



UNIVERSITÀ
DEGLI STUDI
FIRENZE

DOTTORATO DI RICERCA

INTERNATIONAL DOCTORATE
IN STRUCTURAL BIOLOGY

CICLO XXXIII

COORDINATOR Prof. Lucia Banci

¹H-NMR-based metabolomic applications in biomedical research

Settore Scientifico Disciplinare CHIM/03

PhD student

Dott. Cristina Licari

Tutor

Prof. Claudio Luchinat

Coordinator

Prof. Lucia Banci

November 2017 – 2020

***This thesis has been approved by the University of Florence,
the University of Frankfurt and the Utrecht University***



*To whom who cheered me up,
To those who have always encouraged me.*

Abstract

NMR spectroscopy represents a powerful, versatile and reproducible technique for the analysis of complex biological matrices. In fact, virtually, all biologically relevant molecules are characterized by at least one NMR signal with a specific intensity, frequency (or chemical shift) and magnetic relaxation properties, all reflecting the chemical environment surrounding the detected nucleus.

In a high-throughput vision of metabolomic analysis, the very high reproducibility, the minimal sample preparation required, and the possibility to simultaneously detect all metabolites presenting NMR active nuclei, make NMR spectroscopy one of the most suitable techniques for the analysis of any type of biological matrix, enabling the rapid and global evaluation of an NMR spectrum in its entirety or the determination of the concentrations of all metabolic features that are above the μM detection limit.

The NMR versatility allowed a wide variety of metabolomic applications in life science research, especially for both human and veterinary biomedicine. As metabolites indicate intermediate and end-points of gene expression and cell activity, under the combined influence of external stimuli, metabolomics can provide a holistic approach to understand the phenotype of a certain biological system, holding promises for both clinical and precision medicine.

In this context, the presented thesis aims at demonstrating the potential of untargeted NMR-based metabolomic approach in biomedical research, addressing different topics, mainly regarding the use of untargeted NMR-based metabolomic on body fluids to disentangle characteristic fingerprints and/or metabolic markers for different types of both human and animal diseases or healthy conditions; also with the aim of paving the way to personalized individual's healthcare.

Considering the occasional misunderstandings present in the literature about the different aims of "fingerprinting" and "profiling" approaches of the untargeted analysis, and the different tools to achieve them, this thesis also proposes a study where we demonstrate that the criticism on the main drawbacks of the commonly used bucketing procedure of NMR spectra are not valid when an untargeted metabolomic analysis is planned via a fingerprinting approach.

In conclusion, the results presented in this thesis contribute to the demonstration that untargeted NMR-based metabolomics, coupled with biochemistry, analytical chemistry, bioinformatic tools and statistical analysis, can be considered as a comprehensive analytical technique with reasonable and actual prospects of being implemented in biomedical research.

Main abbreviations and acronyms

HPLC: high-performance liquid chromatography

UHPLC: ultra-high-performance liquid chromatography

CSF: cerebrospinal fluid

EBC: exhaled breath condensate

TMSP: trimethylsilylpropanoic acid

PCA: Principal Component Analysis

OPLS-DA: Orthogonal-Partial Least Square Discriminant Analysis

M-PLS: Multilevel Partial Least Squares Analysis

k-NN: k-Nearest Neighbour

SVM: Support Vector Machine

RF: Random Forest

ROC: Receiver Operating Characteristic

AUC: Area under the curve

PQN: Probabilistic Quotient Normalization

FDR: False Discovery Rate

PD: Parkinson's Disease

dnPD: *de novo* Parkinson's disease patients, recently diagnosed

advPD: patients with late PD under pharmacological treatment

CTR: healthy controls

AIS: Acute Ischemic Stroke

rt-PA: recombinant tissue plasminogen activator

3M-nI: not-impaired AIS patients

3M-I: impaired AIS patients

3M-nD: survivors AIS patients

3M-D: deceased AIS patients

PCa: Prostate Cancer

APPs: acute-phase proteins

BO: bilateral orchidectomy

CRCP: castration-resistant prostate cancer

GlycA, GlycB: glycoprotein A, B

MetS: metabolic syndrome

BEF: bioactive-enriched food

DHA: Docosahexaenoic acid (C22:6, *n*-3, DHA)

AC: anthocyanins

O-BG: oat β -glucans

LDA: Left Displaced Abomasum

RDA: Right Displaced Abomasum

Table of contents

Chapter 1	1
Introduction.....	1
1.1. Metabolomics: a tipping point for systems biology	1
1.2. Metabolomics: one science, many experimental approaches.....	2
1.3. NMR and metabolomics.....	4
Chapter 2.....	11
Aims of this thesis	11
Chapter 3.....	13
Methodologies.....	13
3.1. Sample preparation	13
3.2. NMR pulse sequences: tips for 1D metabolomic applications.....	13
3.3. NMR data pre-processing and pre-treatment	16
3.4. Statistical analysis	18
Chapter 4.....	25
Results	25
4.1. NMR-based metabolomics for human biomedicine.....	25
4.1.1. Nuclear Magnetic Resonance-based metabolomics to characterize serum sex-related metabolic profiles of drug-naïve Parkinson’s disease patients with respect to healthy controls and patients with advanced disease and under dopaminergic treatment	31
4.1.2. NMR-based metabolomics to predict three-month adverse outcomes and to estimate metabolic variations in ischemic stroke treated with intravenous thrombolysis.....	63
4.1.3. Differential Network Analysis reveals metabolite and lipid components associated with three-month death and impairment in patients with acute ischemic stroke after thrombolytic treatment with recombinant tissue plasminogen activator	109
4.1.4. Inflammatory metabolic profile of South African patients with prostate cancer.....	135
4.1.5. Untargeted NMR-based metabolomics to investigate the effect of Bioactive Foods enriched with combination of DHA and anthocyanins or oat β -glucan on serum metabolome and lipidome of subjects at risk for metabolic syndrome: Large Intervention Study from the EC FP7 Pathway-27 project...	167
4.2. NMR-based metabolomics for veterinary research.....	193

4.2.1. Nuclear magnetic resonance (NMR)-based metabolome profile evaluation in dairy cows with and without displaced abomasum	195
4.2.2. NMR-based serum metabolomics for monitoring newborn preterm calves' health	215
4.2.3. NMR-based serum extracts metabolomics to evaluate Canine Ehrlichiosis	227
4.3. NMR data pre-processing for metabolomic fingerprinting.....	243
4.3.1. Simple equidistant bucketing as robust and recommended procedure for NMR-based metabolomic fingerprinting.....	247
Chapter 5	275
Conclusions	275
Bibliography.....	279
Appendix A	289
Candidate's contribution	289
PhD publications	289
Appendix B.....	291
Candidate's activities	291

Chapter 1

Introduction

1.1. Metabolomics: a tipping point for systems biology

Nowadays, scientists are increasingly interested in deepening the knowledge and the understanding of biological mechanisms not only at the molecular level, but also focusing on the effects of an ongoing biological process in the organism as a whole. This approach, usually known as *systems biology*, requires the use of physical, molecular, biochemical and chemical approaches to investigate how the interactions among biomolecules impact the functionality of the entire organism.^{1,2}

Omic strategies, such as genomics, transcriptomics and proteomics aim of identifying the whole set of genes, proteins and other biomolecules contained in a biological specimen. For this reason, they have represented real revolutionary tools in the frame of systems biology research. A problem that should not be underestimated is that genomics, protein expression and molecular biology operate on different timescales from one another, making it difficult to find causal linkages.¹ In addition, lifestyle and environmental factors greatly impact on metabolism and it became problematic to unravel their effects from gene-related outcomes.

Metabolomics can overcome these problems. It represents a more recent -omic science which merges optimally with the general philosophy of systems biology; indeed, it provides an integrated picture of biochemistry in a huge variety of organisms (human, animals, plants, microorganisms etc.), dealing with the identification and characterization of $10^3 - 10^4$ different metabolites (small molecules < 1500 Da) in a certain biological sample (*e.g.* cells, tissues, body fluids etc.),³ taking into account both endogenous and exogenous sources of variations (**Figure 1**).

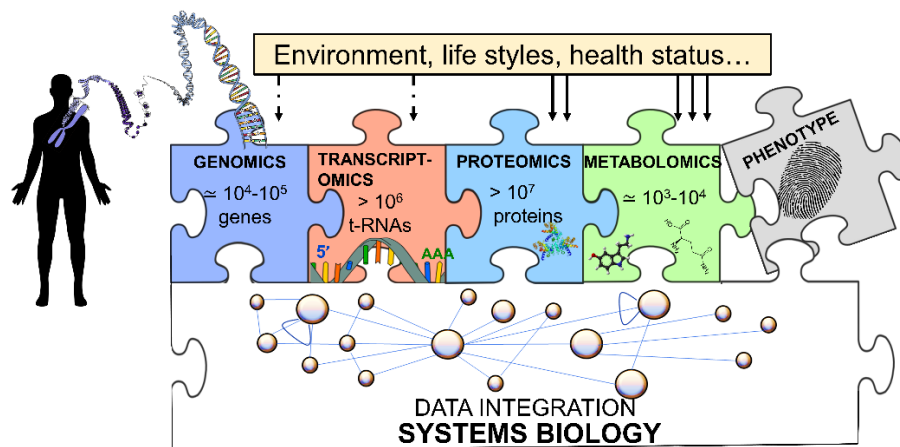


Figure 1. Metabolomics in systems biology: the flow of information proceeds from the genome to the transcriptome, the proteome and finally to the metabolome. From left to right they are increasingly variable during an individual lifespan, and all concur to the definition of the phenotype. Figure taken from Vignoli *et al.*³

Consequently, metabolites, differently from genes and proteins, reflect what is happening in the organism at the time of sampling; they act as direct signatures of biochemical processes, thereby providing crucial information for the understanding of molecular mechanisms of disease and health statuses.

1.2. Metabolomics: one science, many experimental approaches

Metabolomic studies can be planned applying two main methodological approaches: untargeted and targeted (**Table 1**). The former provides a global estimation of a sample by analysing all, or as many as possible, metabolic features present in the biological specimen. The latter involves the monitoring of a panel of metabolites selected *a priori*, on the basis of known biochemical pathways or pre-identified biomarkers that result to be certainly related to the disease or the condition under study.³ The untargeted approach, achievable via metabolic fingerprinting or profiling, allows both sample classification (*fingerprinting*) and the identification and characterization of the metabolic features associated to specific physiological or pathological conditions (*profiling*).⁴

Table 1. Main metabolomic glossary. Adapted from Fiehn-Nielsen,⁵ and Oliver.⁴

Term	Definition
Metabolites	Low molecular weight organic molecules (<1500 Da) involved in metabolic processes as intermediate substrates or end-point products with different biological functions (<i>i.e.</i> fuel, structure, signalling, catalytic and inhibitory effects on enzymes, defence from toxins etc.).
Metabolome	The quantitative ensemble of all low molecular weight organic molecules within an entire organism, an organ, a biofluid, a tissue or a cell in a particular physiological or developmental state.
Metabolomics	The quantitative measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modifications, under the combined effect of environmental stimuli.
Untargeted analysis	The comprehensive analysis of all the measurable analytes in a sample, including chemical unknowns.
Targeted analysis	Measurement and monitoring of a panel of biochemically- and chemically-characterized metabolites, selected <i>a priori</i> on the basis of known metabolic pathways or pre-identified biomarkers, undoubtedly associated with a specific metabolic condition of interest.
Metabolic fingerprinting	Global, high-throughput, rapid analysis of all metabolites that are present in a biological

	sample to provide sample classification (even without metabolite identification).
Metabolic profiling	The identification and quantification of as many as possible metabolites above the detection limit of the chosen analytical platform.

Nuclear magnetic spectroscopy (NMR) and *mass spectrometry* (MS) are the leading analytical techniques for metabolomic research.^{1,3,6} Both of them provide concentrations of metabolites and molecular structures, and they yield information about many molecules in a single measurement, but each technique has its own strengths and weaknesses.

MS studies usually require separation and derivatization of metabolites from the biofluid matrix before the detection, typically through high-performance liquid chromatography (HPLC) or alternatively, gas chromatography (GC). This last one can be applied to make metabolites more volatile. In this light, the MS technique requires several steps for the analytical preparation of samples. Recent methodological improvements allowed LC-MS to analyse various metabolites with different molecular characteristics ranging from hydrophobic to hydrophilic features; while the application of ultra-high performance liquid chromatography (UHPLC) lowered the detection limit from nanomolar down to femtomolar. Nevertheless, MS reproducibility still remains the main limitation of this technique for metabolomic studies; indeed, in MS experiments many compounds give variables responses when present in complex mixtures. Consequently, NMR, although overshadowed by MS in terms of sensitivity, offers key advantages, especially for high-throughput metabolomic analyses. NMR does not damage analytes, thus enabling the investigation of metabolites in intact tissues. Then, it requires a minimal and fast sample preparation, and any molecule containing NMR active nuclei can be detected simultaneously, while for MS several tailored experiments are generally requested for specific chemical species (**Table 2**).³

Table 2. Main strengths and weaknesses of NMR and MS for metabolomic research. Table adapted from Vignoli *et al.*³

Technology	NMR	MS
Reproducibility	Very high	Fair
Detection limit	Micromolar range	Picomolar range
Sample preparation	Minimal	Several steps: metabolites separation/derivatization
Volume of original samples consumed	0.1-0.5 mL	0.01-0.2 mL
Types of detected molecules	Any molecule containing NMR active nuclei	Most of organic molecules and some inorganic
Types of experiments	All metabolites above detection limit can be observed simultaneously	Several: tailored for specific chemical species
Ambiguous/false identification	Origin: compounds with degenerate chemical shifts, chemical shifts variability due	Origin: compounds (<i>e.g.</i> isomers) that can match with a given atomic composition or a

	to experimental conditions (e.g. pH, temperature, ionic strength), presence of only singlet. To overcome these issues, 2D experiments, spiking of authentic reference compounds need to be performed.	parent ion mass. Experimental approaches: LC-MS/MS
--	---	--

Additionally, NMR is an intrinsically quantitative technique: in the spectra, the integral of each peak reflects directly the number of nuclei with non-zero magnetic moment giving rise to that peak.

In conclusion, NMR and MS are two complementary techniques and the weaknesses of one are compensated by the strengths of the other. The peculiarities of NMR, coupled with chemometric tools, make it a more suitable approach for untargeted metabolomic applications, with the aim of characterizing metabolic fingerprint of diseases or any condition of interest, generating new biochemical mechanistic hypothesis. On the contrary, the high sensitivity of MS lends well this technique for confirming or validating pre-existing hypothesis on biochemical pathway in a mostly targeted approach.

1.3. NMR and metabolomics

NMR had contributed very significantly to metabolomic research, despite its lower sensitivity compared to MS. Advantages and strengths of NMR, including its highly reproducibility, quantitative nature, ability to detect metabolites in intact biospecimens, compensate for its main limitations (see §1.2). Recently, improvements of NMR sensitivity and resolution have been reported,^{7,8} making this methodology increasingly useful for the analysis of a huge variety of biological samples, especially for biomedicine research.

1.3.1. Types of samples

Through NMR spectroscopy, we can analyse a wide range of samples, including samples from plants, animals and related based-foods (e.g. fruits and vegetables, wine, oil, milk, meat etc.), microorganisms (e.g. yeasts, bacteria) and human or animal body fluids (i.e. urine, blood serum or plasma), faeces, organ tissues, culture media, cells, etc. The choice of sample depends on the question being asked with, for example, biofluids typically investigated to identify new clinical biomarkers, whereas tissues and cells are commonly studied to explore mechanisms related to pathophysiological processes. Tens-hundreds of small molecules, mainly amino acids, carbohydrates, organic acids, alcohols and other organic compounds (**Figure 2**) can be detected in a sample through NMR which can contribute also to the definition of the overall lipid composition of the biosystem (i.e. high-density and low-density lipoproteins; short-chain fatty acids; mono-, di-, tri- triglycerides; free cholesterol, cholesterol esters; total

cholesterol; glycerophosphocholines; sphingomyelins, etc.) in a so-called lipidomic vision.^{9,10}

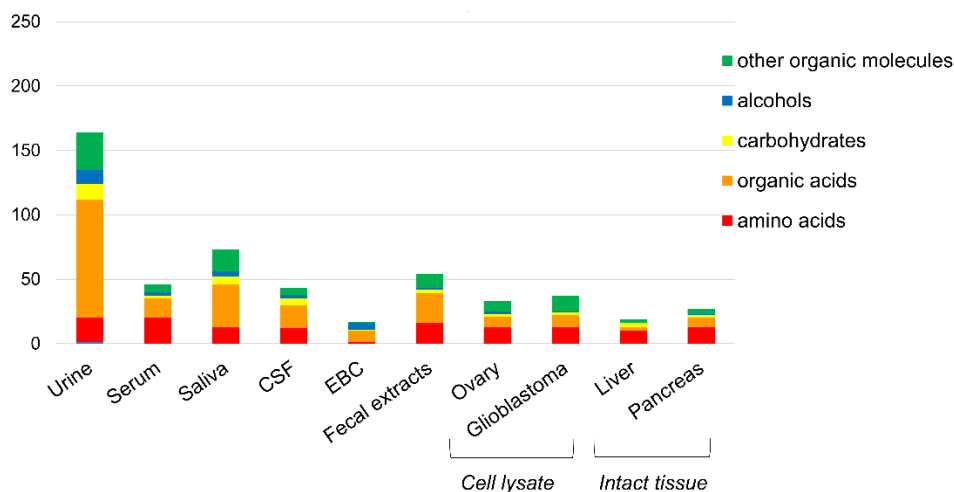


Figure 2. Number of detectable and quantifiable metabolites with >50% occurrence of different biological samples, by ¹H-NMR spectra: urine,¹¹ serum,¹² saliva,¹³ CSF,¹⁴ EBC,¹⁵ faecal extracts, cell lysates (e.g. ovary and glioblastoma cells)^{16,17} and intact tissues (e.g. liver and pancreas).^{18,19} Figure taken from Vignoli *et al.*³

Before performing the analysis, sample-specific considerations need always to be pondered. Different types of cells reflect different metabolomes, and intracellular metabolites vary with respect to endogenous and exogenous stimuli (e.g. treatment with a drug, genetic manipulation or mutation, protein overexpression etc.), evidencing up- or down-regulating specific biochemical pathways. The complementary biological information provided by the exo-metabolome is fundamental to complete the analysis.²⁰ NMR data from tissues are more complicated to interpret: they reflect the organ-specific biochemistry (e.g. aerobic respiration, lipid metabolism etc.), but they are also affected by the heterogeneous composition of tissue extracellular matrix and microenvironment. However, tissue samples directly report the condition of the diseased organ, where metabolic variations with respect to a healthy status are expected to be more pronounced. Compartmentalised biofluids, such as cerebrospinal fluid (CSF) and exhaled breath condensate (EBC) represent the biochemistry of the central nervous system and that of the respiratory tract, respectively. Saliva, another compartmentalised biofluid, report metabolic changes related both to oral disorders and distant pathologies.^{21,22}

Systemic body fluids like urine or blood plasma/serum have a general fainter biochemical correlation with a diseased organ or apparatus, but their simple, minimally invasive collection and their ability to reflect the overall response of the organism to the disease condition, are strength points for high-throughput metabolomic analyses. On the other hand, blood and urine are considerably different in terms of chemical composition, with blood reporting a better defined and stable metabolome, while urine metabolites being heavily influenced by lifestyle factors, such as food and liquid intake, physical activity, use of drugs etc.

In particular, blood represents an optimal and suitable biological matrix for metabolomic analysis. It carries nutrients, dissolved gases, hormones and metabolic wastes¹², but it also regulates pH and ion composition of interstitial fluids; it allows for the defence against toxins and pathogens, and it plays a key role for the stabilization of body temperature. Moreover, through its interaction with every tissue and every organ in the body, it offers a snapshot of the metabolic state of an organism, providing pivotal information for detecting, managing and monitoring virtually all disease statuses or other conditions of interest. For these reasons, blood specimens have been mainly studied in this PhD thesis.

In the light of the above considerations, the preservation of the chemical composition of the *in vivo* metabolome and the feature of spectral reproducibility, are key factors for the global significance and performance of metabolomic analysis by NMR.

1.3.2. NMR-based metabolomic applications: the state of the art

The NMR versatility allowed for a wide variety of metabolomic applications in life science research, ranging from human and veterinary health to the exploratory characterization of plants, animals, related based-foods, microorganisms and environment (**Figure 3**).

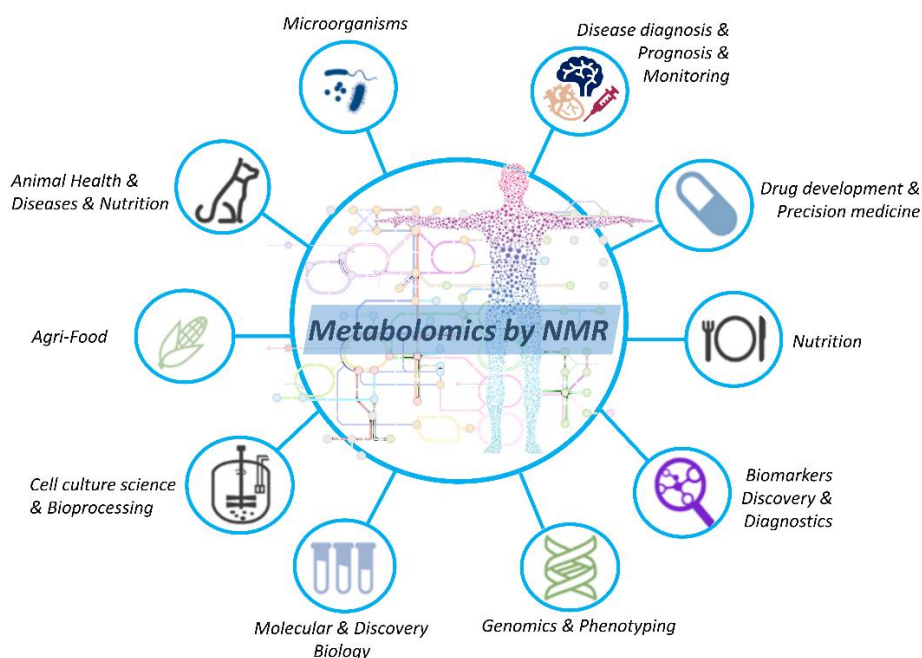


Figure 3. NMR-based metabolomics in life sciences: applications in different fields, ranging from the exploration of microorganisms, animals, agri-food to human research. Figure adapted from web sources.

As metabolites indicate intermediate and end-points of gene expression and cell activity, under the combined influence of external stimuli, metabolomics can provide a holistic approach to understand the phenotype of an organism, holding promises for

both clinical and precision medicine. In this context, many biomedical fields, summarized in three broad areas (**Figure 4**), might benefit from metabolomic studies.

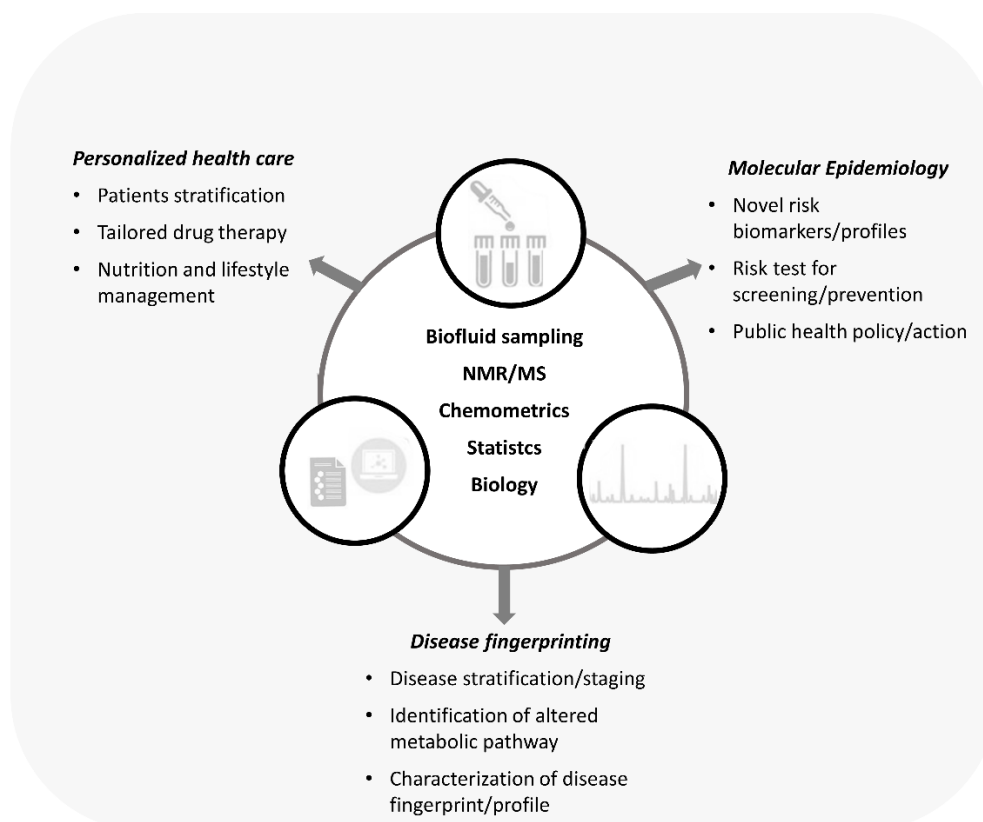


Figure 4. Scheme of the main steps of metabolomic analysis and related applications in biomedicine. Adapted from Nicholson *et al.*¹ and web sources.

As it has been previously stated, NMR-based metabolomics is being increasingly applied to disease diagnosis and characterization (*disease fingerprint*). Cancer is probably the most studied pathology so far via NMR metabolomics.^{23–26} Cardiovascular diseases,^{27–29} cerebrovascular diseases^{30,31} and neurologic disorders,^{32,33} provided other excellent targets for this -omic science.

Metabolomics also offers a cost-effective and productive route for the discovering of new drug targets and to predict and monitor individual response to drug treatments (*pharmacometabolomics*).^{6,34,35} Other metabolomic applications include the monitoring of the effects of surgical and non-surgical interventions,^{36–38} dietary treatments,^{39–41} the understanding of metabolic unbalances underlying pathologies and the mechanisms at the basis of the development of poor patients ‘outcomes. Moreover, metabolomics is promising as a clinical tool for *molecular epidemiology*, with the most challenging goal of detecting early metabolic disturbances even before the manifestation of disease symptoms, providing early diagnosis, tailored therapies and identifying metabolites or fingerprints useful as novel risk biomarkers of a specific pathology, for disease recurrence or stratification.

For the sake of completeness, it is important to underline that metabolomics also proved to be effective for *veterinary medicine*,⁴² in particular in the frame of disease diagnosis, drug discovery and investigation of animal health status.^{43–45}

Considering all the relevant results listed in the literature and the potentiality of NMR-based metabolomics for the biomedical field, the need of deepening researches to identify *i)* new disease fingerprints and *ii)* a panel of small molecules that can be useful to deepen the knowledge of pathological mechanisms is still urgent.

1.3.3. NMR-based metabolomic workflow

Initially, a metabolomic workflow relies on a straightforward formulation of the biological question to be addressed. This step is crucial because it will determine the experimental design that follows. Depending on the biological problem at issue, the type of metabolomic approach (targeted *vs.* untargeted), types of samples (*e.g.* body fluids, tissues and/or intact organisms), sample size, experimental conditions (*e.g.* frequency of sample collection, metabolic quenching to interrupt certain enzymatic activity), storage conditions, analytical platform to be used, must be defined before performing sample collection and the analysis. Considering the comparative character of metabolomic studies, a group of samples that did not undertake the investigated condition (control samples) and test samples (carrying information on the investigated condition) should be usually defined in the experimental design.

A schematic representation of a typical NMR-based metabolomic workflow is illustrated in **Figure 5**.

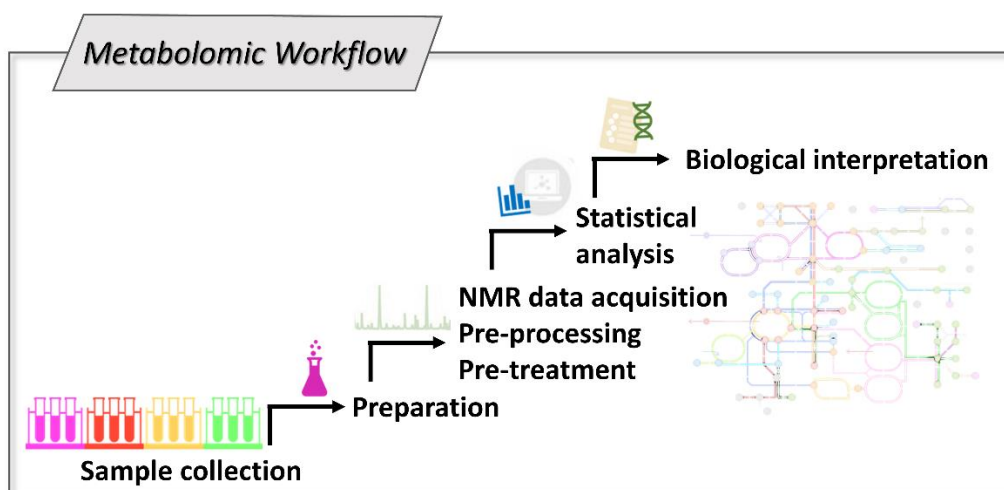


Figure 5. Typical steps followed for metabolomic analyses by NMR.

Sample collection, handling and storage constitute critical phases of metabolomic research, and the adoption of standard operating procedures is fundamental to maintain sample biostability and to guarantee reproducible results.^{46,47}

Sample preparation can vary widely, it is intimately related to the sample type, the metabolomic approach and the elected analytical platform. Processes such as solvent-mediated extractions or buffering could be necessary. The second step regards *data acquisition* by NMR spectroscopy. Differently from other omics sciences, metabolomics imposes a big analytical challenge due to the huge variety of chemical species that biological samples exhibit. In this terms, appropriate NMR pulse sequences need to be implemented and applied. Acquired raw data are subjected to *pre-processing* steps (mainly, baseline correction, phasing, alignment and/or bucketing) and pre-treatment procedure (second step) before the *statistical analysis*. Metabolomic data are quite complex to be interpreted and they require chemometric tools to address the aim of the research (third step). Multivariate analysis, comprising unsupervised method such as Principal Component Analysis (PCA) and supervised method, such as Orthogonal-Partial Least Square Discriminant Analysis (OPLS-DA), Multilevel Partial Least Squares analysis (M-PLS) and machine learning algorithms (*e.g.* support vector machines, random forest) are often employed for sample overview, classification and prediction. Alongside the common multivariate approaches, univariate analysis, based on common Student's t-test, analysis of variance, Kruskal–Wallis test and Wilcoxon signed-rank test, is also used to corroborate multivariate results. Statistical models must be validated and their robustness need to be checked in order to avoid overfitting of data, by using internal cross-validation schemes, permutation tests, receiver operating characteristic (ROC) curve estimation, or when possible, advocating an external and independent validation cohort.

In a more systematic and comprehensive approach, metabolic and lipidic association networks can be useful to infer metabolite-metabolite and metabolite-lipid significant associations. However, Chapter 3 will bring more details about chemometric tools and statistical procedures in NMR-based metabolomics.

Before concluding, untargeted metabolomic studies require metabolite identification. For such purposes, free databases such as HMDB,⁴⁸ KEGG,⁴⁹ PubChem,⁵⁰ and libraries like AMIX (Bruker) and AssureNMR (Bruker) can be used in NMR-based metabolomic research. As a last step, metabolomic data need a *biological interpretation* and several databases (*e.g.* KEGG, MetaboLights⁵¹) or online tools, such as MetaboAnalyst⁵² are available for this purpose.

Chapter 2

Aims of this thesis

Metabolomics is a developing and emerging technique that holds promises for different fields, providing powerful insights into the mechanisms of both human and animal diseases and health status. This work is mainly a methodological thesis covering different topics with the principal aim of demonstrating the potential of untargeted NMR-based metabolomics in the biomedical field.

In the first section of this thesis, particular attention is paid, in the frame of human biomedicine research, to:

- 1) prove the usefulness of untargeted NMR-based metabolomics applied on blood serum or plasma samples to unravel fingerprints/metabolic markers of neurological disorders, *i.e.* Parkinson's disease and ischemic stroke, and prostate cancer; mainly to uncover the underlying molecular mechanisms characterizing different stages of the diseases and to enable patients' stratification and characterization of metabolic changes occurring after the onset of the pathology;
- 2) provide predictive and prognostic metabolic markers of three-months poor outcomes of ischemic stroke (*i.e.* mortality, development of neurological impairment, haemorrhagic transformation of the cerebral lesion and non-response to the commonly applied thrombolytic therapy), to better characterize the post-stroke course of the pathology from a metabolic point of view and to enable a more tailored patient care;
- 3) prove the usefulness of metabolomics in the framework of personalized lifestyle management by unravelling blood serum fingerprints for dietary interventions phenotypes in subjects at risk for metabolic syndrome;

Since metabolomics has so many applications in the biomedical field, the same analytical technique can find potential applications in veterinary medicine, for disease characterization or to monitor animal health status. In this framework, the second section of this thesis is dedicated to:

- 1) prove the effectiveness of untargeted NMR-based metabolomics applied on animal sample extracts (serum and/or liver, urine) to provide fingerprints/biomarkers of displaced abomasum in dairy cows and *Erlichia canis* infection in dogs;
- 2) demonstrate the ability of NMR-based metabolomics in monitoring the health status of premature calves in order to prevent calves' mortality or pathologies.

A third and last section of this PhD thesis is dedicated to demonstrate the robustness of the equidistant bucketing in performing the NMR fingerprinting of biological

samples. During the last years, the bucketing procedure has been improperly criticised and abandoned because of a growing interest in profiling technique; indeed, since most medical and also food analyses require quantifiable properties, bucketing has become less important. The work here reported, shows that equidistant bucketing maintains all the necessary information, especially the one encoded in chemical shifts data, to perform the NMR fingerprint of the most commonly analysed biofluids (blood serum and urine).

Chapter 3

Methodologies

3.1. Sample preparation

As previously stated in Chapter 1 (§1.3.3), pre-analytical processes are crucial for the global performance of molecular analyses of biological samples. Indeed, several steps, such as primary sample collection, transport and storage should avoid the generation of artificial profiles, ensuring the detection of the original metabolome and lipidome of a sample. The impact of non-appropriate pre-analytical procedures could be tremendous and this is one of the main causes that make the comparison of metabolomic data, collected from multicentric studies, difficult. Some molecules, constituting the metabolome, are very sensitive to sample conditions; for example, very low temperature (-80°C) is used to store samples, avoiding their denaturation or to quench enzymatic activity.

Analytical NMR preparation of common biofluids, such as blood and urine, requires minimal steps. Typically, they consist mainly on adding a buffer solution to sample matrix, to avoid pH variation and to easily reference chemical shifts values to existing resonances in published databases. A measured amount of a reference compound such as tetramethylsilane (TMS) for organic solvents and trimethylsilylpropanoic acid (TMSP) or sodium 2,2-dimethyl-2-silapentane-5-sulphonate (DSS) for aqueous solutions is often added to the sample. The addition of small quantity of deuterated solvents, generally D₂O, provides the signal to stabilize the magnetic field, also allowing the optimization of NMR peaks resolution. To extract fractions (*e.g.* serum, urine, liver etc.), the analytical preparation of samples depends on the nature of the extract: hydrophilic or lipophilic. Typically, for organic fractions, chloroform or methanol-chloroform mixtures are added in different proportions, while D₂O is preferred to reconstitute polar fractions.

Usually, NMR analysis requires an original sample volume ranging from 0.1 to 0.5 mL, and it must not be forgotten that several and specific practical consideration (*e.g.* solvent, pH, temperature etc.) should be considered always according to the specific biological matrix under study. Generally, NMR sample preparation must meet the main following criteria: *i)* ease of use; *ii)* reduction of unwanted source of variation and artifacts; *iii)* robustness of the analysis and *iv)* reproducibility.⁵³

3.2. NMR pulse sequences: tips for 1D metabolomic applications

NMR spectroscopy represents a powerful, versatile and reproducible technique for the analysis of complex biological matrices because any biological molecule containing one or more atoms with a non-zero magnetic moment is theoretically detectable by NMR.⁵⁴ Virtually, all biologically relevant molecules are characterized

by at least one NMR signal with a specific intensity, frequency (or chemical shift) and magnetic relaxation properties, all reflecting the chemical environment surrounding the detected nucleus.⁵⁵

One-dimensional (1D) ¹H-NMR spectroscopy is the most widely used technique for metabolomic research, since proton is the most abundant and sensitive nucleus present in biomolecules. For solution-state NMR experiments, the main problem that must be addressed is related to the detection of protons of biomolecules in aqueous samples at μ M concentrations with a background of almost 100 M water protons.⁵³ Indeed, typical biological samples usually contain a large quantity of water and only small amounts of D₂O can be added to offer sufficient signal for the deuterium lock, but without compromising the detection of exchangeable protons in the sample. Therefore, the larger water signal needs to be reduced using water-suppression techniques in order to observe and quantify signals from the metabolites of interest, which are ranging in micromolar concentrations. The simplest strategy to suppress water consists of a method called “pre-saturation” that applies a weak radio-frequency irradiation for a period of the order of the solvent T_2 to selectively saturate the solvent resonance.⁵⁶ Gradient-based solvent suppression techniques can be also used to obtain efficient water signal reduction, such as the WATERGATE scheme, which uses a composite pulse, surrounded by two symmetric pulsed-field gradients with the aim of attenuating the water resonance.⁵⁷

Among the large variety of NMR pulse sequences available to date, only few of them are suitable for the metabolomic analysis. One challenge is to obtain at the same time, access to the measurement of relative or absolute concentrations of the metabolites in their original mixture and to make the analysis as fast as possible.

Indeed, NMR-based metabolomics build up on the high-throughput investigation of large cohorts of biological systems. Urine and blood serum or plasma are analysed considering their natural isotopic abundance, while metabolic fingerprint/profiling strategies largely aim of providing 1D or 2D profiles relevant for the identification/quantification of the greatest number of detectable metabolites in a biological sample. Moreover, in recent years, the standardization of metabolomic NMR experiments has been increasingly requested for constructing reliable spectral databases for metabolite identification, comparisons across different collection centers, and for studies which aim of developing universal sample classification models based on spectral data. Consequently, only few complementary pulse sequences, following standardized protocols, are commonly used for most NMR-based metabolomic applications.^{58–60}

The 1D ¹H NOESY pulse sequence with water pre-saturation is the most popular acquisition scheme, applied in the frame of metabolic fingerprint of biological samples, to detect both small molecules and macromolecules. It has the advantage of needing minimal parameter optimization. The 1D ¹H NOESY pulse sequence involves a series of three 90° radio frequency pulses (**Figure 6 A**), where the first pulse creates transverse spin magnetization, the second and the third ones, separated by a mixing time τ_m , correspond to the NOESY filter, which improves the quality of water

suppression.⁶¹ 1D ^1H NOESY spectra provide a complete and quantitative fingerprints of the observed metabolites and for this reason, they will be widely used in this PhD thesis.

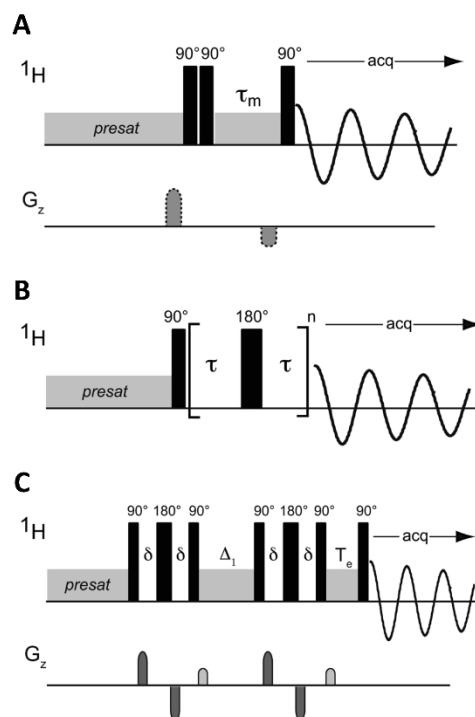


Figure 6. Standard one-dimensional NMR pulse sequences for metabolomic analysis. Figure taken from the “*NMR-based metabolomics*” book.⁵⁶ (A) ^1H NOESY with water pre-saturation during the recycle delay and optional pulsed field gradients to improve water suppression. The mixing time τ_m is typically set to 100 ms. (B) Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence, with water presaturation during the relaxation delays. (C) Stimulated-echo sequence with bipolar-gradient pulse pairs and longitudinal eddy currents delay⁶² for acquisition of diffusion-edited spectra.

A challenging issue to be solved, during the observation of small metabolites from complex mixtures, is the common presence of macromolecules, whose corresponding broad background signals may cover the presence of low-concentrated small metabolites. Considering that large macromolecules usually have relaxation time (T_2) shorter than those of small metabolites, transverse relaxation properties can be used to selectively detect small molecules. This is the aim of the Carr-Purcell-Meiboom-Gill (CPMG) experiment, consisting in a 180° pulse train (**Figure 6 B**) to achieve relaxation editing by the attenuation of the fast-relaxing species.⁶³ Specifically, each 180° pulse generates a spin echo at time 2τ , and the amplitude of these echoes decays exponentially with a time constant corresponding to the transverse relaxation time T_2 . The fast-relaxing signals from macromolecules are allowed to decay without any significant loss for the slow-relaxing small metabolites, by choosing a suitable delay $2\tau n$ before acquisition of the spin echo, typically in the order of 60 to 100 ms.⁵⁶

Another approach for spectral editing according to molecular size is based on the differences in rates of molecular diffusion between the molecular species. In this way, only signals from molecules with restricted translational mobility, *i.e.* macromolecules, can be detected. In other words, diffusion-edited spectroscopy relies on the attenuation of NMR signals of metabolites that diffuse rapidly along the sample matrix in the presence of a pulsed gradient (**Figure 6 C**).

The above-described experiments are commonly used for metabolomic analysis of blood serum or plasma (**Figure 7**), while in the case of urine, the application of 1D ^1H NOESY pulse provides sufficient information. It should be pointed out that some experts prefer physically removing macromolecular components via centrifugation with 3000 MWCO Amicon Ultra-0.5 filters followed by NOESY acquisition, rather than applying the CPMG sequence.^{8,64}

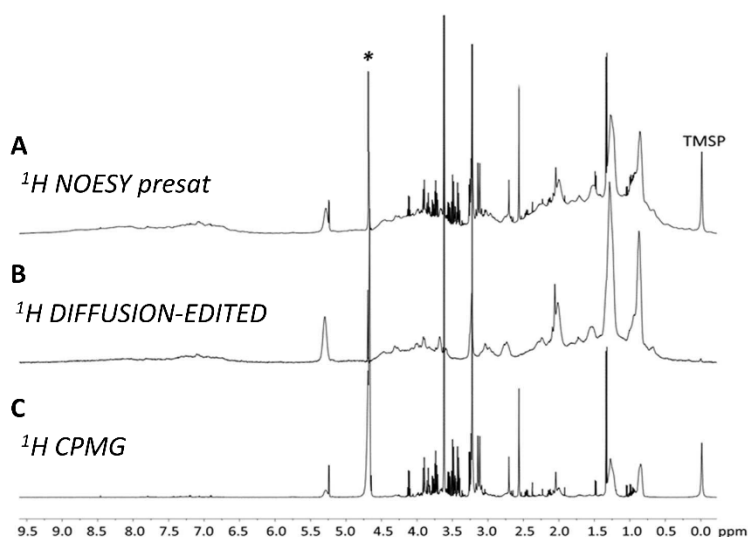


Figure 7. Typical 600 MHz ^1H NMR spectra of a serum sample are reported: **A)** ^1H NOESY presat spectrum; **B)** ^1H Diffusion-edited spectrum; **C)** ^1H CPMG spectrum. Spectra are calibrated at the anomeric glucose signal at 5.24 ppm. After pre-saturation, residual water signal is marked with an asterisk (*).

3.3. NMR data pre-processing and pre-treatment

For any metabolomic study, the intermediate step (*data pre-processing*) between recording NMR raw data and applying statistical analysis, represents a crucial procedure for the final performance of the research. Indeed, data pre-processing, with data pre-treatment methods, allow the reduction of variances caused by measurements, by biology or by combined effects which are not of interest or which might interfere with subsequent data modelling.

Each NMR spectrum must be adequately adjusted for phase and baseline, using automatic procedures; manual adjustments are discouraged for metabolomics because they can add operator-biased artefacts. Spectra need also to be properly aligned to a known resonance of a reference signal, ideally not interacting with any sample

molecule; for example deuterated trimethylsilylpropanoic acid (TMSP) may bind macromolecules (*e.g.* albumin protein) and its use is discouraged in samples like plasma or serum, where alternative internal references are preferred (*e.g.* the anomeric doublet of glucose at 5.24 ppm).⁶⁵ For the same reason, the use of TMSP as a standard for quantification is discouraged. Absolute quantification of molecules can be carried out using alternative approaches, such as the production of an artificial NMR signal based upon PULCON⁶⁶ or ERETIC⁶⁷ methods. Software systems like the B.I. (Bruker IVDr) platform for the analysis and metabolites/lipids quantification in biofluids (*e.g.* serum/plasma and urine) are actually used, also in the “Results” section of this PhD thesis.

When NMR is applied to metabolomic profiling, some practitioners prefer to work with full resolution spectra, applying specific algorithms for peak alignment, such as the *icoshift*,⁶⁸ to assure the comparability of NMR peaks across multiple spectra. Alternatively, NMR metabolomic fingerprinting of samples can be obtained by transforming NMR spectra into data matrix, by means of binning or -bucketing techniques. These methods allow the integration of NMR spectra within small spectral regions, called “bins” or “buckets”. Many sophisticated algorithms exist to bin 1D NMR spectra, but the most commonly and simply method is the equidistant binning of 0.02-0.04 ppm^{53,69}, which allows the division of the spectrum into evenly spaced integral regions with a fixed spectral width.

To compare signal – or bucket – intensities in different biological samples, it is important to refer to the same amount of total sample. Since large physiological variation in metabolite concentration is always present, a preliminary step to correct dilution effect is requested. For example, urines exhibit significant metabolite concentration fluctuations due to the different hydration state of the individual, therefore normalization is generally applied, as pre-treatment step, to correct global signal intensity.⁷⁰ For blood/serum metabolomics, normalization is less crucial; indeed blood matrix is less variable and it is under a stricter physiological control; nevertheless, serum/plasma normalization can compensate for non-physiological sources of variation, especially from experimental inaccuracies and/or artifacts.

The most popular normalization technique is total area normalization, where each NMR variable (*e.g.* a bucket or a data point) is divided by a constant number representing either the integral of the reference peak or the sum of binning of all spectra. In the research for the “best performing” normalization method, a plethora of strategies/algorithms have been developed. For biofluids, especially urine, Probabilistic Quotient Normalization (PQN)⁷¹ is considered a good choice. It works by calculating a reference spectrum (*e.g.* the median spectrum), then for each spectrum of the dataset, each NMR variable is divided by the corresponding variable in the reference spectrum, and the median of all the obtained quotients represents the normalization factor. PQN assumes that variations in the concentration of biological interest, affect only few portions of the NMR spectrum, while dilution effects will influence all signals.

Before concluding this paragraph, few words must be said on scaling methods. Scaling procedures are data pre-treatment approaches that divide each NMR variable (*e.g.* a bucket or single metabolite) by the scaling factor, which is different for each variable, with the aim of adjusting for the fold differences between the different metabolites, making entire profiles more comparable. Different methods exist, like auto-scaling, pareto-scaling etc., and they are generally applied when, within the same dataset, different typologies of data (*e.g.* clinical and metabolomic data) must be analysed together.

3.4. Statistical analysis

NMR-based metabolomic analysis generates a large quantity of complex, multifaceted data. To obtain comprehensive and meaningful results, the statistical analysis represents a key part of the metabolomic workflow. Starting from full or bucketed NMR spectra, or from a list of metabolite concentrations, metabolomic analysis principally aims at: *i*) visualizing the overall differences, trends, relationships among different samples; *ii*) detecting whether there is a significant difference among sample groups under study; *iii*) highlighting spectral regions mostly contributing to mentioned differences; and *iv*) constructing reliable and robust predictive model to correctly classify new samples.^{3,53} Applying *multivariate statistics*, we can achieve these goals, by either *i*) unsupervised (no assumptions made on the samples) or *ii*) supervised methods (samples are defined into classes, or each sample is associated to an outcome y_i value).

Generally, multivariate methods represent samples as points in the space of the initial variables. Then, samples can be projected into a lower dimensionality space, *i.e.* components or latent variables, such as a line, a plane or a hyperplane. In the newly defined latent variables, the coordinates of samples are defined as the “*scores*”, while the direction of variance to which they are projected are named “*loadings*”. The loadings vector for each latent variable contains the weights of each of the initial variables in that latent variable.

Unsupervised methods usually represent the first step for the statistical analysis, helping in the visualization of the data, detection of possible outliers or evident metabolic signatures. Among the different unsupervised methods, Principal Component Analysis (PCA)⁷² is the most popular and commonly used technique. It involves the projection of the data in a new space using just few dimensions (data reduction). Each principal component (PC) represents a linear combination of the original NMR variables (*e.g.* buckets or metabolite concentrations); they are orthogonal and independent from all the others. The first PC (PC1) expresses the maximum percentage of variance inside the dataset and thus, the greatest amount of meaningful information. In the successive PCs, the value of variance decreases in such a way that the last PCs have less significantly importance, thus expressing mainly noise variability.

Geometrically, PCA can be viewed as a rotation of the reference system to maximize the data structure, minimizing the noise. The original data matrix (\mathbf{X}) is factorized into two new matrices: the score matrix \mathbf{T} and the loading matrix \mathbf{P} :

$$\mathbf{T} = \mathbf{X} \mathbf{P}$$

The score matrix \mathbf{T} represents a simplified description of the original observations in a reduced dimensional space, enclosing the coordinates of the original data in the new principal component space. As previously stated, the loading matrix \mathbf{P} contains the weights of the original variables to transform them into the scores. Typically, the first columns of the score matrix can be plotted against the second or the third one, in the so-called “*score plot*” (Figure 8 A), where each sample is represented by a dot. Similarly, the coefficients responsible for sample groups separations are depicted in the “*loading plot*” (Figure 8 B). If in the score plot of PC1 vs. PC2 (Figure 8), a separation between sample groups is evident, the loadings of the corresponding PCs provide evidences on which original variables are responsible for that discrimination.

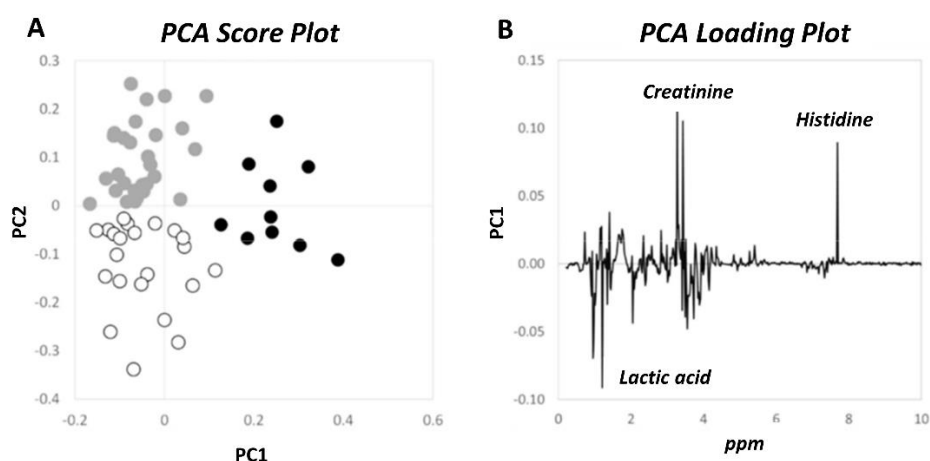


Figure 8. Example of PCA analysis on biofluids. **A)** PCA Score Plot: PC1 vs. PC2, where each dot in the plot represents a different ^1H NMR spectrum corresponding to a sample; grey dots: group 1; black dots: group 2; white dots: group 3. **B)** PCA loadings plot of PC1, where the most significant metabolites in the score plot are evidenced from the highest loadings (*e.g.* lactic acid, creatine, histidine). Figure adapted from our research group material.

Conversely, *supervised methods* use *a priori* knowledge to generate models that are closely focused on evaluating the effects of interest. The main goal of supervised data analysis is to extend the information already available from pre-existing samples to new samples, so that prediction can be performed in the frame of disease diagnosis and/or prognosis.

Supervised analysis includes *methods based on projection and data reduction*, such as Projection to Latent Structures (PLS)^{73–75} and its orthogonal variant (OPLS).⁷⁶

PLS is multivariate linear regression method similar in concept to PCA, which finds the relations between two matrices, data \mathbf{X} and response \mathbf{Y} , by maximizing the covariance of their latent variables. Using this approach, it is possible to understand which variables (*e.g.* NMR bucket or metabolites) of \mathbf{X} are more correlated to the response (*e.g.* disease/health status) and to make predictions of new samples. OPLS is a modification of the previous PLS method; showing the same predictive power as PLS, but providing better interpretation of relevant variables than PLS. This is possible by decomposing the data in the so-called “predictive” information related to the \mathbf{Y} response and the “orthogonal” structured information not related to the response, such as instrumental or unwanted biological variations.

To overcome issues related to the large between-subjects variability, the multilevel PLS⁷⁷ analysis is applied. In a multilevel PLS, between-subject variation is separated from the within-subject variation by subtracting the individual specific average, and only the within-subject variation is considered for the analysis.

Reduction techniques like PCA and PLS can be applied in combination with Canonical Analysis (CA) to enhance sample groups discrimination. CA allows discriminant projection by maximizing between-class distance and minimizing within-class distance. In addition, both PCA, PLS, OPLS and M-PLS can be applied in combination with Discriminant Analysis (DA) to infer the best discrimination among groups and also in the frame of predictive analyses.

For classification and predictive purposes, *machine learning supervised methods* are largely employed, especially k-Nearest Neighbours (k-NN) and Support Vector Machines (SVM) are the widespread algorithms for metabolomic analyses.

To date, k-NN is probably the simplest and computationally easiest classification algorithm: it works in a local neighbourhood around a considered test sample that must be classified. The neighbourhood is generally estimated by the Euclidean distance and the closest k (*e.g.* $3 \leq k$) objects are used to calculate the group membership (based on a “majority voting”) of a new sample⁷⁸ (**Figure 9 A**). The choice of k size largely influences the quality of results: small k values leads to the building of a statistical model where significant statistical fluctuations are present; while large k values can reduce statistical errors, but they flat many details of the sample distribution.

SVM is based on a different concept: it uses a statistical learning paradigm to construct a “borderline” (a line, a plane or a hyper-plane) tract in the sample space (**Figure 9 B**), separating the different classes of samples.⁷⁹

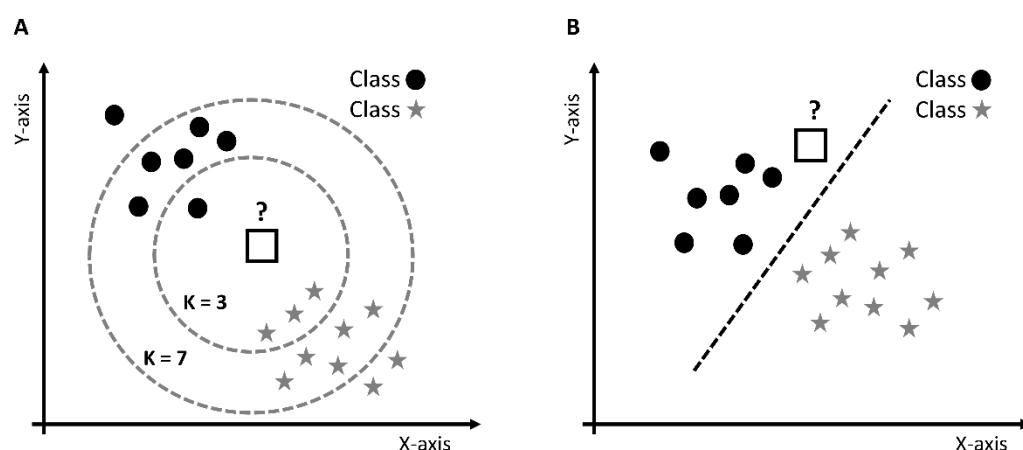


Figure 9. Classification of an unknown sample (white square) among two groups (black dots and grey stars) with different methods: **A)** k-NN classification; **B)** SVM.

A relatively new and powerful machine learning algorithm is the so-called “random forest”. This approach represents a powerful machine learning algorithm with many strengths: *i)* it can deal with large number of predictor variables simultaneously; *ii)* it avoids overfitting problems; *iii)* it can be applied when there are more variables than samples; *iv)* it is relatively insensitive to noise; *v)* it allows visualization of data in a reduced discriminant space using the proximity matrix calculated during the process of forest growing; *vi)* the percentage of trees in the forest that assigns one sample to a specific class can be interpreted as a probability of class belonging; and *vii)* it gives an unbiased estimation of the classification error using the out-of-bag samples, avoiding the need for additional cross validation.

Random Forest algorithm can be used for data classification and predictive modelling. The first goal (*i.e.* classification) is derived from a randomly grown ensemble (forest) of decision trees based on the following steps: *i)* original data are randomly divided using bootstrapping into two separated sets: the training and test set; *ii)* an ensemble of decision trees is grown using the training set, where each of the trees is built on randomly selected NMR variables at each decision node; *iii)* as soon as all the trees are built, an unbiased assessment of the classification error using the out-of-bag samples is performed, enabling the estimation of the model performances.^{80,81} Moreover, for each sample, the percentage of trees in the forest that assign one sample to a specific class can be inferred as a probability of class belonging, making this approach very useful for predictive modelling purposes.^{82,83}

Generally, since the main risk of supervised approaches is the overfitting of data, it is a good practice to validate results by simply using an independent validation set, if available, or with more sophisticated approaches, such as cross-validation schemes. Cross-validation (CV) methods require an initial splitting of the dataset in a training set and in a validation set and secondly, by these techniques, one (Leave-One-Out, LOO) or many (Leave-Many-Out, LMO) samples are randomly removed from the training set and then, they are tested. Indeed, the model is constructed on the training part of the dataset, while excluded samples are used to assess the model performances,

generating a confusion matrix that express sensitivity, specificity and accuracy. *Sensitivity*, also called the “true positive rate”, is defined as the ratio between true positives (TP) and all positives (P), while *specificity*, also defined as “true negative rate”, represent the ratio between true negative (TN) and all negatives (N). Finally, *accuracy* is defined as the ratio between the sum of TP and TN and the total sample population (P + N).

In order to assure the best analysis performance, the use of permutation test to estimate the statistical significance of the results is strongly recommended. The employment of independent training and validation sets is the preferred approach in the medical community, even if much attention should be given to pre-analytical procedures for collecting samples from different hospitals or countries.

The complexity of ^1H -NMR spectra imposes not straightforward or simple processes for the *metabolite identification*. Commonly, many NMR peaks can be directly assigned in one-dimensional spectra on the basis of chemical shifts and multiplicity of resonances. A help is offered by online databases like HMDB⁴⁸ for human metabolites, PubChem,⁵⁰ KEGG⁴⁹ and so on. For doubtful cases, the spiking of authentic standard molecules is advocated, while, sometimes, 2D NMR spectra offer additional information for assigning new metabolites.

The increasing need for a completely automatic assignment and quantification of metabolites leads to the continuous development of automated tools, such as the recent B.I (Bruker IVDr) platform for urine, CSF and serum/plasma samples. Meanwhile, other tools for semi-automatic quantitation have been developed, like BATMAN⁸⁴ and BAYESIL⁶⁴ tools, mostly based on Bayesian inference, ASICS⁸⁵ (a linear model) and the NMR Suite Software Package (Chenomx Inc., Edmonton, Canada) which offers a complete assistance for spectral deconvolution, peak fitting, integration and quantitation of the selected peaks.

With the aim of identifying a metabolite or a panel of metabolites that can represent a possible biomarker for the condition under study, each metabolic feature needs to be investigated in a separated manner from all the others, excluding any possible inter-relationships. This goal is achievable applying *univariate statistical analysis*: statistical tests, correlation analysis and ROC curves are the most employed approaches. Considering the biological assumption that metabolite concentrations do not follow a normal distribution, non-parametric tests are often used. Wilcoxon-Mann-Whitney⁸⁶ test is the non-parametric version of the classical Student t-test, used to compare the distribution of a metabolite between two groups. The null hypothesis (H_0) claims that two randomly selected samples from two populations, actually belong to the same population; the alternative hypothesis (H_a) states that H_0 is retained false, thus the two population are different. When there is the need to compare two related samples, matched samples or repeated measurements on the same specimens (pairwise comparison), the Wilcoxon signed rank test⁸⁷ is applied. When the groups to be

compared are more than two, Kruskal-Wallis⁸⁸ test (the analogue of the parametric Analysis of Variance) or Friedman⁸⁹ test (for paired samples) are employed.

On the whole, the output of all these tests is a *P*-value, that expresses the probability of obtaining an equally extreme result equal or a more extreme one than that what was actually observed, assuming the null hypothesis to be true. Conventionally, when the *P*-value is less than the significance level at 0.05 or 0.01, the null hypothesis is rejected and the considered variable is deemed statistically different in the case group. Because in metabolomic untargeted studies, we look one after thousands of variables, multiple testing corrections must be applied to avoid random false positive. Bonferroni⁹⁰ and Benjamini-Hochberg⁹¹ are the most widely used corrections. It is worth saying that *P*-values are often misused and misinterpreted,⁹² their significance depends strictly from population dimension and distribution. For these reasons, statisticians proposed the use of the effect size in the place of *P*-values, as an alternative measure of evidence.⁹³ Effect size represents the standardized measure of the magnitude of the observed phenomena.

Before concluding, it is important to mention the role of ROC curves and correlation analysis in determining metabolic features associated to the condition of interest.

ROC curves are graphical representations that illustrate the diagnostic ability of a binary classifier. The curve is generated by plotting sensitivity versus one minus specificity for all possible thresholds of the test. The accuracy is derived from the area under the ROC curve (AUC). Area values equal to 1 represent a perfect test, while an area of 0.5 stands for a worthless test. However, ROC analysis finds application also to check a multivariate supervised classifier, not only in the frame of univariate analysis.

Last but not least, Pearson correlations can be calculated to test whether there is a linear association between metabolites and clinical data or other biological features. Correlations are expressed by a coefficient (*R*), ranging from +1 (totally correlated), 0 (no correlation) and -1 (totally anticorrelated).

In conclusion, from metabolite/lipid to metabolite/lipid correlation maps, metabolic networks with a biological meaning can be inferred;⁹⁴⁻⁹⁶ indeed, association networks generally provide interesting information to describe the status of the biosystem or to compare the same across different conditions in the frame of a more systematic and comprehensive approach.

Chapter 4

Results

4.1. NMR-based metabolomics for human biomedicine

Metabolomics offers a snapshot of the metabolic dynamism of an organism, reflecting the response of the living system to pathophysiological stimuli, genetic mutations and external environment, thus providing crucial information for detecting, managing and monitoring any disease or healthy condition of interest.

Particularly in the context of human medicine, metabolomics has been used to deepen the knowledge on the mechanisms of several pathologies, to characterize metabolic profiles of diseases with the aim of discovering new biomarkers or identifying biochemical pathways involved in disease pathogenesis. Moreover, the identification of new metabolites as metabolic markers turned out to be useful for early disease diagnosis or prevention and also to design or to personalize dietary interventions or therapeutic strategies.

In this thesis, I assess the potentiality of untargeted NMR-based metabolomics in characterizing firstly, the metabolic components of two main neurological diseases, *i.e.* Parkinson's disease and ischemic stroke and secondly, an inflammatory metabolic profile of South African men with prostate cancer. Therefore, four studies, with different focuses are here presented. In this section, NMR-based metabolomics was also applied in the field of nutrition and lifestyle management, to investigate, on the human serum metabolome and lipoprotein profiles of subjects at risk for metabolic syndrome (MetS), the effect of the consumption of tailored bioactive enriched foods (BEFs).

The first study here proposed, aims at providing evidences of the existence of serum metabolic fingerprints that differentiate *de novo* drug untreated *Parkinson's disease* (PD) patients from advanced PD patients under dopaminergic treatment and healthy matched controls, also in relation to sex, using training ($n=131$) and validation ($n=198$) cohorts of German subjects (§ 4.1.1). This study is a part of the PROPAG-AGEING project, an H2020 EU-funded project that aims of identifying new molecular signatures for early diagnosis of PD. The main issue of this project is to answer to the central question of why is advancing age the most important risk factor for developing idiopathic PD, proposing a new rational and approach by starting from the hypothesis that the elderly physiology represents the environment which feeds PD onset and progression and that there is an apparent continuum between healthy ageing and neurodegenerative motor disorders. In the PROPAG-AGEING project, PD is assumed to be totally embedded within the basic molecular and cellular mechanisms of the ageing process.⁹⁷ Therefore, the final purpose is to detect the combination of all the alterations able to shifting the phenotype of elderly people from a physiological condition to PD. In this framework, the identification of markers of early PD diagnosis

is possible, before motor symptoms occur, and using common human biofluids such as urine or serum.

To our knowledge, the study proposed in this thesis, represents the first large NMR-based metabolomic application dealing with the characterization of serum profiles of *de novo* Parkinson's disease patients free from dopaminergic treatment compared to advanced PD patients undergoing pharmacological treatment and healthy sex/age-matched volunteers. Our results evidenced that both in the training and in the validation cohorts, 1D NMR fingerprinting analysis succeeded to discriminate between *de novo* PD patients and healthy controls (75% of overall predictive accuracy) and between *de novo* PD and advanced PD patients (89% of overall predictive accuracy), thus identifying characteristic signatures describing early and advanced stages of the disease, although advanced PD patients are pharmacologically treated. A more pronounced fingerprinting of early PD disease was evidenced in men with respect to women (overall discrimination accuracy of 73.5% between male *de novo* PD patients and male healthy controls), confirming the increasing evidences that risk of developing PD is higher in men than in women.⁹⁸ Applying univariate, logistic regression and ROC curve analyses, we have also showed and validated that increased levels of ketone bodies, particularly acetone, and decreased levels of lipoproteins and cholesterol in *de novo* PD patients, define the early stage of PD. Our results suggest that *de novo* PD patients are characterized by increased oxidative defences and a contemporary worsening of the oxidative stress status. Moreover, our results support a potential role of ornithine in monitoring PD progression under dopaminergic treatment. Current results need to be integrated and better interpreted with complete clinical data, but in a future perspective, our findings might allow the development of tailored interventions that meet distinct needs of diseased men and women, generally improving patient care.

In the frame of neurological and cerebrovascular diseases, *ischemic stroke* represents the leading cause of death and neurological impairments and there is an increasing need of deepening the research of effective biomarkers for the clinical practice and for a better understanding of the dysregulation in the pathophysiological mechanisms of the disease, also investigating molecular processes according to patients' outcomes and different phases of the disease. Using retrospective data from the Italian multicentric observational MAGIC study⁹⁹, which involved ischemic stroke patients treated with intravenous thrombolysis with the recombinant tissue plasminogen activator (rt-PA), we firstly aimed at identifying serum metabolic and lipidic predictors of three-month poor outcomes (mortality, developments of neurological and motor impairment, haemorrhagic transformation of the cerebral lesion and the non-response to the thrombolytic therapy), using samples from 243 acute ischemic stroke patients treated with rt-PA. To this purpose, logistic regression and ROC curve analyses were applied on an array of 18 metabolites and 112 lipids, quantified via 1D NOESY ¹H-NMR spectra of serum samples that have been collected before and 24h after the thrombolytic intervention (§ 4.1.2). Adjusting for clinical and demographic determinants of unfavourable outcomes, various metabolomic features, especially

lipids related to HDL, LDL, VLDL particles and ketone bodies, resulted statistically significantly associated with each of the assessed post-acute ischemic stroke poor outcome. Generally, the addition to baseline ROC curve models, related only to clinical determinants of unfavourable outcomes, of the top three statistically significant metabolic features (selected from the previous logistic regression analysis), improved the area under the curve for the prediction of three-month poor outcomes. However, statistically significant improvements were detected only when selected metabolomic features (estimated 24h after the transient cerebral ischemia) have been added to the ROC curve models including baseline clinical characteristics for the detection of impairments development, haemorrhagic transformation of the stroke and non-response to the intravenous thrombolysis with rt-PA.

Secondly, with the aim of investigating, over a period of three-months, metabolic variations with respect to the thrombolytic therapy, within-subject metabolic changes were explored in a subset of the study population ($n = 173$), considering serum samples collected before, 24h after and three-months after the thrombolysis. Many statistically significant changes ($FDR < 0.01$) are described for the majority of quantified metabolites and lipid parameters. In particular, our results suggest that the thrombolytic intervention seems to induce significant variations that reflect a general condition of energy failure, oxidative stress and systemic inflammation.

Thirdly, applying standard multivariate, univariate, network reconstruction and differential analysis of metabolite-metabolite and metabolite-lipid association networks built from an array of 18 serum metabolites and 110 lipids in a set of 248 patients from the same retrospective cohort (of which, 22 died and 82 developed impairments within three-months from acute ischemic stroke treated with intravenous rt-PA), we deeply explored differences in metabolite and lipid connectivity of patients who did not develop impairments and who survived the transient cerebral ischemia from the related opposite conditions (§ 4.1.3). We report statistically significant differences in the connectivity patterns of both low and high-molecular weight metabolites, implying underlying variations in the metabolic pathways involving leucine, glycine, glutamine, tyrosine, phenylalanine, citric, lactic and acetic acids, ketone bodies and different lipids, thus characterizing patients 'outcomes. Our results indicate that dysregulations of the above-mentioned metabolites and lipids connectivity are involved mainly in mechanisms that show how energy failure, glutamate-induced neurotoxicity, oxidative stress and neuroprotection play important roles in the progression of the pathology after the thrombolytic treatment, affecting survivor's outcomes. Acetone emerges as largely involved in the determination of both impairment development and mortality in acute ischemic stroke treated with thrombolytic therapy. In conclusion, these on-going studies provide promising information on underlying metabolic variations which occur in the serum metabolome and lipoprotein profiles of post-acute ischemic stroke patients with respect to poor outcomes. Results here reported also evidence how metabolite-metabolite and metabolite-lipid association networks of acute ischemic stroke patients differ

according to the patient's outcome, providing new insights for their actual use in the clinical practice.

As for neurodegenerative and cardiovascular diseases, cancer research has recently relied on advances in -omics sciences thanks to the progressive understanding of the biology, aetiology and the possibility of developing novel diagnostic test and therapeutic treatments. Among these, metabolomics offers new opportunities due to its important role in connecting the genotype with the phenotype. In § 4.1.4, a first NMR-based metabolomic study for high risk prostate cancer (PCa) in African men is presented. Briefly, the study identifies inflammation as a driving phenotype in the most aggressive form of PCa, and it allows the characterization of the altered metabolic and lipoprotein profiles, paving the way to a better understanding of the metabolic changes occurring in PCa and to new therapeutic strategies. To these aims, plasma samples from a cohort of South African men with PCa have been profiled using the NMR-spectroscopy and we reported that plasma of patients with very high risk and aggressive cancer, have a peculiar metabolic phenotype characterized by higher levels of the inflammatory markers GlycA and GlycB. Therefore, the results achieved in this study answer to the urgent need for new insights into the molecular mechanisms which characterize the increased rate of aggressive and lethal PCa in men of African ancestry. Last but not least, considering the high capability of metabolomics in offering cost-effective and productive route for the personalized lifestyle management, in § 4.1.5, the potential of untargeted NMR-based metabolomics in investigating variations in the human serum metabolome of subjects at risk for Metabolic Syndrome (MetS), which occurs after specific dietary interventions, is demonstrated. The present on-going study belong to the large intervention study (LIS) that was included in the frame of the EC FP7 PATHWAY-27 project with the aim of confirming or better investigating the observations obtained in the pilot studies,^{100,101} where the best enrichments, *i.e.* docosahexanoic acid (DHA), anthocyanins (AC) and oat beta-glucans (O-BG) within the best food matrices (dairy-, egg- and bakery- based foods) were selected. In particular, we applied untargeted NMR-based metabolomics to investigate, on the human serum metabolome and lipoprotein profiles of subjects at risk for MetS: *i*) the effect of dairy-based food enriched with docosahexanoic acid (C22:6, *n*-3, DHA) and O-BG; *ii*) the effect of egg-based food enriched with DHA and AC; *iii*) the effect of bakery-based food enriched with DHA and AC; *iv*) the effect of food matrices in determining the bioavailability of bioactives, administering a combination of dairy-, egg- and bakery-based foods without bioactives (placebo). To these aims, serum samples collected before (t_0) and 12 weeks after (t_1) the dietary interventions have been collected from four different centers: Italy, France, Germany and UK.

DHA, AC and O-BG are well-known bioactives able to have a positive impact on MetS,¹⁰²⁻¹⁰⁴ acting on cell regulation of lipid metabolism, inflammation and being effective in modulating risks factors. However, there are evidences of synergies between the above-mentioned bioactives,^{105,100,106} but the effect of the food matrices on bioactives combinations (*e.g.* DHA + AC or DHA + O-BG) need to be deeply investigated. Applying multivariate and univariate analyses to 1D ¹H-NMR spectra of

serum samples, our preliminary results show a synergism of DHA and AC in inducing changes mainly in the lipoprotein profiles of subjects, independently from the egg- or bakery-based food matrix that was chosen. Moreover, the daily assumption of the three selected food matrices seems to alter the whole metabolic profiles of subjects, since the placebo treatment led to the discrimination between t_0 and t_1 samples considering specific recruitment centres.

In conclusion, this study deepened the current scientific understanding of the impact of different bioactives embedded in various food matrices, especially evidencing their synergies and the role of the food matrix in bioactives delivery and proposing actual use of bakery- and egg-based foods enriched with DHA and AC on subjects at risk for metabolic syndrome.

4.1.1. Nuclear Magnetic Resonance-based metabolomics to characterize serum sex-related metabolic profiles of drug-naïve Parkinson's disease patients with respect to healthy controls and patients with advanced disease and under dopaminergic treatment

In preparation

Candidate's contributions: acquisition of NMR data, statistical analysis of data, interpreting results and writing of the manuscript.

Abstract

Here, we present multivariate and univariate approaches to infer metabolic differences which characterize the serum metabolic profiles of recently diagnosed, drug-naïve Parkinson's disease patients (dnPD), advanced PD patients (advPD) under dopaminergic treatment and healthy controls (CTR), applying Nuclear Magnetic Resonance (NMR)-based metabolomics in large training and validation cohorts of German subjects. We deeply explored metabolite and lipid variations of patient groups, also in relation to sex, evidencing a more pronounced fingerprinting of the pathology in male patients. Moreover, differences among signatures of early and advanced stages of the disease were highlighted, regardless of the dopaminergic treatment. We report and validate statistically significant variations in the concentration of ketone bodies, especially acetone, histidine, lipoproteins and cholesterol among groups under study, underlying differences associated to oxidative stress, which seems to mainly characterize the early stage of the disease from the advanced one. Ornithine was highlighted as a possible marker to follow disease progression.

Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disease, after Alzheimer's disease, affecting 1% of the population above 60 years,¹ and contributing significantly to the increasing of public health costs. PD patients suffer from a various range of motor symptoms, *i.e.* tremor, rigidity, postural instability, bradykinesia, and non-motor manifestations, such as autonomic sensory, neuropsychiatric symptoms and sleep disorders.² These symptoms are mainly caused by a progressive degeneration of dopaminergic neurons from the nigrostriatal pathway, formation of Lewy bodies and microgliosis.³ To date, the diagnosis of PD depends primarily on the onset of motor symptoms, and the rate of misdiagnosis is especially high in the first 5 years due to the clinical overlap to other movement disorders and the lack of objective biomarkers.⁴

Despite many years of investigations on metabolic perturbances in PD,⁵⁻⁹ the mechanisms of PD aetiopathogenesis, progression and the efficacy of drug treatment on the disease evolution need to be deeply explored and validated. Currently, routine analytical tests have not yet provided sufficient information to identify reliable biomarkers to detect early neurodegeneration in PD, to monitor the progression of the disease and to detect the effects of therapy intervention.

Nuclear Magnetic Resonance (NMR)-based metabolomics proved to be efficient in characterizing the metabolic signature of diseases¹⁰⁻¹⁴ and in the context of molecular epidemiology.^{15,16} Metabolomics has the potential to provide valuable insights to deepen the knowledge regarding the aetiopathogenesis of PD, to identify signatures differentiating patient groups and to reveal sensitive PD biomarkers which can be important in the early diagnosis and in monitoring disease progression and the efficacy

of drug treatments.^{17–21} To date, few metabolomic-based studies attempted to characterize blood samples of *de novo* PD patients,^{20,22} and independent, large-scale validation and comparison with other overlapping movement disorders are still needed.

In this work, we aimed of providing evidences of the existence of serum metabolic fingerprints differentiating drug-naïve (or *de novo*) PD patients and advanced PD patients under dopaminergic treatment from healthy control, also in relation to sex, using independent large training and validation cohorts and applying multivariate, univariate, logistic regression and Receiver Operating Characteristics (ROC) curve analyses. The subjects included in this study are a subset of the cohorts analysed in the H2020 Project “PROPAG-AGEING” (www.propag-ageing.eu/project). An overview of the study design is illustrated in **Figure 1**.

Material and Methods

Patient cohorts

The study population consists of a total of 329 German subjects, *i.e.* *de novo* Parkinson’s disease patients, patients in advanced stage of disease under pharmacological treatment and healthy controls. In detail, patient cohorts were divided as follows (see **Figure 1**): firstly, a training cohort consisting of 72 drug-naïve Parkinson’s disease patients (dnPD) and 59 healthy control subjects (CTR) for a total of 131 subjects as part of the longitudinal DeNoPa cohort, as previously published.^{23,24,25,26} Available samples from baseline visit were chosen with enough remaining volume. Secondly, it was selected a validation cohort, consisting of an independent set of 156 dnPD, 20 CTR and 22 samples from subjects with advanced Parkinson’s disease and under dopaminergic treatment (advPD), for a total of 198 subjects, as a part of the cross-sectional Kassel cohort. Patients enrolled in the study were clinically phenotyped before collection of samples. Phenotyping included 1,5 Tesla magnetic resonance imaging (MRI) to determine structural abnormalities, quantitative levodopa testing, smell identification test (Sniffin’ sticks, Burghardt Messtechnik GmbH, Wedel, Germany), Mini Mental Status Examination (MMSE) followed by further cognitive testing and video-supported polysomnography to determine REM sleep behaviour disorder in a subset of patients. The phenotyping was done based on these results and in accordance to established criteria for PD (UK Brain Bank Criteria)²⁷, Multiple System Atrophy (MSA), Dementia with Lewy bodies (DLB), Progressive Supranuclear Palsy, Corticobasal Degeneration (CBD), Alzheimer’s disease and Frontotemporal Dementia (FTD). Subjects with marked vascular lesions in MRI indicative of a vascular comorbidity and subjects with normal pressure hydrocephalus by MRI were excluded.

An overview of the demographic characteristics of analysed patients is reported in **Table 1**.

Ethical Issues

The study was conducted according to the Declaration of Helsinki and with informed written consent provided by all subjects. The study was approved by the ethics committee of the Physician's Board Hesse, Germany (Approval No. FF89/2008 for DeNoPa) and the University Medical Center Goettingen, Germany (Approval No. 9/7/04 and 36/7/02 for Kassel cohort).

NMR sample preparation and analyses

Serum samples were collected in the morning with the subjects fastened, centrifuged, aliquoted and frozen at -80°C until analysis. Samples have been prepared according to common standard procedures for metabolomic studies.²⁸⁻³⁰ The analytical preparation of serum samples and their NMR spectra acquisition followed the protocols detailed elsewhere.²⁹

For each serum specimen, the 1D NOESY, 1D CPMG and 1D DIFFUSION-EDITED pulse sequences were applied to acquire different types of ^1H -NMR spectra, using a Bruker 600 MHz spectrometer, with a proton Larmor frequency of 600.13 MHz and equipped with a 5 mm PATXI 1H-13C-15N and 2H decoupling probe. The instrument includes a z axis gradient coil, an automatic tuning-matching (ATM) and an automatic and refrigerate sample changer (SampleJet). To stabilize approximately, at the level of ± 0.1 K, the sample temperature, a BTO 2000 thermocouple was employed and each NMR tube was kept for about 5 min inside the NMR probe head to equilibrate the acquisition temperature of 310 K.

Spectral Processing

Before applying Fourier transform, raw data were multiplied by an exponential function of 0.3 Hz line-broadening factor. Transformed spectra were automatically corrected for phase and baseline distortions and calibrated to a reference (anomeric glucose proton signal at 5.24 ppm), using Topspin 3.2 software (Bruker BioSpin). Each 1D serum spectrum, in the range of 0.2 – 10.0 ppm, was bucketed into 0.02 ppm chemical shift segments using AMIX (version 3.8.4) software (Bruker BioSpin). Regions containing residual water signal (between 4.68 and 4.84 ppm) were removed.

Serum and lipoprotein identification and quantification

Twenty-seven metabolites (3-hydroxybutyrate, acetate, acetoacetate, acetone, alanine, citrate, creatine, creatinine, dimethylsulfone, formate, glucose, glutamine, glycine, histidine, isoleucine, lactate, leucine, methionine, N,N-dimethylglycine, ornithine, phenylalanine, sarcosine, succinate, trimethylamine N-oxide, tyrosine and valine) and 114 lipids were unambiguously identified and quantified (in terms of absolute concentrations) from 1D ^1H -NOESY NMR spectra using the AVANCE Bruker IVDr (Clinical Screening and In Vitro Diagnostics research, Bruker BioSpin)³¹ software. For all serum samples, different lipoproteins (VLDL, LDL, IDL, HDL) and different lipoprotein subclasses, classified according to density and size, for a total of 15 subclasses (VLDL-1 to VLDL-5, LDL-1 to LDL-6 and HDL-1 to HDL-4), were detected. For each main class and subclass, reported data consist in concentrations of

lipids (total cholesterol, free cholesterol, phospholipids, and triglycerides) contained in each fraction. Concentrations of apolipoproteins Apo-A1 and ApoA2 are estimated for HDL class and each relative subclass, while Apo-B concentrations are calculated for VLDL, IDL classes and all LDL subclasses.

Statistical Analysis

All data analyses were performed using R (version 3.6.1), an open source software for the statistical management of data,³² using NMR data from 1D NOESY spectra.

Exploratory Analysis

Multivariate data analysis was conducted on bucketed 1D NOESY spectra of all available samples. Principal Component Analysis (PCA) was used as a first exploratory approach³³ to investigate, in an unsupervised manner, the data structure and highlighting the possible presence of metabolite and lipids signatures differentiating dnPD, advPD patients and CTR both in the training and in the test cohorts. PCA analysis was performed on data scaled to unit variance.

Predictive modelling

Orthogonal Projections to Latent Structures Discriminant Analysis (OPLS-DA), was employed as supervised technique;³⁴ a total of nine classification models were built to discriminate CTR from dnPD in both cohorts and advPD from dnPD patients in the validation set, also including sex-dependent discrimination analyses. Models were built using bucketed 1D NOESY spectra of all available samples.

When unbalanced number of subjects has been compared, OPLS-DA models were built by reducing groups to the same size by randomly sampling, thus including in the model an equal number of subjects from each group. The procedure was repeated 100 times and results are averaged over the 100 models.

All discriminant and predictive analyses were performed on bucketed 1D-NOESY spectra without prior normalization.

Overall, for different classifications, accuracy, sensitivity and specificity were calculated according to standard definitions, by means of a Monte-Carlo cross-validation scheme (MCCV, R script in-house written). Briefly, 90% of the data were randomly chosen at each iteration as a training set to build the model. Then, the remaining 10% was tested. The full procedure was repeated 100 times to derive an average discrimination accuracy, sensitivity and specificity.

To test the efficacy of the training model in discriminating dnPD from CTR subjects, bucketed 1D-NOESY spectra of the validation serum samples were blindly projected onto the OPLS-DA score plot resulting from the training model.

Univariate Analysis

Univariate Wilcoxon test³⁵ was employed to compare metabolite and lipid concentrations between patient groups (CTR vs. dnPD, CTR vs. advPD and dnPD vs.

advPD) of both training and test cohorts and with respect to sex (we performed male and female independent comparisons). Benjamini & Hochberg method³⁶ was applied to correct for multiple testing and adjusted *P*-values (FDR) < 0.05 were considered statistically significant. Log₂ fold change (FC) ratios of the median intensities were also calculated for all analyses performed. Effect sizes were estimated for groups comparisons using Cliff's delta formulation³⁷ (Cd), which contributes to the characterization of the meaningful signals by giving an estimation of the magnitude of the separation in the different comparisons. Magnitude is evaluated using the thresholds provided in Romano *et al.*³⁸, where Cd values < 0.147 are considered “negligible”, (Cd) < 0.33 are defined “small”, while (Cd) < 0.474 are reported as “medium” and (Cd) > 0.474 are considered “large”.

Using calculated log₂ fold change ratios for all comparisons, an heatmap analysis was performed (“gplots” R package) on the most statistically significant metabolites and lipids. To its reconstruction, we select only the variables that showed FDR < 0.05 and log₂FC ≥ |0.5| values (supplementary **Table S1**) in more than one comparison performed (see **Figure 4** for details).

Logistic regression and ROC analysis

Association between each statistically significant analyte of the training cohort and the disease was assessed and validated using logistic regression in combination with ROC curve analysis. Before performing any analysis, continuous values of the analytes were standardized by centering and dividing by two standard deviations³⁹ using the “rescale” function of the R package “arm”.

Firstly, a binomial logistic regression model was built, for each statistically significant analyte found in the training cohort (*i.e.* a metabolite or a lipid concentration), using the “glm” function in the R package “stats”. These analytes (metabolites and lipids) were used as the predictors (*x*), while the dichotomic variable indicating the status (*i.e.* CTR or dnPD) was used as the dependent variable (*y*) to be predicted.

The fitted values obtained for each analyte and for each subject were used to estimate areas under the ROC curves (AUC values, using the “colAUC” function of the R package “caTools”). Subsequently, the fitted regression models built on the training set were used to predict probabilities of samples in the validation cohort (values between 0 and 1) to be classified as CTR or dnPD (“predict.glm” function in the R package “stats”). These predicted probabilities were used to calculate AUC values for the validation cohort that were further compared with AUC values reported for the training cohort.

Results and Discussion

Exploratory Analysis

As a first unsupervised approach, Principal Component Analysis (PCA) was applied on all available samples (from both training and validation cohorts) to obtain an overview of the variation in the data. **Figure 2** shows the PCA 3D score plot on

bucketed 1D NOESY spectra color-coded by subject status proving the absence of evident outlier samples.

Predictive modelling: OPLS-DA analysis

OPLS-DA analysis was chosen as supervised approach to extrapolate the hidden variables that could be used to characterize the serum metabolomic profile of dnPD patients from healthy volunteers and dnPD patients from subjects with advanced stage of the disease and under pharmacological treatment.

Firstly, the training cohort was used to explore differences between *de novo* patients and healthy subjects and to derive a discriminant serum fingerprint to correctly classify the new test samples to the corresponding category (dnPD or CTR).

The OPLS-DA model built on bucketed serum 1D NOESY spectra of the training cohort showed an evident discrimination between subject groups, reporting an overall predictive accuracy of 74.8%, a specificity of 72.9% and a sensitivity of 76.4% for the classification of dnPD and CTR (**Table 2, Figure S1**).

Since it has been widely demonstrated that differences between two sexes could affect manifestation, epidemiology and pathophysiology of many diseases and sex discrimination is apparent in metabolomics profiles,⁴⁰⁻⁴⁵ two independent cross-validated models were created for male (**Table 2, Figure S1**) and female training subjects (**Table 2, Figure S1**). As shown in **Table 2**, between the two sexes, the male model ($n=76$) performs better, with an overall predictive accuracy of 73.5% compared to the 60.2% predictive accuracy of the female model ($n=55$).

The OPLS-DA model built on the re-sampled bucketed 1D NOESY spectra of advPD and dnPD patients shows an overall mean predictive accuracy of 88.9%, a mean specificity of 90.1 and a mean sensitivity of 87.1, evidencing a clear discrimination between *de novo* Parkinson's disease patients and those under dopaminergic treatment. Considering the heterogeneity of the disease, we can speculate that such differences may be related to different underlying mechanisms which characterize the more cognitive decline of late PD. However, we cannot exclude a potential variation of the serum metabolome of advPD patients induced by the pharmacological treatment; indeed, further investigations should be needed. Recently, the potentiality of multivariate analysis in revealing metabolic changes among patient groups in order to discover signatures differentiating early PD from disease progression, has been reported.⁴⁶

In our study, the availability of a test cohort allowed the validation of the existence of a serum metabolic fingerprint differentiating healthy from *de novo* drug-naïve patients. Indeed, the efficacy of the global training model in discriminating dnPD serum samples from CTR, was tested using bucketed 1D-NOESY spectra from the validation set which were blindly projected onto the OPLS-DA discrimination space obtained from the previously described training model including all female and male subjects (**Figure S2 A, B**). New test samples are predicted with an overall accuracy of 74.4%, a sensitivity of 75.6% and a specificity of 65% (**Table 3, Figure S2 A, B**), thus confirming the presence of a fingerprint discriminating CTR from dnPD drug-free

patients. Further, projecting onto the OPLS-DA discrimination space of the training model, the 22 bucketed serum 1D-NOESY spectra of advPD subjects, serum metabolic signatures characterizing healthy from diseased and different stages of the pathology were definitely highlighted.

Indeed, all the advPD patients resulted to be correctly classified as PD patients, with an accuracy of 100% (**Figure S2 C**). This last result supports the idea of the presence of a specific metabolomic signature of PD in sera of affected patients, regardless of the dopaminergic treatment, since the use of PD medication from advPD patients does not influence the classification of those subjects in a model where only patients free from L-DOPA administration are included.

Afterwards, all dnPD and CTR samples, received for both the training and validation, were used together to increase the number of subjects in the OPLS-DA model. Mixing the two cohorts (228 dnPD and 79 matched CTR), the final re-sampled OPLS-DA model reported a mean accuracy of 76.3%, a mean sensitivity of 76.1% and a mean specificity of 76.5% (**Table 4**).

In addition, to fully confirm previous results related to sex-differences, two separated sex-dependent models were created using the combined cohort. Mean values of accuracy, sensitivity and specificity were obtained for male and female OPLS-DA models. The obtained results confirm what previously reported. Indeed, male dnPD patients resulted to be discriminated from CTR better than women (mean accuracy of male of the overall model: 75.0%; mean accuracy of women of the overall model: 64.7%, as reported in **Table 4**).

Increasing evidence pointed to biological sex as a crucial determining factor in the development and phenotypical expression of PD. It has been reported that the risk of developing PD is twice higher in men than in women.^{47,48} In our case, the observation of a more pronounced discriminating metabolic fingerprint of PD in men could support the idea that the disease development might involve different pathogenetic mechanisms in male and female subjects.

Univariate Analysis

Metabolites and lipids were initially compared across all dnPD and CTR subjects of the training cohort to reveal metabolic changes between the two groups. A total of 25 metabolic features resulted to be statistically significantly different (FDR < 0.05), independently from the sex: acetone, ornithine and phenylalanine appear to be significantly higher in dnPD patients when compared to healthy individuals, while all the 22 statistically significant lipid fractions concentrations are decreased in dnPD group (**Figure 3**).

To deeply explore metabolic variations, serum metabolic profiles of all CTR, dnPD and advPD patients have been investigated, also considering sex-related comparisons. **Figure 4** reports the heatmap and the cluster analysis of 37 selected metabolic features across the different analyses we performed.

We observed increased levels of ketone bodies (acetone, acetoacetate and 3-hydroxybutyrate), decreased levels of LDL particles (especially LDL-3 and LDL-4

related parameters) and free cholesterol related to IDL particles in dnPD patients when compared to healthy subjects, considering both training, test cohorts and sex-dependent comparisons. Saliva ketone bodies, such as acetoacetate and acetoin, were found to be elevated in early PD patients with respect to healthy controls, indicating a probable decreased utilization of these compounds and their association with mitochondrial dysfunction.^{49,50} AdvPD patients show significant higher levels of ornithine and lower levels of histidine when compared to dnPD and CTR of the validation set. These observations are in line with a general condition of oxidative stress and increased oxidative stress defences that seems to characterize the serum profiles of *de novo* drug-untreated patients. Changes in the metabolism of ketone bodies occur during stress conditions and in the past, the biochemistry of ketogenesis and its role in neurological diseases and oxidative stress has been deeply explored.⁵¹ Ketone bodies seems to have the chemical potential to be active antioxidants,⁵¹ proving benefits in diseases associated with oxidative stress, as in the case of Parkinson's disease. However, there are conflicting evidences for the antioxidant role of ketone bodies, especially acetoacetate seems to have a more pronounced pro-oxidant action.⁵¹ The antioxidant/pro-oxidant dichotomy of these substances can depend largely on the complexity of mammalian metabolism that can derive more free radicals from ketone bodies, thus favouring a general body condition of oxidative stress which may exacerbate the progression of the pathology.

To date, the exact mechanisms of serum acetone in determining the onset of PD remain not completely elucidated, but our results support previous evidences of the role of acetone in PD.¹⁷

Higher acetone, 3-hydroxybutyrate and acetate levels have been described also in blood samples of other neurodegenerative diseases, such as multiple sclerosis⁵² and amyotrophic lateral sclerosis⁵³ patients.

Histidine resulted to be statistically significantly higher in dnPD when compared with advPD patients and this metabolite can act as an antioxidant since the presence of its imidazole ring helps in scavenging ROS. Increased concentrations of histidine and phenylalanine have been recently observed also in saliva of PD patients,⁵⁰ thus suggesting alterations in neurotransmitters, especially dopamine.⁵⁴

Regarding serum ornithine, increased levels have been previously detected in the advanced PD group.⁹ By arginase activity on arginine, ornithine lead to the formation of urea whose related pathways seems to be perturbed in PD.⁹ Why elevated serum levels of this metabolite are associated with the pathological mechanisms of PD progression remains unclear, but our results support the hypothesis that serum ornithine concentrations can be a marker of the advancement of the pathology.

Several studies pose lipids as central players in the Parkinson's disease:⁵⁵⁻⁶² various subclasses of fatty acyls, glycerolipids, phospholipids, sterol and lipoproteins contribute to PD pathogenesis, even if sometimes, controversial, fragmented and not always reproducible information is available from the literature. Our results evidenced a statistically significant decrease of concentrations of small low-density lipoproteins in dnPD patients. Lower LDL-related parameters, especially LDL-cholesterol levels

have been associated with higher PD risk^{57,58,62} and generally, small dense LDL particles (sdLDL) are more susceptible to oxidation than larger LDLs. Therefore, we suggest that sdLDL particles may provide an optimal substrate for ROS oxidative action which appears to be increased in dnPD patients. Moreover, univariate analysis reported also decreased levels of free-cholesterol associated to IDL in dnPD when compared both to healthy volunteers and patients with advanced disease. This observation supports previous evidences of a prominent role of cholesterol in PD.^{59–61} In conclusion, from the cluster analysis (**Figure 4**), we obtain two main clusters, corresponding to the levels of advPD patients' metabolites and dnPD patients' metabolites, respectively; corroborating the existence of a different metabolomic profile in this two group of subjects. The second mentioned cluster is in turn composed by two main subgroups. Looking at them it clearly emerges that these groups are categorized according to the cohort of origin. It is clear that differences between the two cohorts exist, and in this lies the importance of external validation.

More in detail, from the clustering analysis (**Figure 4**), we observed that comparisons including only male subjects tend to cluster with the relative sex-independent evaluation, thereby supporting results from the multivariate analysis, where the discrimination accuracy of the training male model was reported as almost similar to that of the overall predictive model (**Table 2**).

Logistic regression and ROC analysis

Several binomial logistic regression models were built using the statistically significant metabolites obtained by comparing the serum profile of CTR and dnPD patients of the training cohort (see **Figure 3**). The corresponding AUCs (Areas Under the Curves) have been calculated (**Table 5**). Looking at the measures of separability between the two groups, all the selected metabolites and lipids have a >60% chance to distinguish between case (dnPD) and control (CTR) groups, but among them, only phenylalanine (AUC: 0.715, OR: 4.91) and acetone (AUC: 0.696, OR: 3.42) appear to be the most associated metabolites to the early PD detection.

The existence of metabolic biomarkers effectively related to the disease condition was confirmed after considering the validation set of samples. Indeed, using each selected variable, we estimated for all subjects of the validation set, the predicted probability to be classified as the control (CTR, probability = 0) or the case group (dnPD, probability = 1) and for each analyte, AUC values were estimated (**Table 5**). Of them, acetone and cholesterol resulted to have higher chance to discriminate dnPD from healthy controls (AUC values are respectively of 0.793 and 0.78), highlighting acetone as a potential biomarker implicated in early onset of Parkinson's disease, and evidencing the prominent role of cholesterol in characterizing dnPD from CTR. Lower serum cholesterol levels were previously reported to occur in PD patients, independent of nutritional status and body mass index (BMI) and dysregulation of cholesterol trafficking was shown to be involved in the pathogenesis of neurodegeneration in PD.²³

Phenylalanine even showing a $\log_2FC < |0.5|$ (see **Figure 3** and **Table S1**), resulted to be able to discriminate dnPD from CTR in the training set, but it is not validated in the test cohort. LDL particles and related parameters showed a >70% chance to distinguish between CTR and dnPD, confirming what previously discussed in the univariate analysis section.

Conclusions

To our knowledge, this is the first large NMR-based metabolomic study dealing with the characterization of serum profiles of *de novo* Parkinson's disease patients free from dopaminergic treatment compared to advanced PD patients undergoing dopaminergic treatment and healthy sex/age-matched volunteers.

Here, 1D NMR fingerprinting analysis succeeded to discriminate between subject groups, identifying characteristic signatures describing early and advanced stages of the disease, although advPD patients are pharmacologically treated. A more pronounced fingerprinting of PD disease was evidenced in men with respect to women. The use of a large number of subjects ($n=329$) and the presence of an external validation cohort represent major strength points of this study.

Applying univariate, logistic regression and receiver operating characteristic analyses, we have also shown that increased levels of ketone bodies, particularly acetone, histidine and decreased levels of lipid (small dense LDL particles and related parameters, free cholesterol related to LDL) in dnPD patients, suggest that the early stage of PD may be characterized by increased oxidative defences and a worsening of the oxidative stress status. However, further biological investigations are needed to deeply explore how an unbalance between body's oxidative defences and increasing in ROS level might lead to the pathology's progression.

Our results also support a potential role of ornithine in monitoring PD progression under dopaminergic treatment.

In conclusion, this study provides new insights regarding the different serum metabolic features involved in the characterization of early and advanced stages of the Parkinson's disease, also in relation to the two sexes and the dopaminergic treatment. In a future perspective, our findings might allow the development of tailored interventions that meet distinct needs of men and women, generally improving patient care.

References

- (1) Tysnes, O.-B.; Storstein, A. Epidemiology of Parkinson's Disease. *J. Neural Transm. Vienna Austria* 1996 2017, 124 (8), 901–905. <https://doi.org/10.1007/s00702-017-1686-y>.
- (2) Xia, R.; Mao, Z.-H. Progression of Motor Symptoms in Parkinson's Disease. *Neurosci. Bull.* 2012, 28 (1), 39–48. <https://doi.org/10.1007/s12264-012-1050-z>.

- (3) Dexter, D. T.; Jenner, P. Parkinson Disease: From Pathology to Molecular Disease Mechanisms. *Free Radic. Biol. Med.* 2013, 62, 132–144. <https://doi.org/10.1016/j.freeradbiomed.2013.01.018>.
- (4) Beach T. G., Adler C.H., Importance of Low Diagnostic Accuracy for Early Parkinson's Disease. *Mov Disord.* 2018, 33(10): 1551–1554. <https://doi:10.1002/mds.27485>.
- (5) Agarwal, P. A.; Stoessl, A. J. Biomarkers for Trials of Neuroprotection in Parkinson's Disease. *Mov. Disord. Off. J. Mov. Disord. Soc.* 2013, 28 (1), 71–85. <https://doi.org/10.1002/mds.25065>.
- (6) Trezzi, J.-P.; Galozzi, S.; Jaeger, C.; Barkovits, K.; Brockmann, K.; Maetzler, W.; Berg, D.; Marcus, K.; Betsou, F.; Hiller, K.; Mollenhauer, B. Distinct Metabolomic Signature in Cerebrospinal Fluid in Early Parkinson's Disease. *Mov. Disord. Off. J. Mov. Disord. Soc.* 2017, 32 (10), 1401–1408. <https://doi.org/10.1002/mds.27132>.
- (7) Chang, K.-H.; Cheng, M.-L.; Tang, H.-Y.; Huang, C.-Y.; Wu, Y.-R.; Chen, C.-M. Alternations of Metabolic Profile and Kynurenine Metabolism in the Plasma of Parkinson's Disease. *Mol. Neurobiol.* 2018, 55 (8), 6319–6328. <https://doi.org/10.1007/s12035-017-0845-3>.
- (8) Phelan, M. M.; Caamaño-Gutiérrez, E.; Gant, M. S.; Grosman, R. X.; Madine, J. Using an NMR Metabolomics Approach to Investigate the Pathogenicity of Amyloid-Beta and Alpha-Synuclein. *Metabolomics* 2017, 13 (12). <https://doi.org/10.1007/s11306-017-1289-5>.
- (9) Hatano, T.; Saiki, S.; Okuzumi, A.; Mohny, R. P.; Hattori, N. Identification of Novel Biomarkers for Parkinson's Disease by Metabolomic Technologies. *J. Neurol. Neurosurg. Psychiatry* 2016, 87 (3), 295–301. <https://doi.org/10.1136/jnnp-2014-309676>.
- (10) Meoni, G.; Lorini, S.; Monti, M.; Madia, F.; Corti, G.; Luchinat, C.; Zignego, A. L.; Tenori, L.; Gragnani, L. The Metabolic Fingerprints of HCV and HBV Infections Studied by Nuclear Magnetic Resonance Spectroscopy. *Sci. Rep.* 2019, 9 (1), 4128. <https://doi.org/10.1038/s41598-019-40028-4>.
- (11) Vignoli, A.; Orlandini, B.; Tenori, L.; Biagini, M. R.; Milani, S.; Renzi, D.; Luchinat, C.; Calabrò, A. S. Metabolic Signature of Primary Biliary Cholangitis and Its Comparison with Celiac Disease. *J. Proteome Res.* 2019, 18 (3), 1228–1236. <https://doi.org/10.1021/acs.jproteome.8b00849>.
- (12) Vignoli, A.; Paciotti, S.; Tenori, L.; Eusebi, P.; Biscetti, L.; Chiasserini, D.; Scheltens, P.; Turano, P.; Teunissen, C.; Luchinat, C.; Parnetti, L. Fingerprinting Alzheimer's Disease by ¹H Nuclear Magnetic Resonance Spectroscopy of Cerebrospinal Fluid. *J. Proteome Res.* 2020, 19 (4), 1696–1705. <https://doi.org/10.1021/acs.jproteome.9b00850>.
- (13) Caracausi, M.; Ghini, V.; Locatelli, C.; Mericio, M.; Piovesan, A.; Antonaros, F.; Pelleri, M. C.; Vitale, L.; Vacca, R. A.; Bedetti, F.; Mimmi, M. C.; Luchinat, C.; Turano, P.; Strippoli, P.; Cocchi, G. Plasma and Urinary Metabolomic Profiles of

- Down Syndrome Correlate with Alteration of Mitochondrial Metabolism. *Sci. Rep.* 2018, 8 (1), 2977. <https://doi.org/10.1038/s41598-018-20834-y>.
- (14) Ghini, V.; Unger, F. T.; Tenori, L.; Turano, P.; Juhl, H.; David, K. A. Metabolomics Profiling of Pre-and Post-Anesthesia Plasma Samples of Colorectal Patients Obtained via Ficoll Separation. *Metabolomics* 2015, 11 (6), 1769–1778. <https://doi.org/10.1007/s11306-015-0832-5>.
- (15) Vignoli, A.; Tenori, L.; Giusti, B.; Takis, P. G.; Valente, S.; Carrabba, N.; Balzi, D.; Barchielli, A.; Marchionni, N.; Gensini, G. F.; Marcucci, R.; Luchinat, C.; Gori, A. M. NMR-Based Metabolomics Identifies Patients at High Risk of Death within Two Years after Acute Myocardial Infarction in the AMI-Florence II Cohort. *BMC Med.* 2019, 17 (1), 3. <https://doi.org/10.1186/s12916-018-1240-2>.
- (16) Di Donato, S.; Mislav, A. R.; Vignoli, A.; Mori, E.; Vitale, S.; Biagioni, C.; Hart, C.; Becheri, D.; Del Monte, F.; Luchinat, C.; Di Leo, A.; Mottino, G.; Tenori, L.; Biganzoli, L. Serum Metabolomic as Biomarkers to Differentiate Early from Metastatic Disease in Elderly Colorectal Cancer (Crc) Patients. *Ann. Oncol.* 2016, 27. <https://doi.org/10.1093/annonc/mdw335.20>.
- (17) Nagesh Babu, G.; Gupta, M.; Paliwal, V. K.; Singh, S.; Chatterji, T.; Roy, R. Serum Metabolomics Study in a Group of Parkinson's Disease Patients from Northern India. *Clin. Chim. Acta Int. J. Clin. Chem.* 2018, 480, 214–219. <https://doi.org/10.1016/j.cca.2018.02.022>.
- (18) Shao, Y.; Le, W. Recent Advances and Perspectives of Metabolomics-Based Investigations in Parkinson's Disease. *Mol. Neurodegener.* 2019, 14 (1), 3. <https://doi.org/10.1186/s13024-018-0304-2>.
- (19) Lei, S.; Powers, R. NMR Metabolomics Analysis of Parkinson's Disease. *Curr. Metabolomics* 2013, 1 (3), 191–209. <https://doi.org/10.2174/2213235X113019990004>.
- (20) Troisi, J.; Landolfi, A.; Vitale, C.; Longo, K.; Cozzolino, A.; Squillante, M.; Savanelli, M. C.; Barone, P.; Amboni, M. A Metabolomic Signature of Treated and Drug-Naïve Patients with Parkinson's Disease: A Pilot Study. *Metabolomics* 2019, 15 (6), 90. <https://doi.org/10.1007/s11306-019-1554-x>.
- (21) Takis, P. G.; Ghini, V.; Tenori, L.; Turano, P.; Luchinat, C. Uniqueness of the NMR Approach to Metabolomics. *TrAC Trends Anal. Chem.* 2019, 120, 115300. <https://doi.org/10.1016/j.trac.2018.10.036>.
- (22) D'Andrea, G.; Pizzolato, G.; Gucciardi, A.; Stocchero, M.; Giordano, G.; Baraldi, E.; Leon, A. Different Circulating Trace Amine Profiles in De Novo and Treated Parkinson's Disease Patients. *Sci. Rep.* 2019, 9 (1), 6151. <https://doi.org/10.1038/s41598-019-42535-w>.
- (23) Mollenhauer B., Trautmann E., Sixel-Döring F., Wicke T., Ebentheuer J., Schaumburg M., Lang E., Focke N. F., Kumar K. R., Lohmann K., Klein C., Schlossmacher M.G, Kohnen R., Friede T, Trenkwalder C., On behalf of the DeNoPa Study Group. Nonmotor and diagnostic findings in subjects with de novo Parkinson disease of the DeNoPa cohort. *Neurology.* 2013, October 01, 2013; 81 (14) <https://doi.org/10.1212/WNL.0b013e3182a6cbd5>.

- (24) Mollenhauer B., Zimmermann J., Sixel-Döring F., Focke N. K., Wicke T., Ebentheuer J., Schaumburg M., Lang E., Trautmann E., Zetterberg H., Taylor P., Friede T., Trenkwalder C. Monitoring of 30 marker candidates in early Parkinson disease as progression markers. *Neurology*. 2016, July 12, 2016; 87 (2) <https://doi.org/10.1212/WNL.0000000000002651>.
- (25) Sixel-Döring F., Zimmermann J., Wegener A., Mollenhauer B., Trenkwalder C. The Evolution of REM Sleep Behavior Disorder in Early Parkinson Disease. *Sleep*. 2016, Volume 39, Issue 9, Pages 1737–1742, <https://doi.org/10.5665/sleep.6102>.
- (26) Stüendl A., Kunadt M., Kruse N., Bartels C., Moebius W., Danzer K. M., Mollenhauer B., Schneider A. Induction of α -synuclein aggregate formation by CSF exosomes from patients with Parkinson's disease and dementia with Lewy bodies. *Brain*. 2016, Volume 139, Issue 2, February 2016, Pages 481–494. <https://doi.org/10.1093/brain/awv346>.
- (27) Hughes A. J., Daniel S. E., Blankson S., Lees A. J. A clinicopathologic study of 100 cases of Parkinson's disease. *Arch. Neurol.* 1993, Feb;50(2):140-8. doi: 10.1001/archneur.1993.00540020018011.
- (28) Bernini, P.; Bertini, I.; Luchinat, C.; Nincheri, P.; Staderini, S.; Turano, P. Standard Operating Procedures for Pre-Analytical Handling of Blood and Urine for Metabolomic Studies and Biobanks. *J. Biomol. NMR* 2011, 49 (3–4), 231–243. <https://doi.org/10.1007/s10858-011-9489-1>.
- (29) Vignoli, A.; Ghini, V.; Meoni, G.; Licari, C.; Takis, P. G.; Tenori, L.; Turano, P.; Luchinat, C. High-Throughput Metabolomics by 1D NMR. *Angew. Chem. Int. Ed Engl.* 2019, 58 (4), 968–994. <https://doi.org/10.1002/anie.201804736>.
- (30) Ghini, V.; Quaglio, D.; Luchinat, C.; Turano, P. NMR for Sample Quality Assessment in Metabolomics. *New Biotechnol.* 2019, 52, 25–34. <https://doi.org/10.1016/j.nbt.2019.04.004>.
- (31) Jiménez, B.; Holmes, E.; Heude, C.; Tolson, R. F.; Harvey, N.; Lodge, S. L.; Chetwynd, A. J.; Cannet, C.; Fang, F.; Pearce, J. T. M.; Lewis, M. R.; Viant, M. R.; Lindon, J. C.; Spraul, M.; Schäfer, H.; Nicholson, J. K. Quantitative Lipoprotein Subclass and Low Molecular Weight Metabolite Analysis in Human Serum and Plasma by ^1H NMR Spectroscopy in a Multilaboratory Trial. *Anal. Chem.* 2018, 90 (20), 11962–11971. <https://doi.org/10.1021/acs.analchem.8b02412>.
- (32) Ihaka, R.; Gentleman, R. R: A Language for Data Analysis and Graphics. *J Comput Stat Graph* 1996, 5, 299–314.
- (33) Serneels, S.; Verdonck, T. Principal Component Analysis for Data Containing Outliers and Missing Elements. *Comput. Stat. Data Anal.* 2008, 52 (3), 1712–1727.
- (34) Ebbels, T. M. D. Chapter 7 - Non-Linear Methods for the Analysis of Metabolic Profiles. In *The Handbook of Metabonomics and Metabolomics*; Elsevier Science B.V.: Amsterdam, 2007; pp 201–226.
- (35) Neuhäuser, M. Wilcoxon–Mann–Whitney Test. In *International Encyclopedia of Statistical Science*; Springer, Berlin, Heidelberg, 2011; pp 1656–1658. https://doi.org/10.1007/978-3-642-04898-2_615.

- (36) Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* 1995, 289–300.
- (37) Cliff, N. *Ordinal Methods for Behavioral Data Analysis*; Psychology Press: Mahwah, N.J, 1996.
- (38) Romano, J.; Kromrey, J.; Coraggio, J.; Skowronek, J. Appropriate Statistics for Ordinal Level Data: Should We Really Be Using t-Test and Cohen'sd for Evaluating Group Differences on the NSSE and Other Surveys? | *BibSonomy*; 2006.
- (39) Gelman, A. Scaling Regression Inputs by Dividing by Two Standard Deviations. *Stat. Med.* 2008, 27 (15), 2865–2873. <https://doi.org/10.1002/sim.3107>.
- (40) Vignoli, A.; Tenori, L.; Luchinat, C.; Saccenti, E. Age and Sex Effects on Plasma Metabolite Association Networks in Healthy Subjects. *J. Proteome Res.* 2018, 17 (1), 97–107. <https://doi.org/10.1021/acs.jproteome.7b00404>.
- (41) Wallner-Liebmann, S.; Gralka, E.; Tenori, L.; Konrad, M.; Hofmann, P.; Dieber-Rotheneder, M.; Turano, P.; Luchinat, C.; Zatloukal, K. The Impact of Free or Standardized Lifestyle and Urine Sampling Protocol on Metabolome Recognition Accuracy. *Genes Nutr.* 2015, 10 (1), 441. <https://doi.org/10.1007/s12263-014-0441-3>.
- (42) Wallner-Liebmann, S.; Tenori, L.; Mazzoleni, A.; Dieber-Rotheneder, M.; Konrad, M.; Hofmann, P.; Luchinat, C.; Turano, P.; Zatloukal, K. Individual Human Metabolic Phenotype Analyzed by (1)H NMR of Saliva Samples. *J. Proteome Res.* 2016, 15 (6), 1787–1793. <https://doi.org/10.1021/acs.jproteome.5b01060>.
- (43) Assfalg, M.; Bertini, I.; Colangiuli, D.; Luchinat, C.; Schäfer, H.; Schütz, B.; Spraul, M. Evidence of Different Metabolic Phenotypes in Humans. *Proc. Natl. Acad. Sci.* 2008, 105 (5), 1420–1424. <https://doi.org/10.1073/pnas.0705685105>.
- (44) Bernini, P.; Bertini, I.; Luchinat, C.; Nepi, S.; Saccenti, E.; Schäfer, H.; Schütz, B.; Spraul, M.; Tenori, L. Individual Human Phenotypes in Metabolic Space and Time. *J. Proteome Res.* 2009, 8 (9), 4264–4271. <https://doi.org/10.1021/pr900344m>.
- (45) Ghini, V.; Saccenti, E.; Tenori, L.; Assfalg, M.; Luchinat, C. Allostasis and Resilience of the Human Individual Metabolic Phenotype. *J. Proteome Res.* 2015, 14 (7), 2951–2962. <https://doi.org/10.1021/acs.jproteome.5b00275>.
- (46) D'Andrea G., Pizzolato G., Gucciardi A., Stocchero M., Giordano G., Baraldi E., Leon A. Different Circulating Trace Amine Profiles in De Novo and Treated Parkinson's Disease Patients. 2019. *Sci. Rep* 9:6151 | <https://doi.org/10.1038/s41598-019-42535-w>.
- (47) Cerri, S.; Mus, L.; Blandini, F. Parkinson's Disease in Women and Men: What's the Difference? *J. Park. Dis.* 2019, 9 (3), 501–515. <https://doi.org/10.3233/JPD-191683>.
- (48) Baldereschi, M.; Di Carlo, A.; Rocca, W. A.; Vanni, P.; Maggi, S.; Perissinotto, E.; Grigoletto, F.; Amaducci, L.; Inzitari, D. Parkinson's Disease and Parkinsonism in a Longitudinal Study: Two-Fold Higher Incidence in Men. ILSA Working Group. *Italian Longitudinal Study on Aging. Neurology* 2000, 55 (9), 1358–1363. <https://doi.org/10.1212/wnl.55.9.1358>.

- (49) Kori, M.; Aydın, B.; Unal, S.; Arga, K. Y.; Kazan, D. Metabolic Biomarkers and Neurodegeneration: A Pathway Enrichment Analysis of Alzheimer's Disease, Parkinson's Disease, and Amyotrophic Lateral Sclerosis. *Omics J. Integr. Biol.* 2016, 20 (11), 645–661. <https://doi.org/10.1089/omi.2016.0106>.
- (50) Kumari, S.; Goyal, V.; Kumaran, S. S.; Dwivedi, S. N.; Srivastava, A.; Jagannathan, N. R. Quantitative Metabolomics of Saliva Using Proton NMR Spectroscopy in Patients with Parkinson's Disease and Healthy Controls. *Neurol. Sci. Off. J. Ital. Neurol. Soc. Ital. Soc. Clin. Neurophysiol.* 2020, 41 (5), 1201–1210. <https://doi.org/10.1007/s10072-019-04143-4>.
- (51) McPherson, P. A. C.; McEneny, J. The Biochemistry of Ketogenesis and Its Role in Weight Management, Neurological Disease and Oxidative Stress. *J. Physiol. Biochem.* 2012, 68 (1), 141–151. <https://doi.org/10.1007/s13105-011-0112-4>.
- (52) Cocco, E.; Murgia, F.; Loreface, L.; Barberini, L.; Poddighe, S.; Frau, J.; Fenu, G.; Coghe, G.; Murru, M. R.; Murru, R.; Del Carratore, F.; Atzori, L.; Marrosu, M. G. ¹H-NMR Analysis Provides a Metabolomic Profile of Patients with Multiple Sclerosis. *Neurol. Neuroimmunol. Neuroinflammation* 2015, 3 (1). <https://doi.org/10.1212/NXI.0000000000000185>.
- (53) Kumar, A.; Bala, L.; Kalita, J.; Misra, U. K.; Singh, R. L.; Khetrapal, C. L.; Babu, G. N. Metabolomic Analysis of Serum by (1) H NMR Spectroscopy in Amyotrophic Lateral Sclerosis. *Clin. Chim. Acta Int. J. Clin. Chem.* 2010, 411 (7–8), 563–567. <https://doi.org/10.1016/j.cca.2010.01.016>.
- (54) Figura, M.; Kuśmierska, K.; Bucior, E.; Szlufik, S.; Koziorowski, D.; Jamrozik, Z.; Janik, P. Serum Amino Acid Profile in Patients with Parkinson's Disease. *PLOS ONE* 2018, 13 (1), e0191670. <https://doi.org/10.1371/journal.pone.0191670>.
- (55) Xicoy, H.; Wieringa, B.; Martens, G. J. M. The Role of Lipids in Parkinson's Disease. *Cells* 2019, 8 (1). <https://doi.org/10.3390/cells8010027>.
- (56) Pizarro, C.; Esteban-Díez, I.; Espinosa, M.; Rodríguez-Royo, F.; González-Sáiz, J.-M. An NMR-Based Lipidomic Approach to Identify Parkinson's Disease-Stage Specific Lipoprotein–Lipid Signatures in Plasma. *Analyst* 2019, 144 (4), 1334–1344. <https://doi.org/10.1039/C8AN01778F>.
- (57) Zhang, L.; Wang, X.; Wang, M.; Sterling, N. W.; Du, G.; Lewis, M. M.; Yao, T.; Mailman, R. B.; Li, R.; Huang, X. Circulating Cholesterol Levels May Link to the Factors Influencing Parkinson's Risk. *Front. Neurol.* 2017, 8, 501. <https://doi.org/10.3389/fneur.2017.00501>.
- (58) Huang, X.; Chen, H.; Miller, W. C.; Mailman, R. B.; Woodard, J. L.; Chen, P. C.; Xiang, D.; Murrow, R. W.; Wang, Y.-Z.; Poole, C. Lower Low-Density Lipid Cholesterol Levels Are Associated with Parkinson's Disease. *Mov. Disord. Off. J. Mov. Disord. Soc.* 2007, 22 (3), 377–381. <https://doi.org/10.1002/mds.21290>.
- (59) Hu, G. Total Cholesterol and the Risk of Parkinson's Disease: A Review for Some New Findings <https://www.hindawi.com/journals/pd/2010/836962/> (accessed May 19, 2020). <https://doi.org/10.4061/2010/836962>.

