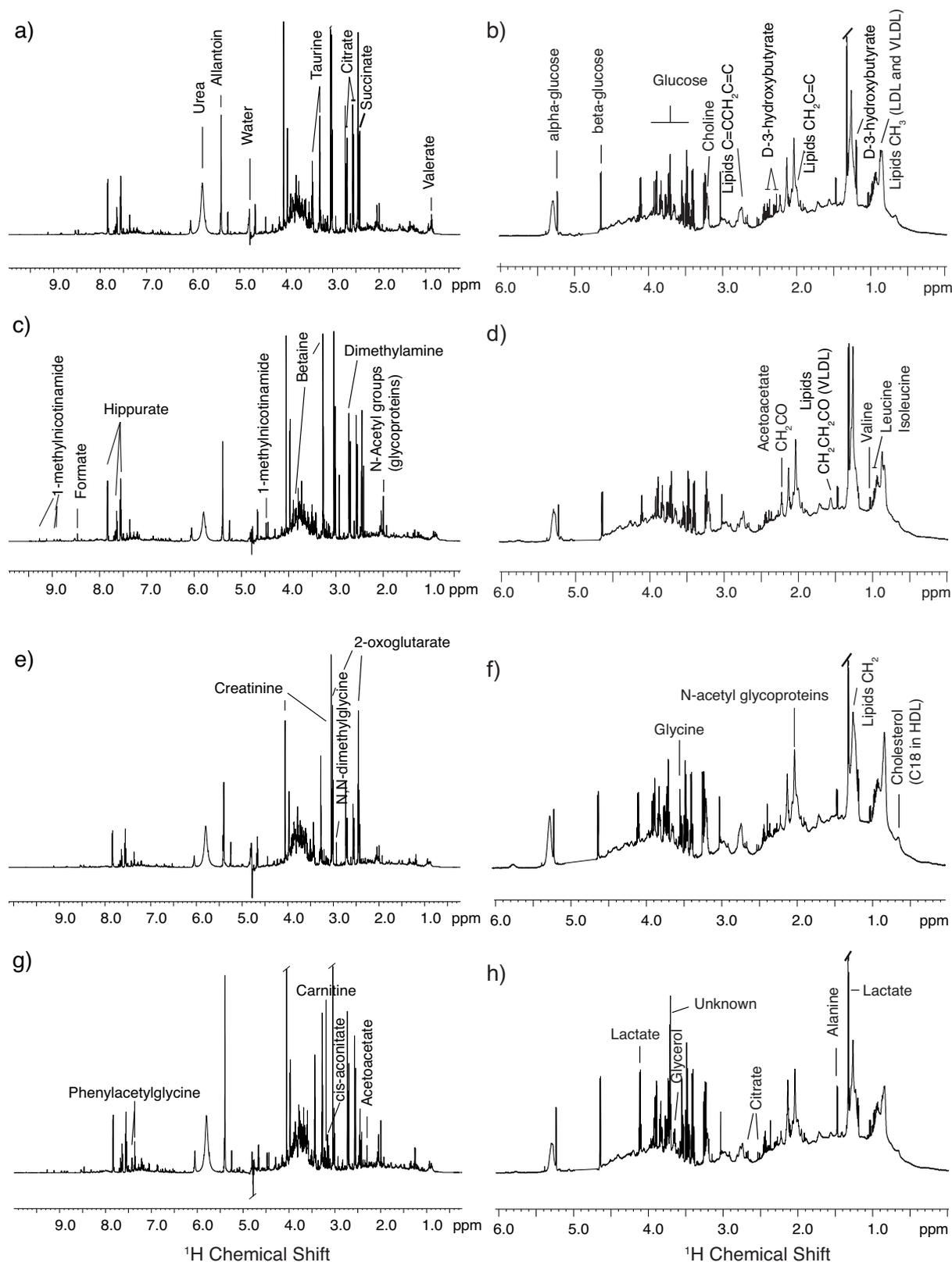


Figure 4.4.4. ^1H 700 MHz NMR spectra of urines from a) LEW, c) F344, e) WKY and g) BN strains, and of plasma from b) LEW, d) F344, f) WKY strain and h) BN strains.

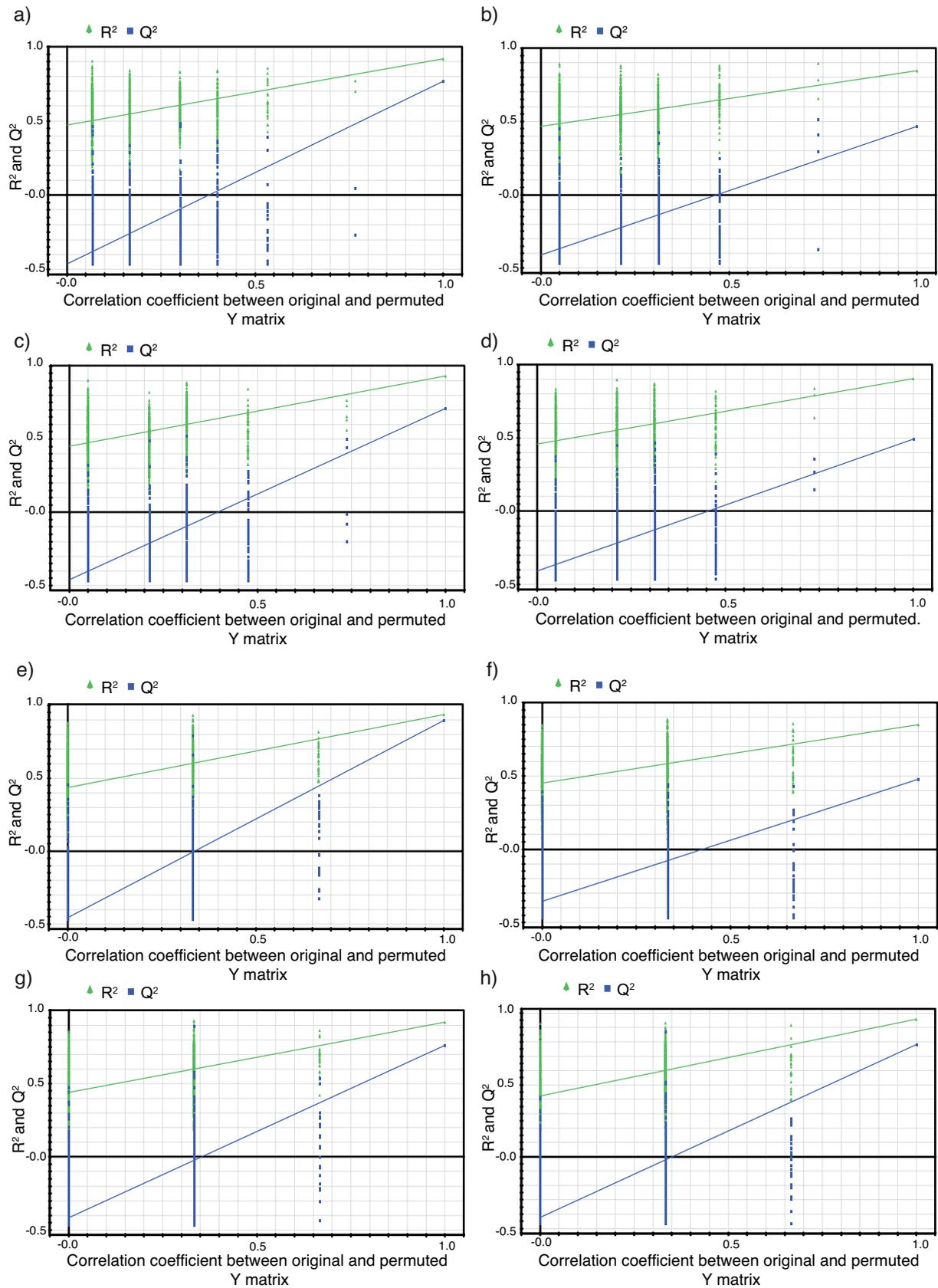


Figure 4.4.5. Cross-validation with 999 permutations for urine OPLS-DA models: a) BN, b) LEW, c) F344 and d) WKY strains; and for plasma OPLS-DA models: e) BN, f) Lewis, g) Fisher and h) WKY plasmas.

Table 4.4.1. Physiological data in inbred rat strains

| | BN | F344 | LEW | WKY | GK |
|-------------------------------|--------------------------------|-------------------------------|---------------------------------|-----------------------------------|-------------------------------------|
| Body weight (g) | 243.8±3.0 (68) | 259.2±8.4 (5) | 309.9±5.0 (5) ^{bd} | 316.9±7.2 (5) ^{ce} | 299.9±6.5 (12) ^{gj} |
| BMI (g.cm ⁻²) | 0.499±0.004 (35) | 0.492±0.027 (5) | 0.602±0.012 (5) ^b | 0.639±0.024 (5) ^{ce} | 0.562±0.013 (12) ^{gj} |
| RFP weight (g) | 0.600±0.041 (7) | 0.932±0.121 (5) | 1.552±0.093 (5) ^d | 3.861±0.487 (5) ^{cef} | 5.213±0.410 (12) ^{ghij} |
| Adiposity index (x1000) | 2.53±0.16 (7) ^{bc} | 3.63±0.53 (5) ^e | 4.99±0.23 (5) ^{bf} | 12.11±1.30 (5) ^{cef} | 17.65±1.71 (12) ^{ghij} |
| Food intake (g) | 15.0±1.4 (53) | 17.6±1.2 (5) | 20.1±0.8 (5) ^b | 17.9±0.1 (5) ^c | 21.2±0.4 (5) ^j |

Body weight, body mass index (BMI), retroperitoneal fat pad (RFP) weight and adiposity index (AI) in 12 week old rats of various inbred strains. RFP weight was used for AI calculation. Data are means ± SEM. Number of rats is reported in parentheses. Differences between groups were assessed by Fisher's LSD and Tamhane's T2 post hoc tests according to Levene's test for variance. Significant differences (P<0.05) between strains are shown as: a, BN vs F344; b, BN vs Lew; c, BN vs WKY; d, F344 vs Lew; e, F344 vs WKY; f, Lew vs WKY; g, GK vs F344; h, GK vs Lew; i, GK vs WKY; j, GK vs BN.

4.5 On metabolic phenotyping of HBV-infected liver cells

Cell lines and cultures. HepG2 (ATCC HB-8065) cells were cultured at 37°C under 5% CO₂, in MEM (minimum essential medium (Invitrogen™, Carlsbad, CA)) supplemented with 10% fetal calf serum (Invitrogen™), 100 UI/ml penicillin, 100 mg/ml streptomycin, 2mM glutamin (Sigma–Aldrich, St. Louis, MO), non-essential amino acids and 1mM sodium pyruvate (both from Invitrogen™). HepG2/2.2.15 cells were cultured at 37°C under 5% CO₂, in DMEM high glucose (4.5 g/l) plus 10% fetal calf serum (Invitrogen™), 100 UI/ml penicillin, 100 mg/ml streptomycin, 2mM glutamin (Sigma–Aldrich, St. Louis, MO), and 2mM sodium pyruvate (both from Invitrogen™) with a final concentration of 200 mg/L G418.

RNA silencing. HepG2/2.2.15 cells were plated in six-well plates with a concentration of 3.10⁵ cells per well. After 24 h, cells were transfected in Opti-MEM (Invitrogen™) containing HiPerfect (6ml/well) (Qiagen) and 20nM p53 small interfering RNA (siRNA) (5'-CAAUGGUUCACUGAAGACCTT-3# and 5'-GGUCUUCAGUGAACCAUUGTT-3#) or a control (Scramble) siRNA (5'-CAUAGAUUACCGUGAGCACTT-3# and 5'-GUGCUCACGGUAAUCUAUGTT-3#) and/or 9.5nM HBx siRNA (5'-ACAUAAGAGGACUCUUGGATT-3# and 5'-UCCAAGAGUCCUCUUAUGUTT-3#) (Eurogentec, Seraing, Belgium), or 20nM E4F1 siRNA (5'-GGAUUUUGUUCAGCACAAGTT-3# and 5'-CUUGUGUCUGAACAAAAUCCTC-3#) (Applied Biosystems, Ambion). Cells were recovered and analyzed 24 or 48 hs after transfection.

