UNIVERSITÀ
DEGLI STUDI
FIRENZE

DOTTORATO DI RICERCA
INTERNATIONAL DOCTORATE IN STRUCTURAL BIOLOGY

CICLO XXXV

COORDINATOR Prof. Lucia Banci

Application of NMR-based metabolomics in biomedical research

Settore Scientifico Disciplinare CHIM/03

PhD student
Dott. Francesca Di Cesare

Francesca Di Cesare

Tutor
Prof. Claudio Luchinat

Coordinator
Prof. Lucia Banci

November 2019 – October 2022

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

Universiteit Utrecht

FRANKFURT AM MAIN

Abstract

Nuclear magnetic resonance (NMR)-based metabolomics is an emerging and robust -*omic* science that deals with the systematic identification, characterization, and quantification of the complete set of metabolites that are present in biological specimens (*i.e.* cells, tissues, biofluids, food-derived matrices, etc.). The metabolome – the object investigated by metabolomics – can be described as a highly complex and organized biochemical network in which metabolites, lipids, and lipoproteins, thanks to their fluctuations in terms of concentration and thanks to their interconnections, are directly responsible for the emerging phenotype of the organism. The metabolome represents also a dynamic and evolving entity arising from the interaction between genome, transcriptome, and proteome, under the combined influence of several endogenous and exogenous stimuli. Considering this dynamic behavior of the metabolome, metabolomic data can be easily correlated with the phenotype and act as a direct signature of biochemical activity, since metabolites play a central role in disease development, cellular signaling, and physiological control. In this light, the very high reproducibility, the minimal requirement in sample preparation, and the possibility to simultaneously detect all metabolites presenting active nuclei (at least above the detection limit) make NMR-based metabolomics one of the most powerful and versatile techniques for the analysis of any type of biological sample, providing a global snapshot of the complex metabolic, biological and biophysical processes that occur in a specific organism at the time of sampling.

In this scenario, the methodological thesis here presented aims to apply and demonstrate the potential of the untargeted metabolomics approach, particularly in the biomedical field, covering various topics, obtaining new insights on different biological and physiological conditions, shedding light on the dimorphic mechanisms of aging, determining how an improving human well-being treatment (*i.e.* probiotics) could affect the metabotypes, characterizing the metabolomic and lipoproteomic profiles associated with the inherited blood types (ABO and Rh systems), characterising the metabolic components of two diseases, acute ischemic stroke and colorectal cancer, providing prognostic and diagnostic biomarkers of these specific pathologies. This thesis also proposes a study in which a robust statistical approach, based on the construction of linear regression Random Forest models, is developed to calculate several chemical parameters and sensory profiles of olive oil using ^{1}H-NMR spectra.

In summary, the results here presented suggest that untargeted NMR-based metabolomics, in combination with biochemical, analytical chemistry, bioinformatics tools,

and robust statistical analysis, is a useful and reasonable candidate for increasing knowledge in various research fields, especially focusing on biomedical research.

Main abbreviation and acronyms

AIS: Acute Ischemic Stroke

ANOVA: Analysis of Variance

BCAAs: branched chains amino acids

AUC: Area Under the Curve

CRC: colorectal cancer

CTR: healthy controls

DDS: 2,2-dimethyl-2-silapentane-5-sulphonate

FN: false negative

FP: false positive

GC: gas chromatography

HPLC: high-performance liquid chromatography

k-NN: k-Nearest Neighbour

LMM: Linear mixed-effects model

LPC: lysophosphatidylcholine

MAG: monoacylglycerol

M-PLS: Multilevel Partial Least Squares Analysis

MPS: muscle protein synthesis

MS: Mass spectrometry

NMR: Nuclear Magnetic Resonance

OPLS-DA: Orthogonal-Partial Least Square-Discriminant Analysis

PC: phosphatidylcholine

PCA: Principal Component Analysis

PCLRC: Probabilistic Context Likelihood of Relatedness based on Correlation

PLS: Partial Least Square

PP: polyposis

PQN: Probabilistic Quotient Normalization

RF: Random Forest

ROC: Receiver Operating Characteristics

rt-PA: recombinant tissue plasminogen activator

SVM: Support Vector Machine

TMS: tetramethylsilane

TMSP: trimethylsilylpropanoic acid

TN: true negative

TP: true positive

Table of contents

Chapter 1

Introduction

1.1 Metabolomics: an emerging field in the *-omics* sciences

In the last few decades, the fields of sciences known as *-omics* have exponentially grown. A common characteristic among disciplines such as genomics, transcriptomics, and proteomics is an integrated approach aimed at comprehensively considering the properties of biological systems and their interactions with the environment, identifying the complete set of genes, proteins, and other biomolecules contained in a biological sample[1].

Lifestyle and environmental variables have a significant impact on the organism, defining its phenotype. From this point of view, a problem that should not be underestimated is that genomics, protein expression, and molecular biology are not sufficient, by themselves, to understand how exogenous variables can affect a biological system as a whole at the molecular level.

In this scenario, metabolomics, a member of the *-omic* club, is the only science capable of solving this limitation[2]. It provides a comprehensive image of biochemistry in several organisms (*i.e.* humans, animals, plants, and microorganisms) and it deals with the characterization of different metabolites (<1500 Da) in a specific sample (*i.e.* cells, tissues, biofluids, etc,), considering the metabolic variations deriving from both exogenous and endogenous factors[3] (**Figure 1**).

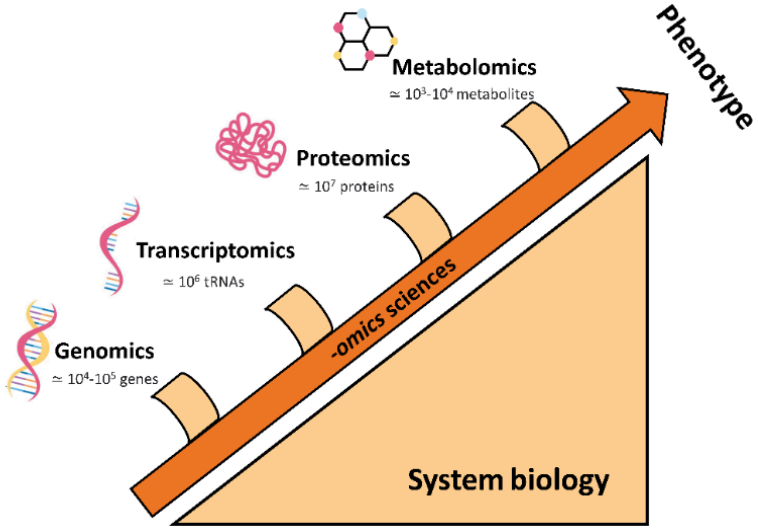


Figure 1: Metabolomics in system biology. According to the central dogma of molecular biology, the flow of information proceeds from the genome to the transcriptome, to the proteome, and the metabolome, contributing together to determine the phenotype.

1.2 Metabolomics and its main experimental approaches

Metabolomic studies can be conducted using two possible methodological approaches: targeted and untargeted. Targeted metabolomics focuses on the measurement and monitoring of a biochemically and chemically characterized set of metabolites selected from known metabolic pathways or previously identified biomarkers[4]. Instead, untargeted metabolomics – applied in this methodological thesis – provides a comprehensive overview of all molecular features detectable in a specific sample, including, also, chemical unknowns[3,5,6]. Fingerprinting and profiling are currently the two approaches mostly used in untargeted metabolomics. The former strategy allows to globally analyse all molecular features present in each sample and it is a strategy mainly used to perform sample classification[7,8]. The latter strategy provides the assignment and the quantification of as many as possible molecular features in a sample, providing more specific information on potential fluctuations or alterations in the activity of specific metabolic pathways in both physiological and pathophysiological processes[9,10].

Once the methodological approach to be followed has been established, it is necessary to choose the most appropriate analytical technique. The two most used are Nuclear Magnetic Resonance (NMR) spectroscopy – mainly used in this methodological thesis – and Mass spectrometry (MS)[3,11–13].

Briefly, MS investigations typically require the separation and derivatization of compounds from biological matrices before detection, which is commonly accomplished using high-performance liquid chromatography (HPLC) or gas chromatography (GC). Recently, the MS analytical technique, in particular LC-MS, is improved to detect various compounds with different chemical characteristics, ranging from hydrophobic to hydrophilic ones. Although MS has higher sensitivity and the volume of sample required is minimal, its reproducibility in the metabolomics field remains a weakness, taking into account that the MS experiments are generally slower and might have to be tailored according to which metabolites need to be detected. In contrast, NMR turns out to be non-destructive and rapid, as any molecule containing active nuclei can be detected simultaneously, and intrinsically quantitative as the integral of each peak is proportional to the number of nuclei with non-zero magnetic moment generating that peak. **Table 1** shows the strengths and weaknesses of both analytical techniques.

	NMR	MS
Destructive technique	No	Yes
Reproducibility	Very high	Fair – depends on separation methods used
Detection limit	Micromolar range	Picomolar range
Sample preparation	Minimal	Onerous – typically requires chromatographic separation and sample derivatization
Sample volume	0.1-0.5 mL	0.01-0.2 mL
Types of compounds detected	Any molecule containing NMR active nuclei, provided concentration is above the detection limit. All metabolites are detected simultaneously	Most organic and some inorganic. Experiments need to be tailored for specific chemical species
Possible sources of error	Compounds with degenerate chemical shifts, chemical shift variability due to experimental conditions (*i.e.* pH, temperature, ionic strength)	Compounds (*i.e.* isomers) that can match the same atomic composition or parent ion mass

Table 1: Main NMR and MS strengths and weaknesses for metabolomics studies (adapted from the review published by Vignoli *et al*.[3]).

In summary, comparing the different characteristics of the two techniques, they are complementary, considering that the NMR technique is more suitable for untargeted metabolomics, characterizing metabolic profiles of organisms under different conditions, while the MS technique is more suitable for targeted metabolomics, confirming or validating the variation or alteration of a metabolic pathway in a specific condition.

1.3 Biological samples analysed by metabolomics

Several types of samples can be analysed using NMR spectroscopy, including human- and animal-derived fluids (*i.e.* serum, plasma, urine, etc.), faeces, cells, tissues from human, animal, and plant, culture media, microorganisms (*i.e.* bacteria and yeast cultures), and food-derived matrices (*i.e.* olive oil, fruit juice, wine, fruit and vegetable extracts, etc.). The choice of the sample is extremely dependent on the biological question posed by the researcher. NMR can identify hundreds of tiny molecules in a sample, including biomolecules belonging to the classes of amino acids, carbohydrates, organic acids, nucleotides, osmolytes (*i.e.* choline),

phenolic compounds, and alcohols[3]. In addition, a large set of large and small lipids and lipoproteins (*i.e.* high-density and low-density lipoproteins, mono-, di-, tri-triglycerides, cholesterol, cholesterol esters, short-chain fatty acids, etc.) are also detected using NMR spectroscopy, thus allowing not only to have an image of the sample metabolome but also of the lipidome and lipoproteome[14].

Furthermore, it is also important to specify that the samples to be analysed have different metabolomic behaviours based on the genetic, biochemical, and environmental conditions in which they are at the time of sampling. For example, the cellular metabolome varies according to both endogenous and exogenous stimuli[15], and, for this reason, in order to have a more complete metabolic image, it is also necessary to analyse the exo-metabolome[16]. Similar considerations could be made by analysing the tissue samples. Tissue NMR data result more complex to interpret because they reflect the organ-specific biochemistry (*i.e.* glycolytic or ketogenic metabolism, etc.) and they are influenced by the extracellular composition. On the other hand, tissue samples provide direct information about diseased organ status. A good informational compromise is given by the evaluation of metabolites and lipoproteins detectable in biofluids, particularly in plasma, serum, and urine samples. These biofluids, although not reflecting a biochemical correlation with a specific organ or apparatus, are simple and minimally invasive to collect and represent the overall response of the organism to a particular physiological or pathophysiological condition. In terms of chemical composition, plasma, serum, and urine samples significantly differ from each other. Urine metabolites are highly influenced by lifestyle factors such as nutritional daily uptake, physical activity, and drug or supplements usage; in contrast, serum and plasma samples are considered more stable biofluids, providing a picture of an organism's metabolic state, offering critical information for recognizing, controlling, and monitoring all potential physiological ad pathophysiological state conditions[17,18]. Blood, in fact, plays a key role in transporting nutrients, hormones, dissolved gases, maintaining the pH and ion content stability, providing defence against infections, and helping to also maintain body temperature. In this methodological thesis, for the motivation described above, the plasma and serum samples have been mainly investigated. Regarding the food-derived matrices, the metabolome results to be extremely complex due to the diversity of chemical compounds found in plants, particularly secondary metabolites that are specific to each species. All food-derived matrices have a unique profile, characterized by the endogenous molecular features produced by the plant and plant-symbiotic microbial community metabolisms, and it is closely connected to their nutritional content, human health-promoting properties, and sensorial characteristics. The heterogenicity between the different food-derived matrices allows using the specific detected metabolites and lipids as potential markers for quality, origin, sensorial characteristics, and authenticity of food[19].

1.4 NMR-based metabolomics workflow

Metabolomic workflow is initially based on the formulation of the biological problem to be addressed, crucially determining the next experimental steps to follow. Based on the biological question chosen, before proceeding with the workflow, it is necessary to establish the type of biological sample to be analysed (*i.e.* biofluid, cells, tissue, food-derived matrices, etc.), the sample size, and the experimental conditions (*i.e.* the number replicates, the frequency of sample collections in longitudinal studies, the choice of an appropriate cohort to analyse and its related metadata, the division into sex- and age-matched case and control groups, etc.).

Taking into account the above-mentioned aspects, a schematic and graphical representation of a typical NMR-based metabolomics workflow is presented in **Figure 2**.

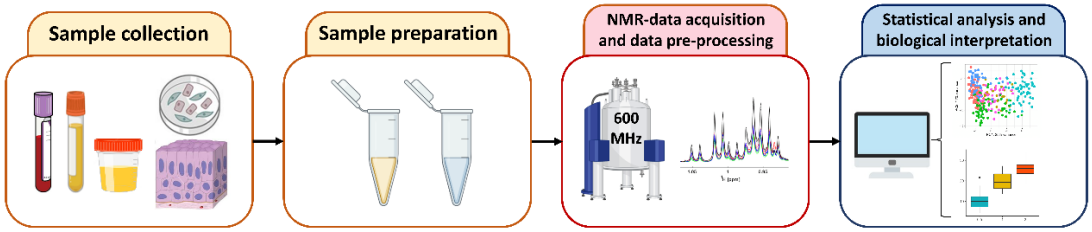


Figure 2: NMR-based metabolomics workflow.

The first and crucial step to be followed is the collection, handling, and storage of a specific sample in which the adoption of Standard Operating Procedures (SOPs)[20–22] results essential to ensure sample stability and data reproducibility. Errors in this phase can irreversibly invalidate the experimental study.

The second step is based on the sample preparation, closely related to the sample type, the analytical technique, and the necessary extraction and/or buffering methods chosen.

The third step regards the NMR-data acquisition in which appropriate NMR pulse sequences need to be implemented and applied. After successfully acquiring the NMR spectra, a pre-processing phase (baseline and phase correction, calibration to an internal reference peak, normalization, and bucketing) is necessary to obtain more comparable metabolomics data. After this step it could be useful, in the untargeted metabolomics data, to proceed with the assignment and quantification of the molecular features detectable in NMR spectra.

Lastly, the statistical analysis of the metabolomic data is performed. In particular, multivariate approaches, comprising unsupervised (*i.e.* Principal Component Analysis (PCA)[23]) and supervised methods (*i.e.* Orthogonal-Partial Least Square Discriminant Analysis (OPLS-DA)[24], Multilevel Partial Least Squares analysis (M-PLS)[25]), and machine learning

algorithms (*i.e.* Support Vector Machines[26], Random Forest[27–29]), are used to have an overview of the samples, to classify a new sample based on two or more known groups, and to make predictions. Univariate approaches (*i.e.* Kruskal–Wallis test[30], Wilcoxon signed-rank test[31,32], mixed-effect models[33,34]) are also used to identify associations between a metabolite or a set of metabolites with a specific condition. In a more system biology holistic approach, molecular features association networks are used to infer metabolic significant associations.

Concluding, the final step is characterized by the biological interpretation of the results obtained, and, for this purpose, several databases (*i.e.* KEGG, MetaboLights[35], etc.) and online tools (*i.e.* MetaboAnalyst[36]) are available.

More details about all these procedures and different approaches are deeply reported in **Chapter 3**.

1.5 NMR-based metabolomics: the state of art

One of the main and extremely relevant characteristics of NMR metabolomics is its versatility regarding several fields of application in life science research, including the investigation of human health, the metabolic characterization of food-derived matrices, plants, animals, microorganisms, and the environment. (**Figure 3**).

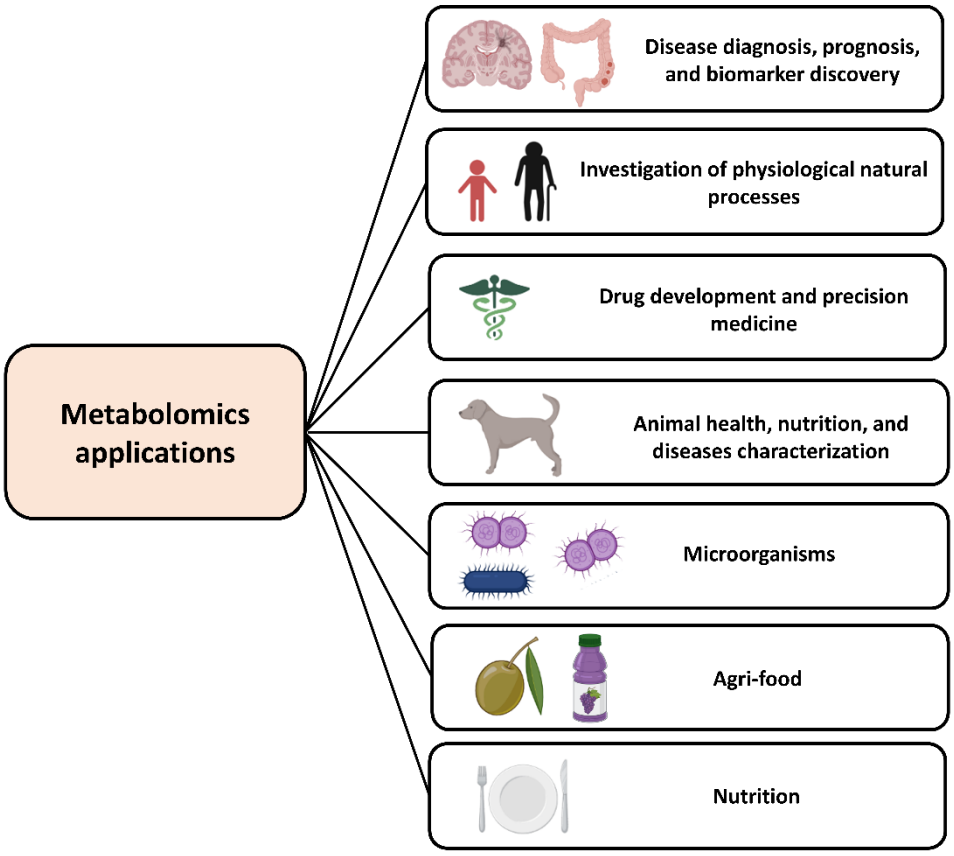


Figure 3: NMR-based metabolomics several applications in life sciences.

Metabolomics provides a comprehensive approach to identify the specific molecular profile, derived from both exogenous and endogenous stimuli, with the final aim to investigate and to understand the phenotype of an organism, both in physiological and pathophysiological conditions.

As known, NMR-based metabolomics is a technique increasingly used to characterize a wide range of human pathophysiological conditions such as cancer[37–41], cardiovascular[42–44], cerebrovascular[45,46], and neurological diseases[47–49] (disease fingerprint). With the aim of achieving a personalized medical approach, metabolomics also offers an important role in discovering new drug targets and correctly predicting a potential individual response to a specific drug (pharmacometabolomics)[50–52]. Moreover, it is promising in *molecular epidemiology*, providing information about novel risk biomarkers, and detecting early metabolic disorders even before the symptomatic manifestation of disease, thus providing early diagnosis[53–55]. Recently, metabolomics embodied a relevant position in characterizing the human physiological processes, such as aging, the endocrine, immune, and gastrointestinal systems functions[11,56,57]. Other applications include the evaluation of dietary uptake (*i.e.* food, supplements, probiotics) to understand the physiological response of the human body to non-pharmacological exogenous intakes[58,59]. In this scenario, the food and nutrition research is of growing interest in metabolomics, focusing on various aspects, including food quality[60], taste[59,61], adulteration[62], traceability[63], contaminants and phytomedical[64], and effects on human metabolism[65,66].

Although this methodological thesis focuses mainly on research in the human biomedical field, it is necessary to remember that metabolomics is important both in veterinary medicine[67–70], characterizing the state of health of the animal, and in the study of plant physiology and pathophysiology[71–73].

Chapter 2

Aim of the thesis

Metabolomics is a rapidly growing valuable technology used to create a more comprehensive picture of the molecular organization of multicellular organisms. The combination of enhanced analytical skills and newly specialized statistical, bioinformatics and data mining methodologies is starting to widen our understanding of physiological and pathophysiological metabolic human mechanisms. In this scenario, this methodological thesis covers various topics with the principal aim of applying and demonstrating the potential of the untargeted metabolomics approach in the biomedical field.

In the first section of this work – in the context of human physiological and pathophysiological research –, particular attention is dedicated to demonstrating the usefulness of the untargeted metabolomics approach to:

1) investigate, using a holistic system biology approach (statistical and metabolite association networking analyses), the complex interconnections and metabolic differences between the individual molecular features associated with a specific clinical condition, *i.e.* colorectal cancer (CRC), and with a dimorphic physiological condition, *i.e*. the aging process, unravelling, also, prognostic and diagnostic markers of distinct pathophysiological status;
2) provide predictive metabolic markers of three-months outcomes of ischemic stroke (*i.e*. mortality, development of neurological impairment, haemorrhagic transformation of the cerebral lesion, and non-response to the thrombolytic therapy), to better characterize the post-stroke course of the pathological condition;
3) characterize the human metabolic phenotype, highlighting how an exogenous non-invasive treatment (*i.e.* the probiotics assumption) could have a strong influence on the complex human metabolome;
4) characterize metabolomic and lipoproteomic profiles specific to different blood group systems (ABO and Rh systems) in a healthy population;

In the context of developing new strategies for NMR-based metabolomics data analyses, the second and last section of this thesis is dedicated to:

1) developing a robust statistical approach, based on the construction of linear regression Random Forest models, to extract several important chemical parameters and sensory profiles of olive oil from ^{1}H-NMR fingerprints of the oil samples.

Chapter 3

Methodologies

3.1 Sample preparation

To structure a robust metabolomic study, the sample preparation procedure must respect important criteria[74]: i) reproducibility; ii) simplicity of usage; iii) reducing variation and artifacts; iv) maintenance and/or improvement of quality resolution and sensitivity of NMR spectra.

In metabolomics, a central and fundamental role is played by the pre-analytical phase which involves several steps, including primary sample collection, processing, transport, and storage. Errors during this phase or inadequate procedural techniques can lead to severe difficulties in comparing metabolomic data collected from different centers or institutes (multicentric studies), reducing the data reproducibility. In this perspective, it is important to adopt and strictly follow Standard Operating Procedures (SOPs)[20–22] related to the different types of samples to be analysed.

Generally, the analytical phase of NMR-sample preparation of biofluids (*i.e.* serum, plasma, urine, etc.) and food-derived fluids (*i.e.* olive oil, wine, fruit juices, etc.) requires minimal essential operative steps, mainly consisting of buffering the solution to avoid pH variation and to easily reference chemical shifts to existing resonances databases. In particular, a known quantity of a reference compound (*i.e.* 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 added to the sample to be analysed. To allow the optimization of NMR peak resolution and to provide a signal for the magnetic field stabilization, deuterated solvents, in particular D2O, are also used. For organic fractions extraction, chloroform – a non-polar solvent – or methanol-chloroform mixtures are added in different proportions, while D2O is preferred to reconstitute polar fractions. Deuterated chloroform (CDCl3) is highly used as a solvent for olive oil sample preparation.

In NMR analysis a volume ranging from 0.1 to 0.5 mL of the original sample is required.

More information about the pre-analytical and analytical procedures of the fluids considered in this methodological thesis is reported in **Table 2**.

Type of sample	Pre-analytical steps	Analytical steps
Urine	Collection of the first urine of the morning (8h fasting). Centrifuge the urine within 120 min after the collection at 1000–3000 RCF for 5 min at +4°C and/or filtrate the samples with 0.20 mm cut-off filter. Recover the urine in sterile condition making 1mL aliquots. Store at -80°C.	Keep at room temperature (RT) and shake. Centrifuge at 14000 RCF for 5min at 4°C. Add 630µL of sample to 70µL of potassiumphosphate buffer (1.5M K_2HPO_4, 100% (v/v) ^2H_2O, 2mM NaN_3, 5.8 mM TMSP; pH 7.4). 600µL of the mixture may be transferred into a 5mm NMR tube.
Plasma	Collect blood into tubes containing an anti-coagulant (preferred EDTA; avoid heparin). Centrifuge within 30 min from the collection at 820 RCF for 10 min at 4°C. Recover supernatant in sterile condition making 0.5 mL aliquots. Store at -80°C.	Keep at RT and shake. Add 350µL of sample to 350 µL of sodiumphosphate buffer (70 mM Na_2HPO_4; 20% (v/v) ^2H_2O, 6.1 mM NaN_3; 4.6 mM TMSP; pH 7.4). Transfer 600µL of each mixture into a 5mm NMR tube.
Serum	Collect blood into anticoagulant-free tubes. Allow the blood to clot in an upright position for 30–60min at RT. Spin centrifuge within 30 min from the collection at 1500 RCF for 10 min at RT. Recover supernatant in sterile condition making 0.5 mL aliquots. Store at -80°C.	
Olive oil	Create aliquot portions of olive oil samples and store them in dark glass vials at 4 °C	Add 60 mg of olive oil to 600 µL of deuterated chloroform (CDCl3).

		After 20s of vortexing, transfer 600 µL of the prepared mixture into a 5 mm NMR tube.

Table 2: NMR sample preparation. In this table (adapted from the review published by Vignoli *et al.*[3]), the different types of samples analyzed in this methodological thesis, and the pre-analytical and analytical procedures are reported.

3.2 NMR-spectra acquisition

NMR may theoretically identify any molecule having one or more atoms with a non-zero magnetic moment. Moreover, because isotopes – including ^{1}H, ^{13}C, ^{14}N, ^{15}N, and ^{31}P – present non-zero magnetic moments, all biomolecules have at least one NMR signal. The frequency (chemical shift), intensity, fine structure, and magnetic relaxation characteristics of these signals all reflect the specific surroundings of the detected nucleus[75,76]. As a result, NMR-spectroscopy is an efficient and versatile technique for analysing complex biological samples, allowing the acquisition of spectra that often include a plethora of information about all molecules detectable in a specific sample. Due to the high sensitivity and natural abundance of ^{1}H isotope (>99%) in biomolecules, mono-dimensional (1D)^{1}H-NMR spectroscopy is the most highly used technique in metabolomics research.

In NMR-based metabolomics studies, a good compromise between sensitivity, spectra resolution, and cost is characterized by the use of a 600 MHz spectrometer for common biofluids[3], cells, and tissues analyses and a 400 MHz spectrometer for food-derived fluids analysis[77].

Considering that most of the 1D ^{1}H-NMR spectra acquired for the metabolomics studies are obtained in water, an important aspect of the solution-state NMR experiment is solvent suppression[74]. Indeed, the biological sample contains a large amount of water and the addition of small quantities of D_2O is sufficient to generate a signal for deuterium lock. In this scenario, the simplest and most suitable strategy for water suppression is known as "pre-saturation" which uses mild radio-frequency irradiation for a duration of the order of the solvent T_2 to selectively saturate the solvent resonance[78]. Gradient-based solvent suppression approaches, such as the WATERGATE system, can also be utilized to provide efficient water signal reduction. This scheme employs a composite pulse surrounded by two symmetric pulsed-field gradients to attenuate the water resonance[79,80]. A modulated-shaped pulse is often used for the study of food matrices, *i.e.* olive oil, to suppress the strong lipid signals[63].