(60) Lau, D.; L, L. M.; Koudstaal, P. J.; Hofman, A.; Breteler, M. M. B. Serum Cholesterol Levels and the Risk of Parkinson's Disease. *Am. J. Epidemiol.* 2006, 164 (10), 998–1002. <https://doi.org/10.1093/aje/kwj283>.

(61) Gudala, K.; Bansal, D.; Muthyala, H. Role of Serum Cholesterol in Parkinson's Disease: A Meta-Analysis of Evidence. *J. Park. Dis.* 2013, 3 (3), 363–370. <https://doi.org/10.3233/JPD-130196>.

(62) Huang, X.; Abbott, R. D.; Petrovitch, H.; Mailman, R. B.; Ross, G. W. Low LDL Cholesterol and Increased Risk of Parkinson's Disease: Prospective Results from Honolulu-Asia Aging Study. *Mov. Disord. Off. J. Mov. Disord. Soc.* 2008, 23 (7), 1013–1018. <https://doi.org/10.1002/mds.22013>.

Tables**Table 1.** Main demographic characteristics of the population under study; mean \pm standard deviation and *P*-values (*P*) are reported.

	Training cohort			Validation cohort					
	dnPD	CTR	<i>P</i>	dnPD	CTR	<i>P</i>	advPD	<i>P</i> (advPD vs dnPD)	<i>P</i> (advPD vs CTR)
Age	65.1 \pm 9.4	64.5 \pm 6.9	0.68	65 \pm 11.4	71.7 \pm 5.1	6 \cdot 10 ⁻⁵	68.9 \pm 7.3	0.05	0.16
Sex (male/tot)	40/72	36/59	/	83/156	8/20	/	15/22	/	/
BMI	27.7 \pm 5	26.7 \pm 3.8	0.25	27.1 \pm 4.9	26.2 \pm 3.2	0.35	25.9 \pm 3.7	0.25	0.8
UPDRS III	19 \pm 10.2	0.4 \pm 0.9	7.4 \cdot 10 ⁻²⁴	23 \pm 12.8	/	/	34.4 \pm 15.8	0.003	/
Hoehn and Yahr stage	1.8 \pm 0.6	0	1.6 \cdot 10 ⁻³⁵	2 \pm 0.8	/	/	3.1 \pm 0.6	6.7 \cdot 10 ⁻⁸	/
MMSE	28.4 \pm 1.3	28.7 \pm 1.2	0.27	28 \pm 1.9	/	/	5.4 \pm 0.002	0.002	/

Table 2. Performances of the OPLS-DA 1D-NOESY models discriminating dnPD patients from CTR subjects of the training cohort. Overall model (considering all the samples from the training cohort), male and female models (considering separately male and female training groups of cases and controls). Accuracy%, specificity% and sensitivity% are reported.

	Overall	Male	Female
Accuracy %	74.8	73.5	60.2
Specificity %	72.9	74.2	58.6
Sensitivity %	76.4	72.9	61.4

Table 3. Confusion matrix of CTR and dnPD subjects tested on the overall OPLS-DA training model.

		Actual Class	
		CTR	dnPD
Predicted Class	CTR	13 <i>True Negative</i>	7 <i>False Negative</i>
	dnPD	38 <i>False Positive</i>	118 <i>True Positive</i>

Table 4. Performances of the OPLS-DA 1D-NOESY models discriminating dnPD patients from CTR subjects combining the training and the test cohorts. Overall model (considering all the samples), male and female models (considering separately male and female groups of cases and controls). Accuracy%, specificity% and sensitivity% are reported.

	Overall	Male	Female
<i>Accuracy %</i>	76.3	75	64.7
<i>Specificity %</i>	76.5	76.6	62.6
<i>Sensitivity %</i>	76.1	73.4	66.9

Table 5. AUC values for training and test samples. For training binomial logistic regression models, Odds Ratio (OR), 95% Confidence Interval (CI), *P*-value and related values adjusted with the Benjamini-Hochberg correction (FDR) are also reported.

	OR (95% CI)	<i>P</i> -value	FDR	AUC TRAINING	AUC TEST
Ornithine	5.74 (2.33 - 14.11)	$1.4 \cdot 10^{-4}$	$2.75 \cdot 10^{-3}$	0.681	0.518
Phenylalanine	4.91 (2.11 - 11.41)	$2.2 \cdot 10^{-4}$	$2.75 \cdot 10^{-3}$	0.715	0.504
Acetone	3.42 (1.49 - 7.87)	$3.75 \cdot 10^{-3}$	$4.94 \cdot 10^{-3}$	0.696	0.793
Cholesterol	0.36 (0.17 - 0.77)	$8.33 \cdot 10^{-3}$	$8.33 \cdot 10^{-3}$	0.635	0.780
LDL - Cholesterol	0.29 (0.13 - 0.63)	$1.91 \cdot 10^{-3}$	$3.9 \cdot 10^{-3}$	0.667	0.740
Apo - B100	0.30 (0.14 - 0.64)	$2.03 \cdot 10^{-3}$	$3.9 \cdot 10^{-3}$	0.654	0.747
LDL – HDL - Cholesterol	0.34 (0.16 - 0.71)	$4.00 \cdot 10^{-3}$	$5 \cdot 10^{-3}$	0.654	0.577
Apo - B100 / Apo - A1	0.36 (0.17 - 0.76)	$7.16 \cdot 10^{-3}$	$7.46 \cdot 10^{-3}$	0.645	0.609

Total Particle Number ApoB	0.30 (0.14 - 0.64)	$2.03 \cdot 10^{-3}$	$3.9 \cdot 10^{-3}$	0.654	0.747
LDL Particle Number	0.26 (0.12 - 0.58)	$9.55 \cdot 10^{-4}$	$3.9 \cdot 10^{-3}$	0.677	0.752
LDL – 4 Particle Number	0.35 (0.16 - 0.71)	$4.75 \cdot 10^{-3}$	$5.39 \cdot 10^{-3}$	0.641	0.691
LDL – 5 Particle Number	0.32 (0.15 - 0.67)	$2.48 \cdot 10^{-3}$	$3.9 \cdot 10^{-3}$	0.655	0.623
Triglycerides - LDL	0.28 (0.13 - 0.62)	$1.63 \cdot 10^{-3}$	$3.9 \cdot 10^{-3}$	0.651	0.721
Cholesterol - LDL	0.29 (0.13 - 0.63)	$1.91 \cdot 10^{-3}$	$3.9 \cdot 10^{-3}$	0.667	0.740
Free Cholesterol - LDL	0.33 (0.15 - 0.72)	$5.11 \cdot 10^{-3}$	$5.55 \cdot 10^{-3}$	0.652	0.745
Phospholipids - LDL	0.28 (0.13 - 0.62)	$1.70 \cdot 10^{-3}$	$3.9 \cdot 10^{-3}$	0.669	0.748
Apo – B - LDL	0.26 (0.12 - 0.58)	$9.55 \cdot 10^{-4}$	$3.9 \cdot 10^{-3}$	0.677	0.752
Triglycerides LDL - 3	0.32 (0.14 - 0.69)	$3.71 \cdot 10^{-3}$	$4.94 \cdot 10^{-3}$	0.636	0.764
Triglycerides LDL - 4	0.31 (0.14 - 0.66)	$2.38 \cdot 10^{-3}$	$3.9 \cdot 10^{-3}$	0.646	0.668
Triglycerides LDL - 5	0.33 (0.16 - 0.69)	$3.54 \cdot 10^{-3}$	$4.94 \cdot 10^{-3}$	0.655	0.633
Cholesterol LDL - 5	0.31 (0.15 - 0.66)	$2.47 \cdot 10^{-3}$	$3.9 \cdot 10^{-3}$	0.653	0.617
Free Cholesterol LDL - 5	0.30 (0.14 - 0.64)	$1.94 \cdot 10^{-3}$	$3.9 \cdot 10^{-3}$	0.656	0.646
Phospholipids LDL - 5	0.31 (0.15 - 0.66)	$2.25 \cdot 10^{-3}$	$3.9 \cdot 10^{-3}$	0.655	0.617
Apo – B LDL - 4	0.33 (0.16 - 0.71)	$4.74 \cdot 10^{-3}$	$5.39 \cdot 10^{-3}$	0.641	0.691
Apo – B LDL - 5	0.32 (0.15 - 0.67)	$2.50 \cdot 10^{-3}$	$3.90 \cdot 10^{-3}$	0.655	0.623

Figures

Figure 1. Study design flowchart illustrating the number participants: *de novo* drug untreated Parkinson's disease patients (dnPD), healthy control subjects (CTR) and advanced Parkinson's disease under dopaminergic treatment (advPD).

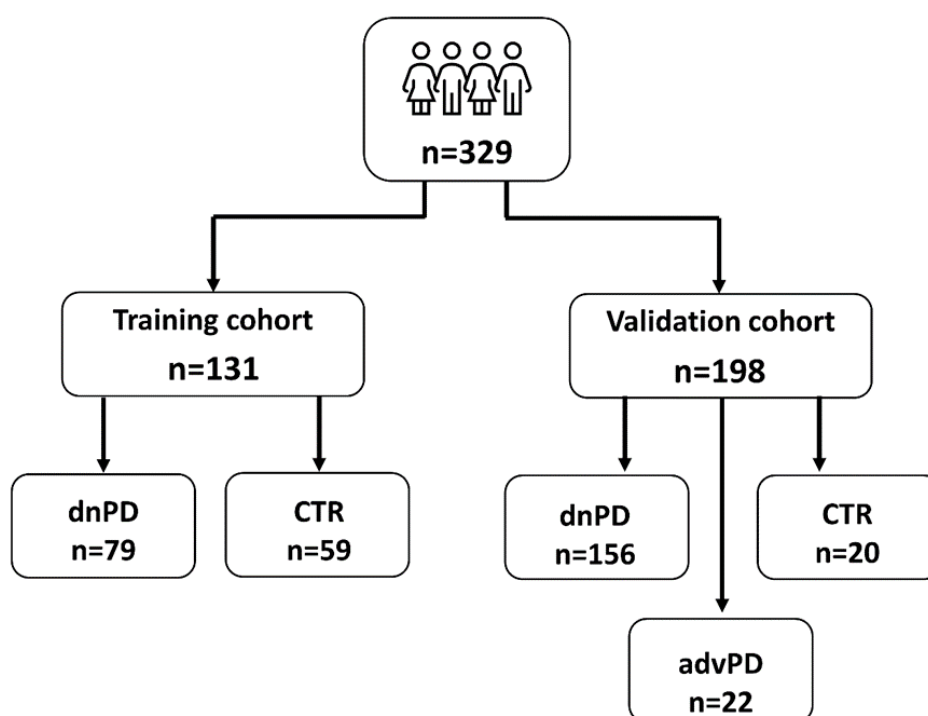


Figure 2. PCA 3D score plot of the whole study population. Each dot represents a 0.02 ppm bucketed 1D-NOESY $^1\text{H-NMR}$ spectrum color-coded by subject groups: green dots, healthy controls ($n=79$); blue dots, *de novo* Parkinson's disease patients ($n=228$) and red dots, advanced PD patients pharmacologically treated ($n=22$).

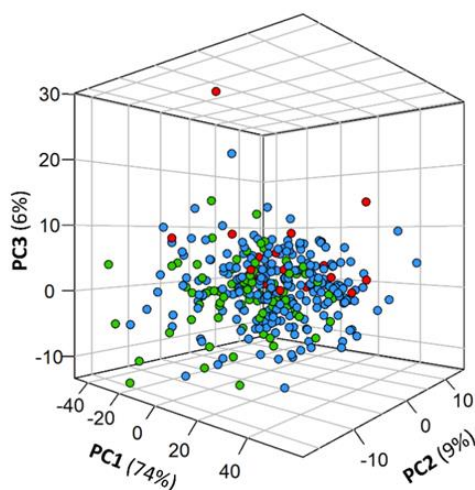


Figure 3. Bar-plot of \log_2 fold changes values (Log_2FC). Statistically significant variables ($\text{FDR} < 0.05$) quantified in serum spectra of training cohort are reported. Negative Log_2FC values mean higher concentrations in CTR subjects, while positive Log_2FC refers to higher concentration levels in dnPD. Effect size (Cliff's delta) is also shown for each variable.

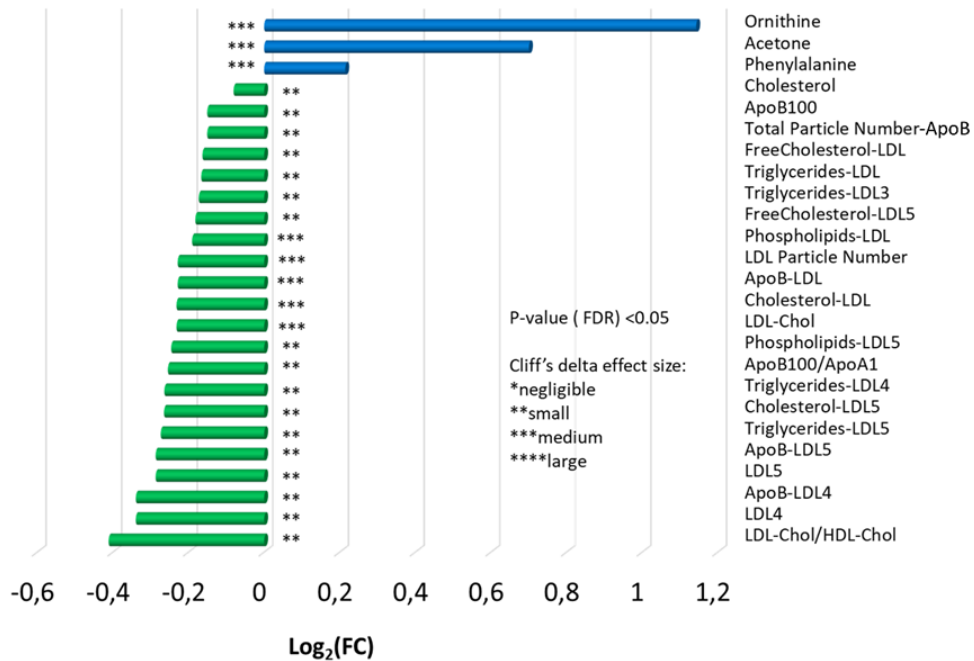
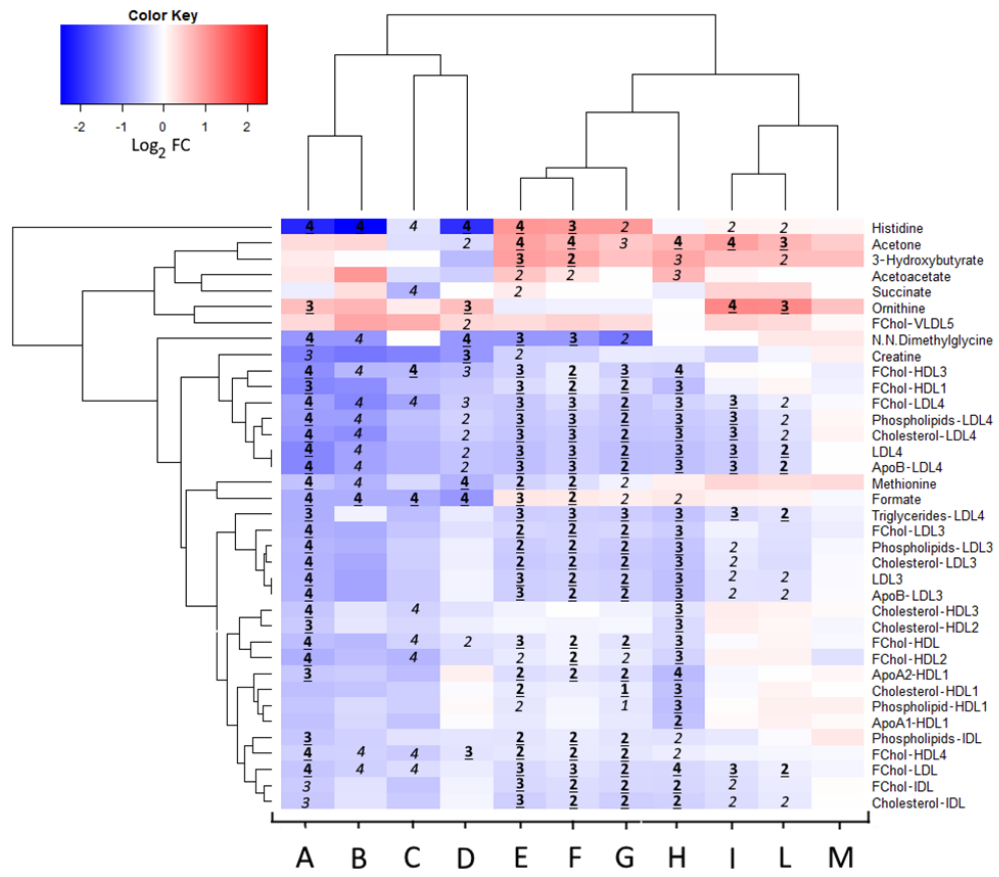


Figure 4. Two-way hierarchical clustering heatmap of significant metabolites. To build the heatmap, only the metabolites that showed significant P -values in more than one comparison were selected. Subsequently, only those metabolites that showed $\log_2\text{FC} \geq |0.5|$ values in at least one comparison were retained. Euclidian distance and complete linkage were used to generate the dendrogram. Color key describes the $\log_2\text{FC}$. Positive $\log_2\text{FC}$ values mean higher metabolite level in case group (dnPD or advPD, depending on the performed analysis); negative $\log_2\text{FC}$ values mean lower metabolite level in case group. Italic numbers refer to P -value < 0.05 , underlined and bolded numbers refer to $\text{FDR} < 0.05$. Numbers from 1 to 4 represent the Cliff's delta effect size, defined as: 1, negligible; 2, small; 3, medium; 4, large effect size. Each column coded with letters from "A" to "M" represents the comparison of metabolites' levels of different group of subjects. A, B, C, D and H are obtained using serum samples of subjects of the validation cohort, where A refers to the comparison of CTR vs. advPD; B, male CTR vs. male advPD; C, female CTR vs. female advPD; D, dnPD vs. advPD and H, CTR vs. dnPD. I, L and M are obtained considering samples of the training cohort: I, male CTR vs. male dnPD; L, CTR vs. dnPD; M, female CTR vs. dnPD. E, F and G are obtained combining subjects of the training and validation cohorts; E, male CTR and male dnPD; F, CTR vs. dnPD and G, female CTR vs. female dnPD.



Supplementary Material**Table S1.** Complete list of log₂(FoldChange) ratios of univariate group comparisons.

	I CTR vs dnPD	II CTR vs dnPD	I M CTR vs dnPD	I W CTR vs dnPD	II dnPD vs advPD	II CTR vs advPD	II M CTR vs advPD	II W CTR vs advPD	I,II CTR vs dnPD	I,II M CTR vs dnPD	I,II W CTR vs dnPD
Acetate	-0.14	0.16	-0.13	-0.18	0.06	0.23	0.14	0.44	0.30	0.32	0.16
Acetoacetate	0.00	0.70	0.07	0.00	-0.45	0.25	1.00	-0.32	0.26	0.56	0.00
Acetone	0.70	0.70	0.91	0.49	-0.36	0.34	0.40	-0.32	0.72	0.89	0.54
Alanine	0.02	-0.10	-0.03	0.18	-0.02	-0.13	-0.10	-0.13	0.03	-0.03	0.10
ApoB10/ApoA1	-0.25	-0.18	-0.28	-0.08	0.07	-0.10	-0.01	-0.23	-0.23	-0.19	-0.19
Citrate	0.08	0.06	0.19	-0.03	0.00	0.00	0.00	0.00	-0.14	-0.08	-0.21
Creatine	-0.09	-0.22	-0.42	0.13	-1.00	-1.22	-1.32	-1.21	-0.42	-0.44	-0.20
Creatinine	0.03	-0.04	-0.01	-0.03	-0.05	-0.10	-0.21	0.00	0.00	-0.02	-0.02
Dimethylsulfone	0.00	-0.15	-0.06	-0.19	0.36	0.22	0.16	0.00	0.00	0.00	0.12
Formate	0.11	0.22	0.13	-0.06	-1.00	-0.78	-0.78	-0.78	0.21	0.23	0.16
Glucose	-0.05	-0.03	-0.06	-0.01	-0.02	-0.05	0.04	-0.19	0.01	0.06	-0.01
Glutamine	0.01	0.00	0.02	-0.01	-0.10	-0.10	-0.05	-0.19	-0.01	-0.02	-0.01
Glycine	0.15	0.10	0.13	0.00	0.25	0.35	0.41	0.29	0.18	0.11	0.14
ApoA1-HDL.1	0.13	-0.58	0.05	0.05	-0.01	-0.60	-0.38	-0.57	-0.08	-0.21	-0.21
ApoA1-HDL2	0.10	-0.33	0.06	-0.02	-0.09	-0.42	-0.22	-0.41	-0.04	-0.09	-0.09
ApoA1-HDL3	0.05	-0.21	0.15	-0.05	-0.22	-0.44	-0.28	-0.41	-0.04	-0.08	-0.08
ApoA1-HDL4	-0.03	-0.04	-0.04	-0.01	-0.15	-0.19	-0.21	-0.26	-0.04	-0.03	-0.03
ApoA2-HDL1	0.00	-0.67	-0.07	0.07	0.14	-0.53	-0.48	-0.64	-0.26	-0.33	-0.33
ApoA2-HDL2	-0.04	-0.39	0.00	-0.17	-0.05	-0.44	-0.33	-0.35	-0.17	-0.18	-0.18

ApoA2-HDL3	0.00	-0.26	0.01	-0.10	-0.16	-0.41	-0.19	-0.41	-0.16	-0.16	-0.16
ApoA2-HDL4	-0.06	-0.04	0.00	-0.09	-0.12	-0.15	-0.06	-0.17	-0.04	0.04	0.04
Chol-HDL1	0.11	-0.59	-0.05	0.00	-0.03	-0.61	-0.57	-0.45	-0.16	-0.21	-0.21
Chol-HDL2	0.08	-0.35	0.15	-0.08	-0.16	-0.51	-0.24	-0.38	-0.10	-0.07	-0.07
Chol-HDL3	0.11	-0.28	0.17	0.04	-0.26	-0.54	-0.26	-0.43	0.00	-0.12	-0.12
Chol-HDL4	-0.05	-0.03	-0.02	-0.05	-0.17	-0.20	-0.20	-0.31	-0.03	-0.03	-0.03
FreeChol-HDL1	0.08	-0.64	-0.12	-0.13	-0.53	-1.16	-1.10	-0.62	-0.22	-0.35	-0.35
FreeChol-HDL2	0.12	-0.42	0.12	-0.29	-0.35	-0.77	-0.64	-0.71	-0.11	-0.20	-0.20
FreeChol-HDL3	0.00	-0.47	0.03	-0.16	-0.62	-1.09	-0.69	-0.68	-0.20	-0.44	-0.44
FreeChol-HDL4	-0.08	-0.17	-0.10	-0.06	-0.26	-0.44	-0.43	-0.49	-0.20	-0.24	-0.24
Phosp-HDL1	0.14	-0.62	0.02	0.14	0.06	-0.56	-0.36	-0.47	-0.09	-0.24	-0.24
Phosp-HDL2	0.13	-0.36	0.14	0.01	-0.07	-0.42	-0.16	-0.36	-0.07	-0.07	-0.07
Phosp-HDL3	0.03	-0.24	0.13	-0.08	-0.19	-0.43	-0.23	-0.47	-0.03	-0.09	-0.09
Phosp-HDL4	-0.05	-0.05	-0.01	-0.08	-0.09	-0.14	-0.15	-0.25	-0.05	-0.06	-0.06
Trigl-HDL1	0.02	-0.22	0.31	-0.13	-0.08	-0.31	-0.01	-0.48	-0.11	-0.09	-0.09
Trigl-HDL2	0.06	-0.12	0.07	0.17	-0.17	-0.30	0.11	-0.34	-0.13	-0.01	-0.01
Trigl-HDL3	-0.02	0.07	0.02	-0.17	-0.20	-0.13	0.13	-0.18	-0.10	-0.06	-0.06
Trigl-HDL4	-0.05	0.22	0.00	0.01	-0.20	0.02	0.15	0.08	0.02	-0.05	-0.05
Histidine	0.09	-0.07	0.10	0.07	-2.03	-2.10	-2.47	-0.30	0.97	1.00	0.97
IDL	-0.19	-0.37	-0.20	-0.18	-0.09	-0.45	-0.07	-0.44	-0.30	-0.35	-0.35
Isoleucine	0.04	0.19	0.02	0.00	-0.18	0.02	-0.07	-0.54	-0.03	-0.08	-0.12
Lactate	0.24	-0.37	0.25	0.26	0.17	-0.20	-0.14	-0.05	-0.03	0.06	-0.21
LDL1	-0.01	-0.41	0.03	-0.11	0.20	-0.21	0.18	-0.26	-0.19	-0.17	-0.17
LDL2	-0.25	-0.43	-0.22	-0.20	0.12	-0.31	-0.14	-0.40	-0.18	-0.06	-0.06
LDL3	-0.29	-0.60	-0.28	-0.04	-0.11	-0.71	-0.87	-0.50	-0.43	-0.46	-0.46

LDL4	-0.34	-0.60	-0.49	0.00	-0.57	-1.17	-0.89	-0.73	-0.50	-0.62	-0.62
LDL5	-0.29	-0.20	-0.37	-0.04	-0.22	-0.42	-0.37	-0.22	-0.28	-0.31	-0.31
LDL6	-0.13	-0.13	-0.18	-0.03	-0.10	-0.23	-0.29	-0.19	-0.14	-0.15	-0.15
LDLChol/HDL Chol	-0.41	-0.07	-0.35	-0.01	-0.07	-0.15	-0.14	-0.15	-0.32	-0.24	-0.24
LDL	-0.23	-0.34	-0.28	-0.05	0.00	-0.33	-0.28	-0.26	-0.25	-0.32	-0.32
ApoB-LDL1	-0.01	-0.41	0.03	-0.11	0.20	-0.21	0.18	-0.26	-0.19	-0.17	-0.17
ApoB-LDL2	-0.25	-0.43	-0.22	-0.20	0.12	-0.31	-0.14	-0.40	-0.18	-0.06	-0.06
ApoB-LDL3	-0.29	-0.60	-0.28	-0.04	-0.11	-0.71	-0.87	-0.50	-0.43	-0.47	-0.47
ApoB-LDL4	-0.34	-0.60	-0.49	0.00	-0.57	-1.17	-0.89	-0.73	-0.50	-0.62	-0.62
ApoB-LDL5	-0.29	-0.20	-0.37	-0.04	-0.22	-0.42	-0.37	-0.22	-0.28	-0.31	-0.31
ApoB-LDL6	-0.13	-0.13	-0.18	-0.03	-0.10	-0.23	-0.29	-0.19	-0.14	-0.15	-0.15
Chol-LDL1	-0.01	-0.28	-0.20	-0.06	0.16	-0.12	0.10	-0.37	-0.13	-0.25	-0.25
Chol-LDL2	-0.28	-0.49	-0.22	-0.13	0.19	-0.30	-0.04	-0.40	-0.19	-0.04	-0.04
Chol-LDL3	-0.34	-0.57	-0.23	-0.05	-0.16	-0.74	-0.83	-0.47	-0.47	-0.52	-0.52
Chol-LDL4	-0.30	-0.54	-0.47	0.11	-0.46	-1.00	-1.09	-0.65	-0.50	-0.55	-0.55
Chol-LDL5	-0.26	-0.16	-0.36	-0.02	-0.22	-0.39	-0.41	-0.08	-0.30	-0.29	-0.29
Chol-LDL6	-0.13	-0.13	-0.18	0.08	-0.09	-0.22	-0.30	-0.17	-0.13	-0.14	-0.14
FreeChol-LDL1	-0.04	-0.29	-0.06	-0.21	0.03	-0.26	0.03	-0.48	-0.16	-0.30	-0.30
FreeChol-LDL2	-0.13	-0.47	0.03	-0.19	0.00	-0.47	-0.16	-0.43	-0.18	-0.02	-0.02
FreeChol-LDL3	-0.29	-0.48	-0.12	-0.17	-0.29	-0.77	-0.80	-0.55	-0.41	-0.40	-0.40
FreeChol-LDL4	-0.17	-0.41	-0.32	-0.05	-0.48	-0.90	-1.16	-0.86	-0.36	-0.53	-0.53
FreeChol-LDL5	-0.18	-0.22	-0.31	-0.05	-0.24	-0.46	-0.45	-0.12	-0.27	-0.33	-0.33
FreeChol-LDL6	-0.07	-0.20	-0.15	0.02	-0.12	-0.32	-0.36	-0.11	-0.13	-0.16	-0.16
Phosps-LDL1	0.04	-0.24	-0.05	-0.11	0.14	-0.10	0.23	-0.31	-0.12	-0.18	-0.18

Phosps-LDL2	-0.29	-0.46	-0.17	-0.14	0.17	-0.29	-0.06	-0.37	-0.18	0.01	0.01
Phosps-LDL3	-0.30	-0.55	-0.21	-0.07	-0.15	-0.71	-0.77	-0.45	-0.42	-0.48	-0.48
Phosps-LDL4	-0.28	-0.50	-0.43	0.07	-0.43	-0.93	-0.93	-0.59	-0.43	-0.52	-0.52
Phosps-LDL5	-0.25	-0.17	-0.34	-0.04	-0.21	-0.37	-0.38	-0.10	-0.29	-0.28	-0.28
Phosps-LDL6	-0.12	-0.17	-0.15	0.02	-0.07	-0.24	-0.32	-0.12	-0.12	-0.13	-0.13
Trigl-LDL1	-0.02	-0.43	-0.19	-0.03	-0.06	-0.49	-0.11	-0.42	-0.23	-0.25	-0.25
Trigl-LDL2	-0.13	-0.47	-0.07	-0.08	0.00	-0.46	-0.33	-0.38	-0.25	-0.33	-0.33
Trigl-LDL3	-0.17	-0.40	-0.34	-0.02	0.07	-0.33	0.00	-0.31	-0.21	-0.40	-0.40
Trigl-LDL4	-0.26	-0.58	-0.36	-0.12	-0.21	-0.79	-0.13	-0.63	-0.44	-0.48	-0.48
Trigl-LDL5	-0.27	-0.11	-0.38	-0.08	-0.09	-0.20	-0.14	-0.25	-0.36	-0.39	-0.39
Trigl-LDL6	-0.11	-0.18	-0.15	-0.10	0.02	-0.15	-0.35	-0.12	-0.11	-0.17	-0.17
Leucine	0.01	0.21	-0.06	0.13	-0.07	0.14	0.25	-0.10	0.10	0.09	0.18
ApoA1-HDL	0.05	-0.20	0.04	0.00	-0.21	-0.40	-0.13	-0.42	-0.04	-0.08	-0.08
ApoA2-HDL	0.00	-0.18	0.01	-0.09	-0.08	-0.26	-0.05	-0.32	-0.09	-0.07	-0.07
ApoB-IDL	-0.19	-0.37	-0.20	-0.18	-0.09	-0.45	-0.07	-0.44	-0.30	-0.35	-0.35
ApoB-LDL	-0.23	-0.34	-0.28	-0.05	0.00	-0.33	-0.28	-0.26	-0.25	-0.32	-0.32
ApoB-VLDL	0.04	0.26	0.10	0.19	-0.15	0.10	0.03	0.19	-0.05	-0.03	-0.03
Chol-HDL	0.05	-0.21	0.03	0.04	-0.15	-0.37	-0.29	-0.35	-0.01	-0.08	-0.08
Chol-IDL	-0.22	-0.43	-0.34	0.00	-0.09	-0.52	-0.27	-0.50	-0.35	-0.47	-0.47
Chol-LDL	-0.23	-0.38	-0.26	-0.08	-0.10	-0.48	-0.38	-0.27	-0.26	-0.34	-0.34
Chol-VLDL	0.02	0.38	0.08	0.42	-0.18	0.20	0.07	0.44	-0.02	-0.05	-0.05
FreeChol-HDL	0.09	-0.35	-0.04	-0.06	-0.29	-0.64	-0.68	-0.43	-0.11	-0.26	-0.26
FreeChol-IDL	-0.19	-0.39	-0.32	0.02	-0.11	-0.50	-0.28	-0.56	-0.34	-0.37	-0.37
FreeChol-LDL	-0.16	-0.38	-0.30	-0.09	-0.18	-0.56	-0.44	-0.34	-0.27	-0.37	-0.37
FreeChol-VLDL	0.05	0.27	0.10	0.50	-0.28	-0.01	-0.18	0.34	-0.02	-0.05	-0.05

Phosps-HDL	0.07	-0.23	0.11	-0.03	-0.16	-0.39	-0.21	-0.39	-0.04	-0.10	-0.10
Phosps-IDL	-0.04	-0.28	-0.22	0.22	-0.25	-0.53	-0.43	-0.28	-0.19	-0.26	-0.26
Phosps-LDL	-0.19	-0.38	-0.25	-0.08	-0.08	-0.46	-0.36	-0.26	-0.22	-0.30	-0.30
Phosps-VLDL	-0.01	0.31	0.05	0.67	-0.29	0.02	-0.29	0.30	-0.03	-0.06	-0.06
Trigl-HDL	0.06	-0.15	0.03	0.06	-0.12	-0.27	0.13	-0.28	-0.06	-0.18	-0.18
Trigl-IDL	-0.09	-0.01	-0.24	0.41	-0.79	-0.80	-1.30	0.10	-0.03	-0.24	-0.24
Trigl-LDL	-0.17	-0.25	-0.20	-0.08	-0.10	-0.36	-0.17	-0.36	-0.24	-0.27	-0.27
Trigl-VLDL	-0.01	0.18	-0.03	0.76	-0.50	-0.33	-0.72	0.05	-0.13	-0.03	-0.03
ApoA1	0.03	-0.19	0.02	0.00	-0.21	-0.40	-0.15	-0.40	-0.05	-0.12	-0.12
ApoA2	0.00	-0.19	0.02	-0.10	-0.09	-0.27	-0.05	-0.32	-0.10	-0.07	-0.07
ApoB100	-0.15	-0.28	-0.19	-0.05	-0.09	-0.37	-0.27	-0.23	-0.22	-0.21	-0.21
Chol	-0.08	-0.27	-0.13	-0.08	-0.09	-0.36	-0.34	-0.22	-0.18	-0.23	-0.23
HDLChol	0.05	-0.21	0.03	0.04	-0.15	-0.37	-0.29	-0.35	-0.01	-0.08	-0.08
LDLChol	-0.23	-0.38	-0.26	-0.08	-0.10	-0.48	-0.38	-0.27	-0.26	-0.34	-0.34
Trigl	-0.03	0.06	-0.11	0.26	-0.32	-0.26	-0.25	0.01	-0.06	-0.09	-0.09
Methionine	0.32	0.15	0.41	0.36	-0.74	-0.58	-0.74	-0.36	-0.32	-0.43	-0.13
N,N-Dimethylglycine	0.22	0.00	0.00	0.22	-1.00	-1.00	-1.00	0.00	-1.00	-1.00	-1.32
Ornithine	1.14	0.00	1.15	0.58	0.63	0.63	0.71	0.19	-0.13	-0.17	-0.13
Phenylalanine	0.21	0.06	0.23	0.17	-0.01	0.05	0.08	-0.05	0.19	0.23	0.10
Pyruvic-acid	-0.01	-0.11	-0.10	0.18	0.16	0.06	0.25	-0.08	-0.01	0.08	-0.07
Sarcosine	0.00	0.13	0.00	0.17	0.00	0.13	0.22	-0.74	0.00	0.00	0.32
Succinic-acid	0.42	-0.17	0.42	0.00	0.00	-0.17	0.32	-0.74	0.00	0.19	0.00
Total	-0.15	-0.28	-0.19	-0.05	-0.09	-0.37	-0.27	-0.23	-0.22	-0.21	-0.21

Trimethylamine -N-oxide	0.00	-0.12	0.37	-0.46	-0.58	-0.70	-0.81	-0.66	0.00	0.25	-0.25
Tyrosine	0.05	-0.13	0.01	0.18	-0.11	-0.24	-0.23	-0.19	0.09	0.04	0.05
Valine	0.03	0.12	0.05	0.17	-0.01	0.10	0.12	-0.24	0.08	0.15	0.02
VLDL	0.04	0.26	0.10	0.19	-0.15	0.10	0.03	0.19	-0.04	-0.03	-0.03
Chol-VLDL1	0.00	0.14	-0.17	0.82	-0.16	-0.02	0.09	0.26	-0.04	-0.16	-0.16
Chol-VLDL2	0.04	0.01	-0.04	0.48	-0.22	-0.21	-0.52	0.34	-0.06	-0.02	-0.02
Chol-VLDL3	0.09	0.32	0.18	0.50	-0.19	0.13	-0.37	0.43	0.07	0.01	0.01
Chol-VLDL4	0.05	0.12	0.13	0.23	0.20	0.32	0.11	0.43	0.00	-0.11	-0.11
Chol-VLDL5	0.18	0.08	0.13	0.27	0.09	0.17	0.43	0.23	0.19	0.13	0.13
FreeChol- VLDL1	-0.29	-0.10	-0.29	0.69	-0.37	-0.47	-0.31	-0.06	-0.21	-0.23	-0.23
FreeChol- VLDL2	0.05	0.24	-0.07	0.40	-0.46	-0.22	-0.58	0.35	-0.01	0.05	0.05
FreeChol- VLDL3	0.14	0.31	0.19	0.73	-0.38	-0.07	-0.10	0.45	0.12	0.02	0.02
FreeChol- VLDL4	0.22	0.28	0.15	0.38	-0.09	0.19	0.13	0.47	0.01	-0.03	-0.03
FreeChol- VLDL5	0.35	-0.01	0.42	0.06	0.38	0.37	0.86	0.79	0.44	0.34	0.34
Phosps-VLDL1	-0.17	0.08	-0.30	0.61	-0.43	-0.35	-0.22	-0.02	-0.03	-0.18	-0.18
Phosps-VLDL2	-0.04	0.15	-0.07	0.35	-0.52	-0.37	-0.65	0.24	0.09	-0.04	-0.04
Phosps-VLDL3	0.04	0.28	0.17	0.39	-0.29	0.00	-0.27	0.42	0.01	0.00	0.00
Phosps-VLDL4	0.07	0.18	0.21	0.12	0.06	0.24	0.17	0.27	0.04	0.01	0.01
Phosps-VLDL5	0.15	0.26	0.15	0.21	0.00	0.26	0.34	0.23	0.11	0.04	0.04

Trigl-VLDL1	-0.18	0.16	-0.30	0.69	-0.65	-0.50	-0.71	-0.10	-0.17	-0.16	-0.16
Trigl-VLDL2	-0.04	0.31	0.02	0.43	-0.83	-0.52	-1.34	0.24	-0.09	-0.06	-0.06
Trigl-VLDL3	0.13	0.23	0.20	0.38	-0.52	-0.29	-0.65	0.38	0.01	0.07	0.07
Trigl-VLDL4	0.18	0.12	0.15	0.24	-0.01	0.11	0.08	0.17	0.04	0.00	0.00
Trigl-VLDL5	0.18	-0.02	0.17	0.28	-0.03	-0.05	0.06	0.13	0.03	-0.05	-0.05
3- Hydroxybutyrate	0.62	0.85	0.62	0.62	-0.66	0.19	-0.01	0.00	0.85	0.92	0.60

*Abbreviations are reported as follows: Trigl: triglycerides; Chol: cholesterol; Phosp: phospholipids; M: males; W: women; I: training cohort; II: validation cohort.

Figure S1. OPLS-DA score plots of bucketed 1D-NOESY ^1H NMR serum spectra. **A)** OPLS-DA score plot of overall training cohort; **B)** OPLS-DA score plot of male training subjects; **C)** OPLS-DA score plot of female training subjects. Colours code for individual status: CTR (green dots) and dnPD patients (blue dots). Ellipses indicate the confidence interval (P -value < 0.05) for each group.

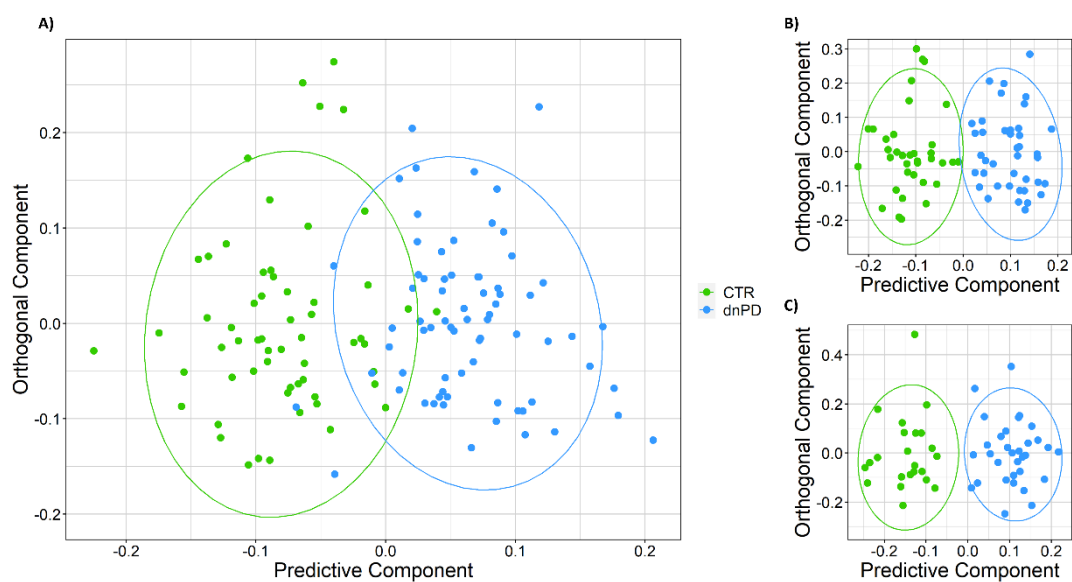
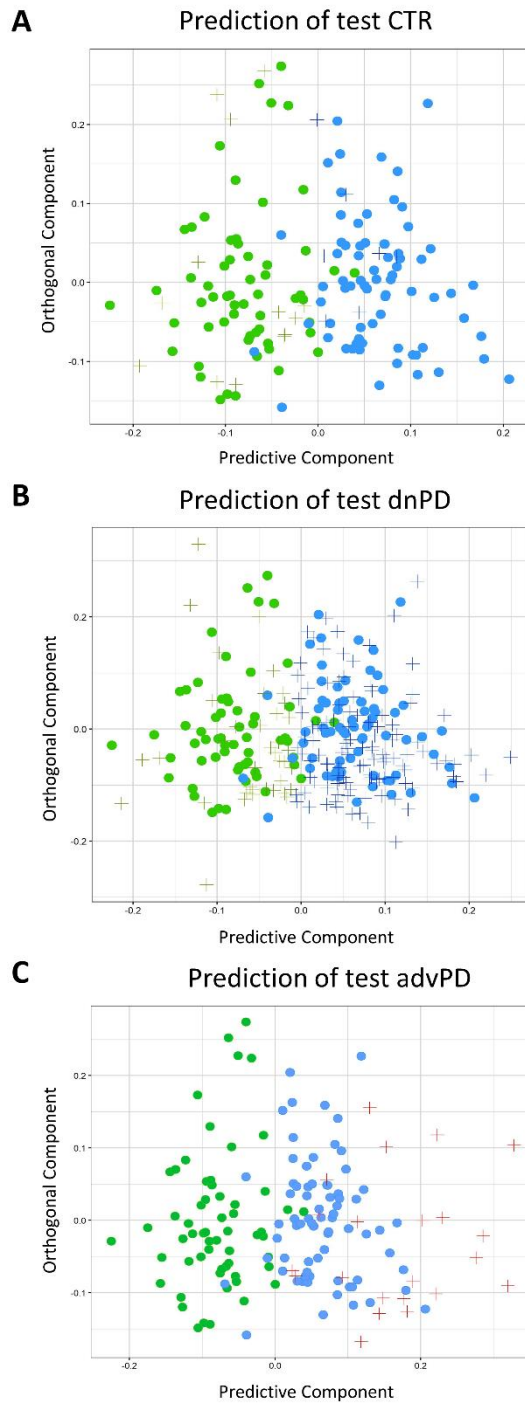


Figure S2. OPLS-DA score plot of training model discriminating CTR (green dots) from dnPD patients (blue dots). “+” light green symbol: validation samples of true healthy controls; “+” dark blue symbol: validation samples of true dnPD; “+” red symbol: validation samples of advPD, all predicted as true PD patients.



4.1.2. NMR-based metabolomics to predict three-month adverse outcomes and to estimate metabolic variations in ischemic stroke treated with intravenous thrombolysis

In preparation

Candidate's contributions: acquisition of NMR spectra, statistical analysis of NMR data, interpreting results and writing of the manuscript.

Introduction

Ischemic stroke (IS) is a leading cause of death and disability continuously increasing,¹ and contributing significantly to health costs. There is an urgent need to deep the research of effective biomarkers useful for the clinical practice and to better understand and interpret occurring dysregulation in the pathophysiological mechanisms of the disease.

Metabolic perturbations are believed to be fundamental events that contribute to the ischemic stroke, to its progression and to the development of subsequent unfavourable outcomes.²⁻⁹ Regarding the knowledge about biomarkers associated with poor prognosis in the setting of stroke patients treated with thrombolysis, few evidences are reported (*e.g.* glucose).¹⁰ Dyslipidaemia is a known risk factor contributing to the onset of ischemic stroke; for example, high levels of total cholesterol and low-density lipoprotein (LDL) cholesterol increase the risk for cerebral ischemia.^{11,12} However, the effects of lipid levels on clinical outcomes after ischemic attack are controversial: high total cholesterol and LDL cholesterol levels have been associated with better functional and vital outcomes after stroke,^{13,14} while low LDL cholesterol level increased the risk of early symptomatic intracranial heamorrhage¹⁵ and total cholesterol was related to worse functional outcome in ischemic stroke patients after the thrombolytic treatment.¹⁶ Therefore, the effective contribution of lipid levels to stroke outcomes, particularly after thrombolysis, needs to be further investigated.

In this framework, comprehensive analytical techniques provide a great chance to identify key metabolic features involved in the onset, in the progression of the disease and also in the development of patient's poor outcomes. Nuclear Magnetic Resonance (NMR)-based metabolomics can provide crucial information. Indeed, it allows a high-throughput analysis of various types of samples (*e.g.* blood, urine, cells and tissues), giving information on hundreds of different metabolites and lipid features present in biological matrices.^{17,18} In particular, multivariate and univariate analyses proved to be efficient in characterizing the metabolic signature of several pathologies,¹⁹⁻²² and in the context of molecular epidemiology.^{23,24}

Here, using a subset of the ischemic stroke patients, treated with intravenous thrombolysis and enrolled in the MAGIC study^{25,26}, we mainly aimed at providing metabolic insights underlying susceptibility to three-month post-acute ischemic stroke mortality, impairments and symptomatic intracerebral haemorrhage, developed after the intravenous (*i.v.*) thrombolytic treatment with recombinant tissue plasminogen activator (rt-PA). Moreover, we retrospectively explored metabolic features associated with the non-response to the thrombolysis. To all these purposes, we applied firstly, logistic regression and Receiver Operating Characteristics (ROC) curve analysis considering: metabolites and lipids quantified in *i*) 1D NOESY NMR spectra from 243 serum samples collected before (t_1) and 24h after the rt-PA therapy (t_2); and *ii*) pre - 24h post metabolite and lipid variations ($\Delta(\text{pre-post})_{\text{rt-PA}}$).

Secondly, with the aim of investigating, in a larger time scale, metabolic variations with respect to the thrombolytic therapy, within-subject metabolic changes were

explored considering serum samples at t_1 , t_2 and samples collected at the outcome evaluation (at three-months (t_3)). An overview of the study design is reported in **Figure 1**.

Material and Methods

Study Population and Outcomes

The study population consists of a total of 243 patients who had an acute ischemic stroke (AIS) and were admitted for thrombolysis treatment with rt-PA in 14 different Italian centres, registered in the Safe Implementation of thrombolysis in Stroke-International Stroke Thrombolysis Register (SITS-ISTR, www.sitsinternational.org), according to SITS-Monitoring Study criteria,²⁷ in the frame of the national, observational and multicentric MAGIC study.^{25,26}

The whole study focuses on the analysis of serum samples collected at three different time-points: before (t_1), 24h after (t_2) and 3 months after (t_3) the administration of rt-PA. Poor outcomes were defined at evaluation (*i.e.* three months after AIS) as follows: *i*) mortality; *ii*) disability or impairment; *iii*) development of symptomatic intracerebral haemorrhage (sICHRICT12); and *iv*) non-response to thrombolytic treatment. Impairment was defined according to the modified Rankin disability score and dichotomized into good (modified Rankin scale, 0–2) or poor (modified Rankin scale, 3–6) outcome. A detailed description of clinical and demographic characteristics of the patients are reported previously in Gori *et al.*²⁶

Ethical Issues

The study protocol was approved by the local Ethical Committee of the Careggi University Hospital (Florence) and it complies with the Declaration of Helsinki. All patients gave informed consent.

NMR serum sample collection and preparation

Whole venous blood was collected in tubes without anticoagulant, before and 24 h after thrombolysis. Tubes were centrifuged at room temperature at 1500 g for 15 min, and the supernatants were stored in aliquots at -80°C until NMR measurements. For metabolomic analyses, serum samples were prepared following the details reported elsewhere.¹⁷

NMR experiments

Serum samples were analysed using a Bruker 600 MHz spectrometer working at 600.13 MHz proton Larmor frequency equipped with a 5 mm PATXI 1H-13C-15N and 2H decoupling probe. This includes a z axis gradient coil, an automatic tuning-matching (ATM) and an automatic and refrigerate sample changer (SampleJet). To stabilize approximately, at the level of ± 0.1 K, the sample temperature (310 K), a BTO 2000 thermocouple was employed and each NMR tube was kept for at least 5 min inside the NMR probe head to equilibrate the acquisition temperature of 310 K.

For each serum specimen, the 1D NOESY, 1D CPMG and 1D DIFFUSION-EDITED pulse sequences were applied to acquire ^1H -NMR spectra. Raw NMR data were multiplied by an exponential function of 0.3 Hz line-broadening factor, before the application of Fourier transform. Phase and baseline distortions were automatically corrected and transformed spectra were calibrated to the glucose doublet at 5.24 ppm using TopSpin 3.2 (BrukerBioSpin).

Metabolite and Lipid identification and quantification

18 metabolites and 112 lipid fractions were unambiguously identified and estimated from ^1H 1D NOESY NMR spectra according to Bruker's B.I.-LISA protocols, version 1.0 (Bruker IVDr Lipoprotein subclass analysis).²⁸ In all serum 1D NOESY NMR spectra, VLDL, LDL, IDL, HDL and 15 different subclasses, (VLDL-1 to VLDL-5, LDL-1 to LDL-6 and HDL-1 to HDL-4) were quantified. For each main class and subclass, reported data consist in concentrations of lipids (total cholesterol, free cholesterol, phospholipids and triglycerides) contained in each fraction. Concentrations of apolipoprotein Apo-A1 and ApoA2 were estimated for HDL class and each relative subclass, while Apo-B concentrations are calculated for VLDL, IDL classes and all LDL subclasses.

Statistical Analysis

Software

All statistical analyses and graphical illustrations were generated using R (version 3.5.3), an open source software for statistical management of data.²⁹

Logistic regression analyses

As main explanatory variables, we considered baseline (t_1), 24h post (t_2) rt-PA and single patient's relative pre-24h post rt-PA variation ($\Delta(\text{pre-post})_{\text{rt-PA}}$) of metabolites and lipids concentrations.

Pre-post thrombolysis variation was calculated using the formula: $\{[(t_2 \text{ metabolite or lipid concentrations}) - (t_1 \text{ metabolite or lipid concentrations})] / [(t_1 \text{ metabolite or lipid concentrations} + t_2 \text{ metabolite or lipid concentrations}) / 2]\}$.

The net effect of each variable's at t_1 , t_2 or considering $\Delta(\text{pre-post})_{\text{rt-PA}}$, on the four poor outcomes, was estimated by logistic regression analysis, including as covariates different patients' characteristics, *i.e.* age, sex, baseline or 24h post-blood glucose (for t_1 and t_2 models, respectively), baseline NIHSS, time onset-to treatment, blood collection center and risk factors and comorbidities (*i.e.* history of atrial fibrillation, congestive heart failure, recent infections or inflammations, hypertension, diabetes, hyperlipidaemia and smoke).

Odd Ratios (OR) values and 95% Confidence Interval (95% CI) were reported for each analysed metabolite or lipid. Benjamini- Hochberg method³⁰ was used to correct for multiple testing. *P*-values and FDR values < 0.05 were deemed both statistically significant, depending on the specific condition.

Receiver Operating Characteristic curves (ROC) analysis

Three-months adverse outcomes were evaluated also applying ROC curve analysis (“roc” function of the R package “pROC”) on selected analytes at t_1 , t_2 and considering specific patient’s relative metabolites and lipids $\Delta(\text{pre-post})_{\text{rt-PA}}$.

In detail, for each analysis, we estimated values of the area under the ROC curve (AUC-ROC) for two different logistic regression models. Firstly, we calculated AUC values for “baseline” models that included only clinical and risk factors known to affect the outcomes. Secondly, to quantify how much the addition of a combination of new metabolomic features (metabolites and/or lipids) correctly increases the risk predicted by the baseline ROC curve for events and non-events; we added to each baseline logistic regression model, a combination among the top three statistical significant metabolites and/or lipids (P -values < 0.05), choosing on the basis of the results of the previous logistic regression analyses.

It is important to point out that, in order to avoid overfitting, before performing any ROC analysis, all built logistic regression models were cross-validated using the leave-one-out scheme (R script in-house developed).

Univariate Analysis

Pairwise metabolic changes in AIS patients, occurring with respect to the thrombolytic therapy, were evaluated over the all three time-points, *i.e.*, t_1 , t_2 and t_3 , applying the Friedman test³¹ followed by post-hoc Nemenyi analysis. It was necessary to use a subset ($n=173$) of the study population included in this study, since not all AIS patients furnished serum samples at t_3 .

Benjamini- Hochberg method³⁰ was used to correct for multiple testing and FDR values < 0.01 were considered statistically significant.

Results

Logistic regression analysis

Metabolites and lipids concentrations were used to build logistic regression models for the evaluation of their net effect on three-months poor outcomes, *i.e.* mortality, impairments, haemorrhagic transformation and non-response to the thrombolytic intervention. All the calculated effects of metabolites and lipids concentrations at t_1 are listed in the supplementary **Table S1**. In **Table 1**, we reported only metabolic variables statistically significant associated to the outcomes. Among them, triglycerides related to the HDL-3 subfraction is the only analyte presenting a statistically significant association (P -value < 0.05) with three-months mortality. Seventeen different lipids, mainly related to small dense VLDL, LDL and HDL particles, resulted to be statistically associated (P -value < 0.05) with the appearance of three-months impairments. Before starting the thrombolytic therapy, acetone and 3-hydroxybutyrate resulted to be the only metabolic variables associated both to the

development of symptomatic intracranial haemorrhage and to the non-response to the intravenous thrombolysis.

Considering analytes concentrations estimated for t_2 serum samples (**Table 1**), we reported many metabolic features statistically associated to each poor outcome. More specifically, HDL triglycerides, LDL-6 related cholesterol, triglycerides and phospholipids, HDL-3 and HDL-4 triglycerides are statistically associated (P -value < 0.05) to three-months mortality; while acetate, 3-hydroxybutyrate and twenty-four different lipids resulted to be statistically related with the development of impairments. Among the twenty-four lipids: triglycerides, total cholesterol and free cholesterol estimated for LDL-6 subfractions appear as the most statistically significant associated variables. Nineteen lipids were associated to the development of intracranial haemorrhage, especially lipids related to bigger and less dense VLDL particles. Regarding the evaluation of non-response to the thrombolysis, we estimated statistically significant association (P -value < 0.05) for phenylalanine, VLDL-5 phospholipids, HDL-4 triglycerides, HDL-2 free cholesterol, acetone and 3-hydroxybutyrate. These last two metabolites also reported FDR values < 0.05. The complete list of all calculated net effects at t_2 is reported in supplementary **Table S2**. Finally, all the net effects of $\Delta(\text{pre-post})_{\text{rt-PA}}$, estimated from each quantified metabolite and lipid fraction, on the evaluated outcomes, are reported in supplementary **Table S3**, where many statistically significant (P -value < 0.05) associations between metabolic features and outcomes were highlighted and summarized in **Table 1**. In particular, when in the logistic regression analysis, we related $\Delta(\text{pre-post})_{\text{rt-PA}}$ metabolites and lipids concentrations to the three-months mortality; triglycerides related to LDL, HDL particles and LDL-1, LDL-2 and HDL-4 sub-particles appear as the statistically associated variables (P -values < 0.05). Glutamate, acetate, free cholesterol, phospholipids and Apo A2 related mainly to HDL-1 and HDL-2 resulted to be statistically associated with three-month impairment (P -value < 0.05). If we consider the logistic regression model built to evaluate the development of intracranial haemorrhage, we found that phenylalanine, pyruvate and glucose are the most statistically associated (P -values < 0.05) metabolites, while among lipid parameters, phospholipids related to VLDL particles are the most statistically associated analytes (P -value < 0.05). In conclusion, alanine, acetone, total particle number of Apo B100, related LDL-3 sub-particles and HDL-4 triglycerides appear as the most statistically significant (P -value < 0.05) predictors of non-response to the therapeutic thrombolytic treatment.

ROC curve analysis

The addition of selected statistically significant (P -value < 0.05) metabolites and/or lipid parameters to baseline models (hereunder referred as “Bas1” for pre rt-PA, $\Delta(\text{pre-post})_{\text{rt-PA}}$ and “Bas2” for 24h post rt-PA), built for t_1 , t_2 , $\Delta(\text{pre-post})_{\text{rt-PA}}$ and including only baseline characteristics, risk factors and comorbidities, generally improved the area under the curve for the prediction of bad events for the AIS patients with respect to cases where the AUC-ROC values have been obtained only considering

baseline clinical characteristics and risk factors (*i.e.* age, gender, blood collection center, time onset-to-treatment, recent infections or inflammations, glycemia, NIHSS and history of atrial fibrillation and congestive heart failure).

AUC values for “baseline” and “baseline plus metabolic features” ROC models, built for t_1 , t_2 and $\Delta(\text{pre-post})_{\text{rt-PA}}$, are listed in **Table 2**, together with relative 95% confidence intervals. *P*-values are also reported to highlight significant changes in the prediction of poor outcomes after considering the association of metabolomic and clinical features.

In detail, statistically significant increases of baseline AUC values for the prediction of impairment, haemorrhagic transformation and non-response to the thrombolysis, were observed considering only t_2 concentrations of metabolites and lipids. For the prediction of three-months impairments, the addition of 24h post rt-PA values of VLDL-5 cholesterol, LDL-6 free cholesterol, LDL-5 Apo B, to the “Bas2” ROC model, led to a statically significant increase in the already good value of baseline AUC [model Bas2: AUC = 0.807 (95% CI 0.748-0.866), model Bas2 + E: AUC = 0.844 (95% CI 0.791 – 0.895), *P*-value = 0.037] (**Table 2**). Instead, for predicting intracranial haemorrhagic transformation of cerebral infarction, the addition of VLDL-2 cholesterol, phospholipids and triglycerides statistically increased the baseline non-informative AUC value of 0.556 to a less accurate value of prediction corresponding to 0.636 [model Bas2: AUC = 0.556, (95% CI 0.438-0.673), model Bas2 + H: AUC = 0.636 (95% CI 0.536 – 0.736), *P*-value = 0.039] (**Table 2**).

Lastly, age, sex, blood-glucose, baseline NIHSS, time onset-to-treatment, history of atrial fibrillation, congestive heart failure, recent infections or inflammations and the blood collection centre are unable to predict the response to intravenous thrombolysis at t_2 [model Bas2: AUC = 0.560, (95% CI 0.486-0.635)]. However, adding 3-hydroxybutyrate and acetone to the above-described ROC curve model, the area under the curve increases to a value of 0.617 (95% CI 0.545-0.69), with a *P*-value = 0.015 (model Bas2 + K, **Table 2**), thus providing a less accurate predictive power of the new model, but encouraging further researches in this direction.

Univariate analysis

To identify within-subject variations in metabolites and lipids levels related to variations introduced by the thrombolytic intervention, the Friedman test followed by post-hoc Nemenyi analysis was applied on assigned metabolites and lipids, considering all time-points of blood-collections (t_1 , t_2 and t_3). Many statistically significant variations of concentrations (FDR < 0.01) are described for the majority of quantified metabolites and lipid parameters (supplementary **Table S4**). Looking at the FDR values reported in **Table 3**, the thrombolytic intervention seems to induce significant variations both in the metabolome and in the lipidome of AIS patients; metabolic variables change mainly between t_1 and t_3 . Specifically, between t_1 and t_2 , lipid parameters related mainly to HDL particles presented the highest statistically significant difference (FDR < 10^{-4}), together with citrate and lactate. Interestingly, between t_2 and t_3 , lipids related mainly to LDL particles and LDL sub-particles change

significantly, reporting FDR values ranging from 10^{-14} to 10^{-5} . Acetone, 3-hydroxybutyrate, glycine, glucose and glutamine also vary with $FDR < 10^{-5}$. Lastly, the significant variation in LDL related parameters, acetone, 3-hydroxybutyrate and glycine concentration at t_3 , has been confirmed, after considering the comparison of metabolites and lipids values at t_1 and t_3 .

Discussion

In this large cohort of AIS patients treated with intravenous thrombolysis, we found that, firstly, for the prediction of three-months death: baseline HDL-3 triglycerides; 24h post rt-PA levels of LDL-6 triglycerides, HDL triglycerides, LDL-6 cholesterol and $\Delta(\text{pre-post})_{\text{rt-PA}}$ variations of triglycerides related to LDL, HDL and LDL-2 particles are statistically associated to post-stroke mortality, but they are not able to statistically significantly improve the AUC of the baseline ROC model for the prediction of the bad event.

Secondly, for the prediction of three-month impairment development: t_1 levels of HDL-3 phospholipids, HDL phospholipids main fraction, HDL/LDL cholesterol improved the AUC of the baseline ROC model; while the addition of t_2 concentrations of VLDL-5 cholesterol, LDL-6 free cholesterol and LDL-5 Apo B values to related baseline ROC models, significantly improved the AUC (P -value < 0.05). Instead, a maintenance in the predictive power of the ROC model was observed when $\Delta(\text{pre-post})_{\text{rt-PA}}$ of the top three statistically significant metabolic features associated to the development of three-months impairments have been added to the baseline ROC model (**Table 2**).

Thirdly, baseline acetone, 3-hydroxybutyrate, pre-post rt-PA variation of VLDL phospholipids, VLDL-2 cholesterol and VLDL-2 phospholipids resulted statistically associated to the onset of intracranial haemorrhage, but they did not improve the AUC values of related ROC predictive models. Only adding the combination of VLDL-2 cholesterol/phospholipids/triglycerides, estimated at t_2 , to baseline ROC model, we obtained a significant improvement in the AUC.

Lastly, t_2 levels of ketone bodies (*i.e.* acetone and 3-hydroxybutyrate) resulted to be statistically associated to the non-response to thrombolysis, and they also improved the ability of the ROC model in predicting the bad event, while the addition of baseline values (t_1) of the same features did not statistically improve the AUC of the baseline ROC model, as for the addition of $\Delta(\text{pre-post})_{\text{rt-PA}}$ variations of acetone, LDL-3 particle number and LDL-3 Apo-B values.

Considering within-subject metabolic variations, we noted that generally, HDL-related lipid parameters tend to decrease after the thrombolysis (t_2) and then, their levels increase again at three-months (t_3) from the acute ischemic stroke, but without returning to the baseline concentrations (t_1). Instead, LDL-related parameters present an opposite behaviour: they tend to increase at t_2 , while at t_3 , their concentrations reach values lower than those at t_1 . Mean values of acetone, 3-hydroxybutyrate and glucose

increase at t_2 , while glycine and glutamine present the highest mean values at t_3 . Lactate and citrate were found to be statistically increased at t_1 .

In the light of the above, we can firstly state that metabolic alterations occurring in the post-AIS course, mainly affect the serum lipidome of patients rather than the serum metabolome. In particular, we observed that t_2 triglycerides levels, associated mainly with HDL and LDL particles, seem to be related to three-months mortality, while alteration of cholesterol and phospholipids levels mainly related to smaller and denser VLDL and LDL sub-particles may be involved in the development of post-stroke impairments and neurological disabilities. Acetone and 3-hydroxybutyrate appear as largely involved in the symptomatic development of intracranial haemorrhage and in the non-response to the thrombolytic therapy.

Many studies evidenced relationships between lipidic features and ischemic stroke,^{3,6,32-34} demonstrating how these molecules play effective roles in the aetiology and progression of the disease. Serum cholesterol has been found to be an independent predictor for long-term functional outcomes and higher serum total cholesterol levels have been associated with better prognosis.³⁵ Triacylglycerols have been significantly associated with ischemic stroke,³⁶ but the biological mechanisms by which they could affect the mortality or the survival of IS patients need further investigations. Since triglycerides are hydrolysed to fatty acids to furnish alternative energy source during stress conditions, we can hypothesize that their statistically significant association with post-stroke mortality may evidence a situation of energy failure, thus leading to an increase of energy demand and to an enhanced transition from aerobic to anaerobic glycolysis. As a consequence, levels of pyruvate and lactate may change; in our case, the statistically significant decrease of lactate concentrations, after the thrombolysis, may suggest its decreased role in providing substitute energy fuel and in metabolic pathways of neuroprotection where it is normally largely involved.³⁷ Moreover, during cerebral ischemia, citrate and ketone bodies levels can change to restore energy homeostasis. Recently, Di Marino and co-workers demonstrated a significant increase of serum ketone bodies in response to angioplasty-induced ischemia applied in patients with stable angina, hypothesizing that changes in the metabolism of ketone bodies could be related to the reperfusion oxidative stress; in other words, it seems that they are playing a fundamental role in free radical homeostasis during ischemia-reperfusion injury.³⁸ Then, there are various studies, where the biochemistry of ketogenesis and its role for cardiovascular diseases,^{39,40} neurological diseases and oxidative stress⁴¹ have been deeply explored. Ketone bodies seem to have the chemical potential to be active antioxidants,⁴¹ proving benefits in diseases associated with oxidative stress, as in the case of ischemic stroke. However, there are conflicting evidences for the antioxidant role of ketone bodies.⁴¹

Other studies demonstrated that acetone is increased in the exhaled breath of patients with heart failure,⁴² a condition which can often occur after stressful situation, as in the case of cerebral ischemia, but the exact mechanisms that lead to increasing concentration of exhaled or serum acetone and other ketone bodies need further biological investigations.

Before concluding, we also reported statistically significant associations of low-density lipoproteins (especially small and denser LDL and VLDL) estimated at t_2 , with three-months death and impairments development. In past studies, it has been shown that AIS is associated with adverse distributions of LDL and HDL subclasses, and particularly, short-term mortality is linked to increased levels of small dense LDL particles (sdLDL).⁴³ Our results also evidence the role of low LDL and VLDL cholesterol, estimated at t_2 , in increasing the risk of early symptomatic intracranial haemorrhage development (*e.g.* VLDL cholesterol: OR = 2.6, P -value = 0.007, all associations are reported completely in **Table 1**).

Moreover, the statistical significant changes in HDL concentrations passing from t_1 to t_2 could reflect a general condition of inflammation, since recently, it has been reported that inflammation may alter the lipoprotein profile as well, for example, modulating the HDL functions.⁴⁴ However, changes in HDL-related parameters may depend on the activity of rt-PA. Indeed, generated plasmin after rt-PA activity on plasminogen can degrade non-target proteins, including Apo A1 which represents the major protein constituent of HDL particles. In a past study,⁴⁵ authors demonstrated that treatment with alteplase and tenecteplase induced apoA1 proteolysis, potentially causing a transient impairment of HDL atheroprotective functions.⁴⁶

Conclusions

In summary, using logistic regression and ROC curve analysis on baseline, 24h post- and $\Delta(\text{pre-post})_{\text{rt-PA}}$ metabolites and lipids concentrations; various metabolomic features, especially lipids related to HDL, LDL, VLDL particles and ketone bodies, resulted statistically significantly associated with each of the assessed post-AIS poor outcome, but to date, few statistically significant metabolites and/or lipids may represent new promising biomarkers to predict especially three-months disability, intracranial haemorrhagic transformation of the stroke and the non-response to the thrombolytic therapy. Instead, applying univariate pairwise analysis in a larger time-scale of three-months with respect to the beginning of the i.v. thrombolysis, many metabolic changes, in both serum metabolome and lipoprotein profiles, reflect a general condition of energy failure, oxidative stress and systemic inflammation.

One limitation of this study is the lack of a control groups of stroke patients not treated with rt-PA and a further limitation rests upon the lack of subsequent blood sampling to determine more in detail, the metabolic variations during the first phase of the post-stroke course. In addition, it must be kept in mind that this study was a retrospective observational study, and for various subjects, clinical information was missing, thus may leading to selection bias.

To conclude, our results support the usefulness of the metabolomic analysis for the identification of a more-detailed risk profile in AIS patients in relation to three-months poor outcomes and at different time-points of blood collection with respect to the thrombolytic treatment. Present findings encourage the application of NMR-based metabolomics for a fast and reproducible definition of the metabolic pattern of AIS

patients associated with three-months mortality, impairments, haemorrhagic stroke transformation and non-response to the thrombolytic therapy, providing additional information compared to what is already known from the clinic. However, further biological explanations are needed to deepen the understanding of the specific role of rt-PA in post-stroke variations on patient's serum metabolome.

References

- (1) Feigin Valery L.; Norrving Bo; Mensah George A. Global Burden of Stroke. *Circulation Research* 2017, 120 (3), 439–448. <https://doi.org/10.1161/CIRCRESAHA.116.308413>.
- (2) Wesley, U. V.; Bhute, V. J.; Hatcher, J. F.; Palecek, S. P.; Dempsey, R. J. Local and Systemic Metabolic Alterations in Brain, Plasma, and Liver of Rats in Response to Aging and Ischemic Stroke, as Detected by Nuclear Magnetic Resonance (NMR) Spectroscopy. *Neurochem. Int.* 2019, 127, 113–124. <https://doi.org/10.1016/j.neuint.2019.01.025>.
- (3) Jung, J. Y.; Lee, H.-S.; Kang, D.-G.; Kim, N. S.; Cha, M. H.; Bang, O.-S.; Ryu, D. H.; Hwang, G.-S. ¹H-NMR-Based Metabolomics Study of Cerebral Infarction. *Stroke* 2011, 42 (5), 1282–1288. <https://doi.org/10.1161/STROKEAHA.110.598789>.
- (4) Szpetnar, M.; Hordyjewska, A.; Malinowska, I.; Golab, P.; Kurzepa, J. The Fluctuation of Free Amino Acids in Serum during Acute Ischemic Stroke. *Current Issues in Pharmacy and Medical Sciences* 2016, 29 (4), 151–154. <https://doi.org/10.1515/cipms-2016-0031>.
- (5) Wang, D.; Kong, J.; Wu, J.; Wang, X.; Lai, M. GC-MS-Based Metabolomics Identifies an Amino Acid Signature of Acute Ischemic Stroke. *Neurosci. Lett.* 2017, 642, 7–13. <https://doi.org/10.1016/j.neulet.2017.01.039>.
- (6) Liu, P.; Li, R.; Antonov, A. A.; Wang, L.; Li, W.; Hua, Y.; Guo, H.; Wang, L.; Liu, P.; Chen, L.; Tian, Y.; Xu, F.; Zhang, Z.; Zhu, Y.; Huang, Y. Discovery of Metabolite Biomarkers for Acute Ischemic Stroke Progression. *J. Proteome Res.* 2017, 16 (2), 773–779. <https://doi.org/10.1021/acs.jproteome.6b00779>.
- (7) Kimberly, W. T.; Wang, Y.; Pham, L.; Furie, K. L.; Gerszten, R. E. Metabolite Profiling Identifies a Branched Chain Amino Acid Signature in Acute Cardioembolic Stroke. *Stroke* 2013, 44 (5), 1389–1395. <https://doi.org/10.1161/STROKEAHA.111.000397>.
- (8) Liu, M.; Zhou, K.; Li, H.; Dong, X.; Tan, G.; Chai, Y.; Wang, W.; Bi, X. Potential of Serum Metabolites for Diagnosing Post-Stroke Cognitive Impairment. *Mol Biosyst* 2015, 11 (12), 3287–3296. <https://doi.org/10.1039/c5mb00470e>.
- (9) Ke, C.; Pan, C.-W.; Zhang, Y.; Zhu, X.; Zhang, Y. Metabolomics Facilitates the Discovery of Metabolic Biomarkers and Pathways for Ischemic Stroke: A Systematic Review. *Metabolomics* 2019, 15 (12), 152. <https://doi.org/10.1007/s11306-019-1615-1>.
- (10) Hasan, N.; McColgan, P.; Bentley, P.; Edwards, R. J.; Sharma, P. Towards the Identification of Blood Biomarkers for Acute Stroke in Humans: A Comprehensive

Systematic Review. *Br J Clin Pharmacol* 2012, 74 (2), 230–240. <https://doi.org/10.1111/j.1365-2125.2012.04212.x>.

(11) Tirschwell, D. L.; Smith, N. L.; Heckbert, S. R.; Lemaitre, R. N.; Longstreth, W. T.; Psaty, B. M. Association of Cholesterol with Stroke Risk Varies in Stroke Subtypes and Patient Subgroups. *Neurology* 2004, 63 (10), 1868–1875. <https://doi.org/10.1212/01.wnl.0000144282.42222.da>.

(12) Kurth, T.; Everett, B. M.; Buring, J. E.; Kase, C. S.; Ridker, P. M.; Gaziano, J. M. Lipid Levels and the Risk of Ischemic Stroke in Women. *Neurology* 2007, 68 (8), 556–562. <https://doi.org/10.1212/01.wnl.0000254472.41810.0d>.

(13) Vauthey, C.; de Freitas, G. R.; van Melle, G.; Devuyst, G.; Bogousslavsky, J. Better Outcome after Stroke with Higher Serum Cholesterol Levels. *Neurology* 2000, 54 (10), 1944–1949. <https://doi.org/10.1212/wnl.54.10.1944>.

(14) Dyker, A. G.; Weir, C. J.; Lees, K. R. Influence of Cholesterol on Survival after Stroke: Retrospective Study. *BMJ* 1997, 314 (7094), 1584–1588. <https://doi.org/10.1136/bmj.314.7094.1584>.

(15) Bang, O. Y.; Saver, J. L.; Liebeskind, D. S.; Starkman, S.; Villablanca, P.; Salamon, N.; Buck, B.; Ali, L.; Restrepo, L.; Vinuela, F.; Duckwiler, G.; Jahan, R.; Razinia, T.; Ovbiagele, B. Cholesterol Level and Symptomatic Hemorrhagic Transformation after Ischemic Stroke Thrombolysis. *Neurology* 2007, 68 (10), 737–742. <https://doi.org/10.1212/01.wnl.0000252799.64165.d5>.

(16) Restrepo, L.; Bang, O. Y.; Ovbiagele, B.; Ali, L.; Kim, D.; Liebeskind, D. S.; Starkman, S.; Vinuela, F.; Duckwiler, G. R.; Jahan, R.; Saver, J. L. Impact of Hyperlipidemia and Statins on Ischemic Stroke Outcomes after Intra-Arterial Fibrinolysis and Percutaneous Mechanical Embolectomy. *Cerebrovasc. Dis.* 2009, 28 (4), 384–390. <https://doi.org/10.1159/000235625>.

(17) Vignoli, A.; Ghini, V.; Meoni, G.; Licari, C.; Takis, P. G.; Tenori, L.; Turano, P.; Luchinat, C. High-Throughput Metabolomics by 1D NMR. *Angew. Chem. Int. Ed. Engl.* 2019, 58 (4), 968–994. <https://doi.org/10.1002/anie.201804736>.

(18) Takis, P. G.; Ghini, V.; Tenori, L.; Turano, P.; Luchinat, C. Uniqueness of the NMR Approach to Metabolomics. *TrAC Trends in Analytical Chemistry* 2019, 120, 115300. <https://doi.org/10.1016/j.trac.2018.10.036>.

(19) Meoni, G.; Lorini, S.; Monti, M.; Madia, F.; Corti, G.; Luchinat, C.; Zignego, A. L.; Tenori, L.; Gragnani, L. The Metabolic Fingerprints of HCV and HBV Infections Studied by Nuclear Magnetic Resonance Spectroscopy. *Scientific Reports* 2019, 9 (1), 4128. <https://doi.org/10.1038/s41598-019-40028-4>.

(20) Vignoli, A.; Orlandini, B.; Tenori, L.; Biagini, M. R.; Milani, S.; Renzi, D.; Luchinat, C.; Calabrò, A. S. Metabolic Signature of Primary Biliary Cholangitis and Its Comparison with Celiac Disease. *J. Proteome Res.* 2019, 18 (3), 1228–1236. <https://doi.org/10.1021/acs.jproteome.8b00849>.

(21) Vignoli, A.; Paciotti, S.; Tenori, L.; Eusebi, P.; Biscetti, L.; Chiasserini, D.; Scheltens, P.; Turano, P.; Teunissen, C.; Luchinat, C.; Parnetti, L. Fingerprinting Alzheimer's Disease by ¹H Nuclear Magnetic Resonance Spectroscopy of

Cerebrospinal Fluid. *J. Proteome Res.* 2020, 19 (4), 1696–1705. <https://doi.org/10.1021/acs.jproteome.9b00850>.

(22) Caracausi, M.; Ghini, V.; Locatelli, C.; Mericio, M.; Piovesan, A.; Antonaros, F.; Pelleri, M. C.; Vitale, L.; Vacca, R. A.; Bedetti, F.; Mimmi, M. C.; Luchinat, C.; Turano, P.; Strippoli, P.; Cocchi, G. Plasma and Urinary Metabolomic Profiles of Down Syndrome Correlate with Alteration of Mitochondrial Metabolism. *Scientific Reports* 2018, 8 (1), 2977. <https://doi.org/10.1038/s41598-018-20834-y>.

(23) Vignoli, A.; Tenori, L.; Giusti, B.; Takis, P. G.; Valente, S.; Carrabba, N.; Balzi, D.; Barchielli, A.; Marchionni, N.; Gensini, G. F.; Marcucci, R.; Luchinat, C.; Gori, A. M. NMR-Based Metabolomics Identifies Patients at High Risk of Death within Two Years after Acute Myocardial Infarction in the AMI-Florence II Cohort. *BMC Med* 2019, 17 (1), 3. <https://doi.org/10.1186/s12916-018-1240-2>.

(24) Di Donato, S.; Mislange, A. R.; Vignoli, A.; Mori, E.; Vitale, S.; Biagioni, C.; Hart, C.; Becheri, D.; Del Monte, F.; Luchinat, C.; Di Leo, A.; Mottino, G.; Tenori, L.; Biganzoli, L. Serum Metabolomic as Biomarkers to Differentiate Early from Metastatic Disease in Elderly Colorectal Cancer (Crc) Patients. *Annals of Oncology* 2016, 27. <https://doi.org/10.1093/annonc/mdw335.20>.

(25) Inzitari Domenico; Giusti Betti; Nencini Patrizia; Gori Anna Maria; Nesi Mascia; Palumbo Vanessa; Piccardi Benedetta; Armillis Alessandra; Pracucci Giovanni; Bono Giorgio; Bovi Paolo; Consoli Domenico; Guidotti Mario; Nucera Antonia; Massaro Francesca; Micieli Giuseppe; Orlandi Giovanni; Perini Francesco; Tassi Rossana; Tola Maria Rosaria; Sessa Maria; Toni Danilo; Abbate Rosanna. MMP9 Variation After Thrombolysis Is Associated With Hemorrhagic Transformation of Lesion and Death. *Stroke* 2013, 44 (10), 2901–2903. <https://doi.org/10.1161/STROKEAHA.113.002274>.

(26) Gori, A. M.; Giusti, B.; Piccardi, B.; Nencini, P.; Palumbo, V.; Nesi, M.; Nucera, A.; Pracucci, G.; Tonelli, P.; Innocenti, E.; Sereni, A.; Sticchi, E.; Toni, D.; Bovi, P.; Guidotti, M.; Tola, M. R.; Consoli, D.; Micieli, G.; Tassi, R.; Orlandi, G.; Sessa, M.; Perini, F.; Delodovici, M. L.; Zedde, M. L.; Massaro, F.; Abbate, R.; Inzitari, D. Inflammatory and Metalloproteinases Profiles Predict Three-Month Poor Outcomes in Ischemic Stroke Treated with Thrombolysis. *J. Cereb. Blood Flow Metab.* 2017, 37 (9), 3253–3261. <https://doi.org/10.1177/0271678X17695572>.

(27) Wahlgren, N.; Ahmed, N.; Dávalos, A.; Ford, G. A.; Grond, M.; Hacke, W.; Hennerici, M. G.; Kaste, M.; Kuelkens, S.; Larrue, V.; Lees, K. R.; Roine, R. O.; Soenne, L.; Toni, D.; Vanhooren, G.; SITS-MOST investigators. Thrombolysis with Alteplase for Acute Ischaemic Stroke in the Safe Implementation of Thrombolysis in Stroke-Monitoring Study (SITS-MOST): An Observational Study. *Lancet* 2007, 369 (9558), 275–282. [https://doi.org/10.1016/S0140-6736\(07\)60149-4](https://doi.org/10.1016/S0140-6736(07)60149-4).

(28) Jiménez, B.; Holmes, E.; Heude, C.; Tolson, R. F.; Harvey, N.; Lodge, S. L.; Chetwynd, A. J.; Cannet, C.; Fang, F.; Pearce, J. T. M.; Lewis, M. R.; Viant, M. R.; Lindon, J. C.; Spraul, M.; Schäfer, H.; Nicholson, J. K. Quantitative Lipoprotein Subclass and Low Molecular Weight Metabolite Analysis in Human Serum and

- Plasma by 1H NMR Spectroscopy in a Multilaboratory Trial. *Anal. Chem.* 2018, 90 (20), 11962–11971. <https://doi.org/10.1021/acs.analchem.8b02412>.
- (29) Ihaka, R.; Gentleman, R. R. A Language for Data Analysis and Graphics. *J Comput Stat Graph* 1996, 5, 299–314.
- (30) Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 1995, 289–300.
- (31) Conover, W. J. *Practical Nonparametric Statistics*, 3rd, 3rd edition.; Wiley: New York, 1999.
- (32) Yang, L.; Lv, P.; Ai, W.; Li, L.; Shen, S.; Nie, H.; Shan, Y.; Bai, Y.; Huang, Y.; Liu, H. Lipidomic Analysis of Plasma in Patients with Lacunar Infarction Using Normal-Phase/Reversed-Phase Two-Dimensional Liquid Chromatography-Quadrupole Time-of-Flight Mass Spectrometry. *Anal Bioanal Chem* 2017, 409 (12), 3211–3222. <https://doi.org/10.1007/s00216-017-0261-6>.
- (33) Ding, X.; Liu, R.; Li, W.; Ni, H.; Liu, Y.; Wu, D.; Yang, S.; Liu, J.; Xiao, B.; Liu, S. A Metabonomic Investigation on the Biochemical Perturbation in Post-Stroke Patients with Depressive Disorder (PSD). *Metab Brain Dis* 2016, 31 (2), 279–287. <https://doi.org/10.1007/s11011-015-9748-z>.
- (34) Makihara, N.; Okada, Y.; Koga, M.; Shiokawa, Y.; Nakagawara, J.; Furui, E.; Kimura, K.; Yamagami, H.; Hasegawa, Y.; Kario, K.; Okuda, S.; Naganuma, M.; Toyoda, K. Effect of Serum Lipid Levels on Stroke Outcome after Rt-PA Therapy: SAMURAI Rt-PA Registry. *Cerebrovasc. Dis.* 2012, 33 (3), 240–247. <https://doi.org/10.1159/000334664>.
- (35) Pan, S.-L.; Lien, I.-N.; Chen, T. H.-H. Is Higher Serum Total Cholesterol Level Associated with Better Long-Term Functional Outcomes after Noncardioembolic Ischemic Stroke? *Arch Phys Med Rehabil* 2010, 91 (6), 913–918. <https://doi.org/10.1016/j.apmr.2010.02.002>.
- (36) Labreuche, J.; Deplanque, D.; Touboul, P.-J.; Bruckert, E.; Amarenco, P. Association between Change in Plasma Triglyceride Levels and Risk of Stroke and Carotid Atherosclerosis: Systematic Review and Meta-Regression Analysis. *Atherosclerosis* 2010, 212 (1), 9–15. <https://doi.org/10.1016/j.atherosclerosis.2010.02.011>.
- (37) Berthet, C.; Castillo, X.; Magistretti, P. J.; Hirt, L. New Evidence of Neuroprotection by Lactate after Transient Focal Cerebral Ischaemia: Extended Benefit after Intracerebroventricular Injection and Efficacy of Intravenous Administration. *Cerebrovasc. Dis.* 2012, 34 (5–6), 329–335. <https://doi.org/10.1159/000343657>.
- (38) Di Marino, S.; Viceconte, N.; Lembo, A.; Summa, V.; Tanzilli, G.; Raparelli, V.; Truscelli, G.; Mangieri, E.; Gaudio, C.; Cicero, D. O. Early Metabolic Response to Acute Myocardial Ischaemia in Patients Undergoing Elective Coronary Angioplasty. *Open Heart* 2018, 5 (1). <https://doi.org/10.1136/openhrt-2017-000709>.

- (39) Cotter, D. G.; Schugar, R. C.; Crawford, P. A. Ketone Body Metabolism and Cardiovascular Disease. *Am. J. Physiol. Heart Circ. Physiol.* 2013, 304 (8), H1060-1076. <https://doi.org/10.1152/ajpheart.00646.2012>.
- (40) Martin-Lorenzo, M.; Zubiri, I.; Maroto, A. S.; Gonzalez-Calero, L.; Posada-Ayala, M.; de la Cuesta, F.; Mourino-Alvarez, L.; Lopez-Almodovar, L. F.; Calvo-Bonacho, E.; Ruilope, L. M.; Padial, L. R.; Barderas, M. G.; Vivanco, F.; Alvarez-Llamas, G. KLK1 and ZG16B Proteins and Arginine–Proline Metabolism Identified as Novel Targets to Monitor Atherosclerosis, Acute Coronary Syndrome and Recovery. *Metabolomics* 2014, 11 (5), 2. <https://doi.org/10.1007/s11306-014-0761-8>.
- (41) McPherson, P. A. C.; McEneny, J. The Biochemistry of Ketogenesis and Its Role in Weight Management, Neurological Disease and Oxidative Stress. *J. Physiol. Biochem.* 2012, 68 (1), 141–151. <https://doi.org/10.1007/s13105-011-0112-4>.
- (42) Marcondes-Braga, F. G.; Gutz, I. G. R.; Batista, G. L.; Saldiva, P. H. N.; Ayub-Ferreira, S. M.; Issa, V. S.; Mangini, S.; Bocchi, E. A.; Bacal, F. Exhaled Acetone as a New Biomarker of Heart Failure Severity. *Chest* 2012, 142 (2), 457–466. <https://doi.org/10.1378/chest.11-2892>.
- (43) Zeljkovic, A.; Vekic, J.; Spasojevic-Kalimanovska, V.; Jelic-Ivanovic, Z.; Bogavac-Stanojevic, N.; Gulan, B.; Spasic, S. LDL and HDL Subclasses in Acute Ischemic Stroke: Prediction of Risk and Short-Term Mortality. *Atherosclerosis* 2010, 210 (2), 548–554. <https://doi.org/10.1016/j.atherosclerosis.2009.11.040>.
- (44) McGarrah, R. W.; Kelly, J. P.; Craig, D. M.; Haynes, C.; Jessee, R. C.; Huffman, K. M.; Kraus, W. E.; Shah, S. H. A Novel Protein Glycan-Derived Inflammation Biomarker Independently Predicts Cardiovascular Disease and Modifies the Association of HDL Subclasses with Mortality. *Clin. Chem.* 2017, 63 (1), 288–296. <https://doi.org/10.1373/clinchem.2016.261636>.
- (45) Gomaschi, M.; Ossoli, A.; Vitali, C.; Pozzi, S.; Vitali Serdoz, L.; Pitzorno, C.; Sinagra, G.; Franceschini, G.; Calabresi, L. Off-Target Effects of Thrombolytic Drugs: Apolipoprotein A-I Proteolysis by Alteplase and Tenecteplase. *Biochem. Pharmacol.* 2013, 85 (4), 525–530. <https://doi.org/10.1016/j.bcp.2012.11.023>.
- (46) Meilhac, O. High-Density Lipoproteins in Stroke. In *High Density Lipoproteins: From Biological Understanding to Clinical Exploitation*; von Eckardstein, A., Kardassis, D., Eds.; Handbook of Experimental Pharmacology; Springer International Publishing: Cham, 2015; pp 509–526. https://doi.org/10.1007/978-3-319-09665-0_16.

Tables

Table 1. Effect of statistically significantly (P -values < 0.05 and/or FDR < 0.05) associated pre (t_1), 24h post (t_2) and Δ (pre-post) $_{rt-PA}$ metabolites and lipids concentrations on three-month mortality, impairments, symptomatic intracerebral haemorrhage (sICHRCT12) and on the non-response to i.v. thrombolysis intervention, adjusting for major determinants for unfavourable outcomes, *i.e.* age, sex, time onset-to-treatment, pre or 24h post rt-PA blood glucose level (for t_1 and t_2 models, respectively), baseline NIHSS, history of atrial fibrillation, congestive heart failure, recent infections or inflammations, hypertension, diabetes, hyperlipidaemia, smoke and blood collection center.

Three-month mortality				Three-month impairments				SICHRCT12				Non-response to thrombolysis			
Analytes	OR (95% CI)	P	FDR	Analytes	OR (95% CI)	P	FDR	Analytes	OR (95% CI)	P	FDR	Analytes	OR (95% CI)	P	FDR
PRE (or BASELINE) rt-PA (t_1)															
SubTrigl	0.55 (0.30-	0.04	0.98	LDL-HDL-	1.69 (1.16-	0.00	0.18	3-HB	0.65 (0.45-	0.02	0.39	3-HB	0.71 (0.52-	0.0	0.27
HDL-3	1.0)			Chol	2.46)	6		0.94)	0.02	1.0)		3			
				Apo-B100-	1.73 (1.16-	0.00	0.18	Acetone	0.69 (0.48-	0.04	0.43	Acetone	0.66 (0.48-	0.0	0.21
				Apo-A1	2.59)	7		1.00)	1.00)	1					
				LDL-6 PN	1.47 (1.03-	0.03	0.27								
					2.09)	5									
				LMF Phosp	0.55 (0.36-	0.00	0.18								
				HDL	0.85)	7									
				LMF ApoA1	0.64 (0.43-	0.03	0.27								
				HDL	0.96)	3									
				SubChol	0.584 (0.39-	0.00	0.18								
				VLDL-5	0.87)	9									
				SubPhosp	0.61 (0.41-	0.01	0.22								
				VLDL-5	0.91)	6									
				SubTrigl	1.53 (1.07-	0.01	0.22								
				LDL-6	2.18)	9									

SubChol	1.52 (1.06-	0.02	
LDL-6	2.18)	4	0.22
SubFreeChol	1.64 (1.13-	0.00	
LDL-6	2.38)	9	0.18
SubPhosp	1.44 (1.01-	0.04	
LDL-6	2.05)	7	0.31
SubApoB	1.47 (1.03-	0.03	
LDL-5	2.09)	5	0.27
SubChol	0.66 (0.43-	0.04	
HDL-3	0.99)	6	0.31
SubPhosp	0.60 (0.39-	0.01	
HDL-2	0.92)	8	0.22
SubPhosp	0.51 (0.33-	0.00	
HDL-3	0.80)	4	0.18
SubApoA1	0.61 (0.40-	0.02	
HDL-2	0.93)	3	0.22
SubApoA1	0.60 (0.39-	0.02	
HDL-3	0.93)	2	0.22

24h POST rt-PA (t₂)

LMF Trigl	0.46 (0.22-	0.03	0.73	Aceticacid	1.77 (1.16-	0.00	0.14	Trigl	2.60 (1.09-	0.03	0.20	Phe	0.7 (0.52-	0.0	0.13
HDL	0.94)				2.69)	8			6.20)				0.95)	2	
SubTrigl	1.97 (1.19-	0.00	0.73	3-HB	1.473(1.009-	0.04	0.40	LMF Trigl	3.076 (1.21-	0.02	0.14	3-HB	0.54 (0.38-	0.0	0.009
LDL-6	3.26)	8			2.15)	5		VLDL	7.79)				0.77)	01	
SubChol	2.004 (1.05-	0.04	0.73	LDL-Chol	1.54 (1.009-	0.04	0.22	LMF Trigl IDL	2.51 (1.08-	0.03	0.205	Acetone	0.58 (0.40-	0.0	0.04
LDL-6	3.83)				2.34)	5	5		5.85)				0.84)	04	
SubPhosp	1.95 (1.02-	0.04	0.73	Apo-B100	1.69 (1.11-	0.01	0.14	LMF Chol	2.64 (1.30-	0.00	0.12	SubPhosp	1.36 (1.002-	0.0	0.64
LDL-6	3.74)				2.58)	5		VLDL	5.35)	7		VLDL-5	1.84)	5	
SubTrigl	0.47 (0.23-	0.04	0.73	LDL-HDL-	1.65 (1.12-	0.01	0.12	LMF FreeChol	2.87 (1.33-	0.00	0.119	SubTrigl	1.62 (1.16-	0.0	0.51
HDL-3	0.95)			Chol	2.44)	1		VLDL	6.19)	7		HDL-4	2.27)	04	

SubTrigl	0.46 (0.23-	0.04	0.73	Apo-B100-	1.87 (1.20-	0.00	0.06	LMF Phosp	3.15 (1.38-	0.00		SubFreeChol	0.72 (0.52-	0.0	
HDL-4	0.95)			Apo-A1	2.90)	5	7	VLDL	7.19)	6	0.119	HDL-2	0.99)	4	0.64
				TPN ApoB	1.69 (1.11-	0.01	0.14	SubTrigl	3.22 (1.43-	0.00	0.12				
					2.58)	5		VLDL-2	7.21)	5					
				LDL PN	1.61 (1.06-	0.02	0.19	SubTrigl	2.60 (1.26-	0.00	0.12				
					2.43)	6		VLDL-3	5.35)	9					
				LDL3 PN	1.51 (1.002-	0.04	0.22	SubChol	3.31 (1.23-	0.01	0.14				
					2.29)	9		VLDL-1	8.91)	8					
				LDL6 PN	1.925 (1.27-	0.00	0.05	SubChol	3.18 (1.46-	0.00	0.12				
					2.91)	2	2	VLDL-2	6.91)	3					
				LMF Trigl	1.56 (1.01-	0.04	0.22	SubChol	2.46 (1.25-	0.00	0.12				
				VLDL	2.41)	4		VLDL-3	4.85)	9					
				LMF Chol	1.58 (1.05-	0.02	0.19	SubChol	1.97 (1.096-	0.02	0.18				
				IDL	2.38)	8	4	VLDL-4	3.53)						
				LMF Chol	1.54 (1.009-	0.04	0.22	SubFreeChol	3.078 (1.14-	0.03	0.19				
				LDL	2.34)	5	5	VLDL-1	8.31)						
				LMF	1.62 (1.06-	0.02	0.19	SubFreeChol	2.64 (1.25-5.6)	0.01	0.13				
				FreeChol LDL	2.47)	6	4	VLDL-2							
				LMF Phosp	0.63 (0.42-	0.03	0.2	SubFreeChol	2.53 (1.21-	0.01	0.14				
				HDL	0.96)	1		VLDL-3	5.29)						
				LMF ApoB	1.61 (1.059-	0.02	0.19	SubPhosp	3.51 (1.28-	0.01	0.14				
				LDL	2.43)	6		VLDL-1	10.085)						
				SubTrigl	1.60 (1.05-	0.02	0.19	SubPhosp	3.27 (1.51-	0.00	0.12				
				VLDL-1	2.45)	9		VLDL-2	7.12)	3					
				SubChol	0.46 (0.30-	0.01	0.05	SubPhosp	2.55 (1.27-	0.00	0.12				
				VLDL-5	0.71)	2		VLDL-3	5.11)	8					
				SubPhosp	0.55 (0.36-	0.00	0.06	SubTrigl HDL-	1.77 (1.01-	0.04	0.27				
				VLDL-5	0.83)	5	7	4	3.089)						

SubTrigl	1.89 (1.25-	0.00	0.05
LDL-6	2.86)	3	2
SubChol	1.87 (1.24-	0.00	0.05
LDL-6	2.81)	3	2
SubFreeChol	1.98 (1.31-	0.00	0.05
LDL-6	3.0)	1	2
SubPhosp	1.81 (1.21-	0.00	0.06
LDL-6	2.71)	4	1
SubApoB	1.51 (1.001-	0.04	0.22
LDL-3	2.29)	9	
SubApoB	1.92 (1.27-	0.00	0.05
LDL-5	2.91)	2	2
SubPhosp	0.67 (0.45-	0.04	0.22
HDL-3	1.0)	8	

$\Delta(\text{pre-post})_{\text{rt-PA}}$

LMF Trigl	0.56 (0.35-	0.02	0.76	Glu	1.56 (1.05-	0.02	0.44	Phe	0.55 (0.34-	0.01	0.21	Ala	1.35 (1.011-	0.0	0.26
LDL	0.90)				2.32)	6			0.88)				1.80)	4	
LMF Trigl	0.54 (0.33-	0.02	0.76	Aceticacid	1.56 (1.003-	0.04	0.44	Pyruvicacid	0.55 (0.32-	0.03	0.21	Acetone	0.71 (0.53-	0.0	0.26
HDL	0.9)				2.42)	8			0.94)				0.94)	2	
SubTrigl	0.53 (0.30-	0.03	0.76	SubFreeChol	1.52 (1.033-	0.03	0.89	Glucose	0.54 (0.31-	0.03	0.213	Apo-B100	0.74 (0.56-	0.0	0.88
LDL-1	0.95)			HDL-2	2.23)	3			0.96)				0.99)	4	
SubTrigl	0.52 (0.29-	0.03	0.76	SubPhosp	1.49 (1.037-	0.03	0.89	Trigl	1.63 (1.015-	0.04	0.27	TPN	0.74 (0.56-	0.0	0.88
LDL-2	0.93)			HDL-1	2.15)	1			2.62)				0.99)	4	
SubTrigl	0.60 (0.37-	0.03	0.76	SubApoA2	1.52 (1.045-	0.02	0.89	VLDL PN	1.81 (1.11-	0.02	0.22	LDL3 PN	0.69 (0.51-	0.0	0.88
HDL-4	0.96)			HDL-1	2.20)	8			2.96)				0.94)	2	
				SubApoA2	1.54 (1.057-	0.02	0.89	LMF Trigl	1.68 (1.027-	0.04	0.27	SubApoB	0.69 (0.51-	0.0	0.88
				HDL-2	2.25)	5		VLDL	2.74)			LDL-3	0.94)	2	
								LMF Trigl IDL	1.78 (1.033-	0.04	0.27	SubTrigl	1.42 (1.038-	0.0	0.88
									3.068)			HDL-4	1.94)	3	

LMF Chol	1.753 (1.094-	0.02	0.22
VLDL	2.808)		
LMF FreeChol	1.797 (1.117-	0.02	0.22
VLDL	2.89)		
LMF Phosp	2.112 (1.268-	0.00	0.22
VLDL	3.518)	4	
LMF ApoB	1.813 (1.11-	0.02	0.22
VLDL	2.96)		
SubTrigl	1.768 (1.048-	0.03	0.27
VLDL-2	2.983)		
SubTrigl	1.815 (1.031-	0.04	0.27
VLDL-4	3.193)		
SubChol	1.957 (1.207-	0.00	0.22
VLDL-2	3.172)	6	
SubChol	1.694 (1.013-	0.04	0.27
VLDL-3	2.833)		
SubChol	1.593 (1.016-	0.04	0.27
VLDL-4	2.497)		
SubFreeChol	1.917 (1.129-	0.02	0.22
VLDL-1	3.254)		
SubPhosp	2.199 (1.316-	0.00	0.22
VLDL-2	3.675)	3	
SubPhosp	2.042 (1.176-	0.01	0.22
VLDL-3	3.546)		
SubPhosp	1.719 (1.055-	0.03	0.27
VLDL-4	2.803)		
SubTrigl HDL-	1.721 (1.103-	0.02	0.22
3	2.683)		

SubTrigl HDL-	1.63 (1.039-	0.03	0.27
4	2.557)		

* Main abbreviations are reported as follows: NIHSS: National Institutes of Health Stroke Scale; OR: odds ratio; CI: confidence interval; P: *P*-values; trigl: triglycerides; chol: cholesterol; phosp: phospholipids; Apo: apolipoprotein; LMF: lipoprotein main fraction; Sub: subfraction; PN: particle number; 3-HB: 3-hydroxybutyrate. Amino acids are reported with the three letters code.

Table 2. ROC curve models for pre- (t_1), 24h post- (t_2) and $\Delta(\text{pre-post})_{\text{rt-PA}}$ metabolites/lipids values considering four different post-stroke poor outcomes, defined at three months from the bad event (*i.e.* mortality, impairments, haemorrhagic transformation and non-response to the i.v. thrombolytic treatment). AUC values are reported both for baseline models (built considering only clinical characteristics, risk factors and comorbidities) and for models obtained after adding selected combinations among the top three statistically significant metabolomic features, selected from previous logistic regression analysis. *P*-values are also reported to evidence significant changes (highlighted in green) in the prediction of poor outcomes after considering the associations of metabolomic and clinical features. After adding metabolomic features to baseline clinical ROC models, we reported in yellow AUC increases.

	Mortality				Impairment				Haemorrhage				No response to thrombolysis			
	model	AUC	95% CI	<i>P</i> -value	model	AUC	95% CI	<i>P</i> -value	model	AUC	95% CI	<i>P</i> -value	model	AUC	95% CI	<i>P</i> -value
t_1	Bas1	0.72	0.612-0.836	0.825	Bas1	0.81	0.758-0.871	0.193	Bas1	0.58	0.454-0.700	0.349	Bas1	0.54	0.467-0.614	0.258
	Bas1 + A	0.72	0.609-0.826		Bas1 + D	0.83	0.774-0.880		Bas1 + G	0.59	0.459-0.713		Bas1 + J	0.55	0.479-0.625	
t_2	Bas2	0.76	0.662-0.854	0.155	Bas2	0.81	0.748-0.866	0.037	Bas2	0.56	0.438-0.673	0.039	Bas2	0.56	0.486-0.635	0.015
	Bas2 + B	0.80	0.730-0.880		Bas2 + E	0.84	0.791-0.895		Bas2 + H	0.64	0.536-0.736		Bas2 + K	0.62	0.545-0.69	
$\Delta(\text{pre-post})_{\text{rt-PA}}$	Bas1	0.72	0.612-0.836	0.29	Bas1	0.81	0.758-0.871	0.57	Bas1	0.58	0.454-0.700	0.063	Bas1	0.54	0.467-0.614	0.062
	Bas1 + C	0.74	0.636-0.85		Bas1 + F	0.81	0.756-0.87		Bas1 + I	0.65	0.531-0.776		Bas1 + L	0.58	0.509-0.654	

Bas1: age, gender, blood collection center, time onset-to-treatment, recent infections or inflammations, baseline values of glycemia, NIHSS and history of atrial fibrillation and congestive heart failure; Bas2: age, gender, blood collection center, time onset-to-treatment, 24h post rt-PA glycemia, recent infections or inflammations, baseline values of NIHSS, history of atrial fibrillation and congestive heart failure. **A**: HDL-3 triglycerides; **B**: LDL-6 triglycerides, HDL triglycerides, LDL-6 cholesterol; **C**: main fraction triglycerides LDL, HDL, LDL-2

triglycerides; **D**: HDL-3 phospholipids, HDL phospholipids main fraction, HDL/LDL cholesterol; **E**: VLDL-5 cholesterol, LDL-6 free cholesterol, LDL-5 apolipoprotein B; **F**: glutamate, HDL-1 apolipoprotein A2, HDL-2 apolipoprotein A2; **G**: 3-HB, acetone; **H**: VLDL-2 cholesterol/phospholipids/triglycerides; **I**: main fraction VLDL phospholipids, VLDL-2 cholesterol, VLDL-2 phospholipids; **J**: 3-HB, acetone; **K**: 3-hydroxybutyrate, acetone; **L**: acetone, LDL-3 particle number, LDL-3 apolipoprotein B.

Table 3. Metabolites and lipids univariate pairwise-variations related to the thrombolytic intervention (rt-PA). Metabolic variations are evaluated in serum samples collected before (t_1), 24h (t_2) and three-month (t_3) after rt-PA, using the Friedman test followed by post-hoc Nemenyi analysis. FDR values are reported for each comparison performed. FDR < 0.01 are highlighted in light grey.

	t_1-t_2	t_2-t_3	t_1-t_3
Creatinine	8.53E-01	8.53E-01	8.53E-01
Ala	4.63E-01	3.09E-04	1.31E-02
Glu	9.44E-01	9.44E-01	9.44E-01
Gln	2.52E-01	6.80E-06	3.37E-03
Gly	4.01E-01	1.69E-09	1.69E-06
His	6.24E-01	3.06E-01	7.16E-01
Ile	8.91E-01	8.91E-01	8.91E-01
Leu	8.79E-01	8.79E-01	8.79E-01
Phe	6.99E-01	3.25E-01	6.71E-01
Tyr	9.53E-01	9.53E-01	9.53E-01
Val	7.69E-01	9.90E-01	7.69E-01
Aceticacid	2.18E-01	4.48E-03	6.80E-06
Citricacid	7.80E-06	1.29E-01	1.20E-02
Lacticacid	1.53E-05	9.62E-02	2.73E-02
3-HB	8.79E-01	1.08E-08	9.98E-10
Acetone	2.56E-04	1.02E-13	5.34E-07
Pyruvicacid	9.35E-01	8.62E-02	1.06E-01
Glucose	4.80E-01	1.65E-06	2.19E-04
Trigl	6.18E-02	1.78E-01	2.45E-04
Chol	2.67E-02	4.79E-04	9.98E-10
LDL-Chol	7.32E-01	2.76E-07	7.67E-06
HDL-Chol	2.16E-01	9.61E-01	2.16E-01
Apo-A1	2.60E-05	8.79E-01	1.22E-04
Apo-A2	1.83E-03	4.00E-02	4.96E-01
Apo-B100	5.97E-01	7.15E-07	4.79E-09
LDL-HDL-Chol	1.44E-01	6.99E-10	1.69E-05
Apo-B100-Apo-A1	1.25E-03	7.23E-14	8.27E-07
Total PN ApoB	6.31E-01	6.18E-07	5.68E-09
VLDL PN	5.63E-01	1.19E-03	2.15E-02
IDL PN	1.10E-02	7.32E-01	1.48E-03
LDL PN	9.03E-01	1.51E-08	2.02E-09
LDL-1 PN	5.42E-02	3.84E-04	4.03E-09
LDL-2 PN	2.25E-03	5.34E-11	1.87E-03
LDL-3 PN	9.25E-01	2.84E-10	1.70E-09
LDL-4 PN	9.90E-01	4.47E-06	4.47E-06
LDL-5 PN	3.86E-01	9.30E-05	8.48E-03
LDL-6 PN	7.95E-01	7.60E-02	2.40E-02

LMF Trigl VLDL	4.64E-02	7.32E-01	8.96E-03
LMF Trigl IDL	3.15E-01	4.00E-02	4.90E-04
LMF Trigl LDL	4.96E-01	1.74E-04	1.43E-06
LMF Trigl HDL	1.65E-06	1.02E-01	6.48E-03
LMF Chol VLDL	1.44E-01	1.44E-01	1.48E-03
LMF Chol IDL	5.40E-03	3.15E-01	2.60E-05
LMF Chol LDL	7.32E-01	2.76E-07	7.67E-06
LMF Chol HDL	2.16E-01	9.61E-01	2.16E-01
LMF FreeChol VLDL	7.05E-02	4.64E-02	2.00E-05
LMF FreeChol IDL	1.31E-02	3.71E-01	1.52E-04
LMF FreeChol LDL	1.00E+00	7.43E-08	7.43E-08
LMF FreeChol HDL	1.95E-06	2.99E-03	9.59E-14
LMF Phosp VLDL	6.18E-02	7.95E-01	1.85E-02
LMF Phosp IDL	1.20E-02	4.01E-01	1.71E-04
LMF Phosp LDL	8.79E-01	1.49E-07	1.10E-06
LMF Phosp HDL	1.17E-05	6.19E-02	4.00E-02
LMF ApoA1 HDL	9.58E-04	9.44E-01	1.69E-03
LMF ApoA2 HDL	2.45E-04	6.18E-02	1.78E-01
LMF ApoB VLDL	5.63E-01	1.19E-03	2.15E-02
LMF ApoB IDL	1.10E-02	7.32E-01	1.48E-03
LMF ApoB LDL	9.03E-01	1.51E-08	2.02E-09
SubTrigl VLDL-1	2.25E-03	9.75E-01	2.52E-03
SubTrigl VLDL-2	2.54E-01	1.63E-01	8.66E-01
SubTrigl VLDL-3	4.32E-01	1.07E-01	4.32E-01
SubTrigl VLDL-4	2.52E-03	2.25E-03	9.75E-01
SubTrigl VLDL-5	1.68E-03	2.97E-12	5.34E-04
SubChol VLDL-1	3.40E-06	9.94E-01	3.40E-06
SubChol VLDL-2	6.14E-01	6.14E-01	2.11E-01
SubChol VLDL-3	5.36E-02	7.16E-01	9.84E-03
SubChol VLDL-4	3.71E-01	4.66E-02	3.71E-01
SubChol VLDL-5	7.15E-07	6.48E-12	1.29E-01
SubFreeChol VLDL-1	1.18E-02	3.86E-01	2.16E-01
SubFreeChol VLDL-2	2.40E-01	2.40E-01	6.13E-03
SubFreeChol VLDL-3	2.33E-02	8.66E-01	8.96E-03
SubFreeChol VLDL-4	9.68E-01	1.55E-02	1.42E-02
SubFreeChol VLDL-5	1.20E-02	1.29E-01	7.80E-06
SubPhosp VLDL-1	6.40E-05	3.01E-01	1.10E-02
SubPhosp VLDL-2	1.00E+00	1.00E+00	1.00E+00
SubPhosp VLDL-3	9.94E-01	9.94E-01	9.94E-01
SubPhosp VLDL-4	1.55E-02	6.13E-03	8.79E-01
SubPhosp VLDL-5	1.48E-05	1.12E-08	3.28E-01
SubTrigl LDL-1	6.65E-01	4.65E-05	1.24E-06
SubTrigl LDL-2	9.61E-01	3.07E-04	1.93E-04
SubTrigl LDL-3	9.90E-01	9.38E-11	1.18E-10
SubTrigl LDL-4	3.86E-01	4.37E-05	4.92E-03

SubTrigl LDL-5	1.00E+00	3.71E-02	3.71E-02
SubTrigl LDL-6	7.64E-01	2.73E-02	5.56E-03
SubChol LDL-1	2.67E-02	2.29E-03	1.32E-08
SubChol LDL-2	2.52E-03	8.36E-10	6.74E-03
SubChol LDL-3	7.95E-01	3.54E-08	7.15E-07
SubChol LDL-4	9.53E-01	1.02E-05	2.19E-05
SubChol LDL-5	5.80E-01	1.34E-04	3.71E-03
SubChol LDL-6	9.44E-01	2.52E-02	1.85E-02
SubFreeChol LDL-1	4.01E-01	2.45E-04	9.21E-07
SubFreeChol LDL-2	1.97E-02	6.75E-08	8.48E-03
SubFreeChol LDL-3	7.32E-01	1.19E-09	6.36E-08
SubFreeChol LDL-4	9.81E-01	3.96E-08	2.56E-08
SubFreeChol LDL-5	6.14E-01	4.97E-05	1.38E-03
SubFreeChol LDL-6	7.80E-01	8.62E-02	2.16E-01
SubPhosp LDL-1	8.00E-02	7.38E-04	3.01E-08
SubPhosp LDL-2	2.99E-03	1.64E-10	2.52E-03
SubPhosp LDL-3	7.64E-01	4.90E-08	1.27E-06
SubPhosp LDL-4	8.39E-01	3.38E-06	2.82E-05
SubPhosp LDL-5	5.63E-01	3.09E-04	7.76E-03
SubPhosp LDL-6	5.80E-01	2.54E-01	4.30E-02
SubApoB LDL-1	5.42E-02	3.84E-04	4.03E-09
SubApoB LDL-2	2.25E-03	5.34E-11	1.87E-03
SubApoB LDL-3	9.25E-01	2.84E-10	1.70E-09
SubApoB LDL-4	9.90E-01	4.47E-06	4.47E-06
SubApoB LDL-5	3.86E-01	9.30E-05	8.48E-03
SubApoB LDL-6	8.10E-01	8.13E-02	2.85E-02
SubTrigl HDL-1	8.21E-05	7.10E-03	4.01E-01
SubTrigl HDL-2	4.14E-03	6.62E-02	5.13E-01
SubTrigl HDL-3	2.75E-04	1.53E-03	8.25E-01
SubTrigl HDL-4	5.08E-01	8.53E-01	6.24E-01
SubChol HDL-1	1.48E-03	1.53E-01	1.98E-01
SubChol HDL-2	2.41E-01	9.98E-01	2.41E-01
SubChol HDL-3	1.83E-03	7.95E-01	9.27E-03
SubChol HDL-4	2.33E-02	5.04E-03	7.80E-01
SubFreeChol HDL-1	1.07E-06	1.44E-01	2.78E-03
SubFreeChol HDL-2	2.56E-08	4.29E-04	1.02E-01
SubFreeChol HDL-3	4.32E-01	8.53E-01	5.56E-01
SubFreeChol HDL-4	7.95E-01	1.42E-02	4.99E-02
SubPhosp HDL-1	6.15E-04	7.38E-04	9.75E-01
SubPhosp HDL-2	7.69E-04	1.44E-01	1.44E-01
SubPhosp HDL-3	1.52E-04	1.15E-01	8.13E-02
SubPhosp HDL-4	5.97E-01	2.72E-01	5.97E-01
SubApoA1 HDL-1	1.12E-03	1.12E-03	1.00E+00
SubApoA1 HDL-2	1.49E-07	1.25E-03	1.08E-01
SubApoA1 HDL-3	1.07E-06	7.32E-01	2.48E-05

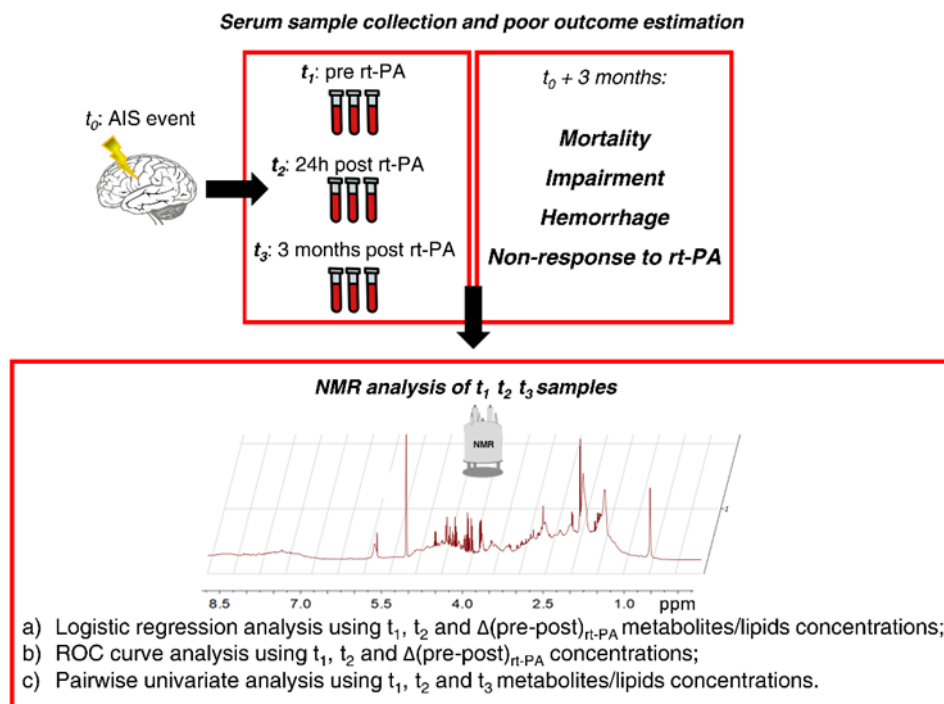
SubApoA1 HDL-4	9.03E-01	4.30E-02	7.10E-02
SubApoA2 HDL-1	9.19E-09	4.08E-10	8.25E-01
SubApoA2 HDL-2	6.99E-10	7.11E-10	9.94E-01
SubApoA2 HDL-3	1.65E-06	7.67E-06	9.03E-01
SubApoA2 HDL-4	8.91E-01	8.91E-01	8.91E-01

* To perform multiple comparison among the three-time points, 173 subjects have been used.