4.6 Personal contribution to the different projects

| Projects | Sample Production | NMR analysis | Data and statistical analysis | Interpretation |
|--|-------------------|--------------|-------------------------------|----------------|
| Targeted projection NMR spectroscopy | T | C | T | T |
| Epic Pilot Study | - | T | T | C |
| Metabolic signature of metastatic breast cancer | - | C | C | C |
| Metabolic signatures of liver pathologies | - | T | T | T |
| Metabolic variations in rat strains | - | T | C | C |
| Aging and dietary restriction in <i>C. elegans</i> | - | T | T | C |
| Characterization of the <i>ahr-1 C. elegans</i> mutant | - | T | T | T |
| Characterization of HepG2/2.2.15 cells | - | T | T | C |

“-”: no contribution, “C”: performed in collaboration, “T”: performed completely

Most of these projects were collaborations between the CRMN, specialized in NMR spectroscopy and data analysis and clinicians or biologists. With a background in Chemistry and following medical studies, I tried in during this thesis to make a link between these different expertises.

References

1. Beyond blueprints. *Nat. Chem. Biol.* **2012**, 8 (6), 495.
2. Dunn, W. B.; Broadhurst, D. I.; Atherton, H. J.; Goodacre, R.; Griffin, J. L., Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chem. Soc. Rev.* **2011**, 40 (1), 387-426.
3. Oliver, S. G.; Winson, M. K.; Kell, D. B.; Baganz, F., Systematic functional analysis of the yeast genome. *Trends Biotechnol.* **1998**, 16 (9), 373-8.
4. Fiehn, O., Metabolomics-the link between genotypes and phenotypes. *Plant Mol. Biol.* **2002**, 48 (1-2), 155-71.
5. Nicholson, J. K.; Lindon, J. C.; 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.
6. Nicholson, J. K.; Lindon, J. C., Systems biology - Metabonomics. *Nature* **2008**, 455 (7216), 1054-1056.
7. Dalgliesh, C. E.; Horning, E. C.; Horning, M. G.; Knox, K. L.; Yarger, K., A gas-liquid-chromatographic procedure for separating a wide range of metabolites occurring in urine or tissue extracts. *Biochem. J.* **1966**, 101 (3), 792-810.
8. Horning, E. C.; Horning, M. G., Metabolic profiles: gas-phase methods for analysis of metabolites. *Clin. Chem.* **1971**, 17 (8), 802-9.
9. Mamer, O. A.; Crawhall, J. C.; Tjoa, S. S., The identification of urinary acids by couples gas chromatography-mass spectrometry. *Clin. Chim. Acta* **1971**, 32 (2), 171-184.
10. Pauling, L.; Robinson, a. B.; Teranishi, R.; Cary, P., Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography. *Proc. Natl. Acad. Sci. U. S. A.* **1971**, 68 (10), 2374-6.
11. Goffeau, A.; Barrell, B.; Bussey, H.; Davis, R.; Dujon, B.; Feldmann, H.; Galibert, F.; Hoheisel, J.; Jacq, C.; Johnston, M., *et al.*, Life with 6000 Genes. *Science* **1996**, 274 (5287), 546-567.

12. Venter, J. C.; Adams, M. D.; Myers, E. W.; Li, P. W.; Mural, R. J.; Sutton, G. G.; Smith, H. O.; Yandell, M.; Evans, C. a.; Holt, R. a., *et al.*, The sequence of the human genome. *Science* **2001**, *291* (5507), 1304-51.
13. Velculescu, V. E.; Zhang, L.; Zhou, W.; Vogelstein, J.; Basrai, M. a.; Bassett, D. E.; Hieter, P.; Vogelstein, B.; Kinzler, K. W., Characterization of the yeast transcriptome. *Cell* **1997**, *88* (2), 243-51.
14. Wilkins, M.; Pasquali, C.; Appel, R.; Ou, K.; Golaz, O.; Sanchez, J.; Yan, J.; Gooley, A.; Hughes, G.; HumpherySmith, I., *et al.*, From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology* **1996**, *14* (1), 61-65.
15. Griffiths, W. J.; Koal, T.; Wang, Y.; Kohl, M.; Enot, D. P.; Deigner, H.-P., Targeted metabolomics for biomarker discovery. *Angew. Chem., Int. Ed.* **2010**, *49* (32), 5426-45.
16. Psychogios, N.; Hau, D. D.; Peng, J.; Guo, A. C.; Mandal, R.; Bouatra, S.; Sinelnikov, I.; Krishnamurthy, R.; Eisner, R.; Gautam, B., *et al.*, The human serum metabolome. *PLoS One* **2011**, *6* (2), e16957.
17. Beckonert, O.; Keun, H. C.; Ebbels, T. M. D.; Bundy, J.; Holmes, E.; Lindon, J. C.; Nicholson, J. K., Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat. Protoc.* **2007**, *2* (11), 2692-703.
18. Dunn, W. B.; Broadhurst, D.; Begley, P.; Zelena, E.; Francis-McIntyre, S.; Anderson, N.; Brown, M.; Knowles, J. D.; Halsall, A.; Haselden, J. N., *et al.*, Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* **2011**, *6* (7), 1060-83.
19. Want, E. J.; Wilson, I. D.; Gika, H.; Theodoridis, G.; Plumb, R. S.; Shockcor, J.; Holmes, E.; Nicholson, J. K., Global metabolic profiling procedures for urine using UPLC-MS. *Nat. Protoc.* **2010**, *5* (6), 1005-18.
20. Chan, E. C. Y.; Pasikanti, K. K.; Nicholson, J. K., Global urinary metabolic profiling procedures using gas chromatography-mass spectrometry. *Nat. Protoc.* **2011**, *6* (10), 1483-99.
21. Duarte, I. F.; Goodfellow, B. J.; Barros, S.; Carreira, I. M.; Couceiro, A. B.; Spraul, M.; Gil, A. M., Potential of NMR Spectroscopy for the Study of Human Amniotic Fluid. *Anal. Chem.* **2007**, *79* (21), 8367-8375.
22. Nicholson, J. K.; Wilson, I. D., High resolution proton magnetic resonance spectroscopy of biological fluids. *Prog. Nucl. Magn. Reson. Spectrosc.* **1989**, *21* (4-5), 449-501.
23. Bertram, H. C.; Eggers, N.; Eller, N., Potential of human saliva for nuclear magnetic resonance-based metabolomics and for health-related biomarker identification. *Anal. Chem.* **2009**, *81* (21), 9188-93.
24. Piñero-Sagredo, E.; Nunes, S.; de Los Santos, M. J.; Celda, B.; Esteve, V., NMR metabolic profile of human follicular fluid. *NMR Biomed.* **2010**, *23* (5), 485-95.
25. Beckonert, O.; Coen, M.; Keun, H. C.; Wang, Y.; Ebbels, T. M. D.; Holmes, E.; Lindon, J. C.; Nicholson, J. K., High-resolution magic-angle-spinning NMR spectroscopy for metabolic profiling of intact tissues. *Nat. Protoc.* **2010**, *5* (6), 1019-32.
26. Duarte, I. F.; Marques, J.; Ladeirinha, A. F.; Rocha, C.; Lamego, I.; Calheiros, R.; Silva, T. M.; Marques, M. P. M.; Melo, J. B.; Carreira, I. M., *et al.*, Analytical approaches toward successful human cell metabolome studies by NMR spectroscopy. *Anal. Chem.* **2009**, *81* (12), 5023-32.
27. Blaise, B. J.; Giacomotto, J.; Elena, B.; Dumas, M.-E.; Toulhoat, P.; Ségalat, L.; Emsley, L., Metabotyping of *Caenorhabditis elegans* reveals latent phenotypes. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (50), 19808-12.

28. Sellick, C. a.; Hansen, R.; Stephens, G. M.; Goodacre, R.; Dickson, A. J., Metabolite extraction from suspension-cultured mammalian cells for global metabolite profiling. *Nat. Protoc.* **2011**, *6* (8), 1241-9.
29. De Vos, R. C. H.; Moco, S.; Lommen, A.; Keurentjes, J. J. B.; Bino, R. J.; Hall, R. D., Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* **2007**, *2* (4), 778-91.
30. Geier, F. M.; Want, E. J.; Leroi, A. M.; Bundy, J. G., Cross-Platform Comparison of *Caenorhabditis elegans* Tissue Extraction Strategies for Comprehensive Metabolome Coverage. *Anal. Chem.* **2011**, *7* (8), 3730-3736.
31. Reaves, M. L.; Rabinowitz, J. D., Metabolomics in systems microbiology. *Curr. Opin. Biotechnol.* **2011**, *22* (1), 17-25.
32. Robertson, D. G., Metabonomics in toxicology: a review. *Toxicol. Sci.* **2005**, *85* (2), 809-22.
33. Bundy, J. G.; Davey, M. P.; Viant, M. R., Environmental metabolomics: a critical review and future perspectives. *Metabolomics* **2008**, *5* (1), 3-21.
34. Madsen, R.; Lundstedt, T.; Trygg, J., Chemometrics in metabolomics-a review in human disease diagnosis. *Anal. Chim. Acta* **2010**, *659* (1-2), 23-33.
35. Kinross, J. M.; Holmes, E.; Darzi, A. W.; Nicholson, J. K., Metabolic phenotyping for monitoring surgical patients. *Lancet* **2011**, *377* (9780), 1817-9.
36. Holmes, E.; Loo, R. L.; Stamler, J.; Bictash, M.; Yap, I. K. S.; Chan, Q.; Ebbels, T.; De Iorio, M.; Brown, I. J.; Veselkov, K. a., *et al.*, Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature* **2008**, *453* (7193), 396-400.
37. Wang, T. J.; Larson, M. G.; Vasan, R. S.; Cheng, S.; Rhee, E. P.; McCabe, E.; Lewis, G. D.; Fox, C. S.; Jacques, P. F.; Fernandez, C., *et al.*, Metabolite profiles and the risk of developing diabetes. *Nat. Med.* **2011**, *17* (4), 448-53.
38. Suhre, K.; Wallaschofski, H.; Raffler, J.; Friedrich, N.; Haring, R.; Michael, K.; Wasner, C.; Krebs, A.; Kronenberg, F.; Chang, D., *et al.*, A genome-wide association study of metabolic traits in human urine. *Nat. Genet.* **2011**, *43* (6), 565-9.
39. Suhre, K.; Shin, S.-Y.; Petersen, A.-K.; Mohnney, R. P.; Meredith, D.; Wägele, B.; Altmaier, E.; Deloukas, P.; Erdmann, J.; Grundberg, E., *et al.*, Human metabolic individuality in biomedical and pharmaceutical research. *Nature* **2011**, *477* (7362), 54-60.
40. Clayton, T. A.; Lindon, J. C.; Cloarec, O.; Antti, H.; Charuel, C.; Hanton, G.; Net, L.; Baker, D.; Walley, R. J.; Everett, J. R., *et al.*, Pharmaco-metabonomic phenotyping and personalized drug treatment. *Nature* **2006**, *440* (7087), 1073-1077.
41. Sreekumar, A.; Poisson, L. M.; Rajendiran, T. M.; Khan, A. P.; Cao, Q.; Yu, J.; Laxman, B.; Mehra, R.; Lonigro, R. J.; Li, Y., *et al.*, Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature* **2009**, *457* (7231), 910-4.
42. Jain, M.; Nilsson, R.; Sharma, S.; Madhusudhan, N.; Kitami, T.; Souza, a. L.; Kafri, R.; Kirschner, M. W.; Clish, C. B.; Mootha, V. K., Metabolite Profiling Identifies a Key Role for Glycine in Rapid Cancer Cell Proliferation. *Science* **2012**, *336* (6084), 1040-1044.
43. Lindon, J. C.; Beckonert, O. P.; Holmes, E.; Nicholson, J. K., High-resolution magic angle spinning NMR spectroscopy: Application to biomedical studies. *Prog. Nucl. Magn. Reson. Spectrosc.* **2009**, *55* (2), 79-100.
44. Blaise, B. J.; Giacomotto, J.; Triba, M. N.; Toulhoat, P.; Piotto, M.; Emsley, L.; Ségalat, L.; Dumas, M.-E.; Elena, B., Metabolic profiling strategy of *Caenorhabditis elegans* by whole-organism nuclear magnetic resonance. *J. Proteome Res.* **2009**, *8* (5), 2542-50.
45. Andrew, E.; Bradbury, A.; Eades, R., Removal of dipolar broadening of nuclear magnetic resonance spectra of solids by specimen rotation. *Nature* **1959**, *183* (4678), 1802-1803.