Standardization of NMR metabolomic experiments has been increasingly requested in recent years mainly for the development of a more complete spectra database and data comparison and reproducibility. As a result, only a few pulse sequences, that follow standard protocols, are applied in the NMR-based metabolomics field[76,80–82].

1D ^{1}H-NOESY-presat (Nuclear Overhauser Effect SpectroscopY)[83] spectra give resonance signals from both low- and high-molecular weight compounds, giving a highly repeatable, robust, and high-quality approach to acquire a thorough overview of metabolites and lipoproteins detectable in a sample. CPMG spin-echo[84] sequence is also frequently used in NMR-based metabolomics, ensuring a more selective acquisition for low-molecular-weight molecules (via T_2 filtering). The ^{1}H Diffusion-Edited pulse sequence[85], on the other hand, allows for the selective detection of macromolecules in solutions containing small molecular components. The three NMR experiments described above are widely used to evaluate metabolites present in the blood (both in serum and plasma samples), since blood contains more macromolecules than other biofluids (**Figure 4**). For the sake of completeness, it should be reported that it is possible to physically remove the macromolecules via centrifugation, avoiding performing a CPMG experiment, but this strategy was not applied in this work.

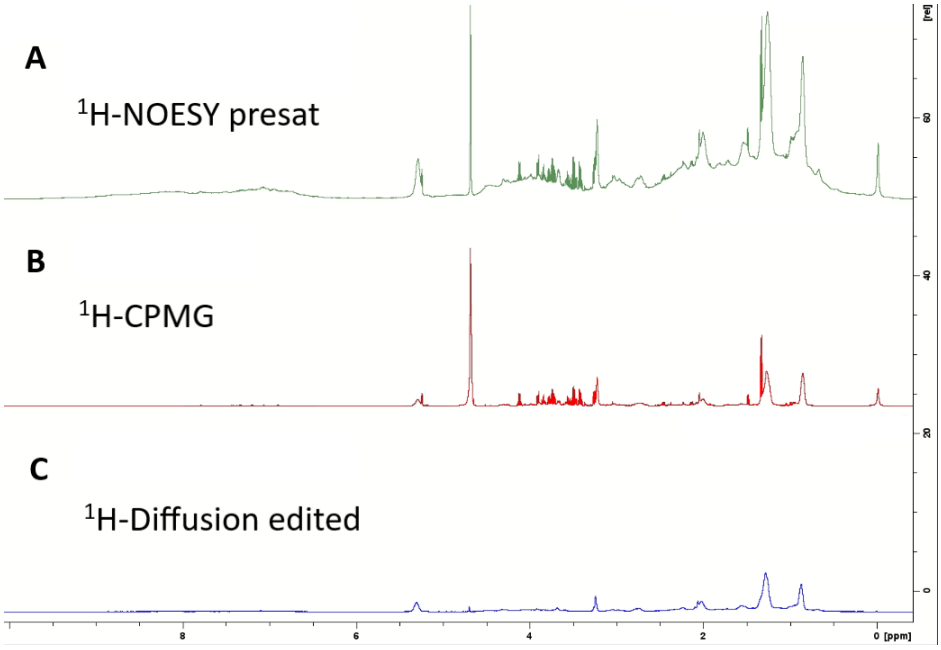


Figure 4: 600 MHz ^{1}H-NMR spectra of a serum sample, obtained performing the three more used experiments: ^{1}H-NOESY-presat (A), ^{1}H-CPMG (B), and ^{1}H-diffusion edited (C).

2D spectra are mostly used in metabolomics research to assign novel compounds or in doubtful circumstances. J-resolved ^{1}H-^{1}H (JRES)[86] represents the most often used 2D experiment offering information on multiplicity and coupling patterns (**Figure 5**). Given the

longer acquisition times to perform a 2D NMR experiment compared to a 1D experiment, it is necessary to take into account the sample stability before proceeding with the acquisitions.

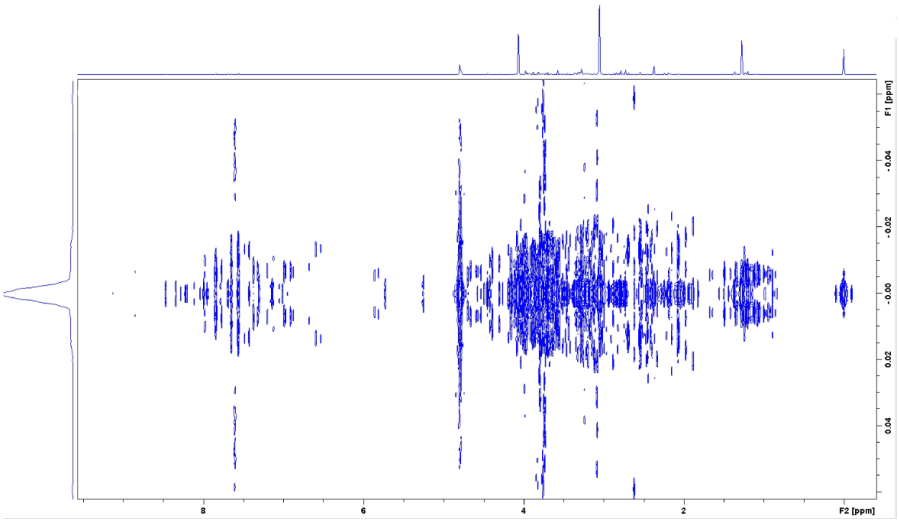


Figure 5: 600 MHz ^{1}H-NMR spectra of a urine sample, obtained performing JRES 2D experiment.

Common pulse sequences for NMR-based metabolomics experiments performed in this methodological thesis are reported in **Table 3**.

Sample Analysed	**Experiment performed**	**Pulse sequence**
Plasma; Serum	CPMG with water pre-saturation	1D experiment with T_2 filter using Carr-Purcell-Meiboom-Gill sequence with presaturation during the relaxation delay
Plasma; Serum	Diffusion-edited	1D sequence for diffusion measurement using stimulated echo and LED with bipolar gradient pulses for diffusion, 2 spoil gradients and presaturation during the relaxation delay
Plasma; Serum; Urine	NOESY with water pre-saturation	1D with spoil gradient using continuous wave presaturation during relaxation delay and mixing time and spoil gradient
Olive oil	NOESY shaped pulse	1D with spoil gradient using shaped pulse for multiple solvent presaturation (main solvent peak on resonance) during relaxation delay and continuous wave presaturation during the mixing time
Plasma; Serum; Urine	J-resolved	homonuclear J-resolved 2D correlation with presaturation during relaxation delay using gradients

Olive oil	ZG	1D ^{1}H standard single pulse experiment, *i.e.* RD-P(90°)-acquisition of the free induction decay (FID)

Table 3: NMR-spectra acquisition. In this table, the different types of samples analyzed in this methodological thesis, the NMR experiments, and the pulse sequences used are reported.

3.3 NMR-data processing

After the NMR spectra acquisition, the data processing results be an indispensable procedure to transform raw NMR data into a more correct form for secondary statistical analysis. It could be divided into different steps: i) baseline and phase correction; ii) calibration to an internal reference peak; iii) normalization; iv) bucketing.

Firstly, phase and baseline corrections are the first processing step in the analysis of raw NMR spectra, generally performed automatically. In this specific case, manual corrections are not recommended as they increase the likelihood of inserting operator-derived bias. After this step, to avoid errors in a chemical shift during the construction of the final matrix, the spectra need to be aligned to a known resonance signal that, ideally, does not interact with other molecules. For example, in serum or plasma samples, the use of deuterated trimethylsilylpropanoic acid (TMSP) is not recommended. TMSP could bind serum/plasma macromolecules (*i.e.* albumin protein) and, for this reason, the use of this signal is not also recommended as standard in the quantification procedures. For these biofluids is used another internal reference, such as the anomeric doublet of glucose at 5.24 ppm[87]. Absolute quantification of molecules can be performed using alternative approaches, such as the production of an artificial NMR signal based upon ERETIC[88] method. Software systems like the B.I. (Bruker IVDr) platform[89] for the analysis and the quantification of metabolites/lipoproteins are employed (for more details about assignment and quantification procedures see **§3.4**).

Calibrated full NMR spectra can be transformed into a multidimensional data matrix, usable for subsequent statistical analysis, by binning (or bucketing) techniques. These methods are applied to reduce the total number of variables and small misalignment in the NMR spectra. The spectra are divided into small spectral regions, called "bins" or "buckets"[74,90,91]. Many complex algorithms for binning 1D NMR spectra are available, but the most common and straightforward method is the equidistant binning of 0.02-0.04 ppm[90]; it allows the spectrum to be divided into equally spaced integration regions with a fixed spectral width.

To compare signals or bins intensities in different biological samples, since considerable physiological fluctuations in terms of metabolite concentration are always present, a procedure

to correct the dilution effect is required. In particular, the different hydration states of an individual could determine a significant variation in metabolites concentration detectable in urine, therefore spectra or binned spectra must be normalized to correct global signal intensity. For plasma and serum spectra, the normalization procedure is less relevant due to the physiological stability of the biofluids. Total area normalization is the most common normalization procedure applied. In more detail, in a NMR spectrum, each variable (*i.e.* a bin or a data point) is divided by a constant number that represents the sum of all bins (or all points) in the selected spectrum. Actually, for biofluids, especially urine, Probabilistic Quotient Normalization (PQN)[92] is considered an efficient normalization method, used also in this methodological thesis. This method is based on the calculation of a most probable dilution factor by looking at the distribution of the quotients of the amplitudes of a test spectrum by those of a reference spectrum. PQN assumes that variations in the concentration of biological interest, affect only a few portions of the NMR spectrum, while dilution effects will influence all signals. For food-derived fluids (*i.e.* olive oil) normalization according to the sample weight is recommended.

3.4 Metabolite assignment and quantification

To identify potential biomarkers attributable to a particular physiological and/or pathophysiological states, regardless of the biofluid or sample examined, it is useful and necessary to identify a metabolite/lipoprotein (or a set of metabolites/lipoproteins) in the NMR spectrum.

Due to the complexity of ^{1}H-NMR spectra, the identification of molecular features is not considered a simple analytical procedure. Firstly, some NMR peaks can be directly assigned in the one-dimensional spectrum (**Figure 6**), taking into account the molecular chemical shifts and the multiplicity of the resonances. This procedure is supported by the use of specific databases or libraries (some of them freely available), like Human Metabolome Database (HMDB)[93] for human metabolites, PubChem[94], Kyoto Encyclopedia of Genes and Genomes (KEGG)[95], ^{1}H-NMR spectra library of pure organic compounds (BBIOREFCODE, Bruker BioSpin)[96], etc. In case of doubtful assignment, the spiking of real standard molecules is recommended, while, in some cases, 2D NMR spectra provide further information for identifying novel metabolites. In recent years the need to develop tools to automatically perform metabolite and/or lipoprotein assignment and quantification in ^{1}H-NMR spectra has increased. Recently, the B.I. (Bruker IVDr) platform[89] for urine, CSF, and serum/plasma samples or the NMRpQuant software for large-scale urinary total protein quantification[97] have been introduced on the market. Meanwhile, other tools for semi-automatic quantitation have been developed (*i.e.* BATMAN[98] and BAYESIL[99] tools, based on Bayesian inference,

ASICS[100], and Chenomx Inc[101] which offers complete assistance for deconvolution, peak fitting, and integration of the selected peaks) but more manual procedures and skilled operators are needed to use them.

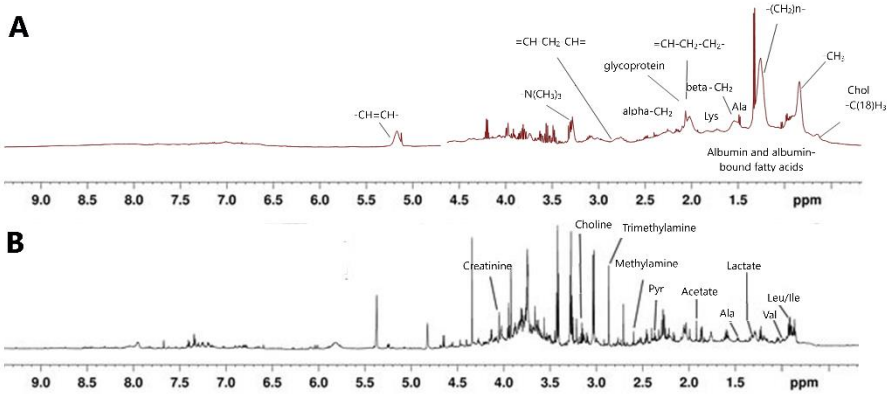


Figure 6: Assignment performed on: A) ^{1}H-NOESY NMR spectrum of a serum sample; B) ^{1}H-NOESY NMR spectrum of a urine sample

3.5 Statistical analysis

Due to the increasing complexity of NMR-based metabolomics data, statistical analysis embodies a very important aspect of the metabolomic workflow for obtaining complete and meaningful results ready to be interpreted from a biomedical and biochemical point of view. Starting independently from bucketed NMR spectra or a list of correctly assigned and quantified metabolite concentrations, the metabolomic analysis aims to:

1) visualize the overall differences, trends, clusters, and relationships among different samples;
2) detect significant metabolic differences potentially present among groups under study (*i.e.* discrimination among healthy and diseased patients, discrimination among samples before and after treatment, metabolic dimorphic discrimination, etc.);
3) highlight all metabolites most responsible for the mentioned differences;
4) build a robust predictive model to correctly classify new samples;
5) build a robust regression model and/or metabolite association networks to demonstrate relationships among potentially biologically related variables.

Depending on the experimental settings, several types of data mining and statistical approaches can be applied to metabolomics data. In the following subparagraphs, the univariate, multivariate, and metabolite association networking statistical approaches, performed in this methodological thesis, were described.

3.5.1 Multivariate analysis

When two or more variables are measured during an experiment, the resulting data is multivariate data. To handle hundreds or thousands of variables simultaneously, as in the case of metabolomics data sets, it is necessary to apply a multivariate statistical approach that can be divided into two different classes:

1) Unsupervised multivariate analysis in which no prior knowledge of the data is needed;
2) Supervised multivariate analysis in which samples are divided into classes or each sample is associated with a specific outcome.

Generally, to have an overview of metabolomics data, unsupervised methods are preferred as they allow to visualize data, potential outliers, similarities, or dissimilarities between metabolic signatures with excellent clarity and low computational cost.

Principal Component Analysis (PCA)[23] results to be the most popular and commonly used unsupervised approach and it involves the projection of the original data on a lower dimensional space (data reduction) capturing the information in the data, corresponding to the observed variation. Briefly, this method expresses variance within a dataset in a new reduced matrix of variables, called principal components (PCs), orthogonal and independent from all the others. A PC results in a linear combination of the original NMR-based metabolomics data (*i.e.* bins of a bucketed spectrum or a list of correctly assigned and quantified metabolites). The first PC (PC1) represents the maximum percentage of variability detectable in the original data. The variance value rapidly decreases, and the subsequent PCs are less important, expressing noise variability. PCA may be interpreted geometrically as a rotation of the reference system to optimize the data structure, reducing noise. The original data matrix (**X**) is factorized into two new matrices, **T**, the score matrix, and **A**, the loading matrix. **T** contains the new coordinates of the original data in a new rotated coordinate system determined by the PCs and **A** contains the weights of the original variables and transforms them into scores. Both score and loadings matrices are reported in graphical form in **Figure 7**. If the separation between sample groups is evident in the PC1 and PC2 score plots (as observable in **Figure 7**), the corresponding PC loadings provide information on the original variable responsible for this discrimination.

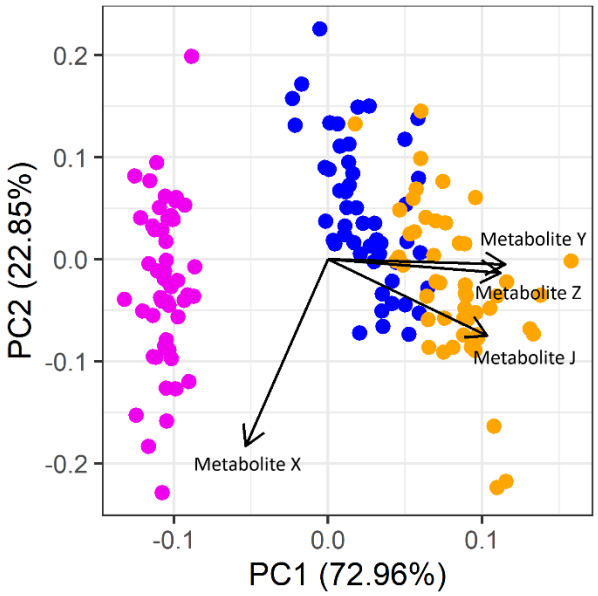


Figure 7: PCA biplot (a graphical representation of both score and loadings PCA plots). Each point in the plot represents a different metabolomic profile corresponding to a specific sample group (*i.e.* magenta points: group A; blue points: group B; orange points: group C); each black arrow represents a higher loading that corresponds to a specific metabolite (i.e. metabolite X, Y, J, and Z) that determine the separation and discrimination between groups.

As reported above, the supervised approaches, in contrast with the unsupervised ones, need previous knowledge to build efficient models able to evaluate the effect of interest. These approaches could also be divided into two main sets: methods based on projection and data reduction, and machine learning methods.

In the first set, Partial Least Squares (PLS)[102–104] and Orthogonal Projections to Latent Structures (OPLS)[24] are mostly used in metabolomics analyses. PLS is a multivariate linear regression approach similar to PCA in that it maximizes the covariance of two matrices, data **X** and response **Y**, to uncover the relationships between them. This method allows us to determine which variables (*i.e.* bins of bucket spectrum or a set of correctly assigned and quantified metabolites) of **X** are associated with the response (*i.e.* specific physiological or pathophysiological condition) and to perform new sample prediction. OPLS is a variant of the PLS approach, with the same predictive ability as PLS but an improved interpretation of relevant variables. This is possible by decomposing the data into so-called "predictive" information linked to the **Y** response matrix and "orthogonal" structural information unrelated to the response, such as instrumental or biological fluctuation. The multilevel PLS (M-PLS)[25] analysis is used to solve difficulties caused by substantial inter-individual variability. The intra-individual variability is separated from the inter-individual variability in an M-PLS by removing the individual specific average, allowing to consider in the analysis only the intra-individual variability. To improve the sample group discriminations, these reduction supervised methods can be applied in combination with Canonical Analysis (CA), which

enables discriminant projection by maximizing the inter-class variability and minimizing the intra-class variability.

In the context of machine learning methods, k-Nearest Neighbours (k-NN)[105], Support Vector Machines (SVM)[26], and Random Forest (RF)[27–29] are mostly used in metabolomics. k-NN is a simple, easy-to-implement supervised machine learning algorithm that can be used to solve both classification and regression problems; it assumes that similar variables exist in close proximity, and, for this reason, it works in a local neighbourhood around the object to be classified. The determination of the neighbourhood depends on the Euclidean distance and the closest *k* objects are used to classify the new sample. In this way, *k* is fundamental to building the model: small *k* numbers can determine a model vulnerable to considerable statistical fluctuation while big *k* numbers flatten many characteristics of the distribution, reducing statistical errors.

SVM is commonly used to solve classification problems, but it could be also used for regression. It uses a statistical paradigm to construct a "borderline" (*i.e.* a line, a plane, or, for more than three classes, a hyper-plane) tract in a sample place able to maximize the discrimination and separation of two (or more) classes.

Actually, RF represents the most important and powerful machine learning algorithm used to solve both classification and regression problems in metabolomics. Many strengths are related to this method: it can work with many predictors simultaneously; it avoids overfitting problems; it is applicable when there are more variables than samples; the percentage of trees in the forest that assigns a sample to a class can be interpreted as a probability of class membership; it provides an unbiased estimate of the classification error using out-of-bag (OOB) samples, eliminating, in this way, the need to perform additional cross-validation. The algorithm works through a few repeated steps:

1) the original data are randomly divided using bootstrapping into training and test sets;
2) an ensemble of decision trees is grown using the training set. In this step, each tree is built on randomly selected NMR variables at each decision node;
3) after the construction of all decision trees, an unbiased assessment of the classification error using the OOB samples is performed;
4) the estimation of the model performances, starting from a confusion matrix (see **Table 4**), is given.

As already reported, except for the RF algorithm, the main risk of the supervised models is the overfitting of data. To overcome this limit, it is necessary to validate the results using an independent validation set, if available, or with the cross-validation (CV) methods. These

methods require an initial splitting of the dataset into training and validation sets. After this step, one (Leave-One-Out (LOO)) or many (Leave-Many-Out (LMO)) samples are randomly removed from the training set and, a second time, they are tested. The model is built on the training set, while excluded samples are used to evaluate the model performances, constructing a confusion matrix that expresses sensitivity, specificity, and accuracy, as reported in **Table 4**.

Confusion matrix			
		Predicted classes	
		Group A (positive)	Group B (negative)
Actual Classes	Group A (positive)	True Positive (TP)	False Negative (FN)
	Group B (negative)	False Positive (FP)	True Negative (TN)

Model performance measures	
Specificity	$\frac{TN}{TN + FN}$
Sensitivity	$\frac{TP}{TP + FP}$
Precision	$\frac{TP}{FP + TP}$
Accuracy	$\frac{TN + TP}{TN + FN + TP + FP}$

Table 4: structure of the confusion matrix and model performance measures.

3.5.2 Univariate analysis

Univariate analysis is fundamental in identifying a potential association between a metabolite or a set of metabolites – separately and independently examined – with a specific clinical outcome or physiological condition.

Taking into account that the concentrations of the molecular features are not normally distributed, it is appropriate to use univariate non-parametric tests. In particular, for comparing two-group data – matched samples or repeated measurements on the same samples (pairwise comparison) – Wilcoxon-Mann-Whitney test[31,32] (the non-parametric version of the classical Student *t*-test[106]) is used. To compare more than two independent group data, originating from the same distribution, Kruskal-Wallis test[30] (the non-parametric version of Analysis of Variance (ANOVA)[107]) or Friedman test[108] are employed. Recently, the Linear mixed-effects

models (LMMs)[33,34] are considered robust and powerful univariate tools useful for the identification of molecular features significantly related to or affected by a specific factor of interest[109]. An LMM is a statistical model that incorporates fixed effects, which are constant across all individuals considered in the specific study, in addition to random effects, which vary across all individuals. The generic formula that describes an LMM is reported below:

$$y_{ij} = \beta_0 + \mu_{0j} + \sum \beta_{ik} X_{jk} + \epsilon_{ij} \quad (1)$$

where:

- y_{ij} corresponds to the concentration of the ith metabolite for the jth sample;
- β_0 is the overall mean or baseline level of metabolite concentration across all samples;
- μ_{0j} is the random effect of the j^{th} sample (or a selective individual variation);
- ϵ_{ij} is the random error of the model associated with metabolites and samples;
- k is the number of fixed effects, which account for the effect of a factor of interest (treatment) or an endogenous factor (age and gender), represented as X_k;
- β_k is the coefficient of the effect associated with factor X_k.

The significance of all these tests is expressed as a *P*-value, which is considered the probability of obtaining an equally extreme result equal to or a more extreme one than that what was observed, assuming the null hypothesis to be true. When the *P*-value is less than the threshold level at 0.05 or 0.01, the null hypothesis is rejected and the specific molecules feature and/or variable is considered statistically different in the observed groups. Due to a large number of variables tested, the application of statistical approaches in an *-omics* context usually needs to proceed with the correction for multiple tests necessary to reduce the probability of having false positives. Bonferroni[110,111] and Benjamini-Hochberg[112] are the most widely used corrections methods.

An important role in metabolomics in determining metabolic features associated with a specific physiological or pathophysiological condition is embodied by the Receiver-Operating Characteristic (ROC) and correlation analyses. In more detail, ROC curves are graphical representations that illustrate the diagnostic ability of a binary classifier. The curve is generated by plotting sensitivity (see **Table 4** in **§3.2.1**) versus [1- specificity] for all possible thresholds of the test (for more details about the measures of the predictive model see **Table 4** in **§3.2.1**). The accuracy is derived from the area under the ROC curve (AUC-ROC). As reported in the schematic **Figure 8,** the area values equal to 1 represent a perfect test, while an area of 0.5 represents a worse test.

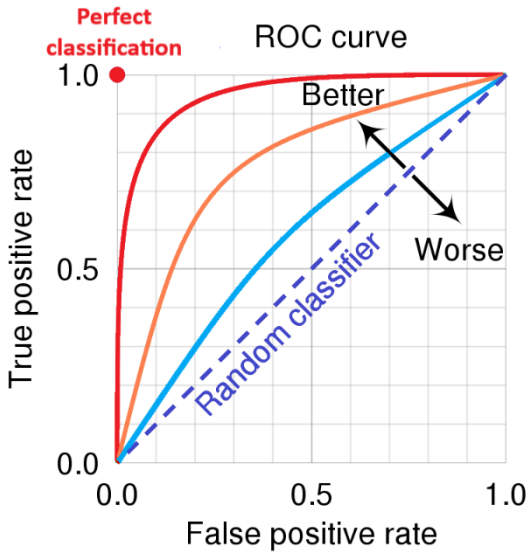


Figure 8: graphical representation of the ROC space that highlights a "better" and a "worse" classification test.

Moreover, to find, if present, a linear association between molecular features and specific clinical or biological data, the Pearson and/or Spearman correlations[113] can be calculated. Correlations are expressed by a coefficient (R, ρ), ranging from +1 (totally correlated), 0 (no correlation), and -1 (totally anticorrelated). Recently, to limit the presence of spurious correlations, robust correlation methods[114] have been developed and are increasingly used in metabolomics.

3.5.3 Molecular features association network analysis

To infer metabolite-metabolite/metabolite-lipoprotein/lipoprotein-lipoprotein association networks, the Probabilistic Context Likelihood of Relatedness based on Correlation (PCLRC)[115] algorithm was used. In order to remove non-significant background correlations, this algorithm provides a robust evaluation of the correlation using a resampling strategy in combination with the previously published Context Likelihood of Relatedness (CLR)[116] approach. The PCLRC algorithm gives like output a probability matrix **P** showing the likelihood p_{ij} for each revealed Spearman correlation r_{ij} (which is the *i*th and *j*th element of the Spearman correlation matrix **R**) between two metabolites *i* and *j*. In metabolomics, as the best compromise between network complexity and interpretability, it is recommended to consider correlation for which the probabilistic value p_{ij} was greater or greater than equal to 95%/96%, setting to 0 for all remaining correlations. The resulting combination between matrices **P** and **R** defines the specific network *A*.

To perform the differential connectivity analysis between molecular features association networks, given a specific network *A*, the connectivity χ_i^A for each metabolite *i* is described as:

$$\chi_i^A = \left(\sum_{j=1}^{J} |r_{ij}| \right) - 1 \quad (2)$$

Moreover, the differential connectivity $\Delta_i^{A,B}$ for each metabolite *i* among two networks *a* and *b* could be calculated as follow:

$$\Delta_i^{A,B} = \chi_i^A - \chi_i^B \quad (3)$$

A schematic procedure of the differential connectivity analysis is reported in **Figure 9**.