Main abbreviations are reported as follows: trigl: triglycerides; chol: cholesterol; phosp: phospholipids; Apo: apolipoprotein; LMF: lipoprotein main fraction; Sub: subfraction; PN: particle number; 3-HB: 3-hydroxybutyrate. Amino acids are reported with the three letters code.

Figures

Figure 1. Graphical representation of the analysis followed to identify, in serum samples, possible predictors of three-months poor outcomes of acute ischemic stroke (AIS) and to estimate metabolic variations with respect to intravenous thrombolysis with recombinant tissue plasminogen activator (rt-PA). Blood samples were collected before (t_1), 24h after (t_2) and 3-months after (t_3) the administration of rt-PA. Poor outcomes were defined at three-months as follow: mortality, impairment development, haemorrhagic transformation of the lesion, non-response to the thrombolysis. For each time-point of blood collection, three different types of 1D NMR spectra have been acquired and 1D NOESY spectra were used to estimate metabolites and lipids concentrations; t_1 , t_2 and $\Delta(t_1-t_2)$ concentrations were employed for logistic regression and ROC curve analysis; while a pairwise univariate analysis was conducted using t_1 , t_2 and t_3 concentrations of metabolic variables.



Supplementary Material**Table S1.** Effect of pre (t_1) rt-PA metabolites and lipids levels on three-months mortality, impairments, symptomatic intracerebral haemorrhage (sICHRCT12) and on the response to i.v. thrombolysis intervention, adjusting *for major determinants for unfavourable outcomes.

	Three-months mortality			Three-months impairments			SICHRCT12			Non-response to thrombolysis		
	OR (95% CI)	P	FDR	OR (95% CI)	P	FDR	OR (95% CI)	P	FDR	OR (95% CI)	P	FDR
Creatinine	1.668 (0.875-3.181)	0.12	0.709	0.833 (0.527-1.318)	0.435	0.771	0.98 (0.602-1.596)	0.935	0.951	1.243 (0.912-1.695)	0.169	0.894
Ala	1.177 (0.718-1.931)	0.517	0.795	1.27 (0.858-1.878)	0.232	0.771	1.251 (0.669-2.338)	0.483	0.951	0.926 (0.693-1.237)	0.602	0.894
Glu	1.305 (0.808-2.108)	0.277	0.795	0.811 (0.546-1.205)	0.299	0.771	1.104 (0.63-1.934)	0.728	0.951	1.103 (0.825-1.476)	0.509	0.894
Gln	1.104 (0.585-2.082)	0.76	0.805	0.866 (0.59-1.27)	0.461	0.771	0.811 (0.437-1.507)	0.508	0.951	0.997 (0.734-1.352)	0.982	0.982
Gly	1.202 (0.71-2.033)	0.493	0.795	1.051 (0.714-1.546)	0.801	0.892	1.311 (0.69-2.491)	0.408	0.951	0.932 (0.694-1.251)	0.638	0.894
His	1.228 (0.867-1.739)	0.248	0.795	1.143 (0.851-1.537)	0.375	0.771	1.4 (0.382-5.133)	0.611	0.951	0.947 (0.728-1.231)	0.682	0.894
Ile	0.992 (0.6-1.641)	0.974	0.974	0.999 (0.684-1.457)	0.995	0.995	1.044 (0.609-1.791)	0.875	0.951	0.967 (0.731-1.277)	0.811	0.973
Leu	1.157 (0.716-1.868)	0.552	0.795	1.087 (0.758-1.557)	0.65	0.78	1.08 (0.63-1.851)	0.78	0.951	0.946 (0.715-1.25)	0.695	0.894
Phe	1.107 (0.697-1.758)	0.667	0.805	1.258 (0.89-1.779)	0.194	0.771	1.05 (0.623-1.769)	0.855	0.951	0.867 (0.653-1.153)	0.327	0.894
Tyr	1.165 (0.711-1.908)	0.544	0.795	1.334 (0.919-1.936)	0.13	0.771	1.164 (0.676-2.003)	0.584	0.951	0.934 (0.699-1.246)	0.642	0.894
Val	1.732 (0.976-3.074)	0.06	0.709	1.108 (0.767-1.601)	0.585	0.78	1.433 (0.808-2.543)	0.219	0.951	0.985 (0.736-1.319)	0.92	0.982
Aceticacid	0.599 (0.1-3.58)	0.575	0.795	0.819 (0.416-1.611)	0.562	0.78	5.628 (0.402-78.742)	0.199	0.951	1.184 (0.835-1.68)	0.342	0.894
Citricacid	0.876 (0.385-1.996)	0.753	0.805	0.76 (0.396-1.458)	0.409	0.771	1.047 (0.414-2.645)	0.923	0.951	1.361 (0.843-2.196)	0.207	0.894
Lacticacid	1.525 (0.849-2.739)	0.157	0.709	1.106 (0.753-1.624)	0.608	0.78	1.297 (0.576-2.919)	0.53	0.951	1.011 (0.751-1.362)	0.941	0.982
3-HB	0.795 (0.449-1.407)	0.43	0.795	1.131 (0.81-1.579)	0.471	0.771	0.648 (0.447-0.939)	0.022	0.392	0.708 (0.519-0.967)	0.03	0.269
Acetone	0.901 (0.533-1.525)	0.698	0.805	1.17 (0.825-1.658)	0.378	0.771	0.692 (0.48-0.997)	0.048	0.434	0.66 (0.477-0.913)	0.012	0.215
Pyruvicacid	1.491 (0.906-2.454)	0.116	0.709	1.042 (0.694-1.564)	0.842	0.892	1.022 (0.514-2.029)	0.951	0.951	1.132 (0.825-1.553)	0.443	0.894
Glucose	1.356 (0.497-3.699)	0.552	0.795	1.405 (0.702-2.813)	0.337	0.771	1.296 (0.441-3.805)	0.638	0.951	0.784 (0.457-1.343)	0.375	0.894
Trigl	0.779 (0.409-1.485)	0.448	0.983	1.105 (0.754-1.621)	0.608	0.771	0.961 (0.578-1.6)	0.88	0.996	1.091 (0.822-1.448)	0.547	0.982

Chol	0.949 (0.514-1.749)	0.866	0.983		1.237 (0.835-1.833)	0.289	0.568		1.172 (0.674-2.036)	0.574	0.996		0.973 (0.727-1.303)	0.856	0.982
LDL-Chol	1.055 (0.584-1.904)	0.86	0.983		1.43 (0.972-2.105)	0.069	0.33		1.077 (0.632-1.836)	0.785	0.996		0.935 (0.7-1.249)	0.649	0.982
HDL-Chol	0.705 (0.394-1.262)	0.24	0.983		0.716 (0.484-1.059)	0.095	0.385		0.999 (0.592-1.685)	0.996	0.996		0.933 (0.696-1.252)	0.645	0.982
Apo-A1	0.657 (0.364-1.185)	0.163	0.983		0.669 (0.444-1.009)	0.055	0.32		1.078 (0.631-1.844)	0.783	0.996		1.082 (0.799-1.464)	0.611	0.982
Apo-A2	0.829 (0.455-1.508)	0.538	0.983		0.806 (0.544-1.196)	0.284	0.568		1.195 (0.699-2.041)	0.515	0.996		1.004 (0.74-1.363)	0.98	0.991
Apo-B100	0.93 (0.492-1.757)	0.822	0.983		1.459 (0.99-2.149)	0.056	0.32		1.041 (0.603-1.799)	0.885	0.996		1.018 (0.766-1.353)	0.901	0.988
LDL-HDL-Chol	1.306 (0.685-2.491)	0.417	0.983		1.689 (1.158-2.462)	0.006	0.177		1.258 (0.707-2.238)	0.435	0.996		0.991 (0.748-1.314)	0.951	0.991
Apo-B100-Apo-A1	1.255 (0.763-2.065)	0.371	0.983		1.735 (1.161-2.593)	0.007	0.177		1.027 (0.633-1.665)	0.915	0.996		0.945 (0.713-1.252)	0.693	0.982
TPN	0.93 (0.492-1.757)	0.822	0.983		1.459 (0.99-2.149)	0.056	0.32		1.041 (0.603-1.799)	0.885	0.996		1.018 (0.766-1.353)	0.901	0.988
VLDL_PN	0.805 (0.438-1.478)	0.484	0.983		1.081 (0.738-1.585)	0.689	0.827		0.936 (0.575-1.524)	0.791	0.996		1.148 (0.865-1.524)	0.338	0.982
IDL_PN	0.983 (0.55-1.757)	0.954	0.983		1.167 (0.798-1.707)	0.426	0.639		1.003 (0.6-1.676)	0.991	0.996		1.092 (0.825-1.446)	0.539	0.982
LDL_PN	1.124 (0.614-2.058)	0.705	0.983		1.446 (0.981-2.13)	0.062	0.323		1.027 (0.606-1.744)	0.92	0.996		0.953 (0.716-1.268)	0.741	0.982
LDL1_PN	0.927 (0.512-1.679)	0.803	0.983		1.013 (0.689-1.489)	0.948	0.974		0.83 (0.488-1.412)	0.493	0.996		0.959 (0.711-1.293)	0.781	0.982
LDL2_PN	0.743 (0.405-1.362)	0.336	0.983		1.171 (0.813-1.686)	0.397	0.639		0.868 (0.511-1.473)	0.6	0.996		0.925 (0.69-1.24)	0.602	0.982
LDL3_PN	1.33 (0.728-2.429)	0.353	0.983		1.381 (0.937-2.036)	0.103	0.387		0.99 (0.58-1.689)	0.969	0.996		0.869 (0.648-1.166)	0.35	0.982
LDL4_PN	1.462 (0.873-2.447)	0.148	0.983		1.122 (0.779-1.616)	0.536	0.711		1.095 (0.664-1.807)	0.722	0.996		0.997 (0.745-1.334)	0.983	0.991
LDL5_PN	1.083 (0.593-1.978)	0.795	0.983		1.275 (0.879-1.847)	0.2	0.514		1.164 (0.679-1.995)	0.58	0.996		1.076 (0.805-1.439)	0.619	0.982
LDL6_PN	0.908 (0.446-1.848)	0.79	0.983		1.466 (1.026-2.094)	0.035	0.27		1.295 (0.687-2.44)	0.424	0.996		0.971 (0.729-1.292)	0.838	0.982
LMF_Trigl_VLDL	0.815 (0.427-1.557)	0.536	0.983		1.18 (0.801-1.737)	0.403	0.639		1.013 (0.594-1.727)	0.962	0.996		1.067 (0.801-1.421)	0.656	0.982
LMF_Trigl_IDL	0.762 (0.401-1.449)	0.407	0.983		1.054 (0.723-1.538)	0.784	0.894		0.929 (0.567-1.523)	0.77	0.996		1.089 (0.822-1.442)	0.553	0.982
LMF_Trigl_LDL	1.009 (0.595-1.712)	0.973	0.983		1.167 (0.802-1.699)	0.419	0.639		0.908 (0.567-1.452)	0.686	0.996		1.092 (0.823-1.448)	0.543	0.982
LMF_Trigl_HDL	0.654 (0.357-1.2)	0.17	0.983		0.769 (0.522-1.132)	0.184	0.514		0.944 (0.575-1.55)	0.82	0.996		1.281 (0.949-1.731)	0.106	0.982
LMF_Chol_VLDL	0.813 (0.422-1.569)	0.538	0.983		1.078 (0.737-1.576)	0.699	0.831		1.134 (0.663-1.941)	0.645	0.996		1.172 (0.886-1.551)	0.267	0.982
LMF_Chol_IDL	1.023 (0.547-1.913)	0.943	0.983		1.272 (0.865-1.87)	0.221	0.514		1.151 (0.663-2)	0.617	0.996		1.083 (0.818-1.434)	0.577	0.982
LMF_Chol_LDL	1.055 (0.584-1.904)	0.86	0.983		1.43 (0.972-2.105)	0.069	0.33		1.077 (0.632-1.836)	0.785	0.996		0.935 (0.7-1.249)	0.649	0.982
LMF_Chol_HDL	0.705 (0.394-1.262)	0.24	0.983		0.716 (0.484-1.059)	0.095	0.385		0.999 (0.592-1.685)	0.996	0.996		0.933 (0.696-1.252)	0.645	0.982
LMF_FreeChol_VLDL	0.739 (0.376-1.455)	0.382	0.983		1.076 (0.731-1.584)	0.71	0.835		1.06 (0.619-1.816)	0.832	0.996		1.139 (0.857-1.514)	0.371	0.982

LMF_FreeChol_IDL	0.967 (0.521-1.796)	0.916	0.983		1.22 (0.828-1.797)	0.314	0.607		1.099 (0.638-1.893)	0.734	0.996		1.096 (0.828-1.451)	0.523	0.982
LMF_FreeChol_LDL	1.24 (0.666-2.309)	0.497	0.983		1.416 (0.953-2.103)	0.085	0.372		1.164 (0.677-2)	0.583	0.996		0.907 (0.675-1.22)	0.52	0.982
LMF_FreeChol_HDL	0.85 (0.477-1.514)	0.58	0.983		0.853 (0.558-1.305)	0.464	0.661		1.363 (0.8-2.325)	0.255	0.996		0.917 (0.679-1.24)	0.575	0.982
LMF_Phosp_VLDL	0.804 (0.426-1.516)	0.5	0.983		1.051 (0.712-1.551)	0.802	0.897		1.016 (0.596-1.73)	0.954	0.996		1.115 (0.835-1.49)	0.459	0.982
LMF_Phosp_IDL	0.914 (0.492-1.696)	0.774	0.983		1.175 (0.801-1.725)	0.41	0.639		1.056 (0.608-1.834)	0.846	0.996		1.08 (0.813-1.434)	0.595	0.982
LMF_Phosp_LDL	1.112 (0.612-2.021)	0.727	0.983		1.345 (0.911-1.986)	0.135	0.467		1.026 (0.6-1.756)	0.924	0.996		0.933 (0.697-1.249)	0.639	0.982
LMF_Phosp_HDL	0.76 (0.416-1.386)	0.37	0.983		0.55 (0.357-0.847)	0.007	0.177		0.917 (0.529-1.591)	0.759	0.996		0.998 (0.742-1.343)	0.991	0.991
LMF_ApoA1_HDL	0.573 (0.314-1.047)	0.07	0.983		0.644 (0.431-0.965)	0.033	0.27		1.01 (0.602-1.696)	0.969	0.996		1.115 (0.825-1.505)	0.48	0.982
LMF_ApoA2_HDL	0.803 (0.438-1.474)	0.48	0.983		0.802 (0.541-1.19)	0.274	0.568		1.15 (0.671-1.972)	0.612	0.996		1.005 (0.741-1.363)	0.972	0.991
LMF_ApoB_VLDL	0.805 (0.439-1.479)	0.485	0.983		1.081 (0.738-1.585)	0.689	0.827		0.936 (0.575-1.524)	0.791	0.996		1.149 (0.865-1.524)	0.338	0.982
LMF_ApoB_IDL	0.983 (0.55-1.757)	0.954	0.983		1.167 (0.798-1.707)	0.426	0.639		1.003 (0.6-1.675)	0.992	0.996		1.092 (0.825-1.446)	0.539	0.982
LMF_ApoB_LDL	1.124 (0.614-2.059)	0.705	0.983		1.446 (0.981-2.13)	0.062	0.323		1.028 (0.606-1.744)	0.92	0.996		0.953 (0.716-1.268)	0.741	0.982
SubTrigl_VLDL-1	0.799 (0.42-1.52)	0.494	0.983		1.188 (0.801-1.762)	0.391	0.639		1.006 (0.586-1.725)	0.984	0.996		1.015 (0.759-1.357)	0.92	0.99
SubTrigl_VLDL-2	1.059 (0.58-1.934)	0.853	0.983		1.266 (0.868-1.846)	0.22	0.514		1.213 (0.708-2.077)	0.483	0.996		1.071 (0.806-1.423)	0.638	0.982
SubTrigl_VLDL-3	0.972 (0.528-1.788)	0.927	0.983		1.249 (0.847-1.842)	0.262	0.568		1.165 (0.679-2)	0.579	0.996		1.124 (0.844-1.496)	0.424	0.982
SubTrigl_VLDL-4	0.865 (0.479-1.563)	0.631	0.983		1.114 (0.762-1.627)	0.578	0.757		0.975 (0.588-1.619)	0.923	0.996		1.139 (0.856-1.516)	0.371	0.982
SubTrigl_VLDL-5	0.722 (0.41-1.274)	0.261	0.983		0.782 (0.537-1.137)	0.197	0.514		0.831 (0.505-1.369)	0.468	0.996		1.069 (0.811-1.41)	0.634	0.982
SubChol_VLDL-1	0.756 (0.386-1.482)	0.416	0.983		1.088 (0.735-1.61)	0.674	0.827		1.023 (0.598-1.749)	0.935	0.996		1.098 (0.822-1.466)	0.528	0.982
SubChol_VLDL-2	0.955 (0.511-1.784)	0.886	0.983		1.062 (0.729-1.546)	0.755	0.869		1.2 (0.707-2.037)	0.499	0.996		1.195 (0.904-1.579)	0.211	0.982
SubChol_VLDL-3	0.98 (0.52-1.846)	0.949	0.983		1.203 (0.823-1.758)	0.34	0.639		1.253 (0.72-2.182)	0.425	0.996		1.174 (0.887-1.553)	0.262	0.982
SubChol_VLDL-4	0.907 (0.492-1.673)	0.755	0.983		1.045 (0.717-1.523)	0.818	0.897		1.209 (0.706-2.071)	0.489	0.996		1.201 (0.906-1.593)	0.204	0.982
SubChol_VLDL-5	0.655 (0.365-1.177)	0.157	0.983		0.584 (0.391-0.874)	0.009	0.177		0.748 (0.451-1.242)	0.262	0.996		1.158 (0.872-1.538)	0.311	0.982
SubFreeChol_VLDL-1	0.742 (0.378-1.458)	0.387	0.983		1.087 (0.735-1.607)	0.676	0.827		0.978 (0.574-1.667)	0.935	0.996		1.088 (0.814-1.456)	0.569	0.982
SubFreeChol_VLDL-2	0.888 (0.468-1.683)	0.715	0.983		1.167 (0.798-1.706)	0.425	0.639		1.079 (0.639-1.821)	0.777	0.996		1.151 (0.871-1.521)	0.323	0.982
SubFreeChol_VLDL-3	0.879 (0.465-1.661)	0.69	0.983		1.2 (0.819-1.759)	0.35	0.639		1.093 (0.639-1.871)	0.745	0.996		1.129 (0.853-1.494)	0.397	0.982
SubFreeChol_VLDL-4	0.941 (0.506-1.75)	0.848	0.983		1.089 (0.748-1.585)	0.657	0.822		1.135 (0.664-1.939)	0.643	0.996		1.156 (0.875-1.526)	0.308	0.982
SubFreeChol_VLDL-5	0.596 (0.301-1.181)	0.138	0.983		0.798 (0.527-1.209)	0.288	0.568		0.815 (0.489-1.359)	0.434	0.996		1.125 (0.836-1.515)	0.436	0.982

SubPhosp_VLDL-1	0.773 (0.4-1.492)	0.442	0.983		1.189 (0.802-1.761)	0.389	0.639		1.039 (0.606-1.781)	0.89	0.996		1.035 (0.776-1.382)	0.814	0.982
SubPhosp_VLDL-2	1.041 (0.568-1.908)	0.897	0.983		1.242 (0.852-1.81)	0.259	0.568		1.204 (0.712-2.037)	0.489	0.996		1.117 (0.843-1.48)	0.44	0.982
SubPhosp_VLDL-3	0.94 (0.512-1.728)	0.843	0.983		1.2 (0.82-1.757)	0.349	0.639		1.1 (0.65-1.861)	0.724	0.996		1.161 (0.875-1.539)	0.301	0.982
SubPhosp_VLDL-4	0.923 (0.51-1.672)	0.792	0.983		1.063 (0.73-1.549)	0.75	0.869		1.059 (0.633-1.772)	0.828	0.996		1.191 (0.898-1.58)	0.225	0.982
SubPhosp_VLDL-5	0.624 (0.344-1.129)	0.119	0.983		0.611 (0.409-0.912)	0.016	0.224		0.843 (0.508-1.398)	0.509	0.996		1.174 (0.885-1.557)	0.265	0.982
SubTrigl_LDL-1	0.877 (0.516-1.49)	0.627	0.983		1 (0.687-1.457)	1	1		0.819 (0.528-1.27)	0.372	0.996		1.08 (0.815-1.432)	0.592	0.982
SubTrigl_LDL-2	1 (0.606-1.651)	1	1		1.037 (0.709-1.517)	0.853	0.908		0.769 (0.491-1.207)	0.254	0.996		1.029 (0.773-1.371)	0.843	0.982
SubTrigl_LDL-3	1.042 (0.618-1.755)	0.878	0.983		1.001 (0.684-1.464)	0.996	1		0.81 (0.505-1.299)	0.382	0.996		0.99 (0.744-1.319)	0.947	0.991
SubTrigl_LDL-4	1.325 (0.787-2.231)	0.289	0.983		1.052 (0.72-1.537)	0.792	0.894		0.956 (0.585-1.563)	0.857	0.996		1.111 (0.828-1.491)	0.481	0.982
SubTrigl_LDL-5	1.323 (0.756-2.314)	0.327	0.983		1.101 (0.763-1.589)	0.608	0.771		1.075 (0.645-1.793)	0.782	0.996		1.111 (0.839-1.472)	0.462	0.982
SubTrigl_LDL-6	1.235 (0.687-2.219)	0.48	0.983		1.527 (1.072-2.176)	0.019	0.224		1.35 (0.728-2.502)	0.341	0.996		0.886 (0.671-1.171)	0.396	0.982
SubChol_LDL-1	0.92 (0.504-1.679)	0.786	0.983		0.994 (0.679-1.456)	0.975	0.993		0.868 (0.501-1.502)	0.612	0.996		0.952 (0.705-1.285)	0.746	0.982
SubChol_LDL-2	0.739 (0.404-1.35)	0.325	0.983		1.182 (0.822-1.699)	0.366	0.639		0.897 (0.53-1.517)	0.685	0.996		0.9 (0.671-1.206)	0.479	0.982
SubChol_LDL-3	1.274 (0.712-2.279)	0.415	0.983		1.37 (0.936-2.005)	0.105	0.387		1.057 (0.63-1.773)	0.835	0.996		0.851 (0.635-1.14)	0.279	0.982
SubChol_LDL-4	1.473 (0.886-2.449)	0.135	0.983		1.136 (0.791-1.632)	0.489	0.687		1.161 (0.69-1.954)	0.574	0.996		0.956 (0.714-1.281)	0.763	0.982
SubChol_LDL-5	1.053 (0.576-1.924)	0.867	0.983		1.266 (0.871-1.839)	0.217	0.514		1.146 (0.672-1.953)	0.617	0.996		1.056 (0.789-1.414)	0.715	0.982
SubChol_LDL-6	0.93 (0.467-1.851)	0.836	0.983		1.517 (1.058-2.176)	0.024	0.224		1.288 (0.693-2.395)	0.424	0.996		0.934 (0.7-1.245)	0.641	0.982
SubFreeChol_LDL-1	0.915 (0.497-1.683)	0.774	0.983		1.04 (0.709-1.524)	0.842	0.908		0.964 (0.552-1.685)	0.898	0.996		0.964 (0.713-1.304)	0.813	0.982
SubFreeChol_LDL-2	0.655 (0.351-1.224)	0.185	0.983		1.181 (0.817-1.706)	0.377	0.639		0.985 (0.583-1.665)	0.956	0.996		0.931 (0.691-1.255)	0.639	0.982
SubFreeChol_LDL-3	1.171 (0.646-2.123)	0.602	0.983		1.298 (0.88-1.914)	0.188	0.514		1.013 (0.588-1.746)	0.963	0.996		0.865 (0.643-1.164)	0.339	0.982
SubFreeChol_LDL-4	1.348 (0.754-2.41)	0.314	0.983		1.255 (0.859-1.834)	0.239	0.546		1.077 (0.649-1.788)	0.773	0.996		0.926 (0.688-1.245)	0.609	0.982
SubFreeChol_LDL-5	1.118 (0.604-2.066)	0.723	0.983		1.32 (0.906-1.923)	0.149	0.485		1.164 (0.679-1.995)	0.581	0.996		1.02 (0.761-1.367)	0.896	0.988
SubFreeChol_LDL-6	1.01 (0.537-1.903)	0.974	0.983		1.64 (1.13-2.382)	0.009	0.177		1.45 (0.813-2.588)	0.208	0.996		0.919 (0.693-1.218)	0.555	0.982
SubPhosp_LDL-1	0.895 (0.49-1.632)	0.717	0.983		0.954 (0.649-1.403)	0.812	0.897		0.828 (0.48-1.428)	0.497	0.996		0.964 (0.713-1.303)	0.811	0.982
SubPhosp_LDL-2	0.763 (0.417-1.393)	0.378	0.983		1.136 (0.789-1.635)	0.494	0.687		0.864 (0.508-1.47)	0.589	0.996		0.913 (0.68-1.224)	0.542	0.982
SubPhosp_LDL-3	1.284 (0.712-2.314)	0.406	0.983		1.333 (0.908-1.958)	0.143	0.478		1.017 (0.601-1.72)	0.95	0.996		0.856 (0.638-1.149)	0.301	0.982
SubPhosp_LDL-4	1.489 (0.897-2.472)	0.124	0.983		1.151 (0.799-1.659)	0.449	0.657		1.113 (0.673-1.841)	0.676	0.996		0.945 (0.704-1.269)	0.707	0.982

SubPhosp_LDL-5	1.068 (0.589-1.936)	0.829	0.983		1.227 (0.844-1.783)	0.284	0.568		1.133 (0.667-1.924)	0.645	0.996		1.066 (0.796-1.426)	0.669	0.982
SubPhosp_LDL-6	0.927 (0.468-1.835)	0.827	0.983		1.436 (1.005-2.051)	0.047	0.315		1.27 (0.686-2.354)	0.447	0.996		0.931 (0.699-1.239)	0.624	0.982
SubApoB_LDL-1	0.927 (0.512-1.68)	0.803	0.983		1.013 (0.689-1.489)	0.948	0.974		0.83 (0.488-1.412)	0.492	0.996		0.959 (0.711-1.293)	0.782	0.982
SubApoB_LDL-2	0.742 (0.405-1.362)	0.336	0.983		1.171 (0.813-1.687)	0.397	0.639		0.868 (0.511-1.473)	0.599	0.996		0.925 (0.69-1.241)	0.603	0.982
SubApoB_LDL-3	1.33 (0.728-2.429)	0.353	0.983		1.382 (0.937-2.036)	0.103	0.387		0.99 (0.58-1.69)	0.97	0.996		0.869 (0.648-1.166)	0.35	0.982
SubApoB_LDL-4	1.462 (0.874-2.448)	0.148	0.983		1.122 (0.779-1.616)	0.537	0.711		1.095 (0.664-1.807)	0.722	0.996		0.997 (0.745-1.334)	0.983	0.991
SubApoB_LDL-5	1.083 (0.593-1.978)	0.795	0.983		1.274 (0.879-1.847)	0.2	0.514		1.164 (0.679-1.995)	0.58	0.996		1.076 (0.805-1.439)	0.619	0.982
SubApoB_LDL-5	0.908 (0.446-1.848)	0.789	0.983		1.466 (1.026-2.094)	0.035	0.27		1.295 (0.687-2.44)	0.424	0.996		0.971 (0.729-1.292)	0.837	0.982
SubTrigl_HDL-1	0.753 (0.415-1.367)	0.351	0.983		0.86 (0.584-1.265)	0.443	0.655		0.934 (0.563-1.55)	0.791	0.996		1.137 (0.846-1.528)	0.394	0.982
SubTrigl_HDL-2	0.684 (0.383-1.22)	0.198	0.983		0.844 (0.576-1.238)	0.387	0.639		0.993 (0.603-1.635)	0.979	0.996		1.19 (0.883-1.603)	0.252	0.982
SubTrigl_HDL-3	0.548 (0.302-0.998)	0.049	0.983		0.786 (0.54-1.144)	0.208	0.514		0.924 (0.576-1.484)	0.745	0.996		1.266 (0.943-1.7)	0.117	0.982
SubTrigl_HDL-4	0.555 (0.288-1.069)	0.078	0.983		0.725 (0.478-1.101)	0.131	0.467		0.949 (0.562-1.603)	0.845	0.996		1.318 (0.978-1.775)	0.069	0.982
SubChol_HDL-1	0.806 (0.431-1.507)	0.5	0.983		0.878 (0.589-1.31)	0.525	0.711		0.984 (0.562-1.725)	0.956	0.996		0.91 (0.676-1.224)	0.532	0.982
SubChol_HDL-2	0.738 (0.412-1.322)	0.307	0.983		0.694 (0.465-1.036)	0.074	0.337		0.894 (0.525-1.524)	0.682	0.996		0.965 (0.717-1.299)	0.816	0.982
SubChol_HDL-3	0.746 (0.407-1.365)	0.342	0.983		0.656 (0.434-0.992)	0.046	0.315		0.954 (0.572-1.59)	0.856	0.996		0.932 (0.69-1.259)	0.648	0.982
SubChol_HDL-4	0.938 (0.511-1.72)	0.836	0.983		0.898 (0.609-1.324)	0.586	0.76		1.134 (0.687-1.87)	0.624	0.996		0.919 (0.68-1.241)	0.581	0.982
SubFreeChol_HDL-1	1.015 (0.569-1.809)	0.961	0.983		0.968 (0.645-1.453)	0.877	0.926		1.3 (0.737-2.293)	0.365	0.996		0.906 (0.676-1.214)	0.507	0.982
SubFreeChol_HDL-2	1.282 (0.69-2.38)	0.432	0.983		0.854 (0.563-1.296)	0.459	0.661		1.361 (0.774-2.392)	0.285	0.996		0.912 (0.674-1.234)	0.551	0.982
SubFreeChol_HDL-3	0.971 (0.508-1.856)	0.929	0.983		0.866 (0.569-1.319)	0.503	0.691		1.337 (0.763-2.345)	0.311	0.996		0.938 (0.684-1.285)	0.689	0.982
SubFreeChol_HDL-4	1.145 (0.629-2.084)	0.659	0.983		1.02 (0.684-1.521)	0.923	0.965		1.257 (0.731-2.159)	0.408	0.996		0.926 (0.678-1.264)	0.626	0.982
SubPhosp_HDL-1	0.828 (0.442-1.553)	0.557	0.983		0.768 (0.506-1.164)	0.213	0.514		0.957 (0.543-1.685)	0.879	0.996		0.964 (0.716-1.298)	0.81	0.982
SubPhosp_HDL-2	0.754 (0.413-1.376)	0.357	0.983		0.598 (0.39-0.916)	0.018	0.224		0.834 (0.481-1.448)	0.52	0.996		1.013 (0.753-1.361)	0.934	0.991
SubPhosp_HDL-3	0.721 (0.387-1.343)	0.303	0.983		0.514 (0.327-0.805)	0.004	0.177		0.835 (0.498-1.4)	0.494	0.996		1.015 (0.758-1.359)	0.921	0.99
SubPhosp_HDL-4	0.956 (0.516-1.771)	0.885	0.983		0.75 (0.501-1.125)	0.164	0.507		1.055 (0.645-1.727)	0.831	0.996		0.974 (0.723-1.311)	0.862	0.982
SubApoA1_HDL-1	0.664 (0.346-1.273)	0.218	0.983		0.743 (0.489-1.128)	0.163	0.507		0.975 (0.567-1.675)	0.926	0.996		1.049 (0.778-1.413)	0.754	0.982
SubApoA1_HDL-2	0.714 (0.397-1.286)	0.262	0.983		0.612 (0.401-0.933)	0.023	0.224		0.873 (0.515-1.482)	0.615	0.996		1.05 (0.787-1.402)	0.739	0.982
SubApoA1_HDL-3	0.626 (0.333-1.178)	0.147	0.983		0.602 (0.39-0.929)	0.022	0.224		0.906 (0.545-1.508)	0.705	0.996		1.038 (0.772-1.396)	0.803	0.982

SubApoA1_HDL-4	0.761 (0.41-1.413)	0.387	0.983		0.846 (0.565-1.267)	0.416	0.639		1.187 (0.701-2.01)	0.523	0.996		1.021 (0.752-1.386)	0.896	0.988
SubApoA2_HDL-1	0.919 (0.514-1.641)	0.774	0.983		0.77 (0.516-1.149)	0.201	0.514		0.975 (0.564-1.683)	0.926	0.996		1.033 (0.778-1.372)	0.822	0.982
SubApoA2_HDL-2	0.92 (0.53-1.594)	0.765	0.983		0.808 (0.549-1.19)	0.281	0.568		0.885 (0.529-1.479)	0.641	0.996		1.036 (0.78-1.377)	0.807	0.982
SubApoA2_HDL-3	0.865 (0.477-1.567)	0.632	0.983		0.755 (0.502-1.134)	0.176	0.514		0.96 (0.568-1.622)	0.878	0.996		1.029 (0.766-1.381)	0.852	0.982
SubApoA2_HDL-4	1.043 (0.555-1.962)	0.896	0.983		0.962 (0.649-1.426)	0.847	0.908		1.251 (0.738-2.119)	0.406	0.996		0.943 (0.693-1.285)	0.712	0.982

*Logistic regression analyses adjustment for age, sex, time onset-to-treatment, pre rt-PA blood glucose level, baseline NIHSS, history of atrial fibrillation, congestive heart failure, recent infections or inflammations, hypertension, diabetes, hyperlipidaemia, smoke and blood collection center. False Discovery Rate correction was applied to *P* values (*P*) using the Benjamini & Hochberg method (FDR).

Main abbreviations: CI: confidence interval; NIHSS: National Institutes of Health Stroke Scale; OR: odds ratio; trigl: triglycerides; chol: cholesterol; phosp: phospholipids; Apo: apolipoprotein; LMF: lipoprotein main fraction; PN: particle number; 3-HB: 3-hydroxybutyrate. Amino acids are reported with the three letters code.

Table S2. Effect of 24h post (t_2) rt-PA metabolites and lipids levels on three-months mortality, impairments, symptomatic intracerebral haemorrhage (sICHRCT12) and on the response to i.v. thrombolysis intervention, adjusting *for major determinants for unfavourable outcomes.

	Three-months mortality			Three-months impairments			SICHRCT12			Non-response to thrombolysis		
	OR (95% CI)	P	FDR	OR (95% CI)	P	FDR	OR (95% CI)	P	FDR	OR (95% CI)	P	FDR
Creatinine	1.429 (0.778-2.624)	0.25	0.643	1.199 (0.818-1.757)	0.352	0.704	0.668 (0.426-1.046)	0.078	0.543	0.789 (0.577-1.08)	0.138	0.447
Ala	1.252 (0.733-2.138)	0.411	0.673	1.073 (0.736-1.563)	0.715	0.74	1.343 (0.824-2.189)	0.237	0.707	1.221 (0.91-1.64)	0.184	0.447
Glu	1.33 (0.72-2.457)	0.363	0.673	1.071 (0.714-1.606)	0.74	0.74	1.267 (0.712-2.256)	0.421	0.768	0.775 (0.56-1.075)	0.127	0.447
Gln	1.125 (0.648-1.951)	0.676	0.8	0.676 (0.445-1.027)	0.066	0.397	0.931 (0.54-1.605)	0.796	0.895	1.188 (0.868-1.625)	0.283	0.463
Gly	1.201 (0.655-2.202)	0.554	0.8	0.921 (0.629-1.35)	0.675	0.74	1.11 (0.618-1.996)	0.727	0.872	0.978 (0.724-1.321)	0.883	0.883
His	1.258 (0.798-1.985)	0.323	0.673	1.141 (0.836-1.558)	0.406	0.732	1.645 (0.427-6.341)	0.47	0.768	0.856 (0.626-1.169)	0.328	0.492
Ile	0.635 (0.326-1.235)	0.181	0.542	0.927 (0.637-1.348)	0.692	0.74	1.315 (0.771-2.242)	0.314	0.707	0.91 (0.689-1.202)	0.505	0.568
Leu	1.134 (0.62-2.073)	0.684	0.8	0.89 (0.602-1.316)	0.559	0.74	0.867 (0.528-1.423)	0.571	0.857	0.844 (0.634-1.125)	0.248	0.447
Phe	1.483 (0.884-2.487)	0.136	0.542	1.276 (0.882-1.845)	0.195	0.586	0.711 (0.444-1.137)	0.154	0.556	0.7 (0.516-0.949)	0.022	0.13
Tyr	1.082 (0.659-1.777)	0.756	0.8	1.355 (0.924-1.986)	0.12	0.431	0.98 (0.639-1.504)	0.927	0.948	0.891 (0.671-1.185)	0.429	0.514
Val	1.447 (0.869-2.411)	0.155	0.542	0.93 (0.649-1.333)	0.692	0.74	1.106 (0.687-1.781)	0.678	0.872	0.886 (0.66-1.19)	0.42	0.514

Aceticacid	0.807 (0.487-1.335)	0.403	0.673		1.77 (1.164-2.692)	0.008	0.137		1.238 (0.718-2.136)	0.442	0.768		1.201 (0.884-1.633)	0.242	0.447
Citricacid	1.476 (0.914-2.381)	0.111	0.542		1.275 (0.844-1.926)	0.249	0.64		1.016 (0.625-1.651)	0.948	0.948		0.901 (0.645-1.259)	0.542	0.574
Lacticacid	1.602 (0.996-2.578)	0.052	0.542		1.359 (0.937-1.971)	0.106	0.431		0.908 (0.568-1.452)	0.687	0.872		0.829 (0.622-1.105)	0.202	0.447
3-HB	1.082 (0.681-1.721)	0.739	0.8		1.473 (1.009-2.152)	0.045	0.397		0.67 (0.434-1.033)	0.07	0.543		0.541 (0.382-0.766)	0.001	0.009
Acetone	1.02 (0.686-1.518)	0.921	0.921		1.201 (0.82-1.759)	0.348	0.704		0.811 (0.556-1.184)	0.279	0.707		0.577 (0.397-0.84)	0.004	0.037
Pyruvicacid	1.545 (0.906-2.635)	0.11	0.542		0.866 (0.595-1.259)	0.451	0.739		0.679 (0.42-1.1)	0.116	0.543		0.816 (0.614-1.085)	0.161	0.447
Glucose	0.832 (0.387-1.788)	0.637	0.8		1.087 (0.666-1.774)	0.738	0.74		0.596 (0.31-1.146)	0.121	0.543		0.832 (0.558-1.239)	0.364	0.504
Trigl	1.292 (0.577-2.89)	0.533	0.994		1.413 (0.921-2.168)	0.113	0.297		2.605 (1.094-6.201)	0.03	0.204		1.179 (0.858-1.618)	0.31	0.637
Chol	1.192 (0.606-2.347)	0.611	0.994		1.434 (0.943-2.182)	0.092	0.275		1.264 (0.733-2.177)	0.399	0.785		0.892 (0.657-1.21)	0.461	0.695
LDL-Chol	1.117 (0.588-2.122)	0.734	0.994		1.536 (1.009-2.34)	0.045	0.225		0.94 (0.55-1.604)	0.82	0.934		0.829 (0.611-1.124)	0.228	0.637
HDL-Chol	1.028 (0.552-1.913)	0.931	0.994		0.818 (0.554-1.207)	0.311	0.507		0.919 (0.562-1.504)	0.738	0.924		0.848 (0.626-1.149)	0.287	0.637
Apo-A1	0.673 (0.359-1.261)	0.216	0.994		0.742 (0.487-1.13)	0.165	0.383		1.226 (0.725-2.073)	0.448	0.837		1.036 (0.752-1.426)	0.829	0.974
Apo-A2	0.909 (0.486-1.701)	0.766	0.994		0.993 (0.658-1.5)	0.975	0.983		1.178 (0.697-1.991)	0.54	0.862		0.988 (0.717-1.361)	0.941	0.983
Apo-B100	1.165 (0.62-2.188)	0.635	0.994		1.691 (1.109-2.576)	0.015	0.138		1.142 (0.672-1.941)	0.624	0.868		0.882 (0.656-1.185)	0.405	0.66
LDL-HDL-Chol	1.066 (0.52-2.186)	0.862	0.994		1.655 (1.124-2.436)	0.011	0.122		1.176 (0.687-2.013)	0.553	0.862		0.974 (0.729-1.301)	0.859	0.974
Apo-B100-Apo-A1	1.323 (0.777-2.253)	0.302	0.994		1.869 (1.204-2.902)	0.005	0.067		0.933 (0.585-1.486)	0.769	0.934		0.894 (0.667-1.198)	0.453	0.695
TPN	1.165 (0.62-2.188)	0.635	0.994		1.691 (1.109-2.576)	0.015	0.138		1.142 (0.671-1.941)	0.625	0.868		0.882 (0.656-1.185)	0.405	0.66
VLDL_PN	1.246 (0.617-2.513)	0.539	0.994		1.231 (0.804-1.884)	0.339	0.533		1.833 (0.95-3.536)	0.071	0.316		1.262 (0.926-1.719)	0.14	0.637
IDL_PN	1.084 (0.606-1.938)	0.786	0.994		1.467 (0.973-2.21)	0.067	0.225		1.345 (0.78-2.319)	0.287	0.666		1.018 (0.762-1.36)	0.905	0.974
LDL_PN	1.16 (0.627-2.147)	0.636	0.994		1.606 (1.059-2.433)	0.026	0.194		0.976 (0.58-1.642)	0.926	0.99		0.835 (0.62-1.123)	0.232	0.637
LDL1_PN	0.99 (0.551-1.779)	0.973	0.994		1.131 (0.753-1.701)	0.553	0.67		0.846 (0.511-1.401)	0.516	0.853		0.846 (0.617-1.162)	0.302	0.637
LDL2_PN	0.706 (0.349-1.431)	0.335	0.994		1.186 (0.788-1.784)	0.413	0.565		0.661 (0.37-1.181)	0.162	0.474		0.845 (0.614-1.164)	0.303	0.637
LDL3_PN	0.976 (0.489-1.946)	0.944	0.994		1.515 (1.002-2.292)	0.049	0.225		0.587 (0.326-1.055)	0.075	0.316		0.763 (0.562-1.037)	0.084	0.637
LDL4_PN	1.197 (0.596-2.405)	0.613	0.994		1.186 (0.803-1.753)	0.391	0.564		0.975 (0.596-1.594)	0.919	0.99		0.961 (0.712-1.296)	0.792	0.974
LDL5_PN	1.02 (0.519-2.007)	0.953	0.994		1.329 (0.878-2.01)	0.178	0.385		1.295 (0.748-2.239)	0.356	0.725		0.997 (0.733-1.355)	0.983	0.992
LDL6_PN	1.991 (0.997-3.974)	0.051	0.726		1.925 (1.271-2.914)	0.002	0.052		1.641 (0.858-3.139)	0.134	0.413		0.888 (0.657-1.202)	0.443	0.691
LMF_Trigl_VLDL	1.701 (0.76-3.81)	0.196	0.994		1.56 (1.012-2.406)	0.044	0.225		3.076 (1.214-7.79)	0.018	0.145		1.182 (0.855-1.634)	0.312	0.637

LMF_Trigl_IDL	1.062 (0.445-2.535)	0.892	0.994		1.227 (0.796-1.89)	0.354	0.538		2.515 (1.08-5.854)	0.032	0.205		1.216 (0.88-1.679)	0.235	0.637
LMF_Trigl_LDL	0.936 (0.563-1.555)	0.798	0.994		1.342 (0.883-2.04)	0.169	0.385		1.009 (0.641-1.588)	0.97	0.99		0.938 (0.706-1.248)	0.662	0.956
LMF_Trigl_HDL	0.457 (0.223-0.938)	0.033	0.726		0.779 (0.524-1.158)	0.218	0.443		1.534 (0.902-2.608)	0.114	0.394		1.353 (0.983-1.863)	0.064	0.637
LMF_Chol_VLDL	1.077 (0.52-2.234)	0.841	0.994		1.184 (0.788-1.781)	0.416	0.565		2.641 (1.304-5.349)	0.007	0.119		1.274 (0.936-1.734)	0.124	0.637
LMF_Chol_IDL	1.077 (0.582-1.994)	0.812	0.994		1.582 (1.052-2.38)	0.028	0.194		1.576 (0.897-2.771)	0.114	0.394		1.021 (0.762-1.37)	0.888	0.974
LMF_Chol_LDL	1.117 (0.588-2.122)	0.734	0.994		1.536 (1.009-2.34)	0.045	0.225		0.94 (0.55-1.604)	0.82	0.934		0.829 (0.611-1.124)	0.228	0.637
LMF_Chol_HDL	1.028 (0.552-1.913)	0.931	0.994		0.818 (0.554-1.207)	0.311	0.507		0.919 (0.562-1.504)	0.738	0.924		0.848 (0.626-1.149)	0.287	0.637
LMF_FreeChol_VLDL	1.442 (0.653-3.184)	0.365	0.994		1.279 (0.837-1.955)	0.254	0.453		2.869 (1.329-6.192)	0.007	0.119		1.255 (0.913-1.725)	0.162	0.637
LMF_FreeChol_IDL	0.99 (0.527-1.858)	0.974	0.994		1.481 (0.984-2.229)	0.06	0.225		1.545 (0.88-2.713)	0.13	0.413		1.052 (0.783-1.412)	0.738	0.974
LMF_FreeChol_LDL	1.202 (0.613-2.357)	0.593	0.994		1.618 (1.06-2.471)	0.026	0.194		0.919 (0.533-1.587)	0.763	0.934		0.785 (0.577-1.068)	0.124	0.637
LMF_FreeChol_HDL	1.204 (0.635-2.284)	0.569	0.994		1.038 (0.699-1.54)	0.854	0.918		1.121 (0.675-1.862)	0.66	0.891		0.838 (0.621-1.132)	0.25	0.637
LMF_Phosp_VLDL	1.682 (0.745-3.797)	0.211	0.994		1.237 (0.798-1.916)	0.341	0.533		3.15 (1.381-7.188)	0.006	0.119		1.326 (0.955-1.841)	0.092	0.637
LMF_Phosp_IDL	1.04 (0.514-2.106)	0.912	0.994		1.386 (0.916-2.098)	0.123	0.311		1.704 (0.925-3.138)	0.087	0.332		1.065 (0.787-1.44)	0.683	0.964
LMF_Phosp_LDL	1.079 (0.573-2.033)	0.814	0.994		1.426 (0.941-2.159)	0.094	0.275		0.888 (0.519-1.519)	0.664	0.891		0.824 (0.607-1.118)	0.213	0.637
LMF_Phosp_HDL	0.71 (0.372-1.357)	0.301	0.994		0.635 (0.42-0.96)	0.031	0.197		0.887 (0.524-1.504)	0.657	0.891		0.887 (0.645-1.221)	0.463	0.695
LMF_ApoA1_HDL	0.628 (0.33-1.198)	0.158	0.994		0.692 (0.456-1.05)	0.084	0.258		1.166 (0.696-1.953)	0.56	0.862		1.021 (0.743-1.404)	0.898	0.974
LMF_ApoA2_HDL	0.937 (0.491-1.789)	0.844	0.994		1.001 (0.661-1.517)	0.996	0.996		1.151 (0.678-1.955)	0.603	0.868		0.983 (0.714-1.353)	0.915	0.975
LMF_ApoB_VLDL	1.246 (0.618-2.514)	0.539	0.994		1.231 (0.804-1.884)	0.338	0.533		1.833 (0.95-3.536)	0.071	0.316		1.262 (0.926-1.72)	0.14	0.637
LMF_ApoB_IDL	1.084 (0.606-1.938)	0.786	0.994		1.467 (0.974-2.211)	0.067	0.225		1.345 (0.78-2.32)	0.286	0.666		1.018 (0.762-1.36)	0.906	0.974
LMF_ApoB_LDL	1.16 (0.627-2.147)	0.636	0.994		1.606 (1.059-2.434)	0.026	0.194		0.976 (0.58-1.642)	0.926	0.99		0.835 (0.62-1.123)	0.232	0.637
SubTrigl_VLDL-1	1.789 (0.778-4.112)	0.171	0.994		1.603 (1.05-2.447)	0.029	0.194		2.931 (0.999-8.603)	0.05	0.287		1.095 (0.791-1.516)	0.584	0.854
SubTrigl_VLDL-2	1.673 (0.772-3.625)	0.192	0.994		1.313 (0.854-2.017)	0.215	0.443		3.216 (1.435-7.208)	0.005	0.119		1.259 (0.913-1.736)	0.161	0.637
SubTrigl_VLDL-3	1.551 (0.731-3.292)	0.253	0.994		1.367 (0.886-2.108)	0.158	0.374		2.601 (1.265-5.349)	0.009	0.119		1.3 (0.947-1.786)	0.105	0.637
SubTrigl_VLDL-4	1.475 (0.728-2.991)	0.281	0.994		1.281 (0.84-1.954)	0.25	0.452		1.71 (0.934-3.132)	0.082	0.324		1.236 (0.912-1.677)	0.172	0.637
SubTrigl_VLDL-5	1.387 (0.692-2.777)	0.356	0.994		0.815 (0.553-1.201)	0.302	0.506		1.154 (0.686-1.941)	0.589	0.868		1.135 (0.844-1.526)	0.402	0.66
SubChol_VLDL-1	1.217 (0.522-2.835)	0.649	0.994		1.405 (0.925-2.137)	0.111	0.297		3.313 (1.232-8.908)	0.018	0.145		1.152 (0.836-1.587)	0.386	0.66
SubChol_VLDL-2	1.019 (0.486-2.134)	0.96	0.994		1.118 (0.747-1.672)	0.588	0.705		3.178 (1.463-6.906)	0.003	0.119		1.278 (0.942-1.732)	0.115	0.637

SubChol_VLDL-3	1.149 (0.575-2.296)	0.693	0.994		1.395 (0.93-2.093)	0.108	0.297		2.46 (1.249-4.847)	0.009	0.119		1.207 (0.893-1.633)	0.221	0.637
SubChol_VLDL-4	0.996 (0.532-1.866)	0.99	0.994		1.09 (0.736-1.614)	0.667	0.775		1.968 (1.096-3.533)	0.023	0.178		1.237 (0.92-1.663)	0.159	0.637
SubChol_VLDL-5	0.73 (0.368-1.45)	0.369	0.994		0.459 (0.297-0.711)	0	0.052		0.99 (0.581-1.685)	0.969	0.99		1.248 (0.924-1.685)	0.148	0.637
SubFreeChol_VLDL-1	1.072 (0.442-2.601)	0.877	0.994		1.338 (0.875-2.047)	0.179	0.385		3.078 (1.14-8.309)	0.027	0.189		1.193 (0.869-1.636)	0.275	0.637
SubFreeChol_VLDL-2	1.298 (0.616-2.732)	0.493	0.994		1.375 (0.906-2.085)	0.134	0.333		2.642 (1.247-5.599)	0.011	0.128		1.265 (0.928-1.724)	0.138	0.637
SubFreeChol_VLDL-3	1.327 (0.634-2.779)	0.453	0.994		1.511 (0.995-2.294)	0.053	0.225		2.531 (1.209-5.295)	0.014	0.142		1.203 (0.884-1.636)	0.24	0.637
SubFreeChol_VLDL-4	1.059 (0.574-1.952)	0.855	0.994		1.235 (0.833-1.833)	0.294	0.5		1.675 (0.962-2.916)	0.068	0.316		1.162 (0.866-1.56)	0.317	0.637
SubFreeChol_VLDL-5	1.41 (0.655-3.035)	0.38	0.994		0.963 (0.611-1.517)	0.87	0.918		1.538 (0.763-3.102)	0.229	0.579		1.173 (0.848-1.625)	0.335	0.637
SubPhosp_VLDL-1	1.408 (0.603-3.289)	0.429	0.994		1.517 (0.985-2.336)	0.058	0.225		3.591 (1.279-10.085)	0.015	0.145		1.156 (0.835-1.599)	0.383	0.66
SubPhosp_VLDL-2	1.376 (0.638-2.969)	0.416	0.994		1.287 (0.847-1.956)	0.237	0.444		3.274 (1.506-7.119)	0.003	0.119		1.306 (0.95-1.794)	0.1	0.637
SubPhosp_VLDL-3	1.31 (0.635-2.703)	0.465	0.994		1.361 (0.893-2.075)	0.152	0.369		2.549 (1.271-5.113)	0.008	0.119		1.324 (0.969-1.808)	0.078	0.637
SubPhosp_VLDL-4	1.128 (0.586-2.171)	0.717	0.994		1.106 (0.739-1.655)	0.624	0.734		1.753 (0.985-3.118)	0.056	0.306		1.249 (0.927-1.683)	0.144	0.637
SubPhosp_VLDL-5	0.812 (0.422-1.562)	0.533	0.994		0.552 (0.365-0.834)	0.005	0.067		1.32 (0.782-2.229)	0.298	0.666		1.356 (1.002-1.836)	0.048	0.637
SubTrigl_LDL-1	0.914 (0.537-1.555)	0.74	0.994		1.174 (0.778-1.773)	0.445	0.59		0.95 (0.616-1.465)	0.815	0.934		0.973 (0.727-1.304)	0.857	0.974
SubTrigl_LDL-2	0.998 (0.623-1.6)	0.994	0.994		1.136 (0.757-1.705)	0.538	0.67		0.862 (0.571-1.301)	0.479	0.841		0.948 (0.703-1.278)	0.726	0.973
SubTrigl_LDL-3	1.058 (0.66-1.696)	0.815	0.994		1.069 (0.734-1.556)	0.729	0.817		0.831 (0.544-1.27)	0.392	0.784		0.872 (0.649-1.173)	0.366	0.66
SubTrigl_LDL-4	0.986 (0.593-1.64)	0.958	0.994		1.07 (0.726-1.578)	0.731	0.817		0.97 (0.617-1.524)	0.894	0.99		0.994 (0.745-1.326)	0.967	0.992
SubTrigl_LDL-5	1.023 (0.586-1.786)	0.936	0.994		1.182 (0.807-1.731)	0.391	0.564		1.291 (0.761-2.187)	0.343	0.725		1.029 (0.775-1.367)	0.843	0.974
SubTrigl_LDL-6	1.972 (1.192-3.262)	0.008	0.726		1.888 (1.248-2.856)	0.003	0.052		1.577 (0.791-3.145)	0.196	0.507		0.759 (0.552-1.042)	0.088	0.637
SubChol_LDL-1	0.927 (0.485-1.773)	0.818	0.994		1.052 (0.698-1.585)	0.809	0.887		0.851 (0.49-1.478)	0.567	0.862		0.87 (0.632-1.198)	0.394	0.66
SubChol_LDL-2	0.769 (0.379-1.563)	0.468	0.994		1.19 (0.795-1.783)	0.398	0.565		0.673 (0.381-1.19)	0.173	0.482		0.832 (0.606-1.143)	0.256	0.637
SubChol_LDL-3	1.056 (0.53-2.102)	0.878	0.994		1.468 (0.975-2.208)	0.066	0.225		0.616 (0.35-1.086)	0.094	0.345		0.787 (0.582-1.065)	0.12	0.637
SubChol_LDL-4	1.199 (0.598-2.404)	0.61	0.994		1.161 (0.784-1.722)	0.456	0.598		0.942 (0.58-1.532)	0.811	0.934		0.956 (0.706-1.295)	0.774	0.974
SubChol_LDL-5	1.061 (0.53-2.125)	0.867	0.994		1.293 (0.848-1.972)	0.233	0.444		1.244 (0.712-2.171)	0.443	0.837		1.01 (0.738-1.383)	0.949	0.983
SubChol_LDL-6	2.004 (1.048-3.834)	0.036	0.726		1.867 (1.241-2.81)	0.003	0.052		1.541 (0.835-2.842)	0.166	0.474		0.877 (0.647-1.188)	0.398	0.66
SubFreeChol_LDL-1	0.903 (0.478-1.706)	0.754	0.994		1.147 (0.761-1.727)	0.513	0.649		0.933 (0.538-1.618)	0.805	0.934		0.864 (0.628-1.189)	0.37	0.66
SubFreeChol_LDL-2	0.854 (0.435-1.677)	0.646	0.994		1.288 (0.859-1.93)	0.221	0.443		0.805 (0.473-1.372)	0.426	0.822		0.84 (0.612-1.153)	0.281	0.637

SubFreeChol_LDL-3	1.115 (0.558-2.227)	0.758	0.994		1.493 (0.984-2.265)	0.06	0.225		0.572 (0.313-1.047)	0.07	0.316		0.736 (0.54-1.003)	0.052	0.637
SubFreeChol_LDL-4	1.108 (0.557-2.203)	0.771	0.994		1.384 (0.926-2.069)	0.113	0.297		0.829 (0.498-1.378)	0.469	0.838		0.845 (0.622-1.146)	0.279	0.637
SubFreeChol_LDL-5	1.073 (0.552-2.086)	0.836	0.994		1.51 (0.987-2.312)	0.058	0.225		1.115 (0.648-1.919)	0.695	0.911		0.901 (0.659-1.232)	0.513	0.76
SubFreeChol_LDL-6	1.33 (0.716-2.468)	0.367	0.994		1.98 (1.309-2.996)	0.001	0.052		1.352 (0.798-2.289)	0.262	0.636		0.827 (0.612-1.117)	0.216	0.637
SubPhosp_LDL-1	0.912 (0.488-1.707)	0.774	0.994		1.013 (0.674-1.524)	0.949	0.974		0.84 (0.496-1.421)	0.515	0.853		0.876 (0.636-1.206)	0.417	0.669
SubPhosp_LDL-2	0.745 (0.359-1.543)	0.427	0.994		1.136 (0.753-1.713)	0.544	0.67		0.63 (0.346-1.146)	0.13	0.413		0.828 (0.599-1.143)	0.251	0.637
SubPhosp_LDL-3	0.969 (0.481-1.952)	0.929	0.994		1.393 (0.923-2.104)	0.115	0.297		0.59 (0.327-1.067)	0.081	0.324		0.78 (0.575-1.059)	0.111	0.637
SubPhosp_LDL-4	1.182 (0.594-2.352)	0.633	0.994		1.152 (0.776-1.709)	0.484	0.619		0.916 (0.566-1.482)	0.72	0.924		0.957 (0.707-1.296)	0.777	0.974
SubPhosp_LDL-5	0.99 (0.497-1.974)	0.978	0.994		1.215 (0.802-1.841)	0.359	0.538		1.226 (0.709-2.121)	0.466	0.838		1.02 (0.747-1.392)	0.902	0.974
SubPhosp_LDL-6	1.95 (1.016-3.743)	0.045	0.726		1.814 (1.212-2.714)	0.004	0.061		1.484 (0.818-2.691)	0.193	0.507		0.862 (0.637-1.166)	0.335	0.637
SubApoB_LDL-1	0.99 (0.551-1.78)	0.974	0.994		1.132 (0.753-1.701)	0.552	0.67		0.846 (0.511-1.401)	0.516	0.853		0.846 (0.617-1.162)	0.302	0.637
SubApoB_LDL-2	0.706 (0.349-1.431)	0.334	0.994		1.186 (0.788-1.784)	0.414	0.565		0.661 (0.37-1.181)	0.162	0.474		0.846 (0.614-1.164)	0.304	0.637
SubApoB_LDL-3	0.976 (0.489-1.946)	0.944	0.994		1.515 (1.001-2.292)	0.049	0.225		0.587 (0.326-1.055)	0.075	0.316		0.763 (0.562-1.037)	0.084	0.637
SubApoB_LDL-4	1.197 (0.596-2.405)	0.613	0.994		1.186 (0.803-1.753)	0.391	0.564		0.975 (0.596-1.595)	0.919	0.99		0.96 (0.712-1.296)	0.792	0.974
SubApoB_LDL-5	1.021 (0.519-2.008)	0.952	0.994		1.329 (0.878-2.01)	0.178	0.385		1.294 (0.748-2.238)	0.356	0.725		0.997 (0.733-1.355)	0.983	0.992
SubApoB_LDL-5	1.991 (0.997-3.975)	0.051	0.726		1.925 (1.271-2.914)	0.002	0.052		1.641 (0.858-3.138)	0.134	0.413		0.888 (0.657-1.202)	0.443	0.691
SubTrigl_HDL-1	0.622 (0.304-1.272)	0.194	0.994		0.942 (0.625-1.421)	0.777	0.86		1.352 (0.768-2.383)	0.296	0.666		1.179 (0.85-1.635)	0.324	0.637
SubTrigl_HDL-2	0.527 (0.265-1.046)	0.067	0.851		0.899 (0.609-1.328)	0.594	0.705		1.347 (0.805-2.254)	0.256	0.635		1.215 (0.887-1.664)	0.225	0.637
SubTrigl_HDL-3	0.467 (0.229-0.951)	0.036	0.726		0.857 (0.58-1.265)	0.437	0.587		1.422 (0.851-2.376)	0.179	0.486		1.362 (0.994-1.868)	0.055	0.637
SubTrigl_HDL-4	0.465 (0.227-0.951)	0.036	0.726		0.657 (0.423-1.019)	0.06	0.225		1.769 (1.013-3.089)	0.045	0.27		1.623 (1.162-2.266)	0.004	0.509
SubChol_HDL-1	1.254 (0.647-2.431)	0.502	0.994		1.08 (0.721-1.619)	0.709	0.816		1.003 (0.595-1.694)	0.99	0.99		0.839 (0.615-1.146)	0.27	0.637
SubChol_HDL-2	0.957 (0.506-1.808)	0.892	0.994		0.784 (0.529-1.16)	0.223	0.443		0.77 (0.462-1.283)	0.315	0.692		0.799 (0.581-1.098)	0.167	0.637
SubChol_HDL-3	0.902 (0.494-1.648)	0.738	0.994		0.873 (0.599-1.274)	0.482	0.619		0.838 (0.507-1.383)	0.488	0.844		0.853 (0.629-1.159)	0.309	0.637
SubChol_HDL-4	1.067 (0.554-2.055)	0.846	0.994		0.819 (0.536-1.251)	0.356	0.538		0.991 (0.582-1.687)	0.973	0.99		0.975 (0.707-1.345)	0.877	0.974
SubFreeChol_HDL-1	1.162 (0.621-2.174)	0.639	0.994		1.277 (0.86-1.898)	0.225	0.443		1.062 (0.634-1.777)	0.82	0.934		0.802 (0.592-1.085)	0.153	0.637
SubFreeChol_HDL-2	1.081 (0.542-2.154)	0.825	0.994		1.271 (0.855-1.888)	0.235	0.444		0.933 (0.539-1.615)	0.804	0.934		0.721 (0.524-0.993)	0.045	0.637
SubFreeChol_HDL-3	0.783 (0.406-1.51)	0.464	0.994		1.016 (0.672-1.537)	0.939	0.973		1.181 (0.684-2.038)	0.551	0.862		0.937 (0.684-1.284)	0.687	0.964

SubFreeChol_HDL-4	0.988 (0.497-1.967)	0.973	0.994		1.036 (0.684-1.57)	0.867	0.918		1.047 (0.611-1.794)	0.868	0.98		0.943 (0.687-1.294)	0.715	0.97
SubPhosp_HDL-1	0.976 (0.491-1.94)	0.945	0.994		0.925 (0.608-1.407)	0.716	0.817		0.99 (0.575-1.704)	0.971	0.99		0.872 (0.633-1.201)	0.401	0.66
SubPhosp_HDL-2	0.757 (0.385-1.488)	0.42	0.994		0.686 (0.456-1.031)	0.07	0.228		0.779 (0.459-1.323)	0.355	0.725		0.852 (0.614-1.181)	0.335	0.637
SubPhosp_HDL-3	0.649 (0.348-1.208)	0.172	0.994		0.668 (0.448-0.997)	0.048	0.225		0.853 (0.507-1.433)	0.547	0.862		0.942 (0.69-1.286)	0.708	0.97
SubPhosp_HDL-4	0.812 (0.419-1.574)	0.537	0.994		0.664 (0.429-1.027)	0.066	0.225		1.012 (0.596-1.72)	0.964	0.99		1.047 (0.759-1.443)	0.781	0.974
SubApoA1_HDL-1	0.676 (0.337-1.358)	0.271	0.994		0.783 (0.51-1.203)	0.264	0.462		1.1 (0.641-1.886)	0.73	0.924		0.973 (0.709-1.336)	0.866	0.974
SubApoA1_HDL-2	0.712 (0.388-1.309)	0.275	0.994		0.696 (0.463-1.045)	0.081	0.256		0.983 (0.587-1.645)	0.949	0.99		1.001 (0.735-1.364)	0.995	0.995
SubApoA1_HDL-3	0.784 (0.404-1.521)	0.472	0.994		0.79 (0.532-1.174)	0.243	0.447		0.997 (0.591-1.681)	0.99	0.99		0.963 (0.705-1.316)	0.812	0.974
SubApoA1_HDL-4	0.881 (0.452-1.716)	0.709	0.994		0.789 (0.508-1.226)	0.292	0.5		1.223 (0.708-2.111)	0.47	0.838		1.069 (0.768-1.487)	0.693	0.964
SubApoA2_HDL-1	0.868 (0.434-1.736)	0.69	0.994		0.989 (0.656-1.49)	0.957	0.974		1.152 (0.662-2.005)	0.617	0.868		0.976 (0.717-1.327)	0.875	0.974
SubApoA2_HDL-2	0.986 (0.49-1.983)	0.969	0.994		1.183 (0.794-1.765)	0.409	0.565		1.128 (0.645-1.973)	0.672	0.891		0.987 (0.726-1.342)	0.932	0.983
SubApoA2_HDL-3	0.778 (0.405-1.494)	0.45	0.994		1.031 (0.689-1.544)	0.881	0.921		1.148 (0.665-1.98)	0.621	0.868		1.05 (0.769-1.433)	0.758	0.974
SubApoA2_HDL-4	1.073 (0.553-2.078)	0.836	0.994		0.96 (0.627-1.47)	0.85	0.918		1.15 (0.672-1.969)	0.61	0.868		1.033 (0.742-1.437)	0.849	0.974

*Logistic regression analyses adjustment for age, sex, time onset-to-treatment, 24h post rt-PA blood glucose level, baseline NIHSS, history of atrial fibrillation, congestive heart failure, recent infections or inflammations, hypertension, diabetes, hyperlipidaemia, smoke and blood collection center. False Discovery Rate correction was applied to *P*-values (*P*) using the Benjamini & Hochberg method (FDR).

Main abbreviations: CI: confidence interval; NIHSS: National Institutes of Health Stroke Scale; OR: odds ratio; trigl: triglycerides; chol: cholesterol; phosp: phospholipids; Apo: apolipoprotein; LMF: lipoprotein main fraction; PN: particle number; 3-HB: 3-hydroxybutyrate. Amino acids are reported with the three letters code.

Table S3. Effect of pre-post-rtPA variations of metabolites and lipids levels on three-months mortality, impairments, symptomatic intracerebral haemorrhage (sICHRCT12) and on the response to i.v. thrombolysis intervention, adjusting *for major determinants for unfavourable outcomes.

	Three-month mortality			Three-month impairments			SICHRCT12			Responder trombolysis		
	OR (95% CI)	P	FDR	OR (95% CI)	P	FDR	OR (95% CI)	P	FDR	OR (95% CI)	P	FDR
Creatinine	0.64 (0.353-1.161)	0.142	0.726	1.201 (0.827-1.744)	0.337	0.689	0.755 (0.487-1.169)	0.207	0.734	0.754 (0.566-1.004)	0.053	0.261
Ala	1.028 (0.615-1.72)	0.916	0.93	0.838 (0.583-1.205)	0.34	0.689	1.104 (0.675-1.808)	0.693	0.818	1.348 (1.011-1.797)	0.042	0.261
Glu	1.035 (0.597-1.794)	0.903	0.93	1.564 (1.054-2.321)	0.026	0.436	1.096 (0.656-1.831)	0.727	0.818	0.855 (0.634-1.153)	0.304	0.497

Gln	1.09 (0.719-1.653)	0.685	0.93	0.919 (0.652-1.296)	0.631	0.811	1.11 (0.677-1.821)	0.68	0.818	1.373 (0.979-1.926)	0.066	0.261
Gly	0.727 (0.406-1.301)	0.282	0.726	0.875 (0.603-1.27)	0.482	0.811	0.78 (0.474-1.284)	0.329	0.734	1.065 (0.805-1.407)	0.661	0.793
His	1.21 (0.747-1.96)	0.44	0.879	1.121 (0.78-1.611)	0.536	0.811	1.005 (0.626-1.613)	0.984	0.984	0.906 (0.688-1.192)	0.48	0.72
Ile	0.627 (0.383-1.028)	0.064	0.726	0.905 (0.634-1.29)	0.58	0.811	1.275 (0.822-1.977)	0.278	0.734	0.999 (0.76-1.312)	0.994	0.994
Leu	0.786 (0.478-1.291)	0.341	0.767	0.84 (0.585-1.206)	0.345	0.689	0.822 (0.516-1.308)	0.408	0.734	0.971 (0.742-1.271)	0.831	0.935
Phe	1.123 (0.67-1.883)	0.66	0.93	0.996 (0.693-1.43)	0.981	0.981	0.546 (0.34-0.877)	0.012	0.213	0.841 (0.633-1.117)	0.231	0.48
Tyr	0.901 (0.558-1.453)	0.668	0.93	0.956 (0.675-1.354)	0.801	0.848	0.882 (0.551-1.413)	0.601	0.818	1.002 (0.759-1.321)	0.991	0.994
Val	0.953 (0.58-1.567)	0.849	0.93	0.913 (0.65-1.284)	0.602	0.811	0.874 (0.548-1.393)	0.572	0.818	0.934 (0.712-1.226)	0.624	0.793
Aceticacid	0.973 (0.524-1.806)	0.93	0.93	1.558 (1.003-2.419)	0.048	0.436	0.993 (0.558-1.77)	0.982	0.984	0.932 (0.693-1.254)	0.642	0.793
Citricacid	1.333 (0.806-2.204)	0.263	0.726	1.425 (0.95-2.139)	0.087	0.498	0.795 (0.482-1.31)	0.367	0.734	0.805 (0.592-1.095)	0.167	0.458
Lacticacid	1.512 (0.825-2.769)	0.181	0.726	1.281 (0.881-1.861)	0.195	0.585	0.621 (0.358-1.078)	0.091	0.408	0.849 (0.636-1.133)	0.267	0.48
3-HB	1.479 (0.781-2.802)	0.23	0.726	1.364 (0.931-1.997)	0.111	0.498	0.797 (0.477-1.329)	0.384	0.734	0.768 (0.576-1.024)	0.072	0.261
Acetone	1.09 (0.621-1.914)	0.763	0.93	1.282 (0.882-1.862)	0.193	0.585	0.839 (0.505-1.392)	0.496	0.812	0.707 (0.533-0.938)	0.016	0.261
Pyruvicacid	1.627 (0.929-2.848)	0.089	0.726	0.934 (0.641-1.362)	0.724	0.848	0.546 (0.316-0.944)	0.03	0.213	0.845 (0.636-1.121)	0.243	0.48
Glucose	1.027 (0.567-1.861)	0.93	0.93	1.052 (0.721-1.535)	0.792	0.848	0.542 (0.306-0.959)	0.035	0.213	0.813 (0.601-1.099)	0.178	0.458
Trigl	1.233 (0.714-2.128)	0.452	0.961	1.144 (0.776-1.686)	0.497	0.954	1.63 (1.015-2.619)	0.043	0.266	0.996 (0.745-1.332)	0.98	0.998
Chol	1.168 (0.717-1.9)	0.533	0.961	1.185 (0.819-1.713)	0.367	0.951	1.121 (0.688-1.827)	0.646	0.945	0.921 (0.697-1.216)	0.561	0.948
LDL-Chol	0.953 (0.599-1.516)	0.838	0.965	0.905 (0.624-1.314)	0.601	0.954	1.169 (0.734-1.861)	0.511	0.945	0.908 (0.685-1.203)	0.502	0.948
HDL-Chol	1.29 (0.734-2.27)	0.376	0.961	1.268 (0.867-1.852)	0.22	0.897	0.873 (0.538-1.416)	0.583	0.945	0.935 (0.703-1.242)	0.641	0.948
Apo-A1	1.006 (0.616-1.643)	0.981	0.989	1.286 (0.877-1.887)	0.198	0.89	1.172 (0.725-1.895)	0.518	0.945	0.967 (0.734-1.274)	0.813	0.968
Apo-A2	0.917 (0.564-1.491)	0.727	0.961	1.269 (0.879-1.832)	0.203	0.89	0.948 (0.589-1.528)	0.827	0.962	0.977 (0.743-1.284)	0.866	0.968
Apo-B100	1.088 (0.67-1.766)	0.734	0.961	1.117 (0.769-1.622)	0.562	0.954	1.117 (0.697-1.79)	0.646	0.945	0.744 (0.558-0.993)	0.044	0.876
LDL-HDL-Chol	0.833 (0.5-1.389)	0.484	0.961	0.773 (0.529-1.13)	0.184	0.89	1.299 (0.8-2.109)	0.29	0.827	0.942 (0.711-1.249)	0.677	0.948
Apo-B100-Apo-A1	1.105 (0.646-1.889)	0.716	0.961	0.909 (0.62-1.332)	0.624	0.954	0.977 (0.578-1.651)	0.931	0.983	0.747 (0.555-1.006)	0.055	0.876
TPN	1.088 (0.67-1.766)	0.734	0.961	1.117 (0.769-1.622)	0.562	0.954	1.117 (0.697-1.79)	0.646	0.945	0.744 (0.558-0.993)	0.044	0.876
VLDL_PN	1.124 (0.649-1.947)	0.676	0.961	1.003 (0.686-1.468)	0.987	0.987	1.813 (1.11-2.961)	0.017	0.221	1 (0.746-1.339)	0.998	0.998
IDL_PN	0.911 (0.559-1.486)	0.709	0.961	1.295 (0.895-1.872)	0.17	0.89	1.168 (0.748-1.822)	0.494	0.945	0.864 (0.656-1.139)	0.299	0.933

LDL_PN	0.896 (0.545-1.472)	0.664	0.961		1.033 (0.705-1.515)	0.868	0.97		0.943 (0.598-1.488)	0.801	0.945		0.768 (0.575-1.025)	0.073	0.876
LDL1_PN	0.845 (0.462-1.546)	0.585	0.961		1.206 (0.829-1.753)	0.328	0.924		0.9 (0.558-1.452)	0.667	0.945		0.924 (0.702-1.217)	0.574	0.948
LDL2_PN	0.72 (0.394-1.314)	0.284	0.961		0.916 (0.64-1.312)	0.633	0.954		0.97 (0.61-1.543)	0.898	0.983		0.809 (0.594-1.102)	0.179	0.933
LDL3_PN	0.675 (0.388-1.174)	0.164	0.961		1.038 (0.693-1.555)	0.857	0.97		0.783 (0.464-1.321)	0.359	0.892		0.691 (0.507-0.941)	0.019	0.876
LDL4_PN	1.219 (0.7-2.126)	0.484	0.961		1.125 (0.767-1.649)	0.547	0.954		1.361 (0.834-2.22)	0.217	0.723		0.943 (0.713-1.247)	0.682	0.948
LDL5_PN	0.71 (0.45-1.122)	0.142	0.961		0.872 (0.62-1.226)	0.431	0.954		1.243 (0.797-1.94)	0.338	0.892		0.905 (0.683-1.2)	0.488	0.948
LDL6_PN	1.404 (0.805-2.449)	0.232	0.961		1.237 (0.846-1.809)	0.274	0.919		1.025 (0.653-1.608)	0.916	0.983		0.838 (0.621-1.13)	0.246	0.933
LMF_Trigl_VLDL	1.265 (0.738-2.168)	0.392	0.961		1.179 (0.797-1.745)	0.409	0.954		1.678 (1.027-2.741)	0.039	0.266		0.998 (0.745-1.339)	0.992	0.998
LMF_Trigl_IDL	0.946 (0.568-1.577)	0.832	0.965		1.242 (0.853-1.808)	0.259	0.919		1.781 (1.033-3.068)	0.038	0.266		1.012 (0.758-1.351)	0.937	0.98
LMF_Trigl_LDL	0.563 (0.35-0.905)	0.018	0.764		1.07 (0.727-1.575)	0.733	0.97		1.048 (0.675-1.626)	0.836	0.962		0.79 (0.592-1.055)	0.11	0.876
LMF_Trigl_HDL	0.542 (0.327-0.9)	0.018	0.764		0.888 (0.647-1.218)	0.46	0.954		1.276 (0.858-1.896)	0.228	0.723		0.978 (0.745-1.284)	0.874	0.968
LMF_Chol_VLDL	1.092 (0.653-1.827)	0.738	0.961		1.088 (0.744-1.589)	0.664	0.958		1.753 (1.094-2.808)	0.02	0.223		1.083 (0.813-1.441)	0.586	0.948
LMF_Chol_IDL	1.068 (0.668-1.706)	0.784	0.961		1.383 (0.963-1.985)	0.079	0.89		1.133 (0.725-1.771)	0.584	0.945		0.89 (0.675-1.173)	0.408	0.948
LMF_Chol_LDL	0.953 (0.599-1.516)	0.838	0.965		0.905 (0.624-1.314)	0.601	0.954		1.169 (0.734-1.861)	0.511	0.945		0.908 (0.685-1.203)	0.502	0.948
LMF_Chol_HDL	1.29 (0.734-2.27)	0.376	0.961		1.268 (0.867-1.852)	0.22	0.897		0.873 (0.538-1.416)	0.583	0.945		0.935 (0.703-1.242)	0.641	0.948
LMF_FreeChol_VLDL	1.283 (0.76-2.165)	0.351	0.961		1.107 (0.752-1.63)	0.605	0.954		1.797 (1.117-2.89)	0.016	0.221		1.057 (0.79-1.413)	0.71	0.948
LMF_FreeChol_IDL	0.984 (0.622-1.557)	0.945	0.984		1.308 (0.916-1.87)	0.14	0.89		1.157 (0.745-1.796)	0.517	0.945		0.918 (0.696-1.21)	0.544	0.948
LMF_FreeChol_LDL	0.979 (0.593-1.618)	0.935	0.984		1.116 (0.762-1.635)	0.572	0.954		0.751 (0.457-1.233)	0.258	0.783		0.792 (0.593-1.058)	0.114	0.876
LMF_FreeChol_HDL	1.739 (0.952-3.176)	0.072	0.961		1.407 (0.944-2.097)	0.094	0.89		0.705 (0.417-1.19)	0.191	0.712		0.898 (0.674-1.195)	0.46	0.948
LMF_Phosp_VLDL	1.206 (0.708-2.055)	0.491	0.961		1.078 (0.73-1.592)	0.707	0.97		2.112 (1.268-3.518)	0.004	0.221		1.151 (0.854-1.551)	0.356	0.933
LMF_Phosp_IDL	0.879 (0.525-1.471)	0.624	0.961		1.02 (0.711-1.465)	0.913	0.987		1.285 (0.795-2.077)	0.306	0.852		1.041 (0.791-1.371)	0.772	0.968
LMF_Phosp_LDL	0.881 (0.54-1.437)	0.612	0.961		0.964 (0.662-1.405)	0.851	0.97		0.93 (0.584-1.48)	0.758	0.945		0.854 (0.644-1.131)	0.271	0.933
LMF_Phosp_HDL	0.914 (0.518-1.612)	0.755	0.961		1.313 (0.897-1.923)	0.162	0.89		0.975 (0.584-1.627)	0.923	0.983		0.945 (0.709-1.259)	0.698	0.948
LMF_ApoA1_HDL	1.041 (0.647-1.676)	0.868	0.977		1.238 (0.863-1.775)	0.246	0.919		1.204 (0.762-1.903)	0.427	0.945		0.88 (0.667-1.16)	0.364	0.933
LMF_ApoA2_HDL	0.932 (0.568-1.53)	0.78	0.961		1.28 (0.882-1.859)	0.194	0.89		0.935 (0.58-1.508)	0.783	0.945		0.972 (0.737-1.28)	0.837	0.968
LMF_ApoB_VLDL	1.124 (0.649-1.948)	0.676	0.961		1.003 (0.686-1.468)	0.986	0.987		1.813 (1.11-2.96)	0.017	0.221		1 (0.746-1.339)	0.998	0.998
LMF_ApoB_IDL	0.911 (0.558-1.485)	0.708	0.961		1.295 (0.895-1.873)	0.17	0.89		1.168 (0.749-1.823)	0.493	0.945		0.864 (0.656-1.138)	0.299	0.933

LMF_ApoB_LDL	0.896 (0.545-1.472)	0.664	0.961		1.033 (0.705-1.515)	0.868	0.97		0.943 (0.598-1.488)	0.801	0.945		0.768 (0.575-1.025)	0.073	0.876
SubTrigl_VLDL-1	1.133 (0.677-1.896)	0.633	0.961		1.038 (0.712-1.514)	0.846	0.97		1.443 (0.865-2.408)	0.16	0.652		1.021 (0.766-1.361)	0.886	0.968
SubTrigl_VLDL-2	1.085 (0.612-1.923)	0.781	0.961		0.945 (0.631-1.414)	0.782	0.97		1.768 (1.048-2.983)	0.033	0.266		1.25 (0.935-1.672)	0.132	0.876
SubTrigl_VLDL-3	1.138 (0.625-2.07)	0.673	0.961		0.991 (0.667-1.473)	0.965	0.987		1.656 (0.963-2.848)	0.068	0.353		1.213 (0.913-1.611)	0.183	0.933
SubTrigl_VLDL-4	1.177 (0.68-2.038)	0.561	0.961		1.018 (0.708-1.463)	0.925	0.987		1.815 (1.031-3.193)	0.039	0.266		0.98 (0.732-1.311)	0.892	0.968
SubTrigl_VLDL-5	1.423 (0.882-2.293)	0.148	0.961		0.943 (0.668-1.331)	0.738	0.97		1.45 (0.874-2.406)	0.15	0.633		0.987 (0.746-1.305)	0.926	0.978
SubChol_VLDL-1	1.126 (0.679-1.867)	0.646	0.961		1.197 (0.817-1.756)	0.356	0.944		1.491 (0.924-2.406)	0.102	0.482		1.03 (0.774-1.372)	0.839	0.968
SubChol_VLDL-2	1.038 (0.583-1.847)	0.899	0.977		1.042 (0.707-1.537)	0.835	0.97		1.957 (1.207-3.172)	0.006	0.221		1.141 (0.862-1.509)	0.357	0.933
SubChol_VLDL-3	1.148 (0.671-1.962)	0.615	0.961		1.201 (0.824-1.75)	0.341	0.927		1.694 (1.013-2.833)	0.044	0.266		1.063 (0.804-1.407)	0.668	0.948
SubChol_VLDL-4	0.971 (0.609-1.548)	0.903	0.977		1.052 (0.733-1.511)	0.784	0.97		1.593 (1.016-2.497)	0.042	0.266		1.059 (0.8-1.4)	0.689	0.948
SubChol_VLDL-5	0.73 (0.414-1.288)	0.278	0.961		0.846 (0.575-1.244)	0.395	0.954		1.208 (0.746-1.958)	0.442	0.945		1.035 (0.78-1.373)	0.814	0.968
SubFreeChol_VLDL-1	1.029 (0.601-1.762)	0.917	0.977		1.037 (0.718-1.497)	0.848	0.97		1.917 (1.129-3.254)	0.016	0.221		1.04 (0.776-1.394)	0.792	0.968
SubFreeChol_VLDL-2	1.093 (0.638-1.873)	0.746	0.961		1.066 (0.733-1.551)	0.737	0.97		1.562 (0.961-2.54)	0.072	0.357		1.102 (0.834-1.456)	0.494	0.948
SubFreeChol_VLDL-3	1.195 (0.681-2.096)	0.535	0.961		1.253 (0.848-1.852)	0.257	0.919		1.606 (0.966-2.67)	0.068	0.353		1.015 (0.763-1.35)	0.919	0.978
SubFreeChol_VLDL-4	1.081 (0.682-1.714)	0.739	0.961		1.244 (0.869-1.782)	0.232	0.914		1.329 (0.866-2.04)	0.194	0.712		0.936 (0.707-1.24)	0.646	0.948
SubFreeChol_VLDL-5	1.385 (0.829-2.314)	0.213	0.961		0.88 (0.617-1.257)	0.483	0.954		1.471 (0.894-2.42)	0.128	0.586		1.069 (0.81-1.41)	0.637	0.948
SubPhosp_VLDL-1	1.158 (0.664-2.018)	0.605	0.961		1.133 (0.766-1.676)	0.533	0.954		1.692 (1.001-2.863)	0.05	0.284		1.043 (0.777-1.399)	0.78	0.968
SubPhosp_VLDL-2	1.17 (0.622-2.203)	0.627	0.961		1.037 (0.693-1.55)	0.86	0.97		2.199 (1.316-3.675)	0.003	0.221		1.18 (0.878-1.585)	0.273	0.933
SubPhosp_VLDL-3	1.068 (0.595-1.914)	0.826	0.965		1.102 (0.747-1.625)	0.626	0.954		2.042 (1.176-3.546)	0.011	0.221		1.079 (0.81-1.436)	0.604	0.948
SubPhosp_VLDL-4	0.964 (0.581-1.602)	0.889	0.977		0.985 (0.68-1.426)	0.935	0.987		1.719 (1.055-2.803)	0.03	0.266		1.029 (0.771-1.373)	0.846	0.968
SubPhosp_VLDL-5	0.952 (0.57-1.592)	0.852	0.971		0.822 (0.571-1.185)	0.294	0.924		1.401 (0.886-2.216)	0.149	0.633		1.156 (0.871-1.534)	0.316	0.933
SubTrigl_LDL-1	0.531 (0.298-0.948)	0.032	0.764		1.046 (0.718-1.524)	0.814	0.97		1.076 (0.678-1.708)	0.755	0.945		0.861 (0.649-1.142)	0.299	0.933
SubTrigl_LDL-2	0.523 (0.293-0.933)	0.028	0.764		1.056 (0.711-1.569)	0.788	0.97		0.992 (0.61-1.612)	0.974	0.983		0.853 (0.642-1.133)	0.272	0.933
SubTrigl_LDL-3	0.832 (0.473-1.461)	0.521	0.961		1.113 (0.759-1.634)	0.583	0.954		1.01 (0.61-1.673)	0.969	0.983		0.8 (0.595-1.075)	0.138	0.876
SubTrigl_LDL-4	0.72 (0.368-1.41)	0.338	0.961		0.988 (0.663-1.472)	0.953	0.987		1.309 (0.756-2.266)	0.337	0.892		0.945 (0.712-1.254)	0.693	0.948
SubTrigl_LDL-5	0.691 (0.4-1.194)	0.186	0.961		1.109 (0.773-1.592)	0.574	0.954		1.324 (0.843-2.077)	0.223	0.723		0.907 (0.689-1.194)	0.485	0.948
SubTrigl_LDL-6	1.563 (0.869-2.811)	0.136	0.961		1.101 (0.752-1.611)	0.62	0.954		1.008 (0.612-1.662)	0.974	0.983		0.846 (0.63-1.135)	0.265	0.933

SubChol_LDL-1	0.8 (0.442-1.447)	0.461	0.961		1.09 (0.749-1.587)	0.652	0.954		0.89 (0.547-1.448)	0.639	0.945		1.033 (0.785-1.36)	0.816	0.968
SubChol_LDL-2	0.928 (0.552-1.561)	0.779	0.961		0.948 (0.663-1.355)	0.769	0.97		0.911 (0.575-1.444)	0.693	0.945		0.884 (0.656-1.191)	0.417	0.948
SubChol_LDL-3	0.816 (0.466-1.429)	0.476	0.961		0.934 (0.62-1.406)	0.743	0.97		0.918 (0.544-1.549)	0.748	0.945		0.908 (0.679-1.213)	0.512	0.948
SubChol_LDL-4	0.968 (0.585-1.601)	0.898	0.977		0.909 (0.627-1.316)	0.612	0.954		1.294 (0.81-2.066)	0.281	0.821		1.136 (0.863-1.495)	0.364	0.933
SubChol_LDL-5	0.785 (0.5-1.234)	0.294	0.961		0.781 (0.552-1.104)	0.162	0.89		1.203 (0.77-1.879)	0.417	0.945		0.933 (0.706-1.233)	0.626	0.948
SubChol_LDL-6	1.276 (0.756-2.152)	0.362	0.961		1.147 (0.791-1.664)	0.469	0.954		1.001 (0.642-1.562)	0.996	0.996		0.873 (0.652-1.17)	0.365	0.933
SubFreeChol_LDL-1	0.794 (0.448-1.407)	0.43	0.961		1.14 (0.794-1.637)	0.478	0.954		0.837 (0.521-1.344)	0.461	0.945		0.949 (0.723-1.247)	0.709	0.948
SubFreeChol_LDL-2	1.085 (0.646-1.825)	0.757	0.961		1.008 (0.707-1.435)	0.966	0.987		0.899 (0.59-1.368)	0.619	0.945		0.871 (0.643-1.181)	0.375	0.933
SubFreeChol_LDL-3	0.91 (0.521-1.589)	0.739	0.961		1.007 (0.676-1.499)	0.973	0.987		0.693 (0.394-1.22)	0.204	0.723		0.844 (0.632-1.126)	0.248	0.933
SubFreeChol_LDL-4	0.923 (0.556-1.534)	0.758	0.961		0.907 (0.62-1.326)	0.614	0.954		1.088 (0.649-1.824)	0.749	0.945		0.902 (0.677-1.202)	0.482	0.948
SubFreeChol_LDL-5	0.828 (0.537-1.278)	0.394	0.961		0.9 (0.643-1.26)	0.54	0.954		1.096 (0.692-1.735)	0.696	0.945		0.885 (0.674-1.161)	0.376	0.933
SubFreeChol_LDL-6	0.879 (0.542-1.425)	0.6	0.961		0.924 (0.639-1.336)	0.674	0.961		0.908 (0.557-1.479)	0.698	0.945		0.934 (0.704-1.239)	0.635	0.948
SubPhosp_LDL-1	0.764 (0.42-1.391)	0.378	0.961		1.095 (0.753-1.593)	0.633	0.954		0.904 (0.555-1.474)	0.686	0.945		1.007 (0.765-1.326)	0.958	0.993
SubPhosp_LDL-2	0.695 (0.383-1.26)	0.23	0.961		0.963 (0.67-1.382)	0.836	0.97		0.916 (0.566-1.48)	0.719	0.945		0.786 (0.575-1.076)	0.132	0.876
SubPhosp_LDL-3	0.759 (0.439-1.312)	0.323	0.961		1.021 (0.677-1.541)	0.919	0.987		0.759 (0.429-1.345)	0.345	0.892		0.768 (0.571-1.033)	0.081	0.876
SubPhosp_LDL-4	1.033 (0.592-1.803)	0.909	0.977		1.077 (0.72-1.611)	0.718	0.97		1.175 (0.712-1.939)	0.528	0.945		1.077 (0.809-1.432)	0.612	0.948
SubPhosp_LDL-5	0.77 (0.482-1.232)	0.276	0.961		0.836 (0.589-1.186)	0.315	0.924		1.162 (0.732-1.843)	0.525	0.945		0.888 (0.667-1.183)	0.417	0.948
SubPhosp_LDL-6	1.356 (0.786-2.337)	0.273	0.961		1.169 (0.8-1.708)	0.42	0.954		1.019 (0.641-1.619)	0.937	0.983		0.878 (0.654-1.179)	0.388	0.94
SubApoB_LDL-1	0.845 (0.461-1.546)	0.584	0.961		1.206 (0.829-1.753)	0.328	0.924		0.901 (0.558-1.453)	0.668	0.945		0.924 (0.701-1.217)	0.573	0.948
SubApoB_LDL-2	0.72 (0.395-1.314)	0.285	0.961		0.917 (0.64-1.312)	0.635	0.954		0.97 (0.609-1.544)	0.897	0.983		0.809 (0.594-1.101)	0.178	0.933
SubApoB_LDL-3	0.675 (0.388-1.174)	0.164	0.961		1.037 (0.692-1.555)	0.859	0.97		0.783 (0.464-1.321)	0.36	0.892		0.691 (0.507-0.941)	0.019	0.876
SubApoB_LDL-4	1.219 (0.699-2.125)	0.485	0.961		1.125 (0.767-1.648)	0.547	0.954		1.361 (0.835-2.221)	0.217	0.723		0.943 (0.713-1.247)	0.682	0.948
SubApoB_LDL-5	0.724 (0.46-1.138)	0.161	0.961		0.841 (0.593-1.193)	0.332	0.924		1.224 (0.785-1.907)	0.373	0.904		0.93 (0.701-1.234)	0.614	0.948
SubApoB_LDL-5	1.404 (0.805-2.449)	0.232	0.961		1.236 (0.845-1.808)	0.274	0.919		1.024 (0.653-1.608)	0.916	0.983		0.838 (0.621-1.13)	0.246	0.933
SubTrigl_HDL-1	0.887 (0.482-1.633)	0.7	0.961		1.264 (0.882-1.813)	0.202	0.89		1.15 (0.727-1.818)	0.551	0.945		0.902 (0.682-1.191)	0.466	0.948
SubTrigl_HDL-2	0.796 (0.469-1.35)	0.398	0.961		1.08 (0.774-1.508)	0.651	0.954		1.336 (0.87-2.053)	0.186	0.712		0.905 (0.681-1.204)	0.493	0.948
SubTrigl_HDL-3	0.703 (0.414-1.192)	0.191	0.961		1.041 (0.745-1.455)	0.812	0.97		1.721 (1.103-2.683)	0.017	0.221		0.976 (0.736-1.294)	0.867	0.968

SubTrigl_HDL-4	0.598 (0.373-0.961)	0.034	0.764		0.776 (0.54-1.113)	0.168	0.89		1.63 (1.039-2.557)	0.033	0.266		1.42 (1.038-1.943)	0.028	0.876
SubChol_HDL-1	1.599 (0.854-2.993)	0.143	0.961		1.414 (0.979-2.044)	0.065	0.89		1.013 (0.631-1.627)	0.956	0.983		0.926 (0.701-1.222)	0.586	0.948
SubChol_HDL-2	1.201 (0.652-2.213)	0.557	0.961		1.35 (0.949-1.92)	0.095	0.89		0.909 (0.596-1.385)	0.657	0.945		0.758 (0.552-1.043)	0.089	0.876
SubChol_HDL-3	0.982 (0.556-1.734)	0.95	0.984		1.354 (0.918-1.998)	0.126	0.89		0.85 (0.51-1.418)	0.534	0.945		0.972 (0.731-1.292)	0.844	0.968
SubChol_HDL-4	1.164 (0.718-1.888)	0.538	0.961		1.152 (0.786-1.687)	0.468	0.954		0.887 (0.556-1.414)	0.614	0.945		1.168 (0.883-1.546)	0.276	0.933
SubFreeChol_HDL-1	1.307 (0.736-2.319)	0.36	0.961		1.368 (0.915-2.045)	0.126	0.89		0.842 (0.498-1.422)	0.519	0.945		0.857 (0.642-1.146)	0.298	0.933
SubFreeChol_HDL-2	0.991 (0.576-1.705)	0.973	0.989		1.518 (1.033-2.229)	0.033	0.89		0.757 (0.466-1.23)	0.261	0.783		0.8 (0.601-1.064)	0.125	0.876
SubFreeChol_HDL-3	0.854 (0.503-1.448)	0.557	0.961		1.213 (0.837-1.758)	0.307	0.924		0.852 (0.485-1.497)	0.577	0.945		0.971 (0.74-1.275)	0.834	0.968
SubFreeChol_HDL-4	1.121 (0.714-1.76)	0.621	0.961		1.11 (0.784-1.572)	0.557	0.954		0.92 (0.557-1.52)	0.746	0.945		1.061 (0.806-1.397)	0.671	0.948
SubPhosp_HDL-1	1.471 (0.756-2.861)	0.256	0.961		1.495 (1.037-2.155)	0.031	0.89		1.008 (0.62-1.638)	0.975	0.983		0.867 (0.654-1.151)	0.324	0.933
SubPhosp_HDL-2	1.013 (0.54-1.899)	0.968	0.989		1.387 (0.983-1.957)	0.062	0.89		0.923 (0.584-1.461)	0.733	0.945		0.792 (0.589-1.066)	0.124	0.876
SubPhosp_HDL-3	0.699 (0.374-1.308)	0.263	0.961		1.331 (0.891-1.987)	0.162	0.89		1.038 (0.617-1.746)	0.889	0.983		0.985 (0.738-1.315)	0.917	0.978
SubPhosp_HDL-4	0.911 (0.513-1.617)	0.751	0.961		1.024 (0.676-1.551)	0.91	0.987		1.089 (0.654-1.812)	0.744	0.945		1.242 (0.934-1.652)	0.136	0.876
SubApoA1_HDL-1	1.086 (0.661-1.783)	0.745	0.961		1.19 (0.85-1.664)	0.311	0.924		1.051 (0.707-1.564)	0.804	0.945		0.951 (0.728-1.244)	0.715	0.948
SubApoA1_HDL-2	1.001 (0.599-1.673)	0.998	0.998		1.303 (0.892-1.903)	0.171	0.89		1.079 (0.647-1.8)	0.77	0.945		0.972 (0.73-1.295)	0.849	0.968
SubApoA1_HDL-3	1.117 (0.653-1.91)	0.686	0.961		1.377 (0.94-2.017)	0.101	0.89		1.075 (0.651-1.775)	0.778	0.945		0.977 (0.734-1.301)	0.875	0.968
SubApoA1_HDL-4	1.114 (0.699-1.774)	0.65	0.961		0.995 (0.673-1.47)	0.978	0.987		1.117 (0.698-1.789)	0.645	0.945		1.208 (0.91-1.602)	0.19	0.933
SubApoA2_HDL-1	1.185 (0.656-2.138)	0.574	0.961		1.517 (1.045-2.202)	0.028	0.89		1.074 (0.633-1.823)	0.792	0.945		0.845 (0.636-1.123)	0.246	0.933
SubApoA2_HDL-2	1.071 (0.614-1.869)	0.809	0.965		1.543 (1.057-2.253)	0.025	0.89		1.049 (0.632-1.741)	0.853	0.973		0.871 (0.653-1.161)	0.346	0.933
SubApoA2_HDL-3	0.743 (0.431-1.281)	0.285	0.961		1.344 (0.918-1.968)	0.128	0.89		1.101 (0.675-1.795)	0.701	0.945		0.974 (0.736-1.289)	0.855	0.968
SubApoA2_HDL-4	1.053 (0.667-1.661)	0.826	0.965		1.089 (0.754-1.571)	0.65	0.954		0.866 (0.538-1.395)	0.555	0.945		1.16 (0.88-1.531)	0.292	0.933

*Logistic regression analyses adjustment for age, sex, time onset-to-treatment, pre rt-PA blood glucose level, baseline NIHSS, history of atrial fibrillation, congestive heart failure, recent infections or inflammations, hypertension, diabetes, hyperlipidaemia, smoke and blood collection center. False Discovery Rate correction was applied to *P*-values (*P*) using the Benjamini & Hochberg method (FDR). Main abbreviations: CI: confidence interval; NIHSS: National Institutes of Health Stroke Scale; OR: odds ratio; trigl: triglycerides; chol: cholesterol; phosp: phospholipids; Apo: apolipoprotein; LMF: lipoprotein main fraction; PN: particle number; 3-HB: 3-hydroxybutyrate. Amino acids are reported with the three letters code.

Table S4. Mean values of metabolites and lipids concentrations at each time-point of blood collection: t_1 , before rt-PA therapy; t_2 , 24h after and t_3 , three months after rt-PA administration. Metabolites and lipids concentrations are reported using the measure unit as reported in the original Bruker IVDr Lipoprotein subclass analysis reports.

	t_1	t_2	t_3
Creatinine	0.09	0.09	0.09
Ala	0.51	0.48	0.52
Glu	0.22	0.19	0.19
Gln	0.62	0.61	0.68
Gly	0.36	0.34	0.39
His	0.12	0.13	0.12
Ile	0.07	0.07	0.07
Leu	0.14	0.14	0.13
Phe	0.08	0.08	0.07
Tyr	0.07	0.06	0.06
Val	0.26	0.27	0.27
Acetic acid	0.11	0.07	0.08
Citric acid	0.2	0.16	0.17
Lactic acid	3.27	2.62	2.76
3-HB	0.25	0.25	0.07
Acetone	0.12	0.21	0.06
Pyruvic acid	0.07	0.07	0.06
Glucose	7.38	7.56	6.52
Trigl	132.8	119.89	111.31
Chol	196.59	192.39	168.42
LDL-Chol	92.81	95.26	73.42
HDL-Chol	53.22	52.59	52.17
Apo-A1	141.09	135.34	136.79
Apo-A2	32.06	30.37	31.87
Apo-B100	86.42	86.48	69.57
LDL-HDL-Chol	1.81	1.86	1.43
Apo-B100-Apo-A1	0.62	0.65	0.51
Total PN ApoB	1571.26	1572.51	1264.93
VLDLPN	161.82	163.64	144.68
IDL PN	139.74	127.16	119.32
LDL PN	1162.46	1177.08	917.76
LDL-1 PN	265.92	256.09	222.84
LDL-2 PN	169.32	186.46	141.87
LDL-3 PN	127.97	128.86	87.31
LDL-4 PN	89.65	85.57	54.88
LDL-5 PN	173.01	180.62	142.38
LDL-6 PN	391.74	380.96	340.71
LMF Trigl VLDL	75.11	66.89	63.22

LMF Trigl IDL	8.79	7.67	6.51
LMF Trigl LDL	23.24	22.7	19.71
LMF Trigl HDL	12.36	11.06	11.7
LMF Chol VLDL	26.56	25.1	23.46
LMF Chol IDL	19.98	17.92	16.48
LMF Chol LDL	92.81	95.26	73.42
LMF Chol HDL	53.22	52.59	52.17
LMF FreeChol VLDL	10.68	10.09	9.09
LMF FreeChol IDL	5.48	4.96	4.56
LMF FreeChol LDL	29.27	29.32	23.32
LMF FreeChol HDL	13.24	12.48	11.72
LMF Phosp VLDL	19.26	17.92	17.36
LMF Phosp IDL	6.55	6.05	5.42
LMF Phosp LDL	53.76	54.77	43.91
LMF Phosp HDL	72.65	68.97	71.01
LMF ApoA1 HDL	140.37	135.97	136.79
LMF ApoA2 HDL	32.82	31.19	32.4
LMF ApoB VLDL	8.9	9	7.96
LMF ApoB IDL	7.68	6.99	6.56
LMF ApoB LDL	63.93	64.74	50.47
SubTrigl VLDL-1	44.73	34.49	34.29
SubTrigl VLDL-2	5.92	6.82	5.82
SubTrigl VLDL-3	8.16	8.96	7.93
SubTrigl VLDL-4	8.24	9.24	7.94
SubTrigl VLDL-5	2.95	3.18	2.6
SubChol VLDL-1	11.05	8.51	8.63
SubChol VLDL-2	3.26	3.12	3
SubChol VLDL-3	4.73	4.38	4.32
SubChol VLDL-4	6.9	7.08	6.56
SubChol VLDL-5	1.15	1.54	1.03
SubFreeChol VLDL-1	2.28	1.75	1.85
SubFreeChol VLDL-2	1.58	1.46	1.37
SubFreeChol VLDL-3	1.97	1.8	1.74
SubFreeChol VLDL-4	3.18	3.11	2.83
SubFreeChol VLDL-5	1.37	1.14	1.05
SubPhosp VLDL-1	6.61	5.25	5.42
SubPhosp VLDL-2	2.04	2.1	2.09
SubPhosp VLDL-3	3.3	3.29	3.33
SubPhosp VLDL-4	5.26	5.62	5.03
SubPhosp VLDL-5	1.52	1.79	1.35
SubTrigl LDL-1	7.99	7.79	6.59
SubTrigl LDL-2	2.52	2.5	2.19
SubTrigl LDL-3	2.75	2.76	2.25
SubTrigl LDL-4	1.9	1.96	1.52
SubTrigl LDL-5	2.48	2.45	2.17

SubTrigl LDL-6	4.99	4.85	4.41
SubChol LDL-1	25.99	25.07	21.73
SubChol LDL-2	16.12	17.83	13.49
SubChol LDL-3	10.47	10.79	6.9
SubChol LDL-4	6.35	6.41	3.74
SubChol LDL-5	11.97	12.7	9.51
SubChol LDL-6	24.13	23.91	21.42
SubFreeChol LDL-1	8.09	7.86	6.85
SubFreeChol LDL-2	6.08	6.43	5.21
SubFreeChol LDL-3	4.05	4.13	2.91
SubFreeChol LDL-4	2.81	2.81	1.74
SubFreeChol LDL-5	3.77	3.89	3.02
SubFreeChol LDL-6	5.52	5.73	5.11
SubPhosp LDL-1	15.13	14.65	12.95
SubPhosp LDL-2	9.23	10.13	7.87
SubPhosp LDL-3	6.42	6.57	4.53
SubPhosp LDL-4	3.88	3.91	2.44
SubPhosp LDL-5	6.57	6.99	5.43
SubPhosp LDL-6	13.49	13.24	12.1
SubApoB LDL-1	14.62	14.08	12.26
SubApoB LDL-2	9.31	10.25	7.8
SubApoB LDL-3	7.04	7.09	4.8
SubApoB LDL-4	4.93	4.71	3.02
SubApoB LDL-5	9.52	9.93	7.83
SubApoB LDL-5	21.55	20.95	18.74
SubTrigl HDL-1	4.72	3.93	4.36
SubTrigl HDL-2	2.2	2	2.12
SubTrigl HDL-3	2.29	2.07	2.24
SubTrigl HDL-4	3.45	3.31	3.38
SubChol HDL-1	18.25	16.26	17.52
SubChol HDL-2	7.92	7.66	7.67
SubChol HDL-3	9.06	8.59	8.77
SubChol HDL-4	16.73	17.99	16.64
SubFreeChol HDL-1	4.83	4.13	4.48
SubFreeChol HDL-2	2.1	1.85	2.02
SubFreeChol HDL-3	1.7	1.61	1.65
SubFreeChol HDL-4	2.96	3.1	2.78
SubPhosp HDL-1	22.76	19.6	22.12
SubPhosp HDL-2	12.55	11.64	12.1
SubPhosp HDL-3	13.78	12.9	13.55
SubPhosp HDL-4	22.23	23.11	22.23
SubApoA1 HDL-1	29.07	25.56	28.97
SubApoA1 HDL-2	19.03	17.01	18.36
SubApoA1 HDL-3	25.5	23.44	23.76
SubApoA1 HDL-4	67.71	69.62	65.77

SubApoA2 HDL-1	3.06	2.35	2.99
SubApoA2 HDL-2	3.48	2.78	3.4
SubApoA2 HDL-3	5.82	5.15	5.8
SubApoA2 HDL-4	18.27	18.44	18.16

* Main abbreviations: trigl: triglycerides; chol: cholesterol; phosp: phospholipids; Apo: apolipoprotein; LMF: lipoprotein main fraction; Sub: subfraction; PN: particle number; 3-HB: 3-hydroxybutyrate. Amino acids are reported with the three letters code.

4.1.3. Differential Network Analysis reveals metabolite and lipid components associated with three-month death and impairment in patients with acute ischemic stroke after thrombolytic treatment with recombinant tissue plasminogen activator

In preparation

Candidate's contributions: conceptualization of the study, statistical analysis of NMR data, interpreting results and writing of the manuscript.

Abstract

Here, we present the standard multivariate, univariate, network reconstruction and differential analysis of metabolite-metabolite and metabolite-lipid association networks built from an array of 18 serum metabolites and 110 lipids identified and quantified through Nuclear Magnetic Resonance (NMR) spectroscopy in a cohort of 248 patients, of which, 22 died and 82 developed impairments within three-months from acute ischemic stroke (AIS) treated with intravenous recombinant tissue plasminogen activator (rt-PA). We deeply explored differences in metabolite and lipid connectivity of patients who did not develop impairments and who survived the transient cerebral ischemia from the related opposite conditions. We report statistically significant differences in the connectivity patterns of both low- and high-molecular weight metabolites, implying underlying variations in the metabolic pathways involving leucine, glycine, glutamine, tyrosine, phenylalanine, citric, lactic and acetic acids, ketone bodies and different lipids, thus characterizing patients' outcomes. Our results evidence the promising and powerful role of the metabolite-metabolite and metabolite-lipid association networks in investigating molecular mechanisms in AIS patients according to their outcomes.

Introduction

Ischemic stroke (IS) is a leading cause of death and disability continuously increasing,¹ and contributing significantly to health costs. There is an urgent need of effective biomarkers useful for the clinical practice and of a better understanding of the dysregulation in the pathophysiological mechanisms of the disease.

Metabolic perturbations are believed to be fundamental events that contribute to the ischemic stroke, to its progression and subsequent unfavourable outcomes²⁻⁹ and comprehensive analytical techniques provide a great chance to identify key metabolic features involved in the onset and progression of this disease.

Nuclear Magnetic Resonance (NMR)-based metabolomics allows a high-throughput analysis of various types of samples, providing information on hundreds of different metabolites and lipid features present in biological matrices.^{10,11} Multivariate and univariate analyses proved to be efficient in characterizing the metabolic signature of diseases¹²⁻¹⁴ and in the context of molecular epidemiology,^{15,16} but integrative systems biology techniques offer a comprehensive representation of the structural and functional characteristics of a certain living organism, helping in the understanding of the inter-relationships among metabolic features at the basis of the system behaviour.¹⁷ Association networks can provide interesting information to describe the status of the biological system under study or to compare the same across different conditions, and correlation among metabolites and lipids levels measured in blood can be used to model and infer, at least partially, the structure of the underlying biological network.¹⁸ In this light, network analysis has proven to be an impressive and powerful tool to deepen the knowledge and interpret the complexity of metabolomic data.¹⁹⁻²³ In

particular, for metabolomic studies, the exploration of association networks revealed to be more efficient when different conditions are compared in the context of a differential network analysis.^{19,22,24}

In this work, we aimed at providing mechanistic insights underlying susceptibility to three-month post-acute ischemic stroke mortality and impairment developed after intravenous (i.v.) thrombolytic treatment with recombinant tissue plasminogen activator (rt-PA), applying a more systematic approach using metabolite-lipid association differential network analysis in combination with standard univariate, exploratory and multivariate analyses. The patients included in this study are a subset of the MAGIC study cohort.²⁵

We investigated, the differences between the serum metabolic and lipid profiles of patients who after three-months from the ischemic event did not develop impairment or had survived and those who developed impairment or died. The analysis was performed in relation to thrombolytic treatment, considering samples collected 24 hours after rt-PA which were retrospectively divided into four groups (not-impaired (3M-nI) /impaired (3M-I) and survivor (3M-nD) /deceased patients (3M-D)) and applying univariate, multivariate and differential network analysis as illustrated in **Figure 1**. Our results showed that dysregulations in the connectivity of triglycerides, HDL, LDL, VLDL fractions and related subfractions, leucine, glycine, glutamine, tyrosine, phenylalanine, citrate, acetate, lactate, acetone and 3-hydroxybutyrate are pivotal for the characterization of post-acute ischemic stroke outcomes.

Material and Methods

Study Population

The study population consists of 248 patients who had an acute ischemic stroke and were admitted for thrombolysis treatment with recombinant tissue plasminogen activator (rt-PA) in 14 different Italian centres, registered in the Safe Implementation of thrombolysis in Stroke-International Stroke Thrombolysis Register (SITS-ISTR, www.sitsinternational.org), according to SITS-Monitoring Study criteria,²⁶ in the frame of the national, observational and multicentric MAGIC study.^{25,27}

The study focuses on the analysis of serum samples collected 24 hours after rt-PA (t_1) and outcomes were defined at evaluation as follows: *i*) mortality at three months after AIS ($t_0 + 3$ months) and *ii*) disability (impairment) at three months after AIS ($t_0 + 3$ months). Impairment was defined according to the modified Rankin disability score^{28,29} and dichotomized into good (modified Rankin scale, 0–2) or poor (modified Rankin scale, 3–6) outcome. An overview of the demographic characteristics and risk factors of analysed patients is reported in **Table 1**.

In detail, we analysed 226 serum samples for survivors, 22 for deceased, 166 for not-impaired and 82 for impaired patients.

Ethical Issues

The study protocol was approved by the local Ethical Committee of the Careggi University Hospital (Florence) and it complies with the Declaration of Helsinki. All patients gave informed consent.

NMR serum sample collection and preparation

Blood samples were collected in tubes with anticoagulants (0.109M sodium citrate at ratio 9:1 or 1.8 mg/ml EDTA). Serum samples were obtained after centrifuging at room temperature at $1500 \times g$ for 15 min; The supernatant was collected in 1.5 ml aliquots and stored at -80°C , until NMR measurements.

NMR experiments

Serum samples were analysed using a Bruker 600 MHz spectrometer working at 600.13 MHz proton Larmor frequency equipped with a 5 mm PATXI 1H-13C-15N and 2H decoupling probe. This includes a z axis gradient coil, an automatic tuning-matching (ATM) and an automatic and refrigerate sample changer (SampleJet). To stabilize approximately, at the level of ± 0.1 K, the sample temperature (310 K), a BTO 2000 thermocouple was employed and each NMR tube was kept for at least 5 min inside the NMR probe head to equilibrate the acquisition temperature of 310 K. The analytical preparation of serum samples and their NMR spectra acquisition followed the procedures detailed elsewhere.¹⁰ For each serum specimen, the 1D NOESY, 1D CPMG and 1D DIFFUSION-EDITED pulse sequences were applied to acquire ^1H -NMR spectra. Raw NMR data were multiplied by an exponential function of 0.3 Hz line-broadening factor, before the application of Fourier transform. Phase and baseline distortions were automatically corrected and transformed spectra were calibrated to the glucose doublet at 5.24 ppm using TopSpin 3.2 (BrukerBioSpin).

Metabolite and Lipid identification and quantification

18 metabolites and lipoprotein fractions were unambiguously identified and quantified using the AVANCE Bruker IVDr (Clinical Screening and In Vitro Diagnostics research, Bruker BioSpin) software.³⁰ In all serum 1D NOESY NMR spectra and VLDL, LDL, IDL, HDL, 15 different subclasses, (VLDL-1 to VLDL-5, LDL-1 to LDL-6 and HDL-1 to HDL-4) were quantified. For each main class and subclass, reported data consist of concentrations of lipids (total cholesterol, free cholesterol, phospholipids and triglycerides) contained in each fraction. Concentrations of apolipoproteins Apo-A1 and ApoA2 were estimated for HDL class and each relative subclass, while Apo-B concentrations are calculated for VLDL, IDL classes and all LDL subclasses.

Statistical analysis

Univariate analysis

Univariate Wilcoxon test³¹ was used to compare metabolite and lipid concentrations between patient groups (3M-I vs. 3M-nI, 3M-nD vs. 3M-D) treated with

the thrombolytic therapy. Benjamini- Hochberg method³² was used to correct for multiple testing; adjusted P -values for FDR <0.05 were deemed statistically significant. Log₂ fold change (FC) ratios of the median intensities were also calculated (**Table S1**).

Multivariate analysis: PCA and Random Forest

Principal Component Analysis (PCA) was applied on all quantified analytes, to investigate, in an unsupervised manner, the data structure and highlighting the possible presence of metabolite and lipid signatures differentiating 3M-nI from 3M-I and 3M-nD from 3M-D AIS patients at t_1 (24h post rt-PA). PCA analysis was performed on data scaled to unit variance.

Random Forest algorithm was employed for sample classification;³³ four classification models were built to discriminate 3M-nI/3M-I and 3M-nD/3M-D patients using sample collected at t_1 . Considering the unbalanced number of subjects in each group to be compared, Random Forest models were built by first making the different groups of the same size by random sampling, from the largest “survivors” or “not-impaired” group, of a comparable number of subjects (20 samples in the “survivor vs. deceased” model and 80 samples in the “not-impaired vs. impaired” model). The procedure was repeated 100 times and results are averaged over the 100 models.

All models mean accuracy, sensitivity, specificity and related 95% CI were calculated.

Network analysis

Networks building

Metabolite-lipid correlation networks were constructed using the PCLRC (Probabilistic Context Likelihood Relatedness on Correlation) algorithm.¹⁹ This algorithm estimates correlation taking into account the background distribution of correlation and using resampling to get solid estimations. The algorithm output a $J \times J$ probability matrix P that is used to filter out spurious and chance correlations. In particular, for the correlation r_{ij} between two metabolites i and j :

$$r_{ij} = \begin{cases} r_{ij} & \text{if } p_{ij} > 0.95 \\ 0 & \text{otherwise} \end{cases}$$

Determining the significance of metabolite and lipid differential connectivity

Differences in terms of connectivity among metabolic features in each couple of networks (3M-I vs. 3M-nI and 3M-nD vs. 3M-D patients) were analysed 24h after rt-PA administration (t_1).