46. Lowe, I., Free induction decays of rotating solids. *Phys. Rev. Lett.* **1959**, 2 (7), 285-287.
47. Cheng, L.; Ma, M.; Becerra, L.; Ptak, T.; Tracey, I.; Lackner, A.; Gonzalez, R., Quantitative neuropathology by high resolution magic angle spinning proton magnetic resonance spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, 94 (12), 6408-6413.
48. McKay, R. T., How the 1D-NOESY Suppresses Solvent Signal in Metabonomics NMR Spectroscopy : An Examination of the Pulse Sequence Components and Evolution. *Concepts Magn. Reson., Part A* **2011**, 38A (5), 197-220.
49. Cui, Q.; Lewis, I. a.; Hegeman, A. D.; Anderson, M. E.; Li, J.; Schulte, C. F.; Westler, W. M.; Eghbalnia, H. R.; Sussman, M. R.; Markley, J. L., Metabolite identification via the Madison Metabolomics Consortium Database. *Nat. Biotechnol.* **2008**, 26 (2), 162-4.
50. Wishart, D. S.; Knox, C.; Guo, A. C.; Eisner, R.; Young, N.; Gautam, B.; Hau, D. D.; Psychogios, N.; Dong, E.; Bouatra, S., *et al.*, HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.* **2009**, 37 (Database issue), D603-10.
51. Ulrich, E. L.; Akutsu, H.; Doreleijers, J. F.; Harano, Y.; Ioannidis, Y. E.; Lin, J.; Livny, M.; Mading, S.; Maziuk, D.; Miller, Z., *et al.*, BioMagResBank. *Nucleic Acids Res.* **2008**, 36 (Database issue), D402-8.
52. Bodenhausen, G.; Ruben, D. J., NATURAL ABUNDANCE N-15 NMR BY ENHANCED HETERONUCLEAR SPECTROSCOPY. *Chem. Phys. Lett.* **1980**, 69 (1), 185-189.
53. Braunschweiler, L.; Ernst, R. R., COHERENCE TRANSFER BY ISOTROPIC MIXING - APPLICATION TO PROTON CORRELATION SPECTROSCOPY. *J. Magn. Reson.* **1983**, 53 (3), 521-528.
54. Aue, W.; Karhan, J.; Ernst, R. R., Homonuclear broad-band decoupling and 2-dimensional J-resolved NMR spectroscopy. *J. Chem. Phys.* **1976**, 64 (10), 4226-27.
55. Nicholson, J. K.; Foxall, P. J.; Spraul, M.; Farrant, R. D.; Lindon, J. C., 750 MHz ¹H and ¹H-¹³C NMR spectroscopy of human blood plasma. *Anal. Chem.* **1995**, 67 (5), 793-811.
56. Bax, A.; Davis, D. G., MLEV-17-BASED TWO-DIMENSIONAL HOMONUCLEAR MAGNETIZATION TRANSFER SPECTROSCOPY. *J. Magn. Reson.* **1985**, 65 (2), 355-360.
57. Xi, Y.; de Ropp, J. S.; Viant, M. R.; Woodruff, D. L.; Yu, P., Improved identification of metabolites in complex mixtures using HSQC NMR spectroscopy. *Anal. Chim. Acta* **2008**, 614 (2), 127-33.
58. Lewis, I. A.; Schommer, S. C.; Hodis, B.; Robb, K. A.; Tonelli, M.; Westler, W. M.; Sussman, M. R.; Markley, J. L., Method for determining molar concentrations of metabolites in complex solutions from two-dimensional ¹H-¹³C NMR spectra. *Anal. Chem.* **2007**, 79 (24), 9385-90.
59. Xi, Y. X.; deRopp, J. S.; Viant, M. R.; Woodruff, D. L.; Yu, P., Automated screening for metabolites in complex mixtures using 2D COSY NMR spectroscopy. *Metabolomics* **2006**, 2 (4), 221-233.
60. Sandusky, P.; Raftery, D., Use of semiselective TOCSY and the pearson correlation for the metabonomic analysis of biofluid mixtures: application to urine. *Anal. Chem.* **2005**, 77 (23), 7717-23.
61. Sandusky, P.; Raftery, D., Use of selective TOCSY NMR experiments for quantifying minor components in complex mixtures: application to the metabonomics of amino acids in honey. *Anal. Chem.* **2005**, 77 (8), 2455-63.
62. Cloarec, O.; Dumas, M.-E.; Craig, A.; Barton, R. H.; Trygg, J.; Hudson, J.; Blancher, C.; Gauguier, D.; Lindon, J. C.; Holmes, E., *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.

63. Blaise, B.; Navratil, V.; Domange, C.; Shintu, L.; Dumas, M.-E.; Elena-Herrmann, B.; Emsley, L.; Toulhoat, P., Two-Dimensional Statistical Recoupling for the Identification of Perturbed Metabolic Networks from NMR Spectroscopy. *J. Proteome Res.* **2010**, *9* (9), 4513-4520.
64. Barjat, H.; Morris, G. A.; Swanson, A. G., A three-dimensional DOSY-HMQC experiment for the high-resolution analysis of complex mixtures. *J. Magn. Reson.* **1998**, *131* (1), 131-8.
65. Shanaiah, N.; Desilva, M. A.; Gowda, G. A. N.; Raftery, M. A.; Hainline, B. E.; Raftery, D., Class selection of amino acid metabolites in body Fluids using chemical derivatization and their enhanced C-13 NMR. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (28), 11540-11544.
66. Ye, T.; Mo, H. P.; Shanaiah, N.; Gowda, G. A. N.; Zhang, S. C.; Raftery, D., Chemoselective N-15 Tag for Sensitive and High-Resolution Nuclear Magnetic Resonance Profiling of the Carboxyl-Containing Metabolome. *Anal. Chem.* **2009**, *81* (12), 4882-4888.
67. Hyberts, S. G.; Heffron, G. J.; Tarragona, N. G.; Solanky, K.; Edmonds, K. A.; Luithardt, H.; Fejzo, J.; Chorev, M.; Aktas, H.; Colson, K., *et al.*, Ultrahigh-resolution (1)H-(13)C HSQC spectra of metabolite mixtures using nonlinear sampling and forward maximum entropy reconstruction. *J. Am. Chem. Soc.* **2007**, *129* (16), 5108-16.
68. Giraudeau, P.; Remaud, G. S.; Akoka, S., Evaluation of ultrafast 2D NMR for quantitative analysis. *Anal. Chem.* **2009**, *81* (1), 479-84.
69. Ludwig, C.; Ward, D. G.; Martin, A.; Viant, M. R.; Ismail, T.; Johnson, P. J.; Wakelam, M. J.; Gunther, U. L., Fast targeted multidimensional NMR metabolomics of colorectal cancer. *Magn. Reson. Chem.* **2009**.
70. Zhang, F.; Bruschweiler, R., Robust deconvolution of complex mixtures by covariance TOCSY spectroscopy. *Angew. Chem., Int. Ed.* **2007**, *46* (15), 2639-42.
71. Snyder, D. A.; Zhang, F.; Robinette, S. L.; Bruschweiler-Li, L.; Bruschweiler, R., Non-negative matrix factorization of two-dimensional NMR spectra: Application to complex mixture analysis. *J. Chem. Phys.* **2008**, *128* (5).
72. Bodenhausen, G.; Ernst, R. R., THE ACCORDION EXPERIMENT, A SIMPLE APPROACH TO 3-DIMENSIONAL NMR-SPECTROSCOPY. *J. Magn. Reson.* **1981**, *45* (2), 367-373.
73. Bodenhausen, G.; Ernst, R. R., DIRECT DETERMINATION OF RATE CONSTANTS OF SLOW DYNAMIC PROCESSES BY TWO-DIMENSIONAL ACCORDION SPECTROSCOPY IN NUCLEAR MAGNETIC-RESONANCE. *J. Am. Chem. Soc.* **1982**, *104* (5), 1304-1309.
74. Nagayama, K.; Bachmann, P.; Wuthrich, K.; Ernst, R. R., USE OF CROSS-SECTIONS AND OF PROJECTIONS IN 2-DIMENSIONAL NMR-SPECTROSCOPY. *J. Magn. Reson.* **1978**, *31* (1), 133-148.
75. Ernst, R. R.; Bodenhausen, G.; Wokaun, A., *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*. Clarendon Press: Oxford, 1987.
76. Szyperski, T.; Wider, G.; Bushweller, J. H.; Wuthrich, K., 3D 13C-15N-heteronuclear two-spin coherence spectroscopy for polypeptide backbone assignments in 13C-15N-double-labeled proteins. *J. Biomol. NMR* **1993**, *3* (1), 127-32.
77. Simorre, J. P.; Brutscher, B.; Caffrey, M. S.; Marion, D., Assignment of NMR spectra of proteins using triple-resonance two-dimensional experiments. *J. Biomol. NMR* **1994**, *4* (3), 325-33.
78. Kupce, E.; Freeman, R., Projection-reconstruction technique for speeding up multidimensional NMR spectroscopy. *J. Am. Chem. Soc.* **2004**, *126* (20), 6429-40.
79. Kim, S.; Szyperski, T., GFT NMR, a new approach to rapidly obtain precise high-dimensional NMR spectral information. *J. Am. Chem. Soc.* **2003**, *125* (5), 1385-1393.