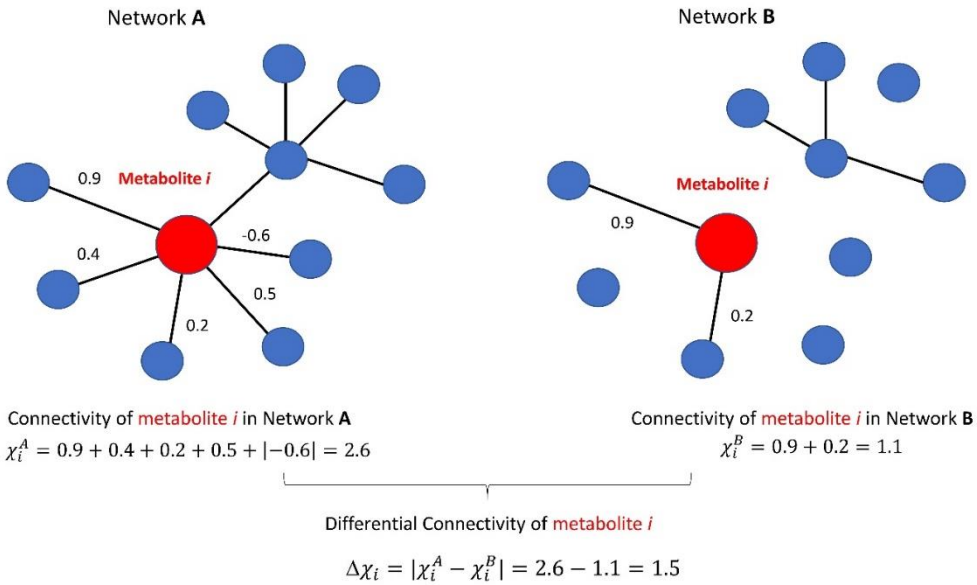


Figure 9: graphical representation of node connectivity and differential connectivity analysis. In this figure, each node represents a specific molecular feature (*i.e.* metabolite, lipid, lipoprotein, etc.) and the edge, connecting two nodes, represents the existence of an association, expressed in terms of correlation, between two nodes. This figure is taken from a published paper reported in **§4.1.1**.

The statistical significance of the differentially connected metabolites is determined by means of a permutation-test. In order to eliminate the relationship between variables and in order to maintain their variance, the columns of each input matrices have to be independently permuted defining a permutated matrix $\mathbf{X}_{(k)}$. The overall network estimation is performed on permutated data matrix, generating the related Spearman correlation $R_{(k)}$ analysis. These estimations were subsequentially used to assess, for each metabolite contained in the permutated matrix $\mathbf{X}_{(k)}$, the permutated connectivity (**Equation (4)**), and the permutated differential connectivity (**Equation (5)**):

$$\chi_{i,k}^{a} = \left(\sum_{j=1}^{J} \left| r_{ij}^{k} \right| \right) - 1 \tag{4}$$

$$\Delta_{i,k}^{a,b} = \chi_{i,k}^{a} - \chi_{i,k}^{b} \tag{5}$$

The permutation step was repeated k = 1000 yielding a null distribution D_i of permutated differential connectivity values. The significance of a given differential connectivity value $\Delta_i^{a,b}$ (estimated from the non-permutated original data) is calculated as a *P*-value, according to the following formula:

$$P - value = \frac{1 + \left(|D_i| > \left| \Delta_i^{a,b} \right| \right)}{k} \tag{6}$$

As reported in **§3.5.2**, to perform a more robust statistical analysis, the *P*-values may be corrected for multiple test comparisons, using, for example, the Benjamini-Hochberg approach[112].

Chapter 4

Results

4.1 Metabolomics for investigating human physiological and pathophysiological processes

Metabolomics provides a dynamic, comprehensive, and accurate image of human phenotype, increasing knowledge on both biological mechanisms of human physiological and pathological processes. In this perspective, metabolomics explains its potential role in characterizing metabolic profiles associated with a specific condition, and, also, in identifying metabolites or lipoproteins that play a key role in both normal physiology and the pathophysiology of several diseases.

In this methodological thesis, firstly, untargeted metabolomics has been applied to obtain new insights on different biological and physiological conditions, shedding light on the dimorphic mechanisms of aging, determining how an improving human well-being treatment (*i.e.* probiotics) could affect the metabotypes, and characterizing the metabolomic and lipoproteomic profiles associated with the inherited blood types (ABO and Rh systems). In this section, untargeted metabolomics has been also applied to characterize the metabolic components of two diseases, acute ischemic stroke and colorectal cancer, providing prognostic and diagnostic biomarkers of these specific pathologies, and to investigate the predictive potential of the lipoproteomic profile in cardiovascular risk.

To investigate the complex physiological aging process, that involves progressive systemic modification, affecting all levels of an organism, from molecules to organs [117,118], in the first studies here proposed, untargeted metabolomics was applied, integrating both standard univariate and multivariate analyses with the molecular features-association networks.

Applying a holistic metabolomic approach, the first study here presented aims at investigating the difference in terms of architecture and connectivity of sex-specific and clinical and biochemical parameters-specific serum metabolic association networks build considering a cohort of *n*=355 nonagenarians from the Italian Mugello study[119] (**§4.1.1**). In particular, comparing the metabolic association networks of nonagenarian women and men, our results evidenced that serum lipoproteins, branched chains amino acids (BCAAs), alanine, and ketone bodies show significant differences in terms of both architecture and connectivity. It is known that BCAAs are associated with a reduction in physical activities and a reduction in muscle protein synthesis (MPS)[120–122]. In human serum, levels of acetic acid, a ketone body,

are correlated with prolonged fasting and diabetic ketosis, potentially frequent in elderly people, but also with the decreasing of estrogens in post-menopausal women. The down-regulation of gonadotropin induces a glycolytic dysregulation, leading women's cells from a physiological aerobic metabolism to a ketogenic phenotype[123,124]. The same approach was also applied to identify, in the dimorphic clinical and biochemical parameters-specific networks, significantly differentially connected molecular features. The results obtained show the importance of lipoproteins in diseases, drug treatments, and familiarity disease, indicating their ability to participate in many pathophysiological mechanisms in nonagenarians; in fact, lipoproteins are not only biomolecules engaged in inter- and intra-cellular signaling regulation pathways, but also they can orchestrate inflammation processes and restore the homeostasis[125]. In the women's disease-specific networks, the rewiring of metabolic activity involves ketone bodies, BCAAs, threonine, and tyrosine, suggesting a potential greater women's predisposition and susceptibility to specific diseases, such as type 1 and 2 diabetes mellitus, sarcopenia, and cognitive impairment[126,127]. Considering the risk factors-specific networks related to women, only lipoproteins, negatively associated with estrogens in the post-menopausal period[128], are statistically differentially connected, suggesting their key role in the emerging of aging's pathologies; glutamine, glucose, proline, and BCAAs, increasing their metabolic activity, could be biomarkers to predict the emergence of neurodegenerative diseases, type 2 diabetes mellitus, and obesity and sarcopenia[129,130]. For bio-humoral parameters-specific networks, lipoproteins and metabolites are unrelated to the aging metabolic intrinsic dimorphisms and the floating of serum parameters[131].

Secondly, using the same cohort described above, applying classical statistical multivariate and univariate analyses, the sex-specific nonagenarian metabolic profiles were deeply investigated to highlight the molecular features directly involved in the elderly sexual dimorphism and to discover potential serum molecules playing important roles in predicting the development and progression of elderly disorders, focusing on cognitive impairment and geriatric depression (**§4.1.2**). Cross-validated (LOO-CV) OPLS-DA models were built on 1D-NOESY, 1D-CPMG, and 1D-diffusion serum bucketed spectra. We observed that the discrimination between nonagenarian women and men is less strong (accuracy ~70%; 70%, 67%, and 68% using 1D-NOESY, 1D-CPMG, and 1D-diffusion experiments respectively) than the same discrimination in a younger population (accuracy ~90%)[132], probably depending on the decreasing of the hormonal activity in elderly subjects. For a better understanding of metabolomic sexual dimorphic differences, the pairwise Wilcoxon test was applied on an array of 20 metabolites, 190 metabolites ratios, and 7 main lipoprotein parameters, correctly

assigned, quantified, and logarithmic scale normalized. In accordance with the results obtained in the previous study (**§4.1.1**), elderly women tend to have a higher level of creatine and lower BCAAs – in particular leucine and valine – concentrations, associated with the decreasing of MPS and with several reductions in motor activity[120,121,133]. In accordance with literature[56,134], higher levels of glycine, cholesterol, triglycerides, HDL, LDL, Apo A1, and Apo B100 concentrations were observed in elderly women compared with men. To identify potential metabolic predictors of the risk of geriatric cognitive impairment, mood disorder, and motor impairment logistic regression and ROC curve analyses were performed; we discovered that phenylalanine and glutamine/glucose ratio were associated with cognitive impairment, suggesting their role in correctly predicting the progress of pathology. Threonine/citric acid and threonine/pyruvic acid ratios were associated with geriatric depression, with a modest, but statistically significant, contribution. No metabolic associations were observed with elderly motor impairments. As a whole, our findings shed light on the sexual dimorphic mechanism of aging and reveal potential biomarkers of elderly cognitive and mood disorders.

The third study here presented, using an untargeted metabolomic approach, aims to evaluate and identify the potential correlation between different age-increasing groups and: i) the molecular features concentrations; ii) the pairwise molecular features correlations; ii) the pairwise molecular features ratios. This innovative approach was proposed to shed light on the dynamic of aging molecular mechanisms. This study is based on the TwinGene project[135] data, characterized by *n*=2139 participants, with an overall age range of 47.6–93.9 years (**§4.1.3**). We observed that linoleic acid, α-linoleic acid, and carnitine have, in the women cohort, a positive correlation trend with age, while monoacylglycerols (MAGs) and lysophosphatidylcholines (LPCs) have, in the men cohort, a negative correlation trend with age. These results highlight, in women, the effect of the reduction of estrogen activity and the increase of testosterone levels on the linoleic acid metabolism and on the energy pool metabolism that induces the overall changes in cell membrane composition and cell cycle mechanisms[136,137]. In men, low levels of LPCs concentrations are directly connected with the reduction of MAGs levels by the MAG lipase enzymatic activity that induces mitochondrial dysfunction[138,139]. Analyzing the pairwise correlations among molecules, PCs/PCs correlations tend to have a positive trend associated with the average ages of women, while MAGs/LPCs correlations tend to have a negative trend associated with the average ages of men. These results, in both cases, suggest an age-dependent remodeling of fatty acid metabolism that induces, overall, remodeling of cell and mitochondrial membranes[140]. Evaluating the results of molecular features ratios, we reported that the decanoyl-L-carnitine/LPC ratio, in women, has a negative association with the increase in the average ages, while, in men, the ratios between L-carnitine/PC and L-acetylcarnitine/PC have a positive association with the increase

of age, suggesting, in both cases, a radical remodeling of the dynamic membrane fluidity and carnitine-shuttle activity[140,141].

To better understand the mechanisms underlying human physiology, it is essential to specify that several microorganisms exist and coexist in the human body, promoting the regulation of human metabolism and physiology[142]. In this perspective, probiotics are increasingly used, with the final aim of manipulating the composition of the gut microbiota and improving balanced microbial communities[58,143]. In **§4.1.4,** an untargeted metabolomics approach was applied to characterize and understand potential metabolic changes that could determine and affect the human phenotype after probiotic assumption. 21 healthy volunteers were enrolled in the study based on two different phases: i) Phase I, characterized by no supplementation of probiotics, in which 20 urine samples and one serum sample from each subject were collected; ii) Phase II, characterized by the supplementation of probiotics, in which 20 urine samples and one serum sample from each subject were collected. In Phase II, the subjects were randomly divided into two groups, named "High-dosage" (n=10 subjects) and "Low-dosage" (n=11 subjects). This study identified, performing PCA-CA models on urine ^{1}H-NMR spectra, that, during the probiotic treatment, individual discrimination decreases by 1% (99% in Phase I; 98% in Phase II). This difference of 1% was also maintained by dividing the cohort into dosage-dependent groups. This result could be related to the probiotic administration, but it is not possible to consider this effect as generating a solid structural change. The global and dosage-dependent effect of the probiotic assumption was investigated by comparing the entire urine spectra using M-PLS analysis. Moderate discrimination (80%) of urine metabolome between the two phases, considering the entire cohort, was observed; investigating a dose-dependent effect, interestingly and unexpectedly, the subjects treated with a low dose of probiotics tended to have a discrimination accuracy higher than the subjects treated with a high dose of probiotics (79% *vs* 61%). To describe metabolic variations, a mixed-effects regression model was implemented for each urinary metabolite. Ascending significant trends were observed for formate in both dosage groups and phases, acetoacetic acid for both dosage groups during phase II, hippurate for the low dosage group during phase I, acetone and 2-hydroxyisobutyric acid for the high dosage group during phase I. Decreasing trends were observed for acetoacetic acid in both dosage groups during phase I, trimethylamine-NO for both dosage groups during phase II, dTTP, and creatinine for the low dosage group during both phases, acetone for the high dosage group during phase II, and isoleucine for the low dosage group during phase I. The same approach was also performed on serum samples. Using an MPLS-DA model, a fair serum metabolomic profile

discrimination (77%) between the two phases, considering the entire cohort, was reported. The two dosage-dependent groups have a comparable discrimination accuracy (~77%). Serum acetone and pyruvate significantly increased in phase II for both dosage groups, while histidine, glutamine, creatinine, acetate, and citrate significantly decreased.

The ABO and Rh systems, which play a fundamental role in transfusion medicine and hematopoietic transplantation, are also related to the pathogenesis and pathophysiology of various human diseases, such as cardiovascular[144,145] and oncological diseases[144,146]. In this scenario, the work **§4.1.5** here presented is the first NMR-based metabolomics work in which the specific associations between circulating levels of 28 specific plasma metabolites, and 114 lipoproteins and the ABO/Rh blood group system were analyzed in a cohort of *n*=840 Italian healthy blood donors. Random Forest (RF) model was applied to investigate whether subjects with different ABO and Rh blood group systems could be discriminated from the metabolic and lipoproteomic profiles. Overall, weak classification models to discriminate between the different ABO groups were obtained (accuracy ~50%). In contrast, subjects with different Rh blood groups can be easily and accurately discriminated based on their metabolite and lipoproteomic profiles (accuracy=81.9%). The association between the levels of circulating plasma metabolites, and lipoproteins and the ABO and Rh groups was assessed using robust linear regression, correcting for sex and age. Significant associations for 8 out of 114 lipoproteins with ABO groups – results confirmed also with a post-hoc test – were reported; in particular, the subfractions of HDL (HDL1 and HDL2) and the subfraction of LDL (LDL2) resulted to be relevant in ABO lipoproteomic differences. It has been shown in both experimental and clinical studies that the higher plasma levels of LDL and cholesterol in non-O blood groups influence the susceptibility of these groups to develop cardiovascular diseases (CVDs), while in the O blood group the higher levels of HDL tend to play a protective role in these systemic pathologies[147–151], but the molecular mechanisms need to be more deeply investigated. The same analyses were performed on Rh groups. Significant associations of 7 out of 114 lipoprotein fractions and subfractions and 2 out of 28 metabolites with the Rh groups were reported. In particular, the lipid main fraction LDL related to triglycerides, and Apo B, creatine, and propylene glycol, the particle number of LDL4 and LDL5, the subfraction LDL4 and LDL5 related to Apo B and the subfraction LDL4 related to free cholesterol were associated with Rh blood factors. Little is known about the association between the Rh phenotype and the metabolomic/lipoproteomic profiles and the molecular role played by LDL4 and LDL5 in Rh^+ and Rh^- group subjects has never been deeply investigated. The presence of the D antigen on the RBCs membrane was found to be significantly associated with lower HDL, higher triglycerides, and, in particular, higher LDL levels than the Rh^- group;

this metabolic behavior could determine the major predisposition of Rh^+ to develop CVDs and lipidic metabolic syndromes[152,153].

Metabolomics, as said above, is an excellent tool for characterizing and preventing a pathophysiological status. In particular, metabolic perturbations are fundamental events that contribute to ischemic stroke – the leading cause of death and disability continuously increasing[154] –, to its progression and development of unfavorable outcomes[155,156]. Using retrospective data from the Italian multicentric observational MAGIC (MArker bioloGici nell'Ictus Cerebrale) study[157,158], an NMR-based metabolomics approach was applied to identify serum metabolic and lipidic predictors of three-month poor outcomes (*i.e.* mortality, development of neurological impairments, hemorrhagic transformation of the cerebral lesion, and non-response to intravenous thrombolysis) in acute ischemic stroke (AIS) patients (*n*=243), treated with intravenous recombinant tissue plasminogen activator (rt-PA) (**§4.1.6**). For this purpose, logistic regression and ROC curve analyses were performed on an array of 18 metabolites and 112 lipoproteins correctly assigned and quantified in serum samples collected before (t_1) and 24h after (t_2) the thrombolytic intervention. Adjusting for clinical and demographic determinants of unfavorable outcomes (*i.e.* age, sex, time onset-to-treatment, diabetes, hyperlipidemia, etc.), acetone and 3-hydroxybutyrate were associated with symptomatic hemorrhagic transformation and with non-response to rt-PA; while 24 hours after rt-PA levels of triglycerides-HDL and triglycerides-LDL were associated with three-month mortality. Cholesterol and phospholipids levels, mainly related to smaller and denser VLDL and LDL subfractions were associated with three-month poor functional outcomes. Also, associations between baseline-24 hours relative variation (Δ) in VLDL subfractions and ΔC-reactive protein, Δinterleukin-10 levels with hemorrhagic transformation were reported. All observed metabolic changes reflect a general condition of energy failure, oxidative stress, and systemic inflammation that characterize the development of adverse outcomes[159–161].

As for neurological disorders, cancer research is one of the most important fields investigated by *-omics* sciences, in particular metabolomics, with the aim to discover new biomarkers, and refine diagnostic tests and therapies. CRC is the third most commonly diagnosed cancer and the second leading cause of cancer-related death worldwide[162]. Early-stage diagnosis is associated with a good prognosis, however, survival declines substantially when the tumor is identified later and is already metastasized[163]. With these ideas in mind, in **§4.1.7**, a metabolomics approach was used to investigate the blood metabolite profiles in patients with CRC, polyposis, and healthy controls, based on the metabolic information reported by Zhu et *al.*[164]. In this work, the aim was to present an evaluation of the serum

metabolomic profiles of *n*=65 CRC and *n*=74 polyposis (PP) patients with respect to *n*=87 healthy controls (CTR) using a metabolite–metabolite association networks approach to investigate and explore the existence of molecular mechanisms underlying these different clinical profiles. Firstly, RF models show good discriminations for the comparisons between CRC and CTR and between CRC and PP obtaining an AUC of 0.875 and 0.871, respectively. Conversely, PP and CTR present only slight differences (AUC of 0.661). These data confirm that CRC patients develop systemic metabolic alterations that are not present, not only in healthy controls but also in PP patients. We compared the networks of CRC, PP, and CTR observing that CRC patients have an architecture completely different from the ones of PP and CTR. This difference is corroborated also by the differential connectivity analysis and by the analysis of the topological parameters. Nodes present in the CRC network are all directly or indirectly related to the different pathways involved in the energetic metabolism, depending on the fact that cancer cells need to meet a high energy demand to support cell proliferation and migration[165,166]. PP and CTR seem to have a similar architecture with more interconnections. Pathway analysis of the differentially connected metabolites reveals the involvement of phenylalanine, tyrosine, and tryptophan metabolism in both CRC and PP patients. Tryptophan metabolism is linked to the production of serotonin while phenylalanine is required to produce tyrosine which is catalyzed by phenylalanine hydroxylase. It has been shown that phenylalanine hydroxylase activity can be altered in inflammation or malignancy[167,168]. The landscape of metabolic alterations associated with polyposis appears to be more heterogeneous than the ones associated with CRC. We observed the involvement of glycolysis and glucogenesis and glycine, serine, and threonine metabolism as a signature for polyposis, together with fructose and mannose metabolism, associated with increased risks of PP and CRC[169].

To conclude, in **§4.1.8** is reported a contribution in a chapter in which the aim is to highlight the various aspects of metabolomics, focusing on: i) the effects on the individual metabolomic fingerprint of non-invasive treatment (*i.e.* diet or probiotic administration); ii) the characterization of metabolic fingerprints of specific diseases (*i.e.* celiac disease, various types of cancer, viral infections, and other diseases); ii) the effects of drugs on the disease fingerprint and on its reversal to a healthy metabolomic status.

4.1.1 Lipid and metabolite correlation networks specific to clinical and biochemical covariate show differences associated with sexual dimorphism in a cohort of nonagenarians

Francesca Di Cesare[1], Leonardo Tenori[1,2], Gaia Meoni[3], Anna Maria Gori[4,5], Rossella Marcucci[4,5], Betti Giusti[4,5], Raffaele Molino-Lova[6], Claudio Macchi[4,6], Silvia Pancani[6], Claudio Luchinat[1,2,7], Edoardo Saccenti[8]

[1] Magnetic Resonance Center (CERM), University of Florence, Sesto Fiorentino, Italy

[2] Department of Chemistry "Ugo Schiff", University of Florence, Sesto Fiorentino, Italy

[3] Giotto Biotech srl, Florence, Italy.

[4] Department of Experimental and Clinical Medicine, University of Florence, Florence, Italy.

[5] Atherothrombotic Unit, Careggi University Hospital, Florence, Italy.

[6] IRCCS Fondazione Don Carlo Gnocchi, Florence, Italy.

[7] Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine (CIRMMP), Sesto Fiorentino, Italy.

[8] Laboratory of Systems and Synthetic Biology, Wageningen University & Research, Wageningen, the Netherlands.

Published

Geroscience, 2022 Apr; 44(2):1109-1128

doi: 10.1007/s11357-021-00404-3

Candidate's contribuitions: NMR-data pre-processing, statistical analysis, interpretation of results, writing, and review of the manuscript.

GeroScience (2022) 44:1109–1128
https://doi.org/10.1007/s11357-021-00404-3

ORIGINAL ARTICLE

Lipid and metabolite correlation networks specific to clinical and biochemical covariate show differences associated with sexual dimorphism in a cohort of nonagenarians

Francesca Di Cesare · Leonardo Tenori · Gaia Meoni · Anna Maria Gori · Rossella Marcucci · Betti Giusti · Raffaele Molino-Lova · Claudio Macchi · Silvia Pancani · Claudio Luchinat · Edoardo Saccenti

Received: 9 April 2021 / Accepted: 13 June 2021 / Published online: 29 July 2021

Abstract This study defines and estimates the metabolite-lipidic component association networks constructed from an array of 20 metabolites and 114 lipids identified and quantified via NMR spectroscopy in the serum of a cohort of 355 Italian nonagenarians and ultra-nonagenarian. Metabolite-lipid association networks were built for men and women and related to an array of 101 clinical and biochemical parameters, including the presence of diseases, bio-humoral parameters, familiarity diseases, drugs treatments, and risk factors. Different connectivity patterns were observed in lipids, branched chains amino acids, alanine, and ketone bodies, suggesting their association with the sex-related and sex-clinical condition-related intrinsic metabolic changes. Furthermore, our results demonstrate, using a holistic system biology approach, that the characterization of metabolic structures and their dynamic interconnections is a promising tool to shed light on the dimorphic pathophysiological mechanisms of aging at the molecular level.

Keywords Aging · Differential network analysis · Lipidomics · Metabolomics · Network inference · Nuclear magnetic resonance · Sexual dimorphism

F. Di Cesare · L. Tenori · C. Luchinat
Magnetic Resonance Center (CERM), University of Florence, Sesto Fiorentino, Italy

L. Tenori · C. Luchinat
Department of Chemistry "Ugo Schiff", University of Florence, Sesto Fiorentino, Italy

G. Meoni
Giotto Biotech srl, Florence, Italy

A. M. Gori · R. Marcucci · B. Giusti · C. Macchi
Department of Experimental and Clinical Medicine, University of Florence, Florence, Italy

A. M. Gori · R. Marcucci · B. Giusti
Atherothrombotic Unit, Careggi University Hospital, Florence, Italy

R. Molino-Lova · C. Macchi · S. Pancani
IRCCS Fondazione Don Carlo Gnocchi, Florence, Italy

C. Luchinat
Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine (CIRMMP), Sesto Fiorentino, Italy

E. Saccenti (✉)
Laboratory of Systems and Synthetic Biology, Wageningen University & Research, Wageningen, the Netherlands
e-mail: edoardo.saccenti@wur.nl

Introduction

Nonagenarian and centenarian people represent a considerable increasing fraction of the world population, concentrated, above all, in economically developed countries [1, 2]. *In Europe, Italy and France hold the record for the number of living centenarians. According to a 2019 statistic, in Italy, 1% of the population is 90 years older or more, and between 2000 and 2019 the number of centenarians (85% women) increased from 11000 to more than 14456, and the number of ultra-centenarians (>105 years, 94% women) increased 136%, from* 472 to 1112 [3].

Aging is associated with irreversible variations in biological, pathophysiological, and psychological dynamics [4–8]: these age-related changes result in a decline of cognitive, motor, and sensory functions and in an increase of susceptibility to disease and disease frequency, a poor quality of life, and increased mortality [9, 10], defining a clinically and biologically heterogeneous population.

Many different fundamental biological processes, such as inflammation, cellular and immune senescence, mitochondrial dysfunction, and reduced resistance to oxidative stress are the main mechanisms at the basis of the aging process [10]. The progressive decline of physiological functions reflects changes happening at the molecular, organelle, cell, tissue, and, finally, the whole organism level [11, 12]. Although individually these biochemical and molecular alterations underlying these processes may have only a modest effect on aging, taken together they involve a complex network of biomolecular mechanisms acting across multiple organs and at different molecular levels [13].

Accumulation of molecular damage has been proposed to be among the mechanisms driving aging [14–16], and this may include not only oxidative stress and DNA mutations, but also errors in protein synthesis and by-products of enzymatic reactions [17]. In this light, the use of *high-throughput omics techniques, like genomics, transcriptomics, proteomics, and metabolomics, offers a great promise for the* understanding of the mechanisms that underlie aging [11, 18, 19]. In particular, the analysis of metabolic signatures associated with age and the comprehensive characterizing and understanding of the structures, functions, and interactions between metabolites and lipidic components, can shed light on the potential mechanism that could influence aging and longevity.

Nuclear magnetic resonance (NMR)-based metabolomics offers the possibility to quantify and investigate hundreds of various metabolites, lipid fractions and sub-fractions [20–23], detectable in biofluids, providing a global image of the complex metabolic, biological and biophysical processes associated with health [10, 22, 24] and disease [23, 25–27].

Integrative analysis of NMR-based metabolomics data using systems biology approaches focusing on the interactions and relationships among biochemical molecules like protein and metabolites can offer a holistic representation of the metabolic structures, indispensable for the understanding of the molecular mechanism underlying aging [28, 29].

Networks and network analysis of blood metabolites, lipid fractions and sub-fractions association networks are fundamental tools to extract information on the status of a biological system since correlation among metabolites and lipids concentration profiles can be used to model and to infer, at least partially, the structure of the underlying biological network [30]. In addition, these network models can be linked to clinical information such as biochemical parameters, risk factors, co-morbidities, allowing the analysis of the relationships existing between different network structures and patient clinical characteristics.

In this work, we take an integrative approach to investigate the associations between metabolite and lipoprotein/lipid networks and the range of clinical, biochemical, environmental, socio-demographic parameters collected on a cohort of 355 nonagenarians from the Italian Mugello study [31]. The goal of the present analysis is twofold: first, we wanted to understand the association between different networks' structures and sex, because of its relevance for gender medicine of aging [32, 33]; second, we wanted to explore the complex web of relationship existing between clinical parameters, risk factors, and comorbidities and blood metabolites and lipids association patterns.

Material and methods

Study description

Samples were collected from the participants in the Mugello study, an epidemiological survey conducted from January 2010 to December 2011 in the Mugello area (north-eastward of Florence, Tuscany, Italy, see Fig. 1) [31]. The original study comprised 356 subjects—of

which 96 men (27%) and 260 women (73%), with an age range of 84–103 years and of 88–105 years and with mean age 92.6 (± 3.4) and 93.2 (± 3.2) years, respectively. All participants, at the time of the cross-sectional survey, were subjected, by a trained physician, to a series of home-based structured interviews and medical examinations regarding clinically relevant geriatric conditions. Blood samples were also collected to perform routine laboratory tests. We refer the reader to the original publication for more details on the study design and the protocols that have been followed [31].