The connectivity of the i metabolite or i lipid is given by:

$$\chi_i = \left(\sum_{j=1}^J |r_{ij}| \right) - 1$$

and differential connectivity is defined as:

$$\Delta\chi_i = |\chi_i^1 - \chi_i^2|$$

where χ_i^1 and χ_i^2 are the connectivity of metabolite or lipid i estimated from metabolite-lipid association networks calculated from data from condition 1 and 2, respectively (*i.e.* 3M-nI and 3M-I, 3M-nD and 3M-D, at t_I).

The statistical significance of the metabolite and lipid differential connectivity was established using a permutation test ($n=1000$) as described in previous publication.²²

Metabolic and lipidic variables that resulted to be statistically significant connected (P -value < 0.05 after FDR correction) were considered to be related to the specific condition under study (post-stroke three-months impairment/death).

Software

All calculation was performed using R (version 3.6.2). Random Forest was performed using the R “randomForest” package,³³ using the default settings. The R code for the PCLRC algorithm and the code to perform differential connectivity analysis are available at the link: semantics.systemsbiology.nl under the SOFTWARE tab.

Results and Discussion

Exploratory analysis

Principal component analysis (PCA) was applied, as an unsupervised multivariate approach, on all quantified metabolic features of all available samples, to obtain an overview of the variation in the data and to check for the presence of metabolic signatures among the compared groups. **Figure 2** shows the PCA 3D score plots of patient samples measured on metabolites and lipids concentrations, colour-coded by patient status: there is no separation among the samples belonging to survivors, deceased and for patients who developed or not developed impairment at 24h post rt-PA.

Random Forest analysis

Since PCA analysis was not able to highlight observable differences in serum profiles of survivor, deceased, impaired and not-impaired patients, we used a supervised method such as the Random Forest algorithm to investigate whether the metabolite and lipid profiles could be employed to discriminate in a predictive manner 3M-nI from 3M-I and 3M-nD from 3M-D, after thrombolysis intervention (t_I).

All models built on serum metabolic and lipidic concentrations resulted to be ineffective to discriminate 3M-I from 3M-nI and 3M-nD from 3M-D AIS patients (see **Table 2**).

Altogether, the common multivariate analysis of serum profiles of AIS patients proved to be unsuccessful to identify multiple spectral characteristics that are different between 3M-I or 3M-nI or between 3M-nD and 3M-D.

Differential Network Analysis

We built serum metabolite-lipid association networks specific for 3M-I/3M-nI and 3M-nD/3M-D patients at t_1 (24h post rt-PA) with the scope of investigating possible perturbations in patients' metabolic status that could be captured by differential network analysis. Metabolic changes were discussed considering that metabolites participating in different metabolic pathways tend to have higher levels of correlations and connectivity,³⁴ while a decreased connectivity of metabolites involved in the association networks in a statistically significant manner, may indicate a reduced role of certain pathways where those metabolites participate.

Differential analysis of not-impaired and impaired specific metabolite and lipid network after thrombolysis treatment

The metabolite and lipid association networks specific to 3M-nI and 3M-I patients are given in **Figure 3A and B**, respectively, while differential connectivity plots are reported in **Figure 4**. Statistically significant differences in connectivity are reported for glutamine, tyrosine, leucine, lactate, acetone, acetate and glycine.

The alteration of lactic acid connections observed in patients with impairment could suggest its decreased role in providing substitute energy fuel and in the metabolic pathways of neuroprotection where lactic acid is normally largely involved; in fact, the transition from aerobic to anaerobic glycolysis is enhanced to support the increasing demand of energy. As a result, the production of pyruvate and lactate increases, and this last one can be shuttled to neurons to guarantee neuron protection and survival.³⁵ This is substantiated also by the observation that 3-hydroxybutyrate (P value = 3×10^{-5} , FDR = 4×10^{-3}) and glucose (P value = 1×10^{-2} , FDR = 7×10^{-2}) median values appeared to be significantly increased for disabled patients. Acetone and acetate were also significantly increased (FDR < 0.05) (**Table S1**) in patients who developed impairment at three months after thrombolysis.

Changes in the connectivity of leucine, citric acid, acetate and acetone, specific to patients who developed neurological impairments treated with thrombolysis after the transient ischemia indicate unbalances in the energy metabolism and oxidative stress-related pathways: alterations in brain energy metabolism are linked to energy deficits associated with ischemia and reperfusion injury, and downregulation of citric acid has been associated to post-stroke cognitive impairment,⁸ while the metabolism of ketone bodies is upregulated to provide alternative energy sources and to maintain free radical homeostasis during ischemia-reperfusion injury.^{36,37}

Increased glutamine connectivity and increased glutamine concentration for patients with impairment suggest alterations in glutamine/glutamate metabolism: glutamine is a main precursor of glutamate and both of them are inter-converted among astrocytes and neurons, guaranteeing glutamine homeostasis and glutamate generation and recycling. However, if glutamate generation and recycling are impaired, glutamate can be an excitatory and possibly toxic neurotransmitter which can lead to glutamate-

induced neurotoxicity. Alterations in glutamine/glutamate metabolism have been observed in patient with ischemic stroke, and increasing levels of serum glutamine were associated to compensatory adaptative mechanisms to counteract glutamate-induced neurotoxicity.³⁸

Moreover, we observed increased levels of phenylalanine (FDR = 0.01) in impaired AIS patients, and differential connectivity of tyrosine. This suggests that phenylalanine and its metabolite tyrosine might be associated with glutamate-induced neurotoxicity, since phenylalanine can suppress the excitatory glutamatergic synaptic transmission.³⁹

Summarizing, the metabolic profiles of 3M-I patients are consistent with the hypothesis of deregulated glutamate metabolism and subsequent glutamate induced neurotoxicity that seems responsible for the worsening of post-stroke quality of life.

We observed increased glycine connectivity for 3M-I patients, indicating a possible role of glycine in impairment. The role of glycine for acute ischemic stroke is quite controversial. Recent studies demonstrated that lower levels of glycine are deleterious for ischemic neuronal injury, while higher level of the same metabolite seem to be neuroprotective.^{40,41}

We also observed altered patterns of connectivity among metabolites and lipid features, indicating that alteration of lipid metabolism may be involved in post-stroke impairments and neurological disabilities. Several studies have explored the relationships between lipidic features and ischemic stroke,^{3,6,42,43} demonstrating how these molecules play dual roles in the aetiology and progression of the disease. Serum cholesterol have been found to be an independent predictor for long-term functional outcomes and higher serum total cholesterol levels have been associated with better prognosis.⁴⁴ Triacylglycerols have been significantly associated with ischemic stroke.⁴⁵

We observed that in the metabolite-lipid association network specific to 3M-I AIS patients, triglycerides show a decreased connectivity. Since triglycerides are hydrolysed to fatty acids to provide alternative energy sources, we associated decrease connectivity to alterations in the triglycerides metabolism leading to a decreased role of triacylglycerols in supporting alternative energy fuel in patients who developed post-stroke impairments. This is line with previous observations regarding the general condition of energy failure that characterizes the topology of the association network of 3M-I patients. All metabolites and lipids highlighted by the differential network analysis are involved in the mechanisms to guarantee energy homeostasis and they show decreased connectivity in the specific network of patients who developed impairment three-months after the transient cerebral ischemia.

Differential analysis of survivor and deceased specific metabolite and lipid network after thrombolysis treatment

The metabolite-lipid association networks specific for 3M-nD and 3M-D AIS patients are shown in **Figure 5**.

There is a statically significant reduction in the connectivity of blood circulating citric acid and acetone and several lipid fractions in the correlation networks of patients who, after three months, did not survive the acute cerebral ischemia (**Figure 6A**). In particular, we observed a loss of structural connections between metabolites and LDL related fractions. Citric acid showed also higher concentrations (P -value = 0.01, FDR = 0.2, **Table S1**) in deceased patients.

As previously discussed, citric acid and ketone bodies are involved in energy metabolism and their levels can change during and after the cerebral ischemia to restore energy homeostasis. Our results suggest that dysregulations in energy metabolism may be associated also with underlying causes related to an increased risk of death at three months from the AIS.

We observed a statistically significant decrease of low-density lipoproteins connectivity (especially VLDL and LDL-5 and LDL-6 sub-particles), in the network specific to deceased patients. Lipoproteins and lipids have been associated with IS,⁴⁶ small dense LDL (sdLDL) and small-sized HDL particles are established risk factors for this disease. It has been shown that AIS is associated with adverse distributions of LDL and HDL subclasses, and short-term mortality is linked to increased levels of small dense LDL particles.⁴⁷ Since sdLDL are more susceptible to oxidation than larger LDLs, we suggest that sdLDL particles may provide an optimal substrate for rt-PA-induced oxidative action and that alterations in the connectivity patterns of lipid sub-fractions reflects an increase in the rt-PA-mediated oxidative damage associated with post-stroke mortality.

We reported that changes in the metabolism of ketone bodies occur during stress conditions and reperfusion oxidative stress, but to date, the exact role and mechanisms of serum acetone in determining the mortality in AIS patients treated with i.v. thrombolysis remain not completely elucidated. Our results indicate strong perturbation in the processes involving serum acetone and VLDL related subfractions, 3-hydroxybutyrate and free cholesterol linked to VLDL-1 subfractions. In particular, disruption of the connectivity of acetone and VLDL again suggests re-modulation of energy metabolisms.

Overall, the reduction of metabolites-lipid connectivity for 3M-D patients may suggest alterations in lipid metabolism during cerebral ischemia which can strongly affect post-stroke mortality, as well as the development of post-stroke impairments.

Conclusions

We have outlined standard multivariate, univariate, network reconstruction and investigation of experimentally identified relationships between metabolites and lipids, and we applied a differential network analysis (paired with a standard univariate and multivariate analysis) to analyse the pattern of correlations of serum circulating metabolites and lipids in AIS patients who did not develop impairment and who survived at three-months from the transient cerebral ischemia, treated with intravenous rt-PA.

While standard multivariate and univariate analysis failed to discriminate between the patient groups, network analysis revealed marked metabolic differences that could be related to development of post-acute ischemic stroke impairment and mortality.

We showed that lipid (triglycerides, HDL, LDL, VLDL fractions and related subfractions), amino acid (leucine, glycine, glutamine, tyrosine, phenylalanine), organic acids (citric, lactic and acetic acids) and ketone bodies (acetone, 3-hydroxybutyrate) metabolisms and their inter-connections, are decisive for characterizing post-stroke functional and neurological outcomes, differentiating serum profiles of 3M-I, 3M-D, 3M-nI and 3M-nD patients.

Our results indicate that dysregulations of the above-mentioned metabolites and lipids connectivity are involved mainly in mechanisms that show how energy failure, glutamate-induced neurotoxicity, oxidative stress and neuroprotection play important roles in the progression of the pathology after the thrombolytic treatment, affecting survivor's outcomes.

Furthermore, acetone emerged as largely involved in the determination of both three-month outcomes (impairment development and mortality) in ischemic stroke treated with thrombolysis.

In conclusion, this study affords important information on how metabolite-metabolite and metabolite-lipid association networks of AIS patients differ according to the patient's outcomes.

References

- (1) Feigin Valery L.; Norrving Bo; Mensah George A. Global Burden of Stroke. *Circulation Research* 2017, 120 (3), 439–448. <https://doi.org/10.1161/CIRCRESAHA.116.308413>.
- (2) Wesley, U. V.; Bhute, V. J.; Hatcher, J. F.; Palecek, S. P.; Dempsey, R. J. Local and Systemic Metabolic Alterations in Brain, Plasma, and Liver of Rats in Response to Aging and Ischemic Stroke, as Detected by Nuclear Magnetic Resonance (NMR) Spectroscopy. *Neurochem. Int.* 2019, 127, 113–124. <https://doi.org/10.1016/j.neuint.2019.01.025>.
- (3) Jung, J. Y.; Lee, H.-S.; Kang, D.-G.; Kim, N. S.; Cha, M. H.; Bang, O.-S.; Ryu, D. H.; Hwang, G.-S. 1H-NMR-Based Metabolomics Study of Cerebral Infarction. *Stroke* 2011, 42 (5), 1282–1288. <https://doi.org/10.1161/STROKEAHA.110.598789>.
- (4) Szpetnar, M.; Hordyjewska, A.; Malinowska, I.; Golab, P.; Kurzepa, J. The Fluctuation of Free Amino Acids in Serum during Acute Ischemic Stroke. *Current Issues in Pharmacy and Medical Sciences* 2016, 29 (4), 151–154. <https://doi.org/10.1515/cipms-2016-0031>.
- (5) Wang, D.; Kong, J.; Wu, J.; Wang, X.; Lai, M. GC-MS-Based Metabolomics Identifies an Amino Acid Signature of Acute Ischemic Stroke. *Neurosci. Lett.* 2017, 642, 7–13. <https://doi.org/10.1016/j.neulet.2017.01.039>.
- (6) Liu, P.; Li, R.; Antonov, A. A.; Wang, L.; Li, W.; Hua, Y.; Guo, H.; Wang, L.; Liu, P.; Chen, L.; Tian, Y.; Xu, F.; Zhang, Z.; Zhu, Y.; Huang, Y. Discovery of

- Metabolite Biomarkers for Acute Ischemic Stroke Progression. *J. Proteome Res.* 2017, 16 (2), 773–779. <https://doi.org/10.1021/acs.jproteome.6b00779>.
- (7) Kimberly, W. T.; Wang, Y.; Pham, L.; Furie, K. L.; Gerszten, R. E. Metabolite Profiling Identifies a Branched Chain Amino Acid Signature in Acute Cardioembolic Stroke. *Stroke* 2013, 44 (5), 1389–1395. <https://doi.org/10.1161/STROKEAHA.111.000397>.
- (8) Liu, M.; Zhou, K.; Li, H.; Dong, X.; Tan, G.; Chai, Y.; Wang, W.; Bi, X. Potential of Serum Metabolites for Diagnosing Post-Stroke Cognitive Impairment. *Mol Biosyst* 2015, 11 (12), 3287–3296. <https://doi.org/10.1039/c5mb00470e>.
- (9) Ke, C.; Pan, C.-W.; Zhang, Y.; Zhu, X.; Zhang, Y. Metabolomics Facilitates the Discovery of Metabolic Biomarkers and Pathways for Ischemic Stroke: A Systematic Review. *Metabolomics* 2019, 15 (12), 152. <https://doi.org/10.1007/s11306-019-1615-1>.
- (10) Vignoli, A.; Ghini, V.; Meoni, G.; Licari, C.; Takis, P. G.; Tenori, L.; Turano, P.; Luchinat, C. High-Throughput Metabolomics by 1D NMR. *Angew. Chem. Int. Ed. Engl.* 2019, 58 (4), 968–994. <https://doi.org/10.1002/anie.201804736>.
- (11) Takis, P. G.; Ghini, V.; Tenori, L.; Turano, P.; Luchinat, C. Uniqueness of the NMR Approach to Metabolomics. *TrAC Trends in Analytical Chemistry* 2019, 120, 115300. <https://doi.org/10.1016/j.trac.2018.10.036>.
- (12) Meoni, G.; Lorini, S.; Monti, M.; Madia, F.; Corti, G.; Luchinat, C.; Zignego, A. L.; Tenori, L.; Gragnani, L. The Metabolic Fingerprints of HCV and HBV Infections Studied by Nuclear Magnetic Resonance Spectroscopy. *Scientific Reports* 2019, 9 (1), 4128. <https://doi.org/10.1038/s41598-019-40028-4>.
- (13) Vignoli, A.; Orlandini, B.; Tenori, L.; Biagini, M. R.; Milani, S.; Renzi, D.; Luchinat, C.; Calabrò, A. S. Metabolic Signature of Primary Biliary Cholangitis and Its Comparison with Celiac Disease. *J. Proteome Res.* 2019, 18 (3), 1228–1236. <https://doi.org/10.1021/acs.jproteome.8b00849>.
- (14) Vignoli, A.; Paciotti, S.; Tenori, L.; Eusebi, P.; Biscetti, L.; Chiasserini, D.; Scheltens, P.; Turano, P.; Teunissen, C.; Luchinat, C.; Parnetti, L. Fingerprinting Alzheimer's Disease by ¹H Nuclear Magnetic Resonance Spectroscopy of Cerebrospinal Fluid. *J. Proteome Res.* 2020, 19 (4), 1696–1705. <https://doi.org/10.1021/acs.jproteome.9b00850>.
- (15) Vignoli, A.; Tenori, L.; Giusti, B.; Takis, P. G.; Valente, S.; Carrabba, N.; Balzi, D.; Barchielli, A.; Marchionni, N.; Gensini, G. F.; Marcucci, R.; Luchinat, C.; Gori, A. M. NMR-Based Metabolomics Identifies Patients at High Risk of Death within Two Years after Acute Myocardial Infarction in the AMI-Florence II Cohort. *BMC Med* 2019, 17 (1), 3. <https://doi.org/10.1186/s12916-018-1240-2>.
- (16) Di Donato, S.; Mislav, A. R.; Vignoli, A.; Mori, E.; Vitale, S.; Biagioni, C.; Hart, C.; Becheri, D.; Del Monte, F.; Luchinat, C.; Di Leo, A.; Mottino, G.; Tenori, L.; Biganzoli, L. Serum Metabolomic as Biomarkers to Differentiate Early from Metastatic Disease in Elderly Colorectal Cancer (Crc) Patients. *Annals of Oncology* 2016, 27. <https://doi.org/10.1093/annonc/mdw335.20>.

- (17) Rosato, A.; Tenori, L.; Cascante, M.; De Atauri Carulla, P. R.; Martins dos Santos, V. A. P.; Saccenti, E. From Correlation to Causation: Analysis of Metabolomics Data Using Systems Biology Approaches. *Metabolomics* 2018, 14 (4), 37. <https://doi.org/10.1007/s11306-018-1335-y>.
- (18) Camacho, D.; de la Fuente, A.; Mendes, P. The Origin of Correlations in Metabolomics Data. *Metabolomics* 2005, 1 (1), 53–63. <https://doi.org/10.1007/s11306-005-1107-3>.
- (19) Saccenti, E.; Suarez-Diez, M.; Luchinat, C.; Santucci, C.; Tenori, L. Probabilistic Networks of Blood Metabolites in Healthy Subjects As Indicators of Latent Cardiovascular Risk. *Journal of Proteome Research* 2015, 14 (2), 1101–1111. <https://doi.org/10.1021/pr501075r>.
- (20) Saccenti, E.; Menichetti, G.; Ghini, V.; Remondini, D.; Tenori, L.; Luchinat, C. Entropy-Based Network Representation of the Individual Metabolic Phenotype. *Journal of Proteome Research* 2016, 15 (9), 3298–3307. <https://doi.org/10.1021/acs.jproteome.6b00454>.
- (21) Vignoli, A.; Tenori, L.; Luchinat, C.; Saccenti, E. Age and Sex Effects on Plasma Metabolite Association Networks in Healthy Subjects. *J. Proteome Res.* 2018, 17 (1), 97–107. <https://doi.org/10.1021/acs.jproteome.7b00404>.
- (22) Vignoli, A.; Tenori, L.; Giusti, B.; Valente, S.; Carrabba, N.; Balzi, D.; Barchielli, A.; Marchionni, N.; Gensini, G. F.; Marcucci, R.; Gori, A. M.; Luchinat, C.; Saccenti, E. Differential Network Analysis Reveals Metabolic Determinants Associated with Mortality in Acute Myocardial Infarction Patients and Suggests Potential Mechanisms Underlying Different Clinical Scores Used To Predict Death. *J. Proteome Res.* 2020, 19 (2), 949–961. <https://doi.org/10.1021/acs.jproteome.9b00779>.
- (23) Afzal, M.; Saccenti, E.; Madsen, M. B.; Hansen, M. B.; Hyldegaard, O.; Skrede, S.; Martins Dos Santos, V. A. P.; Norrby-Teglund, A.; Svensson, M. Integrated Univariate, Multivariate, and Correlation-Based Network Analyses Reveal Metabolite-Specific Effects on Bacterial Growth and Biofilm Formation in Necrotizing Soft Tissue Infections. *J. Proteome Res.* 2020, 19 (2), 688–698. <https://doi.org/10.1021/acs.jproteome.9b00565>.
- (24) Suarez-Diez, M.; Adam, J.; Adamski, J.; Chasapi, S. A.; Luchinat, C.; Peters, A.; Prehn, C.; Santucci, C.; Spyridonidis, A.; Spyroulias, G. A.; Tenori, L.; Wang-Sattler, R.; Saccenti, E. Plasma and Serum Metabolite Association Networks: Comparability within and between Studies Using NMR and MS Profiling. *Journal of Proteome Research* 2017, 16 (7), 2547–2559. <https://doi.org/10.1021/acs.jproteome.7b00106>.
- (25) Gori, A. M.; Giusti, B.; Piccardi, B.; Nencini, P.; Palumbo, V.; Nesi, M.; Nucera, A.; Pracucci, G.; Tonelli, P.; Innocenti, E.; Sereni, A.; Sticchi, E.; Toni, D.; Bovi, P.; Guidotti, M.; Tola, M. R.; Consoli, D.; Micieli, G.; Tassi, R.; Orlandi, G.; Sessa, M.; Perini, F.; Delodovici, M. L.; Zedde, M. L.; Massaro, F.; Abbate, R.; Inzitari, D. Inflammatory and Metalloproteinases Profiles Predict Three-Month Poor Outcomes in Ischemic Stroke Treated with Thrombolysis. *J. Cereb. Blood Flow Metab.* 2017, 37 (9), 3253–3261. <https://doi.org/10.1177/0271678X17695572>.

- (26) Wahlgren, N.; Ahmed, N.; Dávalos, A.; Ford, G. A.; Grond, M.; Hacke, W.; Hennerici, M. G.; Kaste, M.; Kuelkens, S.; Larrue, V.; Lees, K. R.; Roine, R. O.; Soenne, L.; Toni, D.; Vanhooren, G.; SITS-MOST investigators. Thrombolysis with Alteplase for Acute Ischaemic Stroke in the Safe Implementation of Thrombolysis in Stroke-Monitoring Study (SITS-MOST): An Observational Study. *Lancet* 2007, 369 (9558), 275–282. [https://doi.org/10.1016/S0140-6736\(07\)60149-4](https://doi.org/10.1016/S0140-6736(07)60149-4).
- (27) Inzitari Domenico; Giusti Betti; Nencini Patrizia; Gori Anna Maria; Nesi Mascia; Palumbo Vanessa; Piccardi Benedetta; Armillis Alessandra; Pracucci Giovanni; Bono Giorgio; Bovi Paolo; Consoli Domenico; Guidotti Mario; Nucera Antonia; Massaro Francesca; Micieli Giuseppe; Orlandi Giovanni; Perini Francesco; Tassi Rossana; Tola Maria Rosaria; Sessa Maria; Toni Danilo; Abbate Rosanna. MMP9 Variation After Thrombolysis Is Associated With Hemorrhagic Transformation of Lesion and Death. *Stroke* 2013, 44 (10), 2901–2903. <https://doi.org/10.1161/STROKEAHA.113.002274>.
- (28) Sulter Geert; Steen Christel; Jacques De Keyser null. Use of the Barthel Index and Modified Rankin Scale in Acute Stroke Trials. *Stroke* 1999, 30 (8), 1538–1541. <https://doi.org/10.1161/01.STR.30.8.1538>.
- (29) Broderick Joseph P.; Adeoye Opeolu; Elm Jordan. Evolution of the Modified Rankin Scale and Its Use in Future Stroke Trials. *Stroke* 2017, 48 (7), 2007–2012. <https://doi.org/10.1161/STROKEAHA.117.017866>.
- (30) Jiménez, B.; Holmes, E.; Heude, C.; Tolson, R. F.; Harvey, N.; Lodge, S. L.; Chetwynd, A. J.; Cannet, C.; Fang, F.; Pearce, J. T. M.; Lewis, M. R.; Viant, M. R.; Lindon, J. C.; Spraul, M.; Schäfer, H.; Nicholson, J. K. Quantitative Lipoprotein Subclass and Low Molecular Weight Metabolite Analysis in Human Serum and Plasma by 1H NMR Spectroscopy in a Multilaboratory Trial. *Anal. Chem.* 2018, 90 (20), 11962–11971. <https://doi.org/10.1021/acs.analchem.8b02412>.
- (31) Neuhäuser, M. Wilcoxon–Mann–Whitney Test. In *International Encyclopedia of Statistical Science*; Springer, Berlin, Heidelberg, 2011; pp 1656–1658. https://doi.org/10.1007/978-3-642-04898-2_615.
- (32) Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 1995, 289–300.
- (33) Breiman, L. Random Forests. *Machine Learning* 2001, 45 (1), 5–32.
- (34) Pfeiffer, T.; Soyer, O. S.; Bonhoeffer, S. The Evolution of Connectivity in Metabolic Networks. *PLOS Biology* 2005, 3 (7), e228. <https://doi.org/10.1371/journal.pbio.0030228>.
- (35) Berthet, C.; Castillo, X.; Magistretti, P. J.; Hirt, L. New Evidence of Neuroprotection by Lactate after Transient Focal Cerebral Ischaemia: Extended Benefit after Intracerebroventricular Injection and Efficacy of Intravenous Administration. *Cerebrovasc. Dis.* 2012, 34 (5–6), 329–335. <https://doi.org/10.1159/000343657>.

- (36) Early metabolic response to acute myocardial ischaemia in patients undergoing elective coronary angioplasty. - Abstract - Europe PMC <https://europepmc.org/article/pmc/pmc5888439> (accessed Apr 30, 2020).
- (37) Au, A. Chapter Two - Metabolomics and Lipidomics of Ischemic Stroke. In *Advances in Clinical Chemistry*; Makowski, G. S., Ed.; Elsevier, 2018; Vol. 85, pp 31–69. <https://doi.org/10.1016/bs.acc.2018.02.002>.
- (38) Ramonet, D.; Rodríguez, M. J.; Fredriksson, K.; Bernal, F.; Mahy, N. In Vivo Neuroprotective Adaptation of the Glutamate/Glutamine Cycle to Neuronal Death. *Hippocampus* 2004, 14 (5), 586–594. <https://doi.org/10.1002/hipo.10188>.
- (39) Kagiya Tomoko; Glushakov Alexander V.; Sumners Colin; Roose Brandy; Dennis Donn M.; Phillips M. Ian; Ozcan Mehmet S.; Seubert Christoph N.; Martynyuk Anatoly E. Neuroprotective Action of Halogenated Derivatives of L-Phenylalanine. *Stroke* 2004, 35 (5), 1192–1199. <https://doi.org/10.1161/01.STR.0000125722.10606.07>.
- (40) Yao, W.; Ji, F.; Chen, Z.; Zhang, N.; Ren, S.-Q.; Zhang, X.-Y.; Liu, S.-Y.; Lu, W. Glycine Exerts Dual Roles in Ischemic Injury through Distinct Mechanisms. *Stroke* 2012, 43 (8), 2212–2220. <https://doi.org/10.1161/STROKEAHA.111.645994>.
- (41) Lu, Y.; Zhang, J.; Ma, B.; Li, K.; Li, X.; Bai, H.; Yang, Q.; Zhu, X.; Ben, J.; Chen, Q. Glycine Attenuates Cerebral Ischemia/Reperfusion Injury by Inhibiting Neuronal Apoptosis in Mice. *Neurochem. Int.* 2012, 61 (5), 649–658. <https://doi.org/10.1016/j.neuint.2012.07.005>.
- (42) Yang, L.; Lv, P.; Ai, W.; Li, L.; Shen, S.; Nie, H.; Shan, Y.; Bai, Y.; Huang, Y.; Liu, H. Lipidomic Analysis of Plasma in Patients with Lacunar Infarction Using Normal-Phase/Reversed-Phase Two-Dimensional Liquid Chromatography-Quadrupole Time-of-Flight Mass Spectrometry. *Anal Bioanal Chem* 2017, 409 (12), 3211–3222. <https://doi.org/10.1007/s00216-017-0261-6>.
- (43) Ding, X.; Liu, R.; Li, W.; Ni, H.; Liu, Y.; Wu, D.; Yang, S.; Liu, J.; Xiao, B.; Liu, S. A Metabonomic Investigation on the Biochemical Perturbation in Post-Stroke Patients with Depressive Disorder (PSD). *Metab Brain Dis* 2016, 31 (2), 279–287. <https://doi.org/10.1007/s11011-015-9748-z>.
- (44) Pan, S.-L.; Lien, I.-N.; Chen, T. H.-H. Is Higher Serum Total Cholesterol Level Associated with Better Long-Term Functional Outcomes after Noncardioembolic Ischemic Stroke? *Arch Phys Med Rehabil* 2010, 91 (6), 913–918. <https://doi.org/10.1016/j.apmr.2010.02.002>.
- (45) Labreuche, J.; Deplanque, D.; Touboul, P.-J.; Bruckert, E.; Amarenco, P. Association between Change in Plasma Triglyceride Levels and Risk of Stroke and Carotid Atherosclerosis: Systematic Review and Meta-Regression Analysis. *Atherosclerosis* 2010, 212 (1), 9–15. <https://doi.org/10.1016/j.atherosclerosis.2010.02.011>.
- (46) Holmes, M. V.; Millwood, I. Y.; Kartsonaki, C.; Hill, M. R.; Bennett, D. A.; Boxall, R.; Guo, Y.; Xu, X.; Bian, Z.; Hu, R.; Walters, R. G.; Chen, J.; Ala-Korpela, M.; Parish, S.; Clarke, R. J.; Peto, R.; Collins, R.; Li, L.; Chen, Z.; Group, on behalf of the C. K. B. C. Lipids, Lipoproteins, and Metabolites and Risk of Myocardial

Infarction and Stroke. *J Am Coll Cardiol* 2018, 71 (6), 620–632. <https://doi.org/10.1016/j.jacc.2017.12.006>.

(47) Zeljkovic, A.; Vekic, J.; Spasojevic-Kalimanovska, V.; Jelic-Ivanovic, Z.; Bogavac-Stanojevic, N.; Gulan, B.; Spasic, S. LDL and HDL Subclasses in Acute Ischemic Stroke: Prediction of Risk and Short-Term Mortality. *Atherosclerosis* 2010, 210 (2), 548–554. <https://doi.org/10.1016/j.atherosclerosis.2009.11.040>.

(48) Saccenti, E.; Suarez-Diez, M.; Luchinat, C.; Santucci, C.; Tenori, L. Probabilistic Networks of Blood Metabolites in Healthy Subjects as Indicators of Latent Cardiovascular Risk. *J. Proteome Res.* 2015, 14 (2), 1101–1111. <https://doi.org/10.1021/pr501075r>.

Tables**Table 1.** Demographic characteristics and risk factors for 3M-nD, 3M-D, 3M-nI and 3M-I AIS patients.

<i>Parameters</i>	3M-nD (n=226)	3M-D (n=22)	3M-nI (n=166)	3M-I (n=82)
Demographic characteristics				
<i>Age, years, mean and SD</i>	68 ± 11.8	76.4 ± 9.9	67 ± 12.5	72.2 ± 9.8
<i>Sex (male), n (%)</i>	133/226 (58.8%)	6/22 (27.3%)	102/166 (61.4%)	37/82 (45.1%)
Risk factors				
<i>Hypertension, n (%)</i>	132/226 (58.4%)	12/22 (54.5%)	96/166 (57.8%)	48/82 (58.5%)
<i>Diabetes, n (%)</i>	31/226 (13.7%)	6/22 (27.3%)	24/166 (14.4%)	13/82 (15.8%)
<i>Hyperlipidaemia, n (%)</i>	55/226 (24.3%)	2/22 (9%)	35/166 (21.1%)	22/82 (26.8%)
<i>Current smoking, n (%)</i>	35/226 (15.5%)	0/22 (0%)	27/166 (16.3%)	8/82 (9.8%)
<i>Atrial Fibrillation, n (%)</i>	49/226 (23%)	8/22 (36.4%)	37/166 (22.3%)	20/82 (24.4%)
<i>Congestive Heart Failure, n (%)</i>	20/226 (8.8%)	6/22 (27.3%)	11/166 (6.6%)	15/82 (18.3%)

Table 2. Mean values of accuracy, specificity, sensitivity of metabolites and lipids models estimated for 3M-nD/3M-D and 3M-nI/3M-I AIS patients, at 24h post rt-PA (*t*₁).

	24h post rt-PA (<i>t</i>₁)	
	3M-nD vs. 3M-D	3M-I vs. 3M-nI
Mean accuracy % (95% CI)	53.8 (55.6 -51.9)	59.8 (60.4 – 59.1)
Mean specificity % (95% CI)	57.5 (59.5 – 55.6)	60.5 (61.4 – 59.7)
Mean sensitivity % (95% CI)	50.0 (52.2 – 47.8)	59.1 (59.9 – 58.2)

Figures

Figure 1. Graphical representation of the analysis followed to explore differences in serum profiles of AIS patients, using univariate analysis, an unsupervised exploratory approach (PCA analysis), a Random Forest analysis and metabolite-lipid association networks for patients who survived (3M-nD) or not (3M-D) the acute brain ischemia and for those who developed (3M-I) or not (3M-nI) functional and neurological impairments at three-months from the transient event. The AIS is recorded at time t_0 , while serum samples were collected at t_1 , *i.e.* 24 hours the thrombolytic intervention (post rt-PA samples). Survival and absence of impairment were evaluated after three-months, thus retrospectively dividing samples, according to the outcome (survivor *vs.* deceased and not-impaired *vs.* impaired), for the analyses. Differences in metabolite-lipid association networks were estimated using the PCLRC algorithm.⁴⁸

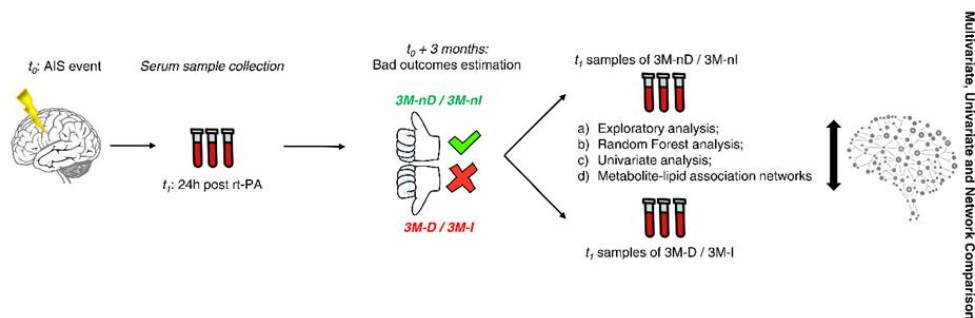


Figure 2. PCA 3D score plots. Each dot represents the serum metabolic profile of AIS patients at 24h post rt-PA (A and B). Colours coded the group of subjects: green dots, 3M-nD ($n=226$); blue dots, 3M-D ($n=22$); orange dots, 3M-nI ($n=166$) and purple dots, 3M-I ($n=82$) patients.

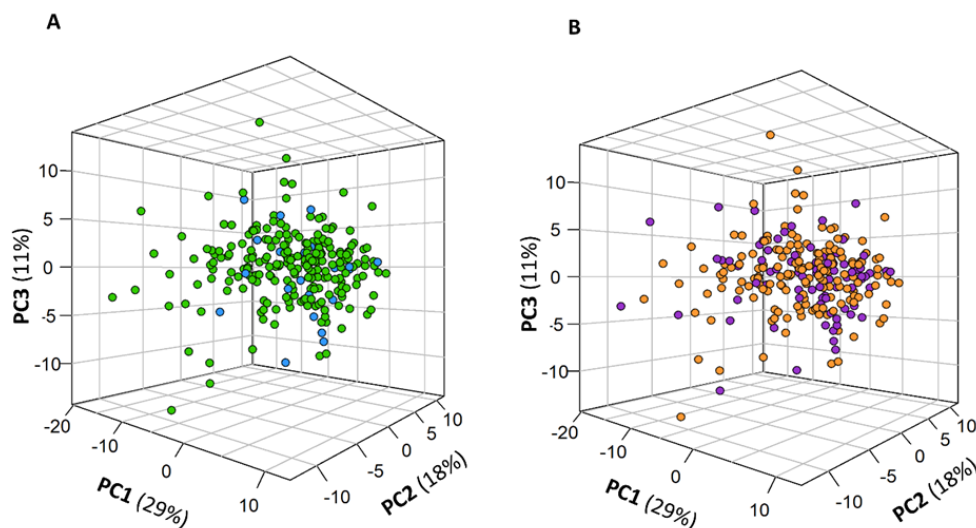


Figure 3. Metabolite-metabolite and metabolite-lipid association specific to 3M-I and 3M-nI AIS patients. **A)** Network specific for 3M-nI. **B)** Network specific for 3M-I. Networks are reconstructed using the PCLRC algorithm from serum metabolites and lipid fractions from sample collected 24h post rt-PA (t_1).

Only the connections of statistically significant differentially connected metabolites (FDR < 0.05) are displayed. Nodes are arranged and coloured according to the increasing metabolite-metabolite or metabolite-lipid degree of connectivity (from pink to purple). Edges represent correlations (black edges display correlation with $|R| > 0.6$). Abbreviations are reported as follows: analytes: 3-HB: 3-hydroxybutyrate, Apo: Apolipoproteins, Chol: cholesterol, LMF: lipoproteins main fractions, Phosp: phospholipids, PN: particle number, Sub: subfractions, Trig: triglycerides. Amino acids are reported with three letter code.

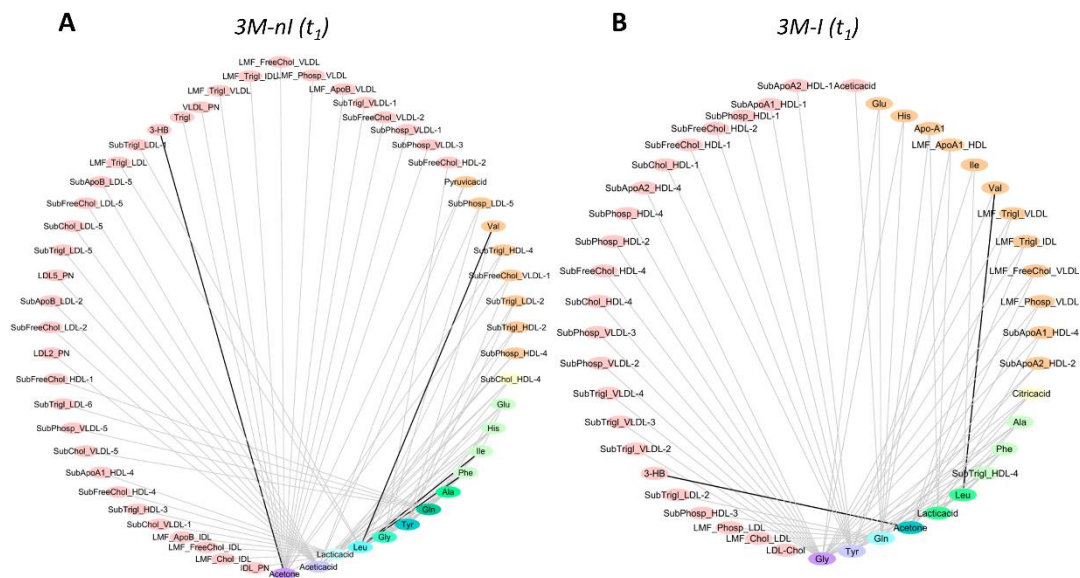


Figure 4. Difference in terms of connectivity in association networks of 3M-nI and 3M-I patients at 24h post rt-PA, t_1 , against each metabolite's (A) and lipid's (B) P -value. The thresholds for significance at 0.05 after FDR correction are reported (grey dashed lines). Colours (from red to blue-violet) code for the increasing difference. Circles represent metabolites, while squares are used for lipids. Differential connectivity of metabolites and lipids, significantly different in terms of concentrations (FDR < 0.05), is also reported both for 3M-nI vs. 3M-I at t_1 (C).

3M-nI vs. 3M-I (t_1)

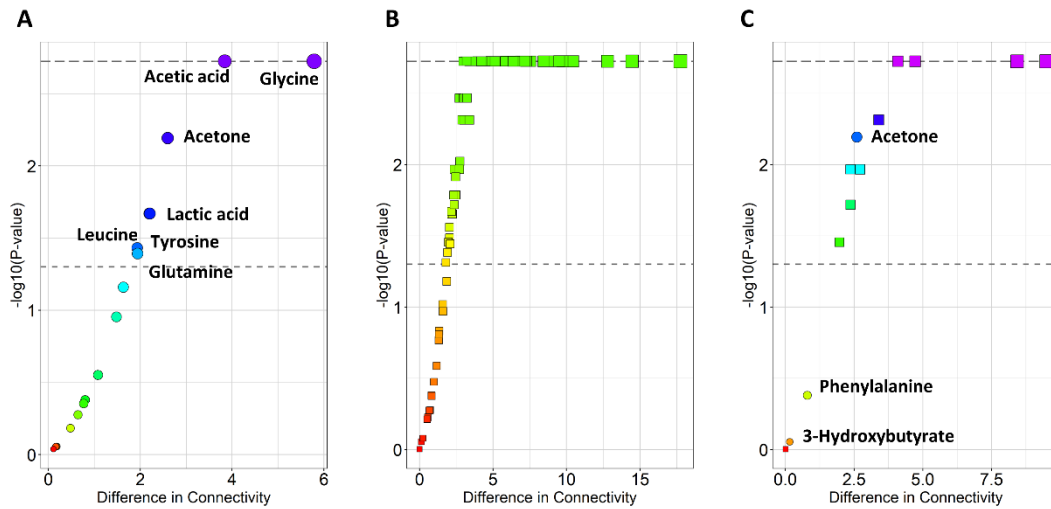


Figure 5. Metabolite-metabolite and metabolite-lipid association networks specific to 3M-nD and 3M-D patients. **A)** Network specific for 3M-nD. **B)** Network specific for 3M-D. Networks are reconstructed using the PCLRC algorithm from serum metabolites and lipid fractions from sample collected 24h post rt-PA (t_1).

Only the connections of statistically significant, differentially connected metabolites (FDR < 0.05) are displayed. Nodes are arranged and coloured according to the increasing metabolite-metabolite or metabolite-lipid degree of connectivity (from pink to purple). Edges represent correlations (black edges display correlation with $|R| > 0.6$). Abbreviations are reported as follows: analytes: 3-HB: 3-hydroxybutyrate, Apo: Apolipoproteins, Chol: cholesterol, LMF: lipoproteins main fractions, Phosp: phospholipids, PN: particle number, Sub: subfractions, Trigl: triglycerides. Amino acids are reported with three letter code.

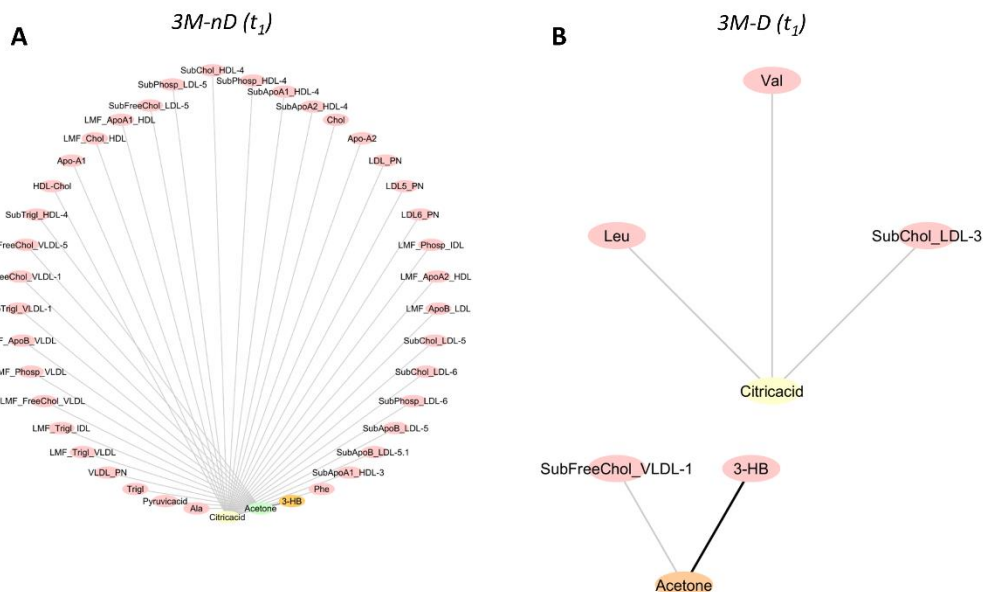
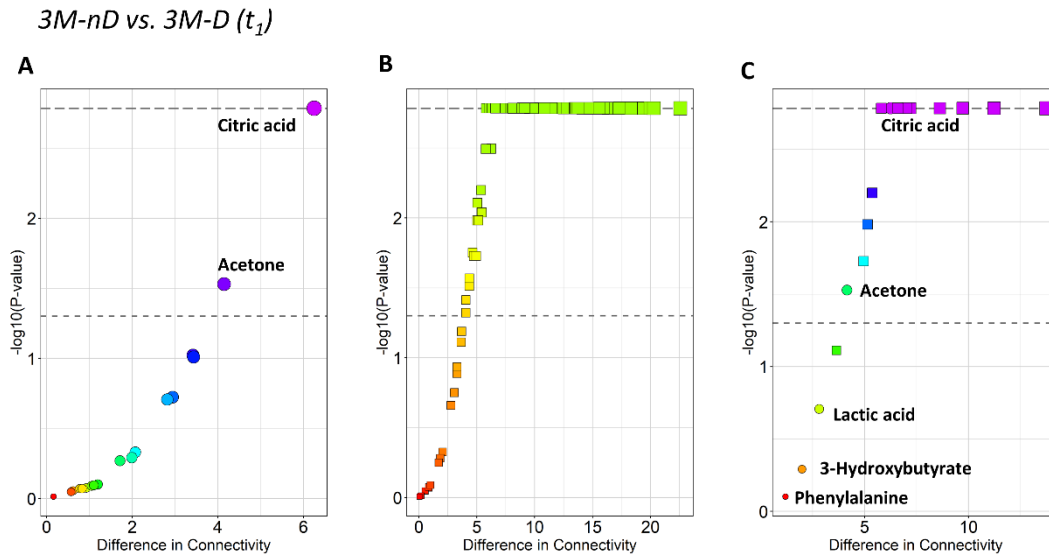


Figure 6. Difference in terms of connectivity in association networks of survivors and deceased patients 24h post rt-PA (t_1), against each metabolite's (A) and lipid's (B) P -value. The thresholds for significance at 0.05 after FDR correction are reported (grey dashed lines). Colours (from red to blue-violet) code for the increasing difference. Circles represent metabolites, while squares are used for lipids. Differential connectivity of metabolites and lipids, significantly different in terms of concentrations (P -value < 0.05), is also reported both for survivor vs. deceased at t_1 (C).



Supplementary Material

Table S1. Metabolite and Lipid Univariate Analysis at 24h post rt-PA (t_I). Metabolic features were assigned and quantified in 1D NOESY NMR spectra and their absolute concentrations are reported as median \pm median absolute deviation. P -value and related adjusted values with the Benjamini-Hochberg correction (P -value < 0.05 and FDR < 0.05) are both deemed significant. Upper (\uparrow) and downward (\downarrow) arrows respectively indicate higher and lower levels in the case group (3M-D or 3M-I patients), while equal symbol (=) means no trend variation.

	3M-nD vs. 3M-D						3M-nI vs. 3M-I					
	3M-nD	3M-D	P	FDR	Log2(FC)	trend	3M-nI	3M-I	P	FDR	Log2(FC)	trend
3-Hydroxybutyrate	0.1 \pm 0.2	0.3 \pm 0.2	0.01	0.3	0.9	\uparrow	0.1 \pm 0.1	0.2 \pm 0.2	3E-05	4E-03	1.3	\uparrow
Acetic acid	0.05 \pm 0.03	0.06 \pm 0.04	0.5	0.8	0.1	\uparrow	0.05 \pm 0.03	0.1 \pm 0.04	2E-02	9E-02	0.3	\uparrow
Acetone	0.1 \pm 0.1	0.2 \pm 0.1	0.05	0.3	0.4	\uparrow	0.1 \pm 0.1	0.2 \pm 0.1	7E-04	1E-02	0.8	\uparrow
Alanine	0.5 \pm 0.1	0.5 \pm 0.2	0.5	0.8	0.00	=	0.5 \pm 0.1	0.5 \pm 0.1	9E-01	1E+00	-0.08	\downarrow
Apo-A1	133.6 \pm 22.7	124.6 \pm 19.3	0.06	0.3	-0.1	\downarrow	134.5 \pm 24.4	127.4 \pm 21.3	1E-01	3E-01	-0.08	\downarrow
Apo-A2	30.1 \pm 5.1	27.9 \pm 6.1	0.04	0.3	-0.1	\downarrow	30.1 \pm 4.9	29.5 \pm 5.1	3E-01	6E-01	-0.03	\downarrow
Apo-B100	83.3 \pm 24.0	83.5 \pm 21.8	0.9	0.9	0.00	=	81.6 \pm 23.1	87.3 \pm 23.5	5E-02	2E-01	0.1	\uparrow
Apo-B100-Apo-A1	0.6 \pm 0.2	0.6 \pm 0.1	0.3	0.7	0.06	\uparrow	0.61 \pm 0.15	0.7 \pm 0.1	2E-03	2E-02	0.1	\uparrow
Chol	188.2 \pm 47.0	195.8 \pm 33.3	0.9	0.9	0.06	\uparrow	183.8 \pm 46.0	195.9 \pm 43.4	1E-01	3E-01	0.09	\uparrow
Citric acid	0.1 \pm 0.05	0.2 \pm 0.1	0.01	0.2	0.4	\uparrow	0.1 \pm 0.04	0.2 \pm 0.1	7E-03	6E-02	0.2	\uparrow
Creatinine	0.08 \pm 0.03	0.09 \pm 0.04	0.4	0.8	0.2	\uparrow	0.1 \pm 0.03	0.1 \pm 0.03	5E-01	7E-01	0.0	=
Glutamine	0.6 \pm 0.2	0.6 \pm 0.2	0.8	0.9	-0.03	\downarrow	0.6 \pm 0.2	0.7 \pm 0.2	3E-02	1E-01	-0.1	\uparrow
Glutamate	0.2 \pm 0.1	0.2 \pm 0.1	0.3	0.7	0.09	\uparrow	0.1 \pm 0.1	0.2 \pm 0.1	6E-01	8E-01	0.2	\uparrow
Glucose	7.0 \pm 1.7	7.5 \pm 2.3	0.3	0.7	0.1	\uparrow	6.7 \pm 1.8	7.4 \pm 1.8	1E-02	7E-02	0.1	\uparrow
Glycine	0.3 \pm 0.09	0.4 \pm 0.1	0.1	0.45	0.2	\uparrow	0.3 \pm 0.1	0.3 \pm 0.1	8E-01	9E-01	0.02	\uparrow
HDL-Chol	51.5 \pm 12.3	50.9 \pm 14.7	0.9	0.9	-0.02	\downarrow	51.5 \pm 12.4	51.2 \pm 12.6	7E-01	9E-01	-0.01	\downarrow
Histidine	0.1 \pm 0.06	0.1 \pm 0.04	0.2	0.5	-0.3	\downarrow	0.1 \pm 0.1	0.1 \pm 0.1	2E-01	4E-01	-0.1	\downarrow
IDL_PN	118.4 \pm 46.5	126.5 \pm 30.0	0.8	0.9	0.1	\uparrow	115.8 \pm 43.7	127.0 \pm 42.6	4E-01	6E-01	0.1	\uparrow
Isoleucine	0.07 \pm 0.03	0.07 \pm 0.02	0.4	0.8	0.0	=	0.1 \pm 0.03	0.1 \pm 0.03	1E+00	1E+00	0.0	=

Lactic acid	2.3 ± 1.0	3.1 ± 0.9	0.01	0.2	0.4	↑	2.3 ± 1.0	2.6 ± 0.9	6E-02	2E-01	0.2	↑
LDL_PN	1137.3 ± 381.2	1182.7 ± 355.9	0.9	0.9	0.06	↑	1094.5 ± 367.1	1218.6 ± 391.8	1E-02	8E-02	0.1	↑
LDL1_PN	249.5 ± 67.6	263.8 ± 78.2	0.5	0.8	0.08	↑	243.1 ± 62.9	262.7 ± 76.0	1E-01	2E-01	0.1	↑
LDL2_PN	189.5 ± 65.7	202.5 ± 68.0	0.6	0.9	0.1	↑	183.3 ± 63.1	207.1 ± 70.0	3E-02	1E-01	0.2	↑
LDL3_PN	128.8 ± 66.0	169.4 ± 56.7	0.1	0.4	0.4	↑	121.8 ± 62.4	158.4 ± 59.3	1E-04	4E-03	0.4	↑
LDL5_PN	160.5 ± 95.2	121.7 ± 94.3	0.1	0.4	-0.4	↓	153.7 ± 89.2	180.2 ± 124.4	8E-01	9E-01	0.2	↑
LDL6_PN	360.3 ± 101.4	362.01 ± 70.2	0.9	0.9	0.01	↑	350.9 ± 92.0	367.9 ± 90.61	6E-01	8E-01	0.07	↑
LDL-Chol	91.8 ± 31.9	96.03 ± 26.0	0.8	0.9	0.06	↑	89.9 ± 29.6	98.2 ± 30.3	2E-02	9E-02	0.1	↑
LDL-HDL-Chol	1.8 ± 0.6	1.8 ± 0.6	0.9	0.9	0.03	↑	1.7 ± 0.6	1.9 ± 0.6	2E-02	1E-01	0.2	↑
Leucine	0.1 ± 0.04	0.1 ± 0.05	0.8	0.9	0.2	↑	0.1 ± 0.04	0.1 ± 0.04	1E+00	1E+00	0.0	=
LMF_ApoA1_HDL	132.3 ± 24.0	122.0 ± 19.1	0.04	0.3	-0.1	↓	134.6 ± 25.6	126.3 ± 18.9	6E-02	2E-01	-0.09	↓
LMF_ApoA2_HDL	30.8 ± 5.1	28.4 ± 5.5	0.04	0.3	-0.1	↓	30.7 ± 5.0	30.2 ± 4.9	3E-01	5E-01	-0.02	↓
LMF_ApoB_IDL	6.5 ± 2.6	7.0 ± 1.6	0.8	0.9	0.09	↑	6.37 ± 2.4	7.0 ± 2.3	3E-01	6E-01	0.1	↑
LMF_ApoB_LDL	62.5 ± 21.0	65.0 ± 19.6	0.9	0.9	0.06	↑	60.2 ± 20.2	67.0 ± 21.5	1E-02	8E-02	0.1	↑
LMF_ApoB_VLDL	8.2 ± 3.0	7.6 ± 2.3	0.4	0.8	-0.1	↓	8.3 ± 2.9	8.0 ± 3.4	4E-01	6E-01	-0.05	↓
LMF_Chol_HDL	51.5 ± 12.3	50.9 ± 14.7	0.9	0.9	-0.02	↓	51.5 ± 12.4	51.2 ± 12.6	7E-01	9E-01	-0.01	↓
LMF_Chol_IDL	17.0 ± 8.0	18.7 ± 3.9	0.7	0.9	0.1	↑	16.7 ± 8.2	18.1 ± 6.3	3E-01	5E-01	0.1	↑
LMF_Chol_LDL	91.8 ± 31.9	96.0 ± 26.0	0.8	0.9	0.06	↑	89.9 ± 29.6	98.2 ± 30.3	2E-02	9E-02	0.1	↑
LMF_Chol_VLDL	24.0 ± 10.5	21.5 ± 4.8	0.3	0.7	-0.2	↓	24.1 ± 10.5	23.5 ± 8.4	3E-01	5E-01	-0.03	↓
LMF_FreeChol_HDL	12.4 ± 3.4	12.5 ± 3.4	0.8	0.9	0.01	↑	12.2 ± 3.5	12.6 ± 2.9	5E-01	7E-01	0.05	↑
LMF_FreeChol_IDL	4.7 ± 2.3	5.0 ± 1.2	0.9	0.9	0.09	↑	4.6 ± 2.4	4.9 ± 1.8	5E-01	7E-01	0.09	↑
LMF_FreeChol_LDL	28.5 ± 8.6	31.5 ± 7.6	0.3	0.7	0.1	↑	27.3 ± 8.9	31.4 ± 8.2	2E-03	2E-02	0.2	↑
LMF_FreeChol_VLDL	9.4 ± 3.9	8.3 ± 2.7	0.3	0.7	-0.2	↓	9.5 ± 4.1	8.8 ± 3.6	2E-01	4E-01	-0.1	↓
LMF_Phosp_HDL	66.1 ± 12.5	69.8 ± 15.0	1.0	1.0	0.08	↑	66.3 ± 13.1	67.1 ± 11.2	9E-01	1E+00	0.02	↑
LMF_Phosp_IDL	5.5 ± 2.9	5.9 ± 1.9	0.5	0.9	0.1	↑	5.2 ± 2.8	5.8 ± 3.4	7E-01	9E-01	0.2	↑
LMF_Phosp_LDL	53.6 ± 15.6	55.4 ± 15.3	0.7	0.9	0.05	↑	51.7 ± 15.5	57.3 ± 16.1	1E-02	8E-02	0.1	↑
LMF_Phosp_VLDL	16.0 ± 7.4	13.1 ± 5.8	0.2	0.6	-0.3	↓	16.7 ± 7.8	14.4 ± 7.5	7E-02	2E-01	-0.2	↓

LMF_Trigl_HDL	10.8 ± 3.1	9.6 ± 3.4	0.3	0.7	-0.2	↓	10.8 ± 3.4	10.6 ± 2.8	5E-01	8E-01	-0.03	↓
LMF_Trigl_IDL	6.1 ± 5.7	3.3 ± 3.2	0.09	0.4	-0.9	↓	6.3 ± 5.6	5.2 ± 5.1	2E-01	4E-01	-0.3	↓
LMF_Trigl_LDL	21.1 ± 6.2	21.8 ± 5.2	0.8	0.9	0.05	↑	20.5 ± 5.9	23.8 ± 6.0	6E-03	5E-02	0.22	↑
LMF_Trigl_VLDL	57.9 ± 28.5	46.7 ± 22.5	0.3	0.7	-0.3	↓	59.0 ± 28.3	52.3 ± 25.9	2E-01	4E-01	-0.2	↓
Phenylalanine	0.08 ± 0.03	0.1 ± 0.04	0.01	0.2	0.3	↑	0.1 ± 0.03	0.1 ± 0.03	8E-04	1E-02	0.2	↑
Pyruvicacid	0.07 ± 0.04	0.08 ± 0.1	0.1	0.4	0.3	↑	0.1 ± 0.04	0.1 ± 0.05	1E+00	1E+00	0.0	=
SubApoA1_HDL-1	24.6 ± 11.9	25.45 ± 12.7	0.5	0.8	0.05	↑	24.7 ± 11.8	24.6 ± 11.7	1E+00	1E+00	0.0	↓
SubApoA1_HDL-2	16.4 ± 4.0	16.1 ± 3.1	0.4	0.8	-0.02	↓	16.7 ± 4.2	16.2 ± 3.0	4E-01	7E-01	-0.04	↓
SubApoA1_HDL-3	23.0 ± 4.9	21.8 ± 4.8	0.1	0.4	-0.08	↓	23.3 ± 5.2	22.6 ± 4.0	2E-01	4E-01	-0.05	↓
SubApoA1_HDL-4	69.2 ± 15.7	57.1 ± 17.0	0.01	0.2	-0.3	↓	70.0 ± 15.0	64.9 ± 16.5	6E-02	2E-01	-0.1	↓
SubApoA2_HDL-1	2.2 ± 1.3	2.7 ± 1.1	0.1	0.4	0.2	↑	2.2 ± 1.3	2.5 ± 1.1	2E-01	4E-01	0.2	↑
SubApoA2_HDL-2	2.7 ± 0.9	3.0 ± 0.7	0.4	0.8	0.1	↑	2.6 ± 0.9	2.8 ± 0.9	3E-01	5E-01	0.09	↑
SubApoA2_HDL-3	5.1 ± 1.2	4.7 ± 0.9	0.2	0.6	-0.1	↓	5.1 ± 1.1	4.9 ± 1.2	9E-01	1E+00	-0.06	↓
SubApoA2_HDL-4	18.4 ± 4.6	14.6 ± 5.4	0.03	0.3	-0.3	↓	18.4 ± 4.3	17.4 ± 5.4	2E-01	4E-01	-0.08	↓
SubApoB_LDL-1	13.7 ± 3.7	14.5 ± 4.3	0.5	0.8	0.08	↑	13.4 ± 3.5	14.4 ± 4.2	1E-01	2E-01	0.2	↑
SubApoB_LDL-2	10.4 ± 3.6	11.1 ± 3.7	0.6	0.9	0.1	↑	10.1 ± 3.5	11.4 ± 3.8	3E-02	1E-01	0.2	↑
SubApoB_LDL-3	7.1 ± 3.6	9.3 ± 3.1	0.1	0.4	0.4	↑	6.7 ± 3.4	8.7 ± 3.3	1E-04	4E-03	0.4	↑
SubApoB_LDL-5	8.8 ± 5.2	6.7 ± 5.2	0.1	0.4	-0.4	↓	8.5 ± 4.9	9.9 ± 6.8	8E-01	9E-01	0.2	↑
SubApoB_LDL-5.1	19.8 ± 5.6	19.9 ± 3.9	0.9	0.9	0.01	↑	19.3 ± 5.1	20.2 ± 5.0	6E-01	8E-01	0.07	↑
SubChol_HDL-1	15.4 ± 7.1	18.0 ± 8.5	0.03	0.3	0.2	↑	15.7 ± 7.6	16.0 ± 6.4	2E-01	3E-01	0.03	↑
SubChol_HDL-2	7.3 ± 2.2	7.8 ± 2.0	0.5	0.9	0.09	↑	7.3 ± 2.3	7.3 ± 2.2	7E-01	9E-01	0.0	=
SubChol_HDL-3	8.4 ± 2.2	8.3 ± 2.0	0.4	0.8	-0.01	↓	8.3 ± 2.3	8.4 ± 1.9	8E-01	9E-01	0.01	↑
SubChol_HDL-4	18.4 ± 5.1	14.4 ± 7.8	0.05	0.3	-0.3	↓	18.5 ± 5.04	17.0 ± 6.3	2E-01	3E-01	-0.1	↓
SubChol_LDL-1	24.6 ± 7.9	26.6 ± 9.9	0.6	0.9	0.1	↑	24.5 ± 7.5	26.6 ± 9.0	2E-01	4E-01	0.1	↑
SubChol_LDL-2	17.78 ± 7.27	19.8 ± 6.6	0.5	0.9	0.2	↑	17.2 ± 7.0	20.2 ± 7.6	5E-02	2E-01	0.2	↑
SubChol_LDL-3	10.5 ± 6.9	13.8 ± 6.1	0.2	0.6	0.4	↑	10.0 ± 6.2	13.5 ± 6.1	6E-04	1E-02	0.4	↑
SubChol_LDL-5	11.1 ± 7.9	6.7 ± 6.4	0.05	0.3	-0.7	↓	10.0 ± 7.4	11.6 ± 9.8	1E+00	1E+00	0.2	↑

SubChol_LDL-6	22.9 ± 6.9	22.3 ± 5.6	0.7	0.9	-0.04	↓	22.5 ± 6.4	23.4 ± 6.8	5E-01	7E-01	0.06	↑
SubChol_VLDL-1	7.3 ± 4.0	5.5 ± 2.8	0.1	0.5	-0.4	↓	7.7 ± 4.1	6.5 ± 2.9	6E-02	2E-01	-0.2	↓
SubChol_VLDL-2	3.0 ± 1.7	2.7 ± 1.02	0.6	0.9	-0.2	↓	3.0 ± 1.7	3.0 ± 1.4	8E-01	9E-01	0.03	↑
SubChol_VLDL-3	4.2 ± 2.4	4.3 ± 1.0	0.8	0.9	0.02	↑	4.1 ± 2.4	4.3 ± 1.9	8E-01	9E-01	0.08	↑
SubChol_VLDL-4	6.78 ± 3.02	6.6 ± 1.4	0.6	0.9	-0.03	↓	6.8 ± 2.9	6.7 ± 2.8	7E-01	9E-01	-0.02	↓
SubChol_VLDL-5	1.5 ± 0.7	1.3 ± 0.6	0.03	0.3	-0.2	↓	1.6 ± 0.8	1.4 ± 0.7	5E-03	5E-02	-0.2	↓
SubFreeChol_HDL-1	3.9 ± 1.7	4.3 ± 1.8	0.1	0.4	0.1	↑	3.8 ± 1.8	4.1 ± 1.6	4E-02	1E-01	0.1	↑
SubFreeChol_HDL-2	1.8 ± 0.5	2.2 ± 0.6	0.04	0.3	0.3	↑	1.7 ± 0.5	2.0 ± 0.7	2E-02	9E-02	0.2	↑
SubFreeChol_HDL-3	1.6 ± 0.7	1.6 ± 0.6	0.6	0.9	-0.02	↓	1.6 ± 0.7	1.6 ± 0.7	6E-01	9E-01	0.02	↑
SubFreeChol_HDL-4	3.2 ± 1.3	2.8 ± 1.4	0.2	0.7	-0.2	↓	3.1 ± 1.3	3.1 ± 1.6	9E-01	1E+00	-0.01	↓
SubFreeChol_LDL-1	7.5 ± 2.4	8.2 ± 3.0	0.6	0.9	0.1	↑	7.4 ± 2.2	8.2 ± 2.8	2E-01	3E-01	0.1	↑
SubFreeChol_LDL-2	6.4 ± 2.3	7.05 ± 2.1	0.4	0.8	0.1	↑	6.2 ± 1.9	6.9 ± 2.6	5E-02	2E-01	0.1	↑
SubFreeChol_LDL-3	4.0 ± 1.8	4.6 ± 1.2	0.1	0.5	0.2	↑	3.8 ± 1.9	4.8 ± 1.6	4E-04	1E-02	0.3	↑
SubFreeChol_LDL-4	2.5 ± 1.9	2.8 ± 1.9	0.8	0.9	0.1	↑	2.2 ± 1.8	3.0 ± 2.2	1E-02	7E-02	0.4	↑
SubFreeChol_LDL-5	3.4 ± 2.0	3.01 ± 1.6	0.2	0.6	-0.2	↓	3.3 ± 1.8	4.0 ± 2.2	3E-01	6E-01	0.3	↑
SubFreeChol_LDL-6	5.6 ± 1.7	5.04 ± 2.05	0.7	0.9	-0.1	↓	5.3 ± 1.5	6.0 ± 1.9	7E-02	2E-01	0.2	↑
SubFreeChol_VLDL-1	1.2 ± 1.2	0.8 ± 1.07	0.1	0.4	-0.7	↓	1.3 ± 1.3	1.0 ± 1.0	6E-02	2E-01	-0.5	↓
SubFreeChol_VLDL-2	1.3 ± 0.8	1.4 ± 0.5	0.9	0.9	0.02	↑	1.3 ± 0.7	1.4 ± 0.7	1E+00	1E+00	0.02	↑
SubFreeChol_VLDL-3	1.6 ± 1.0	1.7 ± 0.5	0.9	0.9	0.04	↑	1.6 ± 1.0	1.6 ± 0.7	9E-01	1E+00	-0.01	↓
SubFreeChol_VLDL-4	3.0 ± 1.6	3.05 ± 0.8	0.8	0.9	0.01	↑	3.0 ± 1.6	3.1 ± 1.3	9E-01	1E+00	0.04	↑
SubFreeChol_VLDL-5	1.0 ± 0.5	0.8 ± 0.4	0.07	0.4	-0.4	↓	1.1 ± 0.5	0.9 ± 0.4	2E-03	2E-02	-0.4	↓
SubPhosp_HDL-1	18.8 ± 8.3	22.3 ± 9.0	0.05	0.3	0.2	↑	18.7 ± 8.6	19.1 ± 8.1	2E-01	4E-01	0.03	↑
SubPhosp_HDL-2	11.2 ± 3.1	12.1 ± 3.7	0.6	0.9	0.1	↑	11.1 ± 3.3	11.6 ± 3.2	7E-01	9E-01	0.06	↑
SubPhosp_HDL-3	12.3 ± 2.8	12.5 ± 2.5	0.4	0.8	0.01	↑	12.5 ± 2.9	12.3 ± 2.7	5E-01	7E-01	-0.03	↓
SubPhosp_HDL-4	23.3 ± 6.4	18.5 ± 7.6	0.02	0.3	-0.3	↓	23.5 ± 6.0	21.7 ± 7.4	9E-02	2E-01	-0.1	↓
SubPhosp_LDL-1	14.3 ± 4.0	15.1 ± 4.8	0.6	0.9	0.08	↑	14.3 ± 3.8	15.2 ± 4.4	2E-01	4E-01	0.09	↑
SubPhosp_LDL-2	10.1 ± 3.7	11.3 ± 3.1	0.4	0.8	0.2	↑	9.9 ± 3.5	11.4 ± 3.6	3E-02	1E-01	0.2	↑

SubPhosp_LDL-3	6.6 ± 3.2	8.1 ± 3.4	0.2	0.6	0.3	↑	6.2 ± 3.0	8.0 ± 3.2	5E-04	1E-02	0.4	↑
SubPhosp_LDL-5	6.2 ± 4.0	3.9 ± 4.1	0.06	0.3	-0.7	↓	5.9 ± 3.8	6.7 ± 5.1	9E-01	1E+00	0.2	↑
SubPhosp_LDL-6	12.8 ± 3.6	12.2 ± 3.1	0.7	0.9	-0.07	↓	12.3 ± 3.5	13.0 ± 3.6	5E-01	7E-01	0.08	↑
SubPhosp_VLDL-1	4.2 ± 2.4	3.4 ± 2.6	0.1	0.4	-0.3	↓	4.4 ± 2.6	3.7 ± 2.1	8E-02	2E-01	-0.2	↓
SubPhosp_VLDL-2	2.0 ± 1.2	1.9 ± 0.7	0.7	0.9	-0.09	↓	2.0 ± 1.2	2.0 ± 1.0	7E-01	9E-01	0.01	↑
SubPhosp_VLDL-3	3.2 ± 1.7	2.7 ± 0.9	0.5	0.9	-0.3	↓	3.2 ± 1.7	3.0 ± 1.5	8E-01	9E-01	-0.1	↓
SubPhosp_VLDL-4	5.3 ± 2.2	5.1 ± 1.4	0.5	0.9	-0.05	↓	5.3 ± 2.1	5.3 ± 2.2	7E-01	9E-01	0.02	↑
SubPhosp_VLDL-5	1.8 ± 0.7	1.5 ± 0.4	0.03	0.3	-0.2	↓	1.9 ± 0.7	1.6 ± 0.5	4E-03	4E-02	-0.2	↓
SubTrigl_HDL-1	3.8 ± 1.5	3.8 ± 2.6	0.8	0.9	0.00	=	3.9 ± 1.6	3.8 ± 1.4	9E-01	1E+00	-0.03	↓
SubTrigl_HDL-2	2.0 ± 0.6	1.9 ± 1.1	0.8	0.9	-0.01	↓	2.0 ± 0.7	2.0 ± 0.7	9E-01	1E+00	0.02	↑
SubTrigl_HDL-3	2.1 ± 0.7	1.9 ± 1.0	0.1	0.4	-0.1	↓	2.1 ± 0.7	2.0 ± 0.6	2E-01	4E-01	-0.08	↓
SubTrigl_HDL-4	3.2 ± 1.0	2.4 ± 1.1	0.001	0.1	-0.4	↓	3.3 ± 1.0	2.9 ± 0.8	2E-03	2E-02	-0.2	↓
SubTrigl_LDL-1	7.5 ± 2.2	6.9 ± 3.3	0.8	0.9	-0.1	↓	7.5 ± 2.4	7.5 ± 2.3	2E-01	4E-01	0.01	↑
SubTrigl_LDL-2	2.4 ± 0.8	2.7 ± 0.8	0.2	0.6	0.2	↑	2.3 ± 0.8	2.6 ± 0.8	1E-02	7E-02	0.2	↑
SubTrigl_LDL-3	2.7 ± 0.7	2.7 ± 0.4	0.8	0.9	0.00	=	2.6 ± 0.7	2.9 ± 0.5	2E-02	9E-02	0.1	↑
SubTrigl_LDL-4	1.9 ± 1.3	1.8 ± 1.2	0.9	0.9	-0.08	↓	1.7 ± 1.2	2.3 ± 1.3	2E-02	9E-02	0.4	↑
SubTrigl_LDL-5	2.3 ± 1.2	2.2 ± 0.8	0.8	0.9	-0.08	↓	2.2 ± 1.0	2.5 ± 1.3	9E-02	2E-01	0.2	↑
SubTrigl_LDL-6	4.7 ± 1.3	5.1 ± 1.3	0.1	0.4	0.1	↑	4.5 ± 1.2	5.0 ± 1.4	7E-03	6E-02	0.1	↑
SubTrigl_VLDL-1	26.2 ± 15.8	20.0 ± 18.8	0.1	0.5	-0.4	↓	27.6 ± 16.1	22.8 ± 15.1	7E-02	2E-01	-0.3	↓
SubTrigl_VLDL-2	6.09 ± 4.5	6.0 ± 3.8	1.0	1.0	-0.02	↓	5.8 ± 4.4	6.2 ± 4.2	4E-01	6E-01	0.09	↑
SubTrigl_VLDL-3	8.1 ± 4.9	7.1 ± 3.9	0.7	0.9	-0.2	↓	7.9 ± 4.5	8.1 ± 5.4	7E-01	9E-01	0.04	↑
SubTrigl_VLDL-4	8.4 ± 3.8	8.4 ± 3.2	0.6	0.9	-0.01	↓	8.5 ± 3.4	8.3 ± 4.2	7E-01	9E-01	-0.03	↓
SubTrigl_VLDL-5	3.2 ± 0.9	3.0 ± 0.5	0.4	0.8	-0.06	↓	3.2 ± 0.9	3.0 ± 0.8	2E-01	4E-01	-0.09	↓
TPN	1514.8 ± 436.9	1518.6 ± 397.4	0.9	0.9	0.001	↑	1483.4 ± 419.5	1586.8 ± 426.8	5E-02	2E-01	0.1	↑
Trigl	107.4 ± 42.3	100.9 ± 29.07	0.4	0.8	-0.09	↓	107.8 ± 42.8	104.8 ± 37.9	4E-01	6E-01	-0.04	↓
Tyrosine	0.06 ± 0.02	0.07 ± 0.01	0.3	0.7	0.1	↑	0.1 ± 0.01	0.1 ± 0.01	5E-02	2E-01	0.2	↑
Valine	0.3 ± 0.06	0.3 ± 0.1	0.2	0.7	0.1	↑	0.3 ± 0.1	0.3 ± 0.1	9E-01	1E+00	-0.05	↓

VLDL_PN	149.8 ± 55.4	137.9 ± 42.5	0.4	0.8	-0.1	↓	150.5 ± 53.0	145.7 ± 61.4	4E-01	6E-01	-0.05	↓
----------------	--------------	--------------	-----	-----	------	---	--------------	--------------	-------	-------	-------	---

*Abbreviations: a) calculated parameters: *P*: *P*-value, FDR: adjusted *P* value with Benjamini-Hochberg correction; b) analytes: Apo: Apolipoproteins, Chol: cholesterol, LMF: lipoproteins main fractions, Phosp: phospholipids, PN: particle number, Sub: subfractions, Trig: triglycerides.

4.1.4. Inflammatory metabolic profile of South African patients with prostate cancer

Stefano Cacciatore^{1,2,†}, Martha Wium^{1,†}, Cristina Licari^{3,†}, Lorenzo Masieri^{4,5}, Chanelle Anderson¹, Azola Samkele Salukazana⁶, Lisa Kaestner⁶, Marco Carini⁴, Giuseppina M. Carbone⁷, Carlo V. Catapano^{7,8,9}, Massimo Loda^{10,11}, Towia A. Libermann¹², Luiz F. Zerbini¹

¹ Cancer Genomics Group, International Centre for Genetic Engineering and Biotechnology, Cape Town, South Africa.

² Institute for Reproductive and Developmental Biology, Imperial College, London, United Kingdom.

³ Magnetic Resonance Center (CERM), University of Florence, Sesto Fiorentino, Italy.

⁴ Department of Urology, Clinica Urologica I, Azienda Ospedaliera Careggi, University of Florence, Florence, Italy.

⁵ Pediatric Urology Unit, Meyer Children Hospital, University of Florence, Florence, Italy.

⁶ Division of Urology, University of Cape Town, Groote Schuur Hospital, Cape Town, South Africa.

⁷ Institute of Oncology Research (IOR), Università della Svizzera italiana, Bellinzona, Switzerland.

⁸ Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland.

⁹ Department of Oncology, Faculty of Biology and Medicine, University of Lausanne, Lausanne, Switzerland.

¹⁰ Department of Oncologic Pathology, Dana-Farber Cancer Institute, Boston, Massachusetts.

¹¹ Department of Pathology and Laboratory Medicine, Weill Cornell Medicine, New York, New York.

¹² BIDMC Genomics, Proteomics, Bioinformatics and Systems Biology Center, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, Massachusetts.

† Contributed equally to this work.

Submitted

Candidate's contributions: acquisition of NMR data, writing and review of the Material and Methods section of the manuscript.

Abstract

Men with African ancestry are more likely to develop aggressive prostate cancer (PCa) and to die from this disease. A distinct genomic landscape associated with specific ethnic groups may lead to different metabolic adaptations and inflammatory responses that permit tumor cells to proliferate and to grow. We hypothesize that a higher risk of lethal PCa in men with African ancestry may be associated with high level of systemic inflammation. In this study, we profile the plasma samples from a cohort of South African men with PCa using Nuclear Magnetic Resonance (NMR) spectroscopy. We found that the plasma of patients with very high risk, aggressive PCa have a peculiar metabolic phenotype (metabotype) characterized by extremely high levels of the inflammatory NMR markers, GlycA and GlycB. The inflammatory processes linked to the higher level of GlycA and GlycB are characterized by a deep change of the plasma metabolome that leads the stratification of patients with PCa.

Systemic inflammation plays a role in the metabolic profile of cancer. This study advances our understanding of the relationship between metabolome and systemic inflammation in the context of PCa and opens the door to a totally innovative approach for biomarker discovery and to the development of new therapies aiming to reduce the systemic inflammation in these patients.

Significance: The first metabolomic study for high-risk prostate cancer in African men identifies inflammation as a driving phenotype in the most aggressive form and simultaneously allows the characterization of their metabolic and lipoprotein profiles.

Introduction

Prostate cancer (PCa) is the second most frequent cancer diagnosis made in men and the fifth leading cause of death worldwide (1). Advancing age, family history of PCa and African ancestry are among established risk factors (2,3). PCa in men of African descent tends to have more aggressive phenotypes compared with other ethnicities (4). Studies investigating the possible role of genetic susceptibility in affecting PCa disparities are limited to the African-American population (5-11). Comparing to European-American men, African-American men have two-thirds higher incidence and two-fold greater risk of dying of PCa (4). Although patterns of genetic similarity among inferred African segments of African-American genomes are mostly similar to non-Bantu Niger-Kordofanian-speaking populations of West Africa (12), the hypothesis that ethnic disparity may be related to biological differences in PCa phenotype is supported by recent studies conducted in Southern Africa. Men from Southern Bantu populations have a 2.1-fold and 4.9-fold greater risk than African Americans for presenting at diagnosis a PCa with Gleason score ≥ 8 and prostate-specific antigen (PSA) ≥ 20 ng/mL (13). Recently, a pilot Whole-Genome Sequencing study conducted on only six African prostate tumors indicated a doubling of the

mutational burden in African men compared to men of European ancestry (14). Disparities in PCa risk and aggressiveness across different ethnicities are poorly understood and likely influenced by genetic factors as well as difficulties to access medical resources in Africa (15).

The South African population is a unique blend of African and non-African ancestry (16). Its ethnic diversity represents both a challenge and an opportunity for biomedical research (17). Currently, the major ethnolinguistic groups in South Africa are Black southeastern Bantu-speakers, an admixed population (including European, Southeast Asian, South Asian, Bantu-speaking African, and hunter-gatherer ancestries) referred to as Coloured (18), Whites of European origin (8.9%), and an Indian population originating from the Indian sub-continent (2.5%) (16).

Lately, more effort has been put on exploring the genetic factors contributing to PCa in men of African ancestry (2,5,19,20). Several genes have been identified as differentially methylated or expressed in PCa between men of European or African ancestry (21). The number of studies linking the association of one or multiple single nucleotide polymorphisms in inflammation-related pathways to PCa risk has greatly increased (22,23). Since inflammation may contribute to PCa development and progression to advanced metastatic disease (24), polymorphisms in immune-related genes could at least partially explain the different incidence and mortality of PCa in African men (25). For example, polymorphisms of cyclooxygenase-2 gene present in the South African population with mixed ancestry were found to be associated with a higher risk of PCa (26). African American men, compared to European American men, show an increased incidence of inflammation in biopsy specimens (27) and increased expression of immune-related genes in tumor tissues (28). A distinct genomic landscape of PCa and immune-related genes associated with specific ethnic groups may lead to different metabolic adaptations that permit tumor cells to proliferate. While metabolic reprogramming is known to play a significant role in both PCa initiation and progression, inflammation seems to be ignored in most biomarker studies (29), including metabolomic studies.

Chronic inflammation has a strong impact on the human metabolome (30-33). Metabolic analysis may illuminate systemic metabolic consequences of inflammation and provide novel targets for intervention. Nuclear magnetic resonance (NMR) spectroscopy is a particularly powerful technique when applied to the high-throughput analysis of biofluids such as blood (34-36), urine (37,38), and saliva (39) which can be collected with minimal impact on the participant. NMR-based metabolomics is a straightforward and useful method for the qualitative and quantitative analysis of a wide range of components in blood samples, including low-molecular-weight metabolites and lipoproteins (different for size and composition) (40,41). Moreover, NMR spectroscopy allows the detection in plasma of signals arising from the glycosylation of circulating acute-phase proteins (APPs), such as fibrinogen, α 1-antichymotrypsin, haptoglobin-1, α 1-antitrypsin, complement C3 and α 1-acid glycoprotein (42-44).

The carbohydrate portions of glycoproteins containing N-acetylglucosamine and N-acetylgalactosamine (hereinafter referred to as GlycA) and N-acetylneuraminic acid (a.k.a., sialic acid; hereinafter referred to as GlycB) moieties are visible as two distinct NMR signals. GlycA and GlycB levels have been associated with common markers of inflammation such as C-reactive protein (CRP, hsCRP as assessed by high-sensitivity assay), fibrinogen, interleukin-6, tumor necrosis factor-alpha, lipoprotein-associated phospholipase A₂ and serum amyloid A (42,45-47). Similar to hsCRP, GlycA is a marker of chronic inflammation (46); despite the similarity, GlycA and hsCRP likely capture different aspects of the inflammatory response (48). CRP is an “early” APP and the proteins that contribute the most to the GlycA and GlycB signal (*i.e.*, alpha1-acid glycoprotein, haptoglobin, and alpha1-antitrypsin) rise later in the acute phase response (49). In response to acute and chronic inflammatory stimuli both the concentrations of APPs (49,50) and their glycan structures are modified (51,52).

This study, to our knowledge, is the first to metabolic profile PCa in men from Africa. Here, we used NMR spectroscopy to quantify a panel of 41 signals, including metabolites, lipid groups, proteins, and the inflammatory markers GlycA and GlycB. Moreover, we used an advanced lipoprotein test based on NMR spectroscopy to characterize the lipoprotein subclasses in each sample. For the first time, we reported the relationship between these inflammatory biomarkers (*i.e.*, GlycA and GlycB) and the most aggressive PCa cases. We provide a clear snapshot of the metabolic alterations during the inflammatory process in PCa paving the way to a better understanding of the metabolic changes occurring in PCa.

Materials and Methods

Patients

Participants were recruited from the Urological clinics of Groote Schuur, Eerste Rivier and New Somerset Hospitals in Cape Town, South Africa. Patients scheduled to undergo transurethral resection of the prostate or prostatectomy were enrolled. The diagnosis of PCa was confirmed by histopathologic examinations. The protocol (*HREC454/2012*) was approved by the Human Research Ethics Committee of the Faculty of Health Science, University of Cape Town, South Africa. Written consent was obtained from all the participants before 5 mL of blood was collected in Vacuette® EDTA tube by medical staff. Blood plasma was separated by centrifugation (1000 g for 10 min at 4°C) and stored at -80°C. This procedure is compatible with the standard operating procedures for metabolomic-grade samples recently defined (53).

NMR sample preparation

Plasma samples were thawed at room temperature. An aliquot of 350 µL of a phosphate sodium buffer (70 mM Na₂HPO₄; 20% (v/v) ²H₂O; 6.1 mM NaN₃; 4.6 mM sodium 3-trimethylsilyl [2,2,3,3-²H₄]-propionate (TMSP); pH 7.4) was added to 350 µL of each sample. The mixture was homogenized by vortexing for 30 s, before 600 µL of this mixture was transferred into a 5 mm NMR tube for analysis (41).

NMR analysis and spectral processing

One-dimensional ^1H -NMR spectra were acquired using a Bruker 600 MHz spectrometer (Bruker BioSpin) operating at 600.13 MHz proton Larmor frequency and equipped with a 5 mm PATXI 1H-13C-15N and 2H-decoupling probe including a z-axis gradient coil, an automatic tuning-matching and an automatic and refrigerated sample changer (SampleJet). A BTO 2000 thermocouple was used at the level of approximately 0.1 K on the sample to stabilize the temperature. Before starting measurements, samples were kept inside the NMR probe head for at least 5 minutes to equilibrate temperature at 310 K.

For each plasma sample, one-dimensional ^1H -NMR spectrum was acquired using a standard Nuclear Overhauser Effect Spectroscopy (NOESY) presat pulse sequence (noesygppr1d.comp; Bruker BioSpin) to detect both signals of small metabolites and high molecular weight macromolecules. Parameters of the experiment were: 32 scans, 98304 data points, a spectral width of 18028.846 Hz, an acquisition time of 2.73 s, a relaxation delay of 4 s and a mixing time of 0.01 s.

Before applying Fourier transform, free induction decays were multiplied by an exponential function of 0.3 Hz line-broadening factor. Transformed spectra were automatically corrected for phase and baseline distortions using Topspin 3.2 (Bruker BioSpin). Transformed spectra were automatically calibrated to the anomeric proton signal of α -glucose at 5.24 ppm.

Molecular profiling and lipoprotein quantification

Lipoprotein parameters were estimated on NOESY spectra according to Bruker's B.I.-LISA protocols (Bruker IVDr Lipoprotein subclass analysis) (54). Information related to the main very-low-density lipoprotein (VLDL), low-density lipoprotein (LDL), intermediate-density lipoprotein (IDL), and high-density lipoprotein (HDL) classes and to their subclasses were extrapolated. In detail, information was extracted of five VLDL subclasses (from VLDL-1 to VLDL-5), six LDL sub-classes (from LDL-1 to LDL-6), and four HDL-subclasses (HDL-1 to HDL-4) sorted according to increasing density and decreasing size.

For each class and subclass, calculated data consist of concentrations of lipids, *i.e.*, cholesterol, free cholesterol, phospholipids, and triglycerides. Instead, concentrations of apolipoproteins Apo-A1 and Apo-A2 were estimated for HDL class and each relative subclass, while Apo-B concentrations are calculated for VLDL, IDL classes and all LDL subclasses.

Identification of signals was undertaken using the SBASE database in Amix (v3.9.11; Bruker BioSpin, Germany) or available assignments in the literature (55). The peaks of the identified metabolites were fitted by a combination of a local baseline and Voigt functions based on the multiplicity of the NMR signal (56). GlycA and GlycB signals were quantified by integrating, respectively, the areas between 2.005 and 2.054 ppm and between 2.086 and 2.054 ppm above a local baseline aimed to remove the signal of the lipoproteins. Fitting methods to quantify GlycA and GlycB signals were not used due to their heterogeneity and due to the impossibility to completely distinguish

them from the lipoprotein signal. The amide protein signals belong to plasma proteins were quantified integrating the area between 6.000 and 10.000 ppm.

Statistical and data analysis

Statistical analysis and graphical illustrations of the data were generated in the R (version 3.6.1) (57) and R studio (version 1.1.456) software using scripts developed in-house.

Wilcoxon Rank Sum test (58) was used to compare differences in numerical covariates (*e.g.*, age and metabolite concentration). Fisher's exact test (59) was used to assess differences between categorical variables (*e.g.*, ethnicity). Spearman's test was used to calculate the correlation coefficient (ρ) between variables. The KODAMA algorithm was used to facilitate the identification of patterns representing underlying metabolic phenotypes (metabotype) on all samples in the data set. Dendrograms were performed using the KODAMA output and Ward linkage. Silhouette median value being used to evaluate the optimal number of clusters with the number of possible clusters varying from 2 to 10 (60). *P*-values less than 0.05 were considered to be significant. To account for multiple testing, a false discovery rate (FDR) of <10% was applied (61).

Regression of metabolic profiles against questionnaire responses and clinical features was performed using partial least-squares (PLS) analysis. To assess the predictive ability of the PLS regression model, a 10-fold cross-validation was conducted as previously described (62). This involved iteratively removing 10% of samples prior to any step of the statistical analysis (including PLS component selection, mean-centering, and univariate scaling) and back-predicting them into the model obtained from the remainder of the data. Parameter selection (*i.e.*, best number of components for PLS) was carried out by means of an inner 10-fold cross-validation on the remaining 90% of the data. The overall procedure was repeated 10 times. The goodness of fit parameter (R^2) and the predictive ability parameter (Q^2) were calculated using standard definitions (63).

Results

South African patient cohort

Although few studies have been performed to investigate the metabolic alterations in the blood of patients with PCa, high-risk populations are underrepresented and limited to African Americans. In this study, we recruited 41 South African patients with PCa in order to generate a better understanding of the metabolic changes in the unique South African setting. The majority of patients were characterized by a unique mixed ancestry (61%) referred to as Coloured; the rest were self-classified as Black (22%) and Whites (17%). We classified the aggressiveness of cancer according to the NCCN classification (version 2.2020): *i*) very low, low, and intermediate risk; *ii*) high risk; and *iii*) very high risk. The clinical and demographic features of the patients with PCa are reported in **Table 1**. Patients with regional or

distant metastasis were classified as a separate group. Patients that received ADT, *i.e.*, bilateral orchiectomy (BO), were considered as two distinct groups based on the evidence of castration-resistant PCa (CRPC). In our cohort, we did not observe statistically significant difference among ethnicity in term of NCCN classification in untreated patients, although, we reported an advanced clinical stage in Black men with 71% of them classified as stage T3 or T4 compared to 19% of Coloured and 33% of White men. This disparity was highlighted even by the PSA level. We reported extremely high values of PSA (>100 ng/mL) in 57% of Black compared to 9% of Coloured and 33% of White. As expected, we observed a higher prevalence of diabetes and hypertension in the post-BO group.

GlycA and GlycB inflammatory biomarkers

Growing evidence implicates chronic inflammation as a contributor to PCa development and progression to advanced metastatic disease (24), and as a driver of CRPC development in ADT (64,65). Recently, GlycA and GlycB have been identified as markers of systemic and chronic inflammation but their association with PCa has not been described yet (66). Here, we used the NMR spectroscopy to quantify the signal associated with GlycA and GlycB and, for the first time, we investigated their association with the aggressiveness of PCa. We noted that the values of both markers are higher in patients with very highly aggressive PCa and metastatic PCa (**Figures 1A, 1D**). Indeed, all patients whose GlycA and GlycB was higher than the 80th percentile were diagnosed with poorly differentiated PCa (*i.e.*, Gleason score higher than or equal to 8). Although the limited number of patients who had BO did not allow for enough statistical power, we observed an increased value of both GlycA and GlycB in patients with CRPC, (**Figures 1B, 1E**).

Several studies have reported the association of GlycA and GlycB with key markers of cancer stratification, such as CRP (67). Here, we report for the first time a statistically significant correlation with PSA (**Figures 1C, 1F**). The correlation with PSA implies that a disparity in the GlycA and GlycB related to ethnicity should be expected in patients with PCa. Noteworthy, the three highest values of GlycA and GlycB were found in patients that identified themselves as Black.

Moreover, we quantified the amides of proteins from the NMR spectra. Albumin is the most concentrated protein in the plasma and consequently, the protein amides concentration is highly related to the albumin level. We reported a negative correlation between the NMR inflammatory marker GlycA ($\rho=-0.41$; $P=0.00747$) and GlycB ($\rho=-0.31$; $P=0.0462$) with protein amides. Albumin is known to be a negative APP that decreases in concentration during inflammation. This further supports our finding that the higher values of GlycA and GlycB could be associated with inflammatory process in patients with PCa.

Metabolic stratification of PCa

Reprogramming of metabolism is a widely accepted hallmark of cancer development (29), however the metabolic changes induced by inflammation in cancer

patients have not been fully characterized. Metabolomics represents an essential tool for the stratification of cancer patients into groups of patients with similar metabolic profiles that could share the same clinicopathologic condition (*e.g.*, systemic inflammation). Here, we quantified the metabolites from each plasma sample using the data collected by the NMR experiments. In order to identify potential underlying metabolic phenotypes (a.k.a., metabotype) in patients with treatment-naïve PCa, we applied the KODAMA method to the quantified metabolite concentrations (GlycA, GlycB, and protein amides were not considered in this analysis). We identified four different metabotypes in the KODAMA score plot (**Figure 2A, 2B**) using the hierarchical clustering (68) on the KODAMA scores.

We clearly observed an association between the PCa aggressiveness (based on the NCCN classification) and the metabotypes that we rank from I to IV in order of aggressiveness. When evaluating GlycB and GlycA in the 4 metabotypes, we noted different levels of GlycB in each metabotypes with the highest levels of both GlycA and GlycB in the metabotype IV. Clinical and demographic features of the metabotypes are shown in **Table 2**.

The metabolic profile of the Metabotype IV is the most peculiar. Samples of the metabotype IV showed an unprecedentedly well-defined fingerprint that may reflect a common biologic process that drives the metabolic changes in the blood of patients with high GlycA and GlycB levels. The metabotype IV is formed almost exclusively by patients categorized as very high risk and also includes a patient with metastatic PCa (**Table 2**).

Noteworthy, a patient classified as low risk PCa based on the NCCN classification showed a metabolic profile typical of Metabotype IV but with lower level of GlycA and GlycB. This patient died only 50 days after sample collection due to pancreatic cancer. On the other hand, patients classified as very high aggressiveness but that do not belong to the Metabotype IV showed a survival time longer than 3 years and a lower Gleason Score compared with patients belong to the Metabotype IV. We were unable to record the date of death for 4 out of 7 patients belong to the Metabotype IV. These patients were lost to follow-up at the hospital cancer center where they were recruited nor did they have any type of diagnostic test at a South African clinic or hospital. Considering the severe condition of these patients and the absence of registered diagnostic tests following the last visit, we assume the latter as a rough estimation of the time of survival. Almost all patients of the Metabotype IV seem to have died within one year after the sample collection (**Table 3**).

Noteworthy, we reported a few clues of possible differences of the prostate tissue inflammation among the metabotype. Of 14 patients in the Metabotype III, 3 patients had mild chronic inflammation and 2 patients had chronic inflammation reported on the histological exam of their prostate tissue. Of the 10 patients of the Metabotype II, 2 patients had mild chronic inflammation, 3 patients had acute-on-chronic inflammation, and 1 had acute prostatitis on the histological exam.

Metabolic profiling of PCa metabotypes

The metabolic differences discriminating between among the four metabotypes appear to be clear (**Figure 2C**). Although we were aware of the low number of patients in this cohort, we built a supervised PLS model to evaluate the accuracy of the identification of the most aggressive metabotype (*i.e.*, Metabotype IV) using the metabolic profile. Using a double cross-validation approach, we calculated an accuracy value of 91.2%, with a 95% coefficient interval of 86.0%-94.1%. Next, we used the Wilcoxon rank-sum test to characterize this metabotype compared to the others (**Table S1**). As previously mentioned, the Metabotype IV is characterized by higher values of the inflammatory markers GlycA ($P=7.06 \times 10^{-6}$; $FDR=7.24 \times 10^{-5}$) and GlycB ($P=2.60 \times 10^{-6}$; $FDR=3.56 \times 10^{-5}$) and lower protein level ($P=9.15 \times 10^{-5}$; $FDR=5.36 \times 10^{-4}$). In addition, we detected a higher level of mannose ($P=4.45 \times 10^{-3}$; $FDR=1.40 \times 10^{-2}$), an important constituent of N-glycans of glycoproteins (69). Mannose residues in N-glycans can be derived from either glycogen/glucose or mannose in the blood (69). Moreover, we detected a reduced level of amino acids and their derivatives in the Metabotype IV. Among them, the reduction of histidine is the most significant ($P=3.72 \times 10^{-7}$; $FDR=1.52 \times 10^{-5}$). We noted an interesting clue of the possible role of microbiota in the inflammatory status of patients with PCa. We found desaminotyrosine, a metabolite produced by human enteric bacteria and able to induce type I interferon (IFN) response (70), prevalent in patients with very high aggressive PCa (**Figure 3**).

Similar to the previous analysis, we also investigated the rearrangement of the lipoprotein profile using the data from the application of the Bruker's B.I.-LISA protocols (**Table S2**). The Bruker's B.I.-LISA protocols were used to characterize 114 parameters related to the lipoproteins, such as HDL, LDL and VLDL, and their relative subclasses. We observed an association between the metabotype IV and lower level of Apolipoprotein Apo-A1 ($P=2.67 \times 10^{-3}$; $FDR=6.10 \times 10^{-2}$) and Apo-A2 ($P=4.04 \times 10^{-3}$; $FDR=5.32 \times 10^{-2}$), attributable to a reduced HDL particle number ($P=1.28 \times 10^{-2}$; $FDR=2.68 \times 10^{-1}$), and higher level of triglycerides in LDL of smaller size, including LDL-1 ($P=4.45 \times 10^{-3}$; $FDR=8.45 \times 10^{-2}$) and LDL-2 ($P=1.40 \times 10^{-3}$; $FDR=5.32 \times 10^{-2}$). This finding completes the snapshot of the metabotype IV as a large regulator of the blood constituents, including metabolites, proteins, and lipoproteins with several implications for the role of inflammation as a confounding factor of PCa.

The levels of inflammatory NMR markers GlycA and GlycB have been shown to be highly correlated, as previously reported in the literature (67). In our study, we reported a Spearman's rank correlation rho of 0.59 ($P=7.16 \times 10^{-5}$). However, the biological meaning of the differences between GlycA and GlycB has not yet been fully explored. In our cohort, we observed that two distinct metabotypes, Metabotype II and Metabotype III, had a similar level of GlycA but different levels of GlycB ($P=1.42 \times 10^{-4}$; $FDR=6.46 \times 10^{-4}$), with the latter showing the higher level (**Table S3**). Moreover, Metabotype III seems to be associated with extremely reduced levels of lipids (**Figure 3**). We discovered a deep difference in the lipoprotein profile between Metabotype III and Metabotype II using the B.I.-LISA protocols (**Table S4**). **Figure 4** shows a graphical representation of the lipoprotein profile changes among metabotypes.

Besides the evident reduction of VLDL, we also noted the lower values of Apolipoprotein Apo-A1 and Apo-A2 that could help to characterize the differences between GlycA and GlycB. Finally, we summarize in a graphical overview the most discriminative features among the four metabotypes (**Figure 5**).

Discussion

PCa is the most common visceral malignancy in men older than 50 years and shows significant ethnic disparity among men with African ancestry representing a well-established risk factor. In a recent study, all patients who underwent a prostate biopsy from July 2008 to July 2014 in one of the hospitals included in our cohort (Groote Schuur Hospital, Cape Town) were recorded (71). Among all patients diagnosed with PCa, 41% and 21% were classified as high and very high risk PCa (NCCN classification), respectively. Although the percentage of clinically advanced cases is already impressive if compared to American or European studies, the relative percentage of cases with very high aggressive PCa is even higher (33%) if we consider only the Black population. Additional research is necessary to understand how a different genomic landscape and inadequate cancer surveillance could contribute to these disparities.

In this first metabolomic study of PCa conducted on an African population, we profiled the plasma samples of men of different ancestry. We observed in men with very high aggressive PCa higher levels of the NMR inflammatory markers GlycA and GlycB, probably due to the increased concentration of positive APPs and the complexity of their glycan structures. Moreover, we noted a simultaneous reduction of the signal from protein source likely attributable to a reduction of the albumin level, a negative APP.

Interestingly, in our cohort we discovered four distinct metabotypes associated with the aggressiveness of PCa, each one characterized by a unique metabolic fingerprint. A metabotype identified as a subgroup of patients with very high aggressive PCa (that we named Metabotype IV) was characterized by the highest values of GlycA and GlycB and by deep changes of the plasma metabolome. We observed a lower level of histidine that could reinforce our hypothesis of inflammatory processes underlying the Metabotype IV. Indeed, histidine has been already associated with inflammatory processes and in particular, it has been negatively correlated with other inflammatory markers, such as IL-6 and CRP (72). In addition, it has been reported that inflammation may alter the lipoprotein profile as well, for example modulating the HDL functions (73-75) as we observed. Recently, it has been shown that high levels of triglycerides and glucose and low levels of HDL cholesterol and Apo-A1 are related to increased PCa risk and its severity (76). Moreover, low HDL was reported to be a risk and prognostic factor for PCa in several epidemiologic studies (77). Elevated serum triglycerides were associated with an increased risk of PCa recurrence (78). Lower levels of Apo-A1 and Apo-A2, and a higher level of triglycerides in LDL, reported in

this study, are consistent with these processes and we suggest that inflammation could be a driving factor of the lipoprotein profile changes observed in Metabotype IV.

Recently, there has been an increase in the efforts to understand the contribution of the microbiome on disease development and inflammation. Desaminotyrosine, a product of human enteric bacteria, enhances the clearance of respiratory virus by inducing type I IFN responses (70). In our study, we found elevated levels of desaminotyrosine in most very high aggressive PCa cases with Metabotype IV. This, therefore, may indicate that there are differences in the microbiome among PCa patients that may contribute to cancer progression since this signal was found elevated in the most aggressive PCa cases. Further studies are needed to elucidate the role of desaminotyrosine and the gut microbiome in men with PCa.

Furthermore, we highlighted two others distinct metabolotypes characterized by large differences in the lipoprotein profile. We noted a higher concentration of GlycB in the Metabotype III, which is characterized by a higher number of patients with very high aggressive PCa compared to the Metabotype II. No differences in the concentration of GlycA and higher levels of GlycB could be due to an elevated sialylation post-translational modification on glycosylated proteins. Complex biantennary glycoforms with α 2,3-sialic acid have been associated with aggressive PCa (79-81). Here, for the first time, we reported the association of lower level of the inflammatory NMR biomarker GlycB with a higher concentration of VLDL. Since the concentration of dietary intake can modulate the VLDL, this finding will further enrich the long-standing debate over the role of dietary fat in promoting PCa (82,83).

In men diagnosed with PCa, the selection of the treatment, including the type of therapy and its aggressiveness, is often based on patient age and life expectancy. In an era of precision medicine, an estimate of the threat of disease and the benefit and the costs of intervention within the context of the patient's characteristics and desires should be taken into consideration regarding the decision of the treatment. Life expectancy is a key factor when weighing the potential beneficial effects of various treatment options, but for life expectancy estimates to be accurate, chronological age must not be the sole or primary factor considered (84).

Local definitive therapy by external beam radiation combined with androgen deprivation is supported by several randomized clinical trials whereas the role of surgery in the very high-risk setting combined with adjuvant radiation / ADT is emerging (85). Growing evidence suggests neoadjuvant taxane-based chemotherapy in the context of a multimodal approach may be beneficial (85). Treatment of high-risk prostate cancer has evolved considerably over the past two decades, yet patients with very high-risk features may still experience poor outcomes despite aggressive therapy.

In this study, we identified a set of patients with very high aggressive PCa with extremely reduced survival time. These patients, belonging to the Metabotype IV, are characterized by a similar metabolic profile predictable with high accuracy. Our results postulate that this subgroup may be most likely to benefit from combination therapy

that associates the androgen deprivation in conjunction with drugs aiming to reduce the level of systemic inflammation.

Corticosteroids, such as prednisone, have been used in the treatment of metastatic CRPC (mCRPC) for more than three decades, particularly, to treat inflammation and pain related to bone metastasis (86). Significant improvement in overall survival in patients has been associated with abiraterone acetate plus low-dose prednisone treatment in both chemotherapy-naïve and chemotherapy-treated patients with mCRPC (87,88). Recently, significant benefits of adding abiraterone acetate plus prednisone to ADT has been proven in high risk metastatic castration-sensitive PCa (mCSPC), by further prolonging the overall survival (89). In non-metastatic PCa, a major reason for the limitation of daily corticosteroids was concerns regarding the high cumulative toxicities associated with long-term use (90). In localized high risk PCa, a notable lowering of prostate tissue androgens has been associated with the addition of abiraterone acetate plus prednisone to neoadjuvant luteinizing hormone-releasing hormone (LHRH) agonists when compared with LHRH agonists alone (91). However, the only clinical settings of corticosteroids treatment with proved clinical utility in PCa treatment is in combination with abiraterone.

In this context, metabolomics could represent an invaluable tool for the stratification of patients with very high aggressive PCa. The life expectancy difference highlights the need to consider an appropriate medical treatment for patients within Metabotype IV. We hypothesize that these patients could largely benefit from daily treatment with corticosteroids to reduce the systemic inflammation improving the overall survival, along with the need for subsequent therapy. We consider the lack of clinical investigation for almost all patients of the presence of distant metastasis as a limitation of this study. Considering the clinical and histopathological features, we are aware that PCa could have spread to distant organs in the patients with very high risk PCa belonging to the Metabotype IV.

Integrating a metabolomic analysis as a tool for patient stratification could be important to identify those patients with very high risk PCa and short life expectancy, as they may benefit from therapeutic interventions, targeting the lowering of systemic inflammation. Thus, further studies are necessary to better characterize this group of patients and determine the costs and benefits of corticosteroid treatment in terms of survival time and quality of life.

In conclusion, this study answers the urgent need for new insights into the molecular mechanisms underlying the remarkably increased rate of aggressive and lethal PCa in men of African ancestry. New non-invasive metabolite biomarkers are necessary to improve the treatment decision, which will improve therapeutic outcomes in African patients.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Grant Support

Support by The International Centre for Genetic Engineering and Biotechnology, ICGEB (LFZ); ICGEB Arturo Falaschi fellowship (SC); Joint Research Grant South Africa/Switzerland Research Partnership Programme Bilateral Agreement (LFZ, CVC, GMC); The South African National Research Foundation (NRF) for Professional Development Programme (MW); NIH/NCI P30 CA006516 (TAL)

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians* 2018;68:394-424
2. Amundadottir LT, Sulem P, Gudmundsson J, Helgason A, Baker A, Agnarsson BA, et al. A common variant associated with prostate cancer in European and African populations. *Nat Genet* 2006;38:652-8
3. Crawford ED. Epidemiology of prostate cancer. *Urology* 2003;62:3-12
4. Powell IJ. The precise role of ethnicity and family history on aggressive prostate cancer: a review analysis. *Arch Esp Urol* 2011;64:711-9
5. Huang FW, Mosquera JM, Garofalo A, Oh C, Baco M, Amin-Mansour A, et al. Exome Sequencing of African-American Prostate Cancer Reveals Loss-of-Function ERF Mutations. *Cancer Discov* 2017;7:973-83
6. Taksler GB, Keating NL, Cutler DM. Explaining racial differences in prostate cancer mortality. *Cancer* 2012;118:4280-9
7. Khani F, Mosquera JM, Park K, Blattner M, O'Reilly C, MacDonald TY, et al. Evidence for molecular differences in prostate cancer between African American and Caucasian men. *Clin Cancer Res* 2014;20:4925-34
8. Petrovics G, Li H, Stumpel T, Tan SH, Young D, Katta S, et al. A novel genomic alteration of LSAMP associates with aggressive prostate cancer in African American men. *EBioMedicine* 2015;2:1957-64
9. Lindquist KJ, Paris PL, Hoffmann TJ, Cardin NJ, Kazma R, Mefford JA, et al. Mutational Landscape of Aggressive Prostate Tumors in African American Men. *Cancer Res* 2016;76:1860-8
10. Cher ML, Lewis PE, Banerjee M, Hurley PM, Sakr W, Grignon DJ, et al. A similar pattern of chromosomal alterations in prostate cancers from African-Americans and Caucasian Americans. *Clin Cancer Res* 1998;4:1273-8
11. Rose AE, Satagopan JM, Oddoux C, Zhou Q, Xu R, Olshen AB, et al. Copy number and gene expression differences between African American and Caucasian American prostate cancer. *J Transl Med* 2010;8:70
12. Bryc K, Auton A, Nelson MR, Oksenberg JR, Hauser SL, Williams S, et al. Genome-wide patterns of population structure and admixture in West Africans and African Americans. *Proceedings of the National Academy of Sciences* 2010;107:786-91
13. Tindall EA, Monare LR, Petersen DC, Van Zyl S, Hardie RA, Segone AM, et al. Clinical presentation of prostate cancer in black South Africans. 2014;74:880-91

14. Jaratlerdsiri W, Chan EK, Gong T, Petersen DC, Kalsbeek AM, Venter PA, et al. Whole-Genome Sequencing Reveals Elevated Tumor Mutational Burden and Initiating Driver Mutations in African Men with Treatment-Naïve, High-Risk Prostate Cancer. *2018;78:6736-46*
15. Dechambenoit G. Access to health care in sub-Saharan Africa. *Surg Neurol Int 2016;7:108*
16. Choudhury A, Ramsay M, Hazelhurst S, Aron S, Bardien S, Botha G, et al. Whole-genome sequencing for an enhanced understanding of genetic variation among South Africans. *Nature communications 2017;8:1-12*
17. Patterson N, Petersen DC, van der Ross RE, Sudoyo H, Glashoff RH, Marzuki S, et al. Genetic structure of a unique admixed population: implications for medical research. *Human Molecular Genetics 2010;19:411-9*
18. de Wit E, Delport W, Rugamika CE, Meintjes A, Möller M, van Helden PD, et al. Genome-wide analysis of the structure of the South African Coloured Population in the Western Cape. *Human genetics 2010;128:145-53*
19. Blattner M, Lee DJ, O'Reilly C, Park K, MacDonald TY, Khani F, et al. SPOP mutations in prostate cancer across demographically diverse patient cohorts. *Neoplasia 2014;16:14-20*
20. Rand KA, Rohland N, Tandon A, Stram A, Sheng X, Do R, et al. Whole-exome sequencing of over 4100 men of African ancestry and prostate cancer risk. *Hum Mol Genet 2016;25:371-81*
21. Devaney JM, Wang S, Furbert-Harris P, Apprey V, Ittmann M, Wang BD, et al. Genome-wide differentially methylated genes in prostate cancer tissues from African-American and Caucasian men. *Epigenetics 2015;10:319-28*
22. Dreussi E, Ecça F, Scarabel L, Gagno S, Toffoli G. Immunogenetics of prostate cancer: a still unexplored field of study. *Pharmacogenomics 2018;19:263-83*
23. Winchester DA, Till C, Goodman PJ, Tangen CM, Santella RM, Johnson-Pais TL, et al. Variation in genes involved in the immune response and prostate cancer risk in the placebo arm of the Prostate Cancer Prevention Trial. *The Prostate 2015;75:1403-18*
24. Stark T, Livas L, Kyprianou N. Inflammation in prostate cancer progression and therapeutic targeting. *Transl Androl Urol 2015;4:455-63*
25. Batai K, Murphy AB, Nonn L, Kittles RA. Vitamin D and immune response: implications for prostate cancer in African Americans. *Frontiers in immunology 2016;7:53*
26. Fernandez P, de Beer PM, van der Merwe L, Heyns CFJC. COX-2 promoter polymorphisms and the association with prostate cancer risk in South African men. *2008;29:2347-50*
27. Eastham JA, May RA, Whatley T, Crow A, Venable DD, Sartor O. Clinical characteristics and biopsy specimen features in African-American and white men without prostate cancer. *J Natl Cancer Inst 1998;90:756-60*

28. Wallace TA, Prueitt RL, Yi M, Howe TM, Gillespie JW, Yfantis HG, et al. Tumor immunobiological differences in prostate cancer between African-American and European-American men. *Cancer Res* 2008;68:927-36
29. Chechlinska M, Kowalewska M, Nowak RJNRC. Systemic inflammation as a confounding factor in cancer biomarker discovery and validation. 2010;10:2
30. Guleria A, Pratap A, Dubey D, Rawat A, Chaurasia S, Sukesh E, et al. NMR based serum metabolomics reveals a distinctive signature in patients with Lupus Nephritis. *Sci Rep* 2016;6:35309
31. Jiang M, Chen T, Feng H, Zhang Y, Li L, Zhao A, et al. Serum metabolic signatures of four types of human arthritis. *J Proteome Res* 2013;12:3769-79
32. Ouyang X, Dai Y, Wen JL, Wang LX. (1)H NMR-based metabolomic study of metabolic profiling for systemic lupus erythematosus. *Lupus* 2011;20:1411-20
33. Alonso A, Julia A, Vinaixa M, Domenech E, Fernandez-Nebro A, Canete JD, et al. Urine metabolome profiling of immune-mediated inflammatory diseases. *BMC Med* 2016;14:133
34. Tenori L, Oakman C, Claudino WM, Bernini P, Cappadona S, Nepi S, et al. Exploration of serum metabolomic profiles and outcomes in women with metastatic breast cancer: a pilot study. *Mol Oncol* 2012;6:437-44
35. Oakman C, Tenori L, Claudino WM, Cappadona S, Nepi S, Battaglia A, 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;22:1295-301
36. Tenori L, Oakman C, Morris PG, Gralka E, Turner N, Cappadona S, et al. Serum metabolomic profiles evaluated after surgery may identify patients with oestrogen receptor negative early breast cancer at increased risk of disease recurrence. Results from a retrospective study. *Mol Oncol* 2015;9:128-39
37. Bertini I, Calabro A, De Carli V, Luchinat C, Nepi S, Porfirio B, et al. The metabonomic signature of celiac disease. *J Proteome Res* 2009;8:170-7
38. Bray R, Cacciatore S, Jimenez B, Cartwright R, Digesu A, Fernando R, et al. Urinary Metabolic Phenotyping of Women with Lower Urinary Tract Symptoms. *J Proteome Res* 2017;16:4208-16
39. Romano F, Meoni G, Manavella V, Baima G, Tenori L, Cacciatore S, et al. Analysis of salivary phenotypes of generalized aggressive and chronic periodontitis through nuclear magnetic resonance-based metabolomics. *J Periodontol* 2018;89:1452-60
40. Cacciatore S, Loda M. Innovation in metabolomics to improve personalized healthcare. *Ann N Y Acad Sci* 2015;1346:57-62
41. Vignoli A, Ghini V, Meoni G, Licari C, Takis PG, Tenori L, et al. High-throughput metabolomics by 1D NMR. 2019;58:968-94
42. Otvos JD, Shalaurova I, Wolak-Dinsmore J, Connelly MA, Mackey RH, Stein JH, et al. GlycA: A Composite Nuclear Magnetic Resonance Biomarker of Systemic Inflammation. *Clin Chem* 2015;61:714-23

43. Bell JD, Brown JC, Nicholson JK, Sadler PJ. Assignment of resonances for 'acute-phase' glycoproteins in high resolution proton NMR spectra of human blood plasma. *FEBS Lett* 1987;215:311-5
44. Muhlestein JB, May H, Winegar D, Rollo J, Connelly M, Otvos J, et al. GlycA and GlycB, novel NMR biomarkers of inflammation, strongly predict future cardiovascular events, but not the presence of coronary artery disease (CAD), among patients undergoing coronary angiography: the intermountain heart collaborative study. 2014;63:A1389
45. Gruppen EG, Connelly MA, Otvos JD, Bakker SJ, Dullaart RP. A novel protein glycan biomarker and LCAT activity in metabolic syndrome. *Eur J Clin Invest* 2015;45:850-9
46. Ritchie SC, Würtz P, Nath AP, Abraham G, Havulinna AS, Fearnley LG, et al. The biomarker GlycA is associated with chronic inflammation and predicts long-term risk of severe infection. 2015;1:293-301
47. Gruppen EG, Connelly MA, Dullaart RP. Higher circulating GlycA, a pro-inflammatory glycoprotein biomarker, relates to lipoprotein-associated phospholipase A2 mass in nondiabetic subjects but not in diabetic or metabolic syndrome subjects. *J Clin Lipidol* 2016;10:512-8
48. Ritchie SC, Würtz P, Nath AP, Abraham G, Havulinna AS, Kangas AJ, et al. Systems medicine links microbial inflammatory response with glycoprotein-associated mortality risk. 2015:018655
49. Gabay C, Kushner I. Acute-phase proteins and other systemic responses to inflammation. *N Engl J Med* 1999;340:448-54
50. Gornik O, Lauc G. Glycosylation of serum proteins in inflammatory diseases. *Dis Markers* 2008;25:267-78
51. van Dijk W, Turner GA, Mackiewicz AJG, Disease. Changes in glycosylation of acute-phase proteins in health and disease: occurrence, regulation and function. 1994;1:5-14
52. Ceciliani F, Pocacqua VJCP, Science P. The acute phase protein α 1-acid glycoprotein: a model for altered glycosylation during diseases. 2007;8:91-108
53. Bernini P, Bertini I, Luchinat C, Nincheri P, Staderini S, Turano P. Standard operating procedures for pre-analytical handling of blood and urine for metabolomic studies and biobanks. *J Biomol NMR* 2011;49:231-43
54. Jiménez B, Holmes E, Heude C, Tolson RF, Harvey N, Lodge SL, et al. Quantitative lipoprotein subclass and low molecular weight metabolite analysis in human serum and plasma by ^1H NMR spectroscopy in a multilaboratory trial. 2018;90:11962-71
55. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, et al. HMDB 3.0—the human metabolome database in 2013. *Nucleic acids research* 2012;41:D801-D7
56. Marshall I, Higinbotham J, Bruce S, Freise A. Use of voigt lineshape for quantification of in vivo ^1H spectra. *Magn Reson Med* 1997;37:651-7

57. Team RC. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>. 2018.
58. Bauer DF. Constructing confidence sets using rank statistics. *Journal of the American Statistical Association* 1972;67:687-90
59. Clarkson DB, Fan Y-A, Joe H. A remark on algorithm 643: FEXACT: An algorithm for performing Fisher's exact test in rxc contingency tables. *ACM Transactions on Mathematical Software (TOMS)* 1993;19:484-8
60. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics* 1987;20:53-65
61. Storey JD. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2002;64:479-98
62. Bertini I, Cacciatore S, Jensen BV, Schou JV, Johansen JS, Kruhoffer M, et al. Metabolomic NMR fingerprinting to identify and predict survival of patients with metastatic colorectal cancer. *Cancer Res* 2012;72:356-64
63. Eriksson L, Jaworska J, Worth AP, Cronin MT, McDowell RM, Gramatica P. Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. *Environ Health Perspect* 2003;111:1361-75
64. Sciarra A, Gentilucci A, Salciccia S, Pierella F, Del Bianco F, Gentile V, et al. Prognostic value of inflammation in prostate cancer progression and response to therapeutic: a critical review. *J Inflamm (Lond)* 2016;13:35
65. Ammirante M, Luo JL, Grivennikov S, Nedospasov S, Karin M. B-cell-derived lymphotoxin promotes castration-resistant prostate cancer. *Nature* 2010;464:302-5
66. Fuertes-Martin R, Correig X, Vallve JC, Amigo N. Title: Human Serum/Plasma Glycoprotein Analysis by (1)H-NMR, an Emerging Method of Inflammatory Assessment. *J Clin Med* 2020;9
67. Lorenzo C, Festa A, Hanley AJ, Rewers MJ, Escalante A, Haffner SM. Novel Protein Glycan-Derived Markers of Systemic Inflammation and C-Reactive Protein in Relation to Glycemia, Insulin Resistance, and Insulin Secretion. *Diabetes Care* 2017;40:375-82
68. Reynolds AP, Richards G, de la Iglesia B, Rayward-Smith VJ. Clustering rules: a comparison of partitioning and hierarchical clustering algorithms. *Journal of Mathematical Modelling and Algorithms* 2006;5:475-504
69. Sharma V, Freeze HH. Mannose Efflux from the Cells, a potential source of mannose in blood. *Journal of Biological Chemistry* 2011;286:10193-200
70. Steed AL, Christophi GP, Kaiko GE, Sun L, Goodwin VM, Jain U, et al. The microbial metabolite desaminotyrosine protects from influenza through type I interferon. 2017;357:498-502
71. Dewar M, Kaestner L, Zikhali Q, Jehle K, Sinha S, Lazarus J. Investigating racial differences in clinical and pathological features of prostate cancer in South African men. *South African Journal of Surgery* 2018;56:54-8

72. Niu YC, Feng RN, Hou Y, Li K, Kang Z, Wang J, et al. Histidine and arginine are associated with inflammation and oxidative stress in obese women. *Br J Nutr* 2012;108:57-61
73. McGarrah RW, Kelly JP, Craig DM, Haynes C, Jessee RC, Huffman KM, et al. A Novel Protein Glycan-Derived Inflammation Biomarker Independently Predicts Cardiovascular Disease and Modifies the Association of HDL Subclasses with Mortality. *Clin Chem* 2017;63:288-96
74. Kim KI, Oh SW, Ahn S, Heo NJ, Kim S, Chin HJ, et al. CRP level and HDL cholesterol concentration jointly predict mortality in a Korean population. *Am J Med* 2012;125:787-95 e4
75. Zewinger S, Drechsler C, Kleber ME, Dressel A, Riffel J, Triem S, et al. Serum amyloid A: high-density lipoproteins interaction and cardiovascular risk. *Eur Heart J* 2015;36:3007-16
76. Arthur R, Møller H, Garmo H, Holmberg L, Stattin P, Malmstrom H, et al. Association between baseline serum glucose, triglycerides and total cholesterol, and prostate cancer risk categories. 2016;5:1307-18
77. Kotani K, Sekine Y, Ishikawa S, Ikpot IZ, Suzuki K, Remaley AT. High-density lipoprotein and prostate cancer: an overview. *J Epidemiol* 2013;23:313-9
78. Allott EH, Howard LE, Cooperberg MR, Kane CJ, Aronson WJ, Terris MK, et al. Serum lipid profile and risk of prostate cancer recurrence: Results from the SEARCH database. *Cancer Epidemiol Biomarkers Prev* 2014;23:2349-56
79. Llop E, Ferrer-Batallé M, Barrabés S, Guerrero PE, Ramírez M, Saldova R, et al. Improvement of prostate cancer diagnosis by detecting PSA glycosylation-specific changes. *Theranostics* 2016;6:1190
80. Ferrer-Batallé M, Llop E, Ramírez M, Aleixandre RN, Saez M, Comet J, et al. Comparative study of blood-based biomarkers, $\alpha 2$, 3-sialic acid PSA and PHI, for high-risk prostate cancer detection. *International journal of molecular sciences* 2017;18:845
81. Ishikawa T, Yoneyama T, Tobisawa Y, Hatakeyama S, Kurosawa T, Nakamura K, et al. An automated micro-total immunoassay system for measuring cancer-associated $\alpha 2$, 3-linked sialyl N-glycan-carrying prostate-specific antigen may improve the accuracy of prostate cancer diagnosis. *International journal of molecular sciences* 2017;18:470
82. Labbé DP, Zadra G, Yang M, Reyes JM, Lin CY, Cacciatore S, et al. High-fat diet fuels prostate cancer progression by rewiring the metabolome and amplifying the MYC program. *Nature communications* 2019;10:1-14
83. Kristal AR, Arnold KB, Neuhaus ML, Goodman P, Platz EA, Albanes D, et al. Diet, supplement use, and prostate cancer risk: results from the prostate cancer prevention trial. *American journal of epidemiology* 2010;172:566-77
84. Daskivich TJ. The Importance of Accurate Life Expectancy Prediction in Men with Prostate Cancer. *Eur Urol* 2015;68:766-7
85. Mano R, Eastham J, Yossepowitch O. The very-high-risk prostate cancer: a contemporary update. *Prostate cancer and prostatic diseases* 2016;19:340-8

86. Haywood A, Duc J, Good P, Khan S, Rickett K, Vayne-Bossert P, et al. Systemic corticosteroids for the management of cancer-related breathlessness (dyspnoea) in adults. *Cochrane Database Syst Rev* 2019;2:CD012704
87. Ryan CJ, Smith MR, Fizazi K, Saad F, Mulders PF, Sternberg CN, et al. Abiraterone acetate plus prednisone versus placebo plus prednisone in chemotherapy-naïve men with metastatic castration-resistant prostate cancer (COU-AA-302): final overall survival analysis of a randomised, double-blind, placebo-controlled phase 3 study. *Lancet Oncol* 2015;16:152-60
88. Fizazi K, Scher HI, Molina A, Logothetis CJ, Chi KN, Jones RJ, et al. Abiraterone acetate for treatment of metastatic castration-resistant prostate cancer: final overall survival analysis of the COU-AA-301 randomised, double-blind, placebo-controlled phase 3 study. *Lancet Oncol* 2012;13:983-92
89. Fizazi K, Tran N, Fein L, Matsubara N, Rodriguez-Antolin A, Alekseev BY, et al. Abiraterone acetate plus prednisone in patients with newly diagnosed high-risk metastatic castration-sensitive prostate cancer (LATITUDE): final overall survival analysis of a randomised, double-blind, phase 3 trial. *Lancet Oncol* 2019;20:686-700
90. Ndibe C, Wang CG, Sonpavde G. Corticosteroids in the management of prostate cancer: a critical review. *Curr Treat Options Oncol* 2015;16:6
91. Taplin ME, Montgomery B, Logothetis CJ, Bubley GJ, Richie JP, Dalkin BL, et al. Intense androgen-deprivation therapy with abiraterone acetate plus leuprolide acetate in patients with localized high-risk prostate cancer: results of a randomized phase II neoadjuvant study. *J Clin Oncol* 2014;32:3705-15

Tables

Table 1. Clinical demographics of PCa patients

Feature	Treatment-naïve				post-BO			
	intermediate, low and very low (n=14)	high (n=7)	very high (n=12)	metastatic (n=1)	Total (n=34)	non-CRPC (n=4)	CRPC (n=3)	Total (n=7)
Age (year), median [95%CI]	65 [56 77]	70 [52 90]	64 [57 86]	75 [75 75]	68 [63 74]	72 [63 73]	66 [64 70]	70 [65 72]
Ancestry, n (%)								
Black	1 (7.1)	1 (14.3)	4 (33.3)	1 (100.0)	7 (20.6)	1 (25.0)	1 (33.3)	2 (28.6)
Coloured	10 (71.4)	5 (71.4)	6 (50.0)	0 (0.0)	21 (61.8)	2 (50.0)	1 (33.3)	3 (42.8)
Coloured/Black	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (25.0)	0 (0.0)	1 (14.3)
White	3 (21.4)	1 (14.3)	2 (16.7)	0 (0.0)	6 (17.6)	0 (0.0)	1 (33.3)	1 (14.3)
PSA (ng/mL), median [95%CI]	9 [3 19]	23 [17 31]	138 [29 3919]	>5000	21 [12 84]	4 [2 40]	332 [49 1128]	34 [4 188]
Diabetes, n (%)								
no	10 (71.4)	5 (71.4)	12 (100.0)	1 (100.0)	28 (82.4)	3 (75.0)	2 (66.7)	5 (71.4)
yes	4 (28.6)	2 (28.6)	0 (0.0)	0 (0.0)	6 (17.6)	1 (25.0)	1 (33.3)	2 (28.6)
Hypertension, n (%)								
no	7 (50.0)	5 (71.4)	10 (83.3)	1 (100.0)	23 (67.6)	3 (75.0)	1 (33.3)	4 (57.1)
yes	7 (50.0)	2 (28.6)	2 (16.7)	0 (0.0)	11 (32.4)	1 (25.0)	2 (66.7)	3 (42.9)
Smoker, n (%)								
no	10 (71.4)	6 (85.7)	9 (75.0)	1 (100.0)	26 (76.5)	4 (100.0)	1 (33.3)	5 (71.4)
yes	4 (28.6)	1 (14.3)	3 (25.0)	0 (0.0)	8 (23.5)	0 (0.0)	2 (66.7)	2 (28.6)

Table 2. Clinical and demographic features of the metabotypes identified through KODAMA analysis.