80. Coggins, B. E.; Venters, R. A.; Zhou, P., Generalized reconstruction of n-D NMR spectra from multiple projections: application to the 5-D HACACONH spectrum of protein G B1 domain. *J. Am. Chem. Soc.* **2004**, *126* (4), 1000-1.
81. Hiller, S.; Wasmer, C.; Wider, G.; Wuthrich, K., Sequence-specific resonance assignment of soluble nonglobular proteins by 7D APSY-NMR spectroscopy. *J. Am. Chem. Soc.* **2007**, *129* (35), 10823-8.
82. Kupce, E.; Freeman, R., Hyperdimensional NMR spectroscopy. *J. Am. Chem. Soc.* **2006**, *128* (18), 6020-1.
83. Lescop, E.; Brutscher, B., Hyperdimensional protein NMR spectroscopy in peptide-sequence space. *J. Am. Chem. Soc.* **2007**, *129* (39), 11916-7.
84. Hiller, S.; Fiorito, F.; Wuthrich, K.; Wider, G., Automated projection spectroscopy (APSY). *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (31), 10876-81.
85. Eghbalian, H. R.; Bahrami, A.; Tonelli, M.; Hallenga, K.; Markley, J. L., High-resolution iterative frequency identification for NMR as a general strategy for multidimensional data collection. *J. Am. Chem. Soc.* **2005**, *127* (36), 12528-12536.
86. Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A., Nmrpipe - a Multidimensional Spectral Processing System Based on Unix Pipes. *J. Biomol. NMR* **1995**, *6* (3), 277-293.
87. <http://www.unio-nmr.eu>.
88. Volk, J.; Herrmann, T.; Wuthrich, K., Automated sequence-specific protein NMR assignment using the memetic algorithm MATCH. *J. Biomol. NMR* **2008**, *41* (3), 127-138.
89. Fiorito, F.; Herrmann, T.; Damberger, F. F.; Wuthrich, K., Automated amino acid side-chain NMR assignment of proteins using C-13- and N-15-resolved 3D [H-1,H-1]-NOESY. *J. Biomol. NMR* **2008**, *42* (1), 23-33.
90. Veselkov, K. a.; Lindon, J. C.; Ebbels, T. M. D.; Crockford, D.; Volynkin, V. V.; Holmes, E.; Davies, D. B.; Nicholson, J. K., Recursive segment-wise peak alignment of biological 1H NMR spectra for improved metabolic biomarker recovery. *Anal. Chem.* **2009**, *81* (1), 56-66.
91. Savorani, F.; Tomasi, G.; Engelsen, S. B., icoshift: A versatile tool for the rapid alignment of 1D NMR spectra. *J. Magn. Reson.* **2010**, *202* (2), 190-202.
92. Cloarec, O.; Dumas, M. E.; Trygg, J.; Craig, A.; Barton, R. H.; Lindon, J. C.; Nicholson, J. K.; Holmes, E., Evaluation of the orthogonal projection on latent structure model limitations caused by chemical shift variability and improved visualization of biomarker changes in 1H NMR spectroscopic metabonomic studies. *Anal. Chem.* **2005**, *77* (2), 517-26.
93. De Meyer, T.; Sinnaeve, D.; Van Gasse, B.; Tsiportkova, E.; Rietzschel, E. R.; De Buyzere, M. L.; Gillebert, T. C.; Bekaert, S.; Martins, J. C.; Van Criekinge, W., NMR-based characterization of metabolic alterations in hypertension using an adaptive, intelligent binning algorithm. *Anal. Chem.* **2008**, *80* (10), 3783-90.
94. Blaise, B. J.; Shintu, L.; Elena, B.; Emsley, L.; Dumas, M.-E.; Toulhoat, P., Statistical recoupling prior to significance testing in nuclear magnetic resonance based metabonomics. *Anal. Chem.* **2009**, *81* (15), 6242-51.
95. Weljie, A. M.; Newton, J.; Mercier, P.; Carlson, E.; Slupsky, C. M., Targeted profiling: quantitative analysis of 1H NMR metabolomics data. *Anal. Chem.* **2006**, *78* (13), 4430-42.
96. Craig, A.; Cloarec, O.; Holmes, E.; Nicholson, J. K.; Lindon, J. C., Scaling and normalization effects in NMR spectroscopic metabonomic data sets. *Anal. Chem.* **2006**, *78* (7), 2262-7.

97. Dieterle, F.; Ross, A.; Schlotterbeck, G.; Senn, H., Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. *Anal. Chem.* **2006**, *78* (13), 4281-90.
98. van den Berg, R. a.; Hoefsloot, H. C. J.; Westerhuis, J. a.; Smilde, A. K.; van der Werf, M. J., Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC genomics* **2006**, *7*, 142.
99. Wold, S.; Sjöström, M.; Eriksson, L., PLS-regression: a basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58* (2), 109-130.
100. Trygg, J.; Wold, S., Orthogonal projections to latent structures (O-PLS). *J. Chemom.* **2002**, *16* (3), 119-128.
101. Broadhurst, D. I.; Kell, D. B., Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics* **2006**, *2* (4), 171-196.
102. Benjamini, Y.; Hochberg, Y., Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Statist. Soc. B* **1995**, *57* (1), 289-300.
103. Pounds, S. B., Estimation and control of multiple testing error rates for microarray studies. *Brief. Bioinform.* **2006**, *7* (1), 25-36.
104. Benjamini, Y.; Yekutieli, D., The Control of the False Discovery Rate in Multiple Testing under Dependency. *Ann. Stat.* **2001**, *29* (4), 1165-1188.
105. Cottret, L.; Wildridge, D.; Vinson, F.; Barrett, M. P.; Charles, H.; Sagot, M.-F.; Jourdan, F., MetExplore: a web server to link metabolomic experiments and genome-scale metabolic networks. *Nucleic Acids Res.* **2010**, *38* (Web Server issue), W132-7.
106. Xia, J.; Wishart, D. S., MSEA: a web-based tool to identify biologically meaningful patterns in quantitative metabolomic data. *Nucleic Acids Res.* **2010**, *38* (Web Server issue), W71-7.
107. Chagoyen, M.; Pazos, F., MBRole: enrichment analysis of metabolomic data. *Bioinformatics* **2011**, *27* (5), 730-1.
108. Persicke, M.; Rückert, C.; Plassmeier, J.; Stutz, L. J.; Kessler, N.; Kalinowski, J.; Goesmann, A.; Neuweger, H., MSEA: metabolite set enrichment analysis in the MeltDB metabolomics software platform: metabolic profiling of *Corynebacterium glutamicum* as an example. *Metabolomics* **2012**, *8* (2), 310-22.
109. Wild, C. P., Environmental exposure measurement in cancer epidemiology. *Mutagenesis* **2009**, *24* (2), 117-25.
110. Jenab, M.; Slimani, N.; Bictash, M.; Ferrari, P.; Bingham, S. a., Biomarkers in nutritional epidemiology: applications, needs and new horizons. *Hum. Genet.* **2009**, *125* (5-6), 507-25.
111. Vineis, P.; Perera, F., Molecular epidemiology and biomarkers in etiologic cancer research: the new in light of the old. *Cancer Epidemiol. Biomarkers Prevent.* **2007**, *16* (10), 1954-65.
112. Dumas, M.-E.; Maibaum, E. C.; Teague, C.; Ueshima, H.; Zhou, B.; Lindon, J. C.; Nicholson, J. K.; Stamler, J.; Elliott, P.; Chan, Q., *et al.*, Assessment of analytical reproducibility of 1H NMR spectroscopy based metabonomics for large-scale epidemiological research: the INTERMAP Study. *Anal. Chem.* **2006**, *78* (7), 2199-208.
113. Barton, R. H.; Nicholson, J. K.; Elliott, P.; Holmes, E., High-throughput 1H NMR-based metabolic analysis of human serum and urine for large-scale epidemiological studies: validation study. *Int. J. Epidemiol.* **2008**, *37* (1), 31-40.
114. Suhre, K.; Meisinger, C.; Döring, A.; Altmaier, E.; Belcredi, P.; Gieger, C.; Chang, D.; Milburn, M. V.; Gall, W. E.; Weinberger, K. M., *et al.*, Metabolic footprint of diabetes: a multiplatform metabolomics study in an epidemiological setting. *PLoS One* **2010**, *5* (11), e13953.

115. Slimani, N.; Fahey, M.; Welch, a. a.; Wirfält, E.; Stripp, C.; Bergström, E.; Linseisen, J.; Schulze, M. B.; Bamia, C.; Chloptsios, Y., *et al.*, Diversity of dietary patterns observed in the European Prospective Investigation into Cancer and Nutrition (EPIC) project. *Public Health Nutr.* **2002**, *5* (6B), 1311-28.
116. Peluso, M.; Hainaut, P.; Airolidi, L.; Autrup, H.; Dunning, A.; Garte, S.; Gormally, E.; Malaveille, C.; Matullo, G.; Munnia, A., *et al.*, Methodology of laboratory measurements in prospective studies on gene–environment interactions: The experience of GenAir. *Mutat. Res.* **2005**, *574* (1-2), 92-104.
117. Ferlay, J.; Shin, H. R.; Bray, F.; Forman, D.; Mathers, C.; D.M., P. Globocan 2008 v1.2. <http://globocan.iarc.fr>.
118. Blanco, M. A.; Kang, Y., Signaling pathways in breast cancer metastasis - novel insights from functional genomics. *Breast Cancer Res.* **2011**, *13* (2), 206.
119. Caldas, C., Translational genomics in breast cancer. *European Journal of Cancer* **2011**, *47 Suppl 3*, S381-2.
120. Kroll, T.; Odyvanova, L.; Clement, J. H.; Platzer, C.; Naumann, a.; Marr, N.; Höffken, K.; Wölfl, S., Molecular characterization of breast cancer cell lines by expression profiling. *J. Cancer Res. Clin. Oncol.* **2002**, *128* (3), 125-34.
121. Warburg, O., Origin of cancer cells. *Science* **1956**, *123* (3191), 309-314.
122. Henneges, C.; Bullinger, D.; Fux, R.; Friese, N.; Seeger, H.; Neubauer, H.; Laufer, S.; Gleiter, C. H.; Schwab, M.; Zell, A., *et al.*, Prediction of breast cancer by profiling of urinary RNA metabolites using Support Vector Machine-based feature selection. *BMC cancer* **2009**, *9*, 104.
123. Keun, H. C.; Sidhu, J.; Pchejetski, D.; Lewis, J. S.; Marconell, H.; Patterson, M.; Bloom, S. R.; Amber, V.; Coombes, R. C.; Stebbing, J., Serum molecular signatures of weight change during early breast cancer chemotherapy. *Clin. Cancer Res.* **2009**, *15* (21), 6716-23.
124. Zhang, S.; Nagana Gowda, G. a.; Ye, T.; Raftery, D., Advances in NMR-based biofluid analysis and metabolite profiling. *Analyst* **2010**, *135* (7), 1490-8.
125. Griffin, J. L.; Shockcor, J. P., Metabolic profiles of cancer cells. *Nature reviews. Cancer* **2004**, *4* (7), 551-61.
126. Raamsdonk, L. M.; Teusink, B.; Broadhurst, D.; Zhang, N.; Hayes, a.; Walsh, M. C.; Berden, J. a.; Brindle, K. M.; Kell, D. B.; Rowland, J. J., *et al.*, A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat. Biotechnol.* **2001**, *19* (1), 45-50.
127. Odunsi, K.; Wollman, R. M.; Ambrosone, C. B.; Hutson, A.; McCann, S. E.; Tammela, J.; Geisler, J. P.; Miller, G.; Sellers, T.; Cliby, W., *et al.*, Detection of epithelial ovarian cancer using 1H-NMR-based metabonomics. *Int. J. Cancer* **2005**, *113* (5), 782-8.
128. Cheng, L. L.; Chang, I. W.; Smith, B. L.; Gonzalez, R. G., Evaluating human breast ductal carcinomas with high-resolution magic-angle spinning proton magnetic resonance spectroscopy. *J. Magn. Reson.* **1998**, *135* (1), 194-202.
129. Sitter, B.; Sonnewald, U.; Spraul, M.; Fjösne, H. E.; Gribbestad, I. S., High-resolution magic angle spinning MRS of breast cancer tissue. *NMR Biomed.* **2002**, *15* (5), 327-37.
130. Bathen, T. F.; Jensen, L. R.; Sitter, B.; Fjösne, H. E.; Halgunset, J.; Axelson, D. E.; Gribbestad, I. S.; Lundgren, S., MR-determined metabolic phenotype of breast cancer in prediction of lymphatic spread, grade, and hormone status. *Breast Cancer Research Treatment* **2007**, *104* (2), 181-9.
131. Whitehead, T. L.; Monzavi-Karbassi, B.; Kieber-Emmons, T., 1H-NMR metabonomics analysis of sera differentiates between mammary tumor-bearing mice and healthy controls. *Metabolomics* **2005**, *1* (3), 269-278.