Overview of clinical variables

For the analysis presented in this article, we selected a sub-set of M=101 clinical covariates which were grouped into 5 categories:

1. Diseases (m=20, dichotomous variables: 0-1): describing the existence of a specific medical condition, *i.e.*, myocardial infarction, congestive heart failure, peripheral vascular disease, hemiplegia, hypertension, dyslipidemia, dementia, cerebrovascular disease, diabetes (with and without organ damage), cancer, leukemia, disability, elderly depression (evaluated using the 15-item Geriatric Depression Scale screening questionnaire validated for the geriatric population [34]), cognitive impairment (evaluated using the Mini-Mental State Examination validated questionnaire [35]) and motor impairment (evaluated using Short Physical Performance Battery and Time up and go questionnaires [36, 37]).
2. Familiarity diseases (m=5, dichotomous variables: 0-1): describing the existence of familiarity for cardiovascular, respiratory, and cerebrovascular diseases, dementia, and cancer.
3. Drugs treatments (m=13, dichotomous variables: 0-1): indicating the presence of ongoing pharmacological treatment for diseases described above.
4. Risk factors (m=12, continuous and dichotomous variables) including socio-demographic variables (age, education, sleep alertness, civil status, living with and smoke habit) and physical parameters (Body Mass Index [38–40], Windsor index or systolic ankle pressure measured using Ankle-brachial index [41, 42], Physical activity scale for elderly score [43], Handrig index [44], Mediterranean Diet Score [45, 46]).

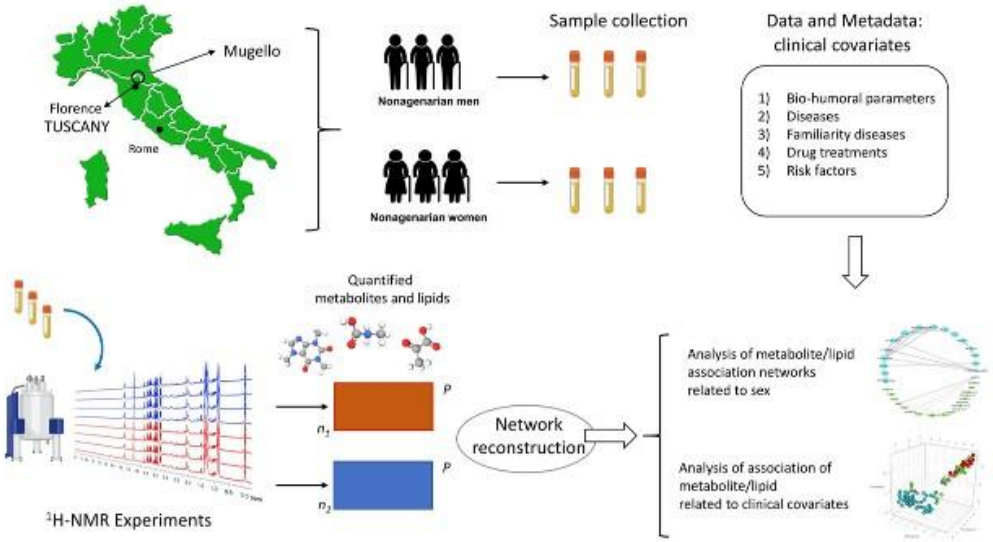


Fig. 1 Graphical overview of the study and the data analysis strategy used in this study. Bio-humoral parameters refer to biochemical information such as blood count, mean cell volume, mean cell hemoglobin, thyroid hormones; see Table 1 for more details. Data are from the Mugello study [31]

5. Bio-humoral parameters (m=51: continuous and dichotomous) including, *i.e.*, complete blood count, mean cell volume, mean cell hemoglobin, thyroid hormones, cholesterol, HDL, LDL, and glycemia.

The variables considered and associated statistics are listed in Table 1.

Ethical considerations

The Mugello study [31] was conducted in agreement with the principles of the Helsinki Declaration on Clinical Research involving human beings (1964) and was approved by the Don Carlo Gnocchi Foundation Ethics Committee. Informed written consent was obtained from all participants or from their delegates before their inclusion in the original study.

Experimental methods

Sample collection

Blood samples were collected after overnight fasting, centrifuged at 2000*g* for 10 min at 4°C, and stored in aliquots at −80° until analyses, following standardized operating procedure as described in Bernini et al. [47].

NMR experiments

Serum samples were prepared for NMR analysis as described by Bernini et al. [47] and acquired using a Bruker 600 MHz spectrometer (Bruker BioSpin s.r.l., Germany). NOESY 1D presat (one-dimensional NOESY) experiments were used to measure selectively low and high molecular weight molecules. Metabolites, lipoproteins, lipid fractions and sub-fractions were assigned, identified, and quantified using the AVANCE IVDr (Clinical Screening and In Vitro Diagnostics (IVD) research with B.I. Methods, Bruker BioSpin) [48], and the principal metabolites and main lipid fractions considered in this study are listed as follow: alanine, creatine, glutamic acid, glutamine, glycine, histidine, isoleucine, leucine, lysine, phenylalanine, proline, threonine, tyrosine, valine, acetic acid, citric acid, lactic acid, acetoacetic acid, pyruvic acid, glucose, main parameter (MP) triglycerides, main parameter (MP) cholesterol, main parameter (MP) LDL, main parameter (MP) HDL, main parameter (MP) Apo A1, main parameter (MP) Apo A2, and main parameter (MP) Apo B100.

A complete list of all lipid fractions and sub-fractions is presented in Supplementary Table S1.

Statistical methods

Data pre-processing

Only covariates with less than 25% missing data were considered; missing data were imputed using a Random Forest approach as implemented in R package missForest [49]; default parameters were used. All variables were log-transformed before analysis. One sample ("F_C_138") was excluded after a check of the spectra due to low quality shimming and remove from all subsequent analyses: the actual number of samples used in the present investigation is n=355.

Extraction of metabolic information related to clinical variables

For each blood metabolite and lipid fraction, we extracted the variation coming from a given clinical covariate, i.e., one of the M=101 covariates (see Table 1) recorded, using the method proposed by Bartzis et al. [50]. This approach extracts, from each metabolite/lipid, the part of concentration that is associated with a given covariate, thus implicitly adjusting for the remaining M-1 covariates. The rationale is that different metabolites/lipids sharing similar correlation/association with the same clinical covariable tend to be close to each other in the network, thus providing a better representation of the underlying biological phenomena [50].

Briefly, be $Y^{(p)}$ the $(n \times 1)$ vector of concentrations of the p-th metabolite or lipid component (with p = 1,2, ..., P) measured on n=355 subjects and be $\mathbf{X}$ the $(n \times M)$ matrix containing the M=101 clinical parameters (covariates) recorded on the n subjects. Let be $X_m$the $n \times 1$ vector containing the values for m-th clinical variables and $X^{(-m)} = \{X_1, X_2, \ldots, X_{M-1}\}$ the remaining $M-1$ clinical variables. The information $\widehat{Y}^{(p)}$ of a metabolite or lipid component p associated with a specific clinical covariate m was estimated by regressing $Y^{(p)}$ on $\mathbf{X}$ and retaining only the main effects and first-order interactions of covariate X_m:

$$\widehat{Y}^{(p)} = \widehat{\beta}^{(p)} X_m + \sum_{\delta \in \Delta} \widehat{\eta}_{\delta}^{(p)} X_m * \prod_{j=1}^{M-1} X_j^{\delta_j} \quad (1)$$

Table 1 Descriptive statistics of clinical variables, divided into five separated categories, stratified by sex. For continuous variables the mean ± SD (standard deviation) is reported. *GDS* geriatric depression scale, *MMSE* Mini-Mental State Examination, *SPPB* short physical performance battery, *MCV* mean corpuscular volume, *MCH* mean corpuscular hemoglobin, *MCHC* mean corpuscular hemoglobin concentration, *RDWCV* red blood cells distribution width coefficient of variation, *RDWSD* red blood cells distribution width standard deviation, *PDW* platelet distribution width, *MPV* mean platelet volume, *PLCR* platelet large cell ratio, *GOT-AST* aspartate aminotransferase, *GPT-ALT* alanine transaminase, *γ-GT* gamma-glutamyl transferase, *TSH* thyroid-stimulating hormone, *WBCs* white blood cells, *RBCs* red blood cells, *HCT* hematocrit, *HbA1C* hemoglobin A1c, *CRP* C-reactive protein, *ACE* Angiotensin-converting enzyme, *ABI* Ankle-Brachial Index, *BMI* Body Mass Index, and *MDS* myelodysplastic syndromes

		women (n = 259, 73.0 %)	men (n = 96, 27.0 %)
Diseases	Myocardial infarction (%)	15.4	12.5
	Congestive heart failure (%)	23.2	15.6
	Peripheral vascular diseases (%)	22.4	11.5
	Hypertension (%)	55.9	58.3
	Dyslipidemia (%)	12.4	6.3
	Dementia (%)	13.5	13.5
	Diabetes (%)	12.7	15.6
	Diabetes without organ damage (%)	7.3	13.5
	Diabetes with organ damage (%)	5.8	2.1
	Cancer (%)	13.9	9.4
	Leukemia (%)	0.4	0.0
	Disability (%)	62.6	88.5
	Motor impairment code	9.0 ± 6.7	7.5 ± 7.4
	GDS code	0.6 ± 0.5	1.8 ± 0.8
	Depression (%)	58.7	77.1
	MMSE (%)	56.4	50.0
	SPPB (%)	63.3	64.6
	Time up and go (%)	58.7	64.6
	Hemiplegia (%)	0.8	1.0
	Cerebrovascular diseases (%)	21.8	20.8
Bio-humoral parameters	MCV (FL)	90.4 ± 5.3	90.3 ± 6.2
	MCH (pg.)	29.7 ± 2.9	29.7 ± 2. 6
	MCHC (g/dL)	33.0 ± 1.0	33.0 ± 1.1
	RDWCV (fL)	14.7 ± 1.3	14.8 ± 1.4
	RDWSD (fL)	47.3 ± 4.1	47.5 ± 5.8
	PDW (fL)	13.3 ± 2.2	13.9 ± 2.5
	MPV (fL)	10.6 ± 1.0	10.9 ± 1.0
	PLCR	29.6 ± 6.9	31.3 ± 6.5
	GOT-AST (IU/L)	20.2 ± 9.4	20.1 ± 5.9
	GPT-ALT (IU/L)	15.0 ± 9.4	14.2 ± 5.2
	γ-GT (IU/L)	25.2 ± 27.6	15.6 ± 4.1
	Neutrophil (x10^3/μL)	4.0 ± 2.7	4.1 ± 1.7
	Lymphocyte (x10^3/μL)	1.8 ± 0.9	1.8 ± 0.7
	Monocyte (×10^3/μL)	0.5 ± 0.2	0.5 ± 0.2
	Eosinophil (x10^3/μL)	0.2 ± 0.1	0.2 ± 0.1
	Basophil (x10^3/μL)	0.02 ± 0.02	0.03 ± 0.03
	Neutrophyl – formula	59.7 ± 9.5	60.7 ± 9.7
	Lymphocyte – formula	28.9 ± 8.8	27.7 ± 8.5
	Monocyte – formula	7.7 ± 2.3	7.9 ± 2.4

Table 1 (continued)

		women (n = 259, 73.0 %)	men (n = 96, 27.0 %)
	Eosinophil – formula	3.2 ± 2.1	3.4 ± 2.0
	Basophil – formula	0.4 ± 0.4	0.4 ± 0.3
	Creatinine (mg/dL)	1.0 ± 0.5	1.1 ± 0.7
	Neutrophyl/lymphocyte	0.5 ± 0.3	0.5 ± 0.3
	Platelets (x10^3/μL)	219.6 ± 93.9	206.5 ± 64.0
	Na^+ (mE/ql)	138.9 ± 3.0	138.6 ± 2.9
	K^+ (mEql)	4.3 ± 0.5	4.4 ± 0.5
	Cl- (mEql)	101.7 ± 7.8	102.0 ± 4.0
	Total proteins (g/dL)	6.4 ± 0.6	6.5 ± 0.7
	Albumin (g/dL)	56.2 ± 4.6	56.1 ± 4.8
	α1-G (g/dL)	3.9 ± 1.3	4.0 ± 1.5
	α2-G (g/dL)	12.0 ± 2.0	12.0 ± 1.8
	β-G (g/dL)	12.2 ± 1.7	12.3 ± 1.9
	γ-G (g/dL)	15.6 ± 3.7	15.6 ± 4.1
	A/G	1.3 ± 0.3	1.3 ± 0.3
	T3 (pg/mL)	2.8 ± 0.5	2.9 ± 0.6
	T4 (ng/dL)	0.9 ± 0.2	0.9 ± 0.3
	TSH (μUI/mL)	2.3 ± 6.3	2.1 ± 4.0
	WBCs (x10^3/(μL)	6.3 ± 1.9	6.7 ± 2.1
	RBCs (x10^6/(μL)	4.3 ± 0.6	4.4 ± 0.6
	Hemoglobin (g/dL)	12.9 ± 1.5	12.9 ± 1.7
	HCT (%)	38.9 ± 4.8	39.0 ± 4.8
	glycemia (mg/dL)	94.6 ± 26.7	93.3 ± 22.7
	HbA1C (g/Hb)	5.6 ± 0.8	5.6 ± 0.8
	Total cholesterol (mg/dL)	190.5 ± 42.0	192.9 ± 41.2
	Cholesterol (mg/dL)	0.4 ± 0.5	0.4 ± 0,5
	HDL (mg/dL)	57.4 ± 16.4	59.0 ± 18.7
	LDL (mg/dL)	110.3 ± 33.7	111.0 ± 32.8
	Triglycerides (mg/dL)	113.9 ± 47.5	115.1 ± 55.1
	CRP (mg/L)	0.9 ± 2.1	1.2 ± 2.8
	CRP (%)	54.4	64.6
	Inflammatory protein (mg/L)	9.2 ± 20.6	11.9 ± 27.5
	Benzodiazepine (%)	15.4	18.8
Drug treatment	Antidepressant (%)	18.5	19.8
	Diuretics (%)	52.5	44.8
	Beta-blockers (%)	10.4	15.6
	Ca^{++} channel blockers (%)	19.7	16.7
	ACE inhibitors (%)	39.0	35.4
	Vasodilators nitrates (%)	24.3	21.9
	Oral anticoagulant (%)	6.2	5.2
	Heparin (%)	11.6	9.4
	Antiplatelet (%)	40.0	35.4
	Antihyperlipidemic (%)	9.3	5.2
	Insulin (%)	4.6	3.1

Table 1 (continued)

		women (n = 259, 73.0 %)	men (n = 96, 27.0 %)
	Oral antidiabetics (%)	9.3	12.5
	Age (years)	93.2 ± 3.2	92.6 ± 3.4
Risk factors	Civil status (% of married person)	95.8	97.9
	Living with (number of person)	2.8 ± 1.2	2.6 ± 1.2
	Education (years)	4.2 ± 2.6	4.3 ± 2.6
	Tobacco exposure (%)	13.9	72.9
	Winsor index	1.0 ± 0.3	1.1 ± 0.3
	ABI code > 1 (%)	23.6	36.5
	Handrig index (kg)	14.3 ± 6.9	15.8 ± 7.
	Pase score (%)	38.6	52.1
	Sleep-alertness (%)	4.3	4.2
	BMI (kg/m^2)	24.8 ± 4.7	25.1 ± 3.4
	MDS index	34.0 ± 3.9	34.5 ± 3.3

where $\sum_{\delta\in\Delta}\widehat{\eta}_{\delta}^{(p)}\mathrm{X}_m{}^{*}\prod_{j=1}^{M-1}\mathrm{X}_j^{\delta_j}$ models all main effects and high-order interactions in terms of clinical variables.

For each clinical parameter m the procedure is repeated for all P metabolites and lipid component to obtain M = 101 $n \times P$ data sets $\mathrm{Y}_m = \{\mathrm{Y}_{(1)},\mathrm{Y}_{(2)}, \ldots, \mathrm{Y}_{(p)}\}$ containing the *part of the measured* metabolite and lipid concentrations associated with each one of the 101 clinical parameters.

Network analysis

Network concepts

We briefly review here some network concepts. A network is a graphical representation of the relationships between objects, called nodes [51]. In a biological network, the nodes are molecular components, like genes, proteins, or, like in this study, metabolites and lipid components. The (existence of a) relationship between two nodes (molecular components) is represented by an edge connecting the two nodes. The type of association among the molecular features can be diverse in nature: in a protein-protein interaction network, edges represent the existence of physical interaction between proteins; in a metabolite-metabolite association network in which two metabolites are connected if their concentration levels are correlated.

Mathematically, a network can be represented as an adjacency (also called connectivity) matrix **A**: the rows and columns of the **A** represent the nodes whereas the entries a_{ij} represent edges. A network is said to be *unweighted* if the edges a_{ij} describing the association between node i and j are either 1 or 0:

$$a_{ij} = \begin{cases} 1 & \text{if } (i,j) \text{ are associated} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

If the strength or magnitude of the relationship can be quantified, a weight can be given to the edge; then, the network is said to be *weighted*: in this case, the elements of a weighted adjacency matrix **A** are real numbers indicating the strength of the interaction, and can vary, for instance, in the [−1, 1] range if the correlation is used as an index for the association.

Reconstruction of metabolite and lipid association networks

The Probabilistic Context Likelihood of Relatedness (PCLRC) [52] algorithm was used to build metabolite and lipid association networks using Spearman correlation as a measure of association [53]. The algorithm allows robust estimation of correlation employing a resampling strategy in combination with a modified version of the Context Likelihood of Relatedness (CLR) [54] to remove non-significant background correlations. The algorithm returns a probability matrix **P** with values between 0 and 1 that was used to filter significant correlation r_{ij} between pairs of metabolites /lipids. In particular

$$r_{ij} = \begin{cases} r_{ij} & \text{if } p_{ij} \geq 0.90 \\ 0 & \text{if } p_{ij} < 0.90 \end{cases} \quad (3)$$

We built a metabolite/lipids association network for each of the 101 $n_1 \times P$ (for women) and $n_2 \times P$ (for men) data sets $Y_m=\{Y_{(1)},Y_{(2)}, ..., Y_{(p)}\}$ containing the *part of the measured* metabolite and lipid concentrations associated with each one of the 101 clinical parameters. We analyzed data for males and females separately, obtaining a total of 202 metabolite/lipids association networks that were divided into 5 categories (*S*= diseases, bio-humoral parameters, familiarity diseases, risk factors, and drug treatments).

Network differential connectivity analysis

Each node in a network can be characterized using measures that can be derived from the patterns of its association. A very common measure is the node degree or connectivity [55, 56], that is the number of its connection. For a $p \times p$ network **A**, the connectivity of the node *i* is given by:

$$\chi_i = \sum_{j>i} |a_{ij}| \quad (4)$$

Given a network A, the connectivity χ_i^A for metabolite/lipid i is defined as

$$\chi_i^A = \left(\sum_{j=1}^{J} |r_{ij}|\right) - 1 \quad (5)$$

If the network is unweighted, it holds $0<\chi_i<p-1$. If the network is weighted, the range of the connectivity depends on the nature of the association measure. If the absolute value of the correlation is used, like in this study, χ_i still ranges between 0 and p−1, in which case, it means that the molecular feature represented by node ai is perfectly correlated with all other nodes in the network.

Two networks **A** and **B** associated with two different conditions or groups (such as those built from men and women samples, or from samples from case-control patients) can be compared, implementing a so-called differential network analysis [28, 57].

The differential connectivity ($\Delta_i^{A,B}$) of a metabolite/lipid *i* between two networks A and B is defined as

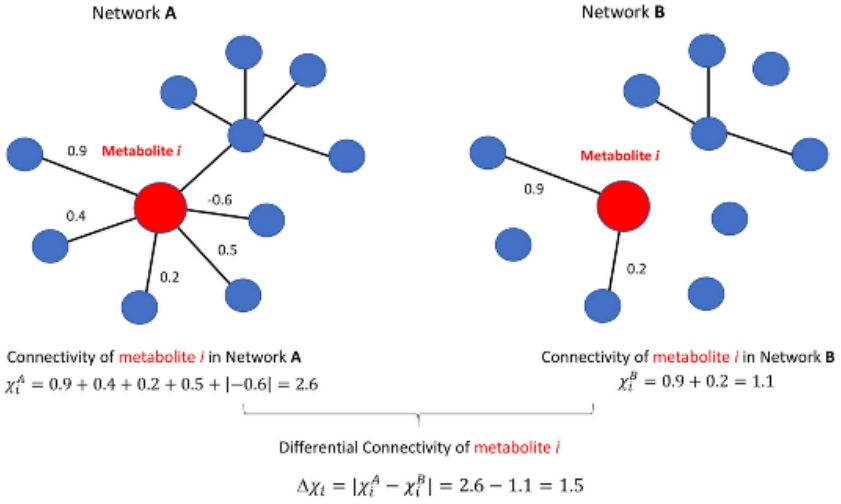


Fig. 2 Graphical illustration of the concept of node connectivity and differential connectivity used in this study. Each node represents a molecular feature (metabolite, lipid). The edge connecting two nodes represent the esistence of an association between two nodes, in this case expressed by correlation; the weight of the edge is given by the (absolute value) of the correlation. Figure adapted from [53]

$$\Delta_i^{A,B} = \chi_i^A - \chi_i^B \quad (6)$$

The concept of differential connectivity is exemplified in Fig. 2.

Estimation of the statistical significance of differential network connectivity

The statistical significance of the differential connectivity ($\Delta_{i,k}^{A,B}$), was assessed by means of a permutation-test. Briefly, the columns of every Y_m matrix are independently permuted to obtain a permutated matrix $\mathbf{X}_{(k)}$whose column mean and variance are unchanged but the association between the elements of different columns is destroyed.

For each metabolite/lipid the differential connectivity was calculated for networks *a* and *b* built from the permuted data:

$$\Delta_{i,k}^{A,B} = \chi_{i,k}^A - \chi_{i,k}^B \quad (7)$$

and the overall procedure was repeated k=100 times to create a null distribution D_i of permutated differential connectivity values. The significance of a given differential connectivity value $\Delta_i^{A,B}$ (calculated on the original data) was calculated as a *P*-value using the following formula (#() indicates the number of elements):

$$P\text{–}value = \frac{1 + \#\left(|D_i| > |\Delta_i^{A,B}|\right)}{k} \quad (8)$$

Multivariate analysis of association networks

Covariance simultaneous component analysis (COVSCA) [58] was performed to analyze simultaneously the (dis)similarities of the sets of K=101 metabolite and lipid association networks. The K association matrices are modeled as the number of low-dimensional prototypes (L << K):

$$\mathbf{S}_k \cong \sum_{l=1}^{L} c_{kl} \mathbf{Z}_l \mathbf{Z}_l^T \quad (9)$$

where $c_{kl} \geq 0$ (l = 1, 2, …, L) are weight coefficients and $\mathbf{Z}_l\mathbf{Z}_l^T$ are the prototypical covariance matrices that characterize the loadings set $\mathbf{Z}$ of dimension $J \times R_l$ that hold together for all C_k.

The COVSCA model was fitted separately for both women's and men's data with 3 rank-1 prototype matrices (R=3) as the best compromise between the goodness-of-fit (82.7%) and model complexity.

In COVSCA, each network becomes a point in an *R* dimensional space and thus the method provides a methodology to represent and visualize a large number of networks in a way akin to standard principal component analysis: points (i.e., metabolite-lipid association networks) close in the R-dimensional space share similar characteristics, i.e., similar patterns of correlation among lipids and metabolites. The relative importance of each metabolite/lipid in shaping the observed network differences is given by the loadings that can also be interpreted in a PCA fashion.

Clustering

T-distributed stochastic neighbor embedding (t-SNE) [59] was applied on the 3-dimensional COVSCA scores to visualize detectable similarities and clusters among the networks.

Software

Calculations were performed using MATLAB (version 2018b R 9.5.0.9) an R (version 3.3.2). The R code for the PCLRC algorithm and the code to perform differential connectivity analysis are available at the link: www.semantics.systemsbiology.nl under the SOFTWARE tab.

Results and discussion

Sex-specific differences of metabolite-lipid association networks

Metabolites and lipidic components take part in many metabolic processes; association networks, that quantify and visualize the interrelationships between molecular features, are representations of the complex web of biochemical reactions and pathways underlying the functioning of an organism: changes in the network structure can be considered to mirror alterations or re-modulation of the underlying network of metabolic reactions [28]. The sex-specific serum metabolite-lipid component association networks were built using separately samples from nonagenarian women (n_1=259) and men (n_2=96), to avoid confounding due to sex. Previous results obtained on this study cohort have observed sex-associated differences between men and women

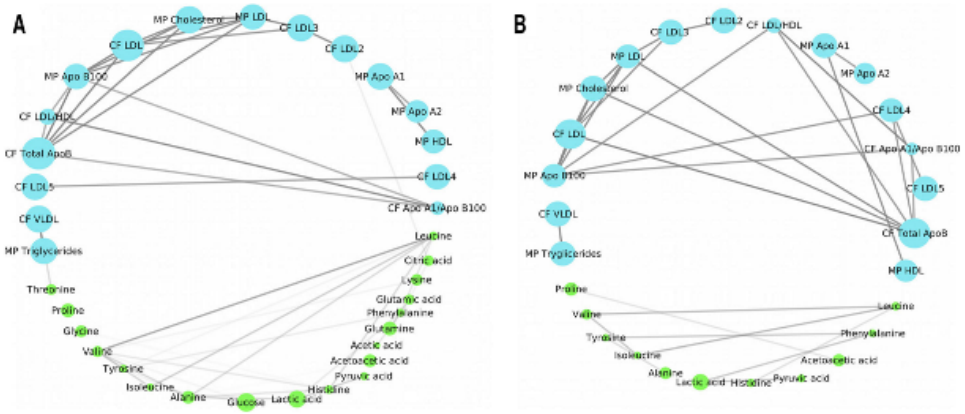


Fig. 3 Metabolite-lipid association networks for women (A) and men (B). Nodes are colored according to compounds' classification groups, light blue for lipid main parameters (MP) and calculated figures (CF), and light green for metabolites. Edges represent correlation with $|r| \geq 0.6$ and their width depends on the likelihood of the connections (see Eq. (3)). For sake of simplicity only metabolites, lipid main parameters, and calculated figures are shown

characteristics [60, 61]. However, although the interplay between sex differences and age-related differences has not been explored fully, accumulating evidence of sex dimorphism in the disease susceptibility [62, 63] aging and longevity phenotypes [64, 65] suggests the necessity of analyzing separately men and women data.

The women- and men-specific networks have markedly different topology. The women-specific network (Fig. 3A) is more densely connected, for what concern metabolite-metabolite associations, with respect to the men-specific one (Fig. 3B). In both networks, lipidic components (lipoproteins, lipid fractions, and sub-fractions) have a strongly inter-connected structure although different lipid species are involved. We quantified and assessed the statistical differences of the metabolite-lipid association networks specific to men and women using node connectivity which quantifies the number and the strength of metabolite-lipid associations, thus representing the importance of given metabolite/lipid in the network.

Differential network analysis results are given in Fig. 4. Seven out of 20 metabolites and 67 out of 114 lipoproteins and lipid fractions and sub-fractions have different connectivity patterns between men and women networks (adjusted P-value ≤ 0.05). Among metabolites, only alanine, isoleucine, leucine, lysine, citric acid, acetoacetic acid, acetic acid, showed altered connectivity, indicating re-modulation of amino acids and ketone bodies' metabolism, a result consistent with other studies focusing on the network-based analysis of the sex-specific difference in metabolite profiles in men and women [66–68]. In elderly women, re-modulation of amino acid metabolism results in a decreased level of branched amino acids (BCAAs) with respect to men [69, 70], and this phenomenon is associated with larger muscle mass loss, depending not only on a reduction in physical activity but also on a reduction of hormone activity, on an inadequate diet, and on the presence chronic diseases [69, 70]. Consistently with these observations, we found disruption of the association between alanine and leucine (present in women but not in the men network): leucine stimulates muscle protein synthesis [70–72] but is an important nitrogen donor for alanine biosynthesis [73].

Disruption of acetoacetic acid and acetic acid connectivity patterns suggests a re-modulation of ketone bodies' metabolism: acetic acid levels associate with prolonged fasting and diabetic ketosis, potentially frequent in elderly people with metabolic diseases [74], and, in women, with post-menopausal downregulation of gonadotropin which induces glycolytic dysregulation, resulting in a shift from physiological aerobic metabolism to a ketogenic phenotype [75, 76].