Feature	Metabotype I	Metabotype II	Metabotype III	Metabotype IV
NCCN classification, n (%)				
very low	1 (33.3)	0 (0.0)	1 (7.1)	0 (0.0)
low	0 (0.0)	1 (10.0)	1 (7.1)	1 (14.3)
intermediate	2 (66.7)	4 (40.0)	3 (21.4)	0 (0.0)
high	0 (0.0)	3 (30.0)	4 (28.5)	0 (0.0)
very high	0 (0.0)	2 (20.0)	5 (35.7)	5 (71.4)
metastatic	0 (0.0)	0 (0.0)	0 (0.0)	1 (14.3)
Gleason Score, n (%)				
3+3	1 (33.3)	1 (10.0)	5 (35.7)	1 (14.3)
3+4	1 (33.3)	4 (40.0)	6 (42.9)	0 (0.0)
4+3	1 (33.3)	1 (10.0)	0 (0.0)	0 (0.0)
3+5	0 (0.0)	0 (0.0)	0 (0.0)	1 (14.3)
4+5	0 (0.0)	4 (40.0)	3 (21.4)	0 (0.0)
5+4	0 (0.0)	0 (0.0)	0 (0.0)	2 (28.6)
5+5	0 (0.0)	0 (0.0)	0 (0.0)	3 (42.9)
Age, median [95%CI]	71 [64 74]	68 [56 90]	65 [51 85]	74 [59 78]
Ancestry, n(%)				
Black	1 (33.3)	1 (10.0)	2 (14.3)	3 (42.9)
Mixed ancestry	1 (33.3)	7 (70.0)	11 (78.6)	2 (28.6)
White	1 (33.3)	2 (20.0)	1 (7.1)	2 (28.6)
PSA, median [95%CI]	9 [3 9]	18 [5 233]	25 [5 126]	738 [26 5000]
Diabetes, n (%)				
no	1 (33.3)	7 (70.0)	13 (92.9)	7 (100.0)
yes	2 (66.7)	3 (30.0)	1 (7.1)	0 (0.0)
Hypertension, n (%)				
no	0 (0.0)	8 (80.0)	9 (64.3)	6 (85.7)
yes	3 (100.0)	2 (20.0)	5 (35.7)	1 (14.3)
Smoker, n (%)				
no	2 (66.7)	8 (80.0)	9 (64.3)	7 (100.0)
yes	1 (33.3)	2 (20.0)	5 (35.7)	0 (0.0)

Table 3. Demographics and clinical features of patients belonging to the Metabotype IV and patients classified as very high aggressiveness that do not belong to the Metabotype IV.

Patient ID	Ethnicity	Collection date	Date of last visit	Date of death	Follow-up (day) ^c	Age (year)	PSA (ng/mL)	Gleeson Score	DRE	NCCN classification	GlycB rank ^d	Metabotype	Note
SAPC0159	Black	2017/03/01	2017/03/23		22	75.2	>5000	5+5	T3/T4	metastatic	3	IV	
SAPC0090	Coloured	2015/05/08		2015/06/27	50	78.2	6.3	3+3	T1c	low	8	IV	a
SAPC0192	Coloured	2017/07/31		2017/12/19	141	80.9	41.68	4+5	T2a	very high	7	III	b
SAPC0080	Black	2014/10/31	2015/04/21		172	77.8	738	3+5	T4	very high	2	IV	
SAPC0180	Black	2017/06/23	2017/12/12		172	58.1	>5000	5+5	T3/T4	very high	1	IV	
SAPC0249	Coloured	2018/03/25	2019/01/07		288	63	1070	5+4	T3	very high	4	IV	
SAPC0078	White	2014/11/07		2015/10/31	358	74.3	135.79	5+5	T4	very high	5	IV	
SAPC0193	Coloured	2017/08/11	2019/04/26		623	63.5	48.85	3+4	T3	very high	9	III	
SAPC0070	Coloured	2014/01/08		2015/12/11	702	65.4	34.8	4+5	T2b	very high	13	II	
SAPC0191	White	2017/07/28		2019/08/03	736	61.8	576	5+4	T3	very high	6	IV	
SAPC0195	Black	2017/08/13	2020/01/30		900	87.2	26.53	4+5	T3	very high	11	III	
SAPC0120	Black	2016/10/03	2019/09/30		1092	82.3	289.9	4+5	T3	very high	14	II	
SAPC0108	Coloured	2016/06/27		2020/04/17	1390	56.6	96.3	3+4	T4	very high	12	III	
SAPC0076	Coloured	2014/09/26	2020/06/04		2078	63.5	140.42	3+4	T3/T4	very high	10	III	

Figures

Figure 1. GlycA concentration in (A) treatment-naïve and (B) post-treatment. (C) Correlation between GlycA and PSA. GlycB concentration in (D) treatment-naïve and (E) post-treatment. (F) Correlation between GlycB and PSA.

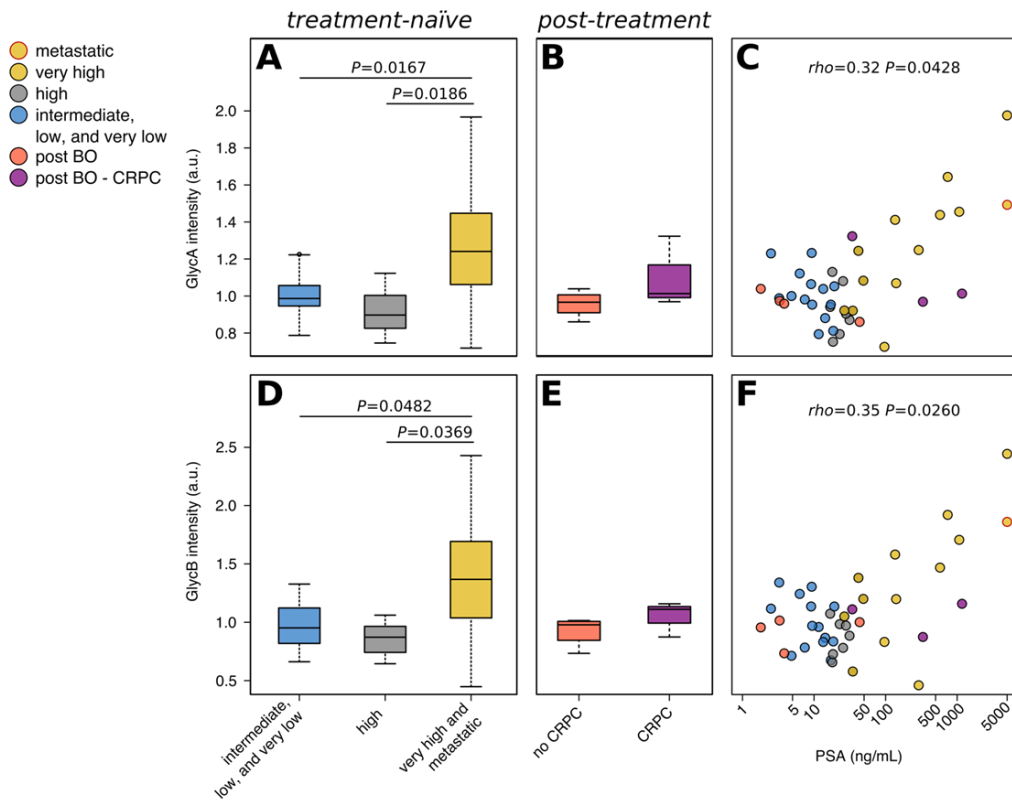


Figure 2. KODAMA score plot of plasma PCa samples colored according to NCCN classification. The size is proportional to (A) the GlycA intensity and (B) the GlycB intensity. (C) Heatmap of the metabolic profiles.

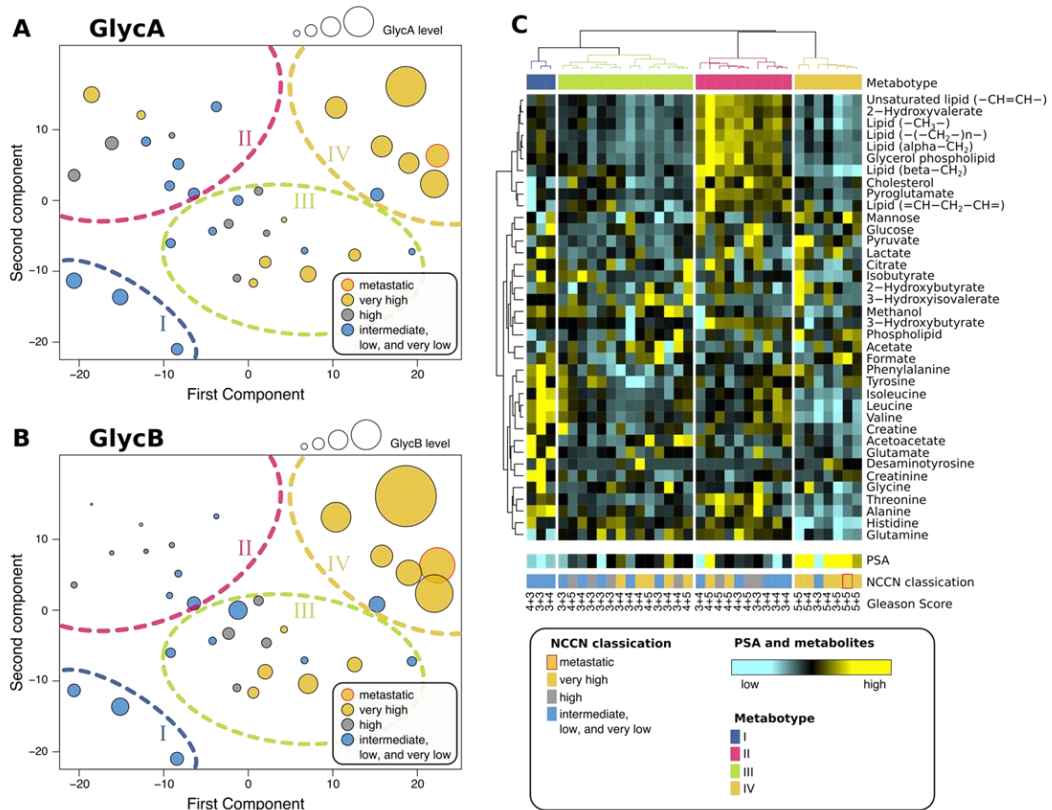


Figure 3. NMR profiles of the plasma of three different spectral regions.

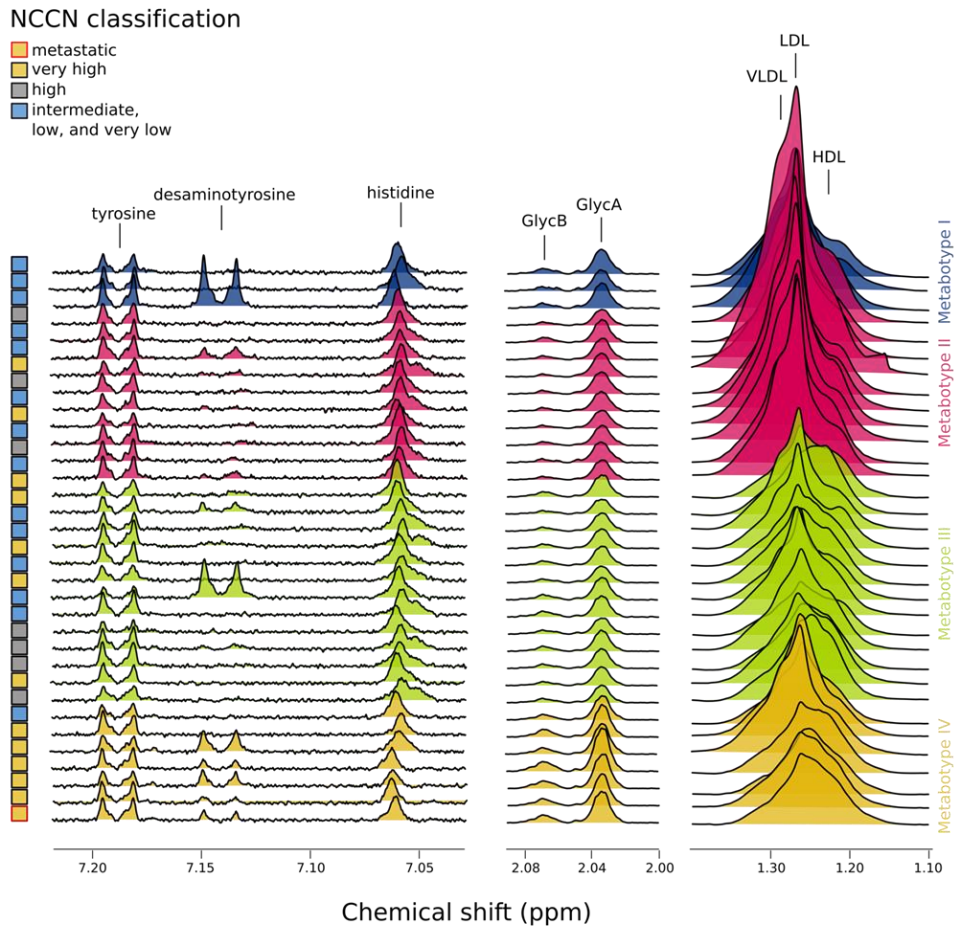


Figure 4. Box-whiskers plots of the intensity of (A) GlycA and (B) GlycB, the concentration of (C) histidine and the number of (D) VLDL particles across the four metabolotypes.

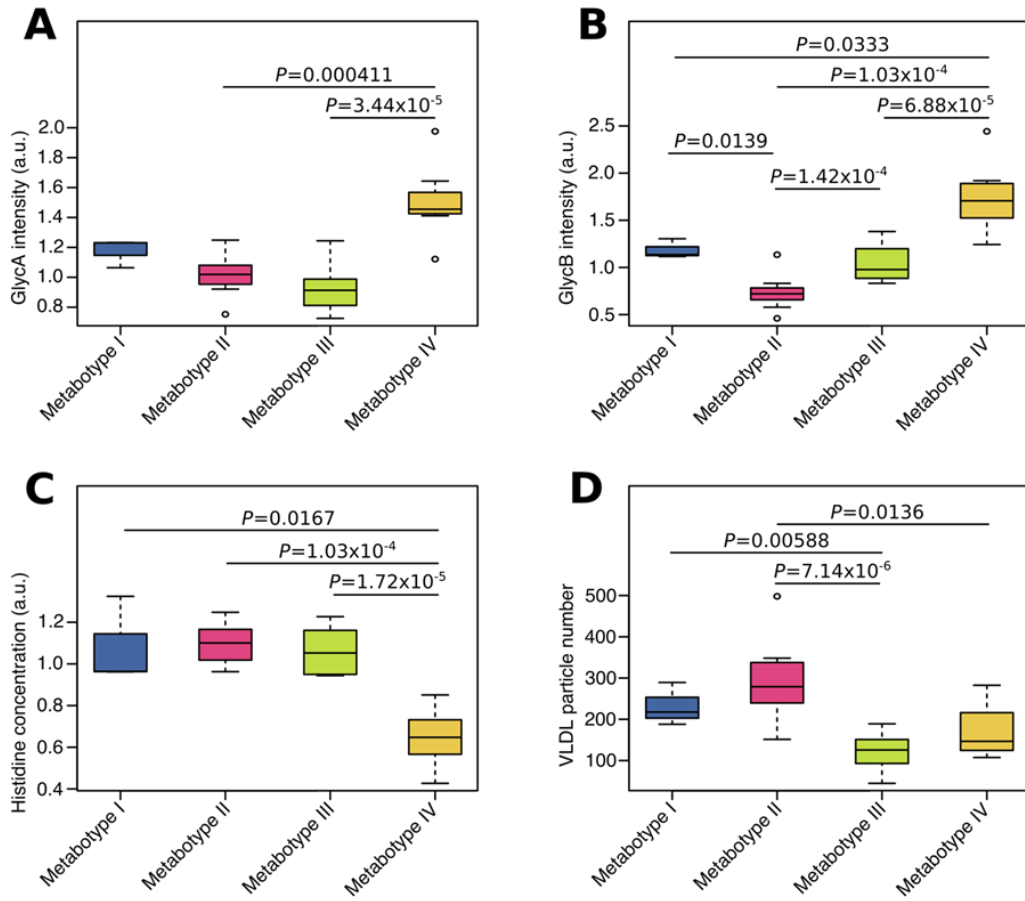
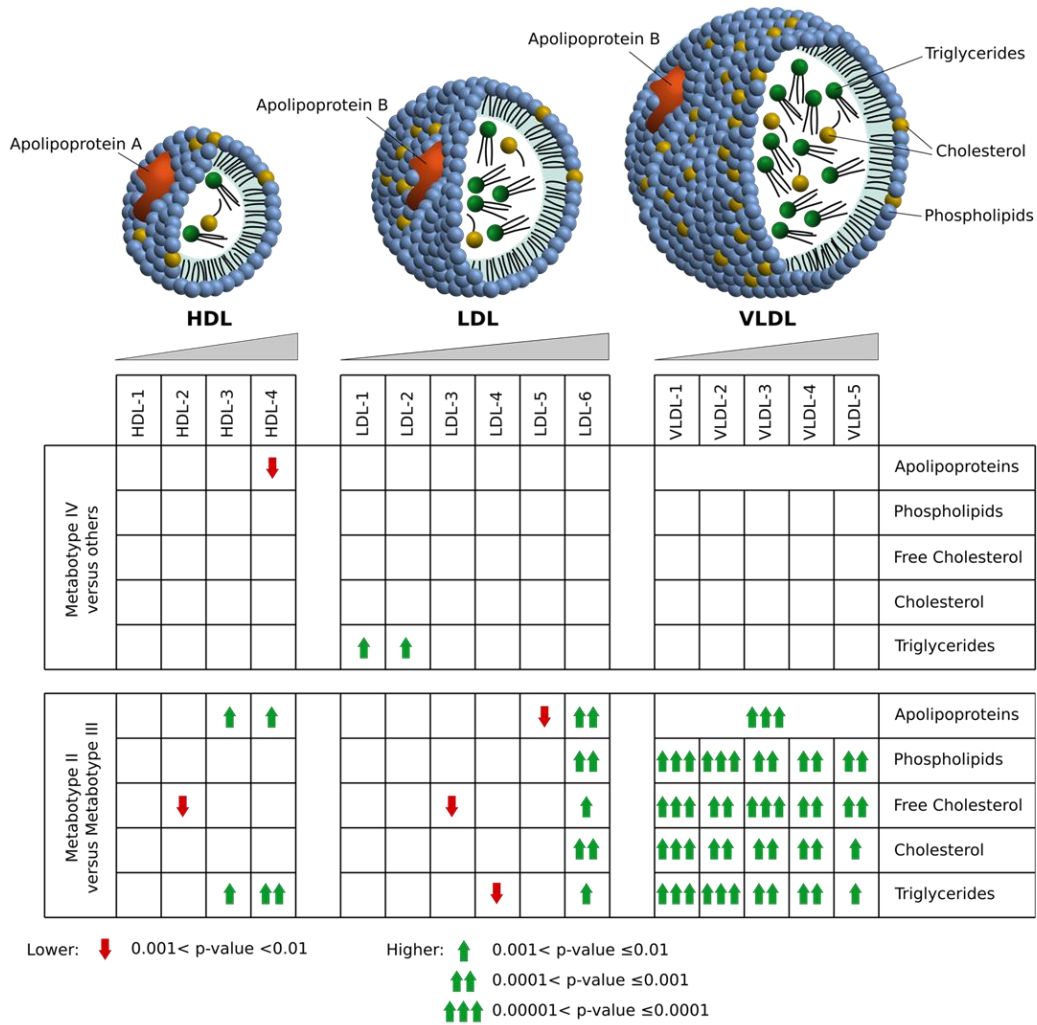


Figure 5. Graphics illustration of the changes of the lipoprotein profile.



Supplementary Material**Table S1.** Statistical comparison between the metabolic profiles of Metabotype IV *versus* the others.

Feature	Metabotype IV, median [IQR]	Others, median [IQR]	log change	P-value	FDR
Histidine	0.647 [0.566 0.732]	1.067 [0.981 1.163]	-0.74	3.72x10 ⁻⁷	1.52x10 ⁻⁵
Lipid (beta-CH2)	0.743 [0.6 0.785]	1.12 [0.989 1.239]	-0.69	7.44x10 ⁻⁷	1.52x10 ⁻⁵
GlycB	1.706 [1.524 1.89]	0.959 [0.782 1.126]	0.89	2.60x10 ⁻⁶	3.56x10 ⁻⁵
GlycA	1.455 [1.425 1.568]	0.981 [0.892 1.075]	0.61	7.06x10 ⁻⁶	7.24x10 ⁻⁵
Valine	0.62 [0.528 0.668]	1.002 [0.915 1.244]	-0.77	6.73x10 ⁻⁵	4.60x10 ⁻⁴
Leucine	0.624 [0.487 0.764]	1.065 [0.878 1.282]	-0.81	6.73x10 ⁻⁵	4.60x10 ⁻⁴
Protein	0.9 [0.87 0.939]	1.021 [0.979 1.071]	-0.19	9.15x10 ⁻⁵	5.36x10 ⁻⁴
Alanine	0.706 [0.578 0.774]	1.075 [0.89 1.234]	-0.69	2.10x10 ⁻⁴	1.07x10 ⁻³
Isoleucine	0.616 [0.548 0.811]	1.076 [0.896 1.224]	-0.78	3.45x10 ⁻⁴	1.57x10 ⁻³
Glutamine	0.842 [0.747 0.976]	1.098 [1.007 1.203]	-0.4	1.54x10 ⁻³	6.33x10 ⁻³
Lipid (-CH3-)	0.868 [0.826 0.946]	1.062 [0.948 1.189]	-0.28	2.67x10 ⁻³	9.97x10 ⁻³
Mannose	1.368 [1.212 1.497]	1.011 [0.754 1.173]	0.46	4.45x10 ⁻³	1.40x10 ⁻²
Threonine	0.663 [0.54 0.829]	1.042 [0.783 1.49]	-0.78	4.45x10 ⁻³	1.40x10 ⁻²
2-Hydroxyvalerate	0.867 [0.853 0.938]	1.053 [0.961 1.222]	-0.3	5.22x10 ⁻³	1.53x10 ⁻²
3-Hydroxybutyrate	0.875 [0.812 0.899]	1.029 [0.965 1.196]	-0.42	6.11x10 ⁻³	1.67x10 ⁻²
Unsaturated lipid (-CH=CH-)	0.769 [0.723 0.922]	1.05 [0.904 1.332]	-0.46	8.27x10 ⁻³	1.99x10 ⁻²
Glycerol phospholipid	0.475 [0.276 0.577]	1.24 [0.552 1.922]	-1.39	8.27x10 ⁻³	1.99x10 ⁻²
Creatine	0.493 [0.362 0.765]	1.103 [0.637 1.575]	-1.08	1.10x10 ⁻²	2.51x10 ⁻²
Pyroglutamate	0.879 [0.864 0.951]	1.023 [0.96 1.239]	-0.25	1.45x10 ⁻²	2.98x10 ⁻²
Phospholipid	1.184 [1.1 1.311]	1.014 [0.926 1.118]	0.23	1.45x10 ⁻²	2.98x10 ⁻²
Desaminotyrosine	13.361 [6.036 19.946]	1 [1 1]	0.79	2.11x10 ⁻²	4.12x10 ⁻²
Isobutyrate	0.643 [0.588 0.868]	0.959 [0.848 1.202]	-0.35	3.89x10 ⁻²	7.25x10 ⁻²
Glutamate	0.702 [0.467 0.938]	1.146 [0.664 1.378]	-0.87	4.85x10 ⁻²	8.64x10 ⁻²
Pyruvate	1.781 [1.131 2.386]	1 [0.643 1.507]	0.74	8.87x10 ⁻²	1.52x10 ⁻¹
Lipid (-(-CH2-)n-)	0.825 [0.773 0.992]	1.051 [0.892 1.512]	-0.43	9.74x10 ⁻²	1.60x10 ⁻¹
Lipid (alpha-CH2)	0.518 [0.491 0.974]	0.943 [0.671 1.98]	-0.86	1.17x10 ⁻¹	1.84x10 ⁻¹
Methanol	0.919 [0.682 1.255]	1.244 [0.924 1.534]	-0.32	1.39x10 ⁻¹	2.10x10 ⁻¹
Acetoacetate	0.833 [0.643 0.94]	0.958 [0.734 1.286]	-0.52	1.77x10 ⁻¹	2.59x10 ⁻¹
Formate	1.119 [0.951 1.352]	1.023 [0.767 1.158]	0.24	2.57x10 ⁻¹	3.63x10 ⁻¹
2-Hydroxybutyrate	0.989 [0.854 1.332]	0.933 [0.759 1.196]	0.21	2.75x10 ⁻¹	3.76x10 ⁻¹
3-Hydroxyisovalerate	0.484 [0.392 2.729]	0.663 [0.489 3.622]	-0.03	3.35x10 ⁻¹	4.29x10 ⁻¹
Lipid (=CH-CH2-CH=)	0.95 [0.819 1.004]	1.05 [0.86 1.102]	-0.1	3.35x10 ⁻¹	4.29x10 ⁻¹
Glycine	1.003 [0.886 1.145]	1.109 [0.979 1.193]	-0.14	4.27x10 ⁻¹	5.30x10 ⁻¹
Cholesterol backbone	1.047 [0.912 1.146]	1.131 [0.93 1.324]	-0.11	5.04x10 ⁻¹	6.08x10 ⁻¹
Glucose	1.015 [0.865 1.093]	1 [0.898 1.247]	-0.13	5.59x10 ⁻¹	6.51x10 ⁻¹
Lactate	1.388 [0.738 1.634]	0.869 [0.698 1.334]	0.27	5.88x10 ⁻¹	6.51x10 ⁻¹
Citrate	0.969 [0.8 1.195]	1.042 [0.805 1.285]	-0.1	5.88x10 ⁻¹	6.51x10 ⁻¹
Creatinine	0.974 [0.873 1.149]	1.004 [0.807 1.14]	-0.03	7.39x10 ⁻¹	7.97x10 ⁻¹
Phenylalanine	0.958 [0.821 1.113]	0.95 [0.799 1.111]	0.03	8.03x10 ⁻¹	8.44x10 ⁻¹
Acetate	1.053 [0.873 1.247]	1.004 [0.902 1.255]	-0.05	9.01x10 ⁻¹	9.23x10 ⁻¹
Tyrosine	1.077 [0.87 1.186]	1.043 [0.865 1.148]	0.01	1.00	1.00

Table S2. Statistical comparison between the lipoprotein profiles of Metabotype IV *versus* the others.

Feature	Metabotype IV, median [IQR]	Others, median [IQR]	log change	P-value	FDR
MP Apo.A2	25.01 [22.865 26.685]	30.93 [27.97 33.845]	-0.32	1.04x10 ⁻³	5.32x10 ⁻²
Apo.A2 HDL.4	12.15 [11.055 13.345]	18.24 [14.905 21.8]	-0.6	1.04x10 ⁻³	5.32x10 ⁻²
Trigl LDL.2	3.59 [3.33 3.905]	2.08 [1.825 2.755]	0.63	1.40x10 ⁻³	5.32x10 ⁻²
MF Apo.A2 HDL	26.37 [24.41 28.245]	31.88 [29.105 34.71]	-0.31	1.87x10 ⁻³	5.34x10 ⁻²
Apo.A1 HDL.4	54.24 [48.105 57.045]	64.89 [60.085 76.18]	-0.35	2.67x10 ⁻³	6.10x10 ⁻²
Trigl LDL.1	8.15 [7.56 9.31]	5.15 [3.7 7.2]	0.57	4.45x10 ⁻³	8.45x10 ⁻²
Phosp HDL.4	18.97 [18.37 19.995]	22.74 [20.58 27.65]	-0.33	5.22x10 ⁻³	8.50x10 ⁻²
MF Trigl LDL	25.6 [23.305 28.935]	20.07 [16.775 23.18]	0.38	8.27x10 ⁻³	1.18x10 ⁻¹
MF Phosp IDL	2.88 [2.1 5.535]	7.1 [4.77 11.555]	-1.13	9.37x10 ⁻³	1.19x10 ⁻¹
Chol HDL.4	13.87 [11.825 14.56]	18.07 [14.09 22.105]	-0.44	1.06x10 ⁻²	1.21x10 ⁻¹
Chol VLDL.1	2.71 [1.69 4.625]	7.76 [4.5 17.16]	-1.52	1.43x10 ⁻²	1.48x10 ⁻¹
MP Chol	151.07 [139.25 184.675]	209.08 [183.615 223.545]	-0.32	1.89x10 ⁻²	1.78x10 ⁻¹
Free Chol					
VLDL.1	0.5 [0.355 1.135]	2.84 [1.345 6.68]	-2.05	2.02x10 ⁻²	1.78x10 ⁻¹
LDL.5 PN	122.22 [44.215 173.705]	235.58 [150.9 303.1]	-0.82	2.75x10 ⁻²	2.08x10 ⁻¹
Apo.B LDL.5	6.72 [2.43 9.555]	12.96 [8.3 16.67]	-0.82	2.75x10 ⁻²	2.08x10 ⁻¹
Phosp VLDL.1	2.89 [2.05 3.805]	7.51 [3.32 14.185]	-1.4	3.09x10 ⁻²	2.08x10 ⁻¹
Chol VLDL.2	2.38 [1.82 2.735]	3.91 [2.935 5.69]	-0.73	3.32x10 ⁻²	2.08x10 ⁻¹
Chol LDL.5	7.27 [1.465 11.7]	18.74 [10.945 22.83]	-1.01	3.32x10 ⁻²	2.08x10 ⁻¹
Phosp LDL.5	4.84 [1.515 6.935]	9.85 [6.405 11.875]	-0.87	3.47x10 ⁻²	2.08x10 ⁻¹
Trigl VLDL.1	11.87 [9.42 26.39]	39.89 [19.25 92.63]	-1.53	4.35x10 ⁻²	2.36x10 ⁻¹
Trigl HDL.1	4.34 [3.785 5.4]	3.33 [2.37 3.97]	0.4	4.35x10 ⁻²	2.36x10 ⁻¹
MF Apo.A1 HDL	126.51 [118.6 131.56]	138.97 [133.715 147.975]	-0.14	4.85x10 ⁻²	2.51x10 ⁻¹
MP Apo.A1	128.49 [120.11 132.985]	137.99 [131.55 147.33]	-0.13	5.98x10 ⁻²	2.97x10 ⁻¹
Free Chol LDL.5	2.63 [1.05 3.635]	4.45 [2.98 5.795]	-0.66	7.36x10 ⁻²	3.50x10 ⁻¹
Trigl HDL.2	2.14 [1.96 2.345]	1.7 [1.365 2.22]	0.32	8.07x10 ⁻²	3.68x10 ⁻¹
Trigl LDL.3	2.9 [2.51 3.37]	2.4 [1.705 2.765]	0.36	8.45x10 ⁻²	3.68x10 ⁻¹
Chol LDL.4	1.51 [0 4.95]	8.94 [2.735 13.525]	-1.03	8.71x10 ⁻²	3.68x10 ⁻¹
MF Trigl IDL	4.43 [2.895 7.345]	9.89 [5.215 25.96]	-1.39	9.25x10 ⁻²	3.70x10 ⁻¹
MP LDL.Chol	70.08 [51.52 95.205]	104.2 [82.98 116.18]	-0.38	9.74x10 ⁻²	3.70x10 ⁻¹
MF Chol LDL	70.08 [51.52 95.205]	104.2 [82.98 116.18]	-0.38	9.74x10 ⁻²	3.70x10 ⁻¹
Trigl LDL.4	2.58 [2.16 2.75]	1.55 [1.11 2.14]	0.56	1.06x10 ⁻¹	3.88x10 ⁻¹
MF Phosp VLDL	14.07 [11.885 23.41]	23.72 [16.435 39.41]	-0.56	1.17x10 ⁻¹	3.95x10 ⁻¹
LDL.4 PN	34.66 [0 72.185]	119.2 [41.125 153.81]	-0.85	1.24x10 ⁻¹	3.95x10 ⁻¹
Phosp LDL.4	2.14 [0 3.865]	5.41 [1.77 7.75]	-0.82	1.24x10 ⁻¹	3.95x10 ⁻¹
Apo.B LDL.4	1.91 [0 3.97]	6.56 [2.265 8.46]	-0.85	1.24x10 ⁻¹	3.95x10 ⁻¹
MF Free Chol					
VLDL	6.27 [5.155 10.92]	11.05 [7.74 17.9]	-0.58	1.25x10 ⁻¹	3.95x10 ⁻¹
LDL.Chol					
HDL.Chol	1.48 [0.985 1.8]	1.82 [1.61 2.12]	-0.23	1.47x10 ⁻¹	3.95x10 ⁻¹
MF Chol VLDL	13.65 [11.82 24.19]	26.46 [17.545 40.495]	-0.57	1.51x10 ⁻¹	3.95x10 ⁻¹
Phosp LDL.6	12.38 [9.875 17.385]	16.01 [13.97 19.735]	-0.31	1.51x10 ⁻¹	3.95x10 ⁻¹
Phosp HDL.1	25.88 [23.355 28.415]	20.69 [14.64 27.49]	0.29	1.51x10 ⁻¹	3.95x10 ⁻¹
MF Free Chol					
LDL	22.25 [19.255 27.525]	28.41 [24.245 32.88]	-0.21	1.60x10 ⁻¹	3.95x10 ⁻¹
MP Apo.B100	69.62 [62.005 86.355]	91.46 [76.055 99.625]	-0.21	1.63x10 ⁻¹	3.95x10 ⁻¹
Total PN	1265.89 [1127.435 1570.115]	1663.05 [1382.915 1811.48]	-0.21	1.63x10 ⁻¹	3.95x10 ⁻¹
MF Trigl VLDL	60.34 [40.14 72.935]	74.42 [49.29 161.715]	-0.77	1.63x10 ⁻¹	3.95x10 ⁻¹
Chol LDL.6	21.31 [16.14 31.8]	29.34 [24.955 37.035]	-0.38	1.63x10 ⁻¹	3.95x10 ⁻¹
Apo.B LDL.1	13.34 [12.155 13.715]	11.37 [8.335 13.33]	0.18	1.73x10 ⁻¹	3.95x10 ⁻¹
LDL PN	1006.8 [780.95 1357.535]	1329.88 [1132.015 1424.785]	-0.23	1.77x10 ⁻¹	3.95x10 ⁻¹
LDL.1 PN	242.55 [220.995 249.37]	206.72 [151.57 242.355]	0.18	1.77x10 ⁻¹	3.95x10 ⁻¹
MF Phosp LDL	46.19 [37.155 58.085]	59.18 [47.095 66.04]	-0.2	1.77x10 ⁻¹	3.95x10 ⁻¹
MF Apo.B LDL	55.37 [42.95 74.66]	73.14 [62.255 78.36]	-0.23	1.77x10 ⁻¹	3.95x10 ⁻¹
Chol HDL.1	22.1 [19.685 24.495]	18.78 [13.855 23.415]	0.24	1.77x10 ⁻¹	3.95x10 ⁻¹
Free Chol					
VLDL.2	1.17 [1.04 1.43]	1.91 [1.16 2.945]	-0.58	1.87x10 ⁻¹	4.09x10 ⁻¹
MF Free Chol					
IDL	2.34 [1.975 4.385]	4.03 [2.78 5.695]	-0.46	2.01x10 ⁻¹	4.28x10 ⁻¹
LDL.6 PN	325.77 [240.52 484.46]	450.55 [346.56 564.44]	-0.39	2.06x10 ⁻¹	4.28x10 ⁻¹
Apo.B LDL.6	17.92 [13.23 26.645]	24.78 [19.06 31.045]	-0.39	2.06x10 ⁻¹	4.28x10 ⁻¹
Phosp VLDL.2	2.64 [2.47 3.195]	3.8 [2.465 5.65]	-0.47	2.17x10 ⁻¹	4.41x10 ⁻¹
MP Trigl	96.35 [80.635 122.505]	114.96 [91.02 219.825]	-0.57	2.39x10 ⁻¹	4.78x10 ⁻¹
MF Trigl HDL	10.91 [9.755 11.665]	8.54 [7.13 12.12]	0.14	2.57x10 ⁻¹	4.92x10 ⁻¹
Free Chol LDL.4	1.95 [0.545 2.795]	2.95 [1.57 4.345]	-0.5	2.59x10 ⁻¹	4.92x10 ⁻¹
Apo.A2 HDL.3	5.51 [5.165 5.945]	6.01 [5.535 6.885]	-0.19	2.59x10 ⁻¹	4.92x10 ⁻¹
MF Chol IDL	9.21 [7.735 15.675]	14.04 [9.71 18.84]	-0.35	2.75x10 ⁻¹	5.14x10 ⁻¹
Free Chol LDL.1	6.71 [5.86 7.34]	5.47 [4.495 7.21]	0.11	2.87x10 ⁻¹	5.28x10 ⁻¹

Free Chol					
VLDL.5	0.39 [0.35 1.16]	0.93 [0.575 1.565]	-0.59	2.94x10 ⁻¹	5.32x10 ⁻¹
Phosp LDL.1	13.28 [11.57 13.34]	11.24 [8.925 13.96]	0.1	3.07x10 ⁻¹	5.46x10 ⁻¹
MP HDL.Chol	52.7 [48.33 54.175]	55.73 [51.2 62.19]	-0.1	3.35x10 ⁻¹	5.79x10 ⁻¹
MF Chol HDL	52.7 [48.33 54.175]	55.73 [51.2 62.19]	-0.1	3.35x10 ⁻¹	5.79x10 ⁻¹
Free Chol LDL.6	5.21 [3.14 7.33]	6.48 [5.045 7.83]	-0.27	3.49x10 ⁻¹	5.90x10 ⁻¹
Free Chol LDL.2	5.75 [5.19 6.645]	5.35 [4.185 6.4]	0.18	3.57x10 ⁻¹	5.90x10 ⁻¹
Trigl HDL.3	2.21 [2.09 2.45]	1.71 [1.585 2.645]	0.17	3.57x10 ⁻¹	5.90x10 ⁻¹
Free Chol HDL.3	1.4 [1.325 1.605]	1.67 [1.375 1.95]	-0.1	3.71x10 ⁻¹	6.04x10 ⁻¹
Trigl VLDL.2	9.48 [9.145 12.295]	12.3 [7.77 20.28]	-0.35	4.03x10 ⁻¹	6.26x10 ⁻¹
Trigl VLDL.4	11.24 [8.685 16.995]	10.15 [7.525 13.39]	0.26	4.03x10 ⁻¹	6.26x10 ⁻¹
Trigl HDL.4	2.8 [2.31 3.11]	2.86 [2.515 3.98]	-0.23	4.03x10 ⁻¹	6.26x10 ⁻¹
Free Chol VLDL.3	1.11 [1.075 2.275]	2.12 [1.16 3.395]	-0.42	4.06x10 ⁻¹	6.26x10 ⁻¹
Apo.B100					
Apo.A1	0.57 [0.475 0.68]	0.66 [0.56 0.745]	-0.06	4.18x10 ⁻¹	6.35x10 ⁻¹
Chol VLDL.3	2.67 [2.43 5.15]	4.67 [2.66 6.615]	-0.32	4.56x10 ⁻¹	6.84x10 ⁻¹
Trigl VLDL.5	3.31 [2.555 3.66]	3.01 [2.105 3.655]	0.19	4.69x10 ⁻¹	6.94x10 ⁻¹
Free Chol HDL.2	1.75 [1.495 1.82]	1.57 [1.03 1.81]	0.1	5.23x10 ⁻¹	7.55x10 ⁻¹
Apo.A2 HDL.1	2.93 [2.72 3.645]	2.86 [2.115 3.64]	0.12	5.23x10 ⁻¹	7.55x10 ⁻¹
Chol HDL.3	8.81 [8.215 9.26]	9 [8.4 10.165]	-0.05	5.31x10 ⁻¹	7.55x10 ⁻¹
Trigl LDL.6	5.11 [4.85 7.015]	5.12 [4.355 6.655]	0.11	5.37x10 ⁻¹	7.55x10 ⁻¹
Free Chol HDL.4	2.4 [2.05 2.655]	2.42 [1.69 3.47]	-0.22	5.59x10 ⁻¹	7.76x10 ⁻¹
Phosp VLDL.3	3.31 [2.965 5.41]	4.97 [3.11 6.755]	-0.22	5.65x10 ⁻¹	7.76x10 ⁻¹
LDL.2 PN	166.03 [151.33 208.055]	164.37 [139.665 205.405]	0.11	5.88x10 ⁻¹	7.84x10 ⁻¹
Apo.B LDL.2	9.13 [8.32 11.44]	9.04 [7.68 11.295]	0.11	5.88x10 ⁻¹	7.84x10 ⁻¹
Free Chol VLDL.4	1.52 [1.075 3.665]	2.14 [1.405 3.315]	-0.14	6.09x10 ⁻¹	7.84x10 ⁻¹
Trigl LDL.5	2.56 [2.1 3.05]	2.94 [2.205 3.695]	-0.07	6.09x10 ⁻¹	7.84x10 ⁻¹
Apo.A1 HDL.1	27.97 [26.28 30.92]	26.72 [18.345 36.18]	0.15	6.17x10 ⁻¹	7.84x10 ⁻¹
Chol VLDL.4	4.74 [3.605 8.725]	5.18 [4.165 7.74]	0	6.39x10 ⁻¹	7.84x10 ⁻¹
Chol VLDL.5	1.21 [0.59 1.64]	1.51 [0.84 1.88]	-0.06	6.39x10 ⁻¹	7.84x10 ⁻¹
VLDL PN	146.61 [124.635 216.05]	174.76 [125.635 260.425]	-0.19	6.47x10 ⁻¹	7.84x10 ⁻¹
IDL PN	85.1 [82.795 107.41]	93.81 [66.715 125.515]	-0.12	6.47x10 ⁻¹	7.84x10 ⁻¹
MF Apo.B VLDL	8.06 [6.855 11.885]	9.61 [6.91 14.32]	-0.19	6.47x10 ⁻¹	7.84x10 ⁻¹
MF Apo.B IDL	4.68 [4.555 5.905]	5.16 [3.67 6.9]	-0.12	6.47x10 ⁻¹	7.84x10 ⁻¹
Chol LDL.3	9.6 [6.69 12.495]	13.12 [7.465 15.125]	-0.13	6.70x10 ⁻¹	8.04x10 ⁻¹
MF Free Chol HDL	10.57 [10.04 12.075]	9.66 [9.24 12.95]	0.06	7.01x10 ⁻¹	8.24x10 ⁻¹
Free Chol HDL.1	3.98 [3.115 4.53]	3.75 [2.475 4.825]	0.02	7.01x10 ⁻¹	8.24x10 ⁻¹
Phosp LDL.2	8.86 [8.125 10.94]	8.8 [7.225 11.315]	0.09	7.39x10 ⁻¹	8.60x10 ⁻¹
Apo.B LDL.3	6.96 [5.93 8.795]	7.25 [5.035 9.525]	0.11	7.66x10 ⁻¹	8.79x10 ⁻¹
LDL.3 PN	126.58 [107.8 159.865]	131.79 [91.545 173.235]	0.11	7.71x10 ⁻¹	8.79x10 ⁻¹
Free Chol LDL.3	3.79 [3.595 4.735]	4.21 [2.53 5.035]	0.15	7.98x10 ⁻¹	8.97x10 ⁻¹
Chol LDL.2	14.59 [13.305 19.375]	16.89 [13.345 20.83]	-0.01	8.03x10 ⁻¹	8.97x10 ⁻¹
Apo.A1 HDL.3	24.55 [22.595 25.55]	23.88 [22.23 27.09]	-0.05	8.15x10 ⁻¹	9.02x10 ⁻¹
Phosp HDL.2	13.27 [12.895 14.535]	14.26 [10.64 15.715]	0.07	8.35x10 ⁻¹	9.15x10 ⁻¹
Phosp VLDL.5	1.36 [1 2.415]	1.97 [1.055 2.29]	-0.06	8.48x10 ⁻¹	9.21x10 ⁻¹
Apo.A1 HDL.2	19.84 [15.235 20.365]	17.65 [15.25 20.59]	0.02	8.98x10 ⁻¹	9.66x10 ⁻¹
Phosp VLDL.4	4.86 [3.69 7.7]	4.87 [3.93 6.625]	0.1	9.15x10 ⁻¹	9.75x10 ⁻¹
Phosp LDL.3	6.36 [5.4 7.79]	7.63 [4.78 8.565]	0.05	9.32x10 ⁻¹	9.76x10 ⁻¹
Chol HDL.2	8.36 [7.58 8.71]	8.56 [6.67 10.025]	-0.01	9.34x10 ⁻¹	9.76x10 ⁻¹
Trigl VLDL.3	9.86 [9.56 16.12]	11.94 [9.045 19.305]	-0.03	9.66x10 ⁻¹	9.84x10 ⁻¹
MF Phosp HDL	73.02 [67.19 75.64]	72.67 [65.82 80.31]	-0.01	9.67x10 ⁻¹	9.84x10 ⁻¹
Chol LDL.1	20.18 [17.485 21.32]	18.55 [14.815 24.82]	-0.07	9.67x10 ⁻¹	9.84x10 ⁻¹
Phosp HDL.3	14.39 [13.985 15.93]	14.33 [13.63 16.59]	0	9.83x10 ⁻¹	9.92x10 ⁻¹
Apo.A2 HDL.2	3.83 [3.415 4.015]	3.61 [3.125 4.435]	-0.06	1.00	1.00

*Abbreviations are reported as follows: MF: Main Fractions, MP: Main Parameters, PN: Particle Number, Chol: Cholesterol, Phosp: Phospholipids, Trig: Triglycerides.

Table S3. Statistical comparison between the metabolic profiles of Metabotype II *versus* Metabotype III.

Feature	Metabotype II, median [IQR]	Metabotype III, median [IQR]	log change	P-value	FDR
Pyroglutamate	1.28 [1.226 1.343]	0.974 [0.943 1.011]	0.45	1.02x10 ⁻⁶	1.05x10 ⁻⁵
2-Hydroxyvalerate	1.263 [1.21 1.325]	0.961 [0.93 0.997]	0.45	1.02x10 ⁻⁶	1.05x10 ⁻⁵
Unsaturated lipid (-CH=CH-)	1.388 [1.331 1.469]	0.917 [0.838 0.998]	0.64	1.02x10 ⁻⁶	1.05x10 ⁻⁵
Lipid (-(-CH2)-n-)	1.569 [1.51 1.742]	0.892 [0.776 1.003]	0.91	1.02x10 ⁻⁶	1.05x10 ⁻⁵
Lipid (-CH3-)	1.226 [1.183 1.272]	0.948 [0.938 1.034]	0.33	2.04x10 ⁻⁶	1.67x10 ⁻⁵
Lipid (alpha-CH2)	2.144 [1.957 2.631]	0.671 [0.494 0.863]	1.76	7.14x10 ⁻⁶	4.18x10 ⁻⁵
Glycerol phospholipid	2.084 [1.919 2.643]	0.552 [0.401 0.809]	1.73	7.14x10 ⁻⁶	4.18x10 ⁻⁵
Threonine	1.596 [1.341 1.786]	0.783 [0.741 0.96]	0.96	4.59x10 ⁻⁵	2.35x10 ⁻⁴
GlycB	0.72 [0.661 0.783]	0.978 [0.903 1.167]	-0.5	1.42x10 ⁻⁴	6.46x10 ⁻⁴
Lipid (beta-CH2)	1.288 [1.193 1.36]	1.032 [0.9 1.121]	0.36	1.98x10 ⁻⁴	8.11x10 ⁻⁴
3-Hydroxybutyrate	1.28 [1.183 1.351]	0.979 [0.949 1.017]	0.4	3.73x10 ⁻⁴	1.39x10 ⁻³
Lipid (=CH-CH2-CH=)	1.111 [1.08 1.156]	0.957 [0.818 1.046]	0.28	1.49x10 ⁻³	5.10x10 ⁻³
Cholesterol backbone	1.324 [1.228 1.43]	1.021 [0.851 1.149]	0.37	1.91x10 ⁻³	6.03x10 ⁻³
Methanol	0.978 [0.826 1.148]	1.41 [1.091 1.572]	-0.46	2.20x10 ⁻²	6.45x10 ⁻²
Pyruvate	1.362 [0.789 1.844]	0.918 [0.581 1.09]	0.75	4.17x10 ⁻²	1.14x10 ⁻¹
Phenylalanine	1.003 [0.942 1.072]	0.834 [0.669 1.01]	0.28	4.84x10 ⁻²	1.24x10 ⁻¹
Isoleucine	1.146 [1.062 1.254]	0.911 [0.777 1.182]	0.28	8.41x10 ⁻²	2.03x10 ⁻¹
GlycA	1.019 [0.96 1.073]	0.913 [0.827 0.979]	0.13	9.56x10 ⁻²	2.18x10 ⁻¹
Creatine	1.401 [0.958 1.613]	0.931 [0.45 1.289]	0.33	1.08x10 ⁻¹	2.34x10 ⁻¹
Mannose	1.11 [0.949 1.27]	0.977 [0.714 1.076]	0.25	1.22x10 ⁻¹	2.51x10 ⁻¹
Valine	1.01 [0.936 1.284]	0.956 [0.806 1.069]	0.19	1.38x10 ⁻¹	2.68x10 ⁻¹
3-Hydroxyisovalerate	0.568 [0.479 0.635]	1.799 [0.498 4.786]	-1.53	1.72x10 ⁻¹	3.21x10 ⁻¹
Glucose	1.011 [0.925 1.471]	0.911 [0.889 1.124]	0.28	1.92x10 ⁻¹	3.41x10 ⁻¹
Glutamine	1.157 [1.108 1.204]	1.07 [0.999 1.213]	0.07	2.35x10 ⁻¹	4.01x10 ⁻¹
Acetate	0.956 [0.895 1.125]	1.149 [0.959 1.291]	-0.15	3.41x10 ⁻¹	5.59x10 ⁻¹
Formate	0.979 [0.758 1.077]	1.051 [0.716 1.253]	-0.18	3.71x10 ⁻¹	5.85x10 ⁻¹
Glutamate	0.802 [0.46 1.302]	1.175 [0.814 1.223]	-0.23	4.03x10 ⁻¹	6.12x10 ⁻¹
Glycine	1.083 [0.933 1.158]	1.151 [0.99 1.195]	-0.1	4.37x10 ⁻¹	6.39x10 ⁻¹
Histidine	1.101 [1.029 1.16]	1.053 [0.967 1.151]	0.04	4.72x10 ⁻¹	6.67x10 ⁻¹
Creatinine	1.023 [0.906 1.071]	0.922 [0.754 1.126]	0.09	5.08x10 ⁻¹	6.72x10 ⁻¹
Leucine	1.069 [0.941 1.199]	0.94 [0.842 1.145]	0.1	5.08x10 ⁻¹	6.72x10 ⁻¹
Tyrosine	1.044 [0.928 1.109]	0.918 [0.784 1.1]	0.11	5.46x10 ⁻¹	6.99x10 ⁻¹
Alanine	1.019 [0.856 1.274]	0.95 [0.879 1.126]	0.16	5.85x10 ⁻¹	7.27x10 ⁻¹
2-Hydroxybutyrate	1.024 [0.746 1.18]	0.865 [0.806 1.063]	0.04	7.96x10 ⁻¹	9.31x10 ⁻¹
Protein	1.013 [0.962 1.072]	1.028 [0.993 1.064]	0	7.96x10 ⁻¹	9.31x10 ⁻¹
Desaminotyrosine	1 [1 1]	1 [1 1]	-1.32	8.37x10 ⁻¹	9.31x10 ⁻¹
Citrate.left	1.003 [0.778 1.279]	1.035 [0.819 1.249]	-0.01	8.41x10 ⁻¹	9.31x10 ⁻¹
Lactate	0.79 [0.689 1.483]	0.82 [0.709 1.035]	0.21	8.86x10 ⁻¹	9.31x10 ⁻¹
Acetoacetate	0.943 [0.724 1.231]	0.904 [0.659 1.244]	-0.18	8.86x10 ⁻¹	9.31x10 ⁻¹
Phospholipid	1.006 [0.925 1.143]	1.035 [0.934 1.099]	-0.02	9.31x10 ⁻¹	9.55x10 ⁻¹
Isobutyrate	0.972 [0.895 1.035]	0.918 [0.817 1.3]	-0.15	9.77x10 ⁻¹	9.77x10 ⁻¹

Table S4. Statistical comparison between the lipoprotein profiles of Metabotype II versus Metabotype III.

Feature	Metabotype II, median [IQR]	Metabotype III, median [IQR]	log change	P-value	FDR
MP Trig1	240.36 [214.002 299.433]	91.02 [71.323 111.78]	1.52	2.04x10 ⁻⁶	8.14x10 ⁻⁵
MF Trig1 VLDL	170.69 [159.183 210.132]	49.29 [38.688 69.142]	1.79	4.08x10 ⁻⁶	8.14x10 ⁻⁵
MF Trig1 IDL	26.57 [25.95 36.155]	5.215 [2.48 8.28]	2.47	4.08x10 ⁻⁶	8.14x10 ⁻⁵
Trigl VLDL.1	107.94 [80.3 134.335]	19.25 [11.785 27.652]	2.45	4.08x10 ⁻⁶	8.14x10 ⁻⁵
Phosp VLDL.1	15.77 [13.717 19.802]	3.32 [2.45 4.475]	-2.2	4.08x10 ⁻⁶	8.14x10 ⁻⁵
VLDL PN	278.955 [243.442 330.25]	125.635 [95.94 150.235]	1.26	7.14x10 ⁻⁶	8.14x10 ⁻⁵
MF Chol VLDL	43.075 [37.285 54.905]	17.545 [11.707 21.458]	1.42	7.14x10 ⁻⁶	8.14x10 ⁻⁵
MF Chol IDL	18.84 [15.553 28.68]	9.94 [6.973 12.57]	1.22	7.14x10 ⁻⁶	8.14x10 ⁻⁵
MF Apo.B VLDL	15.34 [13.39 18.163]	6.91 [5.275 8.262]	1.26	7.14x10 ⁻⁶	8.14x10 ⁻⁵
Chol VLDL.1	20.4 [16.017 25.052]	4.5 [2.473 6.945]	2.2	7.14x10 ⁻⁶	8.14x10 ⁻⁵
MP Apo.B100	105.6 [96.73 118.345]	80.055 [67.373 86.227]	0.45	3.06x10 ⁻⁵	2.49x10 ⁻⁴
Total PN	1920.06 [1758.775 2151.845]	1455.6 [1224.988 1567.833]	0.45	3.06x10 ⁻⁵	2.49x10 ⁻⁴
MF Phosp VLDL	43.66 [36.663 48.425]	16.435 [12.775 21.26]	1.35	3.06x10 ⁻⁵	2.49x10 ⁻⁴
Trigl VLDL.2	23.955 [18.545 28.067]	7.77 [6.612 12.015]	1.31	3.06x10 ⁻⁵	2.49x10 ⁻⁴
FreeChol VLDL.3	3.745 [3.165 4.645]	1.16 [0.662 1.658]	1.65	4.59x10 ⁻⁵	3.27x10 ⁻⁴
Phosp VLDL.2	6.555 [5.123 7.707]	2.465 [1.86 3.462]	1.22	4.59x10 ⁻⁵	3.27x10 ⁻⁴
MF Phosp IDL	13.195 [11.277 13.99]	4.77 [3.405 5.865]	1.44	6.03x10 ⁻⁵	4.04x10 ⁻⁴
FreeChol VLDL.1	6.905 [6.43 8.595]	1.345 [0.48 1.855]	2.52	7.68x10 ⁻⁵	4.87x10 ⁻⁴
MF FreeChol VLDL	19.25 [17.398 23.538]	7.74 [5.652 9.453]	1.43	9.83x10 ⁻⁵	5.90x10 ⁻⁴
MF FreeChol IDL	5.695 [4.505 8.465]	2.78 [1.777 3.518]	1.32	1.40x10 ⁻⁴	6.73x10 ⁻⁴
FreeChol VLDL.2	3.625 [2.663 3.978]	1.16 [0.845 1.642]	1.55	1.41x10 ⁻⁴	6.73x10 ⁻⁴
MP Chol	242.43 [216.257 251.153]	186.395 [170.765 211.16]	0.33	1.42x10 ⁻⁴	6.73x10 ⁻⁴
Chol VLDL.3	7.59 [6.305 9.242]	2.66 [1.96 4.388]	1.34	1.42x10 ⁻⁴	6.73x10 ⁻⁴
FreeChol VLDL.4	3.19 [2.522 5.088]	1.405 [1.087 1.878]	1.38	1.42x10 ⁻⁴	6.73x10 ⁻⁴
Trigl VLDL.4	14.33 [12.16 18.005]	7.765 [5.54 9.773]	0.96	1.98x10 ⁻⁴	9.02x10 ⁻⁴
Phosp VLDL.4	6.795 [5.648 8.938]	3.99 [2.865 4.717]	0.88	2.74x10 ⁻⁴	1.20x10 ⁻³
IDL PN	123.81 [97.71 174.377]	72.05 [59.055 92.15]	0.93	3.73x10 ⁻⁴	1.32x10 ⁻³
LDL.6 PN	627.52 [479.947 741.995]	346.56 [249.407 458.607]	0.82	3.73x10 ⁻⁴	1.32x10 ⁻³
MF Apo.B IDL	6.805 [5.375 9.592]	3.965 [3.25 5.067]	0.93	3.73x10 ⁻⁴	1.32x10 ⁻³
Phosp LDL.6	20.9 [18.147 23.673]	14.04 [11.262 17.135]	0.59	3.73x10 ⁻⁴	1.32x10 ⁻³
Apo.B LDL.6	34.515 [26.395 40.81]	19.06 [13.715 25.225]	0.82	3.73x10 ⁻⁴	1.32x10 ⁻³
Trigl HDL.4	4.755 [3.945 4.992]	2.58 [2.433 2.852]	0.77	3.73x10 ⁻⁴	1.32x10 ⁻³
Chol VLDL.2	7.385 [5.69 7.815]	2.935 [2.002 3.798]	1.24	3.95x10 ⁻⁴	1.32x10 ⁻³
Phosp VLDL.3	7.77 [6.45 8.297]	3.11 [2.045 3.857]	1.21	3.95x10 ⁻⁴	1.32x10 ⁻³
Trigl VLDL.3	21.9 [18.928 23.95]	9.045 [5.5 11.13]	1.15	4.41x10 ⁻⁴	1.44x10 ⁻³
Chol LDL.6	39.615 [33.542 44.188]	24.98 [18.985 31.68]	0.68	5.04x10 ⁻⁴	1.60x10 ⁻³
Phosp VLDL.5	2.29 [2.138 2.743]	1.16 [0.915 1.9]	0.87	6.12x10 ⁻⁴	1.89x10 ⁻³
FreeChol VLDL.5	1.66 [1.202 2.345]	0.575 [0.302 0.908]	1.49	6.70x10 ⁻⁴	2.01x10 ⁻³
Chol VLDL.4	7.425 [5.768 11.15]	4.54 [3.668 5.143]	0.9	8.84x10 ⁻⁴	2.58x10 ⁻³
Trigl VLDL.5	3.69 [3.298 3.907]	2.235 [1.852 2.918]	0.6	1.15x10 ⁻³	3.28x10 ⁻³
Trigl HDL.3	2.645 [2.518 2.85]	1.61 [1.395 1.738]	0.71	3.84x10 ⁻³	1.07x10 ⁻²
Apo.B100 Apo.A1	0.755 [0.705 0.815]	0.59 [0.442 0.62]	0.4	4.46x10 ⁻³	1.21x10 ⁻²
MF Trig1 HDL	12.12 [10.488 13.635]	7.795 [6.595 9.035]	0.63	5.91x10 ⁻³	1.57x10 ⁻²
Apo.A2 HDL.3	6.88 [6.22 7.827]	5.75 [5.462 6.303]	0.3	9.97x10 ⁻³	2.58x10 ⁻²
Trigl LDL.6	6 [5.397 7.277]	4.94 [4.215 5.12]	0.34	1.08x10 ⁻²	2.74x10 ⁻²
FreeChol LDL.6	8.035 [6.635 8.848]	5.59 [4.115 7.095]	0.54	1.18x10 ⁻²	2.92x10 ⁻²
Chol VLDL.5	1.88 [1.268 1.94]	0.935 [0.682 1.532]	0.73	1.28x10 ⁻²	3.10x10 ⁻²
MF Apo.A2 HDL	34.53 [34.012 39.838]	31.145 [28.988 33.998]	0.21	1.51x10 ⁻²	3.58x10 ⁻²
FreeChol HDL.2	1.365 [1.005 1.562]	1.705 [1.578 2.295]	-0.48	1.77x10 ⁻²	4.12x10 ⁻²
LDL.Chol HDL.Chol	2.05 [1.872 2.608]	1.66 [1.365 1.948]	0.35	2.20x10 ⁻²	4.92x10 ⁻²
Chol HDL.2	7.945 [6.66 8.887]	10.025 [8.425 11.535]	-0.32	2.20x10 ⁻²	4.92x10 ⁻²
MP Apo.A2	33.51 [31.752 39.38]	30.085 [28.275 32.907]	0.19	3.06x10 ⁻²	6.46x10 ⁻²
Trigl LDL.4	1.215 [1.105 1.64]	2.055 [1.572 2.675]	-0.75	3.06x10 ⁻²	6.46x10 ⁻²
FreeChol LDL.3	3.41 [1.525 3.945]	4.775 [4.383 5.195]	-0.61	3.06x10 ⁻²	6.46x10 ⁻²
MP HDL.Chol	52.805 [51.48 56.24]	60.32 [55.983 66.567]	-0.19	3.58x10 ⁻²	7.16x10 ⁻²
MF Chol HDL	52.805 [51.48 56.24]	60.32 [55.983 66.567]	-0.19	3.58x10 ⁻²	7.16x10 ⁻²
Apo.A1 HDL.3	25.605 [24.717 30.588]	23 [22.125 27.023]	0.17	3.58x10 ⁻²	7.16x10 ⁻²
LDL.5 PN	300.27 [247.382 321.38]	166.685 [140.99 271.045]	0.53	4.17x10 ⁻²	7.93x10 ⁻²
Apo.B LDL.5	16.515 [13.608 17.677]	9.165 [7.755 14.91]	0.53	4.17x10 ⁻²	7.93x10 ⁻²
Apo.A2 HDL.4	21.8 [19.265 23.615]	16.905 [14.2 20.055]	0.27	4.17x10 ⁻²	7.93x10 ⁻²
MF FreeChol HDL	9.35 [9.21 10.613]	11.74 [9.867 13.873]	-0.26	4.33x10 ⁻²	8.10x10 ⁻²
LDL PN	1413.735 [1368.12 1703.253]	1290.365 [1040.328 1393.905]	0.27	4.84x10 ⁻²	8.76x10 ⁻²
MF Apo.B LDL	77.755 [75.245 93.675]	70.965 [57.218 76.665]	0.27	4.84x10 ⁻²	8.76x10 ⁻²
Chol LDL.5	22.305 [19.075 26.078]	11.495 [8.82 21.198]	0.57	6.43x10 ⁻²	1.11x10 ⁻¹
FreeChol LDL.2	4.665 [3.745 5.432]	6.255 [4.715 6.837]	-0.34	6.43x10 ⁻²	1.11x10 ⁻¹
FreeChol HDL.1	3.275 [2.427 4.258]	4.425 [3.698 5.373]	-0.43	6.43x10 ⁻²	1.11x10 ⁻¹
Trigl LDL.1	6.215 [4.98 7.638]	4.66 [3.415 6.215]	0.41	8.41x10 ⁻²	1.39x10 ⁻¹
Phosp LDL.2	8.03 [5.698 9.823]	10.935 [8.47 11.943]	-0.36	8.41x10 ⁻²	1.39x10 ⁻¹
Apo.A1 HDL.1	21.315 [17.885 26.622]	33.085 [21.492 38.812]	-0.39	8.41x10 ⁻²	1.39x10 ⁻¹
Apo.A2 HDL.2	4.045 [3.553 4.988]	3.53 [2.595 4.11]	0.37	9.50x10 ⁻²	1.54x10 ⁻¹

Phosp LDL.5	11.475 [10.037 13.707]	6.905 [5.468 11.435]	0.45	9.56x10 ⁻²	1.54x10 ⁻¹
MF Phosp HDL	71.915 [67.455 74.508]	78.99 [71.608 86.375]	-0.14	1.22x10 ⁻¹	1.94x10 ⁻¹
MF Trigl LDL	21.375 [19.322 24.075]	19.655 [15.92 20.8]	0.16	1.38x10 ⁻¹	2.15x10 ⁻¹
Trigl LDL.2	1.91 [1.802 2.067]	2.6 [1.895 3.11]	-0.41	1.43x10 ⁻¹	2.18x10 ⁻¹
Trigl HDL.2	2.015 [1.69 2.253]	1.62 [1.118 1.89]	0.37	1.43x10 ⁻¹	2.18x10 ⁻¹
Chol LDL.2	15.645 [10.517 18.652]	20.02 [14.985 21.427]	-0.28	1.54x10 ⁻¹	2.25x10 ⁻¹
Phosp HDL.1	18.775 [15.248 20.535]	24.395 [19.26 30.763]	-0.3	1.54x10 ⁻¹	2.25x10 ⁻¹
Phosp HDL.2	13.235 [10.63 14.83]	15.465 [12.505 17.177]	-0.17	1.54x10 ⁻¹	2.25x10 ⁻¹
Trigl LDL.5	3.56 [2.708 3.827]	2.725 [2.22 3.428]	0.25	1.69x10 ⁻¹	2.43x10 ⁻¹
FreeChol LDL.5	5.4 [4.493 6.308]	3.31 [2.975 5.9]	0.28	1.88x10 ⁻¹	2.66x10 ⁻¹
Phosp LDL.3	6.72 [1.825 8.13]	7.88 [5.605 8.828]	-0.45	1.92x10 ⁻¹	2.66x10 ⁻¹
Chol HDL.1	16.89 [13.783 19.642]	20.415 [15.768 26.862]	-0.28	1.92x10 ⁻¹	2.66x10 ⁻¹
Apo.A1 HDL.4	73.3 [63.96 83.575]	66.545 [59.858 75.243]	0.13	2.35x10 ⁻¹	3.23x10 ⁻¹
MP LDL.Chol	109.14 [99.135 134.655]	103.665 [83.118 113.803]	0.19	2.59x10 ⁻¹	3.39x10 ⁻¹
LDL.2 PN	152.755 [114.168 179.565]	187.735 [143.328 209.08]	-0.25	2.59x10 ⁻¹	3.39x10 ⁻¹
MF Chol LDL	109.14 [99.135 134.655]	103.665 [83.118 113.803]	0.19	2.59x10 ⁻¹	3.39x10 ⁻¹
Apo.B LDL.2	8.4 [6.28 9.872]	10.325 [7.88 11.495]	-0.25	2.59x10 ⁻¹	3.39x10 ⁻¹
MP Apo.A1	143.235 [136.41 150.05]	138.3 [133.245 148.077]	0.07	3.12x10 ⁻¹	4.04x10 ⁻¹
Trigl HDL.1	3.54 [2.73 3.958]	2.96 [1.878 3.848]	0.49	3.41x10 ⁻¹	4.36x10 ⁻¹
FreeChol LDL.3	1.68 [1.388 1.87]	1.76 [1.452 2.138]	-0.16	3.96x10 ⁻¹	5.00x10 ⁻¹
LDL.3 PN	131.705 [37.047 171.588]	146.74 [108.203 174.567]	-0.36	4.03x10 ⁻¹	5.00x10 ⁻¹
Apo.B LDL.3	7.245 [2.04 9.44]	8.07 [5.95 9.598]	-0.36	4.03x10 ⁻¹	5.00x10 ⁻¹
Chol LDL.3	11.685 [2.105 15.377]	13.41 [8.697 15.675]	-0.35	4.29x10 ⁻¹	5.26x10 ⁻¹
Chol HDL.3	8.86 [8.398 10.175]	9.345 [8.625 10.555]	-0.05	4.72x10 ⁻¹	5.72x10 ⁻¹
Trigl LDL.3	2.395 [1.852 2.637]	2.475 [1.735 2.86]	-0.11	5.78x10 ⁻¹	6.94x10 ⁻¹
MF Phosp LDL	59.635 [54.36 74.132]	59.185 [49.73 65.748]	0.09	5.85x10 ⁻¹	6.95x10 ⁻¹
Phosp HDL.3	15.095 [14.135 16.82]	14.365 [13.732 17.263]	0.01	6.25x10 ⁻¹	7.35x10 ⁻¹
FreeChol LDL.4	3.065 [1.095 4.562]	3.54 [2.478 4.383]	-0.14	6.82x10 ⁻¹	7.93x10 ⁻¹
Chol LDL.1	18.885 [15.72 25.035]	20.14 [13.995 26.125]	-0.03	7.52x10 ⁻¹	8.66x10 ⁻¹
Apo.A1 HDL.2	17.805 [15.957 21.42]	18.075 [15.985 20.025]	0.06	8.38x10 ⁻¹	9.43x10 ⁻¹
Phosp HDL.4	26.26 [20.55 29.66]	23.26 [21.895 27.84]	0.06	8.41x10 ⁻¹	9.43x10 ⁻¹
LDL.4 PN	132.31 [49.392 180.19]	126.205 [59.178 156.785]	0.11	8.61x10 ⁻¹	9.43x10 ⁻¹
Phosp LDL.4	5.705 [1.485 8.697]	6.23 [3.59 7.855]	-0.08	8.61x10 ⁻¹	9.43x10 ⁻¹
Apo.B LDL.4	7.275 [2.72 9.91]	6.945 [3.255 8.625]	0.11	8.61x10 ⁻¹	9.43x10 ⁻¹
LDL.1 PN	202.365 [170.42 230.515]	209.91 [145.275 253.295]	-0.05	8.86x10 ⁻¹	9.53x10 ⁻¹
Apo.B LDL.1	11.13 [9.37 12.68]	11.545 [7.992 13.93]	-0.05	8.86x10 ⁻¹	9.53x10 ⁻¹
MF FreeChol LDL	28.15 [25.31 36.675]	29.46 [25.735 33.145]	0	9.07x10 ⁻¹	9.57x10 ⁻¹
Apo.A2 HDL.1	2.71 [2.225 3.487]	2.98 [2.49 3.685]	0.13	9.07x10 ⁻¹	9.57x10 ⁻¹
MF Apo.A1 HDL	140.92 [136.89 147.097]	142.275 [135.855 148.532]	0.01	9.31x10 ⁻¹	9.65x10 ⁻¹
FreeChol HDL.4	3.055 [1.69 3.705]	2.53 [2.2 3.5]	-0.01	9.31x10 ⁻¹	9.65x10 ⁻¹
Chol LDL.4	10.105 [2.053 16.44]	10.1 [5.402 13.578]	0.04	9.53x10 ⁻¹	9.70x10 ⁻¹
Phosp LDL.1	10.87 [9.21 13.52]	11.94 [8.565 14.76]	-0.09	9.53x10 ⁻¹	9.70x10 ⁻¹
FreeChol LDL.1	5.19 [4.58 7.225]	6.675 [4.457 7.775]	-0.09	9.77x10 ⁻¹	9.77x10 ⁻¹
Chol HDL.4	19.02 [14.853 22.042]	18.695 [16.23 22.68]	0.01	9.77x10 ⁻¹	9.77x10 ⁻¹

*Abbreviations are reported as follows: MF: Main Fractions, MP: Main Parameters, PN: Particle Number, Chol: Cholesterol, Phosp: Phospholipids, Trigl: Triglycerides.

- 4.1.5. Untargeted NMR-based metabolomics to investigate the effect of Bioactive Foods enriched with combination of DHA and anthocyanins or oat β -glucan on serum metabolome and lipidome of subjects at risk for metabolic syndrome: Large Intervention Study from the EC FP7 Pathway-27 project.**

In preparation

Candidate's contributions: acquisition of NMR spectra, statistical analysis of NMR data, interpreting results and writing of the manuscript.

Introduction

Genetic factors and aging have important roles in determining the overall risk of many chronic diseases. A considerable proportion of these diseases, mostly cardiovascular disorders, occurs also with various modifiable risk factors which depend on lifestyle habits, including physical exercises and diet.^{1,2}

In the last years, this awareness increased the global interest on improving eating habits and maintaining a healthy lifestyle. In a renewed Hippocratic perception of “*food as a medicament*”, pharmaceutical and food companies pay increasing attention to bioactives and to the preparation of bioactive enriched foods with the overall aim of improving human health.

Bioactives, or nutraceuticals, are any substances that are foods or part of a food, able to provide health benefits, also in the frame of disease treatment and prevention.³ Commonly, bioactives are present in foods at low concentrations, far from the effective dose. To obtain the desired effect, bioactives must be delivered in more concentrated amount, as in bioactive-enriched foods (BEF),⁴ therefore representing a promising approach to prevent and to manage metabolic disorders, including the metabolic syndrome (MetS).⁵⁻⁷

However, the optimal formulation of an effective BEF depends on a better understanding of the complex inter-relationship between the bioactive and the food matrix and between food structure and performance. In this light, there is an increasing need of better understanding whether the single bioactive or related combination with other nutraceuticals can exert beneficial and synergic roles on human health when administered as ingredients of specific BEFs. It cannot be assumed *a priori* that the food matrix by which bioactives are conveyed does not influence the final efficacy of the compound.

Docosahexaenoic acid (C22:6, *n*-3, DHA), anthocyanins (AC) and β -glucans (BG) are well-known bioactives able to have a positive impact on MetS,⁸⁻¹⁰ acting on cell regulation of lipid metabolism, mitigating inflammation, and being effective in modulating risks of metabolic syndrome. There are evidences of synergies between the above-mentioned bioactives,^{7,11,12} but the effects of the food matrices on bioactives combinations (*e.g.* DHA + AC or DHA + O-BG) need to be investigated more deeply. It has been previously evidenced that NMR-based metabolomics, applied to the analysis of human biofluids, is efficient and highly reproducible in assessing novel biomarkers of dietary intake,¹³⁻¹⁵ and in determining the existence of human metabolic phenotypes,¹⁶⁻¹⁹ where nutrition factors and dietary habits can contribute to the movements of an individual in the metabolic space,¹⁷ thus offering a broader understanding of the complex variations due to the diet in a foodomic vision.²⁰

The present large intervention study was included in the frame of the EC FP7 PATHWAY-27 project (ID: 311876) with the aim to confirm or to better investigate the observations obtained in the pilot studies,^{6,7} where the best enrichments (DHA, AC and oat BG (O-BG)) within the optimal food matrices (dairy-, egg- and bakery- based foods) were selected.

In particular, we applied untargeted NMR-based metabolomics to investigate, on the human serum metabolome of subjects at risk for MetS: *i*) the effect of dairy-based food enriched with DHA and O-BG; *ii*) the effect of egg-based food enriched with DHA and AC; *iii*) the effect of bakery-based food enriched with DHA and AC; *iv*) the effect of food matrices in determining the bioavailability of bioactives, administering a combination of dairy-, egg- and bakery-based foods without bioactives (placebo). To this end, we analysed serum samples collected before (t_0) and 12 weeks after (t_1) the dietary interventions.

Material and Methods

Bioactive Enriched Food

Three different foods (milkshake, pancake and biscuits), related to different food matrices (dairy, egg and bakery), were enriched with DHA + O-BG, DHA + AC and DHA + AC respectively.

DHA, O-BG and AC were obtained and BEFs were formulated as detailed in the pilot study.⁷ Placebo dairy, egg and bakery-based foods were also produced.

Study population

The population under study includes a total of 232 healthy male and female subjects aged 18 years and above, and presenting 2, 3 or 4 of the criteria for MetS diagnosis, *i.e.* elevated waist circumference ≥ 102 cm (men) or ≥ 88 cm (women); fasting triglycerides ≥ 150 mg/dL; fasting HDL-cholesterol ≤ 40 mg/dL (men) or ≤ 50 mg/dL (women); systolic blood pressure ≥ 130 mmHg and/or diastolic blood pressure ≥ 85 mmHg or hypotensive treatment; fasting glucose ≥ 100 mg/dL]).^{6,7}

Enrolled subjects are from 4 different recruitment centres: *i*) Italy, University of Bologna (UNIBO), *ii*) France, Centre de Recherche en Nutrition Humaine Auvergne, Clermont-Ferrand (CRNH); *iii*) Germany, Max Rubner-Institut, Karlsruhe (MRI) and *iv*) UK, School of Food Science and Nutrition, University of Leeds (ULE).

The three most effective BEFs (one for each food matrix) from the pilot studies^{6,7} were used in the present randomized, double-blind, placebo-controlled dietary large intervention study (LIS).

Volunteers were randomly divided into 4 treatment-groups from 50 to 70 subjects, each receiving daily, for 12 weeks, one of the following dietary intervention: a) all bakery, egg, dairy-based foods without bioactives (placebo); b) dairy-based BEF + placebo bakery and placebo egg-based foods; c) egg-based BEF + placebo bakery and placebo dairy-based foods; d) bakery-based BEF + placebo dairy and egg-based foods. During the course of the dietary intervention, volunteers were on a free diet, they had only to limit the consumption of foods naturally containing high quantities of the bioactives under study (DHA, AC and O-BG) to one portion per day, and they were required to maintain their usual lifestyle. Prior to the beginning of the study, volunteers' lifestyle and dietary habits were recorded and the volunteers underwent physical examination after 6 and 12 weeks (middle point and endpoint of the study),

and routine blood lab tests (*e.g.* HDL-C, LDL-C, TG, CRP, creatinine etc.). Serum samples for metabolomics were collected before (t_0) and at the end of the study (t_1). An overview of the study design is reported in **Figure 1**, while the number of participants receiving each treatment (a-d) and having suitable samples for the NMR-metabolomics analysis at t_0 and t_1 is reported in **Table 1**.

Ethical Issues

All subjects gave their informed consent for inclusion before they participated in the study, that was conducted in accordance with the Declaration of Helsinki. The study was approved by the local Ethics Committees.

NMR sample preparation and analysis

Both t_0 and t_1 serum samples were collected and prepared according to common standard operating procedures for metabolomic studies.^{21–23} The analytical preparation of serum samples and their NMR spectra acquisition followed the protocols detailed elsewhere.²²

Briefly, for each serum specimen, three one-dimensional proton NMR spectra were acquired with different pulse sequences (*i.e.* 1D NOESY, 1D CPMG and 1D DIFFUSION-EDITED)^{24–26}, allowing the selective detection of different molecular components. All ¹H-NMR spectra were collected using a Bruker 600 MHz spectrometer, with a proton Larmor frequency of 600.13 MHz and equipped with a 5 mm PATXI 1H-13C-15N and 2H decoupling probe. This includes a z axis gradient coil, an automatic tuning-matching (ATM) and an automatic and refrigerate sample changer (SampleJet). To stabilize approximately, at the level of ± 0.1 K, the sample temperature, a BTO 2000 thermocouple was employed and each NMR tube was kept for about 5 min inside the NMR probe head to equilibrate the acquisition temperature of 310 K.

NMR Spectral Processing

Before applying Fourier transform, raw data were multiplied by an exponential function of 0.3 Hz line-broadening factor. Transformed spectra were automatically corrected for phase and baseline distortions and 1D-NOESY and 1D-CPMG spectra were calibrated to a reference (anomeric glucose proton signal at 5.24 ppm), using Topspin 3.2 software (Bruker BioSpin).

Statistical analysis

All data analyses were performed using R (version 3.6.1), an open source software for the statistical management of data.²⁷ Multivariate data analysis was conducted, without prior normalization, on 1D-NOESY NMR spectra, bucketed into 0.02 ppm chemical shift segments in the range of 0.2 – 10.0 ppm, using AMIX (version 3.8.4) software (Bruker BioSpin). Regions containing residual water signal (between 4.4 and 5.0 ppm) were removed.

Exploratory Analysis

Principal Component Analysis (PCA) was used as a first exploratory approach²⁸ to investigate, in an unsupervised manner, the data structure and highlighting the effect of DHA in combination with O-BG or AC in inducing metabolic changes in subjects at risk for MetS. PCA analysis was performed on data scaled to unit variance.

Predictive modelling: Unpaired and Paired Analysis

Firstly, canonical (CA) analysis was used in combination with PCA to perform supervised discrimination among treatments at t_0 at t_1 .

Secondly, Multilevel Partial Least Squares (M-PLS)²⁹ analysis was employed to obtain a paired data reduction and classification in order to consider the effect of the treatments within each subject variation, thus excluding the inter-individual variability.

For each model, global accuracy for classification was estimated by using a Monte Carlo validation scheme. Briefly, 90% of data from each NMR dataset were randomly chosen at each iteration as a training set to build the model. Then, the remaining 10% was tested. The full procedure was repeated 100 times to derive an average discrimination accuracy.

Serum and lipoprotein identification and quantification

24 metabolites (**Table 2**) and 114 lipids were unambiguously identified and quantified (in terms of absolute concentrations) from 1D ¹H-NOESY NMR spectra, using the AVANCE Bruker IVDr (Clinical Screening and In Vitro Diagnostics research, Bruker BioSpin)³⁰ software. For all serum samples, different parameters (*i.e.* triglycerides, cholesterol, phospholipids etc.) related to main lipoproteins (HDL, IDL, LDL and VLDL) and to lipoprotein subclasses (classified according to density and size, for a total of 15 subclasses: VLDL-1 to VLDL-5, LDL-1 to LDL-6 and HDL-1 to HDL-4), were detected. In detail, for each main class and subclass, reported data consist of concentrations of lipids (total cholesterol, free cholesterol, phospholipids, and triglycerides) contained in each fraction. Concentrations of apolipoproteins Apo-A1 and ApoA2 are estimated for HDL class and each relative subclass, while Apo-B concentrations are calculated for VLDL, IDL classes and all LDL subclasses.

Univariate Analysis

Paired univariate Wilcoxon test³¹ was employed to compare metabolite and lipid concentrations between t_0 and t_1 for each treatment group, *i.e.* a) all foods as placebo, b) dairy-based BEF + placebo bakery and placebo egg-based foods, c) egg-based BEF + placebo bakery and d) placebo dairy-based foods, bakery-based BEF + placebo dairy and egg-based foods).

Benjamini & Hochberg method³² was applied to correct for multiple testing and adjusted *P*-values (FDR) < 0.05 were considered statistically significant. Log₂ fold

change (FC) ratios of the median intensities were calculated for all comparisons we considered.

Results

Exploratory analysis

Untargeted NMR-based metabolomic analysis on 1D-NOESY spectra was used to detect the metabolic effect of the above-described dietary interventions on the serum metabolome of subjects at risk for MetS. Principal Component Analysis (PCA) was firstly used to assess the quality of the overall spectral dataset. The resulting 3D score plot (**Figure 2C**) evidences that all samples, although collected in different recruitment centres, were homogenous to each other, and no evident outliers were detected. The uniform distribution of samples can be easily visualized by colouring the scores also according to time-points of blood collections, *i.e.* t_0 , t_1 (**Figure 2A**) and according to the different dietary interventions (**Figure 2B**).

Predictive modelling: unpaired analysis

Applying a supervised multivariate analysis to explore possible differences among each dietary treatment, no meaningful source of variation, in the serum metabolome of recruited subjects, was highlighted at t_0 . Indeed, the PCA-CA discrimination model showed a discrimination accuracy, among each dietary treatment, at most 20% (Supplementary files, **Table S1**). Instead, at t_1 , the overall discrimination among groups raised up to 34% (Supplementary files, **Table S2**), demonstrating the presence of changes in the serum metabolomic profiles of subjects due to the dietary interventions.

Predictive modelling: paired analysis

A paired (before *vs.* after treatment) M-PLS analysis was applied to explore the existence of a strong individual variability as a response to each single dietary treatment.

Considering the whole dataset, we obtained a good classification accuracy of 80% between t_0 and t_1 . An illustration of the M-PLS classification accuracy between t_0 and t_1 for each recruited subject is reported in **Figure 3**. We observed that almost all of the subjects are well classified: this means that a common pattern of biomarkers recorded in the serum spectra of most of the individuals is selected by the predictive model, reflecting a recurrent change in the metabolome profiles when passing from the baseline to the end of the study.

To assess whether the observed metabolic changes are merely unspecific phenomena related to the administration of treatments, or if there are bioactive-enriched foods more active than others, M-PLS analysis was applied considering each recruitment center separately. The resulting classification performances are reported in **Table 3**.

The combination between DHA and AC embedded in bakery-based BEF appeared as the most effective for the French and English cohorts, while for German and Italian

cohorts, the highest discrimination accuracy between t_0 and t_1 was obtained for the combination DHA + AC in egg-based BEFs. These last results suggest that these differences, related to the effect of food matrices by which bio-actives are conveyed, depends on the typical everyday diet of each different country from where subjects are from.

Subsequently, to identify the signals in the NMR spectra mainly related to the DHA + AC effects on the serum metabolome of the individuals, we deeply explored the above-mentioned M-PLS classifications, as illustrated in the loading plot of the first component in **Figure 4** for DHA + AC embedded in bakery-based BEF for French (**Figure 4A**) and English (**Figure 4B**) cohorts and in **Figure 5** for egg-based food matrix enriched with the same combination of bioactives for Italian (**Figure 5A**) and German (**Figure 5B**) cohorts. We observed that, independently of the cohort, the most significant spectral buckets belong to the broad signals of methyl (-CH₃) and methylene (-CH₂-) groups of serum lipoproteins, respectively centred at almost 0.86 and 1.29 ppm. These results evidence that DHA + AC embedded in bakery and egg-based food matrices significantly induces changes in the lipoprotein profiles of the subjects.

Univariate analysis

To deeply investigate the observed variations, metabolites and lipids were compared between t_0 and t_1 , considering firstly, all recruitment centres and treatments together, and secondly separating by the specific treatment.

Considering the whole dataset, the paired analysis of serum profiles of each individual revealed glycine, N,N-dimethylglycine, glutamine, IDL particles and apolipoprotein B related to IDL as the most statistically significant decrease (FDR < 0.05) analytes at t_1 . Subsequently, for each treatment-subgroup, we did not obtain any significant changes between the two time-points in the case of placebo and dairy-based BEFs consumptions. Comparing at t_0 and t_1 , the serum profiles of subjects undergoing egg- and bakery-based BEFs assumptions, we noted a statistically significant decrease of IDL related parameters, *i.e.* IDL cholesterol, free cholesterol, apolipoproteins B and also triglycerides associated to LDL-4 sub-particles.

Instead, N, N-dimethylglycine and phospholipids related to HDL-1 were deemed to be statistically significantly increased (FDR < 0.05) at t_1 when individuals assumed the bakery-based BEFs treatment.

All of these results confirm that DHA + AC administration induces significant variations mostly in lipoproteins profiles (**Table 4**), while only slight changes in the metabolome between the beginning and the end of the study were detected (**Table S3**).

Discussion

In this multicentric study, we characterized, by untargeted NMR-based metabolomics, the metabolic effects of DHA supplementation in combination with O-

BG and AC, in three different BEFs (milkshake, pancake and biscuits), one for each food matrix (dairy, egg and bakery).

Among potentially useful bioactives, DHA is well known to lower triglycerides levels, one of the main components of the lipoproteins, especially VLDL³³; to increase HDL-cholesterol³⁴; to modulate insulin resistance³⁵ and to counteract oxidative stress.³⁶ Moreover, evidences of DHA effects on the lipidome and metabolome of human serum and hepatocytes have been recently reported.^{7,37}

Extensive data from clinical trials and epidemiological studies suggest that *n*-3 PUFA favourably modulate multiple biological processes, but the interpretation on the relevance of DHA-enriched foods is complicated by the interaction of DHA with the food matrix and other bioactives and its effective dose. Synergies among bioactives were demonstrated¹¹ and Toufekstian *et al.*¹² showed that anthocyanin supplementation increased DHA plasma levels, suggesting that AC co-administered with DHA could have beneficial roles via increasing DHA effects.

Our results support a more pronounced synergic effect in the case of co-administration of DHA + AC. The combination between the two bioactives resulted to induce more variations on serum metabolome and lipoprotein profiles of subjects at risk for MetS, differently from what previously observed in the pilot study,⁷ where the addition of O-BG to DHA supplementation in BEFs resulted as the most promising in determining rearrangement of lipoprotein profiles (**Table 4**). Past studies have reported evidences on the actions of AC on cell metabolism, especially concerning the regulation of lipid metabolism and inflammation,^{38,39} highlighting AC as a good candidate for the prevention of metabolic syndrome.

Furthermore, the randomized and double-blinded administration of foods without bioactives (placebo) permitted to explore the effect of different food matrices fully, letting us to hypothesize that the food matrix by which bioactives are conveyed has a role in determining the variation of subjects' serum metabolome between the beginning and the end of the study. This can be justified considering firstly, that the food matrix can exert a modification of bioavailability of DHA, AC, O-BG and their co-presence. For example, it was demonstrated that the presence of DHA in the formulation of bakery products significantly decrease the bioavailability of AC.⁴⁰ Secondly, it cannot be excluded an effect related to the usual everyday diet of subjects enrolled from different country with different dietary habits.

Conclusions

To date, very few intervention studies comparing the effect of bioactives embedded into different food matrices are reported in the literature, and demonstrations of synergies among bioactives are scarce. Here, our results evidenced the potentiality of untargeted NMR-based metabolomics in characterizing variations in the serum metabolome and lipoprotein profiles of enrolled subjects, consequently to the administration of selected bioactives combinations in three different BEFs, *i.e.* milkshake, pancake and biscuits.

Applying multivariate and univariate analyses, we demonstrated a synergism of DHA and AC in inducing changes mainly in the lipid profiles, independently of the egg- or bakery-based food matrix.

Modifications observed in serum profiles were consistent with known clinical effects of DHA and AC, but they did not fully confirm what previously discussed in the pilot study, where a strong and cooperative synergic effect was monitored in the case of co-administration of DHA and O-BG.

In conclusion, this study deepened the current scientific understanding of the impact of different bioactives embedded in various food matrices on the serum metabolome, evidencing their synergies and proposing actual use of bakery- and egg-based foods enriched with DHA and AC on subjects at risk for metabolic syndrome.

References

- (1) Artinian, N. T.; Fletcher, G. F.; Mozaffarian, D.; Kris-Etherton, P.; Van Horn, L.; Lichtenstein, A. H.; Kumanyika, S.; Kraus, W. E.; Fleg, J. L.; Redeker, N. S.; Meininger, J. C.; Banks, J.; Stuart-Shor, E. M.; Fletcher, B. J.; Miller, T. D.; Hughes, S.; Braun, L. T.; Kopin, L. A.; Berra, K.; Hayman, L. L.; Ewing, L. J.; Ades, P. A.; Durstine, J. L.; Houston-Miller, N.; Burke, L. E.; American Heart Association Prevention Committee of the Council on Cardiovascular Nursing. Interventions to Promote Physical Activity and Dietary Lifestyle Changes for Cardiovascular Risk Factor Reduction in Adults: A Scientific Statement from the American Heart Association. *Circulation* 2010, 122 (4), 406–441. <https://doi.org/10.1161/CIR.0b013e3181e8edf1>.
- (2) Galimanis, A.; Mono, M.-L.; Arnold, M.; Nedeltchev, K.; Mattle, H. P. Lifestyle and Stroke Risk: A Review. *Curr. Opin. Neurol.* 2009, 22 (1), 60–68. <https://doi.org/10.1097/WCO.0b013e32831fda0e>.
- (3) Alissa, E. M.; Ferns, G. A. Functional Foods and Nutraceuticals in the Primary Prevention of Cardiovascular Diseases. *J. Nutr. Metab.* 2012, 2012, 569486. <https://doi.org/10.1155/2012/569486>.
- (4) Rajasekaran, A.; Kalaivani, M. Designer Foods and Their Benefits: A Review. *J. Food Sci. Technol.* 2013, 50 (1), 1–16. <https://doi.org/10.1007/s13197-012-0726-8>.
- (5) Khan, M. I.; Anjum, F. M.; Sohaib, M.; Sameen, A. Tackling Metabolic Syndrome by Functional Foods. *Rev. Endocr. Metab. Disord.* 2013, 14 (3), 287–297. <https://doi.org/10.1007/s11154-013-9270-8>.
- (6) Bub, A.; Malpuech-Brugère, C.; Orfila, C.; Amat, J.; Arianna, A.; Blot, A.; Di Nunzio, M.; Holmes, M.; Kertész, Z.; Marshall, L.; Nemeth, I.; Ricciardiello, L.; Seifert, S.; Sutulic, S.; Ulaszewska, M.; Bordoni, A. A Dietary Intervention of Bioactive Enriched Foods Aimed at Adults at Risk of Metabolic Syndrome: Protocol and Results from PATHWAY-27 Pilot Study. *Nutrients* 2019, 11 (8). <https://doi.org/10.3390/nu11081814>.
- (7) Ghini, V.; Tenori, L.; Capozzi, F.; Luchinat, C.; Bub, A.; Malpuech-Brugere, C.; Orfila, C.; Ricciardiello, L.; Bordoni, A. DHA-Induced Perturbation of Human

Serum Metabolome. Role of the Food Matrix and Co-Administration of Oat β -Glucan and Anthocyanins. *Nutrients* 2019, 12 (1). <https://doi.org/10.3390/nu12010086>.

(8) Guo, X.-F.; Li, X.; Shi, M.; Li, D. N-3 Polyunsaturated Fatty Acids and Metabolic Syndrome Risk: A Meta-Analysis. *Nutrients* 2017, 9 (7). <https://doi.org/10.3390/nu9070703>.

(9) Brown, L.; Poudyal, H.; Panchal, S. K. Functional Foods as Potential Therapeutic Options for Metabolic Syndrome. *Obes. Rev. Off. J. Int. Assoc. Study Obes.* 2015, 16 (11), 914–941. <https://doi.org/10.1111/obr.12313>.

(10) Cloetens, L.; Ulmius, M.; Johansson-Persson, A.; Akesson, B.; Onning, G. Role of Dietary Beta-Glucans in the Prevention of the Metabolic Syndrome. *Nutr. Rev.* 2012, 70 (8), 444–458. <https://doi.org/10.1111/j.1753-4887.2012.00494.x>.

(11) García-Alonso, F. J.; Jorge-Vidal, V.; Ros, G.; Periago, M. J. Effect of Consumption of Tomato Juice Enriched with N-3 Polyunsaturated Fatty Acids on the Lipid Profile, Antioxidant Biomarker Status, and Cardiovascular Disease Risk in Healthy Women. *Eur. J. Nutr.* 2012, 51 (4), 415–424. <https://doi.org/10.1007/s00394-011-0225-0>.

(12) Toufektsian, M.-C.; Salen, P.; Laporte, F.; Tonelli, C.; de Lorgeril, M. Dietary Flavonoids Increase Plasma Very Long-Chain (n-3) Fatty Acids in Rats. *J. Nutr.* 2011, 141 (1), 37–41. <https://doi.org/10.3945/jn.110.127225>.

(13) Gibney, M. J.; Walsh, M.; Brennan, L.; Roche, H. M.; German, B.; van Ommen, B. Metabolomics in Human Nutrition: Opportunities and Challenges. *Am. J. Clin. Nutr.* 2005, 82 (3), 497–503.

(14) Brennan, L. Metabolomics in Nutrition Research: Current Status and Perspectives. *Biochem. Soc. Trans.* 2013, 41 (2), 670–673. <https://doi.org/10.1042/BST20120350>.

(15) O'Sullivan, A.; Gibney, M. J.; Brennan, L. Dietary Intake Patterns Are Reflected in Metabolomic Profiles: Potential Role in Dietary Assessment Studies. *Am. J. Clin. Nutr.* 2011, 93 (2), 314–321. <https://doi.org/10.3945/ajcn.110.000950>.

(16) Assfalg, M.; Bertini, I.; Colangiuli, D.; Luchinat, C.; Schäfer, H.; Schütz, B.; Spraul, M. Evidence of Different Metabolic Phenotypes in Humans. *Proc. Natl. Acad. Sci.* 2008, 105 (5), 1420–1424. <https://doi.org/10.1073/pnas.0705685105>.

(17) Bernini, P.; Bertini, I.; Luchinat, C.; Nepi, S.; Saccenti, E.; Schäfer, H.; Schütz, B.; Spraul, M.; Tenori, L. Individual Human Phenotypes in Metabolic Space and Time. *J. Proteome Res.* 2009, 8 (9), 4264–4271. <https://doi.org/10.1021/pr900344m>.

(18) Ghini, V.; Saccenti, E.; Tenori, L.; Assfalg, M.; Luchinat, C. Allostasis and Resilience of the Human Individual Metabolic Phenotype. *J. Proteome Res.* 2015, 14 (7), 2951–2962. <https://doi.org/10.1021/acs.jproteome.5b00275>.

(19) Saccenti, E.; Menichetti, G.; Ghini, V.; Remondini, D.; Tenori, L.; Luchinat, C. Entropy-Based Network Representation of the Individual Metabolic Phenotype. *J. Proteome Res.* 2016, 15 (9), 3298–3307. <https://doi.org/10.1021/acs.jproteome.6b00454>.

- (20) Bordoni, A.; Capozzi, F. Foodomics for Healthy Nutrition. *Curr. Opin. Clin. Nutr. Metab. Care* 2014, 17 (5), 418–424. <https://doi.org/10.1097/MCO.0000000000000089>.
- (21) Bernini, P.; Bertini, I.; Luchinat, C.; Nincheri, P.; Staderini, S.; Turano, P. Standard Operating Procedures for Pre-Analytical Handling of Blood and Urine for Metabolomic Studies and Biobanks. *J. Biomol. NMR* 2011, 49 (3–4), 231–243. <https://doi.org/10.1007/s10858-011-9489-1>.
- (22) Vignoli, A.; Ghini, V.; Meoni, G.; Licari, C.; Takis, P. G.; Tenori, L.; Turano, P.; Luchinat, C. High-Throughput Metabolomics by 1D NMR. *Angew. Chem. Int. Ed Engl.* 2019, 58 (4), 968–994. <https://doi.org/10.1002/anie.201804736>.
- (23) Ghini, V.; Quaglio, D.; Luchinat, C.; Turano, P. NMR for Sample Quality Assessment in Metabolomics. *New Biotechnol.* 2019, 52, 25–34. <https://doi.org/10.1016/j.nbt.2019.04.004>.
- (24) 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. <https://doi.org/10.1002/cmr.a.20223>.
- (25) Meiboom, S.; Gill, D. Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. *Rev. Sci. Instrum.* 1958, 29 (8), 688–691.
- (26) Wu, D. H.; Chen, A. D. Three-Dimensional Diffusion-Ordered NMR Spectroscopy: The Homonuclear COSY-DOSY Experiment. *J Magnen Reson A* 1996, 123, 215–218.
- (27) Ihaka, R.; Gentleman, R. R. A Language for Data Analysis and Graphics. *J Comput Stat Graph* 1996, 5, 299–314.
- (28) Serneels, S.; Verdonck, T. Principal Component Analysis for Data Containing Outliers and Missing Elements. *Comput. Stat. Data Anal.* 2008, 52 (3), 1712–1727.
- (29) Westerhuis, J. A.; van Velzen, E. J.; Hoefsloot, H. C.; Smilde, A. K. Multivariate Paired Data Analysis: Multilevel PLSDA versus OPLSDA. *Metabolomics* 2010, 6 (1573-3890 (Electronic)), 119–128.
- (30) Jiménez, B.; Holmes, E.; Heude, C.; Tolson, R. F.; Harvey, N.; Lodge, S. L.; Chetwynd, A. J.; Cannet, C.; Fang, F.; Pearce, J. T. M.; Lewis, M. R.; Viant, M. R.; Lindon, J. C.; Spraul, M.; Schäfer, H.; Nicholson, J. K. Quantitative Lipoprotein Subclass and Low Molecular Weight Metabolite Analysis in Human Serum and Plasma by 1H NMR Spectroscopy in a Multilaboratory Trial. *Anal. Chem.* 2018, 90 (20), 11962–11971. <https://doi.org/10.1021/acs.analchem.8b02412>.
- (31) Neuhäuser, M. Wilcoxon–Mann–Whitney Test. In *International Encyclopedia of Statistical Science*; Springer, Berlin, Heidelberg, 2011; pp 1656–1658. https://doi.org/10.1007/978-3-642-04898-2_615.
- (32) Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* 1995, 289–300.
- (33) Bays, H. E.; Tighe, A. P.; Sadovsky, R.; Davidson, M. H. Prescription Omega-3 Fatty Acids and Their Lipid Effects: Physiologic Mechanisms of Action and Clinical

Implications. *Expert Rev. Cardiovasc. Ther.* 2008, 6 (3), 391–409. <https://doi.org/10.1586/14779072.6.3.391>.

(34) Jacobson, T. A.; Glickstein, S. B.; Rowe, J. D.; Soni, P. N. Effects of Eicosapentaenoic Acid and Docosahexaenoic Acid on Low-Density Lipoprotein Cholesterol and Other Lipids: A Review. *J. Clin. Lipidol.* 2012, 6 (1), 5–18. <https://doi.org/10.1016/j.jacl.2011.10.018>.

(35) Fedor, D.; Kelley, D. S. Prevention of Insulin Resistance by N-3 Polyunsaturated Fatty Acids. *Curr. Opin. Clin. Nutr. Metab. Care* 2009, 12 (2), 138–146. <https://doi.org/10.1097/MCO.0b013e3283218299>.

(36) Türkez, H.; Geyikoglu, F.; Yousef, M. I. Ameliorative Effect of Docosahexaenoic Acid on 2,3,7,8-Tetrachlorodibenzo-p-Dioxin-Induced Histological Changes, Oxidative Stress, and DNA Damage in Rat Liver. *Toxicol. Ind. Health* 2012, 28 (8), 687–696. <https://doi.org/10.1177/0748233711420475>.

(37) Ghini, V.; Di Nunzio, M.; Tenori, L.; Valli, V.; Danesi, F.; Capozzi, F.; Luchinat, C.; Bordoni, A. Evidence of a DHA Signature in the Lipidome and Metabolome of Human Hepatocytes. *Int. J. Mol. Sci.* 2017, 18 (2). <https://doi.org/10.3390/ijms18020359>.

(38) Qin, Y.; Xia, M.; Ma, J.; Hao, Y.; Liu, J.; Mou, H.; Cao, L.; Ling, W. Anthocyanin Supplementation Improves Serum LDL- and HDL-Cholesterol Concentrations Associated with the Inhibition of Cholesteryl Ester Transfer Protein in Dyslipidemic Subjects. *Am. J. Clin. Nutr.* 2009, 90 (3), 485–492. <https://doi.org/10.3945/ajcn.2009.27814>.

(39) Wei, X.; Wang, D.; Yang, Y.; Xia, M.; Li, D.; Li, G.; Zhu, Y.; Xiao, Y.; Ling, W. Cyanidin-3-O- β -Glucoside Improves Obesity and Triglyceride Metabolism in KK-Ay Mice by Regulating Lipoprotein Lipase Activity. *J. Sci. Food Agric.* 2011, 91 (6), 1006–1013. <https://doi.org/10.1002/jsfa.4275>.

(40) Karakaya, S.; Simsek, S.; Eker, A. T.; Pineda-Vadillo, C.; Dupont, D.; Perez, B.; Viadel, B.; Sanz-Buenhombre, M.; Rodriguez, A. G.; Kertész, Z.; Hegyi, A.; Bordoni, A.; El, S. N. Stability and Bioaccessibility of Anthocyanins in Bakery Products Enriched with Anthocyanins. *Food Funct.* 2016, 7 (8), 3488–3496. <https://doi.org/10.1039/c6fo00567e>.

Tables

Table 1. Number of study participants from each recruitment center (RC) analyzed by NMR-metabolomics according to dietary treatment: *a*) all bakery, egg, dairy-based foods without bioactives (placebo); *b*) dairy-based BEF + placebo bakery and placebo egg-based foods; *c*) egg-based BEF + placebo bakery and placebo dairy-based foods; *d*) bakery-based BEF + placebo dairy and egg-based foods.

RC	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	Tot
All RC	70	55	53	54	232
France (CRNH)	16	12	19	14	61
Germany (MRI)	15	13	8	8	44
Italy (UNIBO)	23	18	15	17	73
UK (ULE)	16	12	11	15	54

Table 2. List of identified metabolites assigned in serum samples. The MSI levels of identification and the compound IDs are provided from the Human Metabolome Database (HMDB).

Number	Metabolite	MSI level of identification	Database	Compound ID
1	Acetate	1	HMDB	HMDB00042
2	Acetoacetate	1	HMDB	HMDB00060
3	Acetone	1	HMDB	HMDB01659
4	Alanine	1	HMDB	HMDB00161
5	Citrate	1	HMDB	HMDB00094
6	Creatine	1	HMDB	HMDB00064
7	Creatinine	1	HMDB	HMDB00562
8	Dimethyl sulfone	1	HMDB	HMDB04983
9	Formate	1	HMDB	HMDB00142
10	Glucose	1	HMDB	HMDB00122
11	Glutamine	1	HMDB	HMDB00641
12	Glycine	1	HMDB	HMDB00123
13	Histidine	1	HMDB	HMDB00177
14	Isoleucine	1	HMDB	HMDB00172
15	Lactate	1	HMDB	HMDB00190
16	Leucine	1	HMDB	HMDB00687

17	Methionine	1	HMDB	HMDB00696
18	N,N-dimethylglycine	1	HMDB	HMDB00092
19	Phenylalanine	1	HMDB	HMDB00159
20	Pyruvate	1	HMDB	HMDB00243
21	Succinate	1	HMDB	HMDB00254
22	Trimethylamine-N-oxide	1	HMDB	HMDB00925
23	Tyrosine	1	HMDB	HMDB00158
24	Valine	1	HMDB	HMDB00883

Table 3. Serum M-PLS classification accuracy values for the discrimination between t_0 and t_1 considering each dietary treatment-subgroup: *a*) all bakery, egg, dairy-based foods without bioactives (placebo); *b*) dairy-based BEF + placebo bakery and placebo egg-based foods; *c*) egg-based BEF + placebo bakery and placebo dairy-based foods; *d*) bakery-based BEF + placebo dairy and egg-based foods.

RC	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>
France (CRNH)	69% (16 subj)	83% (12 subj)	67% (19 subj)	84% (14 subj)
Germany (MRI)	77% (15 subj)	70% (13 subj)	100% (8 subj)	37% (8 subj)
Italy (UNIBO)	55% (23 subj)	74% (18 subj)	100% (16 subj)	82% (17 subj)
UK (ULE)	47% (16 subj)	52% (12 subj)	47% (10 subj)	89% (15 subj)

*Abbreviation: subj: subjects.

Table 4. Bruker IVDr lipoprotein analysis at t_0 and t_1 for enrolled subjects ($n=232$). Lipidic features were assigned and quantified in 1D NOESY NMR spectra and their absolute concentrations are reported as median \pm median absolute deviation. * is used to indicate parameters resulted to be statistically significant, with P -value < 0.05 , in the comparison t_0 vs. t_1 for each of the administered treatment: *a*) all bakery, egg, dairy-based foods without bioactives (placebo); *b*) dairy-based BEF + placebo bakery and placebo egg-based foods; *c*) egg-based BEF + placebo bakery and placebo dairy-based foods; *d*) bakery-based BEF + placebo dairy and egg-based foods). ** is used for parameters reporting both P -value < 0.05 and FDR < 0.05 .