132. Slupsky, C. M.; Steed, H.; Wells, T. H.; Dabbs, K.; Schepansky, A.; Capstick, V.; Faught, W.; Sawyer, M. B., Urine metabolite analysis offers potential early diagnosis of ovarian and breast cancers. *Clin. Cancer Res.* **2010**, *16* (23), 5835-41.
133. Asiago, V. M.; Alvarado, L. Z.; Shanaiah, N.; Gowda, G. a. N.; Owusu-Sarfo, K.; Ballas, R. a.; Raftery, D., Early detection of recurrent breast cancer using metabolite profiling. *Cancer Res.* **2010**, *70* (21), 8309-18.
134. Oakman, C.; Tenori, L.; Claudino, W. M.; Cappadona, S.; Nepi, S.; Battaglia, A.; Bernini, P.; Zafarana, E.; Saccenti, E.; Fornier, M., *et al.*, Identification of a serum-detectable metabolomic fingerprint potentially correlated with the presence of micrometastatic disease in early breast cancer patients at varying risks of disease relapse by traditional prognostic methods. *Ann. Oncol.* **2011**, *22* (6), 1295-301.
135. Bollard, M. E.; Stanley, E. G.; Lindon, J. C.; Nicholson, J. K.; Holmes, E., NMR-based metabolomic approaches for evaluating physiological influences on biofluid composition. *NMR Biomed.* **2005**, *18* (3), 143-62.
136. Oakman, C.; Tenori, L.; Biganzoli, L.; Santarpia, L.; Cappadona, S.; Luchinat, C.; Di Leo, A., Uncovering the metabolomic fingerprint of breast cancer. *Int. J. Biochem. Cell Biol.* **2011**, *43* (7), 1010-20.
137. Chan, E. C. Y.; Koh, P. K.; Mal, M.; Cheah, P. Y.; Eu, K. W.; Backshall, A.; Cavill, R.; Nicholson, J. K.; Keun, H. C., Metabolic Profiling of Human Colorectal Cancer Using High-Resolution Magic Angle Spinning Nuclear Magnetic Resonance (HR-MAS NMR) Spectroscopy and Gas Chromatography Mass Spectrometry (GC/MS). *J. Proteome Res.* **2009**, *8* (1), 352-361.
138. 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.
139. Neurauter, G.; Grahmann, A. V.; Klieber, M.; Zeimet, A.; Ledochowski, M.; Sperner-Unterweger, B.; Fuchs, D., Serum phenylalanine concentrations in patients with ovarian carcinoma correlate with concentrations of immune activation markers and of isoprostane-8. *Cancer Lett.* **2008**, *272* (1), 141-7.
140. Qiu, Y.; Cai, G.; Su, M.; Chen, T.; Zheng, X.; Xu, Y.; Ni, Y.; Zhao, A.; Xu, L. X.; Cai, S., *et al.*, Serum metabolite profiling of human colorectal cancer using GC-TOFMS and UPLC-QTOFMS. *J. Proteome Res.* **2009**, *8* (10), 4844-50.
141. Lai, H.-S.; Lee, J.-C.; Lee, P.-H.; Wang, S.-T.; Chen, W.-J., Plasma free amino acid profile in cancer patients. *Semin. Cancer Biol.* **2005**, *15* (4), 267-276.
142. Maeda, J.; Higashiyama, M.; Imaizumi, A.; Nakayama, T.; Yamamoto, H.; Daimon, T.; Yamakado, M.; Imamura, F.; Kodama, K., Possibility of multivariate function composed of plasma amino acid profiles as a novel screening index for non-small cell lung cancer: a case control study. *BMC cancer* **2010**, *10* (1), 690.
143. Qiu, Y.; Cai, G.; Su, M.; Chen, T.; Liu, Y.; Xu, Y.; Ni, Y.; Zhao, A.; Cai, S.; Xu, L. X., *et al.*, Urinary metabolomic study on colorectal cancer. *J. Proteome Res.* **2010**, *9* (3), 1627-34.
144. Garcia-Caballero, M.; Neugebauer, E.; Campos, R.; Nunez de Castro, I.; Vara-Thorbeck, C., Increased histidine decarboxylase (HDC) activity in human colorectal cancer: results of a study on ten patients. *Agents Actions* **1988**, *23* (3-4), 357-60.
145. Berg, J. M.; Tymoczko, J. L.; Stryer, L., Fatty Acid Metabolism. In *Biochemistry*, Freeman, W. H., Ed. New York, 2002; pp 617-732.
146. Bonuccelli, G.; Tsirigos, A.; Whitaker-Menezes, D.; Pavlides, S.; Pestell, R. G.; Chiavarina, B.; Frank, P. G.; Flomenberg, N.; Howell, A.; Martinez-Outschoorn, U. E., *et al.*, Ketones and lactate "fuel" tumor growth and metastasis: Evidence that epithelial cancer cells use oxidative mitochondrial metabolism. *Cell Cycle* **2010**, *9* (17), 3506-3514.

147. Bell, J. D.; Brown, J. C.; Nicholson, J. K.; Sadler, P. J., Assignment of resonances for 'acute-phase' glycoproteins in high resolution proton NMR spectra of human blood plasma. *FEBS Lett.* **1987**, *215* (2), 311-5.
148. Chen, T.; Xie, G.; Wang, X.; Fan, J.; Qiu, Y.; Zheng, X.; Qi, X.; Cao, Y.; Su, M.; Wang, X., *et al.*, Serum and urine metabolite profiling reveals potential biomarkers of human hepatocellular carcinoma. *Mol. Cell. Proteomics* **2011**, *10* (7), M110.004945.
149. Wang, B.; Chen, D.; Chen, Y.; Hu, Z.; Cao, M.; Xie, Q.; Chen, Y.; Xu, J.; Zheng, S.; Li, L., Metabonomic profiles discriminate hepatocellular carcinoma from liver cirrhosis by ultraperformance liquid chromatography-mass spectrometry. *J. Proteome Res.* **2012**, *11* (2), 1217-27.
150. Zhou, L.; Wang, Q.; Yin, P.; Xing, W.; Wu, Z.; Chen, S.; Lu, X.; Zhang, Y.; Lin, X.; Xu, G., Serum metabolomics reveals the deregulation of fatty acids metabolism in hepatocellular carcinoma and chronic liver diseases. *Anal. Bioanal. Chem.* **2012**, *403* (1), 203-13.
151. Zhou, L.; Ding, L.; Yin, P.; Lu, X.; Wang, X.; Niu, J.; Gao, P.; Xu, G., 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**.
152. Chen, J.; Wang, W.; Lv, S.; Yin, P.; Zhao, X.; Lu, X.; Zhang, F.; Xu, G., Metabonomics study of liver cancer based on ultra performance liquid chromatography coupled to mass spectrometry with HILIC and RPLC separations. *Anal. Chim. Acta* **2009**, *650* (1), 3-9.
153. Shariff, M. I. F.; Ladep, N. G.; Cox, I. J.; Williams, H. R. T.; Okeke, E.; Malu, A.; Thillainayagam, A. V.; Crossey, M. M. E.; Khan, S. a.; Thomas, H. C., *et al.*, Characterization of urinary biomarkers of hepatocellular carcinoma using magnetic resonance spectroscopy in a Nigerian population. *J. Proteome Res.* **2010**, *9* (2), 1096-103.
154. Wu, H.; Xue, R.; Dong, L.; Liu, T.; Deng, C.; Zeng, H.; Shen, X., Metabolomic profiling of human urine in hepatocellular carcinoma patients using gas chromatography/mass spectrometry. *Anal. Chim. Acta* **2009**, *648* (1), 98-104.
155. Yang, J.; Xu, G.; Zheng, Y.; Kong, H.; Pang, T.; Lv, S.; Yang, Q., Diagnosis of liver cancer using HPLC-based metabonomics avoiding false-positive result from hepatitis and hepatocirrhosis diseases. *J. Chromatogr., B* **2004**, *813* (1-2), 59-65.
156. Yin, P.; Wan, D.; Zhao, C.; Chen, J.; Zhao, X.; Wang, W.; Lu, X.; Yang, S.; Gu, J.; Xu, G., A metabonomic 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.
157. Ye, G.; Zhu, B.; Yao, Z.; Yin, P.; Lu, X.; Kong, H.; Fan, F.; Jiao, B.; Xu, G., 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.
158. Amathieu, R.; Nahon, P.; Triba, M.; Bouchemal, N.; Trinchet, J.-C.; Beaugrand, M.; Dhonneur, G.; Le Moyec, L., Metabolomic approach by ¹H NMR spectroscopy of serum for the assessment of chronic liver failure in patients with cirrhosis. *J. Proteome Res.* **2011**, *10* (7), 3239-45.
159. Qi, S.-w.; Tu, Z.-g.; Peng, W.-j.; Wang, L.-x.; Ou-yang, X.; Cai, A.-j.; Dai, Y., ¹H NMR-based serum metabolic profiling in compensated and decompensated cirrhosis. *World J. Gastroenterol.* **2012**, *18* (3), 285-290.
160. Wang, X.; Wang, X.; Xie, G.; Zhou, M.; Yu, H.; Lin, Y.; Du, G.; Luo, G.; Jia, W.; Liu, P., Urinary metabolite variation is associated with pathological progression of the post-hepatitis B cirrhosis patients. *J. Proteome Res.* **2012**, *11* (7), 3838-47.