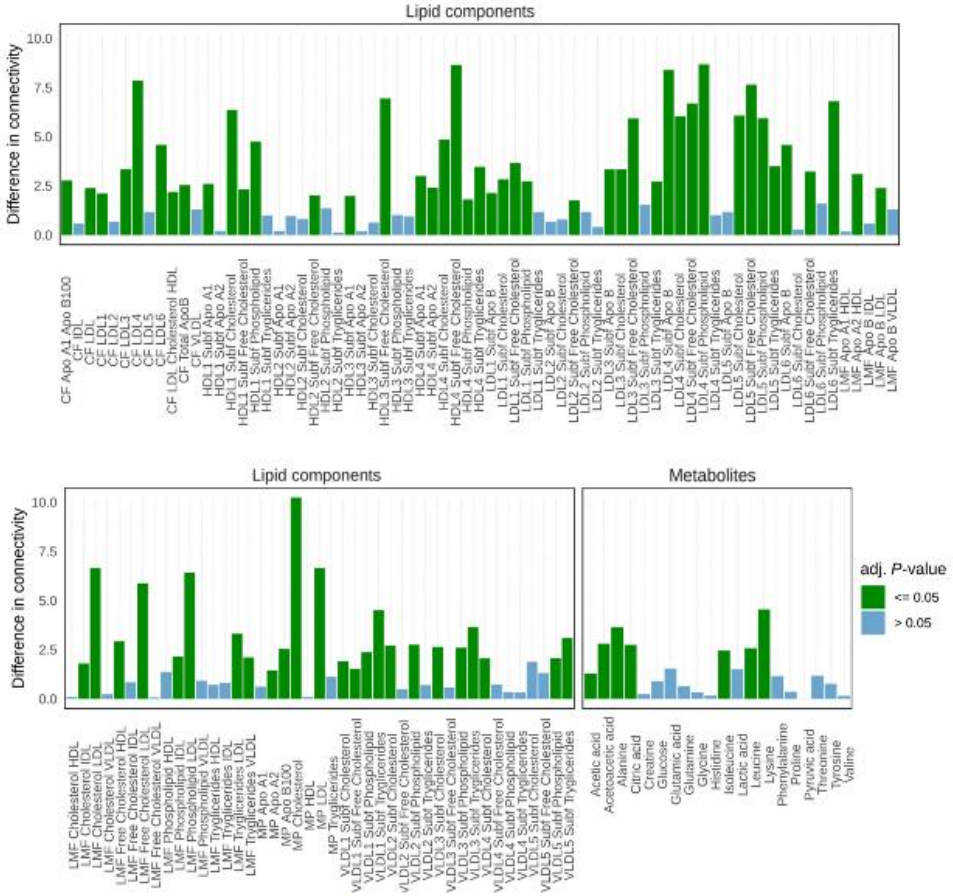


Fig. 4 Differential connectivity (Eq. (6); see Fig. 2 for an overview) from the differential network analysis of sex-related networks (men vs women) given in Fig. 3. For each metabolite and lipid component, adjusted *P*-values (Benjamini-Hochberg) are given.

Similarities and dissimilarities in different sex-related clinical variables-specific networks

We explored in a comprehensive way the relationships between metabolite-lipid association networks and the different 101 clinical covariates describing, for each subject, either the presence of a specific pathophysiological condition, familiarity diseases, pharmacological treatments, risk factors or the levels of 51 bio-humoral parameters (such as complete blood count, mean cell volume, hemoglobin, thyroid hormones. A complete list is reported in Table 1). For each of the 20 metabolites and 114 lipid fractions and sub-fractions, we first extracted the part of the observed variation of the concentration associated with a given clinical covariate (see Eq. (1)), and then we built association networks using only this fraction of the concentration, obtaining 101 (women) + 101 (men) different metabolite-lipid association networks. The rationale is that metabolites/lipids sharing similar relationships with a given covariate tend

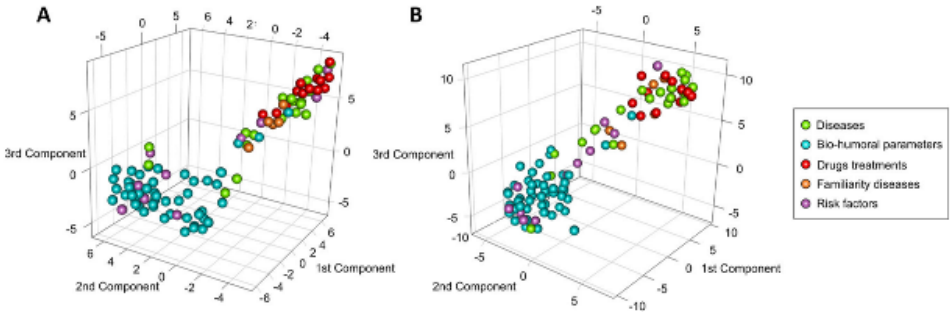


Fig. 5 Multivariate analysis of the metabolite-lipid association networks associated with the 101 clinical covariates (see Table 1) for women (A) and (B). The 101 + 101 networks are analyzed using Covariance Simultaneous Component Analysis (see the "Multivariate analysis of association networks" section). Each sphere corresponds to a network and is coloured according to the clinical variable-specific set: green colour corresponding to diseases, blue colour to bio-humoral parameters, red colour to drugs treatments, light brown colour to familiarity diseases and light violet colour to risk factors. Clustering is performed on the COVSCA score using t-SNE. Metabolite and lipid importance are given in Fig. 6

to be close to each other in the network, and this can provide a clearer representation of the underlying biological phenomena.

Since is not practically possible to compare all 202 networks individually, we used a multivariate component method (COVSCA, see Methods) to model and visualize the (dis)similarity and the relationships among the networks in combination with clustering; results are shown for women in Fig. 5A and for men in Fig. 5B, where each dot represents a metabolite-lipid association network specific to a given covariate.

For both women and men, networks specific to bio-humoral parameters separate from the networks specific to disease-related networks, indicating that the patterns of association among metabolites and lipid fractions and sub-fractions with these covariates are markedly different from those associated with comorbidity. Given a (clinical/biochemical) covariate, the statistical procedure employed can be seen as a correction procedure for confounders: in this light, the bio-humoral associated networks can be seen as representing metabolite/lipid relationship in a healthy condition. Networks associated with diseases tend to cluster with the networks related to ongoing pharmacological treatment of the same diseases suggesting the existence of oh shared information among these two groups of networks. Networks associated with risk factors (like age, smoking habits, and BMI) are scattered, indicating great heterogeneity and possibly reflecting that many risk scores are composite indexes summarizing both clinical and molecular features.

The relative importance of metabolites and lipidic components to explain the network clustering shown in Fig. 5A and B are given in Fig. 6A, B, and C for women and in Fig. 6D, E, F for men and can be interpreted as in standard principal component analysis (PCA). For both men and women, the first two components of the COVSCA model, explaining the variability within bio-humoral associated networks, are dominated LDL and VLDL lipid sub-fractions, while the third component, explaining the variability within disease-associated networks, is dominated by LDL and HDL. Our results show that different patterns of association between LDL and HDL fractions or interaction thereof are associated with comorbidity in this study cohort: LDL and HDL are not only associated with cardiovascular disease [77, 78], type II diabetes [79, 80], peripheral vascular disease [81, 82] and hypertension [83, 84] but also with dementia [85, 86] and cancer [87, 88]: the strong lipidic signature, suggest a potential different manifestation and response to health diseases in the elderly population [89, 90].

Overall, metabolites do not seem to play a significant role in shaping the observed difference among networks: only acetic acid, alanine, glutamine, and, less strongly, pyruvic acid have a relevant contribution to the model. Glucogenic amino acids (glutamine and alanine) have been associated with the regulation of aging and aging-related diseases [69]: in particular, altered levels of glutamine have been associated with higher intima-media thickness of carotid artery and, consequently, with coronary artery

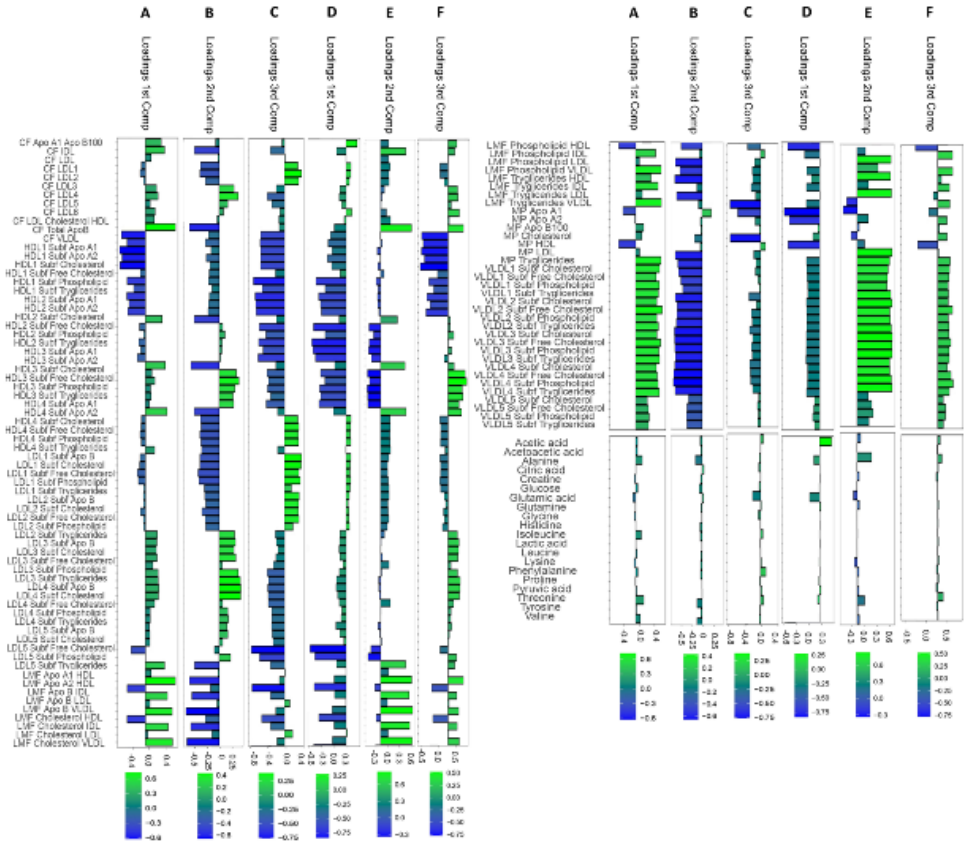


Fig. 6 Multivariate analysis of the metabolite-lipid association networks associated with the 101 clinical covariates (see Table 1) for women (A) and (B). The loadings give the importance of the metabolite and lipids to explain the patterns of network (dis)similarity observed in Fig. 6. Panel A–C: loadings for the analysis of women networks (Fig. 5A); Panel D–F: loadings for the analysis of men networks (Fig. 5B)

atherosclerosis, causing cardio-vascular syndromes [91, 92], and could be responsible for the increase of the activity of the osteoclasts resulting in a reduction of the bone mineral density [93]. Alanine and acetic acids are correlated with protein-energy malnutrition in aged people [94], metabolic syndromes [95], and, in post-menopausal women, are correlated with *a cellular ketogenic phenotypic change* [75, 76].

Differential clinical networks analysis in nonagenarian women and men

Starting from the observation that the metabolite/lipids association networks associated with the same type of covariate (bio-humoral parameters, diseases, drugs treatments, risk factors, and familiarity diseases) tend to share similar but not identical correlation patterns (see Fig. 7), we performed a pairwise comparison among the networks related to similar covariates. For each comparison, we recorded the significantly differentially connected metabolites (*P*-value adjusted ≤ 0.05) and lipoproteins/lipid fractions and sub-fractions (see Eq. (8)) for each comparison. We retained for further investigation only those molecular features that were found to be significant in more than 70% of the comparisons. Results are shown in Fig. 7. For both men and women bio-humoral

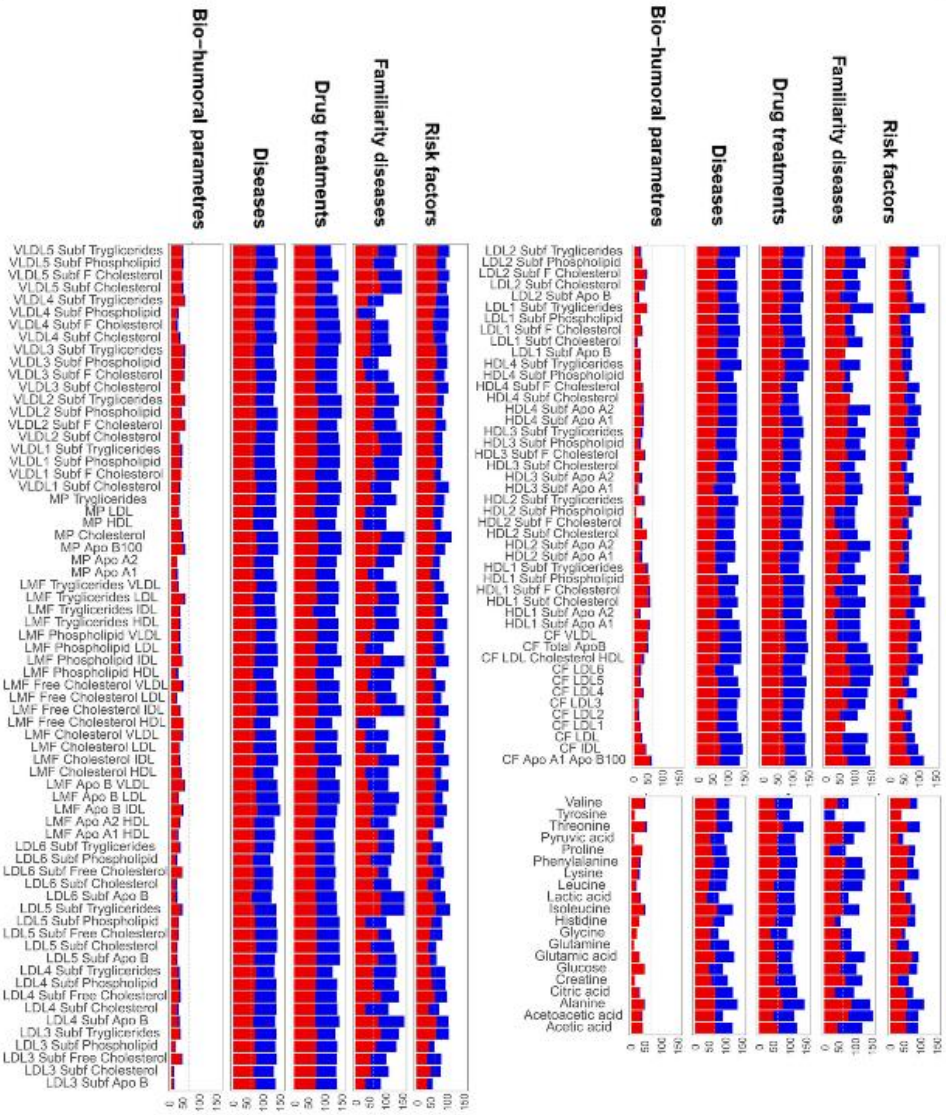


Fig. 7 Results of pairwise comparison of the 101 + 101 networks associated with clinical the covariates. The percentage of time that a metabolite or lipid is found to be significantly differentially connected (adjusted P-value < 0.05) between any two networks belonging to the clinical covariate of the same type is shown. The (overlapping) red bars correspond to the women-related clinical variable-specific networks and blue bars correspond to the men-related clinical variable-specific networks

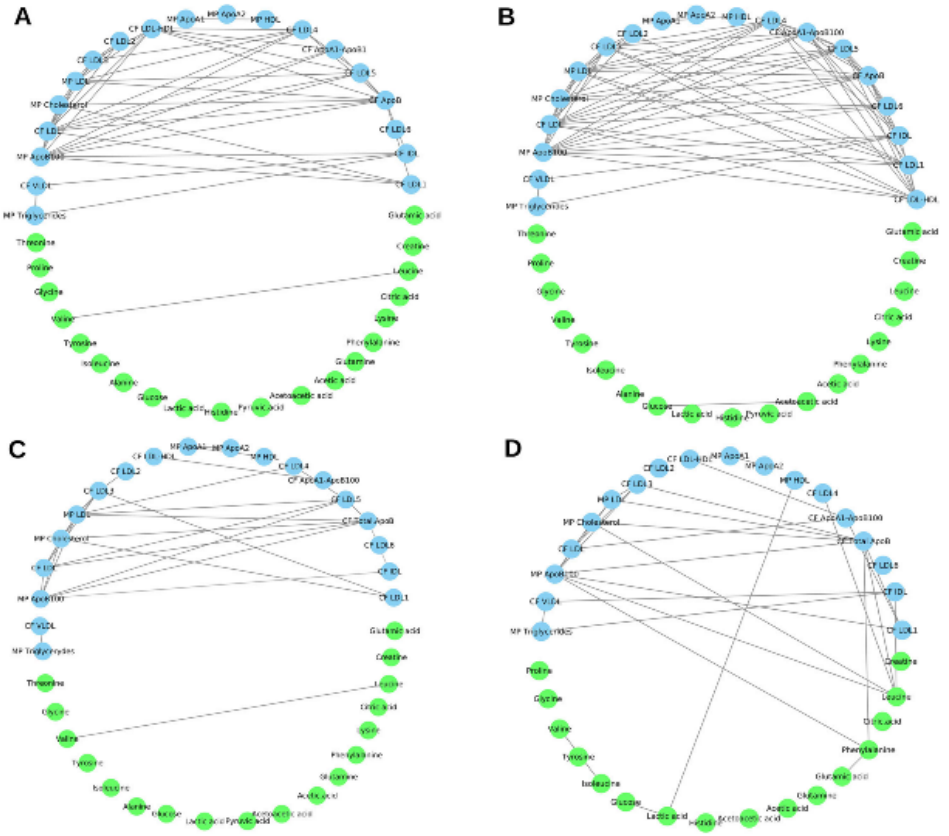


Fig. 8 Metabolite-lipid networks associated with peripheral vascular disease (A: women, B: men) and Diabetes (C: women, D: men). Nodes are colored according to compounds' classification groups, light blue color for lipid main parameters (MP) and calculated figures (CF), and light green color for metabolites. Edges represent correlation with $|r| \geq 0.6$ and their width depends on the likelihood of the connections. For sake of simplicity only metabolites, lipid main parameters (MP), and calculated figures (CF) are shown

parameter networks, we observe that very few metabolites and lipid features show different connectivity among different networks, indicating the similarity of these networks.

The comparison of disease-associated networks shows that for both men and women the full spectrum of (measured) lipids is associated with the differences among networks, suggesting their central role in pathophysiological mechanisms and in their resolutions; lipids are not only engaged in inter- and intra-cellular signaling regulation pathways but are also able to orchestrate inflammation processes and to restore the homeostasis [96], which may explain also the strong lipid signature observed also in drug-treatment associated networks. In women risk factor-networks lipid components, in particular, VLDL, plays a major role: in post-menopausal women, VLDL is negatively associated with estrogens [97–99], and this is related to increased risk of cardiovascular diseases [100, 101], myocardial infarction [102], and hyperlipidemia [103]: its association with diverse risk factors is worthy of mention since VLDL is actually being reconsidered as a potential biomarker [78, 104, 105].

The role of metabolites in explaining network differences is more nuanced, and this probably reflects the large biochemical diversity of these molecules [106]. In diseases-specific networks, key differentiating metabolites are ketone bodies, BCAAs, threonine, and tyrosine (in women networks only), and alanine: this can reflect the association with particular diseases, such as type 1 and 2 diabetes mellitus [91, 107–110], sarcopenia [69, 71, 72, 111] and cognitive impairment [112, 113]. Glutamine, glucose, proline, and BCAAs, increasing their metabolic activity, could be biomarkers to predict the emergence of neurodegenerative diseases [114], type 2 diabetes mellitus, and obesity conditions [91, 108, 109], and sarcopenia [69, 71, 72].

As an example, we show the networks associated with peripheral vascular disease and diabetes in both women and men (Fig. 8) which are two of the most common comorbidities among the subjects in this study (see Table 1).

When comparing the peripheral vascular disease association networks of women (Fig. 8A) and men (Fig. 8B), differential connectivity of leucine with lipidic components and among lipids component: while the latter has a strong association with peripheral vascular disease [82, 115], the remodulation of association between leucine and lipids suggests the existence an interplay between amino acids and lipids.

The difference in lipid correlation observed in diabetes-specific women (Fig. 8C) and men networks (Fig. 8D) further points to sex-specific differences in lipid metabolism [116, 117]. In particular, in women we observe a disruption of the correlation between acetic acids and glucose; it is known that acetic acid can lower glucose level, and can improve insulin resistance and metabolic abnormalities in the atherogenic prediabetic state [118]: the mechanism are not yet fully understood [119] and, as shown here, may be differentially regulated in men and women.

Conclusions

In this study, we presented a differential network analysis approach to highlight sex-related metabolic differences in cohort of nonagenarian subjects. Comparing the networks of nonagenarian women and men, we observed that lipids, branched chains amino acids, alanine, and ketone bodies show significant differences in connectivity in the two groups. In particular, we observed that lipids not only play a central role in the structural robustness of the network but also are directly associated with the intrinsic dynamic metabolic sex-related changes. The same approach was also applied to identify, in the disease-associated networks built in nonagenarian women and men, significantly differentially connected metabolites and lipoproteins/lipid fractions and sub-fractions. Our results show the importance of the lipid components in diseases, drug treatments, and familiarity disease, indicating their ability to participate in many pathophysiological mechanisms in nonagenarians; in the women's disease-specific networks, the rewiring of metabolic activity involves ketone bodies, branched chains amino acids, threonine, and tyrosine. In conclusion, this study provides information about the structure of sex-related networks in nonagenarians, contributing to elucidate the impact of gender on human physiology and pathophysiology in the elderly population and showcases how network analysis may provide a valuable tool in gender medicine.

Acknowledgements The authors acknowledge the support and the use of resources of Instruct-ERIC, a Landmark ESFRI project, and specifically the CERM/CIRMMP Italy Centre.

Author contribution A.M.G., R.M., B.G., R.M-L., C.M., and S.P. recruited patients, collected samples, and managed biological material and clinical data collection. G.M. obtained the NMR data. E.S., L.T., and F.D.C. performed data analyses. C.L., E.S., L.T., and F.D.C. interpreted the data and results, prepared the manuscript, and were responsible for its final content.

Data availability On requestSupplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11357-021-00404-3.

Code availability Code available upon request to E.S.Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11357-021-00404-3.

Declarations

Ethics approval The study protocol was approved by the Ethical Committee of the Don Gnocchi Foundation.

Consent to participate Informed written consent was obtained from all participants or their legal representatives.

Consent for publication Patients signed informed consent regarding publishing their data.

Conflict of interest The authors declare no competing interests.

References

1. Shrestha LB. Population Aging In Developing Countries. Health Aff (Millwood). 2000;19:204–12.
2. Suzman R, Beard JR, Boerma T, Chatterji S. Health in an ageing world—what do we know? Lancet. Elsevier. 2015;385:484–6.
3. I centenari in Italia. :5.n.d..
4. Campisi G, Chiappelli M, De Martinis M, Franco V, Ginaldi L, Guiglia R, et al. Pathophysiology of age-related diseases. Immun Ageing A. 2009;6:12.
5. Pinquart M, Sörensen S. Influences of socioeconomic status, social network, and competence on subjective well-being in later life: a meta-analysis. Psychol Aging. 2000;15:187–224.
6. McLean AJ, Couteur DGL. Aging Biology and Geriatric Clinical Pharmacology. Pharmacol Rev. American Society for Pharmacology and Experimental Therapeutics. 2004;56:163–84.
7. Jazwinski SM, Kim S. Metabolic and Genetic Markers of Biological Age. Front Genet. Front. 2017;8.
8. Metz DH. Mobility of older people and their quality of life. Transp Policy. 2000;7:149–52.
9. Martin JE, Sheaff MT. The pathology of ageing: concepts and mechanisms. J Pathol. 2007;211:111–3.
10. Valenzuela JF, Monterola C, VJC T, Ng TP, Larbi A. Health and disease phenotyping in old age using a cluster network analysis. Sci Rep. Nature Publishing Group. 2017;7:15608.
11. Lorusso JS, Sviderskiy OA, Labunskyy VM. Emerging Omics Approaches in Aging Research. Antioxid Redox Signal. 2018;29:985–1002.
12. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. Cell. 2013;153:1194–217.
13. Kirkwood TBL. Systems biology of ageing and longevity. Philos Trans R Soc B Biol Sci Royal Society. 2011;366:64–70.
14. Kirkwood TB, Austad SN. Why do we age? Nature. 2000;408:233–8.
15. Golubev A, Hanson AD, Gladyshev VN. Non-enzymatic molecular damage as a prototypic driver of aging. J Biol Chem. 2017;292:6029–38.
16. Yin D, Chen K. The essential mechanisms of aging: Irreparable damage accumulation of biochemical side-reactions. Exp Gerontol. 2005;40:455–65.
17. Gladyshev VN. The free radical theory of aging is dead. Long live the damage theory! Antioxid Redox Signal. 2014;20:727–31.
18. Valdes AM, Glass D, Spector TD. Omics technologies and the study of human ageing. Nat Rev Genet. Nature Publishing Group. 2013;14:601–7.
19. Zierer J, Menni C, Kastenmüller G, Spector TD. Integration of 'omics' data in aging research: from biomarkers to systems biology. Aging Cell. 2015;14:933–44.
20. Takis PG, Ghini V, Tenori L, Turano P, Luchinat C. Uniqueness of the NMR approach to metabolomics. TrAC Trends Anal Chem. 2019;120:115300.
21. Vignoli A, Ghini V, Meoni G, Licari C, Takis PG, Tenori L, et al. High-Throughput Metabolomics by 1D NMR. Angew Chem Int Ed. 2019;58:968–94.
22. Ghini V, Saccenti E, Tenori L, Assfalg M, Luchinat C. Allostasis and Resilience of the Human Individual Metabolic Phenotype. J Proteome Res. American Chemical Society. 2015;14:2951–62.
23. Vignoli A, Tenori L, Giusti B, Takis PG, Valente S, Carrabba N, et al. 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;17:3.
24. Chen T, Cao Y, Zhang Y, Liu J, Bao Y, Wang C, et al. Random Forest in Clinical Metabolomics for Phenotypic Discrimination and Biomarker Selection. Evid Based Complement Alternat Med. Hindawi. 2013;2013:e298183.
25. Kim ER, Kwon HN, Nam H, Kim JJ, Park S, Kim Y-H. Urine-NMR metabolomics for screening of advanced colorectal adenoma and early stage colorectal cancer. Sci Rep. Nature Publishing Group. 2019;9:4786.
26. Meoni G, Lorini S, Monti M, Madia F, Corti G, Luchinat C, et al. The metabolic fingerprints of HCV and HBV infections studied by Nuclear Magnetic Resonance Spectroscopy. Sci Rep. Nature Publishing Group. 2019;9:1–13.
27. Hart CD, Vignoli A, Tenori L, Uy GL, Van To T, Adebamowo C, et al. Serum Metabolomic Profiles Identify ER-Positive Early Breast Cancer Patients at Increased Risk of Disease Recurrence in a Multicenter Population. Clin Cancer Res Off J Am Assoc Cancer Res. 2017;23:1422–31.
28. Rosato A, Tenori L, Cascante M, De Atauri Carulla PR, dos Santos VAP M, Saccenti E. From correlation to causation: analysis of metabolomics data using systems biology approaches. Metabolomics. 2018;14:37.
29. Souza LPD, Alseekh S, Brotman Y, Fernie AR. Network-based strategies in metabolomics data analysis and interpretation: from molecular networking to biological interpretation. Expert Rev Proteomics. Taylor & Francis. 2020;17:243–55.
30. Camacho D, de la Fuente A, Mendes P. The origin of correlations in metabolomics data. Metabolomics. 2005;1:53–63.
31. Molino Lova R, Sofi F, Pasquini G, Gori AM, Vannetti F, Abbate R, et al. The Mugello Study, a survey of nonagenarians living in Tuscany: Design, methods and participants' general characteristics. Eur J Intern Med. 2013;24:745–9.
32. Ostan R, Monti D, Gueresi P, Bussolotto M, Franceschi C, Baggio G. Gender, aging and longevity in humans: an update of an intriguing/neglected scenario paving the way to a gender-specific medicine. Clin Sci Lond Engl. 1979;2016(130):1711–25.
33. Almagro P, Ponce A, Komal S, Villaverde M de la A, Castrillo C, Grau G, et al. Multimorbidity gender patterns in hospitalized elderly patients. PLoS One. Public Library of Science. 2020;15:e0227252.

34. Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, et al. Development and validation of a geriatric depression screening scale: a preliminary report. J Psychiatr Res. 1982;17:37–49.
35. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 1975;12:189–98.
36. Treacy D, Hassett L. The Short Physical Performance Battery. Aust J Phys. 2017;64.
37. Beauchet O, Fantino B, Allali G, Muir S, Montero-Odasso M, Annweiler C. Timed Up and Go test and risk of falls in older adults: A systematic review. J Nutr Health Aging. 2011;15:933–8.
38. Watson PE, Watson ID, Batt RD. Obesity indices. Am J Clin Nutr. 1979;32:736–7.
39. Manson JE, Stampfer MJ, Hennekens CH, Willett WC. Body Weight and Longevity: A Reassessment. JAMA. American Medical Association. 1987;257:353–8.
40. Schneider HJ, Friedrich N, Klotsche J, Pieper L, Nauck M, John U, et al. The Predictive Value of Different Measures of Obesity for Incident Cardiovascular Events and Mortality. J Clin Endocrinol Metab. Oxford Academic. 2010;95:1777–85.
41. Newman AB, Siscovick DS, Manolio TA, Polak J, Fried LP, Borhani NO, et al. Ankle-arm index as a marker of atherosclerosis in the Cardiovascular Health Study. Cardiovascular Heart Study (CHS) Collaborative Research Group. Circulation. 1993;88:837–45.
42. Victor A, Criqui MH, Pierre A, Allison MA, Creager MA, Curt D, et al. Measurement and Interpretation of the Ankle-Brachial Index. Circulation. American Heart Association. 2012;126:2890–909.
43. Washburn RA, Smith KW, Jette AM, Janney CA. The physical activity scale for the elderly (PASE): Development and evaluation. J Clin Epidemiol. 1993;46:153–62.
44. Amaral CA, TLM A, GTR M, MTL V, Portela MC. Hand grip strength: Reference values for adults and elderly people of Rio Branco, Acre, Brazil. PLoS One. Public Library of Science. 2019;14:e0211452.
45. Féart C, Samieri C, Rondeau V, Amieva H, Portet F, Dartigues J-F, et al. Adherence to a Mediterranean Diet, Cognitive Decline, and Risk of Dementia. JAMA. American Medical Association. 2009;302:638–48.
46. Panagiotakos DB, Pitsavos C, Stefanadis C. Dietary patterns: A Mediterranean diet score and its relation to clinical and biological markers of cardiovascular disease risk. Nutr Metab Cardiovasc Dis. 2006;16:559–68.
47. 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.
48. 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. Anal Chem. 2018;90:11962–71.
49. Stekhoven DJ, Bühlmann P. MissForest—non-parametric missing value imputation for mixed-type data. Bioinformatics. Oxford Academic. 2012;28:112–8.
50. Bartzis G, Deelen J, Maia J, Ligterink W, Hilhorst HWM, Houwing-Duistermaat J-J, et al. Estimation of metabolite networks with regard to a specific covariable: applications to plant and human data. Metabolomics Off J Metabolomic Soc. 2017;13:129.
51. Pavlopoulos GA, Secrier M, Moschopoulos CN, Soldatos TG, Kossida S, Aerts J, et al. Using graph theory to analyze biological networks. BioData Min. 2011;4:10.
52. Suarez-Diez M, Saccenti E. Effects of Sample Size and Dimensionality on the Performance of Four Algorithms for Inference of Association Networks in Metabonomics. J Proteome Res. 2015;14:5119–30.
53. Jahagirdar S, Saccenti E. On the Use of Correlation and MI as a Measure of Metabolite—Metabolite Association for Network Differential Connectivity Analysis. Metabolites. Multidisciplinary Digital Publishing Institute. 2020;10:171.
54. Akhand MAH, Nandi RN, Amran SM, Murase K. Context likelihood of relatedness with maximal information coefficient for Gene Regulatory Network inference. 2015 18th International Conference on Computer and Information Technology (ICCIT), 2015, pp. 312–316. https://doi.org/10.1109/ICCITechn.2015.7488088.
55. Lawyer G. Understanding the influence of all nodes in a network. Sci Rep. Nature Publishing Group. 2015;5:8665.
56. Wuchty S, Ravasz E, Barabási A-L. The Architecture of Biological Networks. In: Deisboeck TS, Kresh JY, curatori, editors. Complex Syst Sci Biomed. Boston, MA: Springer US; 2006. p. 165–81. https://doi.org/10.1007/978-0-387-33532-2_5.
57. Ideker T, Krogan NJ. Differential network biology. Mol Syst Biol. John Wiley & Sons, Ltd. 2012;8:565.
58. Smilde AK, Timmerman ME, Saccenti E, Jansen JJ, Hoefsloot HCJ. Covariances Simultaneous Component Analysis: a new method within a framework for modeling covariances. J Chemom. 2015;29:277–88.
59. van der Maaten L, Hinton G. Visualizing Data using t-SNE. J Mach Learn Res. 2008;9:2579–605.
60. Padua L, Pasqualetti P, Coraci D, Imbimbo I, Giordani A, Loreti C, et al. Gender effect on well-being of the oldest old: a survey of nonagenarians living in Tuscany: the Mugello study. Neurol Sci Off J Ital Neurol Soc Ital Soc Clin Neurophysiol. 2018;39:509–17.
61. Giovannini S, Macchi C, Liperoti R, Laudisio A, Coraci D, Loreti C, et al. Association of Body Fat With Health-Related Quality of Life and Depression in Nonagenarians: The Mugello Study. J Am Med Dir Assoc. 2019;20:564–8.
62. Sampathkumar NK, Bravo JI, Chen Y, Danthi PS, Donahue EK, Lai RW, et al. Widespread sex dimorphism in aging and age-related diseases. Hum Genet. 2020;139:333–56.
63. Ober C, Loisel DA, Gilad Y. Sex-Specific Genetic Architecture of Human Disease. Nat Rev Genet. 2008;9:911–22.
64. Gems D. Evolution of sexually dimorphic longevity in humans. Aging. 2014;6:84–91.
65. Regan JC, Partridge L. Gender and longevity: Why do men die earlier than women? Comparative and experimental evidence. Best Pract Res Clin Endocrinol Metab. 2013;27:467–79.

66. Vignoli A, Tenori L, Luchinat C, Saccenti E. Age and Sex Effects on Plasma Metabolite Association Networks in Healthy Subjects. J Proteome Res. 2018;17:97–107.
67. Li Z, Zhang Y, Hu T, Likhodii S, Sun G, Zhai G, et al. Differential metabolomics analysis allows characterization of diversity of metabolite networks between males and females. PLoS One. Public Library of Science. 2018;13: e0207775.
68. Krumsiek J, Mittelstrass K, Do KT, Stückler F, Ried J, Adamski J, et al. Gender-specific pathway differences in the human serum metabolome. Metabolomics Off J Metabolomic Soc. 2015;11:1815–33.
69. Canfield C-A, Bradshaw PC. Amino acids in the regulation of aging and aging-related diseases. Transl Med Aging. 2019;3:70–89.
70. Fukagawa NK. Protein and amino acid supplementation in older humans. Amino Acids. 2013;44:1493–509.
71. Borack MS, Volpi E. Efficacy and Safety of Leucine Supplementation in the Elderly. J Nutr. 2016;146:2625S–9S.
72. Hutson SM, Sweatt AJ, Lanoue KF. Branched-chain [corrected] amino acid metabolism: implications for establishing safe intakes. J Nutr. 2005;135:1557S–64S.
73. Haymond MW, Miles JM. Branched chain amino acids as a major source of alanine nitrogen in man. Diabetes. 1982;31:86–9.
74. Puchalska P, Crawford PA. Multi-dimensional roles of ketone bodies in fuel metabolism, signaling, and therapeutics. Cell Metab. 2017;25:262–84.
75. Rettberg JR, Yao J, Brinton RD. Estrogen: A master regulator of bioenergetic systems in the brain and body. Front Neuroendocrinol. 2014;35:8–30.
76. Brinton RD. Estrogen regulation of glucose metabolism and mitochondrial function: therapeutic implications for prevention of Alzheimer's disease. Adv Drug Deliv Rev. 2008;60:1504–11.
77. Gordon T, Kannel WB, Castelli WP, Dawber TR. Lipoproteins, Cardiovascular Disease, and Death: The Framingham Study. Arch Intern Med. 1981;141:1128–31.
78. Barter P, Gotto AM, LaRosa JC, Maroni J, Szarek M, Grundy SM, et al. HDL Cholesterol, Very Low Levels of LDL Cholesterol, and Cardiovascular Events. N Engl J Med. Massachusetts Medical Society. 2007;357:1301–10.
79. Krauss RM. Lipids and Lipoproteins in Patients With Type 2 Diabetes. Diabetes Care. American Diabetes Association. 2004;27:1496–504.
80. Vergès B. Lipid modification in type 2 diabetes: the role of LDL and HDL. Fundam Clin Pharmacol. 2009;23:681–5.
81. Bergmark C, Wu R, de Faire U, Lefvert AK, Swedenborg J. Patients With Early-Onset Peripheral Vascular Disease Have Increased Levels of Autoantibodies Against Oxidized LDL. Arterioscler Thromb Vasc Biol. American Heart Association. 1995;15:441–5.
82. Aday AW, Lawler PR, Cook NR, Ridker PM, Mora S, Pradhan AD. Lipoprotein Particle Profiles, Standard Lipids, and Peripheral Artery Disease Incidence. Circulation. American Heart Association. 2018;138: 2330–41.
83. Gil-Extremera B L. Disorders in Elderly Hypertensive Patients. Int J Hypertens. Hindawi. 2011;2012:e684515.
84. Onuh JO, Aliani M. Metabolomics profiling in hypertension and blood pressure regulation: a review. Clin Hypertens. 2020;26:23.
85. Reitz C, Tang M-X, Luchsinger J, Mayeux R. Relation of Plasma Lipids to Alzheimer Disease and Vascular Dementia. Arch Neurol. 2004;61:705–14.
86. Anstey KJ, Lipnicki DM, Low L-F. Cholesterol as a Risk Factor for Dementia and Cognitive Decline: A Systematic Review of Prospective Studies With Meta-Analysis. Am J Geriatr Psychiatry. 2008;16:343–54.
87. Borgquist S, Butt T, Almgren P, Shiffman D, Stocks T, Orho-Melander M, et al. Apolipoproteins, lipids and risk of cancer. Int J Cancer. 2016;138:2648–56.
88. Kökoğlu E, Karaarslan I, Mehmet Karaarslan H, Baloğlu H. Alterations of serum lipids and lipoproteins in breast cancer. Cancer Lett. 1994;82:175–8.
89. Johnson AA, Stolzing A. The role of lipid metabolism in aging, lifespan regulation, and age-related disease. Aging Cell. 2019;18:e13048.
90. Gonzalez-Covarrubias V, Beekman M, Uh H-W, Dane A, Troost J, Paliukhovich I, et al. Lipidomics of familial longevity. Aging Cell. 2013;12:426–34.
91. Auro K, Joensuu A, Fischer K, Kettunen J, Salo P, Mattsson H, et al. A metabolic view on menopause and ageing. Nat Commun. Nature Publishing Group. 2014;5: 1–11.
92. Würtz P, Raiko JR, Magnussen CG, Soininen P, Kangas AJ, Tynkkynen T, et al. High-throughput quantification of circulating metabolites improves prediction of subclinical atherosclerosis. Eur Heart J. 2012;33:2307–16.
93. You Y-S, Lin C-Y, Liang H-J, Lee S-H, Tsai K-S, Chiou J-M, et al. Association between the metabolome and low bone mineral density in Taiwanese women determined by (1)H NMR spectroscopy. J Bone Miner Res Off J Am Soc Bone Miner Res. 2014;29:212–22.
94. Polge A, Bancel E, Bellet H, Strubel D, Poirey S, Peray P, et al. Plasma amino acid concentrations in elderly patients with protein energy malnutrition. Age Ageing. 1997;26: 457–62.
95. Monnerie S, Comte B, Ziegler D, Morais JA, Pujos-Guillot E, Gaudreau P. Metabolomic and Lipidomic Signatures of Metabolic Syndrome and its Physiological Components in Adults: A Systematic Review. Sci Rep. Nature Publishing Group. 2020;10:1–13.
96. Chiurchiù V, Leuti A, Maccarrone M. Bioactive Lipids and Chronic Inflammation: Managing the Fire Within. Front Immunol. 2018;9.
97. Schaefer EJ, Foster DM, Zech LA, Lindgren FT, Brewer HB, Levy RI. The Effects of Estrogen Administration on Plasma Lipoprotein Metabolism in Premenopausal Females. J Clin Endocrinol Metab. Oxford Academic. 1983;57:262–7.
98. Palmisano BT, Zhu L, Stafford JM. Estrogens in the Regulation of Liver Lipid Metabolism. Adv Exp Med Biol. 2017;1043:227–56.
99. Mauvais-Jarvis F, Clegg DJ, Hevener AL. The role of estrogens in control of energy balance and glucose homeostasis. Endocr Rev. 2013;34:309–38.
100. Ference BA, Ginsberg HN, Graham I, Ray KK, Packard CJ, Bruckert E, et al. Low-density lipoproteins cause atherosclerotic cardiovascular disease. 1. Evidence from

genetic, epidemiologic, and clinical studies. A consensus statement from the European Atherosclerosis Society Consensus Panel. Eur Heart J. 2017;38:2459–72.
101. Ivanova EA, Myasoedova VA, Melnichenko AA, Grechko AV, Orekhov AN. Small Dense Low-Density Lipoprotein as Biomarker for Atherosclerotic Diseases. Oxidative Med Cell Longev. 2017;2017:1273042.
102. Khan HA, Ekhzaimy A, Khan I, Sakharkar MK. Potential of lipoproteins as biomarkers in acute myocardial infarction. Anatol J Cardiol. 2017;18:68–74.
103. Streja D, Streja E. Management of Dyslipidemia in the Elderly. In: Feingold KR, Anawalt B, Boyce A, Chrousos G, Dungan K, Grossman A, et al., curatori. Endotext [Internet]. South Dartmouth (MA): MDText.com, Inc.; 2000 http://www.ncbi.nlm.nih.gov/books/NBK279133/
104. von Zychlinski A, Kleffmann T. Dissecting the proteome of lipoproteins: New biomarkers for cardiovascular diseases? Transl Proteomics. 2015;7:30–9.
105. Ren J, Grundy SM, Liu J, Wang W, Wang M, Sun J, et al. Long-term coronary heart disease risk associated with very-low-density lipoprotein cholesterol in Chinese: the results of a 15-Year Chinese Multi-Provincial Cohort Study (CMCS). Atherosclerosis. 2010;211:327–32.
106. Psychogios N, Hau DD, Peng J, Guo AC, Mandal R, Bouatra S, et al. The Human Serum Metabolome. PLoS One. Public Library of Science. 2011;6:e16957.
107. Nakayama H, Tokubuchi I, Wada N, Tsuruta M, Ohki T, Oshige T, et al. Age-related changes in the diurnal variation of ketogenesis in patients with type 2 diabetes and relevance to hypoglycemic medications. Endocr J. 2015;62:235–41.
108. Park S, Sadanala KC, Kim E-K. A Metabolomic Approach to Understanding the Metabolic Link between Obesity and Diabetes. Mol Cell. 2015;38:587–96.
109. Lo C-J, Tang H-Y, Huang C-Y, Lin C-M, Ho H-Y, Shiao M-S, et al. Metabolic Signature Differentiated Diabetes Mellitus from Lipid Disorder in Elderly Taiwanese. J Clin Med. 2018;8.
110. Garcia E, Shalaurova I, Matyus SP, Oskardmay DN, Otvos JD, Dullaart RPF, et al. Ketone Bodies Are Mildly Elevated in Subjects with Type 2 Diabetes Mellitus and Are Inversely Associated with Insulin Resistance as Measured by the Lipoprotein Insulin Resistance Index. J Clin Med. 2020;9.
111. Agostini D, Zeppa Donati S, Lucertini F, Annibalini G, Gervasi M, Ferri Marini C, Piccoli G, Stocchi V, Barbieri E, Sestili P. Muscle and Bone Health in Postmenopausal Women: Role of Protein and Vitamin D Supplementation Combined with Exercise Training. Nutrients. 2018;10(8):1103.
112. van de Rest O, Bloemendaal M, de Heus R, Aarts E. Dose-dependent effects of oral tyrosine administration on plasma tyrosine levels and cognition in aging. Nutrients. 2017;9(12):1279.
113. Ravaglia G, Forti P, Maioli F, Bianchi G, Martelli M, Talerico T, et al. Plasma amino acid concentrations in patients with amnestic mild cognitive impairment or Alzheimer disease. Am J Clin Nutr. Oxford Academic. 2004;80:483–8.
114. Socha E, Koba M, Kośliński P. Amino acid profiling as a method of discovering biomarkers for diagnosis of neurodegenerative diseases. Amino Acids. 2019;51:367–71.
115. Kou M, Ding N, Ballew SH, Salameh MJ, Martin SS, Selvin E, et al. Conventional and Novel Lipid Measures and Risk of Peripheral Artery Disease. Arterioscler Thromb Vasc Biol. American Heart Association. 2021;41:1229–38.
116. Palmisano BT, Zhu L, Eckel RH, Stafford JM. Sex differences in lipid and lipoprotein metabolism. Mol Metab. 2018;15:45–55.
117. Varlamov O, Bethea CL, Roberts CT. Sex-Specific Differences in Lipid and Glucose Metabolism. Front Endocrinol. 2015;5.
118. Mitrou P, Petsiou E, Papakonstantinou E, Maratou E, Lambadiari V, Dimitriadis P, et al. The role of acetic acid on glucose uptake and blood flow rates in the skeletal muscle in humans with impaired glucose tolerance. Eur J Clin Nutr. 2015;69:734–9.
119. Santos HO, de Moraes WMAM, da Silva GAR, Prestes J, Schoenfeld BJ. Vinegar (acetic acid) intake on glucose metabolism: A narrative review. Clin Nutr. ESPEN. 2019;32:1–7.

Supplementary Materials

Table S1: Complete list of lipid fractions and sub-fractions considered

Lipid fractions and sub-fractions
Calcutated Figures (CF) LDL/HDL
Calcutated Figures (CF) Apo A1/Apo B100
Calcutated Figures (CF) Total ApoB
Calcutated Figures (CF) VLDL
Calcutated Figures (CF) IDL
Calcutated Figures (CF) LDL
Calcutated Figures (CF) LDL1
Calcutated Figures (CF) LDL2
Calcutated Figures (CF) LDL3
Calcutated Figures (CF) LDL4
Calcutated Figures (CF) LDL5
Calcutated Figures (CF) LDL6
Lipoprotein Main Fractions (LMF) Triglycerides VLDL
Lipoprotein Main Fractions (LMF) Triglycerides IDL
Lipoprotein Main Fractions (LMF) Triglycerides LDL
Lipoprotein Main Fractions (LMF) Triglycerides HDL
Lipoprotein Main Fractions (LMF) Cholesterol VLDL
Lipoprotein Main Fractions (LMF) Cholesterol IDL
Lipoprotein Main Fractions (LMF) Cholesterol LDL
Lipoprotein Main Fractions (LMF) Cholesterol HDL
Lipoprotein Main Fractions (LMF) Free Cholesterol VLDL
Lipoprotein Main Fractions (LMF) Free Cholesterol IDL
Lipoprotein Main Fractions (LMF) Free Cholesterol LDL
Lipoprotein Main Fractions (LMF) Free Cholesterol HDL
Lipoprotein Main Fractions (LMF) Phospholipids VLDL
Lipoprotein Main Fractions (LMF) Phospholipids IDL
Lipoprotein Main Fractions (LMF) Phospholipids LDL
Lipoprotein Main Fractions (LMF) Phospholipids HDL
Lipoprotein Main Fractions (LMF) Apo A1 HDL
Lipoprotein Main Fractions (LMF) Apo A2 HDL

Lipoprotein Main Fractions (LMF) Apo B VLDL
Lipoprotein Main Fractions (LMF) Apo B IDL
Lipoprotein Main Fractions (LMF) Apo B LDL
VLDL1 Subfractions Triglycerides
VLDL2 Subfractions Triglycerides
VLDL3 Subfractions Triglycerides
VLDL4 Subfractions Triglycerides
VLDL5 Subfractions Triglycerides
VLDL1 Subfractions Cholesterol
VLDL2 Subfractions Cholesterol
VLDL3 Subfractions Cholesterol
VLDL4 Subfractions Cholesterol
VLDL5 Subfractions Cholesterol
VLDL1 Subfractions Free Cholesterol
VLDL2 Subfractions Free Cholesterol
VLDL3 Subfractions Free Cholesterol
VLDL4 Subfractions Free Cholesterol
VLDL5 Subfractions Free Cholesterol
VLDL1 Subfractions Phospholipids
VLDL2 Subfractions Phospholipids
VLDL3 Subfractions Phospholipids
VLDL4 Subfractions Phospholipids
VLDL5 Subfractions Phospholipids
LDL1 Subfractions Triglycerides
LDL2 Subfractions Triglycerides
LDL3 Subfractions Triglycerides
LDL4 Subfractions Triglycerides
LDL5 Subfractions Triglycerides
LDL6 Subfractions Triglycerides
LDL1 Subfractions Cholesterol
LDL2 Subfractions Cholesterol

LDL3 Subfractions Cholesterol
LDL4 Subfractions Cholesterol
LDL5 Subfractions Cholesterol
LDL6 Subfractions Cholesterol
LDL1 Subfractions Free Cholesterol
LDL2 Subfractions Free Cholesterol
LDL3 Subfractions Free Cholesterol
LDL4 Subfractions Free Cholesterol
LDL5 Subfractions Free Cholesterol
LDL6 Subfractions Free Cholesterol
LDL1 Subfractions Phospholipids
LDL2 Subfractions Phospholipids
LDL3 Subfractions Phospholipids
LDL4 Subfractions Phospholipids
LDL5 Subfractions Phospholipids
LDL6 Subfractions Phospholipids
LDL1 Subfractions Apo B
LDL2 Subfractions Apo B
LDL3 Subfractions Apo B
LDL4 Subfractions Apo B
LDL5 Subfractions Apo B
LDL6 Subfractions Apo B
HDL1 Subfractions Triglycerides
HDL2 Subfractions Triglycerides
HDL3 Subfractions Triglycerides
HDL4 Subfractions Triglycerides
HDL1 Subfractions Cholesterol
HDL2 Subfractions Cholesterol
HDL3 Subfractions Cholesterol
HDL4 Subfractions Cholesterol
HDL1 Subfractions Free Cholesterol
HDL2 Subfractions Free Cholesterol
HDL3 Subfractions Free Cholesterol

HDL4 Subfractions Free Cholesterol
HDL1 Subfractions Phospholipids
HDL2 Subfractions Phospholipids
HDL3 Subfractions Phospholipids
HDL4 Subfractions Phospholipids
HDL1 Subfractions Apo A1
HDL2 Subfractions Apo A1
HDL3 Subfractions Apo A1
HDL4 Subfractions Apo A1
HDL1 Subfractions Apo A2
HDL2 Subfractions Apo A2
HDL3 Subfractions Apo A2
HDL4 Subfractions Apo A2

4.1.2 ^{1}H-NMR-based metabolomics reveals sex-effect in nonagenarian metabolic profiles and is a useful tool for the prediction of cognitive impairment and elderly depression

Francesca Di Cesare[1], Gaia Meoni[3], Leonardo Tenori[1,2], Anna Maria Gori[4,5], Rossella Marcucci[4,5], Betti Giusti[4,5], Raffaele Molino-Lova[6], Claudio Macchi[4,6], Silvia Pancani[6], Claudio Luchinat[1,2,7]

[1] Magnetic Resonance Center (CERM), University of Florence, Sesto Fiorentino, Italy

[2] Department of Chemistry "Ugo Schiff", University of Florence, Sesto Fiorentino, Italy

[3] Giotto Biotech srl, Florence, Italy.

[4] Department of Experimental and Clinical Medicine, University of Florence, Florence, Italy.

[5] Atherothrombotic Unit, Careggi University Hospital, Florence, Italy.

[6] IRCCS Fondazione Don Carlo Gnocchi, Florence, Italy.

[7] Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine (CIRMMP), Sesto Fiorentino, Italy.

.

In preparation

Candidate's contribuitions: NMR-data pre-processing, statistical analysis, interpretation of results, writing the manuscript.

Abstract

Life expectancy has increased over the last century and a growing number of people are reaching the age of 90 years and over. This phenomenon has become one of the most relevant demographic problems in contemporary society. In the nonagenarian population, pathophysiological changes affecting the central nervous system and the musculoskeletal system are frequent. To investigate the sexual dimorphic metabolic behavior in the nonagenarian cohort to understand the complex phenomenon of aging and reveal potential associations between molecular features and specific elderly pathologies, cognitive impairment, depression, and functional motor activity reduction, we firstly analysed the serum metabolic profiles of an Italian nonagenarian population, using an NMR-based metabolomics approach. Firstly, we observed that alteration in terms of metabolic and hormonal activity in elderly people influenced sex discrimination (70% of accuracy). Analysing the sex-related differences in terms of molecular features concentration, we observed that postmenopausal women had a greater serum concentration of lipid main-fraction, creatine, glycine, and less serum concentration of branched-chain amino acids (BCAAs), considered separately and in metabolic ratio, than post-andropausal men. With the aim to investigate the potential association between molecular features and elderly cognitive and motor activity reduction, we discovered that phenylalanine and glutamine/glucose ratio were associated with cognitive impairment, suggesting their role in correctly predicting the progress of pathology. Threonine/citric acid and threonine/pyruvic acid ratios were associated with geriatric depression, with a modest, but statistically significant, contribution. No metabolic associations were observed with elderly motor impairments. As a whole, our findings shed light on the sexual dimorphic mechanism of aging and reveal potential biomarkers of elderly cognitive and mood disorders.

1. Introduction

As the world population is constantly aging, the nonagenarians, often defined as those aged 90 and more tears, present the largest increase in numbers[1,2]. In Europe, the number of nonagenarians (90 and more years), 13.8 million in 2018, is expected to reach 31.8 million by 2050, while centenarians (>105 years), nearly 106000 in 2018, are expected to be almost half a million by 2050[3]. This important growth of the elderly population in developed countries has become one of the most relevant studies of the socio-demographic phenomena of contemporary society.

Aging is a complex biological process characterized by variation in systemic metabolism[4–7]. These multi-structured age-related changes determine in aged population cognitive, motor, and sensory function impairment, determining an increase in geriatric disease frequency, and *a* reduction in terms of life quality[8,9], defining a very intricated clinical and biological picture of the mechanisms of aging.

The steady deterioration of physiological activities reflects irreversible changes occurring at the molecular, cell, tissue, and, eventually, organismal levels[10,11]. It is well known that cognitive and motor activities decline with advancing age, but it has been observed that the reduction or the complete loss of cognitive functionalities and/or mobility represent one of the most significant threats to a nonagenarian's ability to live in autonomy and self-sufficiency[12–14]. The presence and the combination of cognitive and physical disorders increase in the population the risk of dementia, immobility, and mortality in nonagenarians and centenarians[12,15]. These physiological alterations are determined by the accumulation of molecular changes among the aging mechanisms[16–18], including oxidative stress, DNA mutations, errors or alteration in protein synthesis, and by-products of enzymatic reactions[19]. In this light, the use of metabolomics, a powerful *-omic* high-throughput technique, offers great promise for the understanding of the mechanisms that underlie aging [10,20,21]. The investigation of metabolic signatures associated with age and with specific pathophysiological conditions can shed light on the potential mechanism that could influence aging and longevity and can find potential biomarkers to perform early diagnosis of the cognitive and functional activity decline.

In this scenario, Nuclear magnetic resonance (NMR)-based metabolomics is a useful technique able to assign, quantify, and investigate hundreds of various molecular features[22–26], detectable in biological fluids (*i.e*. serum, plasma, urine, etc.), providing a global image of the complex metabolic, biological, and biophysical processes associated with health [24,25,27] and disease [28–31].