	<i>a</i>		<i>b</i>		<i>c</i>		<i>d</i>	
	t_0	t_1	t_0	t_1	t_0	t_1	t_0	t_1
TG	180.4 \pm 61.3	193.5 \pm 71.1	195.2 \pm 67.2	190.7 \pm 66	200.7 \pm 60.5	202.9 \pm 58.1	183.3 \pm 67.6	197.7 \pm 65.6
Chol	244.5 \pm 36.4	238.4 \pm 42.6	253.4 \pm 47.3	246.7 \pm 31.2	251.4 \pm 45.1	239.7 \pm 39.8 *	248.4 \pm 36.1	255 \pm 41.7
LDL-Chol	129.1 \pm 27.4	123.4 \pm 34.6	133.1 \pm 35.2	133.7 \pm 26.6	134.5 \pm 42.5	135.5 \pm 30.9	142.3 \pm 35	140.5 \pm 31.7
HDL-Chol	50.6 \pm 8.6	49.3 \pm 8.4	51.7 \pm 8.9	51.6 \pm 8.9	48.2 \pm 10.8	49.2 \pm 9.2	52.7 \pm 9.6	52.6 \pm 9.5
Apo-A1	145.4 \pm 18.8	143.9 \pm 20.4	152.1 \pm 15.3	152.4 \pm 18.3	147.6 \pm 21.3	147.9 \pm 17.2	148 \pm 23.8	152.4 \pm 22.3
Apo-A2	33.9 \pm 3.9	33.3 \pm 4.8	36.3 \pm 4.1	34.8 \pm 4	34.9 \pm 5.9	34 \pm 5.8 *	34.8 \pm 4.8	33 \pm 4.9
Apo-B100	114.8 \pm 17.8	111.6 \pm 21.2	113.5 \pm 25.4	115.3 \pm 17.4	121.9 \pm 23.5	113.2 \pm 17.5	118.2 \pm 17.3	121.7 \pm 16.7
LDL-Chol/HDL-Chol	2.6 \pm 0.5	2.5 \pm 0.6 *	2.5 \pm 0.7	2.5 \pm 0.6	2.7 \pm 0.8	2.6 \pm 0.7	2.8 \pm 0.7	2.8 \pm 0.6
Apo-B100/Apo-A1	0.7 \pm 0.1	0.8 \pm 0.1	0.8 \pm 0.2	0.8 \pm 0.2	0.8 \pm 0.2	0.8 \pm 0.2	0.8 \pm 0.2	0.8 \pm 0.1
Total ApoB Particle Number	2086.5 \pm 323.9	2029.9 \pm 384.9	2063.2 \pm 462.5	2096.8 \pm 316.3	2216.5 \pm 428	2058 \pm 318.9	2149.9 \pm 315.3	2212.5 \pm 303.4
VLDL Particle Number	211.8 \pm 71.6	218.5 \pm 83.3	231.1 \pm 54.1	257.1 \pm 77.8	247.7 \pm 59.4	244.3 \pm 82.2	227 \pm 71.8	238.5 \pm 73.7
IDL Particle Number	111.6 \pm 39	108.1 \pm 36.9	129.9 \pm 36.6	121.6 \pm 39.2	134.8 \pm 49	120.8 \pm 49.1 **	124.7 \pm 28.9	120.5 \pm 39
LDL Particle Number	1646.2 \pm 333	1608.1 \pm 378	1659.3 \pm 409	1703.9 \pm 306.6	1800.2 \pm 408.3	1674.6 \pm 336.4	1752.4 \pm 316.2	1814.5 \pm 293.7
LDL 1 Particle Number	206.2 \pm 58.4	201.3 \pm 56.7	235.1 \pm 67	230.7 \pm 52.7 *	223.8 \pm 85.4	200.6 \pm 70.7	240.4 \pm 70.3	221.4 \pm 63.3
LDL 2 Particle Number	119.9 \pm 46.4	115.2 \pm 58.4	133.4 \pm 59.1	136.6 \pm 58.5	129.3 \pm 59.3	124.8 \pm 62.3	142.2 \pm 62.2	135.7 \pm 79.1
LDL 3 Particle Number	158 \pm 77	143.6 \pm 66.7 *	171.7 \pm 71.9	144.5 \pm 77.3	165 \pm 116.6	148.6 \pm 92.2	160.1 \pm 98.1	162.6 \pm 74.1
LDL 4 Particle Number	211 \pm 100.2	195.4 \pm 75.2 *	247.6 \pm 98.6	210.3 \pm 119.4	251 \pm 157.8	220.9 \pm 118.2	246.2 \pm 103.9	244.1 \pm 95

LDL 5 Particle Number	343.9 ± 102.8	354.2 ± 87	363.4 ± 100.8	357.2 ± 114.9	380.7 ± 105	366.7 ± 128.7	370.9 ± 110.7	392 ± 101.8
LDL 6 Particle Number	562.1 ± 203.8	578 ± 179.5	589.8 ± 205.9	578.8 ± 166.9	631.4 ± 178.3	591.4 ± 168.9	589.2 ± 244.3	588.1 ± 171.6
TG VLDL	125.4 ± 48.9	136.9 ± 54.9	140.8 ± 53.5	134.2 ± 49.8	141.7 ± 51.8	141.7 ± 54.9	119.9 ± 47.2	134.4 ± 49.7
TG IDL	21.6 ± 11.7	23.7 ± 12.4	24.8 ± 11.5	21.7 ± 10.2	24.1 ± 10	25.5 ± 11.8	20.5 ± 10.5	23.5 ± 11.1
TG LDL	22.5 ± 6	22.5 ± 6	24.5 ± 6.6	25.2 ± 8.3	24.4 ± 7.3	23 ± 5.1 *	23.9 ± 4.8	24.4 ± 5.6
TG HDL	10.3 ± 3.3	10.5 ± 3.9	10.5 ± 3.6	10.9 ± 3.1	10.8 ± 3.1	10.5 ± 3.8	10.6 ± 2.8	11.2 ± 3.3
Chol VLDL	31.3 ± 11.7	32 ± 14.3	34.6 ± 12.1	34.1 ± 14.1	34.9 ± 11.8	34.4 ± 11.2	31.9 ± 13.1	35.1 ± 12.2
Chol IDL	17.2 ± 6.5	16.1 ± 6.6	19.2 ± 6.6	17.9 ± 6.4	19.6 ± 7.4	16 ± 6.7 **	18.1 ± 4.9	18.2 ± 5.2
Chol LDL	129.1 ± 27.4	123.4 ± 34.6	133.1 ± 35.2	133.7 ± 26.6	134.5 ± 42.5	135.5 ± 30.9	142.3 ± 35	140.5 ± 31.7
Chol HDL	50.6 ± 8.6	49.3 ± 8.4	51.7 ± 8.9	51.6 ± 8.9	48.2 ± 10.8	49.2 ± 9.2	52.7 ± 9.6	52.6 ± 9.5
Free Chol VLDL	14.1 ± 5	14.8 ± 5.7	15.4 ± 4.3	16.1 ± 5.7	15.8 ± 4.1	15.5 ± 5.4	14 ± 4.7	15.6 ± 5
Free Chol IDL	4.8 ± 1.8	4.6 ± 1.9	5.5 ± 1.8	5.2 ± 1.9	5.8 ± 1.9	4.7 ± 1.9 **	5.3 ± 1.3	5.3 ± 1.5
Free Chol LDL	37.2 ± 8.2	36.2 ± 9.4	37.9 ± 10.4	38.2 ± 8.9	38.7 ± 11.9	38.1 ± 10.1	40.3 ± 9.3	38.7 ± 7.9
Free Chol HDL	11.6 ± 3.2	11.3 ± 3.3	11.8 ± 3.5	11.9 ± 2.5	10.9 ± 2.8	12.1 ± 2.1	11.8 ± 3	12.6 ± 3.1 *
Phospholipids VLDL	31.8 ± 10.1	32.9 ± 12.7	34.2 ± 10.7	36.3 ± 12.3	35.1 ± 9.2	35.4 ± 11	31.9 ± 11.2	34.8 ± 11
Phospholipids IDL	11.7 ± 4	11.8 ± 4.3	12 ± 3.4	11.8 ± 3.5	13 ± 3.7	12.4 ± 3.2 *	11.5 ± 3.3	12 ± 3.6
Phospholipids LDL	71.7 ± 15.5	68.1 ± 17.7	72.3 ± 17.7	74.2 ± 13.8 *	75.3 ± 21	72.8 ± 16.5	77.6 ± 15.6	77.3 ± 14.9
Phospholipids HDL	67.4 ± 14.3	66.5 ± 12.6	73 ± 11	68.2 ± 10.1	69 ± 13.7	65.4 ± 12.7	71.6 ± 13	72 ± 15.6
Apo A1 HDL	140.1 ± 20.9	138.8 ± 22.4	148.8 ± 18.6	145.3 ± 18.5	143.3 ± 22.3	140.9 ± 19.2	145.8 ± 22.4	148.5 ± 20.8
Apo A2 HDL	34.9 ± 3.9	34.5 ± 4.4	37.5 ± 3.9	35.8 ± 3.8	36.1 ± 6.4	35.1 ± 5.5 *	35.8 ± 4.6	34.3 ± 5.8
Apo B VLDL	11.6 ± 3.9	12 ± 4.6	12.7 ± 3	14.1 ± 4.3	13.6 ± 3.3	13.4 ± 4.5	12.5 ± 3.9	13.1 ± 4.1
Apo B IDL	6.1 ± 2.2	6 ± 2	7.1 ± 2	6.7 ± 2.2	7.4 ± 2.7	6.6 ± 2.7 **	6.9 ± 1.6	6.6 ± 2.1
Apo B LDL	90.5 ± 18.3	88.4 ± 20.8	91.3 ± 22.5	93.7 ± 16.9	99 ± 22.5	92.1 ± 18.5	96.4 ± 17.4	99.8 ± 16.2
TG VLDL 1	68.8 ± 32.2	67.9 ± 36.6	68.8 ± 31.2	63.3 ± 29.2	68.2 ± 26.1	70.4 ± 31.1	60 ± 28.5	64.8 ± 28.5
TG VLDL 2	22.5 ± 10.5	23.7 ± 11.4	24 ± 9.8	25.5 ± 9.5	26.2 ± 9.6	25.2 ± 11	22.5 ± 12.1	25 ± 10.7

TG VLDL 3	18.1 ± 8.1	18.6 ± 9.7	19.4 ± 7.8	21.7 ± 10.3	20.2 ± 7.6	18.9 ± 9.2	18.1 ± 8.5	18.4 ± 8.5
TG VLDL 4	11.8 ± 4.6	11.8 ± 4.4	13.2 ± 2.4	15 ± 5.9	13.5 ± 4.6	12.2 ± 4.8 *	12.8 ± 2.9	12.9 ± 3.9
TG VLDL 5	3.1 ± 0.9	3.2 ± 0.7	3.6 ± 0.9	3.5 ± 1	3.3 ± 1.1	3.1 ± 1.2	3.5 ± 1	3.7 ± 1.1
Chol VLDL 1	11 ± 5.6	12.5 ± 5.8	12.4 ± 5.1	11.4 ± 4	13.4 ± 5.3	12.9 ± 5.2	10.4 ± 5.6	12.4 ± 5.1
Chol VLDL 2	5.2 ± 2.7	5.5 ± 3.1	5.9 ± 2.9	6.2 ± 3.4	6.1 ± 2.9	5.8 ± 3	5.4 ± 3.4	5.9 ± 2.6
Chol VLDL 3	5.4 ± 2.6	5.4 ± 3.1	6.1 ± 2.8	6.3 ± 3.6	6.4 ± 2.6	6 ± 2.9	5.4 ± 2.9	5.8 ± 2.7
Chol VLDL 4	6.1 ± 2.7	6.3 ± 3.1	8 ± 2.8	8.1 ± 3	7.7 ± 3.1	6.8 ± 2.9 *	7.3 ± 2.4	7.5 ± 2.8
Chol VLDL 5	1.5 ± 0.8	1.6 ± 0.8	1.9 ± 0.6	1.9 ± 0.7	1.7 ± 0.8	1.7 ± 0.8	1.8 ± 0.7	1.9 ± 0.9
Free Chol VLDL 1	4.8 ± 2.6	4.8 ± 2.3	4.9 ± 2.1	4.4 ± 1.9	5.2 ± 2	5.1 ± 2.1	4.2 ± 2.1	4.7 ± 2.2
Free Chol VLDL 2	2.3 ± 1	2.3 ± 1.4	2.4 ± 1.1	2.5 ± 1.3	2.7 ± 1.3	2.3 ± 1.3	2.2 ± 1.3	2.4 ± 1.2
Free Chol VLDL 3	2.5 ± 1.2	2.7 ± 1.5	2.7 ± 1.1	2.9 ± 1.6	3 ± 1.3	2.8 ± 1.3	2.4 ± 1.3	2.6 ± 1.2
Free Chol VLDL 4	2.5 ± 1.2	2.5 ± 1.3	3.1 ± 0.9	3.2 ± 1.1	3.1 ± 1.4	2.9 ± 1.5 *	3 ± 1	3.1 ± 1.1
Free Chol VLDL 5	0.7 ± 0.4	0.7 ± 0.5	1 ± 0.4	0.9 ± 0.6	0.8 ± 0.5	0.7 ± 0.4	0.9 ± 0.5	0.9 ± 0.4
Phospholipids VLDL 1	11.3 ± 5.7	11.5 ± 5.4	11.7 ± 4.9	10.6 ± 4.1	11.9 ± 4.3	11.3 ± 4.8	10 ± 4.9	10.8 ± 5
Phospholipids VLDL 2	5.8 ± 2.6	5.9 ± 2.7	6.1 ± 2.5	6.2 ± 2.7	6.6 ± 2.5	6.3 ± 2.7	5.8 ± 3.1	6.1 ± 2.6
Phospholipids VLDL 3	5.6 ± 2.2	5.7 ± 2.8	6.3 ± 2.2	6.9 ± 3.3	6.6 ± 2.3	6.2 ± 2.4 *	5.8 ± 2.8	6 ± 2.3
Phospholipids VLDL 4	5.6 ± 2.1	5.8 ± 2.3	6.8 ± 1.6	6.9 ± 2.4	6.9 ± 2.4	6.3 ± 2.4 *	6.5 ± 1.6	6.3 ± 2
Phospholipids VLDL 5	1.9 ± 0.7	2 ± 0.7	2.3 ± 0.6	2.3 ± 0.7	2.1 ± 0.9	2.1 ± 0.8	2.3 ± 0.7	2.4 ± 0.9
TG LDL 1	6.3 ± 2.3	6.5 ± 2.2	6.8 ± 2.2	6.8 ± 2.3	7 ± 2.4	6.6 ± 2.3	6.5 ± 2.1	6.9 ± 2.1
TG LDL 2	1.9 ± 0.8	1.8 ± 0.8	2.2 ± 0.8	2.2 ± 0.8	2.1 ± 0.9	2.2 ± 0.9	2.2 ± 0.8	2.3 ± 0.9
TG LDL 3	2.4 ± 0.9	2.3 ± 0.9	2.8 ± 0.8	2.7 ± 0.9 *	2.8 ± 1.1	2.6 ± 0.9	2.8 ± 1	2.9 ± 1
TG LDL 4	2.4 ± 1.3	2.4 ± 1.2	2.8 ± 1.1	2.7 ± 1.2	3.1 ± 1.7	2.7 ± 1.2 **	3 ± 1.1	3.1 ± 1.1
TG LDL 5	3.9 ± 1.3	4 ± 1.5	4.2 ± 1.7	4.7 ± 1.3	4.8 ± 1.5	4.3 ± 1.6 *	4.4 ± 1.8	4.5 ± 1.6
TG LDL 6	5.6 ± 1.9	5.9 ± 1.6	5.8 ± 1.7	5.7 ± 2	6.5 ± 2.2	6.2 ± 2.3	5.6 ± 2.3	6 ± 1.8
Chol LDL 1	21.2 ± 6.3	20.4 ± 5.9	24.7 ± 7.3	23.2 ± 6.7 *	22.3 ± 9.7	21.2 ± 8.1	24.2 ± 6.5	23.3 ± 7.9

Chol LDL 2	10.5 ± 4.8	10.7 ± 6.6	12.8 ± 6.7	12.1 ± 6.2	10.7 ± 4.9	11.1 ± 7.7	13.3 ± 6.2	12.1 ± 7.6
Chol LDL 3	14.1 ± 8.4	12.7 ± 6.6	14.5 ± 7.5	12.6 ± 8	14.1 ± 12.1	13.2 ± 9.8	14.4 ± 10.1	14 ± 7.5
Chol LDL 4	18.8 ± 9.7	15.9 ± 7.1 *	21 ± 8	17.2 ± 11.2	20.3 ± 13.8	18.3 ± 10.7	20.5 ± 10	19.8 ± 8.9
Chol LDL 5	26.4 ± 6.7	27.1 ± 6.7	27.4 ± 7.9	27.4 ± 8.6	29.5 ± 8.4	28.1 ± 8.9	29 ± 8.6	31.3 ± 6.7
Chol LDL 6	38.6 ± 12.6	37.8 ± 12.2	39.2 ± 11.2	38.2 ± 9.9	42 ± 11.6	40 ± 11.5	40.1 ± 15.7	39.1 ± 12.8
Free Chol LDL 1	6.8 ± 1.9	6.8 ± 2	7.8 ± 2.3	7.4 ± 1.8	7 ± 2.7	6.9 ± 2.5	7.8 ± 1.7	7.5 ± 2.2
Free Chol LDL 2	3.7 ± 1.6	3.6 ± 2.1	4.2 ± 1.6	4.4 ± 1.9	3.8 ± 1.5	3.8 ± 1.7	4.3 ± 1.7	4.2 ± 2.1
Free Chol LDL 3	5 ± 2	4.7 ± 2	5.1 ± 1.7	4.6 ± 2.1	4.5 ± 3.2	4.6 ± 2.6	5.2 ± 2.6	4.8 ± 2
Free Chol LDL 4	5.6 ± 2.3	5 ± 1.9 *	6 ± 1.9	5.3 ± 2.7	6.1 ± 3.8	5.4 ± 2.7	5.8 ± 2.3	6 ± 2.2
Free Chol LDL 5	7.2 ± 1.8	6.8 ± 1.6	7.2 ± 2	7.3 ± 2.2	7.3 ± 2.4	7.3 ± 2	7.6 ± 2.1	7.9 ± 1.7
Free Chol LDL 6	9.5 ± 2.7	9.4 ± 3.1	9.2 ± 2.5	9.1 ± 2.6	9.7 ± 2.4	9.6 ± 2.8	9.2 ± 3.4	9.4 ± 2.2
Phospholipids LDL 1	11.9 ± 3.2	11.6 ± 3.1	13.8 ± 4.1	13.1 ± 3.2 *	12.6 ± 4.6	11.8 ± 4.4	13.5 ± 3.6	13.1 ± 4.1
Phospholipids LDL 2	6.5 ± 2.8	6.1 ± 3.3	7.1 ± 3.1	6.9 ± 3	6.6 ± 2.7	6.3 ± 3.6	7.8 ± 3.2	7.1 ± 4.3
Phospholipids LDL 3	8.1 ± 4.3	7 ± 3.2	8.4 ± 3.7	7.2 ± 4.3	8.3 ± 6.2	7.3 ± 4.9	8.1 ± 5.1	7.8 ± 3.9
Phospholipids LDL 4	10.3 ± 4.9	8.9 ± 3.4 *	11.6 ± 4.1	9.7 ± 5.7 *	11.5 ± 7	10 ± 5.9	11.4 ± 5.4	11 ± 4.4
Phospholipids LDL 5	13.8 ± 3.5	14.5 ± 3.8	14.5 ± 4.2	14.7 ± 4.3	15.5 ± 4.5	14.8 ± 4.4	15 ± 4.1	16.4 ± 3.6
Phospholipids LDL 6	20.1 ± 6	20.1 ± 5.5	20.5 ± 5.8	19.8 ± 5	21.7 ± 6.4	21 ± 5.4	20.9 ± 8	20.6 ± 5.7
Apo B LDL 1	11.3 ± 3.2	11.1 ± 3.1	12.9 ± 3.7	12.7 ± 2.9 *	12.3 ± 4.7	11 ± 3.9	13.2 ± 3.9	12.2 ± 3.5
Apo B LDL 2	6.6 ± 2.6	6.3 ± 3.2	7.3 ± 3.3	7.5 ± 3.2	7.1 ± 3.3	6.9 ± 3.4	7.8 ± 3.4	7.5 ± 4.3
Apo B LDL 3	8.7 ± 4.2	7.9 ± 3.7 *	9.5 ± 4	8 ± 4.3	9.1 ± 6.4	8.2 ± 5.1	8.8 ± 5.4	9 ± 4.1
Apo B LDL 4	11.6 ± 5.5	10.8 ± 4.1 *	13.6 ± 5.4	11.6 ± 6.6	13.8 ± 8.7	12.1 ± 6.5	13.5 ± 5.7	13.4 ± 5.2
Apo B LDL 5	18.9 ± 5.7	19.5 ± 4.8	20 ± 5.5	19.7 ± 6.3	20.9 ± 5.8	20.2 ± 7.1	20.4 ± 6.1	21.6 ± 5.6
Apo B LDL 6	30.9 ± 11.2	31.8 ± 9.9	32.4 ± 11.3	31.8 ± 9.2	34.7 ± 9.8	32.5 ± 9.3	32.4 ± 13.4	32.3 ± 9.4
TG HDL 1	2.3 ± 1.2	2.4 ± 1.4	2.6 ± 1.4	2.7 ± 1.6	2.5 ± 1.5	2.4 ± 1.7	2.4 ± 1.3	2.7 ± 1.7 *
TG HDL 2	1.5 ± 0.6	1.5 ± 0.7	1.6 ± 0.7	1.6 ± 0.6	1.5 ± 0.6	1.5 ± 0.7	1.5 ± 0.5	1.7 ± 0.7 *

TG HDL 3	2.3 ± 0.7	2.2 ± 0.8	2.4 ± 0.8	2.5 ± 0.6	2.5 ± 0.6	2.3 ± 0.7	2.5 ± 0.7	2.6 ± 0.7
TG HDL 4	4.4 ± 1	4.4 ± 1	4.7 ± 0.9	4.7 ± 1	4.7 ± 0.8	4.6 ± 1	4.6 ± 0.8	4.6 ± 1
Chol HDL 1	12.5 ± 4.6	13.1 ± 4.6	12.4 ± 4.7	13.6 ± 5.1	11.8 ± 4.2	12.4 ± 4.1	12.5 ± 4	14.5 ± 5.3 *
Chol HDL 2	5.7 ± 2	5.8 ± 2.3	6.4 ± 1.8	6.3 ± 2.1	5.8 ± 2.4	6 ± 1.6	6.7 ± 1.9	6.6 ± 1.9
Chol HDL 3	9.5 ± 2.1	9 ± 1.8	10.6 ± 1.9	9.9 ± 2.2	9.5 ± 2.2	9.3 ± 2	10.1 ± 2.4	10 ± 2.2
Chol HDL 4	23 ± 3.8	22.6 ± 2.8	24 ± 4.1	22.7 ± 4.9	22.5 ± 5	21.9 ± 5.1	23.1 ± 2.5	21.8 ± 3.3 *
Free Chol HDL 1	3.2 ± 1.1	3 ± 1.1	3.1 ± 1.2	3.1 ± 0.9	3.1 ± 1.1	3.4 ± 0.8	3.4 ± 1.1	3.6 ± 1 *
Free Chol HDL 2	1.4 ± 0.6	1.4 ± 0.6	1.8 ± 0.6	1.5 ± 0.6	1.6 ± 0.8	1.5 ± 0.5	1.6 ± 0.7	1.7 ± 0.6
Free Chol HDL 3	2.4 ± 0.5	2.3 ± 0.7	2.7 ± 0.5	2.5 ± 0.6	2.5 ± 0.9	2.3 ± 0.6	2.4 ± 0.8	2.5 ± 0.7
Free Chol HDL 4	4.6 ± 0.8	4.7 ± 0.9	5 ± 1	4.6 ± 1.2	4.5 ± 1.2	4.5 ± 1.2	4.7 ± 1	4.5 ± 1.1
Phospholipids HDL 1	12.4 ± 4.8	12.6 ± 5.4	13.7 ± 5.1	13.7 ± 5.2	11.8 ± 4.2	13.1 ± 5.1	13.1 ± 4.5	14.7 ± 5.5 **
Phospholipids HDL 2	9 ± 3.5	9.2 ± 3.8	10.4 ± 3.1	10 ± 3.8	9.7 ± 3.7	9.3 ± 3.1	10.2 ± 3.3	10.5 ± 3.7
Phospholipids HDL 3	15.4 ± 3.4	15.1 ± 2.7	17 ± 3.4	15.6 ± 3	15.3 ± 4.3	14.9 ± 3.8	16.3 ± 4	16 ± 4.2
Phospholipids HDL 4	30.6 ± 3.8	30.7 ± 3.7	32.5 ± 3.9	31 ± 4.3	30.9 ± 4.9	30.4 ± 5.9	32 ± 4.5	30.7 ± 3.9
Apo A1 HDL 1	13.2 ± 9	12.5 ± 6.6	13.3 ± 6	14.2 ± 8.3	12.5 ± 6.4	14.1 ± 7.1 *	14.2 ± 7.1	16.6 ± 8.1 *
Apo A1 HDL 2	15.8 ± 3.9	15.7 ± 4	17 ± 3.7	15.6 ± 3.9	15.6 ± 4.5	14.9 ± 3.4 *	16.9 ± 4.6	17 ± 4.9
Apo A1 HDL 3	24.6 ± 5.2	24.5 ± 5.1	27.5 ± 5.3	26.5 ± 5	25.4 ± 6.8	25.2 ± 5.5 *	26.4 ± 6.3	26.5 ± 5.2
Apo A1 HDL 4	86.5 ± 10.2	85 ± 9.7	89.1 ± 9.9	86.8 ± 11	85.4 ± 13.3	84.3 ± 11.8	87.2 ± 8.4	84.7 ± 10.8
Apo A2 HDL 1	1.4 ± 1	1.3 ± 1	1.9 ± 0.8	1.8 ± 0.9	1.7 ± 0.8	1.5 ± 0.7	1.7 ± 0.7	1.8 ± 1
Apo A2 HDL 2	2.8 ± 1.1	2.7 ± 1.1	3.5 ± 1	3.2 ± 1.2	3.5 ± 1.3	2.8 ± 1.2	3.2 ± 0.7	3.2 ± 1.3
Apo A2 HDL 3	6.5 ± 1.4	6.3 ± 1.5	7.3 ± 1.4	6.8 ± 1.3	6.9 ± 2	6.4 ± 1.7	6.9 ± 1.4	6.7 ± 1.9
Apo A2 HDL 4	23 ± 3.7	23 ± 3.4	24.9 ± 3	23.4 ± 4.4	23.2 ± 4.3	22.2 ± 4.5	23.6 ± 2.7	22.2 ± 3.1 *

§Abbreviations are reported as follows: TG: triglycerides, Chol: cholesterol, Apo: Apolipoproteins

Figures

Figure 1. Graphical representation of the procedure followed to explore differences in serum profiles of subjects at risk for MetS, before (t_0) and 12-weeks after (t_1) the dietary interventions (treatments *a*, *b*, *c* and *d*), using untargeted NMR-based metabolomics in the frame of the EC FP7 PATHWAY-27 project. Enrolled subjects are from 4 different recruitment centers: Italy, France, Germany and UK and they had to consume one of the following treatment: *a*) bakery, egg, dairy-based foods without bioactives (placebo); *b*) dairy-based BEF + placebo bakery and placebo egg-based foods; *c*) egg-based BEF + placebo bakery and placebo dairy-based foods; *d*) bakery-based BEF + placebo dairy and egg-based foods.

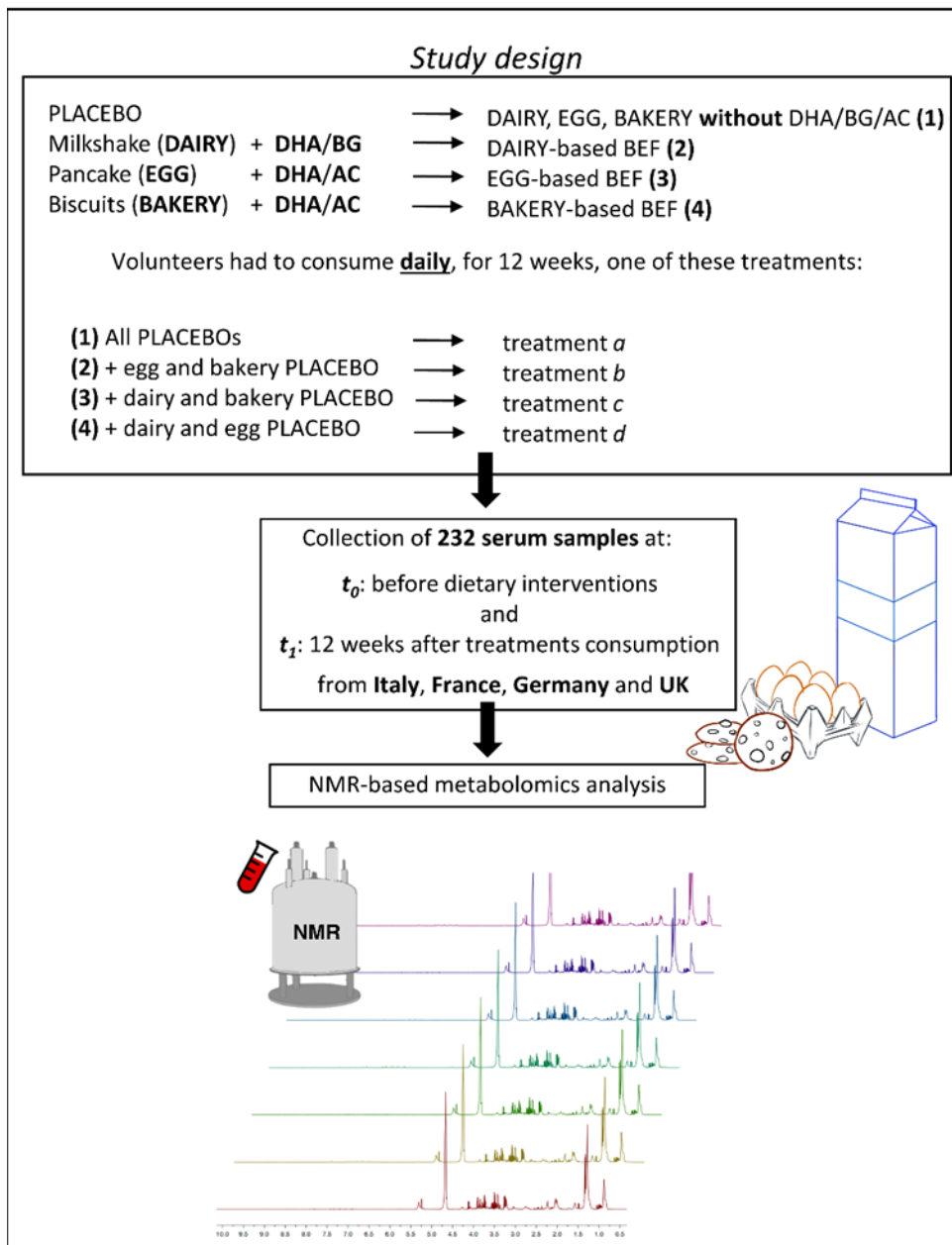


Figure 2. PCA 3D score plot of the whole dataset. Each dot represents a 0.02 ppm bucketed 1D-NOESY $^1\text{H-NMR}$ spectrum color-coded by **A**) time-points of blood collections: t_0 (red dots, $n=232$) and t_1 (blue dots, $n=232$); **B**) dietary intervention: a) all bakery, egg, dairy-based foods without bioactives (pink dots, $n=70$); b) dairy-based BEF + placebo bakery and placebo egg-based foods (yellow dots, $n=55$); c) egg-based BEF + placebo bakery and placebo dairy-based foods (green dots, $n=55$); d) bakery-based BEF + placebo dairy and egg-based foods (cyan dots, $n=54$); **C**) recruitment center: Italy (gray dots, $n=73$); France (yellow dots, $n=61$); Germany (green dots, $n=44$); UK (red dots, $n=54$).

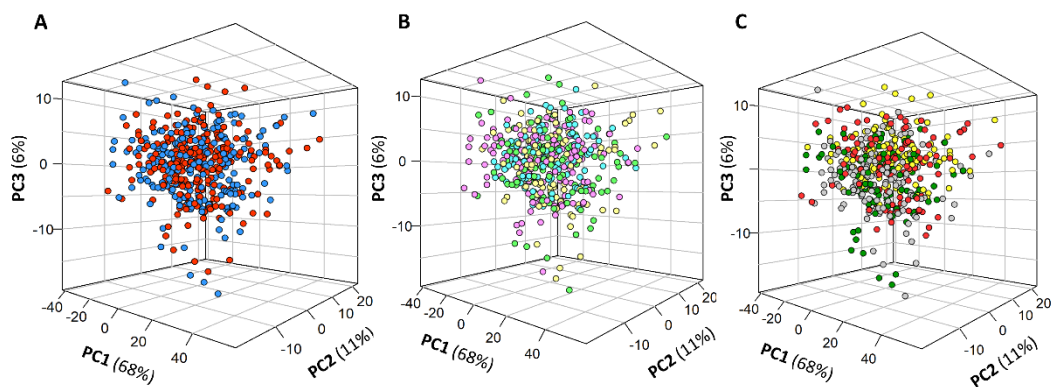


Figure 3. M-PLS discrimination accuracy between t_0 and t_1 for each recruited subject. Each dot represents a different subject, color-coded by the received treatment: black dots = treatment a (Placebo); red dots = treatment b (dairy-based BEF + placebo bakery and placebo egg-based foods); green dots = treatment c (egg-based BEF + placebo bakery and placebo dairy-based foods); blue dots = treatment d (bakery-based BEF + placebo dairy and egg-based foods).

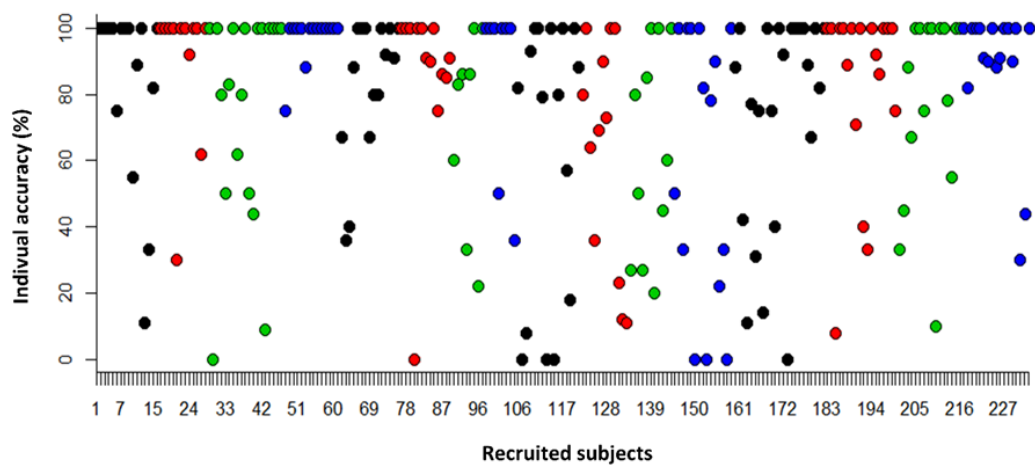


Figure 4. M-PLS loading plot of treatment d (Bakery-based BEF + other foods as placebo) related to the first component, considering **A**) only French cohort and **B**) only

English cohort; the threshold (green lines) to select effective signals was estimated considering bins with values beyond two standard deviations of their averages; **1**: 1.29 ppm ($(-\text{CH}_2-)_n$ VLDL-LDL); **2**: 0.86 ppm (CH_3 VLDL-LDL).

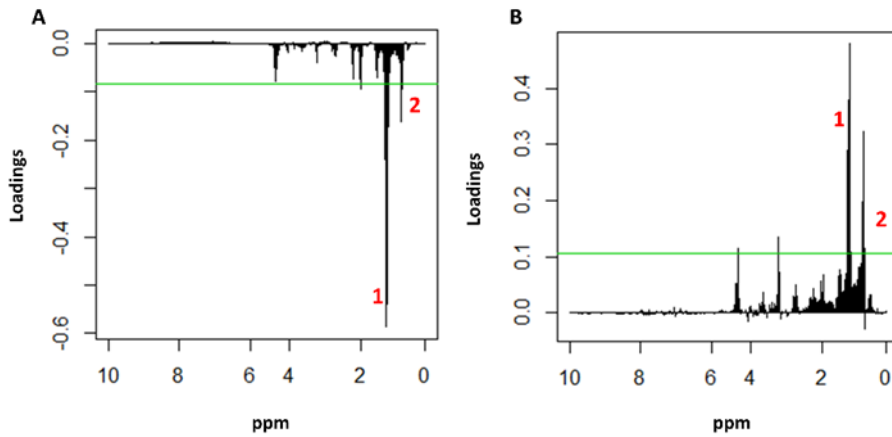
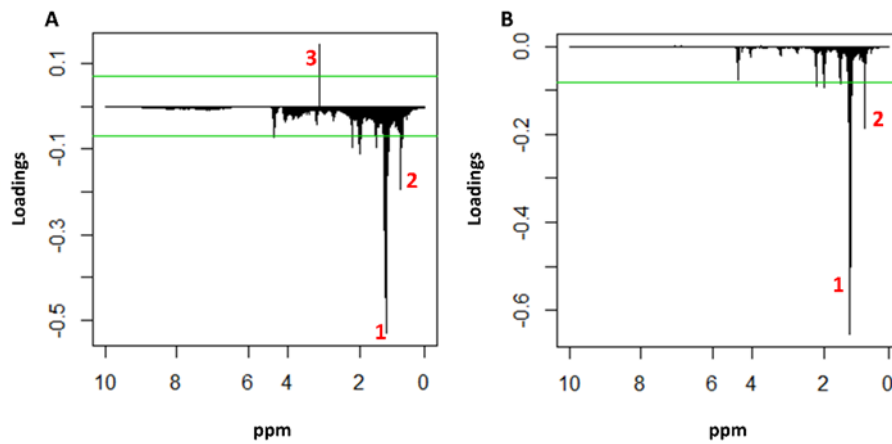


Figure 5. M-PLS loading plot of treatment *c* (egg-based BEF + other foods as placebo) related to the first component, considering **A**) only Italian cohort and **B**) only German cohort; the threshold (green lines) to select effective signals was estimated considering bins with values beyond two standard deviations of their averages; **1**: 1.29 ppm ($(-\text{CH}_2-)_n$ VLDL-LDL); **2**: 0.86 ppm (CH_3 VLDL-LDL); **3**: 3.20 ppm.



Supplementary Material

Table S1. Confusion matrix of PCA-CA model discriminating treatment-groups at t_0 : *a*) all bakery, egg, dairy-based foods without bioactives (placebo); *b*) dairy-based BEF + placebo bakery and placebo egg-based foods; *c*) egg-based BEF + placebo bakery and placebo dairy-based foods; *d*) bakery-based BEF + placebo dairy and egg-based foods. Values and accuracy % are reported.

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>
<i>a</i>	30.4	25.3	21.6	22.7
<i>b</i>	35.3	24.0	22.7	18.0
<i>c</i>	29.0	19.5	17.2	34.3
<i>d</i>	27.2	21.8	25.7	25.3

*Overall discrimination accuracy: **24.6%**

Table S2. Confusion matrix of PCA-CA model discriminating treatment-groups at t_1 : *a*) all bakery, egg, dairy-based foods without bioactives (placebo); *b*) dairy-based BEF + placebo bakery and placebo egg-based foods; *c*) egg-based BEF + placebo bakery and placebo dairy-based foods; *d*) bakery-based BEF + placebo dairy and egg-based foods. Values and accuracy % are reported.

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>
<i>a</i>	48.2	21.6	15.5	14.7
<i>b</i>	28.6	20.2	28.0	23.2
<i>c</i>	22.5	23.3	27.6	26.6
<i>d</i>	18.7	18.2	24.8	38.3

*Overall discrimination accuracy: **34.5%**

Table S3. Bruker IVDr metabolites analysis at t_0 and t_1 for enrolled subjects ($n=232$). Metabolic features were assigned and quantified in 1D NOESY NMR spectra and their absolute concentrations are reported as median \pm median absolute deviation. * is used to indicate parameters resulted to be statistically significant, with P -value < 0.05 , in the comparison t_0 vs. t_1 for each of the administered treatment. *a*) all bakery, egg, dairy-based foods without bioactives (placebo); *b*) dairy-based BEF + placebo bakery and placebo egg-based foods; *c*) egg-based BEF + placebo bakery and placebo dairy-based foods; *d*) bakery-based BEF + placebo dairy and egg-based foods). ** is used for parameters reporting both P -value < 0.05 and FDR < 0.05 .

	<i>a</i>		<i>b</i>		<i>c</i>		<i>d</i>	
	<i>t</i> ₀	<i>t</i> ₁	<i>t</i> ₀	<i>t</i> ₁	<i>t</i> ₀	<i>t</i> ₁	<i>t</i> ₀	<i>t</i> ₁
Trimethylamin e-N-oxide	0.01 ± 0.01	0.01 ± 0.01	0.005 ± 0.01	0.005 ± 0.01	0.01 ± 0.01	0.01 ± 0.01	0.01 ± 0.01	0.01 ± 0.01 *
Alanine	0.48 ± 0.1	0.46 ± 0.11	0.48 ± 0.08	0.47 ± 0.08	0.5 ± 0.09	0.49 ± 0.07	0.48 ± 0.11	0.5 ± 0.07
Creatine	0.02 ± 0.01	0.02 ± 0.01	0.02 ± 0.01	0.02 ± 0.01	0.01 ± 0.01	0.02 ± 0.01	0.02 ± 0.01	0.02 ± 0.01
Creatinine	0.09 ± 0.02	0.09 ± 0.02	0.09 ± 0.02	0.09 ± 0.02	0.1 ± 0	0.09 ± 0.03	0.09 ± 0.02	0.1 ± 0.02
Glutamine	0.79 ± 0.09	0.76 ± 0.13 *	0.76 ± 0.12	0.76 ± 0.14	0.81 ± 0.15	0.81 ± 0.13	0.81 ± 0.13	0.8 ± 0.11
Glycine	0.23 ± 0.04	0.22 ± 0.05 *	0.23 ± 0.04	0.23 ± 0.05	0.24 ± 0.05	0.23 ± 0.06	0.25 ± 0.05	0.24 ± 0.04
Histidine	0.09 ± 0.02	0.08 ± 0.02	0.08 ± 0.01	0.08 ± 0.02	0.09 ± 0.02	0.08 ± 0.01 *	0.08 ± 0.01	0.09 ± 0.02
Isoleucine	0.06 ± 0.01	0.06 ± 0.02	0.06 ± 0.01	0.06 ± 0.02	0.06 ± 0.02	0.06 ± 0.02	0.06 ± 0.01	0.06 ± 0.02
Leucine	0.1 ± 0.02	0.11 ± 0.02	0.1 ± 0.04	0.11 ± 0.03	0.12 ± 0.03	0.11 ± 0.02	0.11 ± 0.02	0.11 ± 0.02
Methionine	0.06 ± 0.03	0.05 ± 0.02	0.05 ± 0.04	0.05 ± 0.03	0.06 ± 0.03	0.06 ± 0.02	0.06 ± 0.03	0.06 ± 0.02
N,N- Dimethylglyci ne	0.005 ± 0.03	0.01 ± 0.003	0.01 ± 0.003	0.01 ± 0.001	0.01 ± 0.003	0.01 ± 0.01	0.01 ± 0.002	0.01 ± 0.003 **
Phenylalanine	0.05 ± 0.02	0.05 ± 0.01	0.05 ± 0.02	0.05 ± 0.01	0.06 ± 0.01	0.07 ± 0.01	0.06 ± 0.01	0.06 ± 0.01
Tyrosine	0.07 ± 0.01	0.07 ± 0.01	0.07 ± 0.01	0.07 ± 0.01	0.07 ± 0.01	0.07 ± 0.01	0.07 ± 0.01	0.07 ± 0.02
Valine	0.26 ± 0.04	0.26 ± 0.06	0.26 ± 0.05	0.26 ± 0.06	0.29 ± 0.04	0.28 ± 0.05	0.28 ± 0.05	0.26 ± 0.04
Acetic acid	0.01 ± 0.01	0.01 ± 0.01	0.01 ± 0.01	0.01 ± 0.01	0.01 ± 0.01	0.01 ± 0.01	0.01 ± 0.01	0.01 ± 0.01
Citric acid	0.15 ± 0.05	0.15 ± 0.04	0.15 ± 0.04	0.14 ± 0.04	0.16 ± 0.04	0.15 ± 0.03	0.16 ± 0.04	0.16 ± 0.03
Formic acid	0.02 ± 0.01	0.02 ± 0.01	0.02 ± 0.01	0.02 ± 0.01	0.02 ± 0.01	0.02 ± 0.01	0.02 ± 0.01	0.02 ± 0.01
Lactic acid	2.01 ± 0.6	2.13 ± 0.66	2.11 ± 0.75	2.06 ± 0.78	2.2 ± 0.5	2.06 ± 0.49	2.12 ± 0.54	2.19 ± 0.75
Succinic acid	0.003 ± 0.001	0.003 ± 0.001	0.002 ± 0.001	0.003 ± 0.001	0.002 ± 0.001	0.002 ± 0.001	0.002 ± 0.001	0.002 ± 0.001

Acetoacetic acid	0.01 ± 0.01	0.01 ± 0.01	0.01 ± 0.004	0.01 ± 0.003	0.01 ± 0.01	0.01 ± 0.004	0.01 ± 0.003	0.01 ± 0.004
Acetone	0.02 ± 0.01	0.02 ± 0.01	0.02 ± 0.01	0.01 ± 0.01	0.01 ± 0.01	0.02 ± 0.01	0.02 ± 0.01	0.01 ± 0.01
Pyruvic acid	0.08 ± 0.04	0.08 ± 0.03	0.1 ± 0.03	0.09 ± 0.04	0.09 ± 0.03	0.09 ± 0.03	0.08 ± 0.03	0.07 ± 0.03
Glucose	5.48 ± 0.76	5.33 ± 0.63	5.61 ± 0.75	5.39 ± 0.67	5.58 ± 0.68	5.38 ± 0.62 *	5.58 ± 0.68	5.48 ± 0.58
Dimethyl sulfone	0.01 ± 0. 00 4	0.01 ± 0.03 *	0.01 ± 0.04	0.01 ± 0.003 *	0.01 ± 0.004	0.01 ± 0.004	0.01 ± 0.004	0.01 ± 0.003

[§]Abbreviations are reported as follows: TMAO: Trimethylamine N-oxide

4.2. NMR-based metabolomics for veterinary research

Although metabolomics mainly deals with human medicine, several potential applications for NMR-based metabolomics in the veterinary field are possible, especially for what concerns the investigation of disease or infection mechanisms, disease diagnosis and monitoring, the characterization of health status, the monitoring of pharmacological treatments and drug discovery, nutrition and food production.

In this PhD thesis, three different NMR-based metabolomic veterinary studies are described, demonstrating the potentiality and the usefulness of this technique in this framework.

In the first proposed study, NMR spectroscopy has been applied to metabolically characterize left and right abomasal displacements of dairy cows (§ 4.2.1). *Displaced abomasum* (DA) is a disease condition of dairy cows that severely impacts animal wellbeing and it causes huge economic losses for food industry. Using serum, urine and liver organic/aqueous extracts, collected from 50 Holstein multiparous cows with DA (42 left and 8 right) and 20 clinically healthy cows, we explored the disease from a metabolic point of view. OPLS-DA models, built on bucketed NMR spectra, revealed metabolic differences between healthy and diseased animals, especially when NMR data related to aqueous extracts have been used. Applying univariate analysis on selected metabolic features, fatty acid fractions and cholesterol were found to be increased in liver samples of cows affected by DA, serum hippuric acid level was significantly higher in healthy animals compared to cows affected by left DA (LDA), whereas serum glycine was reported to be higher in healthy when compared to cows affected by right DA (RDA). Globally, metabolomic profiles combined to clinical analysis revealed that cows with DA (especially LDA) are at higher risk for ketosis and fatty liver. Moreover, a biochemical pathway mapping revealed “valine, leucine and isoleucine biosynthesis” and “phenylalanine, tyrosine and tryptophan biosynthesis” as the most probable altered metabolic pathways in DA condition. In conclusion, our results can find actual applications for clinical practices and they evidenced serum as the optimal and most promising biological matrix for future clinical and $^1\text{H-NMR}$ investigations.

In the second study, NMR-based metabolomics was applied to investigate serum samples collected at set intervals (24h, 48h and 75h) from 25 *premature calves* (§ 4.2.2). Calves born previous to 270 days' pregnancy are at risk for death and increasing morbidities; therefore, the characterization of the premature birth from a metabolic point of view could pave the way to prevent diseases and also to improve food production. Here, within-animal differences were investigated applying M-PLS analysis to bucketed NMR spectra obtained from aqueous and organic serum extracts. For all statistical models built on the bucketed NMR profile of serum aqueous fractions, premature calves can be discriminated at the three different time-points with predictive accuracies higher than 70%. Lower prediction accuracies than 70% have

been obtained when M-PLS models were built on bucketed 1D ^1H NOESY spectra from serum organic fractions. From the univariate approach on quantified metabolites and lipid fractions, we described: *i*) increases in 3-hydroxybutyrate, citrate, leucine and isoleucine levels at 48h and 72h; *ii*) increases in choline, formic acid, fatty acids and polyunsaturated fatty acids levels at 72h and valine concentration at 48h; *iii*) decreases in myo-inositol level at 48h and 72h. Myo-inositol concentration proved to be meaningful for monitoring the recovery, at a molecular level, in premature calves. Overall, these results suggest that NMR-based metabolomics put the basis to deepen future researches in premature calves' clinical pathology and for the monitoring of their therapeutic frame.

In the last study presented in this section, NMR-based metabolomics was used to explore, for the first time, metabolic changes in the serum lipidome and metabolome of dogs affected by *Canine Ehrlichiosis* (§ 4.2.3). Generally, *Ehrlichiosis* is an infection caused by intracellular organisms that affect cells of the immune system in dogs, cats and also people. Here, we tried to identify the metabolic fingerprint of the disease and potentially useful metabolic markers, using NMR spectroscopy for the analysis of serum aqueous and organic extracts from 92 infected and 17 healthy dogs. Univariate results did not show significant changes in the metabolome: the detected metabolites seem to be ineffective in the characterization of the infection, but entire NMR profiles of healthy dogs are discriminated from those of diseased animals with a predictive accuracy around 70% in aqueous extracts, thus showing the existence of an underlying metabolic signature of the infection considering the whole ^1H -NMR profile. However, a more pronounced signature of the pathology was found in the lipidome (predictive accuracy of 76% between healthy and diseased animals) and applying univariate analysis, lipid fractions showed considerable differences between diseased and healthy dogs, thus reflecting a systemic condition of energy deficit during the infection.

To our knowledge, this is the first NMR study dealing with the characterization of *Canine Ehrlichiosis* and our findings may deepen the knowledge on the biological processes of the disease and they could find actual use in the veterinary clinical practice.

4.2.1. Nuclear magnetic resonance (NMR)-based metabolome profile evaluation in dairy cows with and without displaced abomasum

Abdullah Basoglu¹, Nuri Baspinar², Leonardo Tenori³, Cristina Licari⁴, Erdem Gulersoy¹

¹Department of Internal Medicine, Faculty of Veterinary Medicine, Selcuk University, Selcuklu, Konya, Turkey;

²Department of Biochemistry, Faculty of Veterinary Medicine, Selcuk University, Selcuklu, Konya, Turkey;

³Interuniversity Consortium for Magnetic Resonance of Metalloproteins (C.I.R.M.M.P.), Sesto Fiorentino (Florence), Italy;

⁴Magnetic Resonance Center (CERM), University of Florence, Sesto Fiorentino (Florence), Italy.

Published

Veterinary Quarterly, 2020, Vol. 40, No. 1, 1–15

Candidate's contributions: acquisition of NMR data, statistical analysis, interpretation of results, writing and review of the manuscript.

Nuclear magnetic resonance (NMR)-based metabolome profile evaluation in dairy cows with and without displaced abomasum

Abdullah Basoglu^a, Nuri Baspinar^b, Leonardo Tenori^c, Cristina Licari^d and Erdem Gulersoy^a

^aDepartment of Internal Medicine, Faculty of Veterinary Medicine, Selcuk University, Selcuklu, Konya, Turkey; ^bDepartment of Biochemistry, Faculty of Veterinary Medicine, Selcuk University, Selcuklu, Konya, Turkey; ^cInteruniversity Consortium for Magnetic Resonance of Metalloproteins (C.I.R.M.M.P.), Sesto Fiorentino (Florence), Italy; ^dMagnetic Resonance Center (CERM), University of Florence, Sesto Fiorentino (FI), Italy

ABSTRACT

Background: Displaced abomasum (DA) is a condition of dairy cows that severely impacts animal welfare and causes huge economic losses.

Objective: To assess the metabolic status of the disease using metabolomics in serum, urine and liver samples aimed at both water soluble and lipid soluble fractions.

Methods: Fifty Holstein multiparous cows with DA (42 left, 8 right) and 20 clinically healthy Holstein multiparous cows were used. Left DA was associated with concomitant ketosis in 19 animals and right in two. NMR-based metabolomics approach and hematological and biochemical analyses were performed. Statistical analysis was carried out on ¹H-NMR data after they have been normalized using PQN method.

Results: Contrary to generated PCA score plots the OPLS-supervised method revealed differences between healthy animals and diseased ones based on serum water-soluble samples. While water and lipid soluble metabolites decreased in serum samples, fatty acid fractions and cholesterol were increased in liver samples in DA affected cows. The metabolomic and chemical profiles clearly revealed that cows with DA (especially with LDA) were at risk of ketosis and fatty liver. Serum hippuric acid concentration was significantly higher in healthy cows in comparison with LDA, whereas serum glycine concentration was reported higher for healthy when compared to RDA affected animals.

Conclusion: A biochemical network and pathway mapping revealed 'valine, leucine and isoleucine biosynthesis' and 'phenylalanine, tyrosine and tryptophan biosynthesis' as the most probable altered metabolic pathway in DA condition. Serum was advocated as the optimal biological matrix for the ¹H-NMR analysis.

ARTICLE HISTORY

Received 8 July 2019
Accepted 18 December 2019

KEYWORDS



Cattle; cow; metabolomics; NMR; displaced abomasum

1. Introduction


Following parturition, up to 35–50% of high producing dairy cows are affected by metabolic and infectious diseases. Displaced abomasum is a multifactorial disorder usually diagnosed in early lactation dairy cows, and is a common inherited condition in Holstein cows (Zerbin et al. 2015; Doll 2015; Caixeta et al. 2018). The incidence of DA in the United States dairy herds was determined to be approximately 3.5% (NAHMS-USDA 2017). Economic analyses have determined that the average cost per DA diagnosis is more than \$700 when accounting for direct and indirect costs (McArt et al. 2015).

Understanding DA in cattle has been the objective of numerous *in vitro* and *in vivo* studies. However, a complete elucidation of its pathogenesis has still to be achieved (Sickingler 2017). Proteomics

and metabolomics have emerged as valuable techniques to characterize proteins and metabolite assets from tissue and biological fluids, such as milk, blood, and urine (Takis et al. 2018). There is a multitude of studies about the metabolic backgrounds of such so-called production diseases like ketosis, fatty liver, or hypocalcemia, although the investigations aiming to assess the complexity of the pathophysiological reactions are largely focused on gene expression, that is, transcriptomics. For extending the knowledge toward the proteome and the metabolome, the respective technologies are of increasing importance (Vignoli et al. 2019) and can provide an overall view of how dairy cows react to metabolic stress, which is needed for an in-depth understanding of the molecular mechanisms of the related diseases (Cecilian et al. 2018). Displaced abomasum occurs

CONTACT Abdullah Basoglu  abbasoglu@selcuk.edu.tr  Department of Internal Medicine, Faculty of Veterinary Medicine, Selcuk University, Selcuklu, Konya, Turkey.

This article has been produced from E. Gulersoy's doctorate thesis.

 Supplemental data for this article can be accessed at <https://doi.org/10.1080/01652176.2019.1707907>.

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Table 1. Characterizations of the Holstein cows (\pm SD).

Parameters	Cow groups		
	Healthy	LDA	RDA
Postpartum days	13 \pm 5	14 \pm 6	15 \pm 6
Age (yr)	3 \pm 1	3 \pm 1	3 \pm 1
Parity	2 \pm 1	2 \pm 1	2 \pm 1
Milk yield (kg/d)	24.82 \pm 7.55	24.07 \pm 7.38	23.00 \pm 1.68
Milk fat %	3.55 \pm 1	3.65 \pm 1	3.45 \pm 1
Milk protein %	3.06 \pm 1	3.0 \pm 1	3.05 \pm 1

simultaneously with fatty liver, but the relationship between the diseases are not clear (Ingvartsen 2006).

The aim of this study was to reveal new potential biomarkers representing the metabolic status of DA by using NMR-based metabolome profile evaluation and providing possible clues into the pathogenic mechanisms for DA.

2. Materials and methods

The experimental design was approved by the Committee on Use of Animals in Research of the Selcuk University, Faculty of Veterinary Medicine (Protocol No. 68/2017).

2.1. Animals

Fifty with displaced abomasum (42 LDA and 8 RDA), and 20 healthy Holstein multiparous cows within 1 month of parturition were used as animal material (Table 1). Dietary composition and nutrient level daily for the diseased animals were as follows: corn silage 12 kg, sugar beet pulp 10 kg, wheat straw 4.5 kg and concentrate 8.5 kg. The concentrate was consisted of 35% barley, 19.85% wheat, 15% wheat bran, 25% cotton seed meal, 3% limestone, 0.3% salt, and 0.35% vitamin–mineral mixture. It will be contained 21.5% crude protein and 2,850 kcal/kg metabolizable energy.

Displacement diagnosis was based on the history, the presence of the characteristic ping on simultaneous auscultation and percussion and exclusion of other causes of left- or right-sided pings. Ultrasonography was performed in confirming a diagnosis of LDA and RDA. Among cows with LDA, 19 were associated with concomitant ketosis. Two cows with RDA had concomitant ketosis too. Ketosis was detected by urine chemistry analyzer using urine test strips (Bayer Clinitek 50, Leverkusen, Germany) and blood ketone meter using blood ketone test strips Abbott Optium Xceed Pro, Oxford, UK). A Liptak test-needle was placed in the viscus to remove fluid, and pH was measured when needed (Guard 1990). All diagnoses were confirmed during a surgical operation. Control animals were also multiparous within 1 month of lactation and chosen via the same clinical and hematological methods.

2.2. Blood, urine and liver sampling

Blood samples were collected from the coccygeal vein into heparin and K-EDTA coated tubes. Blood samples were immediately used for hematological measurements. Serum samples harvested within an hour by centrifugation for 15 min at 3000 rpm were stored at -80°C until analysis. Urine samples were obtained via a sterile, rigid metal catheter (approximately 0.5 cm in diameter and 40 cm long). Liver biopsies were collected on the right side, proximal within the 11th or 12th intercostal space using a Tru-cut biopsy trocar (Merit Medical, Maastricht, The Netherlands).

Serum, urine, and liver samples were thawed on ice and extracted using a dual methanol-chloroform extraction (for protein precipitation and separation of hydrophilic and lipophilic fractions) as previously described (Serkova et al. 2005). This eliminates macromolecules (e.g., proteins) and establishes a fused metabolic profile for water-soluble and lipid-soluble metabolites. Briefly, 0.5 ml of ice-cold serum or urine (0.5 mg of the liver) was mixed with 0.5 ml of chloroform: methanol (1:1 vol/vol) (0.1 ml of chloroform:0.2 ml of methanol and 0.04 ml of distilled water for liver) and centrifuged. The supernatant (organic phase) was collected, and the pellet was resuspended with 0.5 ml of chloroform/methanol (0.2 ml of chloroform/methanol for liver) and centrifuged. The supernatants were combined, and 0.5 ml of ice-cold water (0.1 ml of ice-cold water for liver) was added to 'wash out' remaining water-soluble metabolites from the organic phase. After 15 min at -20°C , the upper (aqueous) phase was removed and added to the remaining pellet (to wash out remaining water-soluble metabolites from the pellet), 1 ml of water was added, and the sample was centrifuged and lyophilized by evaporation at 42°C .

2.3. Hematological analysis

Hematological analyses including complete blood count (erythrocytes and leukocytes counts, MCV, MCHC, PCV, and Hb) were performed by automated cell counter (MS4e, Melet Schloesing Laboratories, Osny, France) and blood gas analysis (pH, pCO_2 , pO_2 , HCO_3 and BE) by blood gas analyzer (Gem Premier 3000, Instrumentation Laboratory, Bedford, MA 01730, USA).

2.4. Chemical analysis

Serum was harvested within 1 h by centrifugation for 15 min at 3000 rpm, and stored at -80°C until analysis. It was analyzed for glucose, lactate, cholesterol, triglyceride, total protein, albumin, total bilirubin, BUN, creatinine, some minerals (Na, Mg, P, K, Ca and ionized Ca); some enzyme activities (AST, GGT, LDH,

and CPK), NEFA, and CRP by spectrometry (Autoanalyzer/BT3000 Plus, Rome, Italy). Blood BHBA levels were measured from whole blood by using Freestyle Optium H B-ketone, Abbott ® test strips (Allschwil, Switzerland).

2.5. Samples preparation for ¹H-NMR spectroscopy

Dried water-soluble samples were dissolved in 700 µL of ²H₂O (99.9 atom % D, Sigma Aldrich, St. Louis, Missouri, USA) and homogenized by vortexing for 1 min. Then, they were centrifuged (3000 rpm at 4 °C for 15 min) and 630 µL of each supernatant was added to 70 µL of potassium phosphate buffer (1.5 M K₂HPO₄, 100% (v/v) ²H₂O, 2 mM NaN₃, 5.8 mM deuterated trimethylsilyl propanoic acid (TMSP); pH 7.4). After stirring, a total of 600 µL from each mixture was transferred into 5 mm NMR tubes (Bruker BioSpin s.r.l.) for the analysis.

The dried lipid extracts were dissolved in 700 µL of CDCl₃ (99.8 atom % D, Sigma Aldrich) and homogenized by vortexing for 1 min. An aliquot of 600 µL from each sample was transferred into 5 mm NMR tubes (Bruker BioSpin s.r.l.) for the analysis.

2.6. NMR analysis

For each sample one-dimensional ¹H-NMR spectra were acquired using a Bruker 600 MHz spectrometer (Bruker BioSpin s.r.l.; Rheinstetten, Germany) operating at 600.13 MHz proton Larmor frequency and equipped with a 5 mm PATXI ¹H-¹³C-¹⁵N and ²H-decoupling probe including α -z axis gradient coil, an automatic tuning and matching (ATM), and an automatic and refrigerated sample changer (SampleJet, Bruker BioSpin s.r.l.; Darmstadt, Germany). For temperature stabilization, a BTO 2000 thermocouple was used at the level of approximately 0.1 K at the sample. Before starting the measurement, samples were kept for at least 5 minutes inside the NMR probe head to equilibrate temperature, at 300 K for urine water soluble samples and at 310 K for all the other extracts.

For each water and lipid-soluble sample, a one-dimensional ¹H NMR spectrum was acquired using a standard Nuclear Overhauser Effect Spectroscopy pulse sequence (NOESY 1Dpresat; noesygppr1d.com; Bruker BioSpin) using 98,304 data points, a spectral width of 18,028 Hz, an acquisition time of 2.7 s, a relaxation delay of 4 s, a mixing time of 0.01 s and a different number of scans according to the type of extract (128 scans for both serum water/lipid soluble and liver-lipid soluble samples; 256 scans for liver water-soluble extracts; 64 scans for both urine water and lipid soluble samples).

In addition, for serum water-soluble extracts, another ¹H-NMR spectrum was acquired using a standard spin echo Carr-Purcell-Meiboom-Gill pulse sequence (CPMG) (Meiboom and Gill 1958) (cpmgpr1d.com; Bruker BioSpin) with 128 scans, 73,728 data points, a spectral width of 12,019 Hz an acquisition time of 3.1 s, a relaxation delays of 4 s and a total spin-echo delay of 80 ms.

2.7. Spectral processing

Before applying Fourier transform, free induction decays were multiplied by an exponential function equivalent to 0.3 Hz line-broadening factor. Transformed spectra were automatically corrected for phase and baseline distortions and through TopSpin 3.2 (Bruker BioSpin software), they were calibrated using the anomeric glucose doublet at 5.24 ppm for serum water-soluble extracts, the TMSP singlet at 0.00 ppm for hydrophilic extracts of liver and urine and the chloroform singlet at 7.20 ppm for lipid soluble samples.

Using the bucketing procedure each 1D spectrum, in the range of 0.2 and 10.0 ppm, was segmented into 0.02 ppm chemical shift bins and the corresponding spectral areas were integrated using AMIX software (version 3.8.4, Bruker BioSpin). Through this technique, the number of total variables is reduced and small shifts in the signals are compensated, making the analysis more robust and reproducible (Holmes et al. 1994).

For water-soluble extracts, regions between 4.60 and 4.85 ppm containing residual H₂O signal were removed, instead, regions between 6.90 and 7.55 ppm containing chloroform signal were excluded from spectra of lipid-soluble extracts.

On the remaining bins, probabilistic quotient normalization (PQN; Dieterle et al. 2006) was applied prior to pattern recognition.

Of note, systemic biofluids, that is, urine and serum, are able to reflect the global response of an organism to a disease status contrary to tissue samples. They require a simple and minimally invasive collection (Vignoli et al. 2019). Therefore, both urine and serum represent optimal biological matrices for the ¹H-NMR analysis. However it should be realized that urine metabolites are more influenced by factors like age, diet, gut microbiota and/or other pathophysiological stimuli. Thus, urine ¹H-NMR spectra are often more variable and contain crowded regions with a lot of signal overlaps and shifts, while serum samples, being more simple to be analyzed, could actually find a practical veterinary use.

Table 2. Hematological analysis (\pm SD).

Parameters	Cow groups		
	Healthy	LDA	RDA
Leukocytes ($\times 10^9/L$)	13.71 \pm 4.81	16.70 \pm 11.05	11.52 \pm 3.21
Erythrocytes ($\times 10^{12}/L$)	7.35 \pm 1.40 ^b	8.36 \pm 1.22 ^{ab}	8.79 \pm 1.29 ^a
MCV	47.41 \pm 5.73	48.15 \pm 4.74	46.48 \pm 6.04
MCHC ($\times 10$ g/L)	34.62 \pm 2.77 ^a	31.64 \pm 2.47 ^b	30.73 \pm 1.96 ^b
Hematocrit ($\times 10^{-2}$ L/L)	34.15 \pm 4.51 ^b	40.11 \pm 6.61 ^a	41.05 \pm 8.51 ^a
Hemoglobin ($\times 10$ g/L)	11.77 \pm 1.28	12.75 \pm 1.95	12.53 \pm 2.22
Blood pH	7.44 \pm 0.03 ^{ab}	7.42 \pm 0.07 ^b	7.49 \pm 0.06 ^a
pO ₂ (mmHg)	35.72 \pm 6.89	31.80 \pm 5.05	31.87 \pm 2.98
pCO ₂ (mmHg)	36.70 \pm 5.80 ^b	38.67 \pm 5.68 ^b	44.43 \pm 3.92 ^a
Bicarbonate (μ mol/L)	25.30 \pm 2.14 ^b	24.88 \pm 5.21 ^b	33.47 \pm 7.16 ^a
Base excess (μ mol/L)	1.17 \pm 2.41 ^b	1.10 \pm 6.13 ^b	10.18 \pm 7.00 ^a

^{a, b, ab}Means within a row with different superscripts differ ($p < 0.05$).

MCV = Mean corpuscular volume, MCHC = Mean corpuscular hemoglobin concentration.

2.8. Statistical analysis

Statistical significance was determined with one way ANOVA test for all the hematological and biochemical variables.

All metabolomic analysis was performed using R, an open source software for statistical analysis of data (Ihaka and Gentleman 1996). Multivariate analysis was applied on processed data and Principal Component Analysis (PCA) was used as a first exploratory approach. Orthogonal projections to latent structures (OPLS) analysis was applied as a supervised technique. In general, OPLS is a multivariate projection method which is normally used for modeling spectroscopic data. This algorithm is a modification of the PLS method (Wold 1998) and it is based on the idea to separate "response linearly related" and "response unrelated (orthogonal)" in data, providing a simpler method for interpreting them (Trygg and Wold 2002).

Accuracies and confusion matrices for different classifications were assessed by means of 100 cycles of Monte Carlo cross-validation scheme (MCCV, R script in-house developed). For this method, at each iteration, 90% of data are randomly chosen as a training set to build the model. The remaining 10% of data was tested and the accuracy for the classification was established. This procedure is repeated 100 times to derive an average discrimination accuracy for each group of subjects.

Univariate statistical analysis was carried on ¹H-NMR data after they have been normalized using PQN method. In particular, spectral regions related to the different metabolites were assigned by using AMIX 3.8.4 (a Bruker BioSpin software) and published literature data. The same regions were integrated to get concentration values of metabolites in arbitrary units. Resulting values were used to determine discriminating metabolites among the groups of cows. Wilcoxon signed-rank test (Neuhäuser 2011) was applied to deduce metabolite differences among groups on the biological assumption that metabolite

concentrations are not normally distributed. False discovery rate correction was used applying the Benjamini-Hochberg method (FDR) (Benjamini and Hochberg 1995) and the adjusted p value < 0.05 was considered statistically significant.

Biochemical network mapping and related pathway analysis were also performed for serum metabolites (see the Supplementary Material file for details).

3. Results

3.1. Blood and chemistry profile

The main differences in biochemical parameters of LDA group included increased NEFA ($p < 0.001$), BHBA ($p < 0.001$) and triglyceride ($p < 0.001$) as compared to controls. There were significant increases in lactate, BE and CRP in RDA levels, and decreased K in RDA affected animals group. In spite of significant changes in other biochemical, and hematological parameters, they were within normal reference ranges (Tables 2 and 3).

3.2. NMR-Based metabolomic evaluation

NMR spectra of serum, liver and urine samples were acquired. Two samples for serum water-soluble extracts; five for serum lipid extracts; six for liver lipid soluble samples and five for urine lipid soluble samples were removed from the statistical analysis because of the bad quality of the generated spectra.

Processed 1D spectra from all type of extracts including the three different groups of animals (healthy cows, cows with LDA and cows with RDA) have been analyzed firstly using PCA to have an overview of the main differences between healthy cows and those with the displacement of the abomasum. Figure 1a–g shows the PCA score plots on 1D NOESY of all types of extracts and on 1D CPMG of serum water soluble samples.

The resulting score plots are not sufficient to discriminate healthy animals from cows with abomasum displacement and no evident differences are reported for the discrimination between cows with left or right displacement.

To explore differences between healthy and diseased cows, the OPLS-supervised method was used. OPLS models were built on 1D NOESY spectra and on 1D CPMG spectra for serum water soluble samples (Figure 2a–g) and a different number of components were retained in the model depending on the type of samples. All models comparing healthy versus diseased animals, as shown by prediction accuracies of cross-validation analyses in Figure 2a–g, are able to discriminate each group of cows with high accuracies and in particular, healthy animals were discriminated

Table 3. Biochemical analyses (\pm SD).

Parameters	Cow groups		
	Healthy	LDA	RDA
Na (mmol/L)	146.90 \pm 5.41 ^a	139.76 \pm 5.30 ^b	139.75 \pm 5.20 ^b
Cl (mmol/L)	104.60 \pm 3.25 ^a	98.30 \pm 7.67 ^a	85.25 \pm 13.93 ^b
K (mmol/L)	4.24 \pm 0.38 ^a	3.43 \pm 0.64 ^b	2.92 \pm 0.73 ^b
P (mmol/L)	2.3256 \pm 0.42313 ^a	1.59239 \pm 0.54587 ^b	1.9057 \pm 0.9108 ^{ab}
Mg (mmol/L)	0.8877 \pm 0.17673 ^a	0.63705 \pm 0.17673 ^b	0.8918 \pm 0.3945 ^a
Ca (mmol/L)	3.2625 \pm 0.275 ^a	2.7375 \pm 0.33 ^b	2.63 \pm 0.395 ^b
GGT (U/L)	15.90 \pm 5.87 ^b	41.52 \pm 34.26 ^{ab}	70.75 \pm 63.33 ^b
CK (U/L)	326.10 \pm 212.07	487.64 \pm 498.11	395.37 \pm 279.31
AST (U/L)	134.40 \pm 45.56	191.38 \pm 120.86	175.62 \pm 137.23
ALP (U/L)	58.30 \pm 12.97	63.47 \pm 29.59	79.62 \pm 41.76
LDH (U/L)	2070.60 \pm 247.84	2005.59 \pm 571.85	2109.75 \pm 381.83
BUN (mmol)	2.96 \pm 0.6 ^b	3.64 \pm 2.19 ^b	7.81 \pm 5.85 ^a
Creatinine (μ mol/L)	96.36 \pm 13.26 ^b	101.66 \pm 33.59 ^b	143.21 \pm 79.56 ^a
Total bilirubin (μ mol/L)	37.28 \pm 18.1 ^a	12.9 \pm 6.8 ^b	12.9 \pm 7.0 ^b
Glucose (mmol/L)	2.93 \pm 0.54 ^b	4.59 \pm 2.84 ^b	7.06 \pm 3.12 ^a
Lactate (mmol/L)	1.49 \pm 1.01 ^b	2.80 \pm 2.39 ^{ab}	3.97 \pm 3.99 ^a
Cholesterol (mmol/L)	3.390 \pm 1.5 ^o	2.3 \pm 1.04	2.86 \pm 1.21
Triglyceride (mmol/L)	0.35 \pm 0.34 ^{ab}	0.63 \pm 0.51 ^a	0.24 \pm 0.16 ^b
Total protein (g/L)	0.08 \pm 0.01 ^a	0.07 \pm 0.01 ^b	0.07 \pm 0.01 ^b
Albumin (g/L)	0.036 \pm 0.00 ^a	0.0324 \pm 0.01 ^{ab}	0.03 \pm 0.00 ^b
NEFA (mmol/L)	0.28 \pm 0.36 ^b	1.26 \pm 0.60 ^a	1.09 \pm 0.56 ^a
BHBA (mmol/L)	0.82 \pm 0.50 ^b	1.94 \pm 1.35 ^a	0.97 \pm 0.77 ^{ab}
CRP (nmol/L)	5.52 \pm 4.57 ^a	10 \pm 5.81 ^{ab}	12.67 \pm 6.38 ^b

^{a, b, ab}Means within a row with different superscripts differ ($p < 0.05$).

NEFA = Non-esterified fatty acid, BHBA = β -hydroxybutyrate, CRP = C-reactive protein.

from the diseased ones with higher accuracies than 80% regardless of the type of sample under study. However, systemic biofluids, that is, urine and blood serum, are able to reflect the global response of an organism to a disease status with respect to tissues biopsies. They required a simple and minimally invasive collection (Vignoli et al. 2019). Therefore, both urine and blood serum represent optimal biological matrices for the ¹H-NMR analysis of the DA disease, however urine metabolites are more influenced by, for example, age, lifestyle, diet, the activity of gut microflora or another symbiotic organism and/or other pathophysiological stimuli. Thus, urine ¹H-NMR spectra are often more variable and contain crowded regions with a lot of signal overlaps and shifts, while serum blood samples, being more simple to be analyzed, could actually find a practical veterinary use.

¹H-NMR spectra were also analyzed to identify which metabolites are altered in the three groups of cows. The complete list of identified and quantified metabolites from both the lipophilic and the hydrophilic fractions from each type of sample is in Table 4; adjusted p values are reported only for metabolites that differ significantly (p value < 0.05) among the various comparisons performed. In particular, for each type of sample, healthy cows were compared to subjects with both LDA and RDA. To highlight metabolites that are significantly different between the two types of displacements, the comparison between cows with left and right abomasum dislocation was performed too.

From the analysis of serum samples, it appeared that healthy cows showed higher levels of hippuric acid, glycine, citrate, trimethylamine N-oxide,

tyrosine, propionate, 2-aminobutyric acid, acetate, isoleucine, and alanine both when they were compared to LDA and RDA. In particular, hippuric acid is very high in healthy in comparison with LDA, instead, a high level of glycine was reported for healthy when compared to RDA. Furthermore, healthy cows showed higher levels of lipids respect to the diseased animals.

Diseased cows (with left or right displacement) compared with healthy animals appeared to be richer in 2-hydroxybutyrate, inosine, and lactate. Furthermore, glycine and 2-aminobutyrate were significantly higher for LDA than for RDA cows. Serum and liver contained high concentrations of 2-hydroxybutyrate in diseased cows compared with those from healthy ones.

In addition, the liver of diseased cows showed higher levels of myoinositol. Instead, healthy animals appeared to have higher levels of glucose in their liver and glycerol phosphocholine appeared to be higher in LDA subjects when compared to RDA animals.

Interesting information arises also from the analysis of liver lipids. Indeed, it resulted in that signal of glycerol backbone protons $-(CH_2)-$ was very high in cows with LDA when compared to healthy, but not in the case of the comparison with RDA animals. Then, signals of fatty acids arising from $-(CH_2)_n$, $CH_3(CH_2)_n$ and $-CH_2-CO$ protons were higher in diseased both when we considered LDA and RDA groups related to healthy.

At the end, from urine samples, high levels of hippuric acid and trimethylamine N-oxide in healthy subjects were confirmed and high levels of dimethylamine, choline, and creatinine resulted in diseased

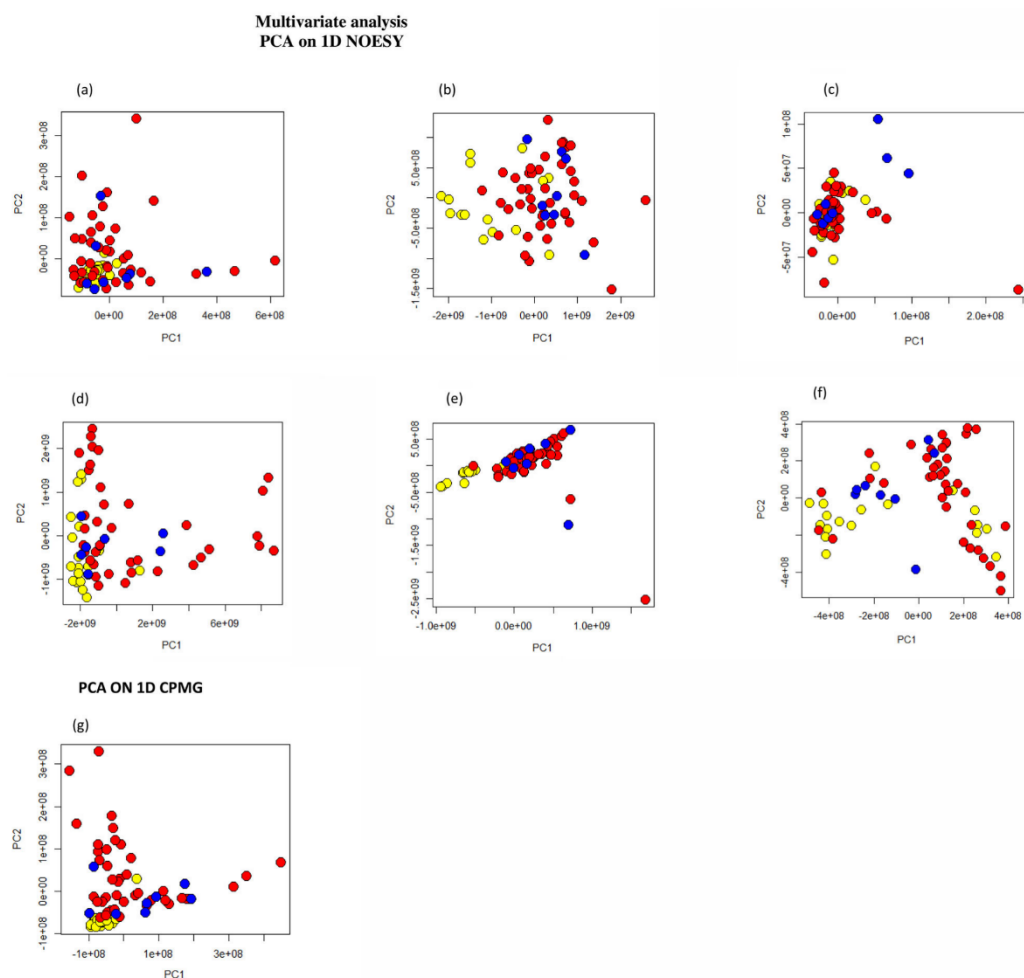


Figure 1. Principal component analysis (PCA) score plots. Each dot represents a single $^1\text{H-NMR}$ spectrum and each color represents a different group of cows: healthy (yellow dots), LDA (red dots) and RDA (blue dots). (a) PCA on 1D NOESY spectra of serum water soluble samples; (b) PCA on 1D NOESY spectra of serum lipid soluble samples; (c) PCA on 1D NOESY spectra of liver water soluble samples; (d) PCA on 1D NOESY spectra of liver lipid soluble samples; (e) PCA on 1D NOESY spectra of urine water soluble samples; (f) PCA on 1D NOESY spectra of urine lipid soluble samples; (g) PCA on 1D CPMG spectra of serum water soluble samples.

animals (both LDA and RDA). Furthermore, cows with LDA showed a very high level of 2-hydroxy-3-methyl valerate when compared to healthy animals; instead, cows with RDA have a very high level of dimethylamine.

The analysis of urine lipids is less straightforward because of the reduced number of lipids in urine, but among the eight lipophilic assigned fractions, it resulted that signals corresponding to glycerol backbone protons $-(\text{CH}_2)-$ were higher for diseased animals, as described for liver samples. Signals arising from $-(\text{CH}_2)_n$, $\text{CH}_3(\text{CH}_2)_n$ protons of fatty acids are higher in RDA group when compared to LDA subjects.

Summarizing the most significant findings obtained from the analysis of the three different matrices (serum, liver and urine samples), it is

obvious that hippuric acid, citrate, dimethyl sulfone, trimethylamine oxide, tyrosine, propionate and 2-aminobutyric acid concentrations were lower (adjusted p values <0.01) in serum in both LDA and RDA cows. In addition, liver glucose concentration was very low (adjusted p value <0.001) in both LDA and RDA cows as compared to controls. Finally, only urine trimethylamine oxide concentration was (adjusted p value <0.05) lower for both right and left abomasal displacements. Moreover, phosphoglycerides, phospholipids, unsaturated and polyunsaturated fatty acids and sphingomyelin were lower (adjusted p values <0.05) in diseased cows' serum only. Finally, the biochemical network and pathway mappings performed on serum metabolites highlighted the "valine, leucine and isoleucine

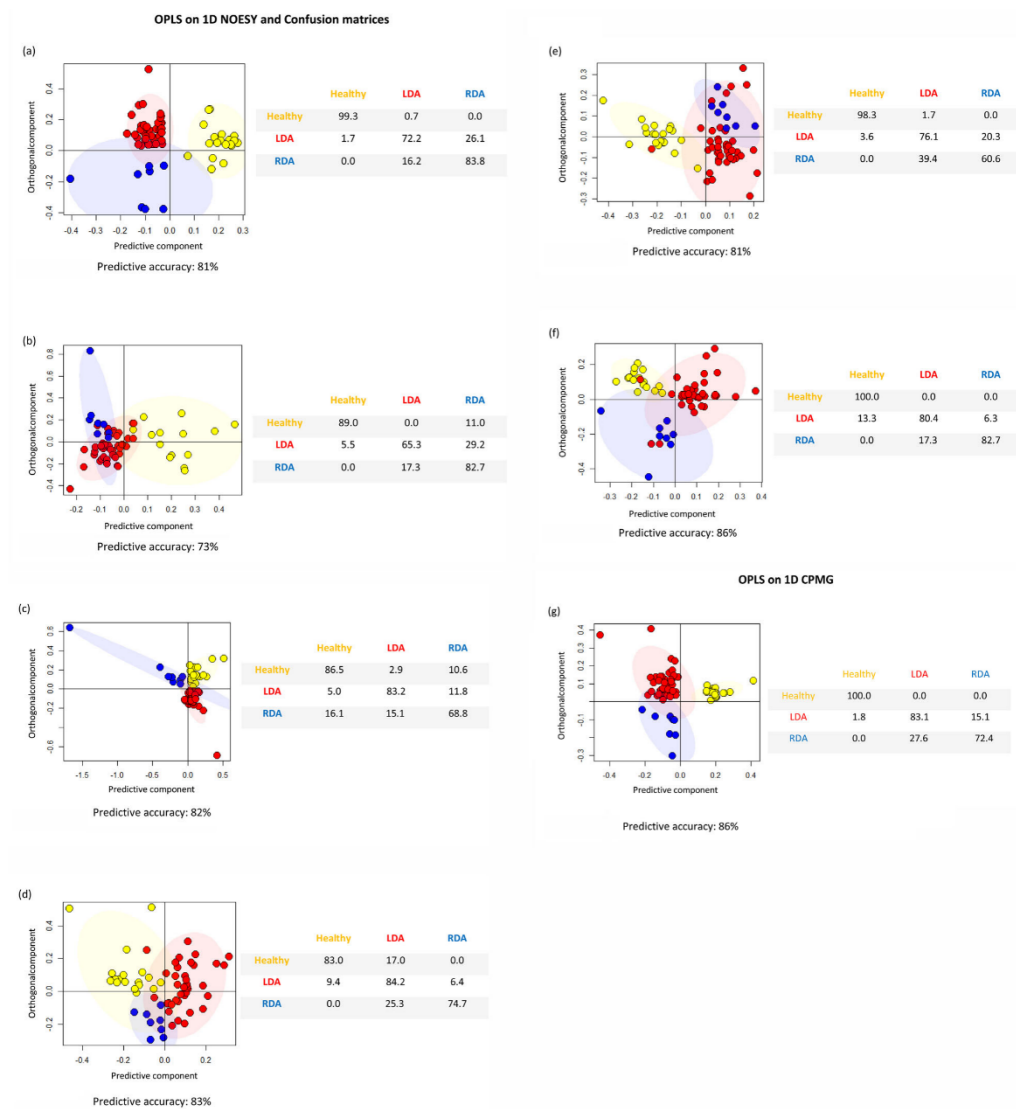


Figure 2. OPLS score plots of all type of extract. Each dot represents a single $^1\text{H-NMR}$ spectrum and each color represents a different group of cows: healthy (yellows dots), LDA (red dots) and RDA (blue dots). (a) OPLS on 1D NOESY spectra of serum water soluble samples; (b) OPLS on 1D NOESY spectra of serum lipid soluble samples; (c) OPLS on 1D NOESY spectra of liver water soluble samples; (d) OPLS on 1D NOESY spectra of liver lipid soluble samples; (e) OPLS on 1D NOESY spectra of urine water soluble samples; (f) OPLS on 1D NOESY spectra of urine lipid soluble samples; (g) OPLS on 1D CPMG spectra of serum water soluble samples. Confusion matrices and related predictive accuracy of cross-validation analysis are reported for each model.

biosynthesis” and the “phenylalanine, tyrosine and tryptophan biosynthesis” as the most probable altered metabolic pathway in DA condition (see Supplementary Figures S1 and S2).

4. Discussion

Displaced abomasum was evaluated more comprehensively using serum, urine and liver metabolome of dairy cows in the present study.

The most commonly used indicators of energy status are NEFA and BHBA concentrations during both the dry and puerperal periods, as well as total protein and albumin during the puerperal period (Puppel and Kuczyńska 2016). A predictive association of elevated concentrations of BHBA with the risk of DA has been reinforced. Furthermore, low Ca and K concentrations, and high apo B100 concentration as well as, AST and GGT activities are related to the subsequent occurrence of DA (Sevinc et al. 2002;

Table 4. Concentrations in arbitrary units (median \pm Median Absolute Deviation (MAD)) of the metabolites assigned in serum, liver and urine samples (both hydrophilic and lipophilic fractions).

	Metabolites	Healthy (arbitrary units)	LDA (arbitrary units)	RDA (arbitrary units)	"adjusted" p value (<0.05)
Serum water soluble (SWS)	2-aminobutyrate	145.0 \pm 63.4	67.9 \pm 28.5	20.6 \pm 10.1	<0.05 (healthy versus LDA) <0.001 (healthy versus RDA) <0.001 (LDA versus RDA)
	2-hydroxybutyrate	3.2 \pm 3.2	201.2 \pm 155.3	311.4 \pm 165.4	<0.0000001 (healthy versus LDA) <0.001 (healthy versus RDA)
	3-hydroxybutyrate	4896.0 \pm 984.3	8454.3 \pm 5252.8	3221.2 \pm 810.1	<0.05 (healthy versus LDA)
	3-hydroxyisobutyrate	212.3 \pm 58.0	228.9 \pm 68.2	265.9 \pm 101.9	<0.05 (healthy versus LDA)
	Acetate	15866.6 \pm 2143.8	0.01 \pm 0.01	0.01 \pm 0.01	<0.00001 (healthy versus LDA) <0.05 (healthy versus RDA)
	Alanine	1603.0 \pm 305.1	1082.9 \pm 337.3	800.8 \pm 230.0	<0.001 (healthy versus LDA) <0.0001 (healthy versus RDA)
	Benzoate	27.2 \pm 14.8	18.5 \pm 18.5	150.2 \pm 130.7	<0.05 (healthy versus LDA)
	Carnitine	681.9 \pm 97.0	1000.0 \pm 451.8	752.3 \pm 354.5	<0.0000001 (healthy versus LDA)
	Citrate	2814.4 \pm 242.9	698.6 \pm 333.4	907.5 \pm 188.2	<0.0001 (healthy versus RDA)
	Choline	456.2 \pm 86.6	531.9 \pm 146.7	393.5 \pm 110.8	<0.05 (healthy versus LDA)
	Creatine	2285.0 \pm 357.1	1788.7 \pm 385.9	1518.3 \pm 703.9	<0.05 (healthy versus LDA)
	Creatinine	943.9 \pm 139.0	819.1 \pm 172.9	1087.8 \pm 449.4	<0.05 (healthy versus LDA)
	Dimethyl sulfone	661.9 \pm 189.9	179.7 \pm 82.8	136.3 \pm 41.3	<0.00000001 (healthy versus LDA) <0.00001 (healthy versus RDA)
	formate	650.8 \pm 77.0	509.9 \pm 155.6	564.1 \pm 97.1	<0.05 (healthy versus LDA)
	Fructose	675.5 \pm 110.5	745.0 \pm 244.7	1466.1 \pm 378.3	<0.05 (healthy versus LDA) <0.05 (healthy versus RDA)
	Glycine	2180.8 \pm 213.5	1270.3 \pm 460.7	184.8 \pm 184.8	<0.05 (healthy versus LDA) <0.05 (healthy versus RDA) <0.05 (LDA versus RDA)
	Glucose	1233.0 \pm 106.3	1105.7 \pm 256.1	1434.0 \pm 342.2	<0.001 (healthy versus LDA)
	Glutamine	19.3 \pm 7.8	40.1 \pm 13.7	15.7 \pm 13.2	<0.00000001 (healthy versus LDA) <0.00001 (healthy versus RDA)
	Hippurate	437.2 \pm 115.7	58.2 \pm 33.2	54.0 \pm 15.8	<0.0001 (healthy versus LDA) <0.00001 (healthy versus RDA)
	inosine	6.13 \pm 6.13	25.1 \pm 18.8	38.3 \pm 10.1	<0.001 (healthy versus LDA) <0.05 (healthy versus RDA)
	Isoleucine	675.8 \pm 106.4	352.1 \pm 118.2	150.2 \pm 50.3	<0.00001 (healthy versus LDA) <0.05 (healthy versus RDA)
	lactate	2240.6 \pm 426.1	3023.0 \pm 1328.1	3024.6 \pm 900.6	<0.05 (healthy versus LDA) <0.05 (healthy versus RDA)
	leucine	1169.5 \pm 195.8	816.5 \pm 318.1	481.4 \pm 136.5	<0.05 (healthy versus LDA)
	Lysine	559.9 \pm 44.9	545.9 \pm 58.3	468.9 \pm 38.1	<0.05 (healthy versus RDA) <0.05 (LDA versus RDA)
	Mannose	205.0 \pm 40.9	265.0 \pm 72.5	457.2 \pm 68.2	<0.001 (healthy versus LDA) <0.05 (healthy versus RDA)
	Myoinositol	83.7 \pm 17.4	83.7 \pm 24.0	93.9 \pm 43.6	<0.001 (healthy versus LDA)
	Phenylalanine	486.9 \pm 60.0	342.7 \pm 79.3	297.2 \pm 98.2	<0.05 (healthy versus LDA)
	Proline	176.9 \pm 27.8	146.0 \pm 35.4	139.4 \pm 61.5	<0.00001 (healthy versus LDA)
	Propionate	60.3 \pm 16.0	27.6 \pm 13.0	24.7 \pm 8.6	<0.001 (healthy versus RDA) <0.05 (healthy versus LDA)
	Succinate	791.9 \pm 142.8	570.7 \pm 170.8	558.5 \pm 194.9	<0.05 (healthy versus LDA)
	Threonine	57.1 \pm 10.1	47.0 \pm 13.0	31.5 \pm 10.4	

(continued)

Table 4. Continued.

	Metabolites	Healthy (arbitrary units)	LDA (arbitrary units)	RDA (arbitrary units)	"adjusted" p value (<0.05)	
Serum lipid soluble (SLS)					<0.05 (healthy versus LDA)	
					<0.05 (healthy versus RDA)	
					<0.000000001 (healthy versus LDA)	
					<0.0001 (healthy versus RDA)	
					<0.000000001 (healthy versus LDA)	
					<0.0001 (healthy versus RDA)	
					<0.05 (healthy versus LDA)	
					<0.00001 (healthy versus LDA)	
					<0.05 (healthy versus RDA)	
					<0.00001 (healthy versus LDA)	
					<0.05 (healthy versus RDA)	
					<0.000001 (healthy versus LDA)	
					<0.05 (healthy versus RDA)	
					<0.05 (healthy versus LDA)	
					<0.0001 (healthy versus LDA)	
					<0.001 (healthy versus RDA)	
					<0.00001 (healthy versus LDA)	
					<0.05 (healthy versus RDA)	
	Liver water soluble (LWS)					<0.05 (healthy versus LDA)
						<0.05 (healthy versus RDA)
					<0.05 (healthy versus LDA)	
					<0.05 (healthy versus RDA)	
					<0.05 (healthy versus LDA)	
					<0.05 (healthy versus RDA)	
					<0.05 (healthy versus LDA)	
					<0.05 (healthy versus RDA)	
					<0.05 (healthy versus LDA)	
					<0.05 (healthy versus RDA)	
					<0.05 (healthy versus LDA)	
					<0.05 (healthy versus RDA)	
					<0.05 (healthy versus LDA)	
					<0.05 (LDA versus RDA)	

(continued)

Table 4. Continued.

	Metabolites	Healthy (arbitrary units)	LDA (arbitrary units)	RDA (arbitrary units)	"adjusted" p value (<0.05)
	Alanine	1684.8 ± 317.7	1251.5 ± 335.9	901.6 ± 486.6	
	Aspartate	223.3 ± 45.0	137.5 ± 40.5	117.9 ± 66.5	<0.05 (healthy versus LDA)
	Choline	5494.4 ± 1219.7	6712.6 ± 1175.0	6621.9 ± 2257.1	
	Creatine	613.5 ± 223.1	663.8 ± 331.4	1152.4 ± 298.8	<0.05 (healthy versus RDA)
	Creatinine	429.1 ± 107.8	463.8 ± 182.7	401.9 ± 153.6	<0.05 (LDA versus RDA)
	Formate	42.7 ± 17.4	51.0 ± 19.8	85.2 ± 64.0	
	Fumarate	51.7 ± 21.6	67.5 ± 18.4	47.6 ± 46.0	
	Glycerol	335.5 ± 78.1	678.9 ± 93.9	638.2 ± 201.4	<0.00000001 (healthy versus LDA)
	sn-Glycero-3-phosphocoline	6496.9 ± 1722.2	5391.4 ± 1085.4	3154.2 ± 1317.5	<0.05 (healthy versus LDA)
	Glycine	1180.1 ± 275.8	2091.9 ± 306.7	1759.6 ± 543.4	<0.001 (LDA versus RDA)
	Glucose	316.3 ± 144.4	25.7 ± 25.7	3.1 ± 3.1	<0.0001 (healthy versus LDA)
	Glutamate	91.8 ± 26.9	116.5 ± 35.7	73.8 ± 59.5	<0.001 (healthy versus RDA)
	Isoleucine	43.3 ± 19.9	31.6 ± 11.8	35.0 ± 7.5	<0.05 (LDA versus RDA)
	Isopropanol	16.5 ± 6.4	19.2 ± 8.7	37.3 ± 10.6	<0.05 (healthy versus RDA)
	Lactate	977.6 ± 247.6	657.6 ± 120.0	837.5 ± 438.1	<0.05 (healthy versus LDA)
	Leucine	27.9 ± 27.9	44.0 ± 20.6	34.2 ± 15.3	
	Myoinositol	39.5 ± 16.9	91.2 ± 28.0	117.0 ± 35.0	<0.00001 (healthy versus LDA)
	O-phosphocoline	1956.1 ± 474.6	2171.6 ± 548.4	3725.5 ± 1229.3	<0.001 (healthy versus RDA)
	Succinate	236.6 ± 42.7	223.0 ± 31.8	193.7 ± 62.6	
	Uracil	52.7 ± 46.5	56.5 ± 37.3	71.1 ± 46.3	
	Uridine	69.3 ± 30.1	125.3 ± 22.6	113.3 ± 39.7	<0.00001 (healthy versus LDA)
	Unknow	278.6 ± 78.2	122.3 ± 37.7	82.6 ± 65.2	<0.05 (healthy versus LDA)
	Valine	111.4 ± 36.9	78.1 ± 23.3	82.3 ± 29.4	<0.05 (healthy versus RDA)
Liver lipid soluble (LLS)	Free cholesterol -C(3)H-	873.5 ± 393.6	1112.8 ± 395.5	1205.4 ± 349.2	
	Total cholesterol -C(18)H ₃ Multiple cholesterol protons	3367.7 ± 818.9	4317.6 ± 1699.1	4667.2 ± 686.1	<0.05 (healthy versus RDA)
	Fatty acids -CH ₃ -(CH ₂) _n -	4930.3 ± 373.3	6027.2 ± 1271.3	6327.8 ± 567.6	<0.05 (healthy versus LDA)
	Fatty acids -(CH ₂) _n -	78874.6 ± 7424.3	130014.7 ± 32490.4	97446.8 ± 9668.3	<0.05(healthy versus RDA)
	Fatty acids -CH ₂ -CH ₂ -CO	672145.9 ± 75460.2	1034204.0 ± 270216.9	853648.9 ± 106179.1	<0.000001 (healthy versus LDA)
	Fatty acids -CH ₂ -CH=	455655.5 ± 113667.9	549820.1 ± 115872.8	601105.4 ± 43428.1	<0.05 (healthy versus RDA)
	Fatty acids -CH ₂ -CO	34891.7 ± 10498.9	57371.2 ± 39746.7	52710.2 ± 16408.1	<0.001 (healthy versus LDA)
	Fatty acids =CH-CH ₂ -CH=	47047.6 ± 10198.8	79889.6 ± 33886.7	64893.3 ± 13046.1	<0.05 (healthy versus RDA)
	Fatty acids =CH-CH ₂ -CH=	26475.0 ± 10383.5	30851.1 ± 15070.1	24588.6 ± 4016.0	
	Fatty acids -CH=CH-	47835.0 ± 17628.7	78192.9 ± 42839.6	56367.3 ± 8661.7	
	Glycerol backbone -(CH ₂)-	199.2 ± 199.2	9267.3 ± 9212.6	1655.4 ± 1438.7	<0.001 (healthy versus LDA)
	Phosphoglycerides -CH-	13916.0 ± 4406.2	13895.6 ± 5307.9	16490.1 ± 3156.4	
	Phosphatidylcholine	12167.2 ± 4147.0	9516.6 ± 3512.6	12207.6 ± 3018.5	
	Sphingomyelin-choline	40154.9 ± 13163.9	29159.9 ± 11177.4	38578.1 ± 9073.0	
	Unknow 1	949.2 ± 156.8	581.2 ± 68.1	654.5 ± 82.1	<0.000001 (healthy versus LDA)
	Unknow 2	3851.5 ± 629.2	3129.0 ± 318.4	3089.0 ± 365.6	<0.05 (healthy versus RDA)
	3-hydroxybutyrate	2057.4 ± 675.8	3756.0 ± 3329.4	1168.0 ± 895.7	<0.05 (healthy versus LDA)

(continued)

Table 4. Continued.

	Metabolites	Healthy (arbitrary units)	LDA (arbitrary units)	RDA (arbitrary units)	"adjusted" p value (<0.05)	
Urine water soluble (UWS)					<0.05 (healthy versus LDA)	
					<0.05 (LDA versus RDA)	
					<0.05 (healthy versus LDA)	
		2-hydroxy-3-methyl-valerate	1526.9 ± 419.4	2988.3 ± 1348.7	1687.1 ± 409.5	<0.001 (healthy versus LDA)
		4-hydroxyphenylacetate	572.2 ± 60.0	388.4 ± 138.6	444.2 ± 183.7	<0.001 (healthy versus LDA)
		acetate	6830.7 ± 2345.6	1482.6 ± 1266.6	4481.2 ± 4150.3	<0.001 (healthy versus LDA)
		Alanine	336.6 ± 108.1	417.8 ± 379.1	243.7 ± 52.1	<0.000001 (healthy versus LDA)
		Allantoin	349.2 ± 61.7	5061.5 ± 3542.5	5761.6 ± 5104.0	<0.05 (healthy versus RDA)
						<0.05 (healthy versus LDA)
		Benzoate	22.8 ± 22.8	212.6 ± 212.6	0.8 ± 0.8	<0.001 (healthy versus LDA)
		Citrate	4678.6 ± 3277.1	1507.7 ± 1286.6	1985.2 ± 1733.5	<0.001 (healthy versus LDA)
		Choline	51.9 ± 51.9	365.0 ± 200.3	389.0 ± 89.4	<0.000001 (healthy versus LDA)
						<0.001 (healthy versus RDA)
		Creatine	8791.9 ± 2611.7	21502.2 ± 8430.5	12218.0 ± 9821.0	<0.0000001 (healthy versus LDA)
		Creatinine	10687.5 ± 2748.9	16234.4 ± 6726.3	23289.9 ± 10904.7	<0.0001 (healthy versus LDA)
						<0.001 (healthy versus RDA)
		Dimethylamine	27.9 ± 18.2	2248.5 ± 1265.4	2272.1 ± 1613.8	<0.00000001 (healthy versus LDA)
						<0.05 (healthy versus RDA)
		Dimethyl sulfone	1348.3 ± 300.8	275.6 ± 196.8	484.4 ± 220.2	<0.0000001 (healthy versus LDA)
						<0.05 (healthy versus RDA)
		Formate	1068.1 ± 143.3	231.5 ± 182.6	1282.3 ± 748.7	<0.00001 (healthy versus LDA)
						<0.05 (LDA versus RDA)
		Fumarate	29.1 ± 26.9	33.2 ± 22.1	26.2 ± 12.5	<0.05 (healthy versus LDA)
	Glyoxilate	49.8 ± 28.7	21.0 ± 20.7	49.3 ± 48.6	<0.000000000001 (healthy versus LDA)	
	Hippurate	95714.3 ± 5107.3	24948.8 ± 14254.0	26049.2 ± 7194.1	<0.00001 (healthy versus RDA)	
	Isobutyrate	347.6 ± 207.2	637.1 ± 346.9	333.2 ± 258.0	<0.05 (healthy versus LDA)	
	Lactate	1473.5 ± 1473.5	5027.5 ± 4525.9	360.1 ± 151.7	<0.05 (healthy versus LDA)	
					<0.05 (healthy versus RDA)	
	Lactose	360.1 ± 151.7	1704.5 ± 1254.4	225.4 ± 225.4	<0.001 (healthy versus LDA)	
					<0.05 (RDA versus LDA)	
	N-phenylacetyl glycine	18749.8 ± 4876.8	16422.8 ± 4249.5	15927.5 ± 2047.1	<0.0000001 (healthy versus LDA)	
	3-hydroxybutyrate	2352.4 ± 1941.6	1657.9 ± 1253.0	1903.9 ± 1513.0	<0.05 (healthy versus RDA)	
	Trimethylamine N-oxide	40396.6 ± 3423.7	11389.4 ± 6067.4	9964.5 ± 3920.7	<0.0000001 (healthy versus LDA)	
					<0.05 (healthy versus RDA)	
	Tyrosine	567.0 ± 125.3	145.2 ± 145.2	497.0 ± 474.1	<0.05 (healthy versus LDA)	
	Unknow 1	7163.5 ± 1254.9	9895.8 ± 3864.0	10944.8 ± 3811.2	<0.05 (healthy versus LDA)	
					<0.05 (healthy versus RDA)	
	Unknow 2	388.2 ± 123.1	1715.2 ± 758.2	539.8 ± 167.9	<0.00001 (healthy versus LDA)	
					<0.05 (LDA versus RDA)	
	Unknow 3	643.9 ± 217.9	1255.1 ± 575.8	670.1 ± 163.3	<0.05 (LDA versus RDA)	
Urine lipid soluble (ULS)	Total cholesterol C(18)H ₃	171.8 ± 89.8	53.6 ± 53.6	0.01 ± 0.01		
	Glycerol backbone -(CH ₂)-	18.4 ± 18.4	197.6 ± 42.3	165.3 ± 37.2	<0.0000001 (healthy versus LDA)	
					<0.001 (healthy versus RDA)	
	Fatty acids -CH ₃ -(CH ₂) _n -	122298.2 ± 1434.7	9413.9 ± 806.0	14060.8 ± 351.8	<0.05 (healthy versus RDA)	
					<0.0001 (RDA versus LDA)	
Fatty acids -(CH ₂) _n	85753.5 ± 12290.1	56153.8 ± 6561.3	94974.8 ± 7035.4	<0.001 (RDA versus LDA)		
Fatty acids -CH ₂ -CH ₂ -CO	274637.5 ± 36310.8	214963.2 ± 26967.3	217487.8 ± 24605.4	<0.0001 (healthy versus LDA)		

(continued)

Table 4. Continued.

Metabolites	Healthy (arbitrary units)	LDA (arbitrary units)	RDA (arbitrary units)	"adjusted" <i>p</i> value (<0.05)
Fatty acids = CH- CH ₂ -CH=	140.6 ± 73.5	51.6 ± 51.6	96.8 ± 17.8	<0.05 (healthy versus RDA) <0.05 (healthy versus LDA)
Fatty acids -CH = CH-	643.0 ± 260.8	384.6 ± 189.4	325.7 ± 290.0	<0.05 (healthy versus LDA)
Unknown	1249.2 ± 93.4	1142.7 ± 126.7	983.2 ± 127.8	<0.001 (healthy versus RDA) <0.05 (LDA versus RDA)

p values adjusted for false discovery rate are reported only for metabolites that differ significantly (*p* value <0.05) among the various comparisons performed. In particular, for each type of sample, healthy cows were compared to subjects with both LDA and RDA. To highlight metabolites that are significantly different between the two types of displacements, the comparison between cows with left and right abomasum dislocation was performed too.

Civelek et al. 2006; Sen et al. 2006; Seifi et al. 2011; Constable et al. 2013). While serum and urine lactate levels from metabolomic analysis were increased in diseased groups, liver lactate was decreased in LDA cows. The timing, magnitude, and duration of peripartum increases in circulating concentrations of NEFA and BHBA are associated with the risk of the displaced abomasum, uterine disease, and reproductive performance from 1 through 20 weeks later (LeBlanc 2010). The generally used cut-off value for the diagnosis of subclinical ketosis is ≥ 1200 and up to $1400 \mu\text{mol/L}$ of blood BHBA (Suthar et al. 2013). Clinical ketosis is generally characterized by concentrations of BHBA in the blood $>3000 \mu\text{mol/L}$ (Oetzel 2007). The different metabolites in cows with milk fever reflected the pathological features of negative energy balance and fat mobilization (Sun et al. 2014). In accordance with the above references, increased lactate, NEFA, BHBA, but decreased K levels in diseased groups were observed in the present study. Although some cows with LDA and RDA had also concomitant ketosis, BHBA level (1.94 ± 1.35) was significantly increased in LDA group as well as 3-hydroxybutyrate in serum and urine samples of LDA cows were increased. The 3-hydroxybutyrate was higher in RDA group compared with LDA cows. Increased blood concentrations of isopropanol are observed in ketotic cows (Sato 2009). In the present study, isopropanol and acetate was found higher in liver samples of RDA group than in healthy and LDA cows. Increased serum amyloid A and haptoglobin were found in cows with LDA or RDA/AV. Such an increase may indicate the presence of hepatic lipidosis in cattle with DA (Guzelbektes et al. 2010). The highest values of CRP and haptoglobin were observed in cows during the first month after calving (Dębski et al. 2016). In the present study, the highest values of CRP was observed in RDA cows. This may be attributed to fatty liver in cows with DA.

In ruminants, the principal gluconeogenic substrates include propionic acid, lactic acid, glycerol, and gluconeogenic amino acids (alanine, asparagine, arginine, aspartic acid, cysteine, glycine, histidine,

methionine, proline, propionic acid, serine, valine). A lack of gluconeogenic substrates is an important risk factor in the pathogenesis of ketosis. Ketogenic amino acids (e.g., leucine and lysine) can also enter the tricarboxylic acid cycle by oxidative deamination. A disturbance in the tricarboxylic acid cycle may contribute to clinical ketosis (Sun et al. 2014). A non-metabolomic approach showed that plasma from LDA cattle exhibited significantly higher free fatty acid and BHBA, lower glucogenic amino acids, such as methionine, alanine, and serine, and higher ratio of ketogenic amino acids among blood free amino acids such as leucine and lysine (Hamana et al. 2010). In the present study, both gluconeogenic and ketogenic amino acids such as phenylalanine, isoleucine, and threonine were found decreased as well as ketogenic (leucine, lysine) and gluconeogenic (alanine, isoleucine, glycine, glutamine, proline, valine) amino acids in diseased cows. Lysine and glycine, and even leucine were lower in RDA cows. Ketosis in cows with LDA may be attributed to a lack of gluconeogenic substrates (propionic acid, glycerol, and gluconeogenic acids) and ketotic cows' consumption of large amounts of ketogenic substances.

Carnitine transports the activated fatty acids from the cytosol into the mitochondrion via their corresponding carnitine ester (Stanley et al. 1992). Acetylcarnitine, as the shortest acylcarnitine, facilitates the movement of acetyl-CoA into the matrices of the mitochondria. Furthermore, in the mitochondria, carnitine acetyl-CoA transferase catalyzes the conversion of acetyl-CoA to C2, a membrane permeable metabolite which facilitates mitochondrial efflux of excess acetyl-CoA. Increased hepatic carnitine concentrations observed in 1 wk postpartum and might be regarded as a physiologic means to provide liver cells with sufficient carnitine required for transport of excessive amounts of NEFA during a negative energy balance (Schlegel et al. 2012). Especially the role of specific glycerophospholipids, sphingolipids, and acylcarnitines as potential biomarkers should be considered for metabolic adaptation of transition dairy cows (Kenéz et al. 2016). In the current study, serum carnitine level in diseased cows (especially in

LDA group) was higher than in a healthy group. This may point to an increase in the mitochondrial oxidation of the ketogenic and gluconeogenic amino acids.

Valine, leucine, and isoleucine biosynthesis plays a vital role in milk protein synthesis. Hippuric acid, nicotinamide, and pelargonic acid are milk protein biomarkers (Wu et al. 2018). Hippuric acid, as an aromatic compound, could be converted into aromatic amino acids such as tryptophan, tyrosine, and phenylalanine, then be transformed into nicotinamide. Higher serum hippuric acid might also indicate more energy supplied by glucose metabolism and hormone regulation (Pero 2010). Urinary hippuric acid might reflect the dietary composition, hepatic function, disease state, and even metabolism alterations. A higher urinary hippuric acid level leads to more nitrogen loss and lower milk protein yield in cows fed low-quality forages, which supports the hypothesis that hippuric acid as excretory products is mainly produced from dietary protein degradation (Sun et al. 2016). In this present study, hippuric acid was one of the most prominent metabolites. Its lower level in serum and urine of diseased groups (especially in LDA group) together with lower valine, leucine, and isoleucine in serum may indicate less energy supplied and less nitrogen loss in a cow with DA.

Leucine, one of the three branch chain amino acids, acts as a signaling molecule in the regulation of overall amino acid and protein metabolism. Leucine is also considered to be a potent stimulus for the secretion of insulin from pancreatic β -cells (Sadri et al. 2017). The three branched-chain amino acids, leucine, isoleucine, and valine that are classified as essential amino acids, play important roles in regulating overall amino acids and protein metabolism. Marczuk et al. (2018) found that higher concentrations of glutamine, glutamic acid, isoleucine, and tyrosine in cows with primary ketosis, and that significant decrease in the concentrations of asparagine, histidine, methionine, and serine, alanine, leucine, lysine, and proline. In the current study, significant decreases in leucine, isoleucine, tyrosine, trimethylamine N-oxide, and valine of diseased groups were observed. Thus, protein and amino acid metabolisms seem to be adversely affected in the present study.

According to Shibano et al. (2005), the improvement of the glycine/alanine ratio in high producing dairy cows may indicate increased milk production and milk quality. Furthermore, the glycine/alanine ratio may be a useful indicator for determining nutritional deficiency from the transition period to the peak lactation period. Measurement of glycine/alanine ratio in serum may be useful for evaluating the nutritional status of a periparturient dairy cow. In

the present study, this ratio was found much more impairment in the RDA group, suggesting their poor nutritional status.

The biopsy of the liver is the most reliable method for accurate estimation of the degree of fatty infiltration. It can be used to determine the concentration of triglycerides and the severity of the fatty liver (Herdt et al. 1983). Due to the rapid development of metabolomics in recent years, the use of this approach for disease biomarker assessment has become popular. A metabolomic approach revealed the primary differences including increases in BHBA, acetone, glycine, valine, trimethylamine-N-oxide, citrulline, and isobutyrate, and decreases in alanine, asparagine in plasma samples of cows with fatty liver (Xu et al. 2016). A study evaluating alterations of the lipid metabolome in dairy cows experiencing excessive lipolysis early postpartum showed that overall, excessive lipolysis in the high group came along with impaired estimated insulin sensitivity and characteristic shifts in acylcarnitine, sphingomyelin, phosphatidylcholine, and lysophospholipid metabolome profiles compared to the low group. From the detected phosphatidylcholines mainly those with diacyl-residues showed differences among lipolysis groups. Furthermore, more than half of the detected sphingomyelins were increased in cows experiencing high lipomobilization. Additionally, strong differences in serum acylcarnitines were noticed among lipolysis groups (Humer et al. 2016). The concentration of triacylglycerides in plasma drops at the day of parturition whereas the plasma level of many phosphatidylcholines and sphingomyelins increases steadily during early lactation (Imhasly et al. 2015). The measurement of specific representatives of phosphatidylcholines in plasma may provide a novel diagnostic biomarker of fatty liver disease in dairy cows (Gerspach et al. 2017). Choline can be metabolized to several other products, including betaine, phosphatidylcholine, and acetylcholine, each with critical biological roles (Garcia et al. 2018). The abnormal decline of certain specific phosphatidylcholines and sphingomyelins could be regarded as a promising biomarker indicative of fatty liver disease (Artegoitia et al. 2014; Imhasly et al. 2015). In the current study, while a lower level of many fatty acid fractions and cholesterol were encountered in serum and urine samples, the similar metabolites tended to be higher in liver samples in diseased groups. Fatty acids $\text{CH}_3\text{-(CH}_2)_n$ and fatty acids $\text{-(CH}_2)_n$ were lower in urine samples of LDA cows. Phosphoglycerides, phospholipids, and unsaturated and polyunsaturated fatty acids and sphingomyelin were found lower in diseased cows' serum samples. The similar decreases in liver samples were not significant. sn-Glycero-3-phosphocoline was lower in liver samples of RDA cows. In

our previous study (Basoglu et al. 2014), where metabolites identified and quantified by NMR analysis only in plasma samples were valine, 3 β -hydroxybutyrate, alanine, glutamine, glutamate, and succinate. Among these parameters, succinate decreased significantly in cows with RDA. Pronounced findings between LDA and RDA groups included significant changes in glutamine, glutamate, and 3 β -hydroxybutyrate. This previous study has been extended by the current one where different metabolites mentioned above were measured in serum, urine, and liver; thus pathogenic mechanisms related to energy metabolism of the disease have been more comprehensible.

5. Conclusions

The integration of several pathophysiological aspects, for example, lipolysis, ketogenesis, and oxidative capacity in cows with DA, by combining gluconeogenic and ketogenic amino acids, fatty acids fractions and cholesterol, ketone bodies, choline products, and carnitine, will likely provide more information than the classical measurement of NEFA and BHBA. The metabolomic profile, in the present study, clearly revealed that cows with DA (especially with LDA) have been at risk ketosis and fatty liver and that analyzing changes of the serum, urine and liver metabolome and identifying new biomarkers by metabolomic approach can help to understand the multifaceted disease. The biochemical network and pathway mappings performed on serum metabolites highlight the 'valine, leucine and isoleucine biosynthesis' and the 'phenylalanine, tyrosine and tryptophan biosynthesis' as the most probable altered metabolic pathway in DA condition.

Disclosure statement

The authors declare there are no competing financial interests.