161. Lian, J.; Liu, W.; Hao, S.; Guo, Y.; Huang, H.; Chen, D.; Xie, Q.; Pan, X.; Xu, W.; Yuan, W., *et al.*, A serum metabonomic study on the difference between alcohol- and HBV-induced liver cirrhosis by ultraperformance liquid chromatography coupled to mass spectrometry plus quadrupole time-of-flight mass spectrometry. *Chin. Med. J.* **2011**, *124* (9), 1367-1373.
162. 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 ¹H NMR-based metabonomics. *Hepatol. Res.* **2012**, *42* (7), 677-85.
163. Villar, S.; Ortiz-Cuaran, S.; Abedi-Ardekani, B.; Gouas, D.; Nogueira da Costa, A.; Plymoth, A.; Khuhaprema, T.; Kalalak, A.; Sangrajrang, S.; Friesen, M. D., *et al.*, Aflatoxin-induced TP53 R249S mutation in hepatocellular carcinoma in Thailand: association with tumors developing in the absence of liver cirrhosis. *PLoS One* **2012**, *7* (6), e37707.
164. Burke, M. D., Liver function: test selection and interpretation of results. *Clinics in laboratory medicine* **2002**, *22* (2), 377-90.
165. Mittelstrass, K.; Ried, J. S.; Yu, Z.; Krumsiek, J.; Gieger, C.; Prehn, C.; Roemisch-Margl, W.; Polonikov, A.; Peters, A.; Theis, F. J., *et al.*, Discovery of sexual dimorphisms in metabolic and genetic biomarkers. *PLoS Genet.* **2011**, *7* (8), e1002215.
166. Pietiläinen, K. H.; Sysi-Aho, M.; Rissanen, A.; Seppänen-Laakso, T.; Yki-Järvinen, H.; Kaprio, J.; Oresic, M., Acquired obesity is associated with changes in the serum lipidomic profile independent of genetic effects—a monozygotic twin study. *PLoS One* **2007**, *2* (2), e218.
167. Yu, Z.; Zhai, G.; Singmann, P.; He, Y.; Xu, T.; Prehn, C.; Römisch-Margl, W.; Lattka, E.; Gieger, C.; Soranzo, N., *et al.*, Human Serum Metabolic Profiles Are Age Dependent. *Aging Cell* **2012**.
168. Teahan, O.; Gamble, S.; Holmes, E.; Waxman, J.; Nicholson, J. K.; Bevan, C.; Keun, H. C., Impact of analytical bias in metabonomic studies of human blood serum and plasma. *Anal. Chem.* **2006**, *78* (13), 4307-18.
169. Shariff, M. I. F.; Gomaa, A. I.; Cox, I. J.; Patel, M.; Williams, H. R. T.; Crossey, M. M. E.; Thillainayagam, A. V.; Thomas, H. C.; Waked, I.; Khan, S. a., *et al.*, Urinary metabolic biomarkers of hepatocellular carcinoma in an Egyptian population: a validation study. *J. Proteome Res.* **2011**, *10* (4), 1828-36.
170. Kahn, R.; Buse, J.; Ferrannini, E.; Stern, M., The metabolic syndrome: time for a critical appraisal. Joint statement from the American Diabetes Association and the European Association for the Study of Diabetes. *Diabetologia* **2005**, *48* (9), 1684-99.
171. Hardy, J.; Singleton, A., Genomewide association studies and human disease. *The New England journal of medicine* **2009**, *360* (17), 1759-68.
172. Aitman, T. J.; Critser, J. K.; Cuppen, E.; Dominiczak, A.; Fernandez-Suarez, X. M.; Flint, J.; Gauguier, D.; Geurts, A. M.; Gould, M.; Harris, P. C., *et al.*, Progress and prospects in rat genetics: a community view. *Nat. Genet.* **2008**, *40* (5), 516-22.
173. Gauguier, D.; Samani, N., Approaches to the analysis of complex quantitative phenotypes and marker map construction based on the analysis of rat models of hypertension. *Methods in molecular biology* **2002**, *195*, 225-51.
174. Damerval, C.; Maurice, A.; Josse, J. M.; Devienne, D., Quantitative Trait Loci Underlying Gene-Product Variation - a Novel Perspective for Analyzing Regulation of Genome Expression. *Genetics* **1994**, *137* (1), 289-301.
175. Klose, J.; Nock, C.; Herrmann, M.; Stuhler, K.; Marcus, K.; Bluggel, M.; Krause, E.; Schalkwyk, L. C.; Rastan, S.; Brown, S. D., *et al.*, Genetic analysis of the mouse brain proteome. *Nat. Genet.* **2002**, *30* (4), 385-93.
176. Rockman, M. V.; Kruglyak, L., Genetics of global gene expression. *Nat. Rev. Genet.* **2006**, *7* (11), 862-72.

177. Nicholson, J. K.; Connelly, J.; Lindon, J. C.; Holmes, E., Metabonomics: a platform for studying drug toxicity and gene function. *Nat. Rev. Drug Discovery* **2002**, *1* (2), 153-61.
178. Keurentjes, J. J.; Fu, J.; de Vos, C. H.; Lommen, A.; Hall, R. D.; Bino, R. J.; van der Plas, L. H.; Jansen, R. C.; Vreugdenhil, D.; Koornneef, M., The genetics of plant metabolism. *Nat. Genet.* **2006**, *38* (7), 842-9.
179. Dumas, M. E.; Wilder, S. P.; Bihoreau, M. T.; Barton, R. H.; Fearnside, J. F.; Argoud, K.; D'Amato, L.; Wallis, R. H.; Blancher, C.; Keun, H. C., *et al.*, Direct quantitative trait locus mapping of mammalian metabolic phenotypes in diabetic and normoglycemic rat models. *Nat. Genet.* **2007**, *39* (5), 666-72.
180. Fearnside, J. F.; Dumas, M. E.; Rothwell, A. R.; Wilder, S. P.; Cloarec, O.; Toye, A.; Blancher, C.; Holmes, E.; Tatoud, R.; Barton, R. H., *et al.*, Phylometabonomic patterns of adaptation to high fat diet feeding in inbred mice. *PLoS One* **2008**, *3* (2), e1668.
181. Martin, B.; Ji, S.; Maudsley, S.; Mattson, M. P., "Control" laboratory rodents are metabolically morbid: why it matters. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (14), 6127-33.
182. Burcelin, R.; Crivelli, V.; Dacosta, A.; Roy-Tirelli, A.; Thorens, B., Heterogeneous metabolic adaptation of C57BL/6J mice to high-fat diet. *American journal of physiology. Endocrinology and metabolism* **2002**, *282* (4), E834-42.
183. Johannesson, M.; Lopez-Aumatell, R.; Stridh, P.; Diez, M.; Tuncel, J.; Blazquez, G.; Martinez-Membrives, E.; Canete, T.; Vicens-Costa, E.; Graham, D., *et al.*, A resource for the simultaneous high-resolution mapping of multiple quantitative trait loci in rats: the NIH heterogeneous stock. *Genome Res.* **2009**, *19* (1), 150-8.
184. Solberg, L. C.; Valdar, W.; Gauguier, D.; Nunez, G.; Taylor, A.; Burnett, S.; Arboledas-Hita, C.; Hernandez-Pliego, P.; Davidson, S.; Burns, P., *et al.*, A protocol for high-throughput phenotyping, suitable for quantitative trait analysis in mice. *Mamm. Genome* **2006**, *17* (2), 129-46.
185. Veselkov, K. A.; Pahomov, V. I.; Lindon, J. C.; Volynkin, V. S.; Crockford, D.; Osipenko, G. S.; Davies, D. B.; Barton, R. H.; Bang, J. W.; Holmes, E., *et al.*, A Metabolic Entropy Approach for Measurements of Systemic Metabolic Disruptions in Patho-Physiological States. *J. Proteome Res.* **2010**, *9* (7), 3537-3544.
186. Kitano, H., Towards a theory of biological robustness. *Mol. Syst. Biol.* **2007**, *3*, 137.
187. Wold, S.; Esbensen, K.; Geladi, P., Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2* (1-3), 37-52.
188. Venables, W. N.; Ripley, B. D., *Modern Applied Statistics with S*. 2002.
189. Sneath, P. H. A.; Sokal, R. R., *Numerical Taxonomy: The Principles and Practice of Numerical Classification*. Freeman: San Francisco, 1973.
190. Subramanian, A.; Tamayo, P.; Mootha, V. K.; Mukherjee, S.; Ebert, B. L.; Gillette, M. A.; Paulovich, A.; Pomeroy, S. L.; Golub, T. R.; Lander, E. S., *et al.*, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (43), 15545-50.
191. Rivals, I.; Personnaz, L.; Taing, L.; Potier, M. C., Enrichment or depletion of a GO category within a class of genes: which test? *Bioinformatics* **2007**, *23* (4), 401-7.
192. Saar, K.; Beck, A.; Bihoreau, M. T.; Birney, E.; Brocklebank, D.; Chen, Y.; Cuppen, E.; Demonchy, S.; Dopazo, J.; Flicek, P., *et al.*, SNP and haplotype mapping for genetic analysis in the rat. *Nat. Genet.* **2008**, *40* (5), 560-6.
193. Kanehisa, M.; Goto, S., KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **2000**, *28* (1), 27-30.
194. Valdar, W.; Solberg, L. C.; Gauguier, D.; Cookson, W. O.; Rawlins, J. N.; Mott, R.; Flint, J., Genetic and environmental effects on complex traits in mice. *Genetics* **2006**, *174* (2), 959-84.

195. Rapp, J. P., Genetic analysis of inherited hypertension in the rat. *Physiological Reviews* **2000**, *80* (1), 135-72.
196. Moisan, M. P.; Courvoisier, H.; Bihoreau, M. T.; Gauguier, D.; Hendley, E. D.; Lathrop, M.; James, M. R.; Mormede, P., A major quantitative trait locus influences hyperactivity in the WKHA rat. *Nat. Genet.* **1996**, *14* (4), 471-3.
197. Galli, J.; Li, L. S.; Glaser, A.; Ostenson, C. G.; Jiao, H.; Fakhrai-Rad, H.; Jacob, H. J.; Lander, E. S.; Luthman, H., Genetic analysis of non-insulin dependent diabetes mellitus in the GK rat. *Nat. Genet.* **1996**, *12* (1), 31-7.
198. Abiola, O.; Angel, J. M.; Avner, P.; Bachmanov, A. A.; Belknap, J. K.; Bennett, B.; Blankenhorn, E. P.; Blizard, D. A.; Bolivar, V.; Brockmann, G. A., *et al.*, The nature and identification of quantitative trait loci: a community's view. *Nat. Rev. Genet.* **2003**, *4* (11), 911-6.
199. Simpson, D. P., Citrate excretion: a window on renal metabolism. *Am. J. Physiol.* **1983**, *244* (3), F223-34.
200. Baracnieto, M., Effects of Ph, Calcium, and Succinate on Sodium-Citrate Cotransport in Renal Microvilli. *Am. J. Physiol.* **1984**, *247* (2), F282-F290.
201. Stryer, L.; Berg, J. M.; L., T. J., *Biochemistry*. Sixth Edition ed.; W. H. Freeman and Company: New York, 2006.
202. Waldram, A.; Holmes, E.; Wang, Y.; Rantalainen, M.; Wilson, I. D.; Tuohy, K. M.; McCartney, A. L.; Gibson, G. R.; Nicholson, J. K., Top-down systems biology modeling of host metabotype-microbiome associations in obese rodents. *J. Proteome Res.* **2009**, *8* (5), 2361-75.
203. Dumas, M. E.; Barton, R. H.; Toye, A.; Cloarec, O.; Blancher, C.; Rothwell, A.; Fearnside, J.; Tatoud, R.; Blanc, V.; Lindon, J. C., *et al.*, Metabolic profiling reveals a contribution of gut microbiota to fatty liver phenotype in insulin-resistant mice. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (33), 12511-6.
204. Li, Z.; Vance, D. E., Phosphatidylcholine and choline homeostasis. *J. Lipid Res.* **2008**, *49* (6), 1187-94.
205. Newgard, C. B.; An, J.; Bain, J. R.; Muehlbauer, M. J.; Stevens, R. D.; Lien, L. F.; Haqq, A. M.; Shah, S. H.; Arlotto, M.; Slentz, C. A., *et al.*, A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. *Cell Metab.* **2009**, *9* (4), 311-26.
206. Nicholson, J. K.; Holmes, E.; Wilson, I. D., Gut microorganisms, mammalian metabolism and personalized health care. *Nat. Rev. Microbiol.* **2005**, *3* (5), 431-8.
207. Eckburg, P. B.; Bik, E. M.; Bernstein, C. N.; Purdom, E.; Dethlefsen, L.; Sargent, M.; Gill, S. R.; Nelson, K. E.; Relman, D. A., Diversity of the human intestinal microbial flora. *Science* **2005**, *308* (5728), 1635-8.
208. Turnbaugh, P. J.; Gordon, J. I., An invitation to the marriage of metagenomics and metabolomics. *Cell* **2008**, *134* (5), 708-13.
209. Yap, I. K.; Li, J. V.; Saric, J.; Martin, F. P.; Davies, H.; Wang, Y.; Wilson, I. D.; Nicholson, J. K.; Utzinger, J.; Marchesi, J. R., *et al.*, Metabonomic and microbiological analysis of the dynamic effect of vancomycin-induced gut microbiota modification in the mouse. *J. Proteome Res.* **2008**, *7* (9), 3718-28.
210. al-Waiz, M.; Mikov, M.; Mitchell, S. C.; Smith, R. L., The exogenous origin of trimethylamine in the mouse. *Metabolism* **1992**, *41* (2), 135-6.
211. Claus, S. P.; Tsang, T. M.; Wang, Y.; Cloarec, O.; Skordi, E.; Martin, F. P.; Rezzi, S.; Ross, A.; Kochhar, S.; Holmes, E., *et al.*, Systemic multicompartmental effects of the gut microbiome on mouse metabolic phenotypes. *Mol. Syst. Biol.* **2008**, *4*, 219.
212. Nicholson, J. K.; Holmes, E.; Lindon, J. C.; Wilson, I. D., The challenges of modeling mammalian biocomplexity. *Nat. Biotechnol.* **2004**, *22* (10), 1268-74.