In this work, we take a statistical approach to investigate, firstly, the sexual dimorphic metabolic differences, relevant for gender medicine of aging[32,33], and, secondly, the

associations between metabolites, ratio metabolites, and main lipoproteins parameters and elderly clinical frequent pathology, in particular cognitive impairment, mood disorder, and motor activity impairment in a cohort of 355 nonagenarians from the Italian Mugello study [34].

2. Materials and methods

2.1 Study population

This study population is based on the framework of the Mugello Study[34], an epidemiological survey conducted from January 2010 to December 2011 on people aged 90 years and more, living in the Mugello area, north-eastward of Florence, in Tuscany (Italy). Individuals enrolled in this study consisted of 356 subjects (mean age 93 ± 3), of which 96 men (27%) and 260 women (73%), with an age range of 84-103 years and 88-105 years and with mean age 92.6 (± 3.4) and 93.2 (± 3.2) years, respectively. Information on demographic and educational data, motor and cognitive status, medical history, and pharmacological therapy used by the study cohort was collected at homes or retirement homes through objective structured clinical examinations and a series of validated questionnaires. We refer the reader to the original publication for more details on the study design and the protocols that have been described elsewhere[34,35].

2.2 Elderly clinical outcomes

The information about the individual health conditions of nonagenarians considered in this study could be listed as follow:

1) to evaluate the elderly cognitive impairment, the cognitive function was measured according to the Mini-Mental State Examination (MMSE)[36] test, in which the higher the score (0–30) indicates an active and healthy cognitive status. MMSE scores were divided, according to the existent literature, into three categories to distinguish people with severe (0–24) and absent (24-30) cognitive impairment; (24–30)[37];

2) to evaluate the elderly mood disorder, the Geriatric Depression Scale (GDS)[38] was used to consider the presence or absence of depression status; as reported above, a higher score (0–15) corresponds to a severe depressive status. GDS scores were divided into three categories to distinguish non-depressed (0–2), moderate and severe depressive status (3–15);

3) The physical performance was evaluated using the Short Physical Performance Battery (SPPB)[39], which assesses the walking speed, standing

balance, and ability to raise from a chair. The total score ranges from 0-8, indicating the bad/worst limb functions, to 8-12, the best motor performance;

4) The probability to fall and, consequently, an indirect measure of motor impairment, was evaluated using the Time up and go test[40]. The rationale for the score interpretation is the same as described above for the other outcomes. The total score is calculated in seconds to perform the specific test: <20 seconds indicates a normal motor function and >20 seconds indicates severely abnormal motor function.

2.3 Ethical considerations

The Mugello study[34] was conducted in agreement with the principles of the Helsinki Declaration on Clinical Research involving human beings (1964) and was approved by the Don Carlo Gnocchi Foundation Ethics Committee. Informed written consent was obtained from all participants or from their delegates before their inclusion in the study.

2.4 Experimental procedures

2.4.1 Sample collection

Blood samples were collected after overnight fasting, centrifuged at 2000 g for 10 minutes at 4°C, and stored in aliquots at -80° until analyses following standardized operating procedures (SOPs)[41–43].

2.4.2 NMR-sample preparation and spectra acquisition

NMR sample preparation and measurements were performed at the CERM/CIRMMP center, University of Florence, Italy. A total of 356 serum samples were prepared according to the validated procedures for NMR serum sample preparation reported by Vignoli et *al.*[23]. A Bruker 600 MHz spectrometer was used to acquire one-dimensional (1D) ^{1}H-NMR spectra. For each serum sample, three different pulse sequences have been applied to selectively detect low and high mass components: NOESY 1D presat (1D-NOESY)[44] in which both metabolites and macromolecular signals are visible; a spin echo Carr-Purcell- Meiboom-Gill (1D-CPMG)[45] for the selective observation of low molecular weight components; and a standard diffusion-edited (1D-diffusion) to permit the isolation of macromolecular components present in the solutions (Supplementary Figure 1).

Details on NMR sample preparation, spectra acquisition parameters, and data processing methods are reported in supplementary methods (Supplementary material §1.1 and §1.2).

Each proton 1D NMR spectrum, within the range of 0.2 to 10 ppm, was segmented into 0.02 ppm buckets 4 5. Spectra regions between 4.40 and 4.87 ppm (1D-CPMG), 4.40 and

4.80 ppm (1D-NOESY and 1D-diffusion), containing residual water signal, were removed. The total spectral area was calculated on the remaining 518 buckets for 1D-CMPG and 522 buckets for 1D-NOESY and 1D-diffusion.

2.4.3 Molecular features assignment and quantification

Metabolites, lipoproteins, lipid fractions, and sub-fractions were assigned, identified, and quantified using the AVANCE IVDr (Clinical Screening and In Vitro Diagnostics (IVD) research with B.I. Methods, Bruker BioSpin)[46].

2.5 Statistical analysis

2.5.1 Molecular features log-transformation and metabolite ratio calculation

The $n(n\text{-}1)/2$ ratios among all n correctly quantified and assigned metabolites (with n=20) have been also calculated. 114 lipoproteins, 20 metabolites concentrations, and 190 ratios of metabolites concentrations were transformed into logarithmic values using the formula: $\log_2(x + 1)$, where x represents the amount of a metabolite, of a ratio, or of a lipoprotein quantified in a serum spectrum. These new 324 features were used in all the following data analyses.

2.5.2 Exploratory analysis

For a preliminary approach, an unsupervised multivariate analysis, Principal component analysis (PCA)[47], was used on bucketed spectra matrices to detect the presence of any potential low-quality spectrum, outliers, and clusters.

2.5.3 Supervised Multivariate analysis

Supervised orthogonal projections to latent structures discriminant analysis (OPLS-DA)[48], cross-validated using Monte Carlo algorithm[49], was applied for data reduction and classification of bucketed spectra; for different output models, confusion matrices and total accuracies were estimated using a mean of 100 runs of Monte Carlo algorithm.

2.5.4 Univariate analysis

The different distribution of each feature in the nonagenarian dimorphic sexual-specific group was evaluated using the non-parametric Wilcoxon rank sum test[50]. To evaluate the probability that individual observations within one sex-related group are likely to be greater than the observations in the other sex-related group, the effect size, using Cliff's δ (delta) statistic[51], was calculated. For each metabolic comparison, the magnitude of the

separations was considered using the following threshold classification: $\delta < 0.147$ "negligible", $\delta < 0.33$ "small", $\delta < 0.474$ "medium" and $\delta > 0.474$ "large"[52]. To perform a descriptive analysis of sociodemographic, cognitive, and motor characteristics in the three principal groups (whole sample, women, and men), continuous variables were reported as mean ± standard deviation (SD); categorical and dichotomous variables were reported as absolute value frequency followed by percentage distribution.

2.5.5 Logistic regression

Logistic regression was performed to assess whether each measured feature, considered the dependent variable, was associated with specific motor and cognitive outcomes (Short Physical Performance Battery (SPPB), Time up and Go scores; Mini-Mental State Examination (MMSE), and Geriatric depression scale (GDS)), adjusted for appropriate confounding variables. For more details see §2.3.

Results obtained were reported as odds ratio (OR), lower and upper confidence limits, *P*-value, and adjusted *P*-value using the Benjamini-Hochberg[53] correction method based on false discovery rate (FDR). *P*-value < 0.01 was deemed statistically significant.

2.5.6 Receiver Operating Characteristic curves (ROC) analysis

Discriminant power of potential metabolite and/or lipoproteins biomarkers was evaluated by performing the receiver operating characteristics curve[54](ROC) analyses, cross-validated using the Leave One Out Cross-Validation (LOOCV)[55] algorithm. In this case, a reference model which included only clinical variables potentially affecting the outcome was constructed. Then the reference model was compared to a second model in which two statistically significant metabolites highlighted by the logistic regression analysis were added to the reference model. The areas under the curve (AUC)[56] of the two models were compared for a statistically significant increment using the AUC test.

2.6 Software

All statistical analyses were performed using R (version 4.0.3), open-source software for statistical management of data[57]. To perform ROC analysis "roc" function from the pROC[54] R package was used.

3. Results and discussion

3.1 Exploratory analysis of serum metabolomic profiles of nonagenarians

Unsupervised multivariate analysis (PCA) was performed on bucketed 1D-CPMG, 1D-NOESY, and 1D-diffusion experiments to obtain an overview of the variation observable in the spectra and to highlight the potential presence of bad quality samples, outliers, and

clusters (Figure 1). The sample coded "F_C_138" (a serum profile of a women subject) was an outlier in the PCA score plot and it was excluded after an appropriate check of the spectra. The presence of this outlier depends on the low-quality shimming of the sample. Using this exploratory, we observed that we are not able to observe clear discrimination between the two sexual dimorphic groups.

3.2 Discrimination and sexual dimorphic metabolic differences in the nonagenarian cohort

Orthogonal projections to latent structures discriminant analysis (OPLS-DA), a supervised statistical method, was built on bucketed 1D-CPMG, 1D-NOESY, and 1D-diffusion experiments and was established to relate the metabolomic data to the gender classes. Due to the unbalanced number of women and men – 259 and 96 respectively –, the OPLS-DA Monte Carlo cross-validation was performed randomly by re-sampling 50 nonagenarian women and 50 nonagenarian men at each step. The models were performed on 1D-CPMG, 1D-NOESY, and 1D-diffusion serum spectra, respectively. The average accuracies obtained for the correct identification of women were 69%, 65%, and 64% and for the correct identification of men were 72%, 69%, and 71%. The overall predictive accuracies of the model were 70%, 67%, and 68%, using, respectively, 1D-CPMG, 1D-NOESY, and 1D-diffusion spectra. In Figure 2 is reported the OPLD-DA score plot and the confusion matrices with the performance measures of the models are reported in Table 1.

Interestingly, using a fingerprinting analytic approach, the discrimination of nonagenarian men and women is less strong (accuracy ~70%) than the same discrimination in the younger population (accuracy ~90%)[58], even if still higher than that obtained for newborns (accuracy ~50%)[59]. This phenomenon is potentially connected with the remodeling, in an elderly population, of the overall metabolic and hormonal activity, generating a slight reduction in sex-dependent metabolic discrimination compared to an adult population. It is necessary to specify that this reduction does not affect the sex-dependent metabolic behavior as a whole but represents an age-derived global metabolic change-dependent similarity[32,60,61].

For a better understanding of metabolomic differences between the sex-specific nonagenarian groups, a pairwise Wilcoxon signed rank sum test was applied on 20 metabolites, on 190 ratios, and on 7 main lipid parameters. 3 metabolites (creatine, glycine and leucine), 24 metabolite ratios (glutamic acid/glycine, glutamine/glycine, creatine/histidine, glycine/histidine, creatine/isoleucine, glycine/isoleucine, creatine/leucine, glycine/leucine, creatine/phenylalanine, glycine/phenylalanine, creatine/tyrosine,

glycine/tyrosine, creatine/valine, glycine/valine, creatine/acetic acid, glutamine/acetic acid, creatine/citric acid, glycine/citric acid, creatine/acetoacetic acid, glycine/acetoacetic acid, isoleucine/acetoacetic acid, leucine/acetoacetic acid, creatine/pyruvic acid, glycine/pyruvic acid) and 5 lipid parameters (cholesterol, LDL cholesterol, HDL cholesterol, ApoA1 and ApoB100) remains significantly different after *P*-value FDR correction. In Figures 3 and Figure 4, the statistically significant differentially distributed serum metabolites (individually and in ratio considered) and lipoproteins, with an adjusted *P*-value < 0.01, were, respectively, reported. The effect size, calculated using Cliff's δ (delta) statistic, was also reported.

The alterations in the ratios between two single metabolites may describe a perturbation or a fluctuation in terms of activity in pathways relevant to a certain specific phenotype. In this work, the pairwise metabolite ratios are considered potential biomarkers of age-related sex metabolic differences[62,63].

Interestingly, we observed that women nonagenarians tend to have higher levels of creatine serum concentration. The reduction of men's serum creatine in older muscle is congruent with research that describes how skeletal muscle becomes more oxidative with aging, observing a decreased reliance on glycolysis and lactate dehydrogenase activity[64–66]. Post-menopausal women tend to have lower levels of BCAAs (Branched-Chain Amino Acids), in particular, leucine and valine, considered separately and in ratio, depending probably on the reduction of Muscle Mass Protein Synthesis (MPS) activity [67–70]. According to the literature[71,72], could observe also higher levels in terms of concentration of glycine concentrations (separately and in ratio) in elderly women, compared with the serum values in men. The biological motivation behind this difference related to sexual dimorphism in a nonagenarian population has yet to be deeply investigated. Probably, from current knowledge of molecular biology, glycine plays a fundamental role in counteracting cellular oxidative stress. It is known that glutathione, an important molecule that prevents cell damage due to oxidative stress, is composed of cysteine, glutamate, and glycine[73]. Reduction in glycine concentration directly determines and higher risk in elderly men being exposed to cell damage from oxidative stress[68].

Moreover, cholesterol, triglycerides, High-Density Lipoprotein (HDL), Low-Density Lipoprotein (LDL), apolipoprotein-A1, and apolipoprotein-B concentrations were higher in elderly women than men; this result, depends on changes in nutrient intakes, health status, and BMI levels[74,75], that tends to be altered in the elderly female population.

3.3 Potential biomarker discovery of elderly cognitive and motor impairment

To predict the risk to increase and/or decrease cognitive and motor activities in the nonagenarian cohort, logistic regression analyses were performed considering metabolites,

ratios, and main lipid component concentrations. Four clinical parameters have been considered: MMSE score, to parametrize cognitive impairment, GDS score, to parametrize the presence or absence of depression status, SPPB value, to evaluate lower limb function, and Time up and Go score, to evaluate function with correlates to balance and fall risk. Supplementary Tables 1, 2, 3, and 4 show the socio-demographic, cognitive, and motor functionality variables in the whole sample, women, and men, associated, respectively, with cognitive functionality impairment (MMSE), elderly depression (GDS), limb function (SPPB), and elderly fall risk (Time up and Go).

The logistic regression model, adjusted for age, education – both expressed in years – and gender shows that the fluctuations in terms of concentrations of phenylalanine and glucose, considered separately and/or related with other different metabolites, are significantly associated (P-value < 0.01) with the risk of cognitive impairment in whole samples enrolled and in nonagenarian women only (Table 2). In this case, no metabolites, ratios of metabolites, or lipids remain significant after correction for multiple testing, suggesting the existence of a statistically significant but not robust association. The same analysis was performed also considering the elderly depression status (Table 3). Threonine concentrations, considered separately and/or related to other metabolites, are significantly associated (P-value < 0.01) with the risk of depression in nonagenarian women only. As reported for MMSE, the associations observed are not statistically relevant after multiple corrections. In contrast, no significant metabolic associations with lower limb functionality, evaluated using the SPPB score, and with fall risk, evaluated using the Time up and Go score, were observed. No main lipid components were significantly associated with clinical parameters analyzed in this study.

To evaluate whether the most relevant features highlighted by logistic regression could provide prognostic information about the reduction of cognitive function and increase of mood disorder, area under the curve (AUC) of receiver operating characteristic (ROC) curve analyses, cross-validated using Leave One Out Cross-Validation (LOOCV) algorithm, were performed. Firstly, a model including only clinical variables was built (age and educational status information for predicting cognitive disorder, and age, BMI, handgrip, MMSE score, and the number of family nuclei per sample for predicting depression disorder). These biomarker-free models were compared to the same models with the addition of selected features. In particular, log-transformed phenylalanine concentration and log transformed glutamine/glucose ratio in terms of concentrations were selected as potential predictors for MMSE Score, while log transformed threonine/citric acid and log transformed threonine/pyruvic acid ratios in terms of concentrations were selected as predictors for GDS

score. As shown in Figure 5, the ROC cross-validated analysis demonstrated that the addition of these potential biomarkers (model 2, red lines in Figure 5) to the base model (model 1, blue lines in Figure 5) improved the area under the curve of cognitive impairment model [model 1 MMSE: AUC = 0.68 (95% CI: 0.61-0.75), model 2 MMSE: AUC = 0.75 (95% CI: 0.69-0.81), *P*-value = 0.003] and (slightly) the area under the curve of mood disorder model [model 1 GDS: AUC = 0.63 (95% CI: 0.54-0.71), model 2 GDS: AUC = 0.68 (95% CI: 0.59-0.76, *P*-value = 0.05)].

It is demonstrated that alteration in glucose and glutamate-glutamine homeostasis and metabolism is associated with an increase in the probability of developing brain pathologies, in particular dementia[76–78], confirming the result observed in this work. Furthermore, high levels of phenylalanine, also, could alter brain functionality in several ways, including competition in situ with the competition with other large neutral amino acids (LNAAs) for transport at the blood-brain barrier[79], decreased brain protein synthesis, and increased myelin turnover [80,81]. Depression disorder is associated with alteration of pyruvate metabolism, via the tricarboxylic acid (TCA) cycle, linked to the metabolism of amino acids (in particular threonine), as observed also in our study[82–84].

4. Conclusion

In this study we first addressed the problem of investigating the serum metabolic profiles of a nonagenarian population, living in an enclosed area – in Tuscany –, using an NMR-based metabolomics approach. Firstly, we examined the sex effect on metabolic profiles, and we observed that alteration and decrease in terms of metabolic and hormonal activity in elderly people influenced the ability to discriminate between the two sex-specific subgroups. Postmenopausal women had a greater serum concentration of lipid main-fraction, creatine, glycine, and less serum concentration of BCAAs (leucine and valine), considered separately and in ratio, than post-andropausal men. These results are in line with the decreasing MPS in women. Moreover, glycine reduction plays a relevant role in increasing the risk of oxidative stress in nonagenarian men. Secondly, the aim was to investigate the potential association between metabolites, ratio metabolites, and main lipid parameters and cognitive and motor activity reduction. We discovered that phenylalanine and glutamine/glucose ratio were associated with cognitive impairment, suggesting their role in correctly predicting the pathology progress. Threonine/citric acid and threonine/pyruvic acid ratios were associated with geriatric depression, with a modest, but statistically significant, contribution. No metabolic associations were observed with elderly motor impairments.

FIGURES

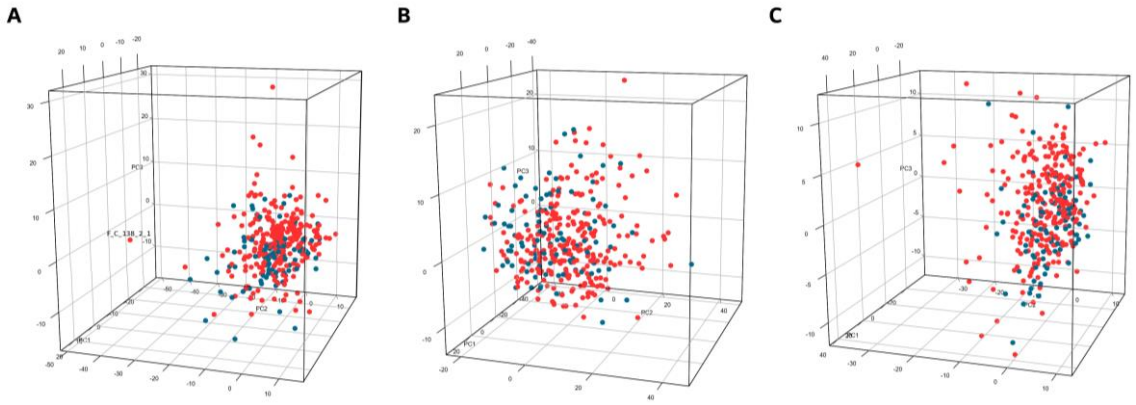


Figure 1: Principal component analysis (PCA) performed on serum nonagenarians 1H NMR spectra. Each dot represents a single NMR serum spectrum: women's samples are represented with red dots and men's samples with blue dots. A) analysis performed on bucketed 1D-CPMG spectra; B) analysis performed on bucketed 1D-NOESY spectra; C) analysis performed on bucketed 1D-diffusion spectra.

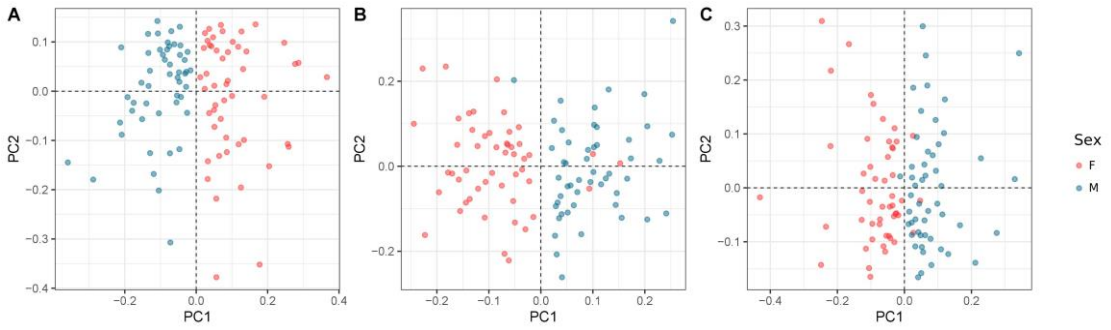


Figure 2: Supervised OPLS-DA score plots. Each dot represents a single NMR serum spectrum: a balanced group of 50 women samples was represented with red dots and 50 men with blue dots. A) analysis performed on bucketed 1D-CPMG spectra; B) analysis performed on bucketed 1D-NOESY spectra; C) analysis performed on bucketed 1D-diffusion spectra.

Figure 3: Metabolites and ratios of metabolites logarithmic concentrations in nonagenarian women and men. Different distribution of metabolites and metabolites ratios in women (red boxplot) and men (blue boxplot); for each comparison the adjusted *P*-value and the effect size were calculated using Wilcoxon rank sum-test and Cliff's δ (delta) statistic algorithm, respectively.

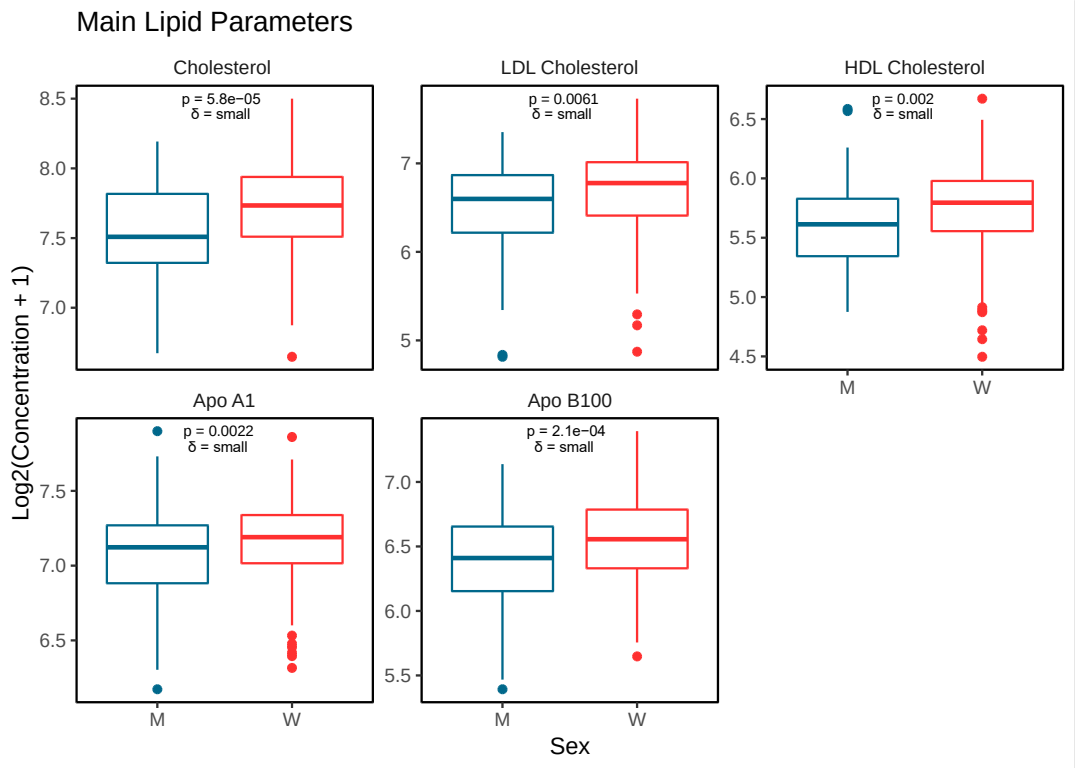


Figure 4: Main lipid parameters logarithmic concentrations in nonagenarian women and me. Different distribution of main lipids components in women (red boxplot) and men (blue boxplot); for each comparison the adjusted *P*-value and the effect size were calculated using Wilcoxon rank sum-test and Cliff's δ (delta) statistic algorithm, respectively.

TABLES

Table 1: 100 times re-sampling average confusion matrices and accuracy values of OPLS-DA cross-validate models performed on: A) 1D-CPMG nonagenarians serum spectra; B) 1D-NOESY nonagenarians serum spectra; C) 1D-diffusion nonagenarians serum spectra.

A	Women	Men
Women	69.29	30.73
Men	28.44	71.56

Accuracy = 70%

B	Women	Men
Women	65.29	34.71
Men	31.18	68.82

Accuracy = 67%

C	Women	Men
Women	64.35	35.65
Men	28.73	71.26

Accuracy = 68%

Table 2: Logistic regression model adjusted for age (years), education (years), and gender showing the association of log-transformed metabolites and ratio concentrations with MMSE.

	MMSE – whole sample				
	odd_ratio	lower confidence limit	upper confidence limit	p-value	p_adj (fdr)
Proline/Glucose	0.706	0.564	0.884	0.002	0.235
Alanine/Proline	1.347	1.074	1.667	0.009	0.235
Phenylalanine	2.660	1.272	5.561	0.009	0.235
	MMSE – Women				
Proline/Glucose	0.620	0.465	0.826	0.001	0.132
Glutamine/Glucose	0.596	0.431	0.823	0.002	0.132
Phenylalanine	5.546	1.832	16.788	0.002	0.132
Leucine	1.646	1.162	2.331	0.005	0.132
Creatine/Glucose	0.644	0.471	0.882	0.006	0.132
Alanine/Proline	1.489	1.119	1.981	0.006	0.132
Alanine/Glutamine	1.498	1.119	2.005	0.007	0.132
Acetic.acid/Glucose	0.651	0.477	0.889	0.007	0.132
Citric.acid/Glucose	0.654	0.480	0.891	0.007	0.132
Glucose	1.607	1.135	2.276	0.008	0.132
Acetoacetic.acid/Glucose	0.654	0.479	0.894	0.008	0.132
Histidine/Glucose	0.657	0.481	0.898	0.008	0.132
Pyruvic.acid/Glucose	0.659	0.482	0.899	0.008	0.132
Isoleucine/Glucose	0.665	0.488	0.906	0.009	0.132
Glycine/Glucose	0.666	0.489	0.907	0.009	0.132

Table 3: Logistic regression model adjusted for age (years), BMI (kg/m²), Handgrip (kg), living with (number of family nuclei), MMSE-score, and gender showing the association of log-transformed metabolites concentrations and ratios with GDS.

	GDS – Women				
	odd_ratio	lower confidence limit	upper confidence limit	p-value	p_adj (fdr)
Threonine/Citric.acid	1.832	1.262	2.659	0.001	0.077
Leucine/Threonine	0.567	0.396	0.811	0.002	0.077
Threonine/Acetoacetic.acid	1.768	1.232	2.537	0.002	0.077
Isoleucine/Threonine	0.571	0.400	0.815	0.002	0.077
Threonine	1.756	1.224	2.519	0.002	0.077
Threonine/Pyruvic.acid	1.775	1.224	2.574	0.002	0.077
Threonine/Tyrosine	1.731	1.212	2.472	0.003	0.077
Histidine/Threonine	0.586	0.409	0.840	0.004	0.093
Threonine/Acetic.acid	1.707	1.186	2.457	0.004	0.093
Creatine/Threonine	0.595	0.416	0.853	0.005	0.098
Threonine/Valine	1.622	1.136	2.319	0.008	0.150

References

(1) Shrestha, L. B. Population Aging In Developing Countries. *Health Affairs* **2000**, *19* (3), 204–212. https://doi.org/10.1377/hlthaff.19.3.204.

(2) Suzman, R.; Beard, J. R.; Boerma, T.; Chatterji, S. Health in an Ageing World—What Do We Know? *The Lancet* **2015**, *385* (9967), 484–486. https://doi.org/10.1016/S0140-6736(14)61597-X.