Funding

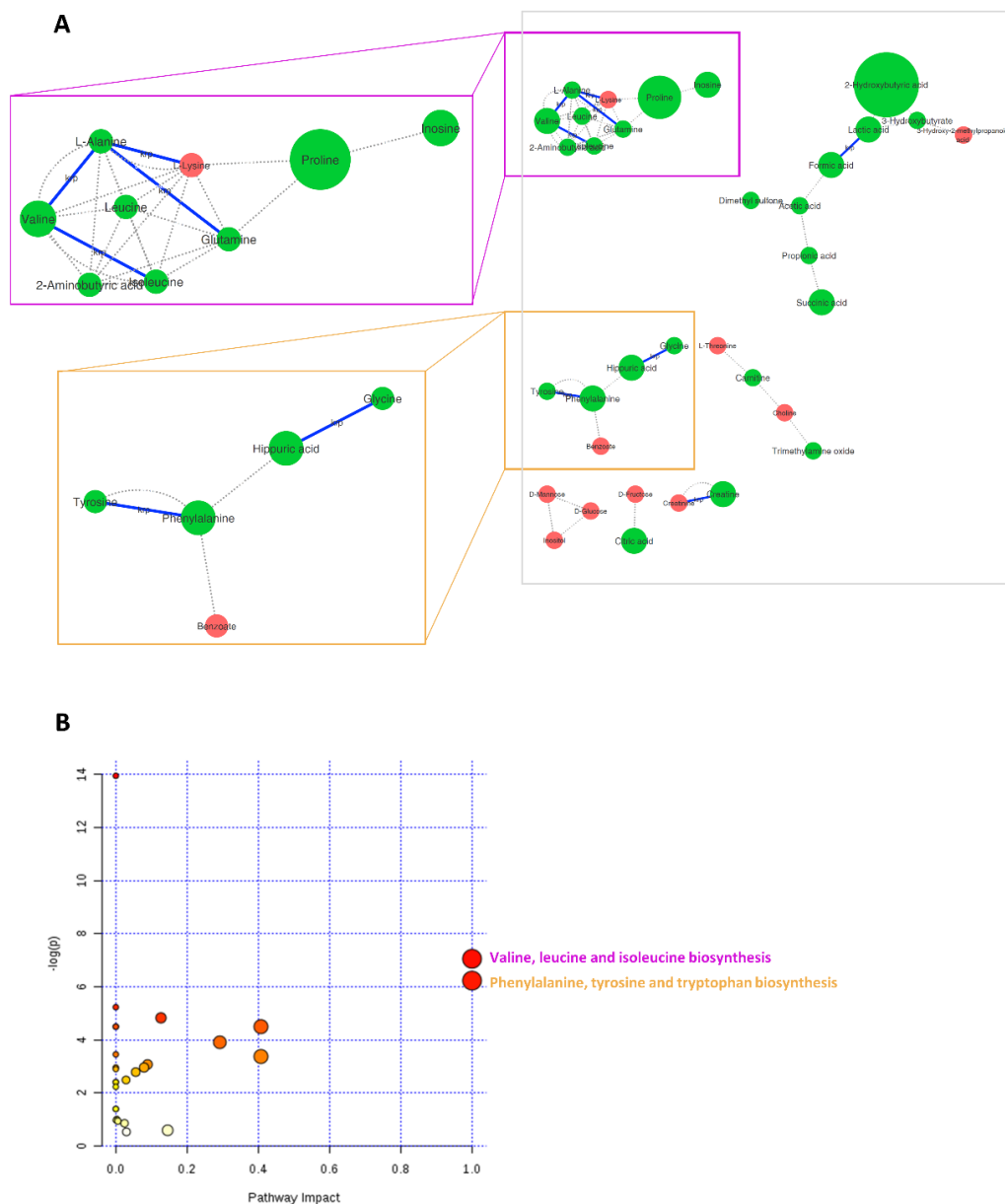
This work was financially supported by the Selcuk University Scientific Research Projects Coordination Unit (Project No: 17102035). CERM/CIRMMP center of the ESFRI Instruct is gratefully acknowledged for the NMR access provision financially supported by the EC Contract INEXT No 653706.

References

Artegoitia VM, Virginia M, Middleton JL, Harte FM, Campagna SR, de Veth MJ. 2014. Choline and choline metabolite patterns and associations in blood and milk during lactation in dairy cows. *PLoS One*. 26:e103412.

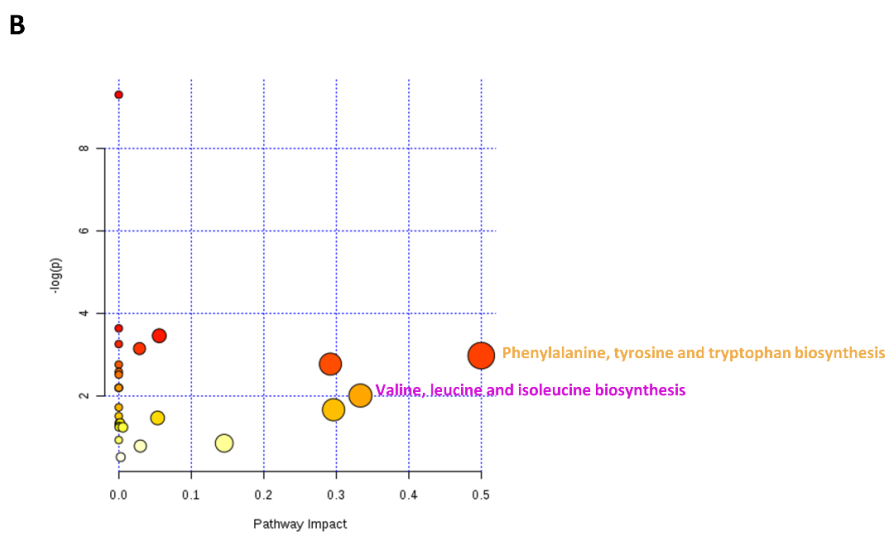
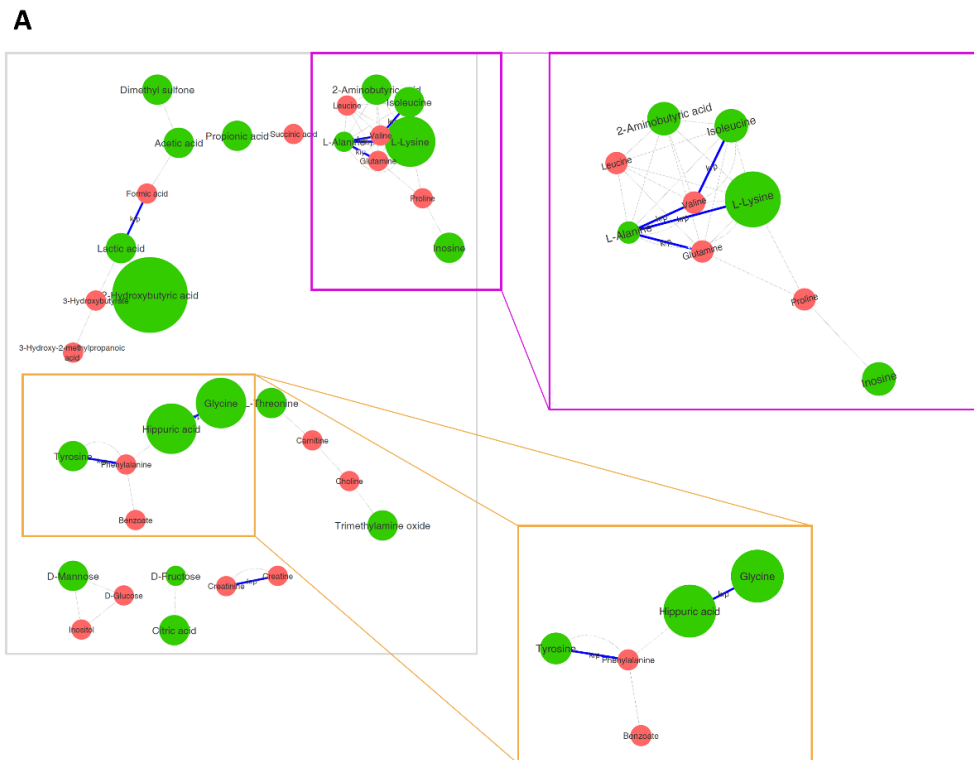
- Basoglu A, Baspinar N, Coskun A. 2014. NMR based metabolomics evaluation in dairy cows with displaced abomasum. *Turkish J Vet Anim Sci*. 38:325–330.
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol*. 57:289–300.
- Caixeta LS, Herman JA, Johnson GW, McArt J. 2018. Herd-level monitoring and prevention of displaced abomasum in dairy cattle. *Vet Clin North Am Food Anim Pract*. 34(1):83–99.
- Ceciliani F, Lecchi C, Urh C, Sauerwein H. 2018. Proteomics and metabolomics characterizing the pathophysiology of adaptive reactions to the metabolic challenges during the transition from late pregnancy to early lactation in dairy cows. *J Proteomics*. 30:92–106.
- Civelek T, Sevinc M, Boydak M, Basoglu A. 2006. Serum apolipoprotein B100 concentrations in dairy cows with left displaced abomasum. *Revue Méd Vét*. 157:361–365.
- Constable P, Grünberg W, Staufienbiel R, Stämpfli HR. 2013. Clinicopathologic variables associated with hypokalemia in lactating dairy cows with abomasal displacement or volvulus. *J Am Vet Med Assoc*. 242(6):826–835.
- Dębski B, Nowicki T, Zalewski W, Ochota M, Mrowiec J, Twardoń J. 2016. Evaluation of acute phase proteins in clinically healthy dairy cows in perinatal period and during lactation. *Pol J Vet Sci*. 19(3):519–523.
- Dieterle F, Ross A, Schlotterbeck G, Senn H. 2006. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in ¹H NMR metabonomics. *Anal Chem*. 78(13):4281–4290.
- Doll K. 2015. Abomasal displacement in dairy cattle: A hereditary disease? *Vet J*. 205(3):329–330.
- Garcia M, Mamedova LK, Barton B, Bradford BJ. 2018. Choline regulates the function of bovine immune cells and alters the mRNA abundance of enzymes and receptors involved in its metabolism in vitro. *Front Immunol*. 25:2448.
- Gerspach C, Imhasly S, Gubler M, Naegeli H, Ruetten M, Laczko E. 2017. Altered plasma lipidome profile of dairy cows with fatty liver disease. *Res Vet Sci*. 110:47–59.
- Guard C. 1990. Abomasal displacement and volvulus. In: Smith BP, editor. *Large animal internal medicine, diseases of horses, cattle, sheep, and goats*. St. Louis, Missouri: The C.V. Mosby Company; p. 792.
- Guzelbektes H, Sen I, Ok M, Constable PD, Boydak M, Coskun A. 2010. Serum amyloid A and haptoglobin concentrations and liver fat percentage in lactating dairy cows with abomasal displacement. *J Vet Intern Med*. 24(1):213–219.
- Hamana M, Ohtsuka H, Oikawa M, Kawamura S. 2010. Blood free amino acids in the postpartum dairy cattle with left displaced abomasum. *J Vet Med Sci*. 72(10):1355–1358.
- Herdt TH, Goeders L, Liesman JS, Emery RS. 1983. Test for the estimation of bovine hepatic lipid content. *J Am Vet Med Assoc*. 182(9):953–955.
- Holmes E, Foxall P.J.D, Nicholson J.K, Neild G.H, Brown S.M, Beddell C.R, Sweatman B.C, Rahr E, Lindon J.C, Spraul M, et al. 1994. Automatic data reduction and pattern recognition methods for the analysis of ¹H nuclear magnetic resonance spectra of human urine from normal and pathological states. *Anal Biochem*. 220(2):284–296.
- Humer E, Khol-Parisini A, Metzler-Zebeli BU, Gruber L, Zebeli Q. 2016. Alterations of the lipid metabolome in

- dairy cows experiencing excessive lipolysis early postpartum. *PLoS One*. 11(7):e0158633.
- Imhasly S, Bieli C, Naegeli H, Nyström L, Ruetten M, Gerspach C. 2015. Blood plasma lipidome profile of dairy cows during the transition period. *BMC Vet Res*. 7: 252.
- Ihaka R, Gentleman RR. 1996. A Language for data analysis and graphics. *J Comput Stat Graph*. 5:299–314.
- Ingvartsen KL. 2006. Feeding- and management-related diseases in the transition cow. Physiological adaptations around calving and strategies to reduce feeding-related diseases. *Anim Feed Sci Technol*. 126(3-4):175–213.
- Kenéz Á, Dánicke S, Rolle-Kampczyk U, von Bergen M, Huber K. 2016. A metabolomics approach to characterize phenotypes of metabolic transition from late pregnancy to early lactation in dairy cows. *Metabolomics*. 12(11):165.
- LeBlanc S. 2010. Monitoring the metabolic health of dairy cattle in the transition period. *J Reprod Dev*. 56:29–35.
- Marczuk J, Brodzki P, Brodzki A, Kurek Ł. 2018. The concentration of free amino acids in blood serum of dairy cows with primary ketosis. *Pol J Vet Sci*. 21(1):149–156.
- McArt JA, Nydam DV, Overton MW. 2015. Hyperketonemia in early lactation dairy cattle: a deterministic estimate of component and total cost per case. *J Dairy Sci*. 98(3): 2043–2054.
- Meiboom S, Gill D. 1958. Modified spin-echo method for measuring nuclear relaxation times. *Rev Sci Instrum*. 29: 68–691.
- NAHMS-USDA. 2017 National Animal Health Monitoring System: dairy cattle management practices in the United States. In: Agriculture UDO, ed. [accessed on 2017 May 22]. https://www.aphis.usda.gov/aphis/ourfocus/animalhealth/monitoring-and-surveillance/nahms/nahms_dairy_studies2007.
- Neuhäuser M. 2011. Wilcoxon–Mann–Whitney Test. In *International Encyclopedia of statistical science* 1656–58. Berlin: Springer.
- Oetzel GR. 2007. Herd-level ketosis – diagnosis and risk factors. *American Association of Dairy Sci*. 96, 2925–38. *Bovine Practitioners*. 40th Annual Conference; Vancouver, BC, Canada.
- Pero RW. 2010. Health consequences of catabolic synthesis of hippuric acid in humans. *Curr Clin Pharmacol*. 5(1): 67–73.
- Puppel K, Kuczyńska B. 2016. Metabolic profiles of cow's blood; a review. *J Sci Food Agric*. 96(13):4321–4328.
- Sadri H, von Soosten D, Meyer U, Kluess J, Dánicke S, Saremi B, Sauerwein H. 2017. Plasma amino acids and metabolic profiling of dairy cows in response to a bolus duodenal infusion of leucine. *PLoS One*. 12(4):e0176647.
- Sato H. 2009. Increased blood concentration of isopropanol in ketotic dairy cows and isopropanol production from acetone in the rumen. *Anim Sci J*. 80(4):381–386.
- Schlegel G, Keller J, Hirche F, Geiszler S, Schwarz F J, Ringseis R, Stangl G I, Eder K. 2012. Expression of genes involved in hepatic carnitine synthesis and uptake in dairy cows in the transition period and at different stages of lactation. *BMC Vet Res*. 8(1):28.
- Seifi HA, Leblanc SJ, Leslie KE, Duffield TF. 2011. Metabolic predictors of post-partum disease and culling risk in dairy cattle. *Vet J*. 188(2):216–220.
- Sen I, Ok M, Coskun A. 2006. The level of serum ionized calcium, aspartate aminotransferase, insulin, glucose, betahydroxybutyrate concentrations and blood gas parameters in cows with left displacement of abomasums. *Pol J Vet Sci*. 9:227–232.
- Serkova N, Fuller TF, Klawitter J, Freise CE, Niemann CU. 2005. H-NMR-based metabolic signatures of mild and severe ischemia/reperfusion injury in rat kidney transplants. *Kidney Int*. 67(3):1142–1151.
- Sevinc M, Ok M, Basoglu A. 2002. Liver function in dairy cows with abomasal displacement. *Revue Méd Vét*. 153: 477–480.
- Shibano K, Kawamura S, Hakamada R, Kawamura Y. 2005. The relationship between changes in serum glycine and alanine concentrations in non-essential amino acid and milk production in the transition period in dairy cows. *J Vet Med Sci*. 67(2):191–193.
- Sickingler M. 2017. [Abomasal displacement in cattle - short overview of recent research results]. *Tierarztl Prax Ausg G Grosstiere Nutztiere*. 45(3):187–190.
- Stanley C A, Hale D E, Berry G T, Deleeuw S, Boxer J, Bonnefont J-P. 1992. A deficiency of carnitine-acylcarnitine translocase in the inner mitochondrial membrane. *N Engl J Med*. 327(1):19–23.
- Sun HZ, Wang B, Wang JK, Liu HY, Liu JX. 2016. Biomarker and pathway analyses of urine metabolomics in dairy cows when corn Stover replaces alfalfa hay. *J Anim Sci Biotechnol*. 7(1):49.
- Sun LW, Zhang HY, Wu L, Shu S, Xia C, Xu C, Zheng JS. 2014. (1)H-Nuclear magnetic resonance-based plasma metabolic profiling of dairy cows with clinical and sub-clinical ketosis. *J Dairy Sci*. 97(3):1552–1562.
- Sun Y, Xu C, Li C, Xia C, Xu C, Wu L, Zhang H. 2014. Characterization of the serum metabolic profile of dairy cows with milk fever using 1H-NMR spectroscopy. *Vet Q*. 34(3):159–163.
- Suthar VS, Canelas-Raposo J, Deniz A, Heuwieser W. 2013. Prevalence of subclinical ketosis and relationships with postpartum diseases in European dairy cows. *J Dairy Sci*. 96(5):2925–2938.
- Takis PG, Ghini V, Tenori L, PaolaTurano P, Luchinat C. 2018. Uniqueness of the NMR approach to metabolomics. *Trends Anal Chem*. 120:115300.
- Trygg J, Wold S. 2002. Orthogonal projections to latent structures (O-PLS). *J Chemometrics*. 16(3):119–128.
- Vignoli A, Ghini V, Meoni G, Licari C, Takis PG, Tenori L, Turano P, Luchinat C. 2019. High-throughput metabolomics by 1D NMR. *Angew Chem Int Ed Engl*. 58:968–994.
- Wold S. 1998. PLS in chemistry. In *Encyclopedia of computational chemistry*. Chichester: Wiley.
- Xu C, Sun LW, Xia C, Zhang HY, Zheng JS, Wang JS. 2016. (1)H-Nuclear Magnetic Resonance-Based Plasma Metabolic Profiling of Dairy Cows with Fatty Liver. *Asian-Australas J Anim Sci*. 29(2):219–229.
- Wu X, Sun H, Xue M, Wang D, Guan LL, Liu J. 2018. Serum metabolome profiling revealed potential biomarkers for milk protein yield in dairy cows. *J Proteomics*. 184: 54–61.
- Zerbin I, Lehner S, Distl O. 2015. Genetics of bovine abomasal displacement. *Vet J*. 204(1):17–22.

Supplementary Material

Supplemental Fig. S1. Biochemical network mapping and related pathway analysis for serum water-soluble metabolites from the comparison between healthy and left displaced abomasum cows. The global network graph of metabolites (on the right-side of panel A of the figure) was obtained using the MetaMapp online tool¹⁰⁷ where green nodes represent metabolites whose concentration is significantly different (adjusted¹⁰⁸ P -value < 0.05) in the comparison, while red nodes represent metabolites whose concentrations are not statistically relevant. Nodes size reflects Fold-Change values. Biochemical and chemical relationships among metabolites are represented by KRP (KEGG Reaction Pairs) and TMSIM (Tanimoto similarity) bold blue edges and black dashed links respectively. In panel B, the MetaboAnalyst¹⁰⁹ pathway mapping is reported only for statistically significant metabolites (adjusted P values < 0.05). In

detail, each dot represents a specific metabolic pathway which is plotted depending on the “pathway impact” and related “ $-\log(P\text{-values})$ ”. The plot highlights as the most significant metabolic pathways for impact and $-\log(P\text{-values})$ the “valine, leucine and isoleucine biosynthesis” and the “phenylalanine, tyrosine and tryptophan biosynthesis” (reported on the right side of the plot) whose respective biochemical networks are enlarged in the magenta and orange boxes depicted on the left-side of panel A of the figure.



Supplemental Fig. S2. Biochemical network mapping and related pathway analysis for serum water-soluble metabolites from the comparison between healthy and right displaced abomasum cows. The global network graph of metabolites (on the left-side

of panel A of the figure) was obtained using the MetaMapp online tool¹⁰⁷ where green nodes represent metabolites whose concentration is significantly different (adjusted¹⁰⁸ P -value < 0.05) in the comparison, while red nodes represent metabolites whose concentrations are not statistically relevant. Nodes size reflects Fold-Change values. Biochemical and chemical relationships among metabolites are represented by KRP (KEGG Reaction Pairs) and TMSIM (Tanimoto similarity) bold blues edges and black dashed links respectively. In panel B, the MetaboAnalyst¹⁰⁹ pathway mapping is reported only for statistically significant metabolites (adjusted P -values < 0.05). In detail, each dot represents a specific metabolic pathway which is plotted depending on the “pathway impact” and related “ $-\log(P$ -values)”. The plot highlights as the most significant metabolic pathways for impact and $-\log(P$ -values) the “valine, leucine and isoleucine biosynthesis” and the “phenylalanine, tyrosine and tryptophan biosynthesis” (reported on the right side of the plot) whose respective biochemical networks are enlarged in the magenta and orange boxes depicted on the right-side of panel A of the figure.

4.2.2. NMR-based serum metabolomics for monitoring newborn preterm calves' health

Abdullah Basoglu¹, Nuri Baspinar², Cristina Licari³, Leonardo Tenori⁴, Amir Naseri¹

¹Department of Internal Medicine, Faculty of Veterinary Medicine, Selcuk University, Selcuklu, Konya, Turkey;

²Department of Biochemistry, Faculty of Veterinary Medicine, Selcuk University, Selcuklu, Konya, Turkey;

³Magnetic Resonance Center (CERM), University of Florence, Sesto Fiorentino (Florence), Italy;

⁴Interuniversity Consortium for Magnetic Resonance of Metalloproteins (C.I.R.M.M.P.), Sesto Fiorentino (Florence), Italy.

Published

Japanese Journal of Veterinary Research 68(2): 105-116, 2020

Candidate's contributions: acquisition of NMR data, statistical analysis, interpretation of data, writing and review of the manuscript.

NMR based serum metabolomics for monitoring newborn preterm calves' health

Abdullah Basoglu^{1,3)}, Nuri Baspinar²⁾, Cristina Licari³⁾,
Leonardo Tenori⁴⁾ and Amir Naseri¹⁾

¹⁾ Department of Internal Medicine, Faculty of Veterinary Medicine, Selcuk University, Aleaddin Keykubat Campus 42030 Konya, Turkey

²⁾ Department of Biochemistry, Faculty of Veterinary Medicine, Selcuk University, Aleaddin Keykubat Campus 42030 Konya, Turkey

³⁾ Magnetic Resonance Center (CERM), University of Florence, Sesto Fiorentino (FI), Italy

⁴⁾ Interuniversity Consortium for Magnetic Resonance of Metalloproteins (C.I.R.M.M.P.), Sesto Fiorentino (Florence), Italy

Received for publication, June 20, 2019; accepted, March 4, 2020

Abstract

It was aimed to detect the novel future biomarkers using a metabolomics approach in premature calves. Calves born previous to 270 days' pregnancy are at risk, and the earlier the calving is, the higher the risk. More trials are needed in neonatology field as it little known almost the generally metabolic status of preterm neonates. To date, this is the first NMR (nuclear magnetic resonance) based study on serum metabolomics at set intervals in premature calves. Biochemical health profiles and NMR based metabolomic analysis were performed in twenty-five premature dairy calves. The whole animals partly recovered following 72h. Clinical data were compatible with those of premature animals. Increased levels of AST and CPK may be attributed to subclinic trauma at birth. Alterations in metabolites, increases in 3-hydroxybutyrate, citrate, leucine and isoleucine at 48th and 72h; choline, formate, fatty acids and polyunsaturated fatty acids at 72h, and valine at 48h; and decreases in myo-inositol at 48h and 72h were meaningful for monitoring the recovery at a molecular level in premature calves. Metabolomics became an important tool for identification of premature calves' clinical pathology and monitoring therapeutic picture.

Key Words: Premature calf, Omics, Metabolomics, NMR.

Introduction

Although the preterm birth rate of calves is not known in the world, premature calving remains a major participant in neonatal calf mortality and morbidity. While studying biomarkers in premature calves^{30,1} their utility has not been accepted widely, because a stronger signature of varying metabolites, suggestive of

disturbances in nucleotide metabolism, lung surfactants biosynthesis and renal function, along with enhancement of tricarboxylic acid cycle activity, fatty acids oxidation, and oxidative stress are features of premature newborns because a stronger signature of varied metabolites, implicational disturbances in nucleotid metabolism, respiratory organ surfactants biogenesis and excretory organ perform,

* Corresponding author: Abdullah Basoglu, Department of Internal Medicine, Faculty of Veterinary Medicine, Selcuk University, Aleaddin Keykubat Campus 42030 Konya, Turkey

Email: abbasoglu@selcuk.edu.tr, Fon: +90 332 2233578, Fax: +903322410063

doi: 10.14943/jjvr.68.2.105

beside improving of tricarboxylic acid cycle activity, fatty acids reaction, and oxidative stress are options of premature newborns¹¹⁾. Hence, more understanding approximately the metabolic and improvement forms of neonates which can be assist valuable within the change of their clinical administration is required. Metabolomics offers the capability for identifying the variations in metabolite profiles that can be applied for disease discrimination. This is notably necessary for neonatal and pediatric studies particularly for severely sick patient diagnosing and early identification²¹⁾. In previous studies^{3,4} we took advantage of metabolomics to the way better understanding of extreme sepsis in neonatal calves, and bronchopneumonia in calves. Metabolomics may be useful for understanding the longer-term effects of biomarkers of preterm birth¹⁴⁾. In spite of metabolomic studies in human medicine for premature infants, a metabolomic approach is not encountered in premature calves. This study aimed at evaluating the NMR based metabolomics at set intervals in serum of neonatal premature calves.

Materials and Methods

Ethical Committee: The Committee on Animal Research and Ethics of Faculty of Veterinary Medicine at Selcuk University, approved this experimental design (Protocol Number: 2018/18). All therapy protocol carried out following the Committee rules was needed for supporting by the European Commission.

Calves and Prematurity Criteria: Twenty-five neonatal premature calves, gestation length of around 7 to 8 months, were used. General weakness, difficulty or unable to stand, weakening or absence of sucking reflex, soft nails, an incomplete eruption of the incisor teeth from gums, short and silky hair coat, low body weight, dyspnea, and hypothermia made up inclusion criteria. The calves were treated by

oxygen (administered at 10 L/min.) and antibiotic (aerosolized ceftiofur sodium, 1 mg/kg, IM, for five days), vitamins (vitamin A 1.000.000 IU, vitamin E 100 mg) and minerals (Ca 0.5 g, P 0.25 g, Se 1 mg), and critical care (setting and including dextrose, amino acid and lipid emulsion solutions without electrolytes) was performed, as needed.

Blood Sampling: Serum samples were obtained by centrifuging (1000 ×g for 15 min at 4°C) the blood collected at 24, 48, and 72 hours.

Biochemical Analyses: Biochemical parameters that reflect health profile (glucose, cholesterol, triglyceride, total protein, albumin, bilirubin, urea nitrogen, creatinine, enzymes (AST, ALT, ALP, GGT) and minerals (Ca, P, Mg) in serum samples was performed by using commercial kits (Mindray Chemistry Reagents) via an automated analyzer in our clinical laboratory.

Nuclear Magnetic Resonance Analysis: Thawing serum samples on ice and extracting them eliminate macromolecules (e.g., proteins) and establish a fused metabolic profile for water-soluble and lipid-soluble metabolites²⁵⁾. After dissolving water-soluble extracts in 700 µL of ²H₂O and homogenization by vortexing for 1 min., and centrifugation (3000 rpm at 4°C for 15 min), each supernatant (630 µL) was added to 70 µL of potassium phosphate buffer (1.5 M K₂HPO₄, 100% (v/v) ²H₂O, 2 mM NaN₃, 5.8 mM deuterated trimethylsilyl propanoic acid (TMSP); pH 7.4). After stirring, a total of 600 µL from each mixture was transferred into 5 mm NMR tubes. After dissolving lipid extracts in 700 µL of CDCl₃, and homogenization by vortexing for 1 min., an aliquot (600 µL) from each sample was transferred into 5 mm NMR tubes.

Using a Bruker 600 MHz spectrometer operating at 600.13 MHz proton Larmor frequency 1D ¹H-NMR spectra were acquired and equipped with a 5 mm PATXI ¹H-¹³C-¹⁵N and ²H-decoupling probe including a-z axis gradient coil, an automatic tuning and matching (ATM),

and an automatic and refrigerated sample changer. For temperature stabilization at the level of approximately 0.1K at the sample, a BTO 2000 thermocouple was used. After keeping the samples for at least 5 minutes inside the NMR probe head for temperature equilibration at 310K, they were measured.

All experiments are based on the use of a fully automated and standardized 600 MHz platform which is developed ensuring high reproducibility in metabolomic studies (<https://www.bruker.com/products/mr/nmr/avance-ivdr.html>). This platform includes the “B.I. Methods” which comprises all the automatisms necessary to standardize the generation of metabolomic data from biofluids. It also includes efficient quality control procedures ensuring the transferability of spectral information generated. In detail, for each water and lipid-soluble sample, a 1D ^1H -NMR spectrum was acquired using a standard Nuclear Overhauser Effect Spectroscopy pulse sequence (NOESY 1D presat) (noesygppr1d.com; Bruker BioSpin) using 98,304 data points, a spectral width of 18,028 Hz, an acquisition time of 2.7 s, a relaxation delay of 4 s, a mixing time of 0.01 s and 128 scans.

Also, for all the extracts, using a standard spin echo Carr-Purcell-Meiboom-Gill pulse sequence (CPMG) another ^1H -NMR spectrum was obtained¹⁹ (cpmgpr1d.com; Bruker Bio-Spin) with 128 scans, 73,728 data points, a spectral width of 12,019 Hz and an acquisition time of 3.1 s.

Spectral Processing: Before practising Fourier transform, free induction decays were multiplied by an exponential function equivalent to 0.3 Hz line-broadening factor. Transformed spectra were automatically corrected for phase and baseline distortions and calibrated (anomeric glucose doublet at 5.24 ppm for serum water-soluble extracts and chloroform singlet at 7.20 ppm for lipid-soluble samples) using TopSpin 3.2.

For water-soluble samples, each 1D spectrum (0.2 and 10.0 ppm range) was segmented into 0.02 ppm chemical shift bins and the corresponding

spectral areas were integrated using AMIX software (version 3.8.4, Bruker BioSpin). Through the binning technique, the number of total variables is reduced and small shifts in the signals are compensated, making the analysis more robust and reproducible¹⁶. Then, regions between 4.62 and 4.77 ppm containing residual H_2O signals were removed.

Instead of serum lipid-soluble extracts, only regions between 0.2 and 6.70 ppm were considered and due to the presence of various shifts in the NMR signals of lipids, many regions were removed in this range of ppm.

On remaining bins, Probabilistic Quotient Normalization¹² (PQN) was applied before pattern recognition both for water-soluble and lipid-soluble fractions.

Statistical Analysis: To explain whether all data are parametric or nonparametric normality test was performed for biochemical data. For comparing parametric values which calculated as mean \pm SD, ANOVA and Tukey tests were used. Mann-Whitney U test was performed and presented as median (SPSS 21.0) for comparing the nonparametric ones.

Using R an open-source software for statistical analysis of data, all metabolomic analyses were performed¹⁷. On processed data, multivariate analysis and Principal Component Analysis (PCA)²⁷ were used as a first exploratory approach. To perform supervised data reduction and classification, multilevel PLS (M-PLS)²⁸ was employed.

For all classifications, the global accuracy was evaluated utilizing 100 cycles of Monte Carlo cross-validation scheme (MCCV, R scrip in-house developed). To build the model, for this method, 90% of data are randomly chosen as the training set, at each iteration. After testing the remaining 10% of data and establishing sensitivity, specificity, and accuracy for the classification, for deriving a mean discrimination accuracy in each group this procedure was repeated 100 times.

After performing univariate statistical

Table 1. Biochemical profile in premature calves.

Parameters	Hours		
	24	48	72
AST (U/L)	79.60±34.82 ^a	66.76±32.06 ^{ab}	57.92±28.54 ^b
ALP (U/L)	322±339.84	244.48±239.44	217.08±125.65
ALT (U/L)	18.92±13.06	21.64±15.64	21.64±17.95
CPK (U/L)	503 (70/2626) ^a	242 (31/2874) ^{ab}	162 (31/2268) ^b
GGT (U/L)	14 (7/883)	90 (6/842)	74 (8/798)
Glucose (mg/dL)	83.52±38.10	97.04±43.54	82.60±29.40
Protein (g/dL)	4.53±1.16	4.64±1.34	4.56±0.97
Albumin (g/dL)	2.16±0.33	2.29±0.28	2.31±0.24
Total Bilirubin (mg/dL)	1.80±0.74	2.11±1.19	2.01±1.52
BUN (mg/dL)	11.48±5.76	13.28±10.47	14.80±10.04
Creatinine (mg/dL)	1.16±0.67	0.9±0.42	0.84±0.33
Ca (mg/dL)	9.48±1.34	9.48±1.08	9.83±1.13
Mg (mg/dL)	2.10±0.47	2.14±0.45	1.98±0.36
P (mg/dL)	6.34±1.15	6.90±1.32	7.10±0.99

^{a, b, ab} Means with different superscripts in the same row differ significantly ($P < 0.05$).

analysis on ¹H-NMR spectra, normalized with PQN method, well defined and resolved spectral regions related to the different metabolites were particularly assigned by using matching routines of AMIX 3.8.4 (Bruker BioSpin) and published literature data. To obtain concentrations of metabolites in arbitrary units, after integration of the same regions these concentrations were analyzed to determine discriminating metabolites among the groups. When two groups are constituted by subjects that could be compared in pairs (e.g the same individual before and after a specific treatment) a pairwise analysis can be particularly performed. What has just been said, justifies the previous use of multilevel PLS analysis and the pairwise Wilcoxon signed-rank test for the determination of meaningful metabolites in the context of univariate statistical analysis.

After the pairwise Wilcoxon test, false discovery rate correction was particularly used applying the Benjamini-Hochberg method (FDR)⁶, and adjusted P -value < 0.05 was considered statistically significant. Then, changes in metabolites levels between two groups of NMR

spectra at a time were calculated as the log₂ fold change (FC) ratio of the normalized median intensities of the corresponding signals in the spectra of the two groups (fold change).

Results

Partly clinical recovery was observed at the 72h in all the calves. Biochemical parameters between the hours were meaningful, AST and CPK were decreased gradually with statistical significance (Table 1).

¹H-NMR spectra of serum samples have been acquired. For serum water-soluble fractions, five subjects were removed from the statistical analysis because of the bad quality of related spectra and three subjects have been removed in the analysis of serum lipid-soluble samples for the same reasons.

To have an overview of the main differences between the three-time of blood collection, processed 1D spectra including all groups that differ in the collection time of serum sample (after 24h, 48h, and 72h) were analyzed firstly using

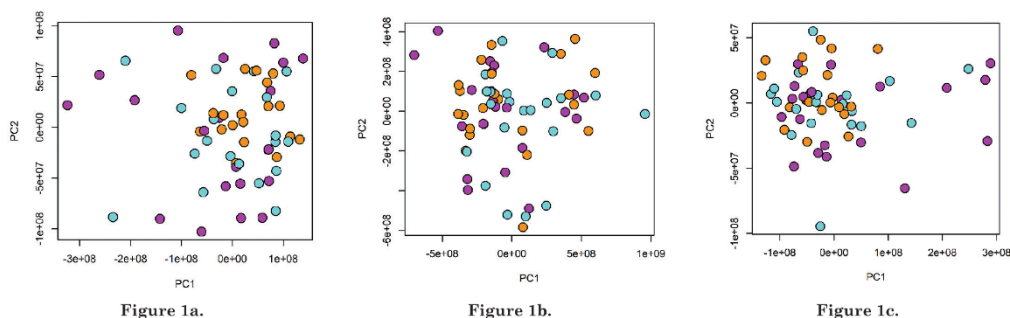


Fig. 1 (a – c). Principal component analysis (PCA) score plots. Each dot represents a single $^1\text{H-NMR}$ spectrum and each colour represents a group of premature calves that differ in the time of blood sample collection: 24h samples (magenta dots), 48h samples (cyan dots) and 72h samples (orange dots). a PCA on 1D NOESY spectra of serum water-soluble samples; b PCA on 1D NOESY spectra of serum lipid-soluble samples; c PCA on 1D CPMG spectra of serum water-soluble samples.

the unsupervised multivariate method (PCA). PCA score plots on 1D NOESY of serum water and lipid-soluble extracts and 1D CPMG of serum water-soluble samples are shown in Fig. 1 (a-c).

To explore changes in the metabolic profile after 24h, 48h and 72h, serum samples were compared in groups of two: first of all, we performed a comparison between samples collected at 24h and 48h, then between samples collected at 48h and 72h and in the end, samples at 24h and 72h have been compared to evaluate global metabolic changes during the entire period of blood collection.

To characterize the within-subject changes presented in the personal metabolic profile the M-PLS approach was used for all comparisons. Between-subject varieties were evaluated and as it were the within-subject varieties were considered with this approach. For serum water-soluble samples, M-PLS models on 1D NOESY spectra and 1D CPMG spectra (Fig. 2(a–f)) were built and a different number of components were retained in the model depending on the type of samples under analysis.

For serum water-soluble extracts, all the built models, as shown by prediction accuracies of cross-validation analyses in Fig. 2(a–f), were able to discriminate premature calves in the different time of blood collection with greater accuracies that 70% and in particular, we obtained higher accuracy when samples at 24h were compared

both with 48h and 72h samples, suggesting a probable change in the metabolic profile in the first phase of blood collection.

Lower prediction accuracies than 70% have been obtained when M-PLS models were built on 1D NOESY spectra from serum lipid-soluble fractions (Fig. 3(a–c)). This can be justified considering that the presence of various shifts in the NMR signals of lipids led to consider, for the statistical analysis, limited portions of the NMR spectra, mainly related to alkyl chains of fatty acids.

To identify discriminating metabolites among the groups, NMR spectra were also analyzed. The complete list of identified and quantified metabolites from each type of sample (lipophilic and hydrophilic fractions) is in Table 2; adjusted P values are reported only for significantly different metabolites (P -value < 0.05) among the various comparisons performed. In particular, for both types of extract, samples collected at 24h were compared with 48h samples and this second group was compared also with 72h samples.

To explore metabolites being significantly different between the beginning and the end of the blood collection, the comparison between serum samples collected from premature calves at 24h and 72h was performed too.

From the analysis of serum water-soluble samples, it appeared that the 24h group showed higher levels of myo-inositol both for

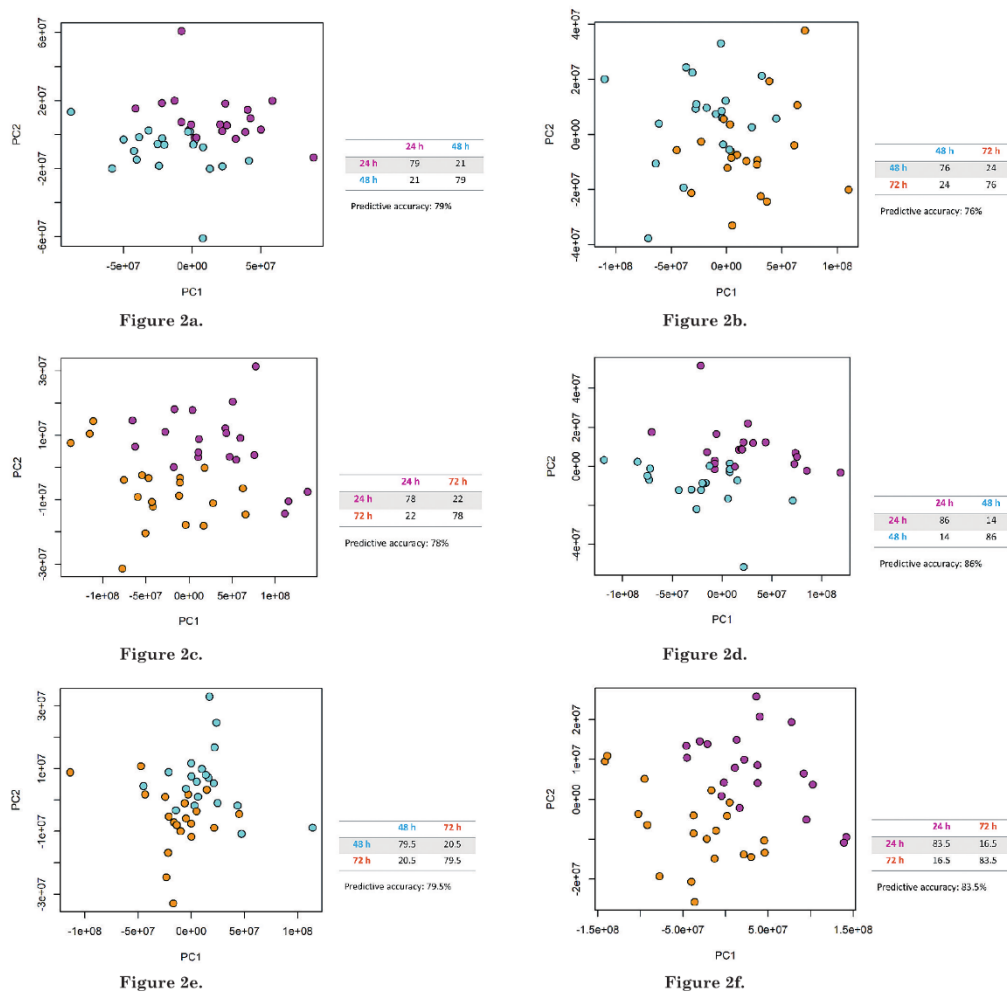


Fig. 2 (a – f). M-PLS score plots of serum water-soluble extracts. Each dot represents a single $^1\text{H-NMR}$ spectrum and each colour represents a group of premature calves that differ in the time of blood sample collection: 24h samples (magenta dots), 48h samples (cyan dots) and 72h samples (orange dots). **a** M-PLS on 1D NOESY spectra comparing 24h vs 48h samples; **b** M-PLS on 1D NOESY spectra comparing 48h vs 72h samples; **c** M-PLS on 1D NOESY spectra comparing 24h vs 72h samples; **d** M-PLS on 1D CPMG spectra comparing 24h vs 48h samples; **e** M-PLS on 1D CPMG spectra comparing 48h vs 72h samples; **f** M-PLS on 1D CPMG spectra comparing 24h vs 72h samples. Confusion matrices and related predictive accuracy of cross-validation analysis are reported for each model.

the comparison with the group of samples collected at 48h and 72h. Instead, higher levels of 3-hydroxybutyrate, citrate, leucine, and isoleucine were reported both for 48h and 72h groups when they were compared to 24h group.

Higher levels of myo-inositol have been reported also in the 48h group when it was com-

pared with serum samples collected at 72h.

From the analysis of serum lipids, it resulted that signals of alkyl chains of fatty acids arising from $\text{CH}_3(\text{CH}_2)_n$, $-\text{CH}_2-\text{CO}$, $=\text{CH}-\text{CH}_2-\text{CH}_2$ protons were higher in the 72h group compared with the one whose samples have been collected at 24h. For the same comparison, also signals arising from

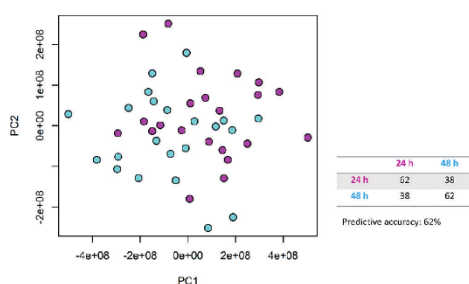


Figure 3a.

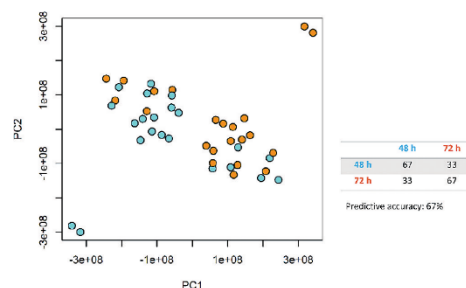


Figure 3b.

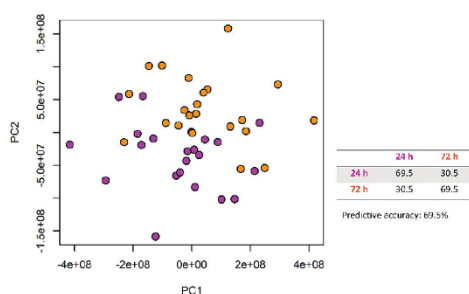


Figure 3c.

Fig. 3 (a – c). M-PLS score plots of serum lipid soluble extracts. Each dot represents a single $^1\text{H-NMR}$ spectrum and each colour represents a group of premature calves that differ in the time of blood sample collection: 24h samples (magenta dots), 48h samples (cyan dots) and 72h samples (orange dots). **a** M-PLS on 1D NOESY spectra comparing 24h vs 48h samples; **b** M-PLS on 1D NOESY spectra comparing 48h vs 72h samples; **c** M-PLS on 1D NOESY spectra comparing 24h vs 72h samples. Confusion matrices and related predictive accuracy of cross-validation analysis are reported for each model.

cholesterol $\text{C}(18)\text{H}_3$ and polyunsaturated fatty acids (18:2) bis allylic protons resulted higher for the 72h group.

Discussion

In premature calves, metabolomics was assessed for the first time. Our data showed that metabolomics was a favorable mean for metabolite identification and meaningful biomarker quantification for indicating the recovery at a molecular level for premature calves. Calves born alive with a gestation length of fewer than 270 days are considered as premature. Infections, poisonings, acute rumen acidosis, fatty liver syndrome, nutritional deficiencies, hereditary factors, induction of premature parturition, situations that cause compulsory slaughtering and procedures applied in that cows exposed to in the last period of pregnancy plays a role in the etiology of premature calving. Diagnosis is based on a short gestation period, clinical and

laboratory findings²⁹. For premature calves with respiratory difficulty, blood lactate and pCO_2 are prognostic³⁰. In this study, clinical data were compatible with most references. Difficult births lead to elevated indicators of subclinical trauma and decreased vigour in the neonate, which can be quantified by measuring serum CPK, AST, and vigour parameters, respectively²⁴. In the present study, however AST changes were between reference ranges CPK levels were decreased gradually. This may be attributed to subclinical trauma at birth, and indicates the recovery with therapy.

Neonatal health care using all the present resources to increase the survival rate of all especially those with critical conditions is carried out for the primary target of neonatology²¹. Treatments of premature calves should be an immediate and powerful treatment and be made for hypothermia, hypoglycemia, metabolic and respiratory acidosis, opportunistic infections, respiratory distress syndrome and failure of passive transfer². In the current study, all the

Table 2. Concentrations in arbitrary units (median \pm Median Absolute Deviation (MAD)) of the metabolites assigned in serum samples (both hydrophilic and lipophilic fractions). Statistically significant adjusted (13) P value < 0.05 from all comparisons are also reported.

Metabolites	24 h (arbitrary units)	48 h (arbitrary units)	72 h (arbitrary units)	P value
1,2-propanediol	143.1 \pm 126.0	195.4 \pm 101.4	203.6 \pm 75.2	
2-hydroxybutyrate	117.4 \pm 42.2	145.0 \pm 39.8	132.8 \pm 50.0	
3-hydroxybutyrate	1013.1 \pm 365.5	1545.3 \pm 445.9	1623.3 \pm 485.9	<0.05 (24h versus 48h) <0.05 (24h versus 72h)
3-hydroxyisobutyrate	225.4 \pm 62.7	249.2 \pm 65.1	264.2 \pm 79.0	
Acetate	6377.6 \pm 2189.4	7315.6 \pm 2210.4	6664.9 \pm 1285.8	
Alanine	3172.8 \pm 856.5	3667.9 \pm 1228.9	3420.9 \pm 776.0	
Carnitine	1700.4 \pm 761.2	1466.5 \pm 604.4	1309.0 \pm 443.8	
Citrate	1303.0 \pm 364.1	1857.6 \pm 815.8	2024.7 \pm 642.8	<0.05 (24h versus 48h) <0.05 (24h versus 72h)
Choline	639.4 \pm 200.7	862.8 \pm 423.1	963.4 \pm 197.3	<0.05 (24h versus 72h)
Creatine	2476.2 \pm 851.6	2354.6 \pm 621.7	2672.4 \pm 1043.8	
Creatinine	861.6 \pm 223.5	824.7 \pm 180.2	790.5 \pm 216.4	
Dimethylsulfone	184.7 \pm 137.1	223.9 \pm 161.0	262.4 \pm 164.4	
Formate	507.2 \pm 183.2	618.9 \pm 151.0	592.3 \pm 135.4	<0.05 (24h versus 72h)
Fructose	1068.0 \pm 699.6	1182.9 \pm 558.9	1021.4 \pm 305.9	
Glycine	942.5 \pm 652.0	1120.7 \pm 369.6	1200.1 \pm 454.6	
Glucose	2736.3 \pm 1349.4	2799.4 \pm 1039.3	2953.0 \pm 739.8	
Hippurate	140.0 \pm 68.4	143.7 \pm 57.2	96.6 \pm 68.6	
inosine	8.3 \pm 8.3	24.2 \pm 10.9	21.4 \pm 14.5	
Isoleucine	201.5 \pm 74.3	280.6 \pm 105.3	339.6 \pm 115.5	<0.05 (24h versus 48h) <0.05 (24h versus 72h)
Lactate	5998.8 \pm 2543.2	5775.0 \pm 1997.9	4996.2 \pm 1276.3	
Leucine	477.1 \pm 155.7	589.4 \pm 178.0	675.6 \pm 167.6	<0.05 (24h versus 48h) <0.05 (24h versus 72h)
Lysine	848.0 \pm 118.4	851.6 \pm 125.2	857.5 \pm 95.1	
Mannose	154.8 \pm 82.3	177.9 \pm 82.5	130.4 \pm 48.3	
Myo-inositol	1911.9 \pm 504.7	967.3 \pm 303.3	619.0 \pm 210.7	<0.05 (24h versus 48h) <0.05 (48h versus 72h) <0.001 (24h versus 72h)
Phenylalanine	415.5 \pm 141.1	482.9 \pm 95.8	602.2 \pm 106.3	
Proline	111.5 \pm 43.9	134.1 \pm 23.8	116.2 \pm 23.5	
Succinate	710.9 \pm 283.5	946.9 \pm 307.4	774.5 \pm 172.6	
Threonine	79.0 \pm 19.7	69.9 \pm 33.2	82.9 \pm 23.5	
Trimethylamine N-oxide	8872.2 \pm 3160.9	7375.3 \pm 1869.8	7167.7 \pm 1631.0	
Tyrosine	362.3 \pm 149.3	414.4 \pm 114.5	492.6 \pm 111.6	
Unknow	181.7 \pm 48.2	183.3 \pm 44.9	154.2 \pm 36.9	
Valine	931.6 \pm 418.8	1169.3 \pm 324.4	1324.0 \pm 393.1	<0.05 (24h versus 48h)
Cholesterol C(18)H ₃₆	19585.5 \pm 7180.7	19353.0 \pm 4525.2	22563.4 \pm 3462.6	<0.05 (24h versus 72h)
Fatty acid-CH=CH-CH ₂ -CH=CH-	11995.0 \pm 4849.9	12106.2 \pm 2319.8	11202.4 \pm 2563.2	
Fatty acid-(CH ₂) _n -	669798.0 \pm 81373.8	684309.2 \pm 43494.3	669677.9 \pm 52028.9	
Fatty acid-CH ₃	34133.1 \pm 12367.9	33866.0 \pm 6847.8	36961.7 \pm 5745.7	<0.05 (24h versus 72h)
Fatty acid-CH ₂ -CO	7580.0 \pm 2945.9	9145.3 \pm 1700.6	8497.5 \pm 2151.2	<0.05 (24h versus 72h)
Fatty acid=CH-CH ₂ -CH ₂	14142.1 \pm 2787.1	15335.8 \pm 1375.5	15840.2 \pm 1907.8	<0.05 (24h versus 72h)
Polysaturated fatty acids (18:2, bis allylic protons)	4382.9 \pm 1081.4	5530.3 \pm 680.4	5872.1 \pm 1079.3	<0.05 (24h versus 72h)
Unsaturated fatty acid -CH=CH-	40993.4 \pm 11224.4	45087.2 \pm 5496.9	44814.6 \pm 9431.5	

calves recovered clinically without monitoring the picture of clinical pathology. Knowledge obtained in this respect would be significant for planning an effective prevention and treatment strategy for the future because the pathogenesis is not yet completely understood⁹. Metabolomics is applying for monitoring of postnatal metabolic maturation,

and for identification of biomarkers as early predictors of outcome and the diagnosis, and the "tailored" management of neonatal disorders¹³. Metabolomics is a promising approach to identify novel specific biomarkers²⁵. Novel roads for biomarkers revelation totally different branches of medicine, containing perinatology are opened

utilizing metabolomics²⁶⁾. In human medicine, metabolomics significantly enhances the relation to 7-day survival in premature infants²²⁾. In this current study, alterations in metabolites, increases in 3- hydroxybutyrate, citrate, leucine and isoleucine at 48th and 72nd hours; choline, formate, fatty acids and poly-unsaturated fatty acids at 72nd hour, and valine at 48th hour; and decreases in myoinositol at 48th and 72nd hours were meaningful for monitoring the recovery at a molecular level in premature calves. The metabolites and their functions in the present study are the following: *citrate* is an intermediate in the krebs cycle, a central metabolic pathway for animals; *leucine*: protein turnover regulation through cellular mammalian target of rapamycin signaling and gene expression; glutamate dehydrogenase activator; branched-chain amino acid balance; flavor enhancer; *isoleucine*: glutamine and alanine synthesis; balance among branched-chain amino acid; *valine*: glutamine and alanine synthesis; balance among branched-chain amino acid²⁹⁾. The 3 main branched-chain amino acids (leucine, isoleucine, and valine) are initially catabolized by common pathways and later diverge into complex pathways⁷⁾. *Formate* is important in embryonic development. Increased formate levels may have a role in fetal development and suggest that extracellular formate may affect the interorgan distribution of one-carbon groups⁸⁾. *Myo-inositol* is available in totally different sorts of cells in free form or is consolidated into inositol phospholipids, additionally other inositol subordinations and in this way plays an imperative administrative part¹⁰⁾, and seems to be involved in lung maturation during antenatal life²⁰⁾ and could be implicated in glucose homeostasis⁸⁾. For fetal development especially of the brain, *choline* is essential¹⁸⁾. In spite of adequate macronutrient supply and weight gain, insufficient choline may contribute to the impeded incline body mass development and respiratory and neurocognitive advancement of preterm newborn children. In this setting a reevaluation of current proposals

for choline supply to preterm newborn children is required⁵⁾. Long-chain PUFAs, similar to choline, are critical for brain improvement¹⁵⁾. Obtained metabolites above in this present study suggest that metabolomics have therapeutic importance in premature calves, and that premature calves need energy, especially amino acid and myo-inositol supplementation. After a premature birth, parenteral amino acid nutrition is needed to be insure sufficient growth and neurodevelopment⁷⁾.

In conclusion this first study showed the metabolomics could indeed be a strong mean providing information on the factors in charge for the metabolic changes in premature calves. The metabolites in the present study may become potential biomarkers for monitoring therapy, however, the predictive and prognostic value extents of this given set of metabolites will be needed for more clinical trials.

Acknowledgements

CERM/CIRMMP center of the ESFRI Instruct is gratefully acknowledged for the NMR access provision financially supported by the EC Contract iNEXT No 653706.

References

- 1) Aydogdu U, Yildiz R, Guzelbektes H, Coskun A, Sen I. Cardiac biomarkers in premature calves with respiratory distress syndrome. *Acta Vet Hung* 64, 38-46, 2016.
- 2) Aytug N, Basbugan Y. Premature calves. *Turkiye Klinikleri J Vet Sci* 4, 53-61, 2013.
- 3) Basoglu A, Baspinar N, Tenori L, Vignoli A, Yildiz R. Plasma metabolomics in calves with acute bronchopneumonia. *Metabolomics* 12, 1-10, 2016.
- 4) Basoglu A, Sen I, Meoni G, Tenori L, Naseri A. NMR based plasma metabolomics at set intervals in newborn dairy calves with severe sepsis. *Mediators of Inflammation* 2018.

- doi:10.1155/2018/8016510.
- 5) Bernhard W, Poets CF, Franz AR. Choline and choline-related nutrients in regular and preterm infant growth. *Eur J Nutr* 8, 1834-1837, 2018.
 - 6) Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol* 289-300, 1995.
 - 7) Blanco C, McGill-Vargas L, Li C, Winter L, Nathanielsz P. High branched-chain amino acid concentrations are found in preterm baboons receiving intravenous amino acid solutions and mimic alterations found in preterm infants. *JPEN* 7, 1507, 2019.
 - 8) Brosnan ME, Brosnan JT. Formate: the neglected member of one-carbon metabolism. *The Annual Review of Nutrition* 17, 369-388, 2016.
 - 9) Capasso L, Vento C, Loddo C, Iavarone F, Raimondi F, Dani C, Fanos V. Oxidative stress and bronchopulmonary dysplasia: evidences from microbiomics, metabolomics, and proteomics. *Frontiers in Pediatrics* 13, 30, 2019.
 - 10) Croze ML, Soulage CO. Potential role and therapeutic interests of myo-inositol in metabolic diseases. *Biochimie* 95, 1811-1827, 2013.
 - 11) Diaz SO, Pinto J, Barros AS, Morais E, Duarte D, Negrão F, Pita C, Almeida Mdo C, Carreira IM, Spraul M, Gil AM. Newborn urinary metabolic signatures of prematurity and other disorders: a case control study. *J Proteome Res* 4, 311-325, 2016.
 - 12) 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, *Analytical Chemistry* 78, 4281-4290, 2006.
 - 13) Fanos V. Pediatric and neonatal individualized medicine. *JPNIM* 1, 7-10, 2018.
 - 14) Gil AM, Duarte D. Biofluid metabolomics in preterm birth research. *Reproductive Sciences* 25, 967-977, 2018.
 - 15) Haggarty P. Fatty acid supply to the human fetus. *Annual Review of Nutrition* 30, 237-255, 2010.
 - 16) Holmes E, Foxall PJ, Nicholson JK, Neild GH, Brown SM, Beddell CR, Sweatman BC, Rahr E, Lindon JC, Spraul M. Automatic data reduction and pattern recognition methods for analysis of ¹H nuclear magnetic resonance spectra of human urine from normal and pathological states. *Analytical Biochemistry* 220, 284-296, 1994.
 - 17) Ihaka R, Gentleman RR. A language for data analysis and graphics. *The Journal of Computational and Graphical Statistics* 5, 299-314, 1996.
 - 18) Maas C, Franz AR, Shunova A, Mathes M, Bleeker C, Poets CF, Schleicher E, Bernhard W. Choline and polyunsaturated fatty acids in preterm infants' maternal milk. *European Journal of Nutrition* 56, 1733-1742, 2017.
 - 19) Meiboom S, Gill D. Modified spin - echo method for measuring nuclear relaxation times. *Rev Sci Instrum* 29, 688-691, 1958.
 - 20) Mimmi MC, Ballico M, Nakib G, Calcaterra V, Peiro JL, Marotta M, Pelizzo G. Altered metabolic profile in congenital lung lesions revealed by ¹H nuclear magnetic resonance spectroscopy. *Analytical Chemistry* 391836, 2014.
 - 21) Noto A, Fanos V, Dessì A. Metabolomics in newborns. *Adv Clin Chem* 74, 35-61, 2016.
 - 22) Oltman SP, Rogers EE, Baer RJ, Anderson JG, Steurer MA, Pantell MS, Partridge JC, Rand L, Ryckman KK, Jelliffe-Pawłowski LL. Initial metabolic profiles are associated with 7-day survival among infants born at 22-25 weeks of gestation. *J Pediatr* 198, 194-200, 2018.
 - 23) Parfieniuk E, Zbucka-Kretowska M, Ciborowski M, Kretowski A, Barbas C. Untargeted metabolomics: an overview of its usefulness and future potential in prenatal diagnosis. *Expert Rev Proteomics* 15, 809-

- 816, 2018.
- 24) Pearson JM, Homerosky ER, Caulkett NA, Campbell JR, Levy M, Pajor EA, Windeyer MC. Quantifying subclinical trauma associated with calving difficulty, vigour, and passive immunity in newborn beef calves. *Vet Rec Open* 6, e000325, 2019.
 - 25) Piersigilli F, Lam TT, Vernocchi P, Quagliariello A, Putignani L, Aghai ZH, Bhandari V. Identification of new biomarkers of bronchopulmonary dysplasia using metabolomics. *Metabolomics* 15, 2, 20, 2019.
 - 26) Stringer KA, Serkova NJ, Karnovsky A, Guire K, Paine R, Standifor TJ. Metabolic consequences of sepsis induced acute lung injury revealed by plasma ¹H-nuclear magnetic resonance quantitative metabolomics and computational analysis. *Am J Physiol Lung Cell Mol Physiol* 300, L4–L11, 2011.
 - 27) Vignoli A, Ghini V, Meoni G, Licari C, Takis PG, Tenori L, Turano P, Luchinat C. High-throughput metabolomics by 1D NMR. *Angew Chem Int Ed Engl* 58, 968–994, 2019.
 - 28) Westerhuis JA, van Velzen EJ, Hoefsloo HC, Smilde AK. Multivariate paired data analysis: multilevel PLSDA versus OPLSDA. *Metabolomics* 6, 119–128, 2010.
 - 29) Wu G. Amino acids: metabolism, functions, and nutrition. *Amino Acids* 37, 1–17, 2009.
 - 30) Yildiz R, Aydogdu U, Guzelbektes H, Coskun A, Sen I. Venous lactate, pH and partial pressure of carbondioxide levels as prognostic indicators in 110 premature calves with respiratory distress syndrome. *Vet Rec* 180, 611, 2017.

4.2.3. NMR-based serum extracts metabolomics to evaluate Canine Ehrlichiosis

Abdullah Basoglu¹, Kursad Turgut², Nuri Baspinar³, Leonardo Tenori⁴, Cristina Licari⁵, M. Ege Ince², Merve Ertan², Havva Suleymanoglu², Serkan Sayiner⁶

¹Department of Internal Medicine, Faculty of Veterinary Medicine, Selcuk University, Selcuklu, Konya, Turkey;

²Department of Internal Medicine, Faculty of Veterinary Medicine, Near East University, Nicosia, TRNC;

³Department of Biochemistry, Faculty of Veterinary Medicine, Selcuk University, Selcuklu, Konya, Turkey;

⁴Interuniversity Consortium for Magnetic Resonance of Metalloproteins (C.I.R.M.M.P.), Sesto Fiorentino (Florence), Italy;

⁵Magnetic Resonance Center (CERM), University of Florence, Sesto Fiorentino (Florence), Italy;

⁶ Department of Biochemistry, Faculty of Veterinary Medicine, Near East University, Nicosia, TRNC.

Accepted for Publication

In the Japanese Journal of Veterinary Research

Candidate's contributions: acquisition of NMR data, statistical analysis, interpretation of data, writing and review of the manuscript.

Abstract

Ehrlichiosis is an infection caused by obligate, intracellular organisms that primarily affect cells of the immune system in dogs, cats and people. The aim of this study was to determine the changes in the serum lipidome profiling of dogs with Canine Ehrlichiosis (*E. Canis*) and try to identify potentially useful metabolic markers. Our study animals included infected (92) and healthy (17) dogs. Indirect fluorescent-antibody assay (IFA) was used for the diagnosis of Ehrlichiosis. Anorexia, depression, hemorrhagic tendencies, enlarge lymph nodes are variable clinical signs of *Ehrlichia*. The hemogram reflected anemia and thrombocytopenia. There were no significant changes in other biochemical parameters. The individually identified metabolites seemed to be not effective in the characterization of the Canine Ehrlichiosis. However, results from the analysis of lipid fractions lead to the hypothesis that considerable differences among diseased and healthy animals could be found in their lipidome instead of the metabolome. This reflects a great systemic energy deficit during the infection.

Introduction

Ehrlichiosis is a tick-borne infection caused by obligate, intracellular organisms that primarily affect cells of the immune system in dogs, cats and humans. In Europe published reports on this infection have increased in recent years. The prevalence of *Ehrlichia* infection in dogs is high and different among European countries.¹ In humans, *Ehrlichia* may cause hematologic malignancies, such as acute leukemia.² The disease can evolve into a severe multisystem disease such as sepsis, meningoenzephalitis or acute respiratory distress syndrome.³ An acute phase of disease, which develop during the subacute phase, starts in 2–4 weeks after exposure to *Erlichia canis*. The chronic phase of the disease can develop in a few dogs. The main clinical symptoms of the disease are weakness, weight loss, and a tendency to bleed. If effective treatment is not applied, infected dogs are most likely to die.^{4,5} Diagnosis of Canine Ehrlichiosis can be difficult to determine through blood smears, serology and even PCR tests; this because of their limitations to deal with the interpretation of test results according to the state of disease. IFA assay is still the diagnostic gold standard protocol for this disease.⁶ A better understanding of the dynamics of Canine Ehrlichiosis is necessary to reach the therapeutic targets. Metabolomics offers an objective approach to determine and to understand the global metabolic pathways of infections. Variations in the metabolome reflect changes in the regulation of biochemical reactions, due to internal/external stimuli. Metabolite profiling can identify changes in host metabolism in response to infections. This will be the first NMR-based metabolomic study for Canine Ehrlichiosis. The goal of this proposal is to obtain a deeper understanding of Canine Ehrlichiosis from a metabolomic point of view.

Materials and Methods

The experimental design was approved by the Committee on Use of Animals in Research of the Near East University.

Animals

The study animals included 92 infected and 17 healthy dogs. The diseased dogs presented some clinical signs, such as a loss of appetite, depression, loss of stamina, stiffness and reluctance to walk, swelling of the limbs or scrotum and coughing or difficulty in breathing.

Laboratory Analysis

Complete blood counts (leukocytes (WBC), erythrocytes (RBC), platelets (PLT), mean cell volume (MCV), mean corpuscular haemoglobin concentration (MCHC), haematocrit (Ht), haemoglobin (Hb)), and biochemical profiles including total protein (TP), albumin (Alb), globulin, cholesterol, triglyceride, alanine amino transaminase (ALT), aspartate amino transaminase (AST), alkaline phosphatase (ALP), lactate dehydrogenase (LDH), blood urea nitrogen (BUN), creatinine and phosphorus were performed by routine automated cell counter and spectrometric methods, respectively. The infection in the animals was diagnosed by indirect fluorescent-antibody assay.

Samples preparation for ¹H-NMR spectroscopy

Serum samples were thawed on ice and extracted for protein precipitation and separation of hydrophilic and lipophilic fractions with a dual methanol-chloroform extraction as described by Stringer *et al.*⁷ As a result of this, macromolecules (*e.g.* proteins) were eliminated and a fused metabolic profile for water-soluble and lipid metabolites was established.

A Bruker 600 MHz spectrometer (with a proton Larmor frequency of 600.13 MHz) was employed to acquire 1D NMR spectra for all analysed samples. The instrument was assembled with a 2H-decoupling probe including a-z axis gradient coil, a 5 mm PATXI 1H-13C-15N, an automatic and refrigerated sample changer and an automatic unit for tuning and matching (ATM). To avoid temperature variations (limited to ± 0.1 K at sample), a BTO 2000 thermocouple was used. Before initiating measurements, NMR tubes were maintained at 310K inside the NMR probe head for at least 5 minutes to equilibrate temperature. A standard Nuclear Overhauser Effect Spectroscopy pulse sequence (noesygppr1d.com; Bruker BioSpin; NOESY 1D) was employed for both aqueous and organic serum extracts, as described in detail in Basoglu *et al.*⁸ For the aqueous fractions, a standard spin echo Carr-Purcell-Meiboom-Gill pulse sequence (cpmgpr1d.comp; Bruker BioSpin; CPMG) was also applied to acquire a second one dimensional experiment using the same parameters described earlier.⁸

Spectral Processing

Before carrying out Fourier transform, free induction decays were multiplied by an exponential function equivalent to 0.3 Hz line-broadening factor. Obtained spectra were automatically corrected for phase and baseline distortions and calibrated using the software TopSpin 3.2 (Bruker BioSpin), considering as references, the anomeric glucose doublet signal at 5.24 ppm for serum aqueous extracts and the chloroform singlet at 7.24 ppm for the organic fractions. One-dimensional spectra from aqueous extracts, in the range of 0.2-10.0 ppm, were divided into 0.02 ppm chemical shifts buckets and their corresponding spectral areas were integrated using AMIX software (version 3.8.4, Bruker BioSpin). Using the bucketing techniques, the global number of variables is decreased and small shifts in the signals can be compensated allowing more reproducible and more robust statistical analysis.⁹ Both regions between 4.62 and 4.75 ppm containing residual H₂O signal and bins related to signals present only for a restricted number of samples were excluded for the multivariate statistical analyses. Instead for organic fractions, only bins between 0.2 and 6.70 ppm were considered for the analysis, due to the presence of various shifts in the aromatic NMR signals of lipid molecules. On remaining bins, Probabilistic Quotient Normalization¹⁰ (PQN) was applied before performing the pattern recognition both for aqueous and organic extracts.

Statistical Analysis

In order to determine the normality for clinical and haematological parameters, the Kolmogorov-Smirnov test was used, while for comparison between groups, the Mann-Whitney-U test was performed for non-parametric cases and the Independent-T test was used for parametric ones.

All metabolomic analysis were done using R (version 3.5), an open source software for statistical manipulation of data.¹¹ Multivariate analysis was applied on processed NMR data and to preliminarily explore the dataset, Principal Component Analysis (PCA) technique was employed.¹² When unsupervised approach was not able to discriminate between the conditions of interest, Orthogonal Projection to Latent Structures (OPLS) analysis is applied in combination with Discriminant Analysis (DA) as a supervised method. In general, this algorithm uses information in the categorical response **Y** matrix to separate, in the **X** matrix of data, the predictive from non-predictive (**Y**-orthogonal) variation, providing a better model interpretation with respect to PCA or to the PLS techniques.¹³ All the accuracies and confusion matrices reported for the various classifications were assessed by means of 100 cycles of Monte Carlo cross-validation scheme (MCCV, R script in-house written). Further explanations are reported elsewhere.⁸

Univariate statistical analysis was performed on ¹H-NMR spectra. In particular, well defined and resolved spectral regions associated to the different metabolites/lipid fractions were assigned by using matching routines of Assure NMR (Bruker BioSpin) and published literature data. The same regions were integrated to obtain concentrations of metabolites and lipidic fractions in arbitrary units. Resulting values

were analysed to find discriminating metabolites between the diseased and healthy dogs using a non-parametric Wilcoxon-Mann-Whitney test¹⁴ on the biological assumption that metabolites and lipids concentrations are not normally distributed.

When several metabolites/lipids are tested together, to avoid random false positives, multiple testing corrections need to be adopted; here, the Benjamini-Hochberg method (FDR)¹⁵ was applied. Then, changes in metabolites/lipids levels between the two compared groups are calculated as the \log_2 fold change (FC) ratio of the normalized median intensities of the corresponding signals in the spectra.

Commonly, to express correlation among different metabolites expression levels and clinical and/or other biological data, correlation coefficients must be calculated. Spearman's test was used to express the correlation coefficients (ρ) among the metabolite concentrations of diseased animals and the clinical features. Values between +1 and -1 are reported, where +1 indicates a total positive correlation, 0 means no correlation, and -1 is related to a total negative correlation. All correlation coefficients were calculated using the "cor.test" function of R software. Graphical representations of correlation matrices of metabolites and lipid fractions were displayed using the R "corrplot" package.

Results

Anorexia, depression, haemorrhagic tendencies, large lymph nodes were variable clinical symptoms of *Ehrlichia*. The hemogram reflected anaemia and thrombocytopenia. There were no significant changes in biochemical parameters (**Table 1**). Serum indirect fluorescent antibody titers were positive in all the diseased animals.

One-dimensional ¹H-NMR spectra of serum samples were acquired. Eleven diseased animals were not considered for the statistical analysis of serum aqueous extracts because of the inadequate quality of the related NMR spectra, while a total of thirty-two dogs were removed from the analysis of organic fractions for same reasons.

Diseased animals were compared with healthy ones using firstly, an unsupervised approach, *i.e.* the PCA analysis. No apparent differences or clusters were highlighted among the groups, but from the analysis of 1D NOESY and 1D CPMG spectra of aqueous extracts, one and three subjects were respectively identified as evident outliers and therefore eliminated for the subsequent analyses. PCA score plots on 1D NOESY and 1D CPMG of serum aqueous fractions and on 1D NOESY of serum organic extracts are reported in **Figure 1 (a-c)**.

To explore differences in the metabolic profile of diseased and healthy animals, a Monte Carlo cross-validated OPLS-DA model was built for both aqueous and organic extracts, randomly sampling among diseased group a comparable number of samples to the healthy group. Models built on bucketed 1D NOESY and 1D CPMG spectra of aqueous fractions provides average accuracies of 68% and 72% respectively, suggesting the presence of slight metabolic differences between healthy and diseased animals. Instead, an average discrimination accuracy of 76% between the two groups of subjects was defined considering the bucketed 1D NOESY spectra of organic

fractions, leading to the hypothesis that this disease affects more the lipidome of the dogs.

In order to identify the presence of discriminating metabolites among the two groups of interest, $^1\text{H-NMR}$ spectra were also examined. The complete list of identified and quantified metabolites from each type of samples (lipophilic and hydrophilic fractions) is presented in **Table 2**, where adjusted P -values are reported only for variables that differ significantly (adjusted P -value < 0.05) after performing the comparison using the Wilcoxon-Mann-Whitney test. Among the eighteen identified metabolites, none of them resulted in being statistically different between diseased and healthy dogs (all adjusted P -values > 0.05). From the univariate analysis of serum lipids, results showed that signals of alkyl chains of fatty acids arising from unsaturated fatty acid protons -CH=CH and -CH₂-CO, -CH=CH-CH₂-CH-CH protons, were higher and statistically different (adjusted P -values < 0.05) for healthy animals, while glycerol resulted in being significantly higher for the diseased dogs (adjusted P -value < 0.05). These last results are in line with the fact that OPLS-DA model built on bucketed 1D NOESY of organic fractions showed a higher predictive accuracy than that reported for statistical models of aqueous extracts.

In summary, the obtained results demonstrate that we can define a likely fingerprint of the disease considering the whole metabolic profile of animals (NMR spectra in their entirety), since the individually identified metabolites seem to be not effective in the characterization of the Canine Ehrlichiosis. However, the existence of four statistically significant lipid fractions leads to the hypothesis that considerable differences among diseased and healthy animals could be found in their lipidome instead of the metabolome (**Table 2**).

Subsequently, correlations among identified metabolites/lipid fractions and clinical data were expressed for the diseased animals. Spearman's test was used to calculate the correlation coefficients (ρ) between metabolomic data and blood parameters. In particular, it was found that albumin positively correlate with proton signals of alkyl chains from fatty acids arising from =CH-CH₂-CH₂, UF=CH-CH groups and from PUFA ($\rho > 0.60$), while tyrosine positively correlates with ALT values ($\rho = 0.60$). Instead, globulins negatively correlate with all the identified lipid fractions except for glycerol and cholesteryl ester fractions ($\rho < -0.60$). Negative correlations were also identified for total proteins values with phospholipids-N(CH₃)₃ and unsaturated -CH=CH- protons. All calculated correlation coefficients (ρ) are reported in **Table 3**. Graphical representations of correlation matrices of metabolites and lipid fractions are reported in **Figure 2 (a-b)** where only statistically significant correlations (P -values < 0.05) are highlighted and represented through glyphs that are coloured according to the calculated ρ coefficients and whose respective values (from -1 to +1) are shown on the gradient coloured bar located in the right-side of the figures.

Discussion

This is the pioneer metabolomic and lipidomic study carried out on dogs suffering from Ehrlichiosis. Identified lipidic profile in this study will give an opportunity for further mechanistic studies to better understand the host responses in *Ehrlichia* infection. Ehrlichiosis is a bacterial illness that affects humans and animals causing flu-like symptoms. Many subjects have mild symptoms and never seek medical attention. However, life-threatening cases of ehrlichiosis manifest as meningoencephalitis or acute respiratory distress syndrome together with sepsis.³ *Ehrlichia* can also lead to hematologic malignancies.² Most often, a diagnosis is performed by a combination of clinical signs, positive serum IFA titer and response to treatment.¹⁶ In the current study, either clinical signs or haematological parameters of seropositive dogs were in accordance with most references (**Table 1**). Metabolites with extraordinary array of physicochemical properties, produced by microbial and host cells, may be found in any body tissue or fluid at various concentrations. Analytical determination from the host-response to bacterial infection, by -omic sciences, could provide new insights for the comprehension of this pathology. Metabolomics have a great potential to determine new biomarkers of diseases, useful to identify for example, early stage diseases, therefore potentially addressing an important clinical need. Performing new studies on metabolomics, low-cost biomarkers from body fluids that indicate infection, therapeutic efficacy, or drug resistance might be identified. In particular, with advantages of minimal sample preparation, high throughput, high reproducibility and high accuracy, metabolomic analyses provide great potentialities for diseases diagnosis and treatment with respect to classical clinic tests. Moreover, with respect to traditional techniques used to explore biomarker profiles, metabolomics offers complete information on low- and high-molecular-weight metabolites present in biofluids. Therefore, metabolomics gives the possibility to generate innovative and non-invasive diagnostic tests providing a unique insight into already known and novel metabolic pathways, which are simple and cost-effective yet retaining high sensitivity and specificity properties.¹⁷ Despite the impact of metabolomics on infectious diseases,^{18,19} no study has been done in regards to Ehrlichiosis infection.

Lipids have important roles in various cellular processes. Changes in the lipidome, in addition of nucleic acid and proteins, can be evaluated as biomarkers. The role of lipids, especially for bacterial infections, is well recognized by the human innate immune response, such as lipopolysaccharide in Gram-negative bacteria, lipoteichoic acid in Gram-positive bacteria, and lipoglycans in mycobacteria. When we compared nucleic acids and protein analysis, a complete analysis of the lipidome is usually very difficult due to the heterogeneity of lipid classes and their intrinsic physical properties caused by variations in the constituents of each class. Therefore, their biological relevance and their use as potential biomarkers for non-infectious and infectious diseases is crucial. Sepsis and tuberculosis are the two primary diseases in which lipids can be diagnosed using biomarkers.²⁰ In septic patients, lipid profiles may be a

predictor of survival. In metabolomic studies, most of the changes from baseline in septic patients are related to lipid metabolism.²¹

NMR-based metabolomics has been evaluated in diarrheic and presumed septic calves, where significant decreases in the whole lipid soluble metabolites such as sphingomyeline and fatty acids including PUFA were found. Other characteristic metabolites, such as increases in niacinamide, choline and phosphocholine, 2-methylglutarate and isopropanol, and decreases in formate, lysine, arginine, acetate, creatine also reflected the systemic inflammatory response syndrome, organ dysfunction and organ failure.²² Moreover, NMR metabolomics provided an optimal tool for faster identification of sepsis in new-born calves.²³

In conclusion, in the present study, among the eighteen metabolites which were identified, none of them resulted to be statistically different between diseased and healthy dogs (**Table 2**), but from our results, it seems that Ehrlichiosis affected more the lipidome of the dogs. Alkyl chains of fatty acids arising from UFA-CH=CH, -CH₂-CO, -CH=CH-CH₂-CH-CH protons were higher and statistically different for healthy animals, while glycerol resulted in being significantly higher in the diseased ones. A decrease in the whole lipid fraction may indicate a great systemic energy deficit occurring in Canine Ehrlichiosis. The correlations of lipid metabolites with blood proteins may be meaningful in this regard (**Table 3**).

Acknowledgments

CERM/CIRMMP center of the ESFRI Instruct is gratefully acknowledged for the NMR access provision financially supported by the EC Contract iNEXT No 653706.

References

- 1) Schwartz C, Katz DA, Larson M, Licciardi N, Kallick C, Timothy M, et al. The relationship between ehrlichiosis and the development of hematologic malignancies. *Medical Hypotheses* 121, 57–59, 2018.
- 2) Sainz Á, Roura X, Miró G, Estrada-Peña A, Kohn B, Harrus S, et al. Guideline for veterinary practitioners on canine ehrlichiosis and anaplasmosis in Europe. *Parasit Vectors* 8,75, 2015
- 3) Buzzard SL, Bissell BD, Melissa L. Bastin T. Ehrlichiosis presenting as severe sepsis and meningoencephalitis in an immunocompetent adult. *JMM Case Rep.* 5, e005162, 2018.
- 4) Chochlios TA, Angelidou E, Kritsepi-Konstantinou M, Koutinas CK, Mylonakis ME. Seroprevalence and risk factors associated with *Ehrlichia canis* in a hospital canine population. *Vet. Clin. Pathol.* 48, 305–309, 2019.
- 5) Mylonakis ME, Harrus S, Breitschwerdt EB. An update on the treatment of canine monocytic ehrlichiosis (*Ehrlichia canis*). *Vet J* 246, 45-53, 2019.

- 6) Bélanger M, Heather L, Sorenson HL, France MK, Bowie MV, Barbet AF, et al. Comparison of serological detection methods for diagnosis of *Ehrlichia canis* infections in dogs. *Journal of Clinical Microbiology* 40, 3506–3508, 2002.
- 7) Stringer KA, Serkova NJ, Karnovsky A, Guire K, Paine R, Standiford TJ. Metabolic consequences of sepsis-induced acute lung injury revealed by plasma ¹H-nuclear magnetic resonance quantitative metabolomics and computational analysis. *Am J Physiol Lung Cell Mol Physiol*. 300, L4–L11, 2011.
- 8) Basoglu A, Baspinar N, Tenori L, Licari C, Gulersoy E. Nuclear magnetic resonance (NMR)-based metabolome profile evaluation in dairy cows with and without displaced abomasum. *Veterinary Quarterly* 40, 1-15, 2020
- 9) Holmes E, Foxall PJ, Nicholson JK, Neild GH, Brown SM, Beddell CR, et al. Automatic data reduction and pattern recognition methods for analysis of ¹H nuclear magnetic resonance spectra of human urine from normal and pathological states. *Anal Biochem* 220, 284–296, 1994.
- 10) 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 metabolomics. *Anal. Chem.* 78, 4281–90, 2006.
- 11) Ihaka R and Gentleman RR. A language for data analysis and graphics. *J Comput Stat Graph* 5, 299–314, 1996.
- 12) Vignoli A, Ghini V, Meoni G, Licari C, Takis PG, Tenori L, et al. High-throughput metabolomics by 1D NMR. *Angew. Chem. Int. Ed Engl.* doi:10.1002/anie.201804736, 2019.
- 13) Westerhuis JA, van Velzen EJ, Hoefsloot HC, Smilde AK. Multivariate paired data analysis: multilevel PLS-DA versus OPLS-DA. *Metabolomics* 6, 119–128, 2010.
- 14) Neuhäuser M. Wilcoxon–Mann–Whitney Test. in *International Encyclopedia of Statistical Science* 1656–1658 (Springer, Berlin, Heidelberg, 2011). doi:10.1007/978-3-642-04898-2_615, 2011.
- 15) Benjamini Y and Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* 289–300, 1995.
- 16) Csokai J, Klas EM, Heusinger A, Müller E. Occurrence of *Ehrlichia canis* in dogs living in Germany and comparison of direct and indirect diagnostic methods. *Tierarztl Prax Ausg K Kleintiere Heimtiere* 45, 301-307, 2017.
- 17) Zhang A, Sun H, Yan G, Wang P, Wang X. Metabolomics for biomarker discovery: Moving to the clinic. *BioMed Research International* 2015, 1-6, 2015.
- 18) Eoh H. Metabolomics: a window into the adaptive physiology of *Mycobacterium tuberculosis*. *Tuberculosis* 94, 538–543, 2014.
- 19) Fuchs TM, Eisenreich W, Heesemann J, Goebel W. Metabolic adaptation of human pathogenic and related nonpathogenic bacteria to extra- and intracellular habitats. *FEMS Microbiol Rev* 36, 435–462, 2012.
- 20) Larrouy-Maumus G. Lipids as biomarkers of cancer and bacterial infections. *Curr Med Chem* 26, 1924-32, 2019.

- 21) Green P, Theilla M, Singer P. Current opinion in clinical nutrition and metabolic care. 19, 111–115, 2016.
- 22) Basoglu, Baspinar N, Tenori L, Hu X, Yildiz R. NMR based metabolomics evaluation in neonatal calves with acute diarrhea and suspected sepsis: a new approach for biomarkers, *Metabolomics Open acces* 2, 134, 2014.
- 23) Basoglu A, Sen I, Meoni G, Tenori L, Naeari A. NMR based plasma metabolomics at set intervals in newborn dairy calves with severe sepsis. *Mediators of Inflammation* 2018; doi:10.1155/2018/8016510, 2018.

Tables**Table 1.** Hematological and biochemical parameters between the animal groups.

	Diseased (n=92)	Healthy (n=17)	P value<0.05
WBCs (x10 ³ /μL)	9.9 (0.2/5390)	12.9 (7.70/38.2)	0,165
RBCs (x10 ⁶ /μL)	4.976±1.583	6.669±8.819	0,03
Platelets (x10 ³ /μL)	93 (0/1102)	277 (107/441)	0,001
Hgb (g/dL)	11 (2.2/18.6)	15.8 (6.2/18.2)	0,01
PCV (%)	33.436±11.643	48.318±10.087	0,01
Total protein (g/dL)	7.014±1.926	5.792±0.952	0,09
Albumin (g/dL)	2.058±0.575	2.8±0.372	0
BUN (mg/dL)	25.332±22.538	15.071±3.536	0,386
Urea (mg/dL)	22.69±1.590	47.525±9.326	0,159
Phosphorus (mg/dL)	4.06 (2.25/15.64)	4.4 (3.49/513)	0,781
Creatinine (mg/dL)	0.640 (0.18/3.86)	1.15 (0.44/1.15)	0,043
Triglyceride (mg/dL)	87.943 (41/115.940)	70.481 (8/21.405)	0,905
Cholesterol (mg/dL)	198.050±76.350	223.663±72.923	0,388
ALP (U/L)	115.35 (13.10/2461.65)	46.198 (12/277.04)	0,153
ALT (U/L)	61.385 (15/1472.24)	65.495 (39/100.17)	0,929
LDH (U/L)	212.791±151.811	142.828±100.768	0,124

WBC, PLT, ALP, ALT, BUN, phosphorus, creatinin, triglyceride = nonparametric (median (min/max));
RBC, Hgb, PCV, albumin, globulin, LDH, total protein, urea, cholesterol = parametric (mean±std).

Table 2. Concentrations in arbitrary units (median ± Median Absolute Deviation (MAD)) of metabolites and lipid fractions assigned in serum samples (both aqueous and organic extracts). Statistically significant adjusted *P*-value < 0.05 are also reported.

Metabolites		Diseased (n=92) (arbitrary units)	Healthy (n=17) (arbitrary units)	Adjusted P value
Serum aqueous fraction (SWS)	3-hydroxybutyrate	0.02 ± 0.01	0.02 ± 0.00	
	3-hydroxyisobutyrate	0.01 ± 0.002	0.01 ± 0.002	
	Acetate	0.385 ± 0.065	0.37 ± 0.08	
	Alanine	0.1 ± 0.03	0.12 ± 0.03	
	Citrate	0.07 ± 0.02	0.08 ± 0.02	
	Choline	0.01 ± 0.00	0.02 ± 0.01	
	Creatine	0.01 ± 0.005	0.007 ± 0.002	
	Creatinine	0.03 ± 0.01	0.04 ± 0.01	
	formate	0.07 ± 0.01	0.06 ± 0.01	
	Glycine	0.04 ± 0.01	0.06 ± 0.015	
	D-Glucose	0.18 ± 0.06	0.17 ± 0.065	
	Isoleucine	0.02 ± 0.01	0.02 ± 0.005	
	Lactate	1.015 ± 0.22	1.08 ± 0.255	
	Leucine	0.05 ± 0.02	0.05 ± 0.01	
	Phenylalanine	0.03 ± 0.02	0.04 ± 0.015	
Pyruvate	0.06 ± 0.01	0.08 ± 0.02		

	Tyrosine	0.02 ± 0.01	0.02 ± 0.01	
	Valine	0.06 ± 0.02	0.07 ± 0.02	
Serum organic fraction (SLS)	Cholesterol C(18) H ₃	0.011 ± 0.0033	0.0135 ± 0.002	
	Cholesteryl ester	0.0016 ± 0.0005	0.048 ± 0.008	
	Fatty acid -CH=CH- CH ₂ -CH=CH-	0.0065 ± 0.002	0.011 ± 0.0032	<0.05 (diseased versus healthy)
	Fatty acid -(CH ₂) _n -	0.19 ± 0.05	0.245 ± 0.02	
	Fatty acid -CH ₃	0.017 ± 0.005	0.022 ± 0.005	
	Fatty acid -CH ₂ -CO	0.0052 ± 0.0015	0.00735 ± 0.0012	<0.05 (diseased versus healthy)
	Fatty acid =CH-CH ₂ - CH ₂	0.016 ± 0.005	0.02 ± 0.002	
	Glycerol backbone	0.00051 ± 0.00017	0.00027 ± 0.000225	<0.05 (diseased versus healthy)
	Polyunsaturated fatty acids (18:2, bis allylic protons)	0.0066 ± 0.0018	0.008 ± 0.0016	
	Unsaturated fatty acid -CH=CH-	0.034 ± 0.009	0.048 ± 0.008	<0.05 (diseased versus healthy)

Table 3. Spearman's rho correlation coefficients between metabolites/lipids concentrations and measured blood parameters of diseased animals. Positive correlation coefficient values higher than 0.60 are colored in light gray, while negative rho coefficients smaller than 0.60 are colored in dark gray.

	WBC	RBC	HGB	HCT	PLT	Albumin	Globulin	ALP	ALT	BUN	P	Creatinine	LDH	T. P	Trigl	Chol
Alanine	0.06	0.14	0.18	0.18	0.07	0.04	-0.17	0.19	0.28	0.10	0.04	-0.01	0.38	-0.15	0.29	-0.07
Valine	0.01	0.01	0.03	0.04	0.14	0.02	-0.10	0.03	0.14	0.33	0.21	0.03	0.25	0.00	0.22	-0.05
Leucine	0.02	0.09	0.08	0.09	0.09	0.12	-0.25	0.05	0.19	0.15	0.12	-0.07	0.22	-0.10	0.16	0.00
L-Isoleucine	0.02	-0.07	-0.06	-0.05	-0.03	0.11	-0.18	0.15	0.08	0.16	0.11	-0.10	0.22	-0.10	0.11	0.08
Lactate	0.26	-0.08	-0.05	-0.05	-0.07	-0.23	0.14	0.30	0.19	0.14	0.07	0.02	0.30	0.08	0.29	0.06
Acetate	0.17	0.05	0.05	0.02	0.14	-0.02	-0.25	0.17	0.53	-0.10	0.11	-0.03	0.29	-0.24	0.04	0.12
Citrate	0.09	0.09	0.18	0.12	0.48	-0.07	0.09	-0.09	0.18	0.11	0.18	0.17	-0.07	0.11	-0.03	0.06
Creatinine	0.31	0.11	0.18	0.16	0.14	-0.07	-0.06	-0.34	-0.03	0.20	0.23	0.58	0.03	-0.08	0.13	0.28
Creatine	0.30	0.09	0.09	0.08	0.03	0.12	-0.42	-0.03	-0.17	0.25	0.06	0.20	0.16	-0.35	0.20	0.34
L-Pyroglutamate	-0.06	0.03	0.08	0.07	0.07	0.19	-0.14	-0.17	0.16	-0.29	-0.07	-0.15	0.07	-0.11	-0.14	-0.17
D-Glucose	0.12	-0.25	-0.34	-0.32	-0.13	-0.35	0.04	0.05	0.07	-0.18	0.06	-0.28	0.21	-0.12	-0.08	-0.10
3-Hydroxybutyrate	-0.10	-0.24	-0.17	-0.20	0.27	-0.12	0.04	-0.10	0.18	0.20	0.27	0.01	-0.12	0.06	0.10	0.02
Choline	0.18	0.06	0.03	0.05	0.17	0.09	-0.33	-0.06	-0.02	0.24	0.15	0.15	0.28	-0.31	0.21	0.13
Glycine	0.12	-0.05	-0.05	-0.07	-0.19	0.07	-0.19	0.26	0.05	-0.11	-0.12	0.02	0.24	-0.25	0.31	0.24
Tyrosine	-0.05	0.28	0.30	0.27	0.01	0.05	-0.12	0.34	0.60	0.05	-0.09	0.03	0.25	-0.12	-0.01	0.15
Phenylalanine	0.11	0.03	0.01	0.00	-0.11	-0.02	-0.22	0.24	0.39	0.02	-0.03	0.06	0.35	-0.24	0.28	0.42
Xanthine	-0.11	-0.36	-0.37	-0.39	-0.27	-0.23	-0.08	0.12	0.32	0.07	0.24	0.07	0.26	-0.24	0.17	-0.05
Formate	0.11	-0.01	-0.02	-0.06	0.15	-0.08	-0.11	0.14	0.26	0.03	0.07	-0.19	0.27	-0.10	0.08	-0.03
Pyruvate	0.06	0.30	0.31	0.31	0.19	0.35	-0.23	-0.07	0.01	-0.45	-0.18	-0.35	0.13	-0.17	-0.13	-0.04
3-Hydroxyisobutyrate	0.13	-0.45	-0.42	-0.45	0.02	-0.34	0.19	0.12	0.17	0.44	0.31	0.20	0.12	0.28	0.25	-0.10
Isobutyrate	-0.06	0.07	0.10	0.07	-0.08	0.06	-0.11	0.03	0.19	-0.12	-0.05	-0.18	-0.01	-0.04	0.06	-0.17
FAs-CH ₃	0.25	0.17	0.15	0.18	0.12	0.53	-0.61	-0.14	-0.22	0.16	-0.02	0.18	0.04	-0.56	0.16	0.70

FAs-(CH ₂) _n -	0.20	0.14	0.10	0.14	0.07	0.58	-0.66	-0.13	-0.27	0.15	-0.05	0.06	0.13	-0.59	0.24	0.59
FAs=CH-CH ₂ -CH ₂	0.22	0.22	0.19	0.22	0.13	0.66	-0.71	-0.14	-0.23	0.14	-0.10	0.06	0.11	-0.59	0.21	0.60
FAs-CH ₂ -CO	0.22	0.26	0.23	0.25	0.21	0.58	-0.64	-0.15	-0.22	0.08	-0.03	0.12	0.05	-0.59	0.18	0.71
PUFA	0.26	0.18	0.16	0.18	0.14	0.65	-0.69	-0.15	-0.22	0.15	-0.10	0.07	0.14	-0.57	0.24	0.61
FAs-CH=CH-CH ₂ - CH=CH	0.04	0.39	0.35	0.39	0.08	0.46	-0.69	-0.04	-0.06	0.19	0.05	0.12	0.08	-0.59	0.04	0.55
Phosp-N(CH ₃) ₃ -	0.17	0.24	0.21	0.25	0.08	0.58	-0.70	-0.13	-0.16	0.12	-0.01	0.11	0.14	-0.62	0.13	0.62
Glycerol	0.11	-0.11	-0.17	-0.13	0.01	0.31	-0.48	-0.03	-0.02	0.43	0.13	0.01	0.22	-0.39	0.68	0.25
Cholesteryl ester	0.25	0.22	0.22	0.24	0.16	0.54	-0.57	-0.14	-0.22	0.13	-0.06	0.16	0.01	-0.52	0.14	0.69
UFA-CH=CH	0.18	0.28	0.24	0.27	0.11	0.62	-0.72	-0.08	-0.18	0.12	-0.08	0.07	0.13	-0.61	0.20	0.64
Chol (C18)H ₃	0.27	0.15	0.13	0.16	0.11	0.49	-0.61	-0.16	-0.17	0.19	0.01	0.22	0.07	-0.57	0.18	0.70

*Abbreviations: Trigl: Triglycerides; Chol: Cholesterol; T. P: Total Protein; FAs: Fatty Acids; PUFA: Polyunsaturated fatty acids; UFA: unsaturated fatty acids; Phosp: Phospholipids

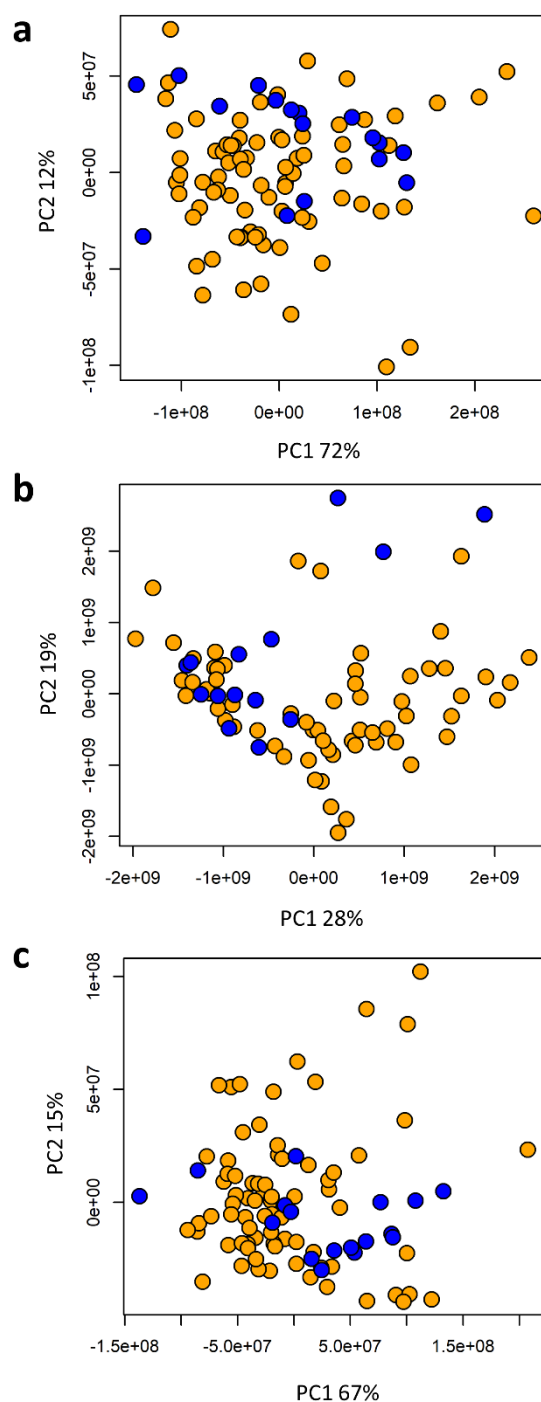
Figures

Figure 1 (a – c). Principal component analysis (PCA) score plots. Each dot represents a single $^1\text{H-NMR}$ spectrum. Diseased animals ($n=92$) are represented in orange, while the healthy ($n=17$) in blue. The variance related to the first principal component (PC1) is reported in the X-axis, while the variance related to the second principal component is reported in the Y-axis. **a)** PCA on bucketed 1D NOESY spectra of serum aqueous fractions; **b)** PCA on bucketed 1D NOESY spectra of serum organic fractions; **c)** PCA on bucketed 1D CPMG spectra of serum aqueous extracts.

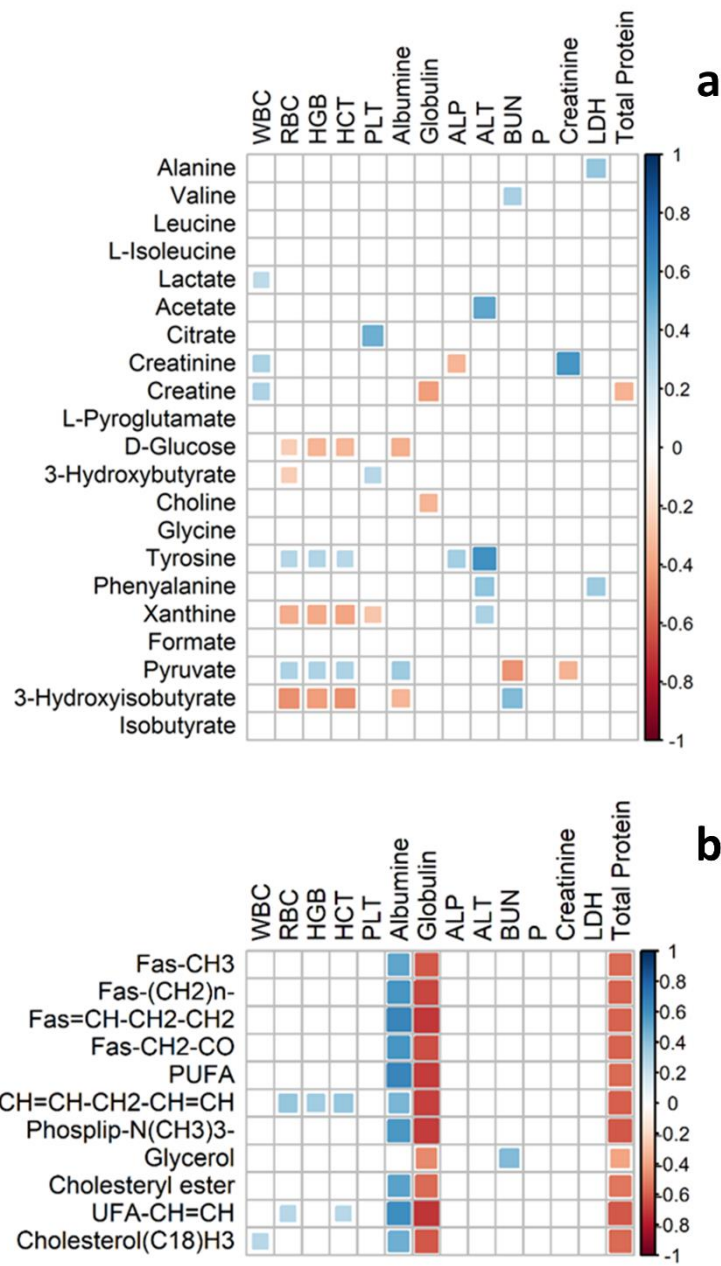


Figure 2 (a – b). Graphical representations of correlation matrices between metabolites (a), lipid fractions (b) (reported in the rows) and blood parameters (reported in columns). Only glyphs that are representatives for a statistically significant correlation (P -values < 0.05) are reported and coloured following a gradient that ranges from dark red to dark blue according to Spearman’s rho correlation coefficients from -1 to +1.

4.3. NMR data pre-processing for metabolomic fingerprinting

Most NMR-based metabolomic studies, especially in the clinical setting, are performed with large number of samples. Some differences between samples groups may be due to biology, to experimental measurements or to their combined effects which are not of interest or which might interfere with the subsequent statistical analysis. Data pre-processing methods allow the reduction of this variability, but depending on the biological problems at issue, not all data pre-processing techniques (*e.g.* spectral bucketing, alignment etc.) are equally able to keep all the necessary biological information to classify samples, *i.e.* to perform sample fingerprinting, or to identify new biomarkers when a profiling approach is chosen.

During past years, different mathematical approaches were developed to pre-treat NMR data, after their acquisition as raw data. Among them, the equidistant bucketing represents an efficient pre-processing method to reduce the total number of variables by dividing NMR spectra into small buckets, typically spanning 0.02 – 0.04 ppm, also in the light of the availability of more computational power to analyse full spectra.

Recently, since most medical and food analyses require quantifiable properties, bucketing became less important, while the interest in the profiling approach continued to grow. Maybe, a bit confusion arose between the specific aim of the two distinct approaches of untargeted analyses (*i.e.* fingerprinting and profiling) and relative tools to achieve them. Indeed, criticisms on the main drawbacks of bucketing for profiling analyses have been implicitly extended to the fingerprinting ones.

In the study proposed in this PhD thesis, we demonstrated that equidistant bucketing does represent the most appropriate and recommended pre-processing procedure for the performance of NMR-based metabolomics fingerprinting (§ 4.3.1). This was possible through the evaluation of how different ways to perform the equidistant bucketing of 1D ¹H-NMR spectra could affect the results of NMR-based metabolomic fingerprinting analysis and how the subsequent pattern recognition performance (using the Random Forest algorithm)⁸⁰ could be impacted. These key-points were addressed using large datasets of serum 1D NOESY/1D CPMG NMR spectra; urine 1D NOESY spectra, and estimating the effects, on the final accuracy of sample classification, of *i*) shifting the buckets boundaries, as many times as the number of spectral points in a 0.02 ppm segment, with the consequent shift of all adjacent data points, and *ii*) changing the bucket widths. The sensitivity of PCA score and loadings plot on binning details (*i.e.* change of bucket size and boundaries) was also reported for both sample datasets. Non uniform binning is also compared.

Globally, our results showed that, in the case of both serum 1D NOESY and 1D CPMG experiments, the location of bucket boundaries does not influence the classification of samples to the right category. By modifying the size of equidistant bucket widths, it seems that converting 1D NMR spectra into matrices of buckets with size increasingly smaller or bigger than the usual 0.02 ppm size, down to using full resolution spectra, provides only marginally better accuracy values. On the other hand, we can clearly demonstrate that final resolutions of multivariate statistical analyses can be strongly

affected by the use of bucketed spectra into segments larger than 0.4 ppm. Instead, challenging the large urine dataset (a total of 1167 samples), we found that changes in the position of segments boundaries do not significantly affect the pattern recognition, as already demonstrated for the serum dataset. After varying the buckets widths, discrimination accuracies between the subjects remain similarly high, with the 0.01, 0.02 and 0.04 ppm bucketing actually providing the best results. This means that, despite urines present a larger shift variability, for their fingerprinting, equidistant bucketing is able to keep all the necessary information, including the one encoded in the chemical shifts data.

With respect to the classical equidistant bucketing procedure, results obtained after applying an optimized bucketing are comparable. Moreover, PCA analysis, carried out on selected bucket matrices obtained after shifting bucket boundaries, changing the bucket width or using optimized bucket matrices, showed that the global shape of the respective score and loadings plots resulted to be almost the same, with only minor differences, finally corroborating the hypothesis that although binning parameters are changed, the clustering of sample cohorts remains unaffected.

As suggested by a reviewer of the presented manuscript, we also evaluated how the bucketing procedure behaves when, in an extreme case, only one single NMR signal serves as discriminating feature. To discuss this point, we created a very simple artificial dataset of one-hundred simulated 600 MHz ^1H NMR spectra, hypothesizing the existence of two distinct groups (A and B) discriminated by the different area of a singlet resonating in the interval 7.19-7.16 ppm. While for group B (low area, let's say "healthy") the singlet shift in the interval 7.19-7.175 ppm, for group A (high area, let's say "diseased") the singlet shifts follow two uniform distributions, one spanning the interval 7.19-7.175 ppm (group A1) and the other spanning the interval 7.18-7.16 ppm (group A2). Thus, while the "biomarker" amount clearly differentiates between "healthy" and "diseased", the latter are also characterized by a more pronounced shift of the signal, that, in our metaphor, could represent different "manifestations" of the disease. If we want to accurately integrate the peak to quantify the "biomarker", we undoubtedly need to chase it along the ppm axis. However, in doing so, we lost the A1 and A2 separation, neglecting other (maybe "clinically" relevant) information. The same is true if we forcefully align the spectra before multivariate statistics.

The application of a fingerprinting approach, based on an unsupervised analysis (PCA) and a supervised one (Random Forest algorithm), on the equidistantly bucketed simulated spectra, confirmed our hypothesis, that was further validated, by repeating the multivariate analysis on the same dataset, but after applying the non-uniform bucketing algorithm.

To recap, the proposed study shows that, although small perturbations in NMR chemical shifts can complicate the recovery of biomarker data, they could be pivotal in characterizing the whole fingerprint, especially in complex biological matrices, like urine samples. In other words, if some spectra are characterized by systematic shifts of some peaks, this is an information that constitutes an integral part of the fingerprint of that spectra. Because it could be due to biological reasons, it cannot be *a priori*

neglected. To this perspective, we advocate that to perform NMR fingerprinting, the equidistant bucketing represents the optimal procedure to follow, avoiding alternative procedures, such as for example, the segmental alignment.

Translating these results for future perspectives, we can also advocate a hypothetical role of NMR-based metabolomics fingerprinting performance using low-resolution, low-field NMR spectroscopy, because a bucketed high-field NMR spectrum is, to a certain extent, a simulation for an NMR spectrum acquired at low-field, since bucketing is a procedure that intrinsically degrades the spectral resolution, but apparently still provides relevant results for fingerprinting.

4.3.1. Simple equidistant bucketing as robust and recommended procedure for NMR-based metabolomic fingerprinting

Cristina Licari,¹ Leonardo Tenori,² Claudio Luchinat^{1,3}

¹Magnetic Resonance Center (CERM), University of Florence, Sesto Fiorentino (Florence), Italy;

²Interuniversity Consortium for Magnetic Resonance of Metalloproteins (C.I.R.M.M.P.), Sesto Fiorentino (Florence), Italy;

³Department of Chemistry “Ugo Schiff”, University of Florence, Via della Lastruccia 3–13, 50019, Sesto Fiorentino, Florence, Italy.

Submitted

Candidate's contributions: Conceptualization of the study, statistical analysis, interpretation of data, writing and review of the manuscript.

Abstract

Equidistant bucketing of NMR spectra represents a common procedure used for metabolomic fingerprinting. However, since most medical and food analyses require quantifiable properties, bucketing has been progressively abandoned, while the interest in profiling increased. Perhaps, a bit of confusion arose between the specific aim of a fingerprinting and a profiling approach and relative tools to reach them. Indeed, criticisms on the main drawbacks of bucketing for profiling analyses have been implicitly extended to the fingerprinting approach, also in the light of the availability of more computational power to analyse full datapoints in the spectra. Here, we show that equidistant bucketing does represent the most appropriate and recommended pre-processing procedure for NMR-based metabolomics fingerprinting. To do so, we evaluated how different ways to perform the equidistant bucketing of the spectra could affect the results of NMR-based metabolomic fingerprinting analysis and how the subsequent pattern recognition performance could be impacted. We address these key-points using large datasets of serum and urine NMR spectra and estimating the effects, on the final accuracy of sample classification, of *i*) shifting the buckets boundaries with the consequent shift of all adjacent data points, and *ii*) changing the bucket widths. Non uniform binning is also compared. Obtained results confirm the robustness and the efficiency of the equidistant bucketing procedure, and support its use, especially in the case of urine, whose shifts are significantly variable, as the most adequate pre-processing method to preserve all the necessary information to perform NMR fingerprinting and sample classification.

Introduction

NMR spectroscopy represents a powerful, versatile and reproducible technique for the analysis of complex biological matrices because any biological molecule containing one or more atoms with a non-zero magnetic moment is theoretically detectable by NMR.¹ Therefore, virtually all biologically relevant molecules are characterized by at least one NMR signal with a specific intensity, frequency (or chemical shift) and magnetic relaxation properties, all reflecting the chemical environment surrounding the detected nucleus.²

In a high-throughput vision of metabolomic analysis, the very high reproducibility, the minimal sample preparation required, and the possibility to simultaneously detect all metabolites presenting NMR active nuclei³, make NMR spectroscopy one of the most suitable techniques for the analysis of any type of biological matrix, enabling the rapid and global evaluation of an NMR spectrum in its entirety (sample fingerprinting) or the determination of the concentrations of all metabolic features that are above the detection limit (sample profiling)¹ (**Figure 1**).

When NMR is applied to profiling, the intrinsic μM detection limit of ^1H -NMR turns into the possibility to ideally measure concentrations of many metabolites, according to the biological matrix under study. However, the number of molecules quantifiable

via profiling is significantly lower than that contributing to the fingerprint (typically <50% in the case of urine)⁴ of the sample, and the spectral processing, necessary to deconvolute 1D NMR spectra to obtain concentrations, may not be completely automated and straightforward.³

As the matrix of biofluids, especially urine, is variable, the local environment of protons also differs and many parameters, *e.g.* changes of pH, salt concentration, overall dilution of sample, relative concentration of metabolites, can affect the results of profiling.⁵ Since all these parameters can influence the shift of many NMR signals, different mathematical approaches were developed over the years with the main aim to overcome these problems, to align the peaks, to integrate and to quantify them.

On the other hand, NMR fingerprinting of samples can be obtained by manipulating NMR spectra after their transformation into data matrices. The procedures most commonly used to convert spectra into data matrices are the so-called “bucketing” or “binning” methods, which allow the integration of NMR spectra within small spectral regions, called “buckets” or “bins”. Many sophisticated algorithms exist to bin 1D NMR spectra. The most commonly and simply used is the equidistant binning of 0.02-0.04 ppm^{5,6}, which allows the division of the spectrum into evenly spaced integral regions with a fixed spectral width.

However, some practitioners argue about the presence of significant difficulties in the analysis of NMR spectral data after their transformation into bucketed matrices, *i.e.* the susceptibility to inter-sample chemical shift variations occurring even when ionic strength and pH are well-controlled, especially in the case of urine,⁷ while other authors believe that strong shifts may lead to non-corresponding peaks incorrectly ending up in the same bin.⁸ Then, there are also experts that reinforce the idea that loss of spectral resolution, intrinsic to the bucketing procedure, can introduce strong artefacts on the border of buckets boundaries.^{4,5} Moreover, despite the simplicity of performing equidistant bucketing, criticisms were made on the main drawback of this method: the absence of adaptability of the bin boundaries. Indeed, some authors retain that if a peak is divided between two bins, small differences in peak frequency among different samples may significantly affect the subsequent data analysis.^{5,6,8} To deal with these problems, alternative methods, based on non-equidistance arrangement, *e.g.* adaptive-intelligent binning,⁹ Gaussian-binning,¹⁰ adaptive-binning through wavelet transform¹¹ and dynamic adaptive binning,¹² have been proposed, while some practitioners prefer to work using full resolution spectra, totally excluding the use of binning methods, but proposing algorithms able to rapidly align NMR peaks, thus assuring the comparability of NMR peaks across multiple spectra.^{8,13}

Although these last methods have been developed for NMR profiling, for which they are valid and highly recommendable, they have been progressively considered to be superior to the simple equidistant bucketing even in the framework of fingerprinting analyses. Perhaps a bit of confusion is born between the aim (and the tools employed to reach it) of a fingerprinting and a profiling approach.

Indeed, spectral alignment procedures greatly simplify the identification, quantification, and comparison of the same peak, from the same compound, across

multiple NMR spectra. The application of these algorithms fits well when the untargeted NMR-based metabolomic analysis is planned via a “profiling” strategy to research biomarker information related to all quantifiable metabolites in a biological sample. Conversely, for fingerprinting purposes, the whole NMR spectrum constitutes the “fingerprint” of the sample and alignment processes can hide important information encoded in chemical shift data, regarding for example, pH, ionic strength and metal ion complement that can be characteristics for the sample under study (especially in the case of urine). Therefore, although small perturbations in NMR chemical shifts can complicate the recovery of biomarker data, they could be pivotal in characterizing the whole fingerprint, especially in complex biological matrices. In other word, if some spectra are characterized by systematic shifts of some peaks, this is an information that constitutes an integral part of the fingerprint of that spectra. Because it could be due to biological reasons, it cannot be a priori neglected. To this perspective, we advocate that to perform NMR fingerprinting, the equidistant bucketing represents the optimal procedure to follow.

Thus, this paper aims to demonstrate that to perform NMR-based metabolomic fingerprinting, it is not necessary to use pre-processing techniques other than simple equidistant bucketing. Indeed, here we prove the robustness, the efficiency and the uniqueness of the equidistant bucketing procedure for NMR-based “fingerprinting” of biological samples by evaluating how sample classification is largely unaffected when using different ways of performing the bucketing of 1D NMR spectra. We trust that the present work will avoid future misunderstandings or non-specific criticisms on the use of this technique to process NMR data for metabolomic fingerprinting purposes.

Material and Methods

In this study, we used 1D ^1H NOESY and 1D ^1H CPMG NMR spectra of serum samples from the AMI-Florence II cohort (for a total of 126 samples) and 1D ^1H NOESY spectra from the first and the second MetRef urine collection (for a total of 1167 samples). Detailed procedures on how samples were collected and prepared are reported in the respective original publications,¹⁴⁻¹⁷ while a brief summary of the acquisition and processing parameters of related NMR experiments are listed in the Supporting Information (**Table S1**).

For both datasets, buckets matrices were obtained as described below.

Through an R¹⁸ in-house developed script, 131 equidistant bucket tables, using 0.02 ppm bucket width, were respectively generated for serum 1D CPMG and urine 1D NOESY experiments, by subsequently starting the bucketing procedure one point rightmost (first one starting at 10 ppm, final one starting at 9.98 ppm), and ending 9.78 ppm afterwards (first one ending at 0.22 ppm, final one ending at 0.2 ppm. See **Figure 2** for an overview of the procedure. Using the same approach, 87 equidistant bucket tables were generated for serum 1D NOESY experiments. More specifically, for serum 1D NOESY spectra, a shift of one point corresponds to 0.14 Hz, while for serum 1D CPMG and urine 1D NOESY spectra, one point-shift corresponds to 0.092 Hz.

Further, for the experiments using different bucket size, each 1D spectrum, in the range of 0.2-10.0 ppm, was segmented into equidistant buckets of different widths (*i.e.* 0.001, 0.002, 0.004, 0.01, 0.02, 0.04, 0.1, 0.2, 0.4, 1 ppm) and the corresponding spectral areas were integrated through AMIX software (version 3.8.4, Bruker BioSpin); obtaining a total of ten data matrices that were used for feeding statistical algorithms. Full resolution spectra were also used (131072 data points for both serum 1D-NOESY/CPMG spectra and urine 1D-NOESY spectra).

All serum and urine 1D ^1H -NMR spectra were also bucketed applying an optimized bucketing method proposed by Sousa *et al.*¹⁹ Briefly, this method optimizes buckets sizes by setting their boundaries at the local minima determined by the average NMR spectrum.

In all cases, the residual H₂O signal was removed for serum 1D NMR spectra (region between 4.5 to 5.1 ppm), while regions between 4.5 and 6.2 ppm, containing residual H₂O and urea signals, were excluded for urine spectra.

Serum bucket matrices and full resolution spectra have been used without normalization, while total area and PQN²⁰ normalizations were applied in the case of urines.