213. Schilling, C. H.; Palsson, B. O., The underlying pathway structure of biochemical reaction networks. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (8), 4193-8.
214. Atanur, S. S.; Birol, I.; Guryev, V.; Hirst, M.; Hummel, O.; Morrissey, C.; Behmoaras, J.; Fernandez-Suarez, X. M.; Johnson, M. D.; McLaren, W. M., *et al.*, The genome sequence of the spontaneously hypertensive rat: Analysis and functional significance. *Genome Res.* **2010**, *20* (6), 791-803.
215. Brenner, S., The Genetics of *Caenorhabditis elegans*. *Genetics* **1974**, *77* (1), 71-94.
216. Brenner, S., Nature's gift to science (Nobel lecture). *ChemBioChem* **2003**, *4* (8), 683-7.
217. Hulme, S. E.; Whitesides, G. M., Chemistry and the worm: *Caenorhabditis elegans* as a platform for integrating chemical and biological research. *Angewandte Chemie-International Edition* **2011**, *50* (21), 4774-807.
218. Altun, Z.; Hall, D. H., Handbook of *C. elegans* Anatomy. *WormAtlas* **2012**, <http://www.wormatlas.org/hermaphrodite/hermaphroditehomepage.htm>.
219. Sulston, J. E.; Horvitz, H. R., Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Dev. Biol.* **1977**, *56* (1), 110-56.
220. White, J.; Southgate, E.; Thomson, J.; Brenner, S., The structure of the nervous-system of the nematode *Caenorhabditis-elegans*. *Philos. Trans. R. Soc. London Ser. B* **1986**, *314* (1165), 1-340.
221. Fire, A.; Xu, S.; Montgomery, M. K.; Kostas, S. A.; Driver, S. E.; Mello, C. C., Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *nature* **1998**, *391* (6669), 806-811.
222. Consortium, T. C. e. S., Genome Sequence of the Nematode *C. elegans*: A Platform for Investigating Biology. *Science* **1998**, *282* (5396), 2012-2018.
223. Braeckman, B. P.; Houthoofd, K.; Vanfleteren, J. R., Intermediary metabolism. In *WormBook*, The *C. elegans* Research Community, W., Ed. 2009.
224. Holt, S. J.; Riddle, D. L., SAGE surveys *C. elegans* carbohydrate metabolism: evidence for an anaerobic shift in the long-lived dauer larva. *Mechanisms of Ageing and Development* **2003**, *124* (7), 779-800.
225. Rea, S.; Johnson, T. E., A Metabolic Model for Life Span Determination in *Caenorhabditis elegans*. *Dev. Cell* **2003**, *5* (2), 197-203.
226. Fuchs, S.; Bundy, J. G.; Davies, S. K.; Viney, J. M.; Swire, J. S.; Leroi, A. M., A metabolic signature of long life in *Caenorhabditis elegans*. *BMC biology* **2010**, *8*, 14.
227. Butler, J. a.; Ventura, N.; Johnson, T. E.; Rea, S. L., Long-lived mitochondrial (Mit) mutants of *Caenorhabditis elegans* utilize a novel metabolism. *FASEB J.* **2010**, *24* (12), 4977-88.
228. Mouchiroud, L.; Molin, L.; Kasturi, P.; Triba, M. N.; Dumas, M. E.; Wilson, M. C.; Halestrap, A. P.; Roussel, D.; Masse, I.; Dalli ere, N., *et al.*, Pyruvate imbalance mediates metabolic reprogramming and mimics lifespan extension by dietary restriction in *Caenorhabditis elegans*. *Aging Cell* **2011**, *10* (1), 39-54.
229. Martin, F.-P. J.; Spanier, B.; Collino, S.; Montoliu, I.; Kolmeder, C.; Giesbertz, P.; Affolter, M.; Kussmann, M.; Daniel, H.; Kochhar, S., *et al.*, Metabotyping of *Caenorhabditis elegans* and their culture media revealed unique metabolic phenotypes associated to amino acid deficiency and insulin-like signaling. *J. Proteome Res.* **2011**, *10* (3), 990-1003.
230. Zeitoun-Ghandour, S.; Leszczyszyn, O. I.; Blindauer, C. a.; Geier, F. M.; Bundy, J. G.; St urzenbaum, S. R., *C. elegans* metallothioneins: response to and defence against ROS toxicity. *Mol. BioSyst.* **2011**, *7* (8), 2397-406.
231. Falk, M. J.; Zhang, Z.; Rosenjack, J. R.; Nissim, I.; Daikhin, E.; Sedensky, M. M.; Yudkoff, M.; Morgan, P. G., Metabolic pathway profiling of mitochondrial respiratory chain mutants in *C. elegans*. *Mol. Genet. Metab.* **2008**, *93* (4), 388-97.

232. Atherton, H. J.; Jones, O. a. H.; Malik, S.; Miska, E. a.; Griffin, J. L., A comparative metabolomic study of NHR-49 in *Caenorhabditis elegans* and PPAR-alpha in the mouse. *FEBS Lett.* **2008**, *582* (12), 1661-6.
233. Castro, C.; Sar, F.; Shaw, W. R.; Mishima, M.; Miska, E. a.; Griffin, J. L., A metabolomic strategy defines the regulation of lipid content and global metabolism by $\Delta 9$ desaturases in *Caenorhabditis elegans*. *BMC genomics* **2012**, *13*, 36.
234. Jones, O. a. H.; Swain, S. C.; Svendsen, C.; Griffin, J. L.; Sturzenbaum, S. R.; Spurgeon, D. J., Potential new method of mixture effects testing using metabolomics and *Caenorhabditis elegans*. *J. Proteome Res.* **2012**, *11* (2), 1446-53.
235. Hughes, S. L.; Bundy, J. G.; Want, E. J.; Kille, P.; Stürzenbaum, S. R., The Metabolomic Responses of *Caenorhabditis elegans* to Cadmium Are Largely Independent of Metallothionein Status , but Dominated by Changes in Cystathionine and Phytochelatins. *J. Proteome Res.* **2009**, *8* (7), 3512-3519.
236. Reinke, S. N.; Hu, X.; Sykes, B. D.; Lemire, B. D., *Caenorhabditis elegans* diet significantly affects metabolic profile, mitochondrial DNA levels, lifespan and brood size. *Mol. Genet. Metab.* **2010**, *100* (3), 274-82.
237. Pungaliya, C.; Srinivasan, J.; Fox, B. W.; Malik, R. U.; Ludewig, A. H.; Sternberg, P. W.; Schroeder, F. C., A shortcut to identifying small molecule signals that regulate behavior and development in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (19), 7708-13.
238. Srinivasan, J.; von Reuss, S. H.; Bose, N.; Zaslaver, A.; Mahanti, P.; Ho, M. C.; O'Doherty, O. G.; Edison, A. S.; Sternberg, P. W.; Schroeder, F. C., A modular library of small molecule signals regulates social behaviors in *Caenorhabditis elegans*. *PLoS Biol.* **2012**, *10* (1), e1001237.
239. von Reuss, S. H.; Bose, N.; Srinivasan, J.; Yim, J. J.; Judkins, J. C.; Sternberg, P. W.; Schroeder, F. C., Comparative metabolomics reveals biogenesis of ascarosides, a modular library of small-molecule signals in *C. elegans*. *J. Am. Chem. Soc.* **2012**, *134* (3), 1817-24.
240. Kenyon, C.; Chang, J.; Gensch, E.; Rudner, A.; Tabtlang, R., A *C. elegans* mutant that lives twice as long as wild type. *Nature* **1993**, *366* (6454), 461-464.
241. Fontana, L.; Partridge, L.; Longo, V. D., Extending healthy life span-from yeast to humans. *Science* **2010**, *328* (5976), 321-6.
242. Wang, Y.; Lawler, D.; Larson, B.; Ramadan, Z.; Kochhar, S.; Holmes, E.; Nicholson, J. K., Metabonomic investigations of aging and caloric restriction in a life-long dog study. *J. Proteome Res.* **2007**, *6* (5), 1846-54.
243. Rezzi, S.; Martin, F.-P. J.; Shanmuganayagam, D.; Colman, R. J.; Nicholson, J. K.; Weindruch, R., Metabolic shifts due to long-term caloric restriction revealed in nonhuman primates. *Experimental gerontology* **2009**, *44* (5), 356-62.
244. Houtkooper, R. H.; Argmann, C.; Houten, S. M.; Cantó, C.; Jenning, E. H.; Andreux, P. a.; Thomas, C.; Doenlen, R.; Schoonjans, K.; Auwerx, J., The metabolic footprint of aging in mice. *Scientific reports* **2011**, *1*, 134.
245. Wijeyesekera, A.; Selman, C.; Barton, R. H.; Holmes, E.; Nicholson, J. K.; Withers, D. J., Metabotyping of long-lived mice using 1H NMR spectroscopy. *J. Proteome Res.* **2012**, *11* (4), 2224-35.
246. Davies, S. K.; Leroi, A. M.; Bundy, J. G., Fluorodeoxyuridine affects the identification of metabolic responses to *daf-2* status in *Caenorhabditis elegans*. *Mechanisms of ageing and development* **2012**, *133* (1), 46-9.
247. Herndon, L. a.; Schmeissner, P. J.; Dudaronek, J. M.; Brown, P. a.; Listner, K. M.; Sakano, Y.; Paupard, M. C.; Hall, D. H.; Driscoll, M., Stochastic and genetic factors influence tissue-specific decline in ageing *C. elegans*. *Nature* **2002**, *419* (6909), 808-14.