All generated data matrices and full resolution spectra of serum and urine samples were employed for multivariate analysis. Firstly, unsupervised Principal Component Analysis (PCA) was applied on full resolution spectra and on both equidistant and optimized bucket matrices to explore the sensitivity of score and loadings plots on binning details (*i.e.* bucket boundaries and different width). Secondly, samples classification was performed using the Random Forest algorithm²¹⁻²³ (R package “RandomForest”)²⁴ on all available data. In particular, for the serum dataset, discrimination accuracies were established using the settings reported in the original publication of Vignoli *et al.*¹⁷ Instead, for urine samples, predictive accuracies for the individual discrimination of the 31 healthy donors were estimated after building a forest with 5000 trees, applying the default settings of the “Random Forest” function.

Results and Discussion

To demonstrate the robustness of the equidistant bucketing procedure to perform NMR-based metabolomic fingerprinting, a study on two large datasets of 600 MHz ^1H -NMR spectra of serum and urine samples was carried out.

In detail, we planned to test the hypothesis that transforming NMR spectra in data matrices through simple equidistant bucketing represents a highly recommendable way to obtain an efficient sample classification in relation to different biological conditions,^{3,4} when the fingerprinting approach is chosen as the analytical strategy; we want to show that bucketing retains all the necessary information, also including the one encoded in chemical shifts data, to highlight the presence of clusters among samples under study, despite an intrinsic resolution loss.

To this aim, we used two different schemes to challenge the simple equidistant bucketing procedure. The first strategy is based on evaluating the impact of shifting

the boundaries of each equidistant bucket by point-to-point horizontal translations. The reason of this test is to explore how the position of bucket boundaries may influence multivariate statistical discrimination models. In practice, we want to understand to which extent the subdivision of peaks or multiplets across different buckets may affect the classification performance of the final model. For this purpose, buckets have been shifted one point at a time across the usual bucket width of 0.02 ppm. In other words, for each dataset, we obtained as many NMR data matrices as the number of data points in a 0.02 ppm bucket. Then, each data matrix was employed to perform PCA analysis and pattern recognition using the Random Forest algorithm.

As a second approach, we evaluated the effect of varying the commonly used equidistant bucket width of 0.02 ppm on PCA score and loadings plots and on final sample classification, starting from using full resolution spectra to NMR spectra segmented into buckets of increasing size, up to 1 ppm buckets.

Our hypotheses were demonstrated starting from considering a set of one-hundred and twenty-six 600.04 MHz 1D-NOESY and 1D-CPMG spectra of serum samples from the AMI-Florence II cohort,¹⁷ where patients who died within two years after acute myocardial infarction are discriminated from survivors with a predictive accuracy higher than 70%.¹⁷

Since 1D NOESY and 1D CPMG spectra reflect different information related to the biological components of samples, we generally expected that statistical models built on bucketed CPMG or NOESY spectra perform differently in terms of discrimination accuracies. However, for both types of experiments, the shift of buckets boundaries and consequently of all adjacent data points, as many times as the number of spectral points in a 0.02 ppm segment, leads to the achievement of different Random Forest based classification models all having very similar accuracy values (**Figure 3** and **Figure 4**). Therefore, regardless of the use of the 1D-NOESY or 1D-CPMG experiments, the location of bucket boundaries does not influence the classification of samples to the right category (in this specific case, survivors and deceased patients).

On the other hand, observing the results obtained after modifying the size of equidistant bucket widths (**Figure 5** and **Figure 6**), it seems that converting NMR spectra into matrices made of increasing number of buckets with size increasingly smaller than the usual 0.02 ppm size, down to using buckets made of single datapoints, *i.e.* using the full resolution spectra, provides only marginally better accuracy values. These results can be considered reasonable taking into account random errors.

Conversely, we can clearly demonstrate that final resolutions of multivariate statistical analyses can be strongly affected by the use of bucketed spectra into segments of 0.4-1.0 ppm (**Figure 5** and **Figure 6**).

The strength and the robustness of equidistant bucketing for NMR metabolomic fingerprinting is even better demonstrated in the case of urine. To this aim, the analysis was carried out using a large cohort of samples from the first and the second MetRef collection,^{14,15} selecting a total of one-thousand one-hundred sixty-seven urine 1D-NOESY spectra, collected from thirty-one healthy donors. The individuals are recognised from their urine samples with over 95% accuracy, using the Random Forest

algorithm. In previous works,^{14–16} the spectra were converted into 0.02 ppm equidistant buckets to obtain a matrix containing intensity-based descriptors of the original spectrum, and this strategy proved to be effective in discriminating individuals. However, the robustness of this data processing step was not previously assessed.

Interestingly, in this new study, when using data matrices obtained by shifting all the buckets one point at time, the thirty-one healthy donors can be always discriminated with accuracies higher than 95% (**Figure 7**). Despite each final classification model has a different predictive accuracy value, a slightly oscillation of the global predictive accuracies around a same average value is observed (**Figure 7**).

These small variations of final results can be explained since each used data matrix differs by the position of buckets boundaries, thus giving rise to a slightly different final discrimination accuracy of the related predictive statistical model. However, what emerges is that, despite these small changes in the final performances of classification models, all calculated discrimination accuracies are comparable, demonstrating, first of all, the efficiency and robustness of the bucketing technique itself in carrying out NMR fingerprinting of urine samples. Moreover, these results highlight that modifications in the positions of segments boundaries do not significantly affect final classification and pattern recognition in the case of urine, as already demonstrated for the serum dataset.

Crucial interesting information were obtained after evaluating the effect of varying the buckets widths. In detail, using the Random Forest algorithm, different multivariate statistical models were built to discriminate healthy donors considering full resolution spectra and bucketed spectra that were equidistantly segmented into different bucket widths of 0.001, 0.002, 0.004, 0.01, 0.02, 0.04, 0.1, 0.2, 0.4 and 1 ppm, respectively (**Figure 8**).

What is worth of noting in our results is that, through the equidistant bucketing of NMR spectra into narrower segments (from 0.04 ppm bucket size down to full resolution spectra), discrimination accuracies remain similarly high, with the 0.01, 0.02 and 0.04 ppm bucketing actually providing the best results (**Figure 8**). This means that, despite urines present a larger shift variability, for their fingerprinting, equidistant bucketing is able to keep all the necessary information, including the one related to the position of NMR peaks.

Actually, urine peak shifts are an important source of chemical information to be considered for sample fingerprinting, but they can complicate the assignment and the quantification of urine metabolites signals. In this last perspective, in past years, urine peak shifts have been also successfully exploited, by our group, to derive interrelationships between concentrations and chemical shifts in urine.²⁵

The robustness of equidistant bucketing was further assessed by comparing discrimination accuracies obtained after equidistantly bucketing all serum and urine NMR spectra into 0.02 ppm buckets with bucket matrices obtained after applying an optimized bucketing approach.¹⁹ The latter was possible starting from considering, for both sample datasets, bucket widths of 0.02 ppm and then determining, through the definition of the “slackness”, how far the bin boundaries can move by adjusting them

to local minima determined through the average NMR spectrum, in order to provide optimized bucket matrices of different sizes.

Using the above-described method, three different optimized bucket matrices with slackness of 25%, 50% and 75% have been generated for both serum 1D NOESY and 1D CPMG spectra. Applying the Random Forest algorithm on optimized bucketed NOESY spectra, serum samples were classified with predictive accuracies of 68%, 68.2% and 67.2% when a slackness of 25%, 50% and 75% has been respectively used. Instead, 1D-CPMG predictive models, built with the same optimized data matrices, reported accuracies of 73.3%, 74.1% and 74.30%. With respect to the predictive accuracies of 67.7% and 73.8%, obtained respectively for 1D-NOESY and 1D-CPMG statistical models when a classical equidistant bucketing of 0.02 was performed; results obtained after applying an optimized bucketing are comparable.

We derived the same observations in the case of the urine dataset; indeed, using the three optimized buckets matrices with slackness of 25%, 50% and 75%, healthy donors were correctly recognised with predictive accuracies of 97.1%, 96.9% and 96.6%, with respect to the overall predictive accuracy of 96.6%, obtained after feeding the statistical algorithm with an equidistant bucket matrix of 0.02 ppm segments.

PCA analysis, carried out on selected bucket matrices obtained after shifting bucket boundaries, changing the bucket width or using optimized bucket matrices, showed that the global shape of the respective score and loadings plots resulted to be almost the same, with only minor differences, finally corroborating the hypothesis that although binning parameters are changed, the clustering of sample cohorts remains unaffected. Selected PCA score and loadings plot are reported in the Supporting Information (**Figure S1-S6**).

In the light of the out-turn, we recommend users to check the effect of varying the binning parameters on their own analysis to better enhance the peculiarities of the used dataset. However, as demonstrated by our results, binning spectra into 0.02 or 0.01 ppm segments appears as a reasonable choice for the performance of NMR fingerprinting analysis.

Before concluding, as suggested by a reviewer of this manuscript, we evaluated how the equidistant bucketing procedure impacts the clustering of samples, when, in an extreme case, only one single NMR signal characterizes the fingerprint of the samples, *i.e.* it serves as a discriminating feature. To discuss this point, we created a very simple simulated dataset of one-hundred 600 MHz ^1H NMR spectra (see Supplementary **Figure S7** and its caption), hypothesizing the existence of two groups (A and B) discriminated by the different area of a singlet resonating in the interval 7.19-7.16 ppm. While for group B (low area, let's say "healthy") the singlet shift in the interval 7.19-7.175 ppm, for group A (high area, let's say "dis-eased") the singlet's shift follows two uniform distributions, one spanning the interval 7.19-7.175 ppm (group A1) and the other spanning the interval 7.18-7.16 ppm (group A2). Thus, while the "biomarker" amount clearly differentiates between "healthy" and "diseased", the latter are also characterized by a more pronounced shift of the signal, that, in our metaphor, could represent different "manifestations" of the disease. If we want to accurately integrate

the peak to quantify the “biomarker”, we undoubtedly need to chase it along the ppm axis. However, in doing so, we lost the A1 and A2 separation, neglecting other (maybe “clinically” relevant) information. The same is true if we forcefully align the spectra before multivariate statistics.

Using a fingerprinting approach based on an equidistant bucketing into 0.02 ppm segments, PCA analysis clearly evidenced the presence of three different clusters, with “diseases” well separated from “healthy”, but also showing the presence of two subgroups corresponding to A1 and A2 samples (**Figure S7 B**). Further, a supervised Random Forest analysis performed on the same bucket matrix is able to discriminate the three groups with an accuracy of about 90%.

To further stress this point we applied the optimized bucketing algorithm (0.02 ppm) with slackness of 50% and 75% to the same artificial dataset (**Figure S7 C and D**). As expected, the information about sub-groups A1 and A2 is masked by the bucket’s variable size that is increased as much as possible to include the signal in only one bucket.

In conclusion, all these results show that peak shifts are characteristics for sample classification and the equidistant bucketing succeed in maintaining the information encoded in the chemical shifts data; conversely, having aligned spectra, this peculiar information would have been completely lost.

Conclusions

In summary, we can state that, independently of the type of biofluid under analysis and with respect to an optimized procedure of the spectral bucketing, the equidistant bucketing processing is as simple as robust to perform the NMR fingerprinting of biological samples. Relevant differences are not highlighted between PCA score/loadings plots and classification accuracies obtained using full resolution spectra or segmented NMR spectra into narrower buckets in the case of serum; while the transformation of NMR data spectra into a smaller set of variables is actually optimal for urine fingerprinting. The results achieved in this paper also highlight the main drawback related to the application of the bucketing procedure: the consistent loss of spectral resolution using buckets larger than 0.4 ppm, but at the same time, they can be interpreted as new perspective.

Indeed, it is known that the reduction of spectral resolution can substantially affect final performances in NMR sample profiling (where the aim is to assign and integrate individual signals), but our results suggest that the same statement is not valid for NMR fingerprinting. Thus, translating the results in terms of future perspectives, we can advocate a hypothetical role of NMR-based metabolomic fingerprinting performance using low-resolution, low-field NMR spectroscopy, because a bucketed high-field NMR spectrum is, to a certain extent, a mock for an NMR spectrum acquired at low-field. As an example, using bucket widths of 0.1 or 0.2 ppm for both serum and urine sample classification, we obtained predictive models where the information retrieved from the fingerprint is still (sub-optimally) maintained.

In other words, since bucketing is a procedure that intrinsically degrades the spectral resolution, but apparently still provides relevant results for fingerprinting, it may not be always mandatory to perform metabolic fingerprinting using high resolution spectra. Thus, since some authors suggest the possible role for small benchtop low-field NMR instruments in biofluid analysis for point-of-care applications,²⁶ our findings confirm the possibility of performing fast and cheap low-field fingerprinting of diseases, especially in areas where availability, accessibility and affordability of common and more expensive analytical techniques are not granted. Of course, these considerations are not valid for metabolic profiling: to accurately assign and quantify as many metabolites as possible, the high resolution and the high sensitivity typically obtained at high field are indeed needed.

References

- (1) Takis, P. G.; Ghini, V.; Tenori, L.; Turano, P.; Luchinat, C. Uniqueness of the NMR Approach to Metab-olomics. *TrAC Trends in Analytical Chemistry* 2019, 120, 115300. <https://doi.org/10.1016/j.trac.2018.10.036>.
- (2) Krishnan, P.; Kruger, N. J.; Ratcliffe, R. G. Metabolite Fin-gerprinting and Profiling in Plants Using NMR. *J. Exp. Bot.* 2005, 56 (410), 255–265. <https://doi.org/10.1093/jxb/eri010>.
- (3) Vignoli, A.; Ghini, V.; Meoni, G.; Licari, C.; Takis, P. G.; Tenori, L.; Turano, P.; Luchinat, C. High-Throughput Metab-olomics by 1D NMR. *Angew. Chem. Int. Ed. Engl.* 2019, 58 (4), 968–994. <https://doi.org/10.1002/anie.201804736>.
- (4) Griffiths, W. J. *Metabolomics, Metabonomics and Metabolite Profiling*; Royal Society of Chemistry, 2008.
- (5) Lindon, J. C.; Nicholson, J. K.; Holmes, E. *The Handbook of Metabonomics and Metabolomics*; Elsevier, 2011.
- (6) Smolinska, A.; Blanchet, L.; Buydens, L. M. C.; Wijmenga, S. S. NMR and Pattern Recognition Methods in Metabolom-ics: From Data Acquisition to Biomarker Discovery: A Re-view. *Analytica Chimica Acta* 2012, 750, 82–97. <https://doi.org/10.1016/j.aca.2012.05.049>.
- (7) Barry, S.; Sengupta, A.; Weljie, A. Chapter 3 - NMR Spec-troscopy of Urine. In *NMR-based metabolomics*; Hector C. Keun; p 368.
- (8) Vu, T. N.; Laukens, K. Getting Your Peaks in Line: A Re-view of Alignment Methods for NMR Spectral Data. *Metabo-lites* 2013, 3 (2), 259–276. <https://doi.org/10.3390/metabo3020259>.
- (9) De Meyer, T.; Sinnaeve, D.; Van Gasse, B.; Tsiporkova, E.; Rietzschel, E. R.; De Buyzere, M. L.; Gillebert, T. C.; Beka-ert, S.; Martins, J. C.; Van Criekeing, W. NMR-Based Char-acterization of Metabolic Alterations in Hypertension Using an Adaptive, Intelligent Binning Algorithm. *Anal. Chem.* 2008, 80 (10), 3783–3790. <https://doi.org/10.1021/ac7025964>.
- (10) Anderson, P. E.; Reo, N. V.; DelRaso, N. J.; Doom, T. E.; Raymer, M. L. Gaussian Binning: A New Kernel-Based Method for Processing NMR Spectroscopic

Data for Metabo-lomics. *Metabolomics* 2008, 4 (3), 261–272. <https://doi.org/10.1007/s11306-008-0117-3>.

(11) Davis, R. A.; Charlton, A. J.; Godward, J.; Jones, S. A.; Har-rison, M.; Wilson, J. C. Adaptive Binning: An Improved Binning Method for Metabolomics Data Using the Undeci-mated Wavelet Transform. *Chemometrics and Intelligent La-boratory Systems* 2007, 85 (1), 144–154. <https://doi.org/10.1016/j.chemolab.2006.08.014>.

(12) Anderson, P. E.; Mahle, D. A.; Doom, T. E.; Reo, N. V.; DelRaso, N. J.; Raymer, M. L. Dynamic Adaptive Binning: An Improved Quantification Technique for NMR Spectro-scopic Data. *Metabolomics* 2011, 7 (2), 179–190. <https://doi.org/10.1007/s11306-010-0242-7>.

(13) Savorani, F.; Tomasi, G.; Engelsen, S. B. Icoshift: A Versa-tiler Tool for the Rapid Alignment of 1D NMR Spectra. *Jour-nal of Magnetic Resonance* 2010, 202 (2), 190–202. <https://doi.org/10.1016/j.jmr.2009.11.012>.

(14) Bernini, P.; Bertini, I.; Luchinat, C.; Nepi, S.; Saccenti, E.; Schäfer, H.; Schütz, B.; Spraul, M.; Tenori, L. Individual Human Phenotypes in Metabolic Space and Time. *J. Proteo-me Res.* 2009, 8 (9), 4264–4271. <https://doi.org/10.1021/pr900344m>.

(15) Assfalg, M.; Bertini, I.; Colangiuli, D.; Luchinat, C.; Schäfer, H.; Schütz, B.; Spraul, M. Evidence of Different Metabolic Phenotypes in Humans. *PNAS* 2008, 105 (5), 1420–1424. <https://doi.org/10.1073/pnas.0705685105>.

(16) Ghini, V.; Saccenti, E.; Tenori, L.; Assfalg, M.; Luchinat, C. Allostasis and Resilience of the Human Individual Metabolic Phenotype. *J. Proteome Res.* 2015, 14 (7), 2951–2962. <https://doi.org/10.1021/acs.jproteome.5b00275>.

(17) Vignoli, A.; Tenori, L.; Giusti, B.; Takis, P. G.; Valente, S.; Carrabba, N.; Balzi, D.; Barchielli, A.; Marchionni, N.; Gen-sini, G. F.; Marcucci, R.; Luchinat, C.; Gori, A. M. NMR-Based Metabolomics Identifies Patients at High Risk of Death within Two Years after Acute Myocardial Infarction in the AMI-Florence II Cohort. *BMC Med* 2019, 17 (1), 3. <https://doi.org/10.1186/s12916-018-1240-2>.

(18) Ihaka, R.; Gentleman, R. R. A Language for Data Analysis and Graphics. *J Comput Stat Graph* 1996, 5, 299–314.

(19) Sousa, S. A. A.; Magalhães, A.; Ferreira, M. M. C. Optimized Bucketing for NMR Spectra: Three Case Studies. *Chemomet-rics and Intelligent Laboratory Systems* 2013, 122, 93–102. <https://doi.org/10.1016/j.chemolab.2013.01.006>.

(20) Dieterle, F.; Ross, A.; Schlotterbeck, G.; Senn, H. Probabilis-tic Quotient Normalization as Robust Method to Account for Dilution of Complex Biological Mixtures. Application in 1H NMR Metabonomics. *Anal. Chem.* 2006, 78 (13), 4281–4290. <https://doi.org/10.1021/ac051632c>.

(21) Breiman, L. Random Forests. *Machine Learning* 2001, 45 (1), 5–32. <https://doi.org/10.1023/A:1010933404324>.

(22) Touw, W. G.; Bayjanov, J. R.; Overmars, L.; Backus, L.; Boekhorst, J.; Wels, M.; Hijum, V.; T, S. A. F. Data Mining in the Life Sciences with Random Forest: A Walk in the Park or Lost in the Jungle? *Brief Bioinform* 2013, 14 (3), 315–326. <https://doi.org/10.1093/bib/bbs034>.

- (23) Verikas, A.; Gelzinis, A.; Bacauskiene, M. Mining Data with Random Forests: A Survey and Results of New Tests. *Pattern Recognition* 2011, 44 (2), 330–349. <https://doi.org/10.1016/j.patcog.2010.08.011>.
- (24) Liaw, A.; Wiener, M. Classification and Regression by Random Forest. *R news* 2002, 2 (3), 18–22.
- (25) Takis, P. G.; Schäfer, H.; Spraul, M.; Luchinat, C. Deconvoluting Interrelationships between Concentrations and Chemical Shifts in Urine Provides a Powerful Analysis Tool. *Nat Commun* 2017, 8 (1), 1662. <https://doi.org/10.1038/s41467-017-01587-0>.
- (26) Percival, B. C.; Grootveld, M.; Gibson, M.; Osman, Y.; Molinari, M.; Jafari, F.; Sahota, T.; Martin, M.; Casanova, F.; Mather, M. L.; Edgar, M.; Masania, J.; Wilson, P. B. Low-Field, Benchtop NMR Spectroscopy as a Potential Tool for Point-of-Care Diagnostics of Metabolic Conditions: Validation, Protocols and Computational Models. *High Throughput* 2018, 8 (1). <https://doi.org/10.3390/ht8010002>.

Figures

Graphic abstract

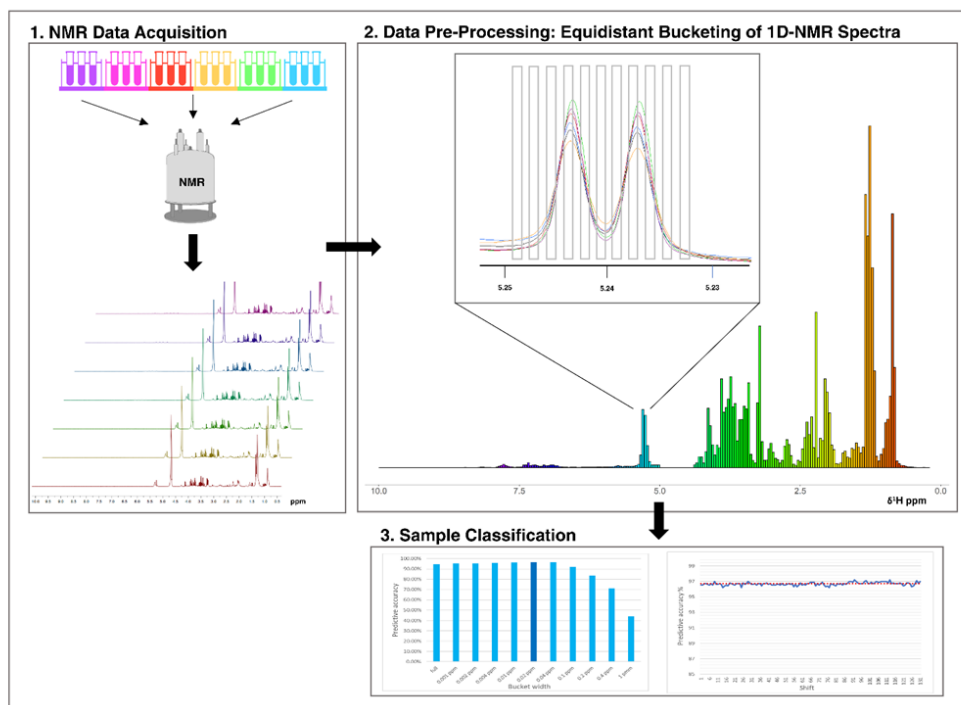


Figure 1. Different untargeted metabolomic approaches using NMR spectroscopy. Fingerprinting is used to globally evaluate all of the features of a bucketed spectrum, without identifying single metabolites, but the whole fingerprint of the sample. Instead, profiling deals with the quantification of concentrations of all metabolites above the μM detection limit. Figure adapted from ref³.

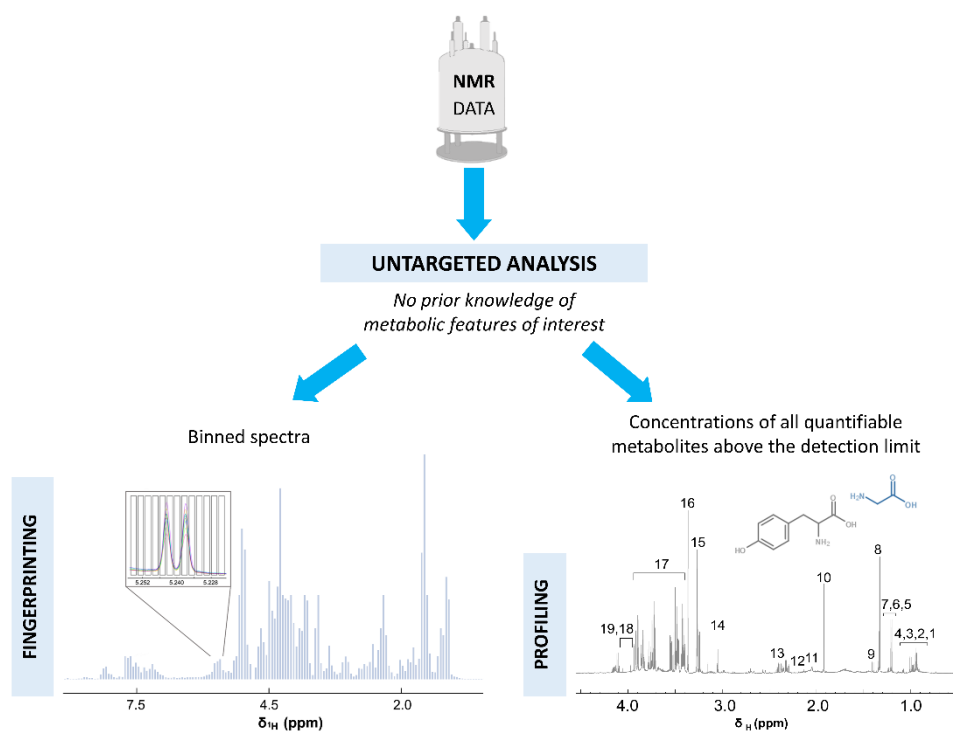


Figure 2. Graphical representation of the procedure followed to obtain as many shifted NMR bucket tables as the number of data points in a 0.02 ppm bucket. 131 shifted NMR data matrices were respectively generated for urine 1D NOESY spectra and serum 1D CPMG spectra, by subsequently starting the bucketing procedure one point rightmost: first one starting at 10 ppm, final one starting at 9.98 ppm (bold numbers in black), and ending 9.78 ppm afterwards: first one ending at 0.22 ppm and the final one ending at 0.2 ppm (bold numbers in black). On the right side of the figure, an enlargement better represents the procedure followed. The same procedure was then applied to obtain 87 shifted NMR data matrices in the case of serum 1D NOESY experiments.

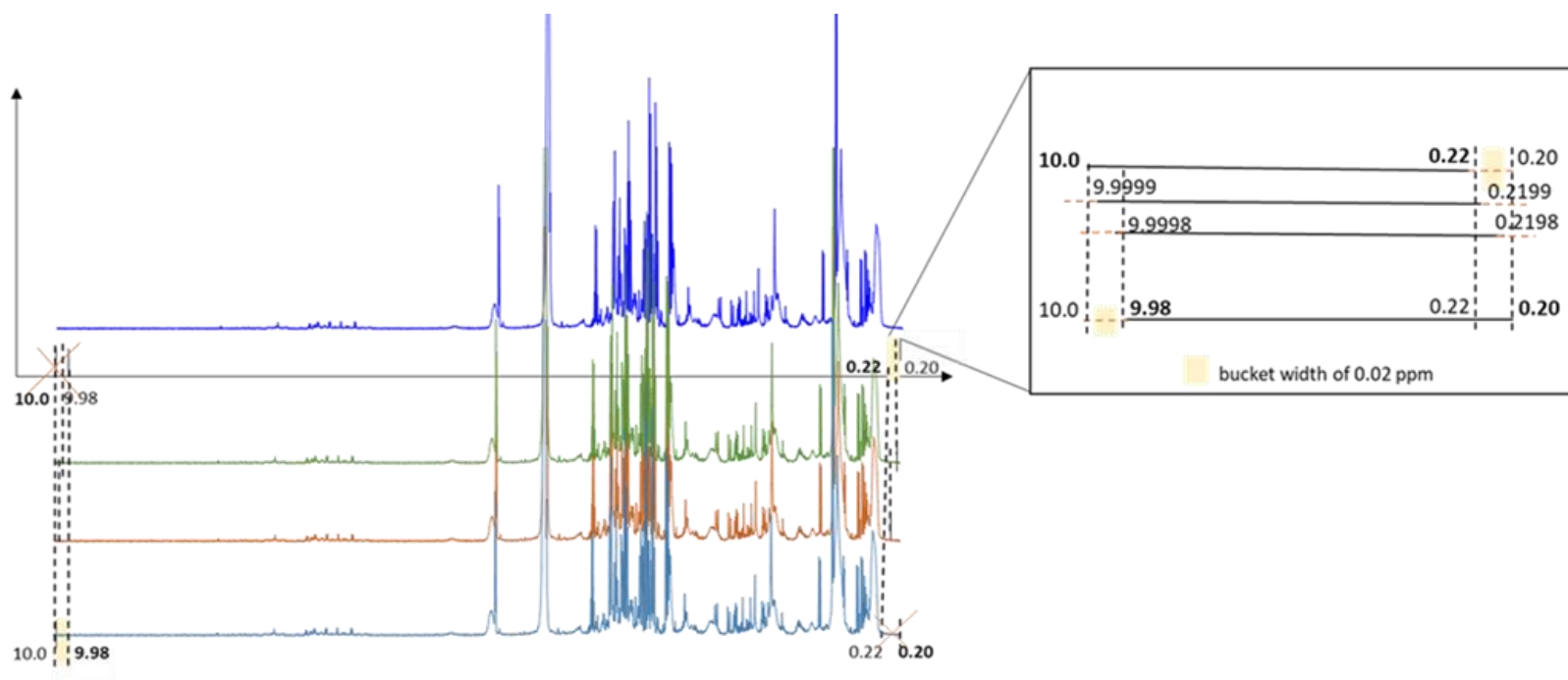


Figure 3. Trend of predictive accuracies (y axis) of classification models calculated for each data matrix shifted point by point (x axis), in the case of the 1D CPMG NMR spectra of the serum dataset. Classification models and relative accuracies have been estimated using the Random Forest classifier.

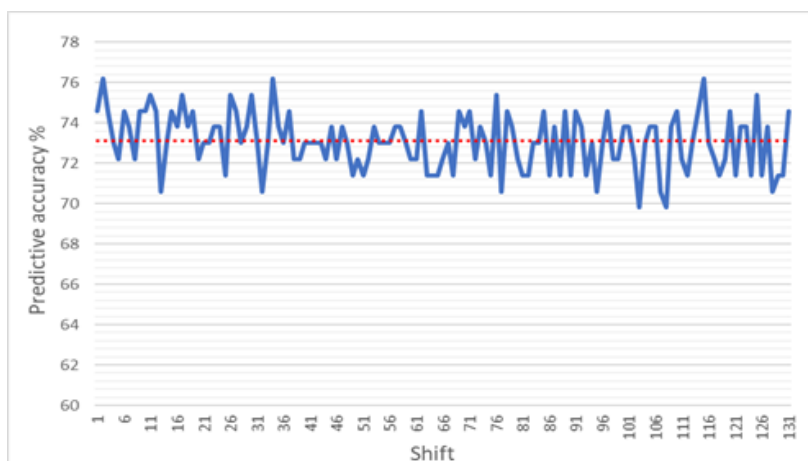


Figure 4. Trend of predictive accuracies (y axis) of classification models calculated for each data matrix shifted point by point (x axis), in the case of the 1D NOESY NMR spectra of the serum dataset. Classification models and relative accuracies have been estimated using the Random Forest classifier.

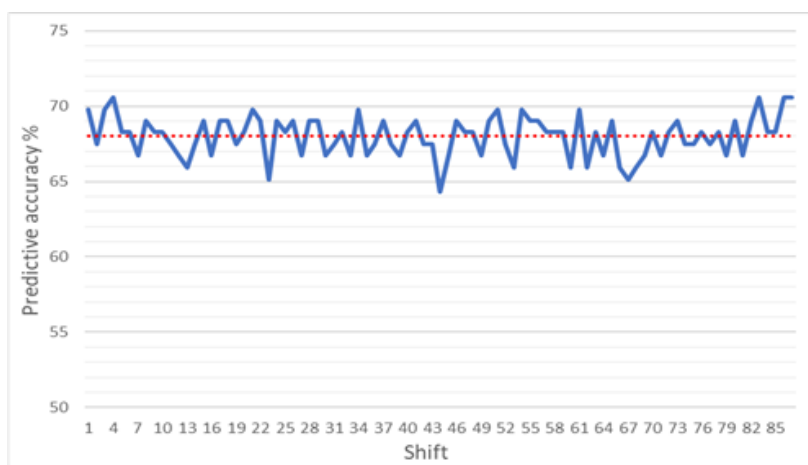


Figure 5. Histogram indicating the trend of the predictive accuracies calculated for serum samples classification using full resolution NMR spectra and spectra bucketed into different equidistant bucket widths (reported on the x axis). All data refer to 1D-CPMG experiments.

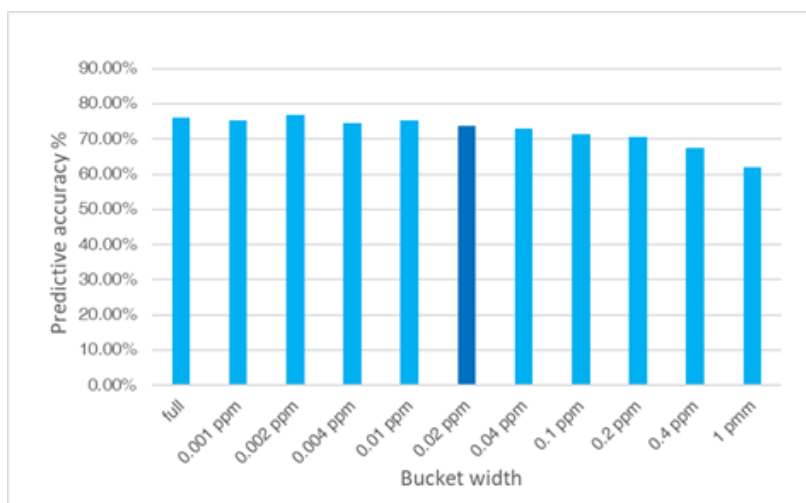


Figure 6. Histogram indicating the trend of the predictive accuracies calculated for serum samples classification using full resolution NMR spectra and spectra bucketed into different equidistant bucket widths (reported on the x axis). All data refer to 1D-NOESY experiments.

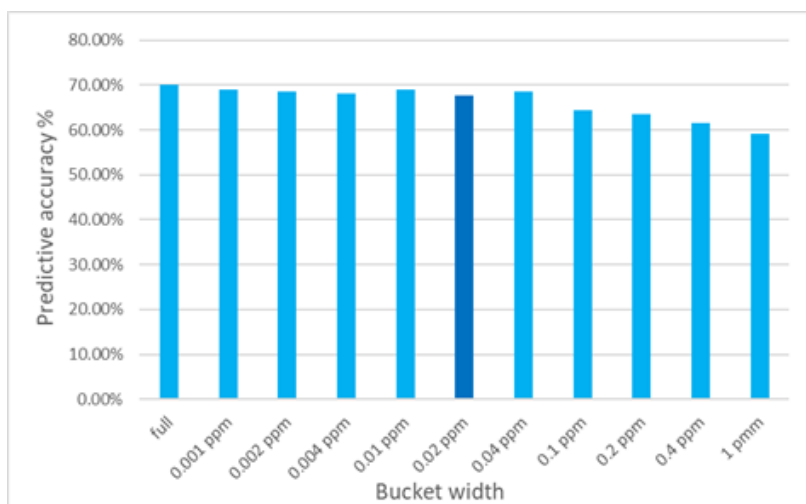


Figure 7. Trend of predictive accuracies (y axis) of classification models calculated for each data matrix shifted point by point (x axis), in the case of the urine dataset. Classification models and relative accuracies have been estimated using the Random Forest algorithm (number of forest tree = 5000).

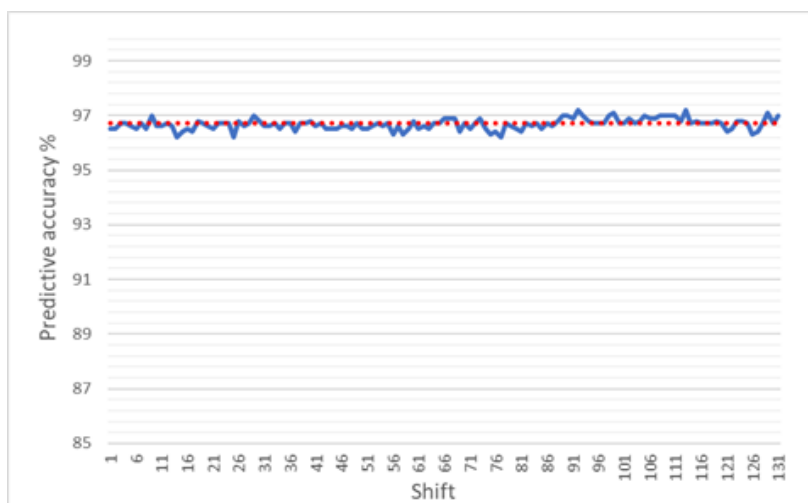
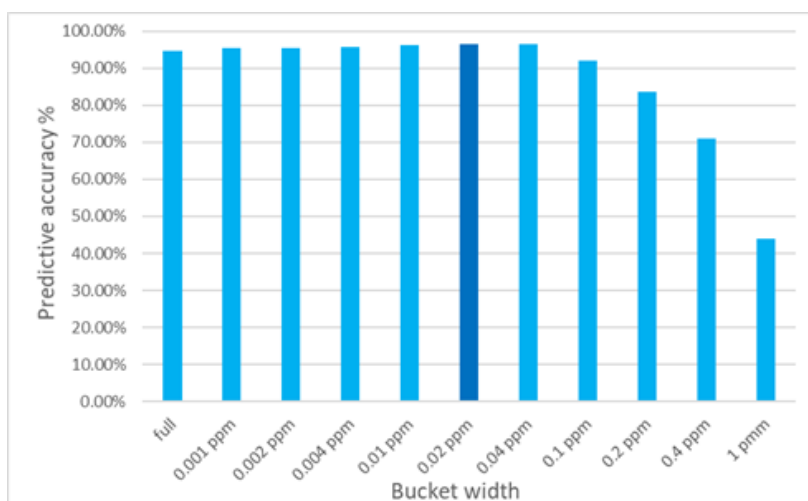


Figure 8. Histogram indicating the trend of the predictive accuracies calculated for urine samples classification using full resolution NMR spectra and spectra bucketed into different equidistant bucket widths (reported on the x axis).



Supplementary Material**Table S1.** Summary of acquisition and processing parameters of urine 1D-NOESY experiments, serum 1D-NOESY and 1D-CPMG experiments.

URINE - 1D NOESY			
Acquisition parameters		Processing parameters	
Size of fid	65536	Size of real spectrum	131072
Number of dummy scans	4	Spectrometer frequency (MHz)	600.13
Number of scans	64	Apodization window function	Exponential multiplication (em)
Spectral width (ppm)	20.0276	Line broadening for em (Hz)	1.00
Acquisition time (sec)	2.73		
Mixing time (sec)	0.1		
relaxation delay (sec)	4		
Fid resolution (Hz)	0.37		
Dwell time (μ sec)	41.6		
SERUM - 1D NOESY			
Acquisition parameters		Processing parameters	
Size of fid	98304	Size of real spectrum	131072
Number of dummy scans	4	Spectrometer frequency (MHz)	600.04
Number of scans	64	Apodization window function	Exponential multiplication (em)
Spectral width (ppm)	30.0459	Line broadening for em (Hz)	1.00
Acquisition time (sec)	2.73		
Mixing time (sec)	0.01		
relaxation delay (sec)	4		
Fid resolution (Hz)	0.37		
Dwell time (μ sec)	27.73		
SERUM - 1D CPMG			
Acquisition parameters		Processing parameters	
Size of fid	73728	Size of real spectrum	131072
Number of dummy scans	4	Spectrometer frequency (MHz)	600.04
Number of scans	64	Apodization window function	Exponential multiplication (em)
Spectral width (ppm)	20.0306	Line broadening for em (Hz)	1.00
Acquisition time (sec)	3.07		
Mixing time (sec)	0.01		
relaxation delay (sec)	4		
Fid resolution (Hz)	0.37		
Dwell time (μ sec)	27.73		

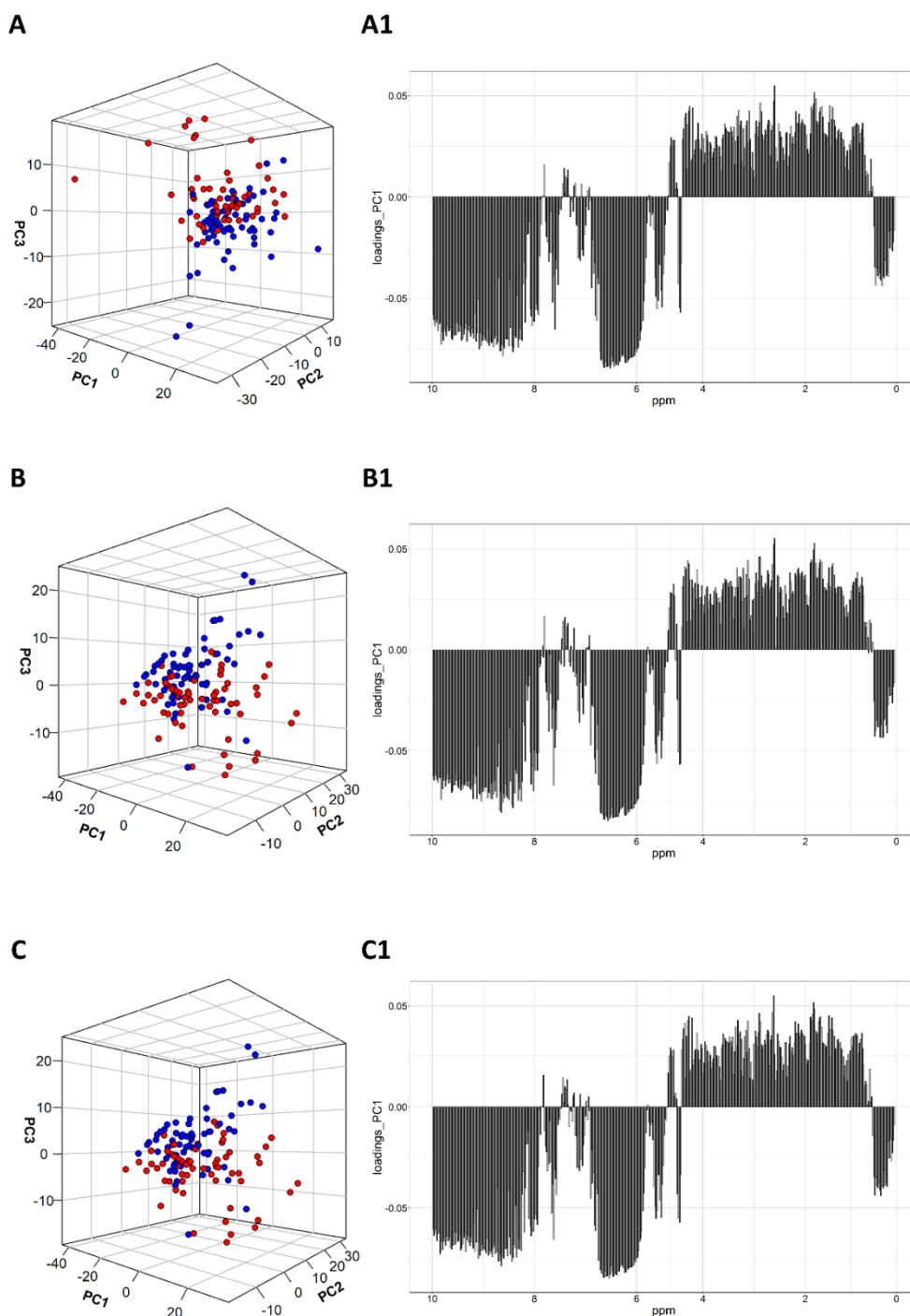


Figure S1. PCA 3D score and loading plots obtained after shifting the bucket boundaries of one (A-A1), sixty-five (B-B1) and one-hundred thirty-one (C-C1) points with respect to the first starting point at 10.0 ppm, for serum 1D-CPMG spectra. In the score plots (A, B, C), each dot represents the serum 1D-CPMG bucketed spectrum, colour-coded by the group of subjects (red = dead; blue = survived). In the loadings plot (A1, B1, C1), loadings of the first component (PC1) are plotted against each related ppm.

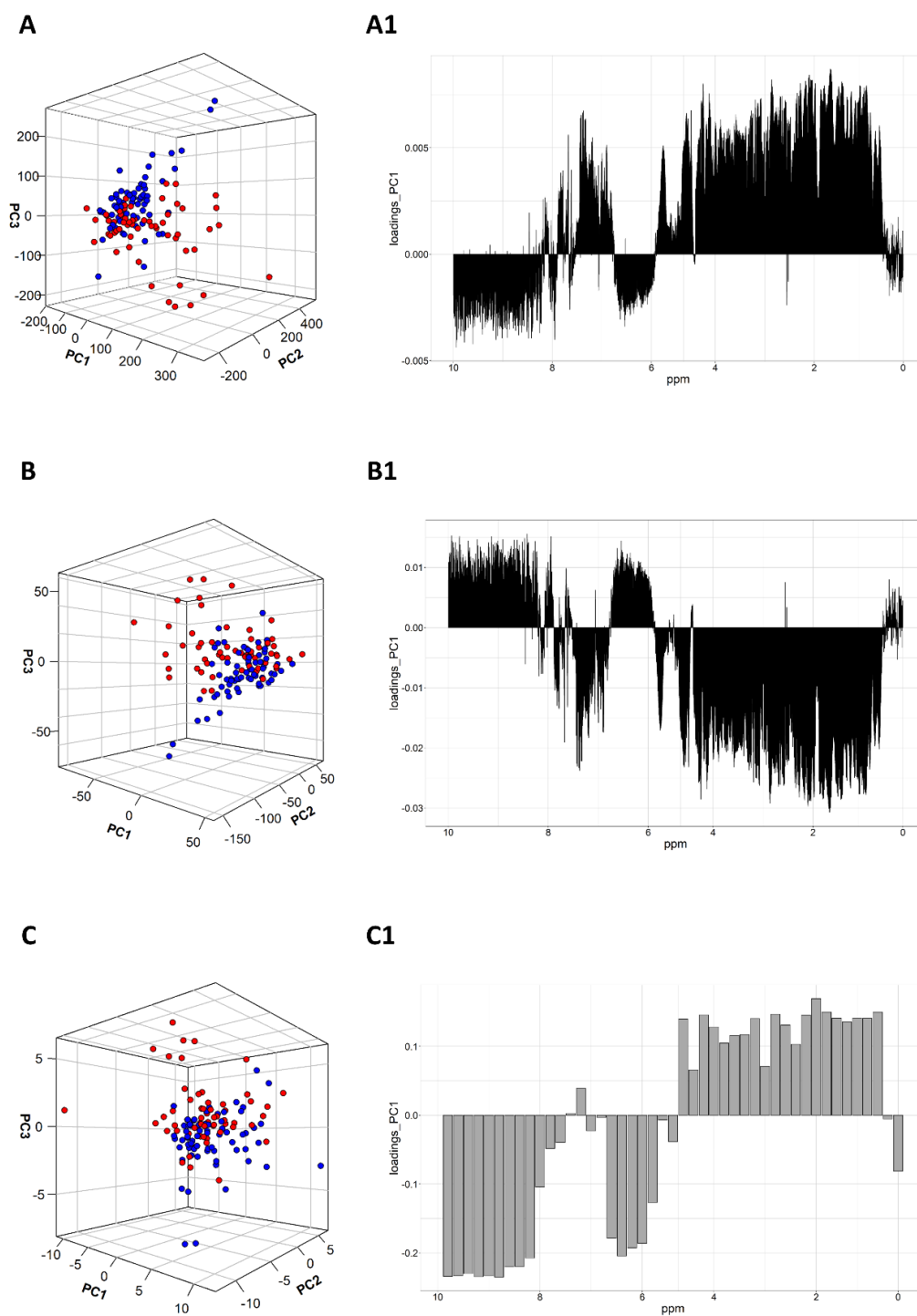


Figure S2. PCA 3D score and loading plots obtained for serum 1D-CPMG spectra, after using full resolution spectra (**A, A1**), bucketed spectra into 0.002 ppm segments (**B, B1**) and 0.2 ppm segments (**C, C1**). In the score plots (**A, B, C**), each dot represents the serum 1D-CPMG bucketed spectrum, colour-coded by the group of subjects (red = dead; blue = survived). In the loadings plot (**A1, B1, C1**), loadings of the first component (PC1) are plotted against each related ppm.

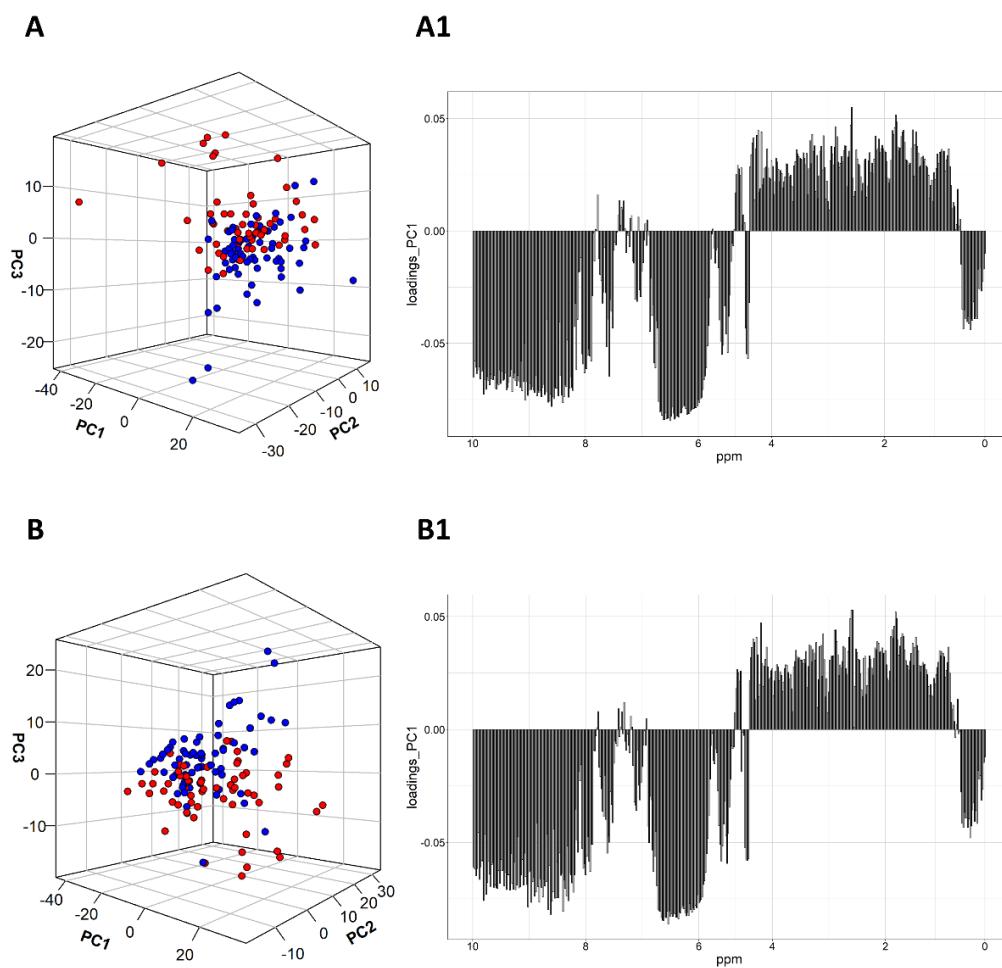


Figure S3. PCA 3D score and loading plots obtained after equidistantly bucketing 1D-CPMG serum spectra into 0.02 ppm segments (**A**, **A1**), and after applying an optimized bucketing procedure, starting from a bucket width of 0.02 ppm and using a slackness of 50% (**B**, **B1**). In the score plots (**A**, **B**), each dot represents the serum 1D-CPMG bucketed spectrum, colour-coded by the group of subjects (red = dead; blue = survived). In the loadings plot (**A1**, **B1**), loadings of the first component (PC1) are plotted against each related ppm.

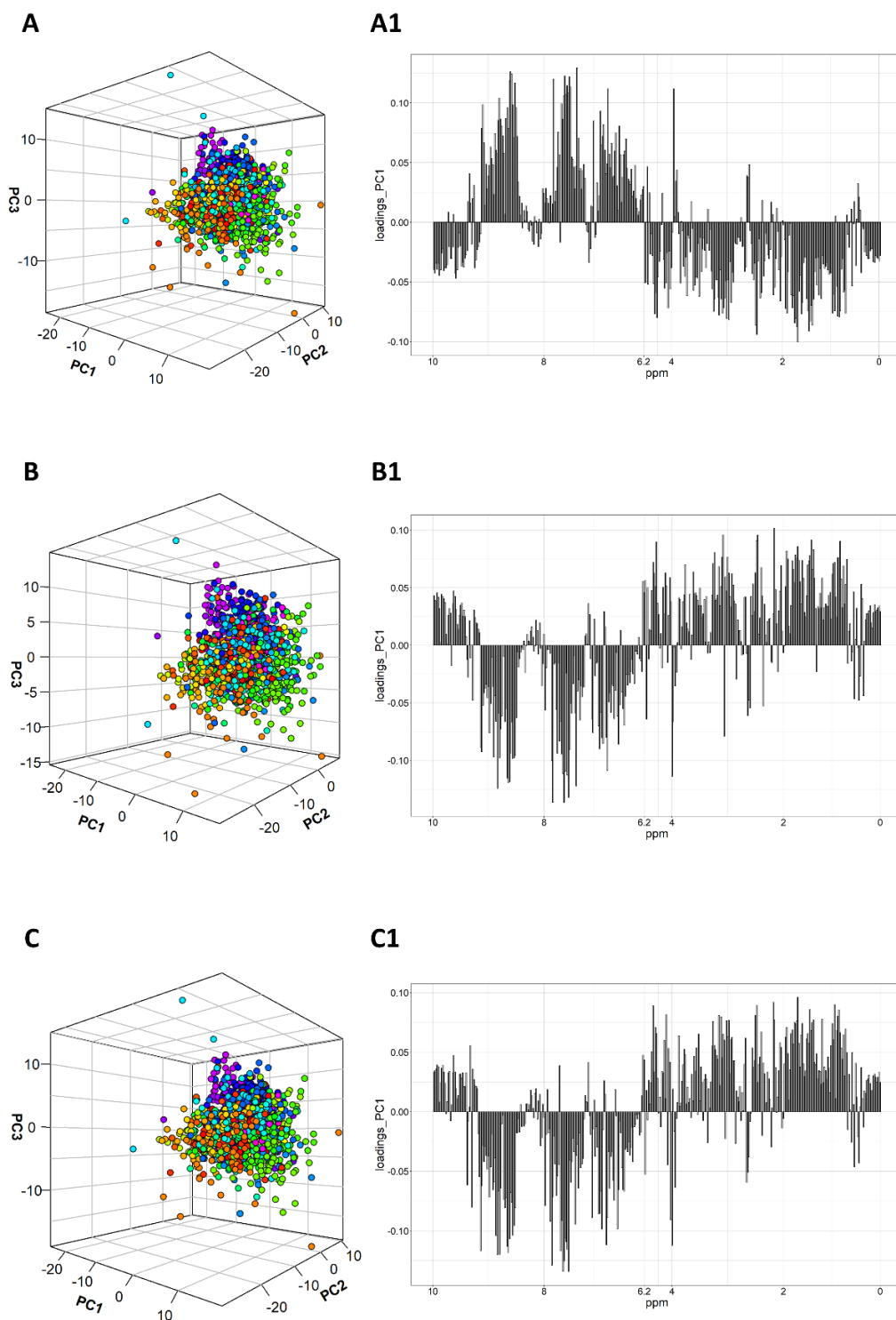


Figure S4. PCA 3D score and loading plots obtained after shifting the bucket boundaries of one (**A-A1**), sixty-five (**B-B1**) and one-hundred thirty-one (**C-C1**) points with respect to the first starting point at 10.0 ppm, for urine 1D-NOESY spectra. In the score plots (**A**, **B**, **C**), each dot represents the urine 1D-NOESY bucketed spectrum, colour-coded by the different 31 healthy donors (for a total of 1167 NMR spectra). In

the loadings plot (A1, B1, C1), loadings of the first component (PC1) are plotted against each related ppm.

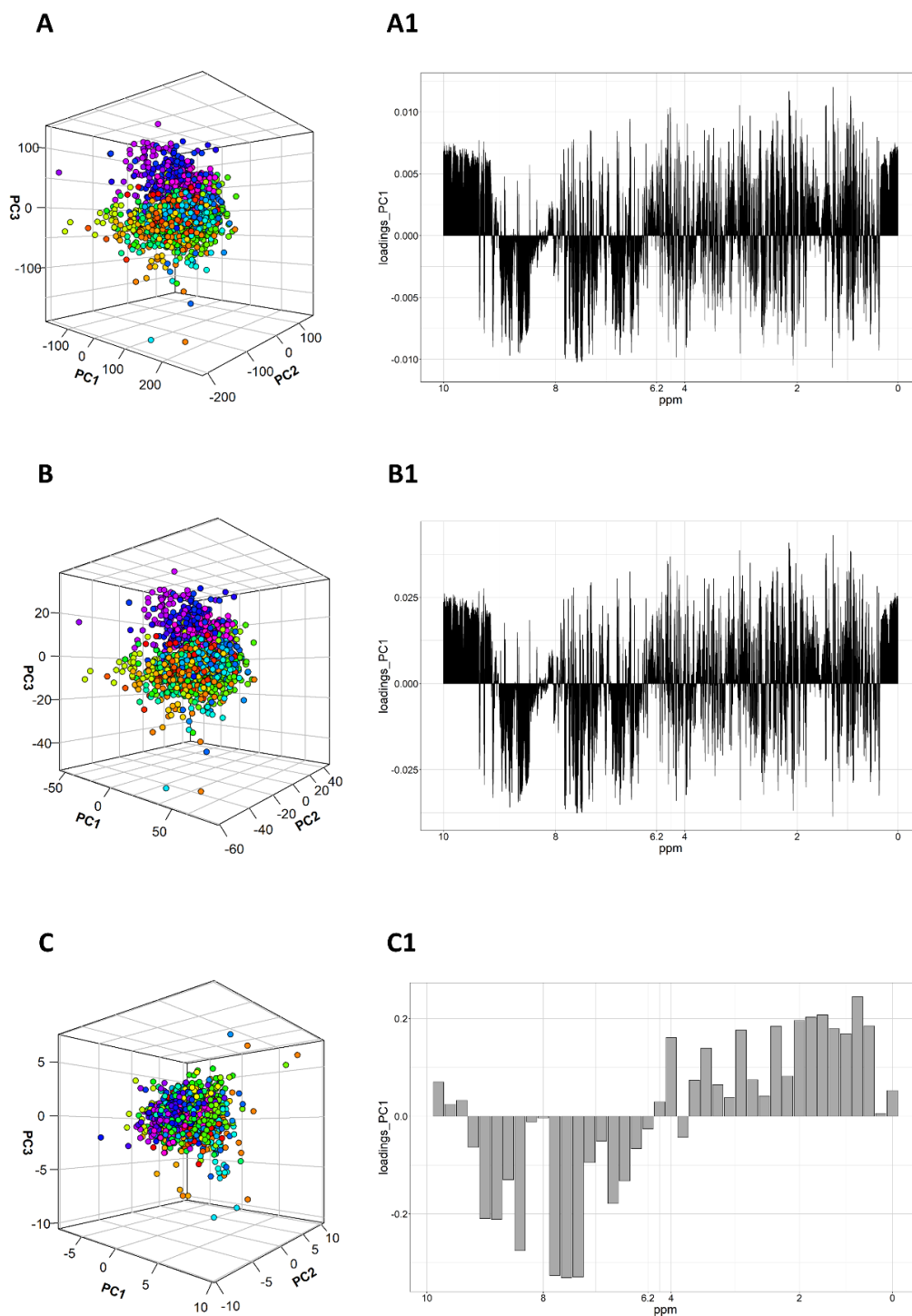


Figure S5. PCA 3D score and loading plots obtained for urine 1D-NOESY spectra, after using full resolution spectra (**A**, **A1**), bucketed spectra into 0.002 ppm segments (**B**, **B1**) and 0.2 ppm segments (**C**, **C1**). In the score plots (**A**, **B**, **C**), each dot represents

the urine 1D-NOESY bucketed spectrum, colour-coded by the different 31 healthy donors (for a total of 1167 NMR spectra). In the loadings plot (**A1**, **B1**, **C1**), loadings of the first component (PC1) are plotted against each related ppm.

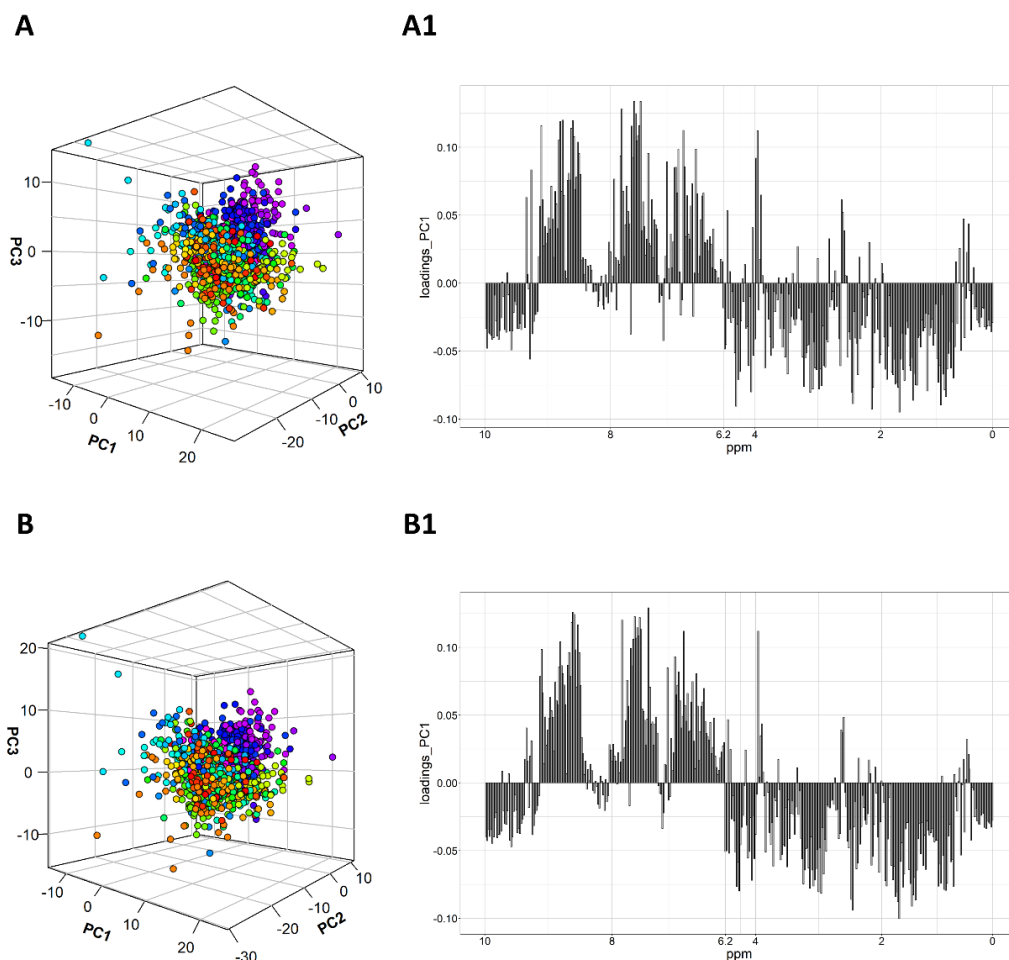


Figure S6. PCA 3D score and loading plots obtained after equidistantly bucketing 1D-NOESY urine spectra into 0.02 ppm segments (**A**, **A1**), and after applying an optimized bucketing procedure, starting from a bucket width of 0.02 ppm and using a slackness of 50% (**B**, **B1**). In the score plots (**A**, **B**), each dot represents the urine 1D-NOESY bucketed spectrum, colour-coded by the different 31 healthy donors (for a total of 1167 spectra). In the loadings plot (**A1**, **B1**), loadings of the first component (PC1) are plotted against each related ppm.

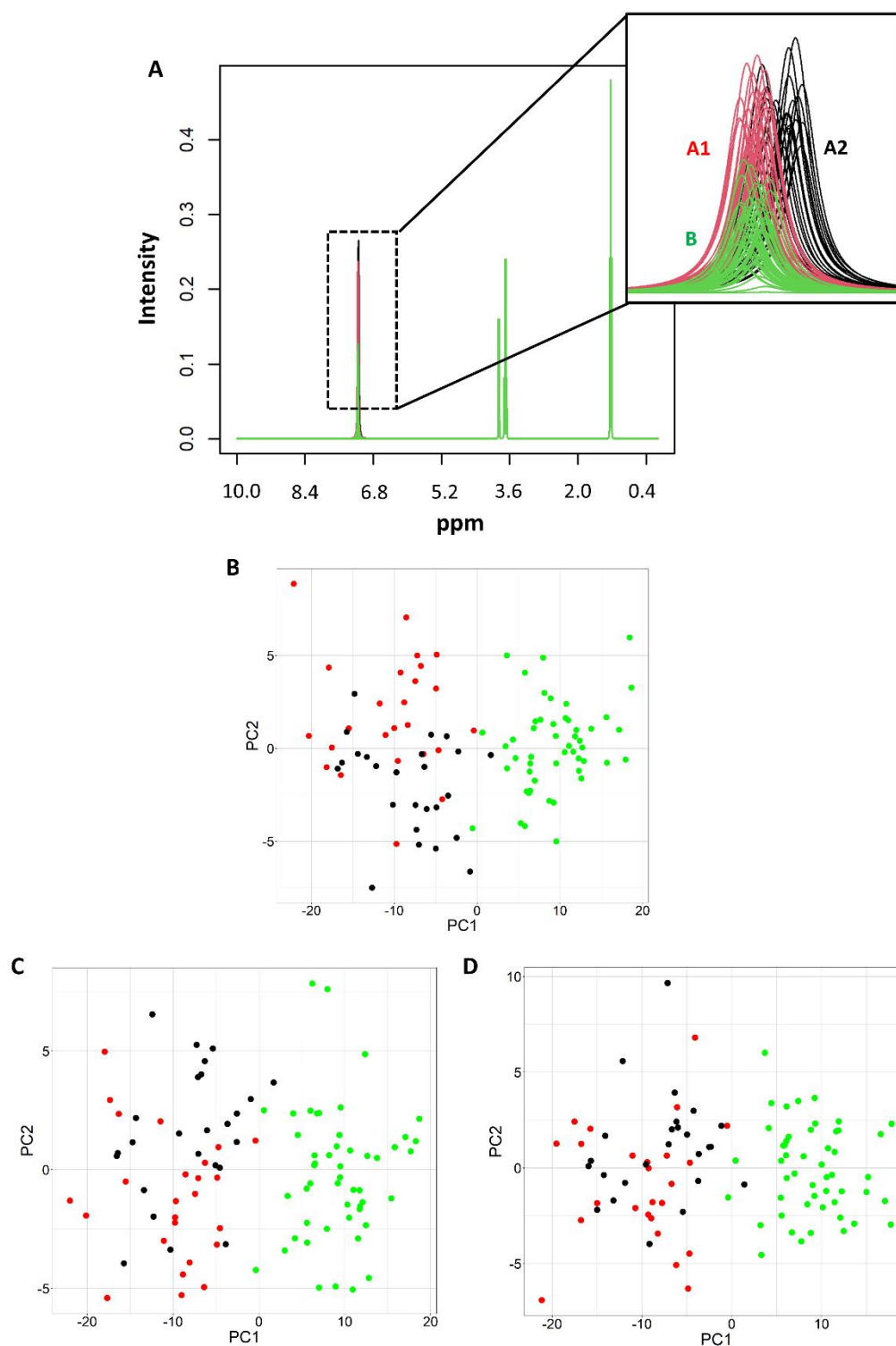


Figure S7. One-hundred 600 MHz simulated ^1H NMR spectra (21 points/Hz; 123,480 points for each spectrum), composed by a singlet, a doublet, a quartet and a triplet (from left to right) with a J-coupling of 0, 14, 14, 14 Hz, a peak width of 10, 2, 2, 2 Hz and resonating in the interval 7.19-7.16 and at 3.9, 3.75, 1.3 ppm, respectively. All spectra have been generated using the R function “plotNMRspec” (“SpecHelpers” R package, <https://cran.r-project.org/web/packages/SpecHelpers/index.html>).

We assumed that the single NMR peak, resonating in the interval 7.19-7.16 ppm, serves as a characteristic discriminating “biomarker” with intensities markedly lower in a group of 50 spectra (green, B). Further, the position of the singlet shifts in the interval 7.19-7.16 ppm, following two uniform distributions: one spanning the interval 7.19-7.175 ppm (red (A1) and green) and the other spanning the interval 7.18-7.16 ppm (black, A2). Thus, the spectra with the higher peak area of the singlet are clustered in two sub-groups depending on the shift. B) PCA score plot of equidistantly bucketed simulated ^1H NMR spectra; C) PCA score plot of simulated ^1H NMR spectra bucketed with the optimized algorithm and using a slackness of 50%; D) PCA score plot of simulated ^1H NMR spectra bucketed with the optimized algorithm and using a slackness of 75%. All spectra have been bucketed into 0.02 ppm segments. Each dot represents a bucketed spectrum, colour-coded by the different shifts distribution of the singlet resonance.

Chapter 5

Conclusions

Biomedical research continuously relies on advances in -omics sciences thanks to the progressive understanding of the biology, aetiology of diseases and the possibility of developing novel diagnostic test and/or therapeutic treatments. In this framework, metabolomics is featured prominently, due to its important role in connecting the genotype with the phenotype. The results illustrated in this three-years PhD thesis mainly exalted the potentiality of untargeted NMR-based metabolomics for different biomedical applications, covering both human and veterinary sciences, also highlighting the fundamental synergy between chemistry, biochemistry and bioinformatic tools for a better understanding of the biomedical goals.

Results obtained in each presented study answered important questions and opened interesting perspectives, either based on the whole molecular fingerprint of the condition of interest or based on selected metabolic features.

First of all, results here reported, bring metabolomics by NMR closer to its adoption in the clinical field, although some limitations should be mentioned.

Overall, presented results showed how identified NMR-based signatures could potentially find future clinical applications. For example, we can envisage applications to early diagnosis of a disease condition, or to the monitoring of the health status of a subject in the frame of a more personalized healthcare. Indeed, a variegated ensemble of both low-molecular-weight metabolites and high-molecular-weight molecules can indicate biological changes in the host due to perturbations in different metabolic pathways. However, to become a clinically approved test, a potential biomarker, both considered as a specific single NMR feature or as the whole metabolic profile, should be confirmed and validated using hundreds of specimens and it should be reproducible, robust, specific and sensitive. In this perspective, some of the presented studies showed limitations, due to the small number of available samples or the lack of optimal control groups and external validation cohorts. Moreover, even if body fluids such as plasma, serum and urine are considered to be ideal for biomarker discovery and monitoring, a number of potential metabolomics pitfalls lurk in these fluids. Indeed, some factors, like the patients' lifestyle or the nutritional status might be reflected in the metabolite composition of the chosen body fluid, potentially masking the molecular changes caused by the disease; but at the same time, the disease itself will influence the patients' habits and food intake, thus representing for many disease signatures the classical "chicken and eggs" paradox, like for example in the case of Parkinson's disease.

More in detail, from the results reported in § 4.1.1., it clearly appeared that NMR fingerprinting succeeded in characterizing different signatures of Parkinson's disease, describing, from a molecular point of view, the early and the advanced stage of the same pathology. In particular, serum signature differentiating *de novo* untreated

Parkinson's disease patients from healthy matched controls has been shown to be strong, since our results have been validated by the use of an external validation cohort. This has represented the major strength of this study, highlighting the potential concrete use of serum NMR fingerprints of Parkinson's disease for further clinical applications, especially for early diagnosis of the disease before motor symptoms occur.

To our knowledge, this study also represented the first large-scale study in this field, therefore representing a step further towards the increasing request of analytical validation for metabolomic studies. Indeed, despite metabolomics is on-going in terms of technology and computational improvements, the literature features relatively high numbers of small-scale or preliminary-type studies, with many of them suffering from a lack of statistical robustness and validity.

Results reported in § 4.1.2 and § 4.1.3 highlighted the importance of undertaking large multi-centre cohort studies to enhance the discovery of metabolic features that have good prospects to be translated into point of care and rapid diagnostics. Indeed, in these studies serum samples collected from more than ten different hospitals have been analysed. By applying both standard statistical techniques and a more systematic networking approach, we reported various analytes able to statistically increase the predictive ability of already known clinical factors, for instance regarding the prediction of poor outcomes after acute ischemic stroke. Moreover, our results pointed out some dysregulated mechanisms, potentially involved in the progression of the pathology after the thrombolytic treatment, affecting survivors' outcomes and their quality of life. Again, the studies presented in the above-mentioned sections also suffer from limitations due to the obvious lack of a control group of "not-treated" subjects. Despite this, the reported results do support the usefulness of NMR-based metabolomics in identifying a more-detailed risk profile in stroke patients.

Personalised medicine is relatively new to the field of healthcare research, and for decades it has been practiced within a so-called "evidence-based framework", where the individual is treated for the condition of interest, mostly on the basis of popular medicine. More recently, personalized medicine, also called precision medicine, involves assessing the genotype and the phenotype of the patient, before they undergo any treatment. Metabolomics plays a key role in this field, as a potential approach to provide pivotal biomarkers, to test their detectability within large and diverse cohorts and then translating results into cheap, fast and reliable methods. Data reported in § 4.1.4 are an example of how metabolomics fits well with precision medicine goals. In particular, results here reported, have exalted the role of NMR-based metabolomics in phenotyping aggressive prostate cancer in South African men, evidencing inflammation as a key-driving factor and paving the way to new tailored therapeutic strategies. A further limitation of this study rests upon the small number of subjects enrolled, but it preliminarily answers the urgent need for new insights into the molecular mechanisms underlying the remarkably increased rate of aggressive and lethal prostate cancer (PCa) in men of African ancestry that are more likely to develop aggressive PCa and to die from this disease.

However, the focus of metabolic signatures or specific biomarker discovery should not only be for pathological cures, but also for preventive screening of healthy individuals who may be susceptible to a certain disease. The study reported in § 4.1.5 is in line with this aim; indeed, the achieved results evidenced the potentiality of untargeted NMR-based metabolomics in characterizing serum metabolic variations of healthy volunteers at risk for Metabolic Syndrome (MetS), subsequent to the administration of selected nutraceuticals combinations, embedded in different bioactive-enriched foods. The randomized and double-blinded administration of placebo foods represented a major strength point of this study, permitting to fully explore the effect of different food matrices. Ideally, our findings should be integrated with clinical information to better assess how the usual dietary habits of the volunteers or any other lifestyle factor might influence the discovered serum signatures.

Beyond human medicine applications, this thesis proved the great potentiality of untargeted NMR-based metabolomics also for veterinary research; indeed results reported in section 4.2 have evidenced the role of the untargeted approach by NMR in: *i*) characterizing, from a metabolic point of view, the left and right abomasal displacement of dairy cows; *ii*) monitoring the health status of preterm calves, to prevent early death and incoming pathologies; and *iii*) discovering metabolic signatures able to discriminate healthy dogs from dogs affected by *Ehrlichia Canis*.

In the light of the outcomes of the different studies, this thesis has demonstrated the crucial role of NMR-based metabolomics to determine new metabolic features of diseases or the whole metabolic fingerprints that, taking the definition to the extreme, it would be “the best biomarker”, both simple and robust, to be used in the medical field. In particular, with advantages of minimal sample preparation, high throughput, high reproducibility and high accuracy, NMR-based metabolomics analyses provide great potentialities for early disease diagnosis with respect to classical clinic tests. Also considering that NMR instruments are long-lived and procedures are generally cost-effective, the overall cost per sample (when all operations are optimized) is affordable. Therefore, a metabolomic NMR analysis on, for example, a serum sample would add only a modest amount to the cost of a routine blood test, but a significant amount of new information.

Given the high potentialities of the untargeted metabolomics and the huge breath of its applications, ranging from medical to food and environmental researches, two different approaches, *i.e.* fingerprinting and profiling, have been developed to satisfy different aims over past years. However, the increasing interest in quantifiable properties, especially from the medical community, probably created a bit of confusion between the specific aim of the two distinct approaches and the related tools to achieve them. As an example, criticisms on the main drawbacks of the bucketing of NMR spectra for profiling analyses have been implicitly extended to the fingerprinting goals. Therefore, we reasoned that there might be a place in the literature for an update in this field, demonstrating how the equidistant bucketing procedure is simple and robust to perform NMR-based fingerprinting analyses.

Results presented in § 4.3.1 reached this point. Translating them into future perspectives, we advocate a possibility of performing fast and cheap low-field fingerprinting of diseases, especially in those areas where the availability and the affordability of more expensive analytical techniques are not granted.

However, it should not be forgotten that one of the most important challenges of metabolomic research is its integration with classical clinical tools and, in this perspective, the on-going studies presented here have the potential to be fruitfully integrated with specific clinical information.

In conclusion, even if a significant fraction of the presented material is still in preparation and touches on variegated research areas, this thesis may contribute to the demonstration that untargeted NMR-based metabolomics, coupled with bioinformatic tools and statistical analyses, can be considered as a comprehensive analytical technique with reasonable and actual prospects of being implemented in biomedical research.

Bibliography

- (1) Nicholson, J. K.; Lindon, J. C. Systems Biology: Metabonomics. *Nature* **2008**, *455* (1476-4687 (Electronic)), 1054–1056.
- (2) Westerhoff, H. V.; Alberghina, L. Systems Biology: Did We Know It All Along? In *Systems Biology: Definitions and Perspectives*; Alberghina, L., Westerhoff, H. V., Eds.; Topics in Current Genetics; Springer: Berlin, Heidelberg, 2005; pp 3–9. <https://doi.org/10.1007/b137744>.
- (3) Vignoli, A.; Ghini, V.; Meoni, G.; Licari, C.; Takis, P. G.; Tenori, L.; Turano, P.; Luchinat, C. High-Throughput Metabolomics by 1D NMR. *Angew. Chem. Int. Ed Engl.* **2019**, *58* (4), 968–994. <https://doi.org/10.1002/anie.201804736>.
- (4) Nielsen, J.; Oliver, S. The next Wave in Metabolome Analysis. *Trends Biotechnol.* **2005**, *23* (11), 544–546. <https://doi.org/10.1016/j.tibtech.2005.08.005>.
- (5) Fiehn, O. Combining Genomics, Metabolome Analysis, and Biochemical Modelling to Understand Metabolic Networks. *Int. J. Genomics* **2001**, *2* (3), 155–168. <https://doi.org/10.1002/cfg.82>.
- (6) Wishart, D. S. Emerging Applications of Metabolomics in Drug Discovery and Precision Medicine. *Nat. Rev. Drug Discov.* **2016**, *15* (7), 473–484. <https://doi.org/10.1038/nrd.2016.32>.
- (7) Jones, O. A. H. Illuminating the Dark Metabolome to Advance the Molecular Characterisation of Biological Systems. *Metabolomics* **2018**, *14* (8), 101. <https://doi.org/10.1007/s11306-018-1396-y>.
- (8) Nagana Gowda, G. A.; Raftery, D. Recent Advances in NMR-Based Metabolomics. *Anal. Chem.* **2017**, *89* (1), 490–510. <https://doi.org/10.1021/acs.analchem.6b04420>.
- (9) Barbosa, B. S.; Martins, L. G.; Costa, T. B. B. C.; Cruz, G.; Tasic, L. Qualitative and Quantitative NMR Approaches in Blood Serum Lipidomics. *Methods Mol. Biol. Clifton NJ* **2018**, *1735*, 365–379. https://doi.org/10.1007/978-1-4939-7614-0_25.
- (10) Li, J.; Vosegaard, T.; Guo, Z. Applications of Nuclear Magnetic Resonance in Lipid Analyses: An Emerging Powerful Tool for Lipidomics Studies. *Prog. Lipid Res.* **2017**, *68*, 37–56. <https://doi.org/10.1016/j.plipres.2017.09.003>.
- (11) Bouatra, S.; Aziat, F.; Mandal, R.; Guo, A. C.; Wilson, M. R.; Knox, C.; Bjorn Dahl, T. C.; Krishnamurthy, R.; Saleem, F.; Liu, P.; Dame, Z. T.; Poelzer, J.; Huynh, J.; Yallou, F. S.; Psychogios, N.; Dong, E.; Bogumil, R.; Roehring, C.; Wishart, D. S. The Human Urine Metabolome. *PLoS ONE* **2013**, *8* (9), e73076. <https://doi.org/10.1371/journal.pone.0073076>.
- (12) Psychogios, N.; Hau, D. D.; Peng, J.; Guo, A. C.; Mandal, R.; Bouatra, S.; Sinelnikov, I.; Krishnamurthy, R.; Eisner, R.; Gautam, B.; Young, N.; Xia, J.; Knox, C.; Dong, E.; Huang, P.; Hollander, Z.; Pedersen, T. L.; Smith, S. R.; Bamforth, F.; Greiner, R.; McManus, B.; Newman, J. W.; Goodfriend, T.; Wishart, D. S. The Human Serum Metabolome. *PLoS ONE* **2011**, *6* (2), e16957. <https://doi.org/10.1371/journal.pone.0016957>.

- (13) Dame, Z. T.; Aziat, F.; Mandal, R.; Krishnamurthy, R.; Bouatra, S.; Borzouie, S.; Guo, A. C.; Sajed, T.; Deng, L.; Lin, H.; Liu, P.; Dong, E.; Wishart, D. S. The Human Saliva Metabolome. *Metabolomics* **2015**, *11* (6), 1864–1883. <https://doi.org/10.1007/s11306-015-0840-5>.
- (14) Wishart, D. S.; Lewis, M. J.; Morrissey, J. A.; Flegel, M. D.; Jeroncic, K.; Xiong, Y.; Cheng, D.; Eisner, R.; Gautam, B.; Tzur, D.; Sawhney, S.; Bamforth, F.; Greiner, R.; Li, L. The Human Cerebrospinal Fluid Metabolome. *J. Chromatogr. B* **2008**, *871* (2), 164–173. <https://doi.org/10.1016/j.jchromb.2008.05.001>.
- (15) Airoldi, C.; Ciaramelli, C.; Fumagalli, M.; Bussei, R.; Mazzoni, V.; Viglio, S.; Iadarola, P.; Stolk, J. 1H NMR To Explore the Metabolome of Exhaled Breath Condensate in A1-Antitrypsin Deficient Patients: A Pilot Study. *J. Proteome Res.* **2016**, *15* (12), 4569–4578. <https://doi.org/10.1021/acs.jproteome.6b00648>.
- (16) Bernacchioni, C.; Ghini, V.; Cencetti, F.; Japtok, L.; Donati, C.; Bruni, P.; Turano, P. NMR Metabolomics Highlights Sphingosine Kinase-1 as a New Molecular Switch in the Orchestration of Aberrant Metabolic Phenotype in Cancer Cells. *Mol. Oncol.* **2017**, n/a-n/a. <https://doi.org/10.1002/1878-0261.12048>.
- (17) Cuperlovic-Culf, M.; Ferguson, D.; Culf, A.; Morin, P.; Touaibia, M. 1H NMR Metabolomics Analysis of Glioblastoma Subtypes CORRELATION BETWEEN METABOLOMICS AND GENE EXPRESSION CHARACTERISTICS. *J. Biol. Chem.* **2012**, *287* (24), 20164–20175. <https://doi.org/10.1074/jbc.M111.337196>.
- (18) Battini, S.; Faitot, F.; Imperiale, A.; Cicek, A. E.; Heimbürger, C.; Averous, G.; Bachellier, P.; Namer, I. J. Metabolomics Approaches in Pancreatic Adenocarcinoma: Tumor Metabolism Profiling Predicts Clinical Outcome of Patients. *BMC Med.* **2017**, *15* (1), 56. <https://doi.org/10.1186/s12916-017-0810-z>.
- (19) Martínez-Granados, B.; Monleón, D.; Martínez-Bisbal, M. C.; Rodrigo, J. M.; Olmo, J. del; Lluch, P.; Ferrández, A.; Martí-Bonmatí, L.; Celda, B. Metabolite Identification in Human Liver Needle Biopsies by High-Resolution Magic Angle Spinning 1H NMR Spectroscopy. *NMR Biomed.* **2006**, *19* (1), 90–100. <https://doi.org/10.1002/nbm.1005>.
- (20) Palama, T. L.; Canard, I.; Rautureau, G. J. P.; Mirande, C.; Chatellier, S.; Elena-Herrmann, B. Identification of Bacterial Species by Untargeted NMR Spectroscopy of the Exo-Metabolome. *The Analyst* **2016**, *141* (15), 4558–4561. <https://doi.org/10.1039/c6an00393a>.
- (21) Romano, F.; Meoni, G.; Manavella, V.; Baima, G.; Tenori, L.; Cacciatore, S.; Aimetti, M. Analysis of Salivary Phenotypes of Generalized Aggressive and Chronic Periodontitis through Nuclear Magnetic Resonance-Based Metabolomics. *J. Periodontol.* **2018**. <https://doi.org/10.1002/JPER.18-0097>.
- (22) Kumari, S.; Goyal, V.; Kumaran, S. S.; Dwivedi, S. N.; Srivastava, A.; Jagannathan, N. R. Quantitative Metabolomics of Saliva Using Proton NMR Spectroscopy in Patients with Parkinson's Disease and Healthy Controls. *Neurol. Sci. Off. J. Ital. Neurol. Soc. Ital. Soc. Clin. Neurophysiol.* **2020**, *41* (5), 1201–1210. <https://doi.org/10.1007/s10072-019-04143-4>.
- (23) Čuperlović-Culf, M. *NMR Metabolomics in Cancer Research*; Elsevier, 2012.

- (24) Giskeødegård, G. F.; Madssen, T. S.; Euceda, L. R.; Tessem, M.-B.; Moestue, S. A.; Bathen, T. F. NMR-based metabolomics of biofluids in cancer. *NMR Biomed.* **0** (0), e3927. <https://doi.org/10.1002/nbm.3927>.
- (25) Cheng, L. L.; Pohl, U. Chapter 13 - The Role of NMR-Based Metabolomics in Cancer. In *The Handbook of Metabonomics and Metabolomics*; Elsevier Science B.V.: Amsterdam, 2007; pp 345–374.
- (26) Ludwig, C.; Ward, D. G.; Martin, A.; Viant, M. R.; Ismail, T.; Johnson, P. J.; Wakelam, M. J. O.; Günther, U. L. Fast Targeted Multidimensional NMR Metabolomics of Colorectal Cancer. *Magn. Reson. Chem. MRC* **2009**, *47 Suppl 1*, S68-73. <https://doi.org/10.1002/mrc.2519>.
- (27) Tenori, L.; Hu, X.; Pantaleo, P.; Alterini, B.; Castelli, G.; Olivotto, I.; Bertini, I.; Luchinat, C.; Gensini, G. F. Metabolomic Fingerprint of Heart Failure in Humans: A Nuclear Magnetic Resonance Spectroscopy Analysis. *Int. J. Cardiol.* **2013**, *168* (4), e113-115. <https://doi.org/10.1016/j.ijcard.2013.08.042>.
- (28) Vignoli, A.; Tenori, L.; Giusti, B.; Takis, P. G.; Valente, S.; Carrabba, N.; Balzi, D.; Barchielli, A.; Marchionni, N.; Gensini, G. F.; Marcucci, R.; Luchinat, C.; Gori, A. M. NMR-Based Metabolomics Identifies Patients at High Risk of Death within Two Years after Acute Myocardial Infarction in the AMI-Florence II Cohort. *BMC Med.* **2019**, *17* (1), 3. <https://doi.org/10.1186/s12916-018-1240-2>.
- (29) Stryeck, S.; Gastrager, M.; Degoricija, V.; Trbušić, M.; Potočnjak, I.; Radulović, B.; Pregartner, G.; Berghold, A.; Madl, T.; Frank, S. Serum Concentrations of Citrate, Tyrosine, 2- and 3- Hydroxybutyrate Are Associated with Increased 3-Month Mortality in Acute Heart Failure Patients. *Sci. Rep.* **2019**, *9* (1), 6743. <https://doi.org/10.1038/s41598-019-42937-w>.
- (30) Wesley, U. V.; Bhute, V. J.; Hatcher, J. F.; Palecek, S. P.; Dempsey, R. J. Local and Systemic Metabolic Alterations in Brain, Plasma, and Liver of Rats in Response to Aging and Ischemic Stroke, as Detected by Nuclear Magnetic Resonance (NMR) Spectroscopy. *Neurochem. Int.* **2019**, *127*, 113–124. <https://doi.org/10.1016/j.neuint.2019.01.025>.
- (31) Jung, J. Y.; Lee, H.-S.; Kang, D.-G.; Kim, N. S.; Cha, M. H.; Bang, O.-S.; Ryu, D. H.; Hwang, G.-S. ¹H-NMR-Based Metabolomics Study of Cerebral Infarction. *Stroke* **2011**, *42* (5), 1282–1288. <https://doi.org/10.1161/STROKEAHA.110.598789>.
- (32) Zhang, A.; Sun, H.; Wang, X. Recent Advances in Metabolomics in Neurological Disease, and Future Perspectives. *Anal. Bioanal. Chem.* **2013**, *405* (25), 8143–8150. <https://doi.org/10.1007/s00216-013-7061-4>.
- (33) Lei, S.; Powers, R. NMR Metabolomics Analysis of Parkinson's Disease. *Curr. Metabolomics* **2013**, *1* (3), 191–209. <https://doi.org/10.2174/2213235X113019990004>.
- (34) Everett, J. R. Pharmacometabonomics in Humans: A New Tool for Personalized Medicine. *Pharmacogenomics* **2015**, *16* (7), 737–754. <https://doi.org/10.2217/pgs.15.20>.
- (35) Everett, J. R. NMR-Based Pharmacometabonomics: A New Paradigm for Personalised or Precision Medicine. *Prog. Nucl. Magn. Reson. Spectrosc.* **2017**, *102–103*, 1–14. <https://doi.org/10.1016/j.pnmrs.2017.04.003>.
- (36) Gralka, E.; Luchinat, C.; Tenori, L.; Ernst, B.; Thurnheer, M.; Schultes, B. Metabolomic Fingerprint of Severe Obesity Is Dynamically Affected by

- Bariatric Surgery in a Procedure-Dependent Manner. *Am. J. Clin. Nutr.* **2015**, *102* (6), 1313–1322. <https://doi.org/10.3945/ajcn.115.110536>.
- (37) Casella, G.; Cavarretta, E.; Tenori, L.; Frati, G.; Luchinat, C.; Banci, L.; Redler, A.; Giannotti, D.; Condorelli, G.; Basso, N. Circulating Amino Acids Profile Before and After Sleeve Gastrectomy: A Metabolomic Study. *Obes. Surg.* **2013**, *23* (8), 1048–1049.
- (38) Tenori, L.; Oakman, C.; Morris, P. G.; Gralka, E.; Turner, N.; Cappadona, S.; Fornier, M.; Hudis, C.; Norton, L.; Luchinat, C.; Di Leo, A. Serum Metabolomic Profiles Evaluated after Surgery May Identify Patients with Oestrogen Receptor Negative Early Breast Cancer at Increased Risk of Disease Recurrence. Results from a Retrospective Study. *Mol. Oncol.* **2015**, *9* (1), 128–139. <https://doi.org/10.1016/j.molonc.2014.07.012>.
- (39) Gibney, M. J.; Walsh, M.; Brennan, L.; Roche, H. M.; German, B.; van Ommen, B. Metabolomics in Human Nutrition: Opportunities and Challenges. *Am. J. Clin. Nutr.* **2005**, *82* (3), 497–503.
- (40) Brennan, L. Metabolomics in Nutrition Research: Current Status and Perspectives. *Biochem. Soc. Trans.* **2013**, *41* (2), 670–673. <https://doi.org/10.1042/BST20120350>.
- (41) O’Sullivan, A.; Gibney, M. J.; Brennan, L. Dietary Intake Patterns Are Reflected in Metabolomic Profiles: Potential Role in Dietary Assessment Studies. *Am. J. Clin. Nutr.* **2011**, *93* (2), 314–321. <https://doi.org/10.3945/ajcn.110.000950>.
- (42) Kirwan, J. Metabolomics for the Practising Vet. *In Pract.* **2013**, *35* (8), 438–445. <https://doi.org/10.1136/inp.f5259>.
- (43) Basoglu, A.; Baspinar, N.; Tenori, L.; Vignoli, A.; Yildiz, R. Plasma Metabolomics in Calves with Acute Bronchopneumonia. *Metabolomics* **2016**, *12* (8), 128. <https://doi.org/10.1007/s11306-016-1074-x>.
- (44) Basoglu, A.; Baspinar, N.; Tenori, L.; Vignoli, A.; Gulersoy, E. Effects of Boron Supplementation on Peripartum Dairy Cows’ Health. *Biol. Trace Elem. Res.* **2017**, 1–8. <https://doi.org/10.1007/s12011-017-0971-9>.
- (45) Basoglu, A.; Baspinar, N.; Tenori, L.; Licari, C.; Gulersoy, E. Nuclear Magnetic Resonance (NMR)-Based Metabolome Profile Evaluation in Dairy Cows with and without Displaced Abomasum. *Vet. Q.* **2020**, *40* (1), 1–15. <https://doi.org/10.1080/01652176.2019.1707907>.
- (46) Bernini, P.; Bertini, I.; Luchinat, C.; Nincheri, P.; Staderini, S.; Turano, P. Standard Operating Procedures for Pre-Analytical Handling of Blood and Urine for Metabolomic Studies and Biobanks. *J. Biomol. NMR* **2011**, *49* (3–4), 231–243. <https://doi.org/10.1007/s10858-011-9489-1>.
- (47) Ghini, V.; Quaglio, D.; Luchinat, C.; Turano, P. NMR for Sample Quality Assessment in Metabolomics. *New Biotechnol.* **2019**, *52*, 25–34. <https://doi.org/10.1016/j.nbt.2019.04.004>.
- (48) Wishart, D. S.; Tzur, D.; Knox, C.; Eisner, R.; Guo, A. C.; Young, N.; Cheng, D.; Jewell, K.; Arndt, D.; Sawhney, S.; Fung, C.; Nikolai, L.; Lewis, M.; Coutouly, M.-A.; Forsythe, I.; Tang, P.; Shrivastava, S.; Jeroncic, K.; Stothard, P.; Amegbey, G.; Block, D.; Hau, D. D.; Wagner, J.; Miniaci, J.; Clements, M.; Gebremedhin, M.; Guo, N.; Zhang, Y.; Duggan, G. E.; Macinnis, G. D.; Weljie, A. M.; Dowlatabadi, R.; Bamforth, F.; Clive, D.; Greiner, R.; Li, L.; Marrie, T.; Sykes, B. D.; Vogel, H. J.; Querengesser, L. HMDB: The Human Metabolome

- Database. *Nucleic Acids Res.* **2007**, *35* (Database issue), D521-526. <https://doi.org/10.1093/nar/gkl923>.
- (49) Kanehisa, M.; Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **2000**, *28* (1), 27–30. <https://doi.org/10.1093/nar/28.1.27>.
- (50) Wang, Y.; Xiao, J.; Suzek, T. O.; Zhang, J.; Wang, J.; Bryant, S. H. PubChem: A Public Information System for Analyzing Bioactivities of Small Molecules. *Nucleic Acids Res.* **2009**, *37* (Web Server issue), W623-633. <https://doi.org/10.1093/nar/gkp456>.
- (51) Kale, N. S.; Haug, K.; Conesa, P.; Jayseelan, K.; Moreno, P.; Rocca-Serra, P.; Nainala, V. C.; Spicer, R. A.; Williams, M.; Li, X.; Salek, R. M.; Griffin, J. L.; Steinbeck, C. MetaboLights: An Open-Access Database Repository for Metabolomics Data. *Curr. Protoc. Bioinforma.* **2016**, *53*, 14.13.1-18. <https://doi.org/10.1002/0471250953.bi1413s53>.
- (52) Xia, J.; Psychogios, N.; Young, N.; Wishart, D. S. MetaboAnalyst: A Web Server for Metabolomic Data Analysis and Interpretation. *Nucleic Acids Res.* **2009**, *37* (suppl 2), W652–W652.
- (53) Lindon, J. C.; Nicholson, J. K.; Holmes, E. *The Handbook of Metabonomics and Metabolomics*; Elsevier, 2011.
- (54) Takis, P. G.; Ghini, V.; Tenori, L.; Turano, P.; Luchinat, C. Uniqueness of the NMR Approach to Metabolomics. *TrAC Trends Anal. Chem.* **2019**, *120*, 115300. <https://doi.org/10.1016/j.trac.2018.10.036>.
- (55) Krishnan, P.; Kruger, N. J.; Ratcliffe, R. G. Metabolite Fingerprinting and Profiling in Plants Using NMR. *J. Exp. Bot.* **2005**, *56* (410), 255–265. <https://doi.org/10.1093/jxb/eri010>.
- (56) Keun, H. C.; Liu, M.; Price, W.; Dona, A.; Elena-herrmann, B. *Nmr-based Metabolomics*; Royal Society of Chemistry: Cambridge, United Kingdom?, 2018.
- (57) Piotto, M.; Saudek, V.; Sklenář, V. Gradient-Tailored Excitation for Single-Quantum NMR Spectroscopy of Aqueous Solutions. *J. Biomol. NMR* **1992**, *2* (6), 661–665. <https://doi.org/10.1007/BF02192855>.
- (58) 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), 2692–2703. <https://doi.org/10.1038/nprot.2007.376>.
- (59) Barding, G. A.; Salditos, R.; Larive, C. K. Quantitative NMR for Bioanalysis and Metabolomics. *Anal. Bioanal. Chem.* **2012**, *404* (4), 1165–1179. <https://doi.org/10.1007/s00216-012-6188-z>.
- (60) McKenzie, J. S.; Donarski, J. A.; Wilson, J. C.; Charlton, A. J. Analysis of Complex Mixtures Using High-Resolution Nuclear Magnetic Resonance Spectroscopy and Chemometrics. *Prog. Nucl. Magn. Reson. Spectrosc.* **2011**, *59* (4), 336–359. <https://doi.org/10.1016/j.pnmrs.2011.04.003>.
- (61) 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. <https://doi.org/10.1002/cmr.a.20223>.

- (62) Wu, D. H.; Chen, A. D.; Johnson, C. S. An Improved Diffusion-Ordered Spectroscopy Experiment Incorporating Bipolar-Gradient Pulses. *J. Magn. Reson. A* **1995**, *115* (2), 260–264. <https://doi.org/10.1006/jmra.1995.1176>.
- (63) Carr, H. Y.; Purcell, E. M. Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments. *Phys. Rev.* **1954**, *94* (3), 630–638. <https://doi.org/10.1103/PhysRev.94.630>.
- (64) Ravanbakhsh, S.; Liu, P.; Bjordahl, T. C.; Mandal, R.; Grant, J. R.; Wilson, M.; Eisner, R.; Snelnikov, I.; Hu, X.; Luchinat, C.; Greiner, R.; Wishart, D. S. Accurate, Fully-Automated NMR Spectral Profiling for Metabolomics. *PLoS ONE* **2015**, *10* (5), e0124219. <https://doi.org/10.1371/journal.pone.0124219>.
- (65) Pearce, J. T. M.; Athersuch, T. J.; Ebbels, T. M. D.; Lindon, J. C.; Nicholson, J. K.; Keun, H. C. Robust Algorithms for Automated Chemical Shift Calibration of 1D 1H NMR Spectra of Blood Serum. *Anal. Chem.* **2008**, *80* (18), 7158–7162. <https://doi.org/10.1021/ac8011494>.
- (66) Wider, G.; Dreier, L. Measuring Protein Concentrations by NMR Spectroscopy. *J. Am. Chem. Soc.* **2006**, *128* (8), 2571–2576. <https://doi.org/10.1021/ja055336t>.
- (67) Akoka, S.; Barantin, L.; Trierweiler, M. Concentration Measurement by Proton NMR Using the ERETIC Method. *Anal. Chem.* **1999**, *71* (13), 2554–2557. <https://doi.org/10.1021/ac981422i>.
- (68) 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. <https://doi.org/10.1016/j.jmr.2009.11.012>.
- (69) Smolinska, A.; Blanchet, L.; Buydens, L. M. C.; Wijmenga, S. S. NMR and Pattern Recognition Methods in Metabolomics: From Data Acquisition to Biomarker Discovery: A Review. *Anal. Chim. Acta* **2012**, *750*, 82–97. <https://doi.org/10.1016/j.aca.2012.05.049>.
- (70) Wu, Y.; Li, L. Sample Normalization Methods in Quantitative Metabolomics. *J. Chromatogr. A* **2016**, *1430*, 80–95. <https://doi.org/10.1016/j.chroma.2015.12.007>.
- (71) 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 Metabolomics. *Anal. Chem.* **2006**, *78* (13), 4281–4290. <https://doi.org/10.1021/ac051632c>.
- (72) Wold, S.; Esbensen, K.; Geladi, P. Principal Component Analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2* (1–3), 37–52. [https://doi.org/10.1016/0169-7439\(87\)80084-9](https://doi.org/10.1016/0169-7439(87)80084-9).
- (73) Geladi, P. Herman Wold, the Father of PLS. *Chemom. Intell. Lab. Syst.* **1992**, *15* (1), vii–viii.
- (74) Wold, S.; Eriksson, L. PLS-Regression: A Basic Tool of Chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58* (2), 109–130.
- (75) Abdi, H. Partial Least Squares Regression and Projection on Latent Structure Regression (PLS Regression). *Wiley Interdiscip. Rev. Comput. Stat.* **2010**, *2* (1), 97–106.
- (76) Trygg, J.; Wold, S. Orthogonal Projections to Latent Structures (O-PLS). *J. Chemom.* **2002**, *16* (3), 119–128. <https://doi.org/10.1002/cem.695>.

- (77) Westerhuis, J. A.; van Velzen, E. J.; Hoefsloot, H. C.; Smilde, A. K. Multivariate Paired Data Analysis: Multilevel PLSDA versus OPLSDA. *Metabolomics* **2010**, *6* (1573-3890 (Electronic)), 119–128.
- (78) Cover, T.; Hart, P. Nearest Neighbor Pattern Classification. *IEEE Trans. Inf. Theory* **1967**, *13* (1), 21–27. <https://doi.org/10.1109/TIT.1967.1053964>.
- (79) Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20* (3), 273–297. <https://doi.org/10.1007/BF00994018>.
- (80) Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45* (1), 5–32. <https://doi.org/10.1023/A:1010933404324>.
- (81) Liaw, A.; Wiener, M. Classification and Regression by RandomForest. *R News* **2002**, *2* (3), 18–22.
- (82) Chen, T.; Cao, Y.; Zhang, Y.; Liu, J.; Bao, Y.; Wang, C.; Jia, W.; Zhao, A. Random Forest in Clinical Metabolomics for Phenotypic Discrimination and Biomarker Selection. *Evid.-Based Complement. Altern. Med. ECAM* **2013**, *2013*, 298183. <https://doi.org/10.1155/2013/298183>.
- (83) Touw, W. G.; Bayjanov, J. R.; Overmars, L.; Backus, L.; Boekhorst, J.; Wels, M.; Hijum, V.; T, S. A. F. Data Mining in the Life Sciences with Random Forest: A Walk in the Park or Lost in the Jungle? *Brief. Bioinform.* **2013**, *14* (3), 315–326. <https://doi.org/10.1093/bib/bbs034>.
- (84) Hao, J.; Liebeke, M.; Astle, W.; De Iorio, M.; Bundy, J. G.; Ebbels, T. M. D. Bayesian Deconvolution and Quantification of Metabolites in Complex 1D NMR Spectra Using BATMAN. *Nat. Protoc.* **2014**, *9* (6), 1416–1427. <https://doi.org/10.1038/nprot.2014.090>.
- (85) Tardivel, P. J. C.; Canlet, C.; Lefort, G.; Tremblay-Franco, M.; Debrauwer, L.; Concordet, D.; Servien, R. ASICS: An Automatic Method for Identification and Quantification of Metabolites in Complex 1D ¹H NMR Spectra. *Metabolomics* **2017**, *13* (10), 109. <https://doi.org/10.1007/s11306-017-1244-5>.
- (86) Neuhäuser, M. Wilcoxon–Mann–Whitney Test. In *International Encyclopedia of Statistical Science*; Springer, Berlin, Heidelberg, 2011; pp 1656–1658. https://doi.org/10.1007/978-3-642-04898-2_615.
- (87) Snedecor, G. W.; Cochran, W. G. *Statistical Methods*, 8 edition.; Iowa State University Press: Ames, Iowa, 1989.
- (88) Kruskal, W. H.; Wallis, W. A. Use of Ranks in One-Criterion Variance Analysis. *J. Am. Stat. Assoc.* **1952**, *47* (260), 583–621. <https://doi.org/10.1080/01621459.1952.10483441>.
- (89) Conover, W. J. *Practical Nonparametric Statistics, 3rd*, 3rd edition.; Wiley: New York, 1999.
- (90) Bonferroni, C. E. Il Calcolo Delle Assicurazioni Su Gruppi Di Teste. In *Studi in Onore del Professore Salvatore Ortu Carboni*; Rome, 1935; pp 13–60.
- (91) Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **1995**, 289–300.
- (92) Vidgen, B.; Yasseri, T. P-Values: Misunderstood and Misused. *Front. Phys.* **2016**, *4*. <https://doi.org/10.3389/fphy.2016.00006>.
- (93) Sullivan, G. M.; Feinn, R. Using Effect Size—or Why the P Value Is Not Enough. *J. Grad. Med. Educ.* **2012**, *4* (3), 279–282. <https://doi.org/10.4300/JGME-D-12-00156.1>.

- (94) Vignoli, A.; Tenori, L.; Luchinat, C.; Saccenti, E. Age and Sex Effects on Plasma Metabolite Association Networks in Healthy Subjects. *J. Proteome Res.* **2018**, *17* (1), 97–107. <https://doi.org/10.1021/acs.jproteome.7b00404>.
- (95) Vignoli, A.; Tenori, L.; Giusti, B.; Valente, S.; Carrabba, N.; Balzi, D.; Barchielli, A.; Marchionni, N.; Gensini, G. F.; Marcucci, R.; Gori, A. M.; Luchinat, C.; Saccenti, E. Differential Network Analysis Reveals Metabolic Determinants Associated with Mortality in Acute Myocardial Infarction Patients and Suggests Potential Mechanisms Underlying Different Clinical Scores Used To Predict Death. *J. Proteome Res.* **2020**, *19* (2), 949–961. <https://doi.org/10.1021/acs.jproteome.9b00779>.
- (96) Saccenti, E.; Menichetti, G.; Ghini, V.; Remondini, D.; Tenori, L.; Luchinat, C. Entropy-Based Network Representation of the Individual Metabolic Phenotype. *J. Proteome Res.* **2016**, *15* (9), 3298–3307. <https://doi.org/10.1021/acs.jproteome.6b00454>.
- (97) Franceschi, C.; Campisi, J. Chronic Inflammation (Inflammaging) and Its Potential Contribution to Age-Associated Diseases. *J. Gerontol. A. Biol. Sci. Med. Sci.* **2014**, *69 Suppl 1*, S4-9. <https://doi.org/10.1093/gerona/glu057>.
- (98) Cerri, S.; Mus, L.; Blandini, F. Parkinson's Disease in Women and Men: What's the Difference? *J. Park. Dis.* **2019**, *9* (3), 501–515. <https://doi.org/10.3233/JPD-191683>.
- (99) Gori, A. M.; Giusti, B.; Piccardi, B.; Nencini, P.; Palumbo, V.; Nesi, M.; Nucera, A.; Pracucci, G.; Tonelli, P.; Innocenti, E.; Sereni, A.; Sticchi, E.; Toni, D.; Bovi, P.; Guidotti, M.; Tola, M. R.; Consoli, D.; Micieli, G.; Tassi, R.; Orlandi, G.; Sessa, M.; Perini, F.; Delodovici, M. L.; Zedde, M. L.; Massaro, F.; Abbate, R.; Inzitari, D. Inflammatory and Metalloproteinases Profiles Predict Three-Month Poor Outcomes in Ischemic Stroke Treated with Thrombolysis. *J. Cereb. Blood Flow Metab. Off. J. Int. Soc. Cereb. Blood Flow Metab.* **2017**, *37* (9), 3253–3261. <https://doi.org/10.1177/0271678X17695572>.
- (100) Ghini, V.; Tenori, L.; Capozzi, F.; Luchinat, C.; Bub, A.; Malpuech-Brugere, C.; Orfila, C.; Ricciardiello, L.; Bordoni, A. DHA-Induced Perturbation of Human Serum Metabolome. Role of the Food Matrix and Co-Administration of Oat β -Glucan and Anthocyanins. *Nutrients* **2019**, *12* (1). <https://doi.org/10.3390/nu12010086>.
- (101) Bub, A.; Malpuech-Brugère, C.; Orfila, C.; Amat, J.; Arianna, A.; Blot, A.; Di Nunzio, M.; Holmes, M.; Kertész, Z.; Marshall, L.; Nemeth, I.; Ricciardiello, L.; Seifert, S.; Sutulic, S.; Ulaszewska, M.; Bordoni, A. A Dietary Intervention of Bioactive Enriched Foods Aimed at Adults at Risk of Metabolic Syndrome: Protocol and Results from PATHWAY-27 Pilot Study. *Nutrients* **2019**, *11* (8). <https://doi.org/10.3390/nu11081814>.
- (102) Guo, X.-F.; Li, X.; Shi, M.; Li, D. N-3 Polyunsaturated Fatty Acids and Metabolic Syndrome Risk: A Meta-Analysis. *Nutrients* **2017**, *9* (7). <https://doi.org/10.3390/nu9070703>.
- (103) Brown, L.; Poudyal, H.; Panchal, S. K. Functional Foods as Potential Therapeutic Options for Metabolic Syndrome. *Obes. Rev. Off. J. Int. Assoc. Study Obes.* **2015**, *16* (11), 914–941. <https://doi.org/10.1111/obr.12313>.
- (104) Cloetens, L.; Ulmius, M.; Johansson-Persson, A.; Akesson, B.; Onning, G. Role of Dietary Beta-Glucans in the Prevention of the Metabolic Syndrome.

- Nutr. Rev.* **2012**, *70* (8), 444–458. <https://doi.org/10.1111/j.1753-4887.2012.00494.x>.
- (105) García-Alonso, F. J.; Jorge-Vidal, V.; Ros, G.; Periago, M. J. Effect of Consumption of Tomato Juice Enriched with N-3 Polyunsaturated Fatty Acids on the Lipid Profile, Antioxidant Biomarker Status, and Cardiovascular Disease Risk in Healthy Women. *Eur. J. Nutr.* **2012**, *51* (4), 415–424. <https://doi.org/10.1007/s00394-011-0225-0>.
- (106) Toufektsian, M.-C.; Salen, P.; Laporte, F.; Tonelli, C.; de Lorgeril, M. Dietary Flavonoids Increase Plasma Very Long-Chain (n-3) Fatty Acids in Rats. *J. Nutr.* **2011**, *141* (1), 37–41. <https://doi.org/10.3945/jn.110.127225>.
- (107) Barupal, D. K.; Haldiya, P. K.; Wohlgemuth, G.; Kind, T.; Kothari, S. L.; Pinkerton, K. E.; Fiehn, O. MetaMapp: Mapping and Visualizing Metabolomic Data by Integrating Information from Biochemical Pathways and Chemical and Mass Spectral Similarity. *BMC Bioinformatics* **2012**, *13* (1), 99. <https://doi.org/10.1186/1471-2105-13-99>.
- (108) Benjamini, Y.; Hochberg, Y. On the Adaptive Control of the False Discovery Rate in Multiple Testing with Independent Statistics. *J. Educ. Behav. Stat.* **2000**, *25* (1), 60–83.
- (109) Chong, J.; Soufan, O.; Li, C.; Caraus, I.; Li, S.; Bourque, G.; Wishart, D. S.; Xia, J. MetaboAnalyst 4.0: Towards More Transparent and Integrative Metabolomics Analysis. *Nucleic Acids Res.* **2018**, *46* (W1), W486–W494. <https://doi.org/10.1093/nar/gky310>.

Appendix A

Candidate's contribution

PhD publications

PUBLISHED

1. Basoglu A, Baspinar N, Tenori L, Licari C, Gulersoy E. “*Nuclear magnetic resonance (NMR)-based metabolome profile evaluation in dairy cows with and without displaced abomasum*”. Vet Q. 2020, 40(1), 1-15 (J.I.F. 2.34)
2. Basoglu A, Baspinar N, Licari C, Tenori L, Naseri A. “*NMR-based serum metabolomics for monitoring newborn preterm calves' health*”. JJVR. 2020, 68(2),105-116 (J.I.F. 0.32)
3. Basoglu A, Turgut K, Baspinar N, Tenori L, Licari C, Ege Ince M, Ertan M, Suleymanoglu H. “*NMR-based serum extracts' metabolomics for evaluation of Canine Ehrlichiosis*”. Accepted for publication in JJVR (J.I.F. 0.32).
4. Vignoli A, Ghini V, Meoni G, Licari C, Takis P.G, Tenori L, Turano P, Luchinat C. “*High-Throughput metabolomics by 1D NMR*”. Angew. Chem. Int. Ed Engl. 2019, 58(4), 968-994 (J.I.F. 12.26)

SUBMITTED

1. Cacciatore S.*, Wium M.*, Licari C.*, Masieri L., Anderson C., Salukazana A. S., Kaestner L., Carini M., Carbone G. M., Catapano C. V., Loda M., Libermann T. A., Zerbini L. F. “*Inflammatory metabolic profile of South African patients with prostate cancer*”.
2. Licari C, Tenori L, Luchinat C. “*Simple equidistant bucketing as robust and recommended procedure for NMR-based metabolomic fingerprinting*”.

IN PREPARATION

1. Licari C*, Meoni G, Tenori L, Turano P, Luchinat C. *et al.* “*Nuclear Magnetic Resonance-based metabolomics to characterize serum sex-related metabolic profiles of drug-naïve Parkinson's disease patients with respect to healthy controls and patients with advanced disease and under dopaminergic treatment*”.
2. Licari C*, Tenori L, Luchinat C. *et al.* “*NMR-based metabolomics for the prediction of three-month outcomes in ischemic stroke treated with thrombolysis*”.
3. Licari C*, Tenori L, Luchinat C, Saccenti E. *et al.* “*Differential Network Analysis reveals metabolite and lipid components associated with three-month*

death and impairment in patients with acute ischemic stroke after thrombolytic treatment with recombinant tissue plasminogen activator”.

4. Licari C*, Ghini V, Tenori L, Luchinat C. *et al.* “*Untargeted NMR-based metabolomics to investigate the effect of Bioactive Foods enriched with combination of DHA and anthocyanins or oat β -glucan on serum metabolome and lipidome of subjects at risk for metabolic syndrome: Large Intervention Study from European Union H2020 Pathway-27 project*”.

*First author or first co-author.

Appendix B

Candidate's activities

TOTAL CREDITS from attended seminars, courses and other training activities:
53.24 CFU

ATTENDED CONFERENCES AND CONTRIBUTIONS

Keynote Lecture: 1st Conference on Innovative Researches in Pharmaceutical and Environmental Sciences – 27th November 2019, Pisa, Italy. “*NMR-based metabolomics: applications and challenges for clinical and pharmaceutical research*”.

Oral communication: PhD day 10th edition, 23rd May 2019, Sesto Fiorentino, Italy. “*Nuclear Magnetic Resonance-based Metabolomic approach to study acute ischemic stroke*”.

Poster: GIDRM-IMASS Advances in NMR and MS-based Metabolomics 2019, 20th-22th November 2019, Lucca, Italy. “*The importance of bucketing procedure for NMR-based metabolomic fingerprinting*”.

Poster: GIDRM XLVIII National Congress on Magnetic Resonance, 11th-13th September 2019, L'Aquila, Italy. “*The importance of bucketing procedure for NMR-based metabolomic fingerprinting*”.

Poster: EMBO workshop: Challenge for Magnetic Resonance in Life Sciences, 27th-31th May 2018, Grosseto, Italy. “*Application of the NMR technique for the analysis of sera of subjects with diet enriched in DHA*”.

Poster: GIDRM XLVII National Congress on Magnetic Resonance, 19th-21th September 2018, Torino, Italy. “*Application of the NMR technique for the analysis of sera of subjects with diet enriched in DHA*”.

Workshop: GIDRM “*Metabolomics in cancer*”, 28th November 2018, Firenze, Italy.

Meeting: CERM “*Fingerprinting in Metabolomics by NMR*”, 12th-13th June 2019, Firenze, Italy.

Meeting: “*The Use of NMR in biobanking, from standardization to Quality Control, generation of high value meta information and support of epidemiological studies and clinical trials*”, 23rd September 2019, Firenze, Italy.

GRANTS AND AWARDS

Sept. 2018: **GIDRM Young Fellowship grant** to participate in the XLVII GIDRM National Congress on Magnetic Resonance. Participation with a poster.

Sept. 2019: **GIDRM Young Fellowship grant** to participate in the XLVIII GIDRM National Congress on Magnetic Resonance. Participation with a poster.

Nov. 2019. **First awards** for the best keynote lecture in the memory of Prof. Cinzia Chiappe at the conference “*NMR-based metabolomics: applications and challenges for clinical and pharmaceutical research*”.

"Just as a flower blooms after enduring the harsh winter cold, a dream can only come true if you are prepared to endure the torments that accompany its realization and to make all the necessary efforts!"
[from D.Ikeda]

Whatever flower you are, the time will come when you will blossom.

The end