248. Lakowski, B.; Hekimi, S., The genetics of caloric restriction in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (22), 13091-6.
249. Solari, F.; Bourbon-Piffaut, A.; Masse, I.; Payrastra, B.; Chan, A. M.-L.; Billaud, M., The human tumour suppressor PTEN regulates longevity and dauer formation in *Caenorhabditis elegans*. *Oncogene* **2005**, *24* (1), 20-7.
250. Al-Saffar, N. M. S.; Jackson, L. E.; Raynaud, F. I.; Clarke, P. a.; Ramírez de Molina, A.; Lacal, J. C.; Workman, P.; Leach, M. O., The phosphoinositide 3-kinase inhibitor PI-103 downregulates choline kinase alpha leading to phosphocholine and total choline decrease detected by magnetic resonance spectroscopy. *Cancer Res.* **2010**, *70* (13), 5507-17.
251. Garcia-Cao, I.; Song, M. S.; Hobbs, R. M.; Laurent, G.; Giorgi, C.; de Boer, V. C. J.; Anastasiou, D.; Ito, K.; Sasaki, A. T.; Rameh, L., *et al.*, Systemic elevation of PTEN induces a tumor-suppressive metabolic state. *Cell* **2012**, *149* (1), 49-62.
252. Ortega-Molina, A.; Efeyan, A.; Lopez-Guadamillas, E.; Muñoz-Martin, M.; Gómez-López, G.; Cañamero, M.; Mulero, F.; Pastor, J.; Martinez, S.; Romanos, E., *et al.*, Pten positively regulates brown adipose function, energy expenditure, and longevity. *Cell Metab.* **2012**, *15* (3), 382-94.
253. Gee, P.; Kent, C., Multiple isoforms of choline kinase from *Caenorhabditis elegans*: cloning, expression, purification, and characterization. *Biochim. Biophys. Acta* **2003**, *1648* (1-2), 33-42.
254. Selman, C.; Kerrison, N. D.; Cooray, A.; Piper, M. D. W.; Lingard, S. J.; Barton, R. H.; Schuster, E. F.; Blanc, E.; Gems, D.; Nicholson, J. K., *et al.*, Coordinated multitissue transcriptional and plasma metabolomic profiles following acute caloric restriction in mice. *Physiological genomics* **2006**, *27* (3), 187-200.
255. Barouki, R.; Coumoul, X.; Fernandez-Salguero, P. M., The aryl hydrocarbon receptor, more than a xenobiotic-interacting protein. *FEBS Lett.* **2007**, *581* (19), 3608-15.
256. Powell-Coffman, J. a.; Bradfield, C. a.; Wood, W. B., *Caenorhabditis elegans* orthologs of the aryl hydrocarbon receptor and its heterodimerization partner the aryl hydrocarbon receptor nuclear translocator. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (6), 2844-9.
257. Qin, H.; Powell-Coffman, J. A., The *Caenorhabditis elegans* aryl hydrocarbon receptor, AHR-1, regulates neuronal development. *Dev. Biol.* **2004**, *270* (1), 64-75.
258. Aarnio, V.; Storvik, M.; Lehtonen, M.; Asikainen, S.; Reisner, K.; Callaway, J.; Rudgalvyte, M.; Lakso, M.; Wong, G., Fatty acid composition and gene expression profiles are altered in aryl hydrocarbon receptor-1 mutant *Caenorhabditis elegans*. *Comparative biochemistry and physiology, Part C* **2010**, *151* (3), 318-24.
259. Ruiz-Aracama, A.; Peijnenburg, A.; Kleinjans, J.; Jennen, D.; van Delft, J.; Hellfrisch, C.; Lommen, A., An untargeted multi-technique metabolomics approach to studying intracellular metabolites of HepG2 cells exposed to 2,3,7,8-tetrachlorodibenzo-p-dioxin. *BMC genomics* **2011**, *12* (1), 251.
260. Forgacs, A. L.; Kent, M. N.; Makley, M. K.; Mets, B.; DelRaso, N.; Jahns, G. L.; Burgoon, L. D.; Zacharewski, T. R.; Reo, N. V., Comparative metabolomic and genomic analyses of TCDD-elicited metabolic disruption in mouse and rat liver. *Toxicol. Sci.* **2012**, *125* (1), 41-55.
261. Lin, S.; Yang, Z.; Shen, Y.; Cai, Z., LC/MS-based non-targeted metabolomics for the investigation of general toxicity of 2,3,7,8-tetrachlorodibenzo-p-dioxin in C57BL/6J and DBA/2J mice. *Int. J. Mass Spectrom.* **2011**, *301* (1-3), 29-36.
262. Lin, S.; Yang, Z.; Zhang, X.; Bian, Z.; Cai, Z., Hippocampal metabolomics reveals 2,3,7,8-tetrachlorodibenzo-p-dioxin toxicity associated with ageing in Sprague-Dawley rats. *Talanta* **2011**, *85* (2), 1007-12.

263. Dai, Y.; Cros, M.-P.; Pontoizeau, C.; Elena-Herrmann, B.; Bonn, G. K.; Hainaut, P., Down-regulation of transcription factor E4F1 in hepatocarcinoma cells: HBx-dependent effects on autophagy, proliferation and metabolism. *submitted*.
264. Le Cam, L.; Linares, L. K.; Paul, C.; Julien, E.; Lacroix, M.; Hatchi, E.; Triboulet, R.; Bossis, G.; Shmueli, A.; Rodriguez, M. S., *et al.*, E4F1 is an atypical ubiquitin ligase that modulates p53 effector functions independently of degradation. *Cell* **2006**, *127* (4), 775-88.
265. Rizos, H.; Diefenbach, E.; Badhwar, P.; Woodruff, S.; Becker, T. M.; Rooney, R. J.; Kefford, R. F., Association of p14ARF with the p120E4F transcriptional repressor enhances cell cycle inhibition. *J. Biol. Chem.* **2003**, *278* (7), 4981-9.
266. Hatchi, E.; Rodier, G.; Lacroix, M.; Caramel, J.; Kirsh, O.; Jacquet, C.; Schrepfer, E.; Lagarrigue, S.; Linares, L. K.; Lledo, G., *et al.*, E4F1 deficiency results in oxidative stress-mediated cell death of leukemic cells. *J. Exp. Med.* **2011**, *208* (7), 1403-17.
267. Lacroix, M.; Caramel, J.; Goguet-Rubio, P.; Linares, L. K.; Estrach, S.; Hatchi, E.; Rodier, G.; Lledo, G.; de Bettignies, C.; Thépot, A., *et al.*, Transcription factor E4F1 is essential for epidermal stem cell maintenance and skin homeostasis. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (49), 21076-21081.
268. Caramel, J.; Lacroix, M.; Le Cam, L.; Sardet, C., E4F1 connects the Bmi1-ARF-p53 pathway to epidermal stem cell-dependent skin homeostasis. *Cell Cycle* **2011**, *10* (6), 866-867.
269. Cuperlović-Culf, M.; Barnett, D. a.; Culf, A. S.; Chute, I., Cell culture metabolomics: applications and future directions. *Drug discovery today* **2010**, *15* (15-16), 610-21.
270. Bayet-Robert, M.; Loiseau, D.; Rio, P.; Demidem, A.; Barthomeuf, C.; Stepien, G.; Morvan, D., Quantitative two-dimensional HRMAS 1H-NMR spectroscopy-based metabolite profiling of human cancer cell lines and response to chemotherapy. *Magn. Reson. Med.* **2010**, *63* (5), 1172-83.
271. Shintu, L.; Baudoin, R.; Navratil, V.; Prot, J.-M.; Pontoizeau, C.; Defernez, M.; Blaise, B. J.; Domange, C.; Péry, A. R.; Toulhoat, P., *et al.*, Metabolomics-on-a-Chip and Predictive Systems Toxicology in Microfluidic Bioartificial Organs. *Anal. Chem.* **2012**, *84* (4), 1840-8.
272. Massimi, M.; Tomassini, A.; Sciubba, F.; Sobolev, A. P.; Devirgiliis, L. C.; Micheli, A., Effects of resveratrol on HepG2 cells as revealed by 1H-NMR based metabolic profiling. *Biochim. Biophys. Acta* **2012**, *1820* (1), 1-8.
273. Birungi, G.; Li, S. F. Y., Investigation of the effect of exposure to non cytotoxic amounts of microcystins. *Metabolomics* **2010**, *7* (4), 485-499.
274. Sells, M. a.; Chen, M. L.; Acs, G., Production of hepatitis B virus particles in Hep G2 cells transfected with cloned hepatitis B virus DNA. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84* (4), 1005-9.
275. Hannon, G. J., RNA interference. *Nature* **2002**, *418* (6894), 244-251.
276. Griffin, J. L.; Mann, C. J.; Scott, J.; Shoulders, C. C.; Nicholson, J. K., Choline containing metabolites during cell transfection: an insight into magnetic resonance spectroscopy detectable changes. *FEBS Lett.* **2001**, *509* (2), 263-6.
277. Waris, G.; Huh, K.-W.; Siddiqui, A., Mitochondrially Associated Hepatitis B Virus X Protein Constitutively Activates Transcription Factors STAT-3 and NF- κ B via Oxidative Stress. *Mol. Cell. Biol.* **2001**, *21* (22), 7721-30.
278. Wang, J.-H.; Yun, C.; Kim, S.; Lee, J.-H.; Yoon, G.; Lee, M.-O.; Cho, H., Reactive oxygen species modulates the intracellular level of HBx viral oncoprotein. *Biochem. Biophys. Res. Commun.* **2003**, *310* (1), 32-39.

Communications

Publications

1. “Targeted Projection NMR Spectroscopy for Unambiguous Metabolic Profiling of Complex Mixtures”, C. Pontoizeau, T. Herrmann, P. Toulhoat, B. Elena-Herrmann and L. Emsley, *Magn. Reson. Chem.* **48**, 727 (2010).
2. “Broad-Ranging Natural Metabotype Variation Drives Physiological Plasticity in Healthy Control Inbred Rat Strains”, C. Pontoizeau, J.F. Fearnside, V. Navratil, C. Domange, J.-B. Cazier, C. Fernandez-Santamaria, P.J. Kaisaki, L. Emsley, P. Toulhoat, M.-T. Bihoreau, J.K. Nicholson, D. Gauguier and M.E. Dumas, *J. Proteome Res.* **10**, 1675 (2011).
3. “Metabolomics-on-a-Chip and Predictive Systems Toxicology in Microfluidic Bioartificial Organs”, L. Shintu, R. Baudoin, V. Navratil, J.M. Prot, C. Pontoizeau, M. Defernez, B.J. Blaise, C. Domange, A.R. Pery, P. Toulhoat, C. Legallais, C. Brochot, E. Leclerc and M.E. Dumas, *Anal. Chem.* **84**, 1840 (2012).
4. “Down-regulation of transcription factor E4F1 in hepatocarcinoma cells: HBx-dependent effects on autophagy, proliferation and metabolism”, Y. Dai, M.P. Cros, C. Pontoizeau, B. Elena-Herrmann, G.K. Bonn and P. Hainaut, *submitted*.

5. “A Serum Nuclear Magnetic Resonance-based Metabolomic Signature of Advanced Metastatic Human Breast Cancer”, E. Jobard, C. Pontoizeau, B.J. Blaise, T. Bachelot, B. Elena-Herrmann and O. Tredan, *submitted*.
6. “Dietary restriction buffers metabolic changes associated to ageing in *Caenorhabditis elegans*”, C. Pontoizeau, L. Mouchiroud, L. Molin, N. Dallièrè, K. Gieseler, B. Elena-Herrmann and F. Solari, *submitted*.

Posters

1. “Metabolic profiling of uropathies by ^1H Nuclear Magnetic Resonance Spectroscopy on human amniotic fluid and paediatric urine”, C. Pontoizeau *et al.*, Metabomeeting 2008, 28-29 April 2008, Lyon, France.
2. “Applications of reduced-dimensionality NMR methods for elucidation of complex mixture composition”, C. Pontoizeau *et al.*, SMASH 2009, 20-23 September 2009, Chamonix, France.
3. “Functional genomics of *Caenorhabditis elegans* by whole organism NMR spectroscopy in applications to ageing and Toxicology”, C. Pontoizeau *et al.*, Metabolomics 2010, 27 June-1 July 2010, Amsterdam, The Netherlands.
4. “Assessing the EPIC blood sample bio-bank for NMR-based metabolomics”, C. Pontoizeau *et al.*, 6èmes Journées Scientifiques du CLARA, 28-29 March 2011, Lyon, France.
5. “Metabolic Phenotyping of *Caenorhabditis elegans* by Whole Organism NMR Spectroscopy, Applications for Functional Genomics in Aging and Toxicology”, C. Pontoizeau *et al.*, International Conference on Systems Biology 2011, 28 August-1 September 2011, Heidelberg, Germany.
6. “Plasma Metabolic Profiling of HBV-related Hepatocellular Carcinoma by ^1H NMR Spectroscopy in a Thai Cohort”, C. Pontoizeau *et al.*, 7èmes Journées Scientifiques du CLARA, 20-21 March 2012, Lyon, France.

Oral presentations

1. “Metabolic Perturbations in *Caenorhabditis elegans* monitored by HR-MAS NMR Spectroscopy, Applications to Toxicology and Aging”, Journée Scientifique de l’Institut de Chimie de Lyon 2011, 28 June 2011, Lyon, France.

2. “Metabolic Phenotyping of *Caenorhabditis elegans* by Whole Organism NMR Spectroscopy, Applications for Functional Genomics in Aging and Toxicology”, Metabomeeting 2012, 25-27 September 2012, Manchester, Great Britain.