(3) Teixeira, L.; Araújo, L.; Paúl, C.; Ribeiro, O. *Centenarians: A European Overview*; Springer International Publishing, 2020.

(4) Adav, S. S.; Wang, Y. Metabolomics Signatures of Aging: Recent Advances. *Aging Dis* **2021**, *12* (2), 646–661. https://doi.org/10.14336/AD.2020.0909.

(5) Jazwinski, S. M.; Kim, S. Metabolic and Genetic Markers of Biological Age. *Front. Genet.* **2017**, *8*. https://doi.org/10.3389/fgene.2017.00064.

(6) Metz, D. H. Mobility of Older People and Their Quality of Life. *Transport Policy* **2000**, *7* (2), 149–152. https://doi.org/10.1016/S0967-070X(00)00004-4.

(7) Mäkinen, V.-P.; Ala-Korpela, M. Metabolomics of Aging Requires Large-Scale Longitudinal Studies with Replication. *Proceedings of the National Academy of Sciences* **2016**, *113* (25), E3470–E3470. https://doi.org/10.1073/pnas.1607062113.

(8) Martin, J. E.; Sheaff, M. T. The Pathology of Ageing: Concepts and Mechanisms. *J. Pathol.* **2007**, *211* (2), 111–113. https://doi.org/10.1002/path.2122.

(9) Valenzuela, J. F.; Monterola, C.; Tong, V. J. C.; Ng, T. P.; Larbi, A. Health and Disease Phenotyping in Old Age Using a Cluster Network Analysis. *Scientific Reports* **2017**, *7* (1), 15608. https://doi.org/10.1038/s41598-017-15753-3.

(10) Lorusso, J. S.; Sviderskiy, O. A.; Labunskyy, V. M. Emerging Omics Approaches in Aging Research. *Antioxid Redox Signal* **2018**, *29* (10), 985–1002. https://doi.org/10.1089/ars.2017.7163.

(11) López-Otín, C.; Blasco, M. A.; Partridge, L.; Serrano, M.; Kroemer, G. The Hallmarks of Aging. *Cell* **2013**, *153* (6), 1194–1217. https://doi.org/10.1016/j.cell.2013.05.039.

(12) Pancani, S.; Vannetti, F.; Sofi, F.; Cecchi, F.; Pasquini, G.; Fabbri, L.; Mosca, I. E.; Macchi, C.; Group, T. M. S. W. Association between Physical Activity and Functional and Cognitive Status in Nonagenarians: Results from the Mugello Study. *International Psychogeriatrics* **2019**, *31* (6), 901–908. https://doi.org/10.1017/S1041610218001424.

(13) Dinu, M.; Colombini, B.; Pagliai, G.; Vannetti, F.; Pasquini, G.; Molino Lova, R.; Cecchi, F.; Sorbi, S.; Sofi, F.; Macchi, C. BMI, Functional and Cognitive Status in a Cohort of Nonagenarians: Results from the Mugello Study. *Eur Geriatr Med* **2021**, *12* (2), 379–386. https://doi.org/10.1007/s41999-020-00417-9.

(14) Mänty, M.; Thinggaard, M.; Christensen, K.; Avlund, K. Musculoskeletal Pain and Physical Functioning in the Oldest Old. *European Journal of Pain* **2014**, *18* (4), 522–529. https://doi.org/10.1002/j.1532-2149.2013.00386.x.

(15) Pais, R.; Ruano, L.; Moreira, C.; Carvalho, O. P.; Barros, H. Prevalence and Incidence of Cognitive Impairment in an Elder Portuguese Population (65–85 Years Old). *BMC Geriatrics* **2020**, *20* (1), 470. https://doi.org/10.1186/s12877-020-01863-7.

(16) Kirkwood, T. B.; Austad, S. N. Why Do We Age? *Nature* **2000**, *408* (6809), 233–238. https://doi.org/10.1038/35041682.

(17) Golubev, A.; Hanson, A. D.; Gladyshev, V. N. Non-Enzymatic Molecular Damage as a Prototypic Driver of Aging. *J Biol Chem* **2017**, *292* (15), 6029–6038. https://doi.org/10.1074/jbc.R116.751164.

(18) Yin, D.; Chen, K. The Essential Mechanisms of Aging: Irreparable Damage Accumulation of Biochemical Side-Reactions. *Exp Gerontol* **2005**, *40* (6), 455–465. https://doi.org/10.1016/j.exger.2005.03.012.

(19) Gladyshev, V. N. The Free Radical Theory of Aging Is Dead. Long Live the Damage Theory! *Antioxid Redox Signal* **2014**, *20* (4), 727–731. https://doi.org/10.1089/ars.2013.5228.

(20) Valdes, A. M.; Glass, D.; Spector, T. D. Omics Technologies and the Study of Human Ageing. *Nature Reviews Genetics* **2013**, *14* (9), 601–607. https://doi.org/10.1038/nrg3553.

(21) Zierer, J.; Menni, C.; Kastenmüller, G.; Spector, T. D. Integration of 'Omics' Data in Aging Research: From Biomarkers to Systems Biology. *Aging Cell* **2015**, *14* (6), 933–944. https://doi.org/10.1111/acel.12386.
(22) 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.
(23) Vignoli, A.; Ghini, V.; Meoni, G.; Licari, C.; Takis, P. G.; Tenori, L.; Turano, P.; Luchinat, C. High-Throughput Metabolomics by 1D NMR. *Angewandte Chemie International Edition* **2019**, *58* (4), 968–994. https://doi.org/10.1002/anie.201804736.
(24) 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.
(25) Di Cesare, F.; Tenori, L.; Meoni, G.; Gori, A. M.; Marcucci, R.; Giusti, B.; Molino-Lova, R.; Macchi, C.; Pancani, S.; Luchinat, C.; Saccenti, E. Lipid and Metabolite Correlation Networks Specific to Clinical and Biochemical Covariate Show Differences Associated with Sexual Dimorphism in a Cohort of Nonagenarians. *GeroScience* **2022**, *44* (2), 1109–1128. https://doi.org/10.1007/s11357-021-00404-3.
(26) Licari, C.; Tenori, L.; Giusti, B.; Sticchi, E.; Kura, A.; De Cario, R.; Inzitari, D.; Piccardi, B.; Nesi, M.; Sarti, C.; Arba, F.; Palumbo, V.; Nencini, P.; Marcucci, R.; Gori, A. M.; Luchinat, C.; Saccenti, E. Analysis of Metabolite and Lipid Association Networks Reveals Molecular Mechanisms Associated with 3-Month Mortality and Poor Functional Outcomes in Patients with Acute Ischemic Stroke after Thrombolytic Treatment with Recombinant Tissue Plasminogen Activator. *J Proteome Res* **2021**, *20* (10), 4758–4770. https://doi.org/10.1021/acs.jproteome.1c00406.
(27) Wishart, D. S. Metabolomics for Investigating Physiological and Pathophysiological Processes. *Physiological Reviews* **2019**, *99* (4), 1819–1875. https://doi.org/10.1152/physrev.00035.2018.
(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) Kim, E. R.; Kwon, H. N.; Nam, H.; Kim, J. J.; Park, S.; Kim, Y.-H. Urine-NMR Metabolomics for Screening of Advanced Colorectal Adenoma and Early Stage Colorectal Cancer. *Scientific Reports* **2019**, *9* (1), 4786. https://doi.org/10.1038/s41598-019-41216-y.
(30) 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), 1–13. https://doi.org/10.1038/s41598-019-40028-4.
(31) Hart, C. D.; Vignoli, A.; Tenori, L.; Uy, G. L.; Van To, T.; Adebamowo, C.; Hossain, S. M.; Biganzoli, L.; Risi, E.; Love, R. R.; Luchinat, C.; Di Leo, A. Serum Metabolomic Profiles Identify ER-Positive Early Breast Cancer Patients at Increased Risk of Disease Recurrence in a Multicenter Population. *Clin. Cancer Res.* **2017**, *23* (6), 1422–1431. https://doi.org/10.1158/1078-0432.CCR-16-1153.
(32) Ostan, R.; Monti, D.; Gueresi, P.; Bussolotto, M.; Franceschi, C.; Baggio, G. Gender, Aging and Longevity in Humans: An Update of an Intriguing/Neglected Scenario Paving the Way to a Gender-Specific Medicine. *Clin Sci (Lond)* **2016**, *130* (19), 1711–1725. https://doi.org/10.1042/CS20160004.
(33) Almagro, P.; Ponce, A.; Komal, S.; Villaverde, M. de la A.; Castrillo, C.; Grau, G.; Simon, L.; Sierra, A. de la. Multimorbidity Gender Patterns in Hospitalized Elderly Patients. *PLOS ONE* **2020**, *15* (1), e0227252. https://doi.org/10.1371/journal.pone.0227252.
(34) Molino Lova, R.; Sofi, F.; Pasquini, G.; Gori, A. M.; Vannetti, F.; Abbate, R.; Gensini, G.; Macchi, C. The Mugello Study, a Survey of Nonagenarians Living in Tuscany: Design, Methods and Participants' General Characteristics. *European journal of internal medicine* **2013**, *24*. https://doi.org/10.1016/j.ejim.2013.09.008.
(35) Padua, L.; Pasqualetti, P.; Coraci, D.; Imbimbo, I.; Giordani, A.; Loreti, C.; Marra, C.; Molino-Lova, R.; Pasquini, G.; Simonelli, I.; Vannetti, F.; Macchi, C.; Mugello Study Working Group. Gender Effect on Well-Being of the Oldest Old: A Survey of Nonagenarians Living in Tuscany: The Mugello Study. *Neurol Sci* **2018**, *39* (3), 509–517. https://doi.org/10.1007/s10072-017-3223-z.

(36) Folstein, M. F.; Folstein, S. E.; McHugh, P. R. "Mini-Mental State": A Practical Method for Grading the Cognitive State of Patients for the Clinician. *Journal of Psychiatric Research* **1975**, *12* (3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6.
(37) Folstein, M.; Anthony, J. C.; Parhad, I.; Duffy, B.; Gruenberg, E. M. The Meaning of Cognitive Impairment in the Elderly. *Journal of the American Geriatrics Society* **1985**, *33* (4), 228–235. https://doi.org/10.1111/j.1532-5415.1985.tb07109.x.
(38) Yesavage, J. A.; Brink, T. L.; Rose, T. L.; Lum, O.; Huang, V.; Adey, M.; Leirer, V. O. Development and Validation of a Geriatric Depression Screening Scale: A Preliminary Report. *J Psychiatr Res* **1982**, *17* (1), 37–49. https://doi.org/10.1016/0022-3956(82)90033-4.
(39) Treacy, D.; Hassett, L. The Short Physical Performance Battery. *Journal of Physiotherapy* **2017**, *64*. https://doi.org/10.1016/j.jphys.2017.04.002.
(40) Beauchet, O.; Fantino, B.; Allali, G.; Muir, S.; Montero-Odasso, M.; Annweiler, C. Timed Up and Go Test and Risk of Falls in Older Adults: A Systematic Review. *The journal of nutrition, health & aging* **2011**, *15*, 933–938. https://doi.org/10.1007/s12603-011-0062-0.
(41) 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.
(42) Ghini, V.; Quaglio, D.; Luchinat, C.; Turano, P. NMR for Sample Quality Assessment in Metabolomics. *N Biotechnol* **2019**, *52*, 25–34. https://doi.org/10.1016/j.nbt.2019.04.004.
(43) Ghini, V.; Abuja, P. M.; Polasek, O.; Kozera, L.; Laiho, P.; Anton, G.; Zins, M.; Klovins, J.; Metspalu, A.; Wichmann, H.-E.; Gieger, C.; Luchinat, C.; Zatloukal, K.; Turano, P. Impact of the Pre-Examination Phase on Multicenter Metabolomic Studies. *New Biotechnology* **2022**, *68*, 37–47. https://doi.org/10.1016/j.nbt.2022.01.006.
(44) Macias, S.; Kirma, J.; Yilmaz, A.; Moore, S. E.; McKinley, M. C.; McKeown, P. P.; Woodside, J. V.; Graham, S. F.; Green, B. D. Application of 1H-NMR Metabolomics for the Discovery of Blood Plasma Biomarkers of a Mediterranean Diet. *Metabolites* **2019**, *9* (10), 201. https://doi.org/10.3390/metabo9100201.
(45) 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.
(46) *AVANCE IVDr - Clinical Screening and In Vitro Diagnostics (IVD) | Bruker*. https://www.bruker.com/products/mr/nmr/avance-ivdr.html (accessed 2020-03-16).
(47) Hotelling, H. Analysis of a Complex of Statistical Variables into Principal Components. *Journal of Educational Psychology* **1933**, *24* (6), 417–441. https://doi.org/10.1037/h0071325.
(48) Westerhuis, J. A.; van Velzen, E. J. J.; Hoefsloot, H. C. J.; Smilde, A. K. Multivariate Paired Data Analysis: Multilevel PLSDA versus OPLSDA. *Metabolomics* **2010**, *6* (1), 119–128. https://doi.org/10.1007/s11306-009-0185-z.
(49) Xu, Q.-S.; Liang, Y.-Z.; Du, Y.-P. Monte Carlo Cross-Validation for Selecting a Model and Estimating the Prediction Error in Multivariate Calibration. *Journal of Chemometrics* **2004**, *18* (2), 112–120. https://doi.org/10.1002/cem.858.
(50) Wilcoxon, F. Individual Comparisons by Ranking Methods. *Biometrics Bulletin* **1945**, *1* (6), 80–83. https://doi.org/10.2307/3001968.
(51) Macbeth, G.; Razumiejczyk, E.; Ledesma, R. D. Cliff's Delta Calculator: A Non-Parametric Effect Size Program for Two Groups of Observations. *Universitas Psychologica* **2011**, *10* (2), 545–555.
(52) Romano, J.; Kromrey, J. Appropriate Statistics for Ordinal Level Data: Should We Really Be Using t-Test and Cohen's d for Evaluating Group Differences on the NSSE and Other Surveys? **2006**.
(53) 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**, *57* (1), 289–300.
(54) Sing, T.; Sander, O.; Beerenwinkel, N.; Lengauer, T.; Unterthiner, T.; Ernst, F. G. M. ROCR: Visualizing the Performance of Scoring Classifiers, 2020.
(55) Kneser, R.; Essen, U.; Ney, H. On the Use of the Leaving-One-Out Method in Statistical Language Modelling. In *Speech Recognition and Coding*; Ayuso, A. J. R., Soler, J. M. L., Eds.;

NATO ASI Series; Springer: Berlin, Heidelberg, 1995; pp 256–259. https://doi.org/10.1007/978-3-642-57745-1_35.

(56) Senaratna, D. M.; Sooriyarachchim, M. R.; Meyen, N. Bivariate Test for Testing the EQUALITY of the Average Areas under Correlated Receiver Operating Characteristic Curves (Test for Comparing of AUC's of Correlated ROC Curves). *American Journal of Applied Mathematics and Statistics* **2015**, *3* (5), 190–198. https://doi.org/10.12691/ajams-3-5-3.

(57) *RStudio | Open source & professional software for data science teams*. https://www.rstudio.com/ (accessed 2022-06-14).

(58) 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.

(59) Diaz, S. O.; Pinto, J.; Barros, A. S.; Morais, E.; Duarte, D.; Negrão, F.; Pita, C.; Almeida, M. do C.; Carreira, I. M.; Spraul, M.; Gil, A. M. Newborn Urinary Metabolic Signatures of Prematurity and Other Disorders: A Case Control Study. *J. Proteome Res.* **2016**, *15* (1), 311–325. https://doi.org/10.1021/acs.jproteome.5b00977.

(60) Nakamura, E.; Miyao, K. Sex Differences in Human Biological Aging. *The Journals of Gerontology: Series A* **2008**, *63* (9), 936–944. https://doi.org/10.1093/gerona/63.9.936.

(61) Saito, K.; Maekawa, K.; Kinchen, J. M.; Tanaka, R.; Kumagai, Y.; Saito, Y. Gender- and Age-Associated Differences in Serum Metabolite Profiles among Japanese Populations. *Biological and Pharmaceutical Bulletin* **2016**, *39* (7), 1179–1186. https://doi.org/10.1248/bpb.b16-00226.

(62) Zelezniak, A.; Sheridan, S.; Patil, K. R. Contribution of Network Connectivity in Determining the Relationship between Gene Expression and Metabolite Concentration Changes. *PLoS Comput Biol* **2014**, *10* (4). https://doi.org/10.1371/journal.pcbi.1003572.

(63) Krumsiek, J.; Stückler, F.; Suhre, K.; Gieger, C.; Spector, T. D.; Soranzo, N.; Kastenmüller, G.; Theis, F. J. Network-Based Metabolite Ratios for an Improved Functional Characterization of Genome-Wide Association Study Results. *bioRxiv* **2016**, 048512. https://doi.org/10.1101/048512.

(64) Lanza, I. R.; Befroy, D. E.; Kent-Braun, J. A. Age-Related Changes in ATP-Producing Pathways in Human Skeletal Muscle in Vivo. *Journal of Applied Physiology* **2005**, *99* (5), 1736–1744. https://doi.org/10.1152/japplphysiol.00566.2005.

(65) Larsson, L.; Grimby, G.; Karlsson, J. Muscle Strength and Speed of Movement in Relation to Age and Muscle Morphology. *J Appl Physiol Respir Environ Exerc Physiol* **1979**, *46* (3), 451–456. https://doi.org/10.1152/jappl.1979.46.3.451.

(66) Rawson, E. S.; Venezia, A. C. Use of Creatine in the Elderly and Evidence for Effects on Cognitive Function in Young and Old. *Amino Acids* **2011**, *40* (5), 1349–1362. https://doi.org/10.1007/s00726-011-0855-9.

(67) Borack, M. S.; Volpi, E. Efficacy and Safety of Leucine Supplementation in the Elderly. *J. Nutr.* **2016**, *146* (12), 2625S-2629S. https://doi.org/10.3945/jn.116.230771.

(68) Canfield, C.-A.; Bradshaw, P. C. Amino Acids in the Regulation of Aging and Aging-Related Diseases. *Translational Medicine of Aging* **2019**, *3*, 70–89. https://doi.org/10.1016/j.tma.2019.09.001.

(69) Fukagawa, N. K. Protein and Amino Acid Supplementation in Older Humans. *Amino Acids* **2013**, *44* (6), 1493–1509. https://doi.org/10.1007/s00726-013-1480-6.

(70) Hutson, S. M.; Sweatt, A. J.; Lanoue, K. F. Branched-Chain [Corrected] Amino Acid Metabolism: Implications for Establishing Safe Intakes. *J. Nutr.* **2005**, *135* (6 Suppl), 1557S-64S. https://doi.org/10.1093/jn/135.6.1557S.

(71) Mittelstrass, K.; Ried, J. S.; Yu, Z.; Krumsiek, J.; Gieger, C.; Prehn, C.; Roemisch-Margl, W.; Polonikov, A.; Peters, A.; Theis, F. J.; Meitinger, T.; Kronenberg, F.; Weidinger, S.; Wichmann, H. E.; Suhre, K.; Wang-Sattler, R.; Adamski, J.; Illig, T. Discovery of Sexual Dimorphisms in Metabolic and Genetic Biomarkers. *PLoS Genet* **2011**, *7* (8), e1002215. https://doi.org/10.1371/journal.pgen.1002215.

(72) Darst, B. F.; Koscik, R. L.; Hogan, K. J.; Johnson, S. C.; Engelman, C. D. Longitudinal Plasma Metabolomics of Aging and Sex. *Aging (Albany NY)* **2019**, *11* (4), 1262–1282. https://doi.org/10.18632/aging.101837.

(73) Grimble, R. F.; Jackson, A. A.; Persaud, C.; Wride, M. J.; Delers, F.; Engler, R. Cysteine and Glycine Supplementation Modulate the Metabolic Response to Tumor Necrosis Factor α in Rats Fed a Low Protein Diet. *The Journal of Nutrition* **1992**, *122* (11), 2066–2073. https://doi.org/10.1093/jn/122.11.2066.

(74) Lamon-Fava, S.; Jenner, J. L.; Jacques, P. F.; Schaefer, E. J. Effects of Dietary Intakes on Plasma Lipids, Lipoproteins, and Apolipoproteins in Free-Living Elderly Men and Women. *Am. J. Clin. Nutr.* **1994**, *59* (1), 32–41. https://doi.org/10.1093/ajcn/59.1.32.

(75) Ettinger, W. H.; Wahl, P. W.; Kuller, L. H.; Bush, T. L.; Tracy, R. P.; Manolio, T. A.; Borhani, N. O.; Wong, N. D.; O'Leary, D. H. Lipoprotein Lipids in Older People. Results from the Cardiovascular Health Study. The CHS Collaborative Research Group. *Circulation* **1992**, *86* (3), 858–869. https://doi.org/10.1161/01.cir.86.3.858.

(76) Aldana, B. I.; Zhang, Y.; Jensen, P.; Chandrasekaran, A.; Christensen, S. K.; Nielsen, T. T.; Nielsen, J. E.; Hyttel, P.; Larsen, M. R.; Waagepetersen, H. S.; Freude, K. K. Glutamate-Glutamine Homeostasis Is Perturbed in Neurons and Astrocytes Derived from Patient IPSC Models of Frontotemporal Dementia. *Molecular Brain* **2020**, *13* (1), 125. https://doi.org/10.1186/s13041-020-00658-6.

(77) Nugent, S.; Croteau, E.; Pifferi, F.; Fortier, M.; Tremblay, S.; Turcotte, E.; Cunnane, S. C. Brain and Systemic Glucose Metabolism in the Healthy Elderly Following Fish Oil Supplementation. *Prostaglandins, Leukotrienes and Essential Fatty Acids* **2011**, *85* (5), 287–291. https://doi.org/10.1016/j.plefa.2011.04.008.

(78) Yang, Z.; Wang, J.; Chen, J.; Luo, M.; Xie, Q.; Rong, Y.; Wu, Y.; Cao, Z.; Liu, Y. High-Resolution NMR Metabolomics of Patients with Subjective Cognitive Decline plus: Perturbations in the Metabolism of Glucose and Branched-Chain Amino Acids. *Neurobiology of Disease* **2022**, *171*, 105782. https://doi.org/10.1016/j.nbd.2022.105782.

(79) Research, I. of M. (US) C. on M. N. *Amino Acid and Protein Requirements: Cognitive Performance, Stress, and Brain Function*; National Academies Press (US), 1999.

(80) Surtees, R.; Blau, N. The Neurochemistry of Phenylketonuria. *Eur J Pediatr* **2000**, *159* (2), S109–S113. https://doi.org/10.1007/PL00014370.

(81) Ravaglia, G.; Forti, P.; Maioli, F.; Bianchi, G.; Martelli, M.; Talerico, T.; Servadei, L.; Zoli, M.; Mariani, E. Plasma Amino Acid Concentrations in Patients with Amnestic Mild Cognitive Impairment or Alzheimer Disease. *Am J Clin Nutr* **2004**, *80* (2), 483–488. https://doi.org/10.1093/ajcn/80.2.483.

(82) Hung, C.-I.; Lin, G.; Chiang, M.-H.; Chiu, C.-Y. Metabolomics-Based Discrimination of Patients with Remitted Depression from Healthy Controls Using 1H-NMR Spectroscopy. *Sci Rep* **2021**, *11* (1), 15608. https://doi.org/10.1038/s41598-021-95221-1.

(83) F. Guerreiro Costa, L. N.; Carneiro, B. A.; Alves, G. S.; Lins Silva, D. H.; Faria Guimaraes, D.; Souza, L. S.; Bandeira, I. D.; Beanes, G.; Miranda Scippa, A.; Quarantini, L. C. Metabolomics of Major Depressive Disorder: A Systematic Review of Clinical Studies. *Cureus 14* (3), e23009. https://doi.org/10.7759/cureus.23009.

(84) Hashimoto, K.; Yoshida, T.; Ishikawa, M.; Fujita, Y.; Niitsu, T.; Nakazato, M.; Watanabe, H.; Sasaki, T.; Shiina, A.; Hashimoto, T.; Kanahara, N.; Hasegawa, T.; Enohara, M.; Kimura, A.; Iyo, M. Increased Serum Levels of Serine Enantiomers in Patients with Depression. *Acta Neuropsychiatrica* **2016**, *28* (3), 173–178. https://doi.org/10.1017/neu.2015.59.

4.1.3 Age and sex dependent changes of free circulating blood metabolite and lipid abundances, correlations and ratios

Francesca Di Cesare[1], Claudio Luchinat[1,2], Leonardo Tenori[1,2], Edoardo Saccenti[3]

[1] Magnetic Resonance Center (CERM), University of Florence, Via Luigi Sacconi 6, 50019, Sesto Fiorentino, Firenze, Italy

[2] Department of Chemistry "Ugo Schiff", University of Florence, Via della Lastruccia 3, 50019, Sesto Fiorentino, Italy

[3] Laboratory of Systems and Synthetic Biology, Wageningen University & Research, Stippeneng 4, 6708 WE, Wageningen, the Netherlands.

Published

Journals of Gerontology: Biological Sciences, **2022**, Vol. 77, No. 5, 918–926

https://doi.org/10.1093/gerona/glab335

Candidate's contribuitions: Statistical analysis, interpretation of results, writing, and review of the manuscript.

THE GERONTOLOGICAL SOCIETY OF AMERICA®

Journals of Gerontology: Biological Sciences
cite as: *J Gerontol A Biol Sci Med Sci*, 2022, Vol. 77, No. 5, 918–926
https://doi.org/10.1093/gerona/glab335
Advance Access publication November 8, 2021

OXFORD

Original Article

Age- and Sex-Dependent Changes of Free Circulating Blood Metabolite and Lipid Abundances, Correlations, and Ratios

Francesca Di Cesare, MSc,[1] Claudio Luchinat, PhD,[1,2] Leonardo Tenori, PhD,[1,2] and Edoardo Saccenti, PhD[3,*]

[1]Magnetic Resonance Center (CERM), University of Florence, Firenze, Italy. [2]Department of Chemistry "Ugo Schiff," University of Florence, Sesto Fiorentino, Italy. [3]Laboratory of Systems and Synthetic Biology, Wageningen University & Research, Wageningen, The Netherlands.

*Address correspondence to: Edoardo Saccenti, PhD, Laboratory of Systems and Synthetic Biology, Wageningen University & Research, Stippeneng 4, 6708 WE Wageningen, The Netherlands. E-mail: edoardo.saccenti@wur.nl

Received: August 3, 2021; Editorial Decision Date: October 30, 2021

Decision Editor: David G. Le Couteur, MBBS, FRACP, PhD

Abstract

In this study, we investigated how the concentrations, pairwise correlations and ratios of 202 free circulating blood metabolites and lipids vary with age in a panel of $n = 1\ 882$ participants with an age range from 48 to 94 years. We report a statistically significant sex-dependent association with age of a panel of metabolites and lipids involving, in women, linoleic acid, α-linoleic acid, and carnitine, and, in men, monoacylglycerols and lysophosphatidylcholines. Evaluating the association of correlations among metabolites and/or lipids with age, we found that phosphatidylcholines correlations tend to have a positive trend associated with age in women, and monoacylglycerols and lysophosphatidylcholines correlations tend to have a negative trend associated with age in men. The association of ratio between molecular features with age reveals that decanoyl-L-carnitine/lysophosphatidylcholine ratio in women "decrease" with age, while L-carnitine/phosphatidylcholine and L-acetylcarnitine/phosphatidylcholine ratios in men "increase" with age. These results suggest an age-dependent remodeling of lipid metabolism that induces changes in cell membrane bilayer composition and cell cycle mechanisms. Furthermore, we conclude that lipidome is directly involved in this age-dependent differentiation. Our results demonstrate that, using a comprehensive approach focused on the changes of concentrations and relationships of blood metabolites and lipids, as expressed by their correlations and ratios, it is possible to obtain relevant information about metabolic dynamics associated with age.

Keywords: Correlation analysis, Gender differences, Human aging, Lipids, Metabolomics

Aging is a very complex process, influenced by genetic, environmental, and lifestyle factors (1,2), and involves progressive systemic dysregulation, affecting all levels of an organism, from molecules to organs (3,4). Metabolomics, that is the comprehensive analysis of small molecule profiles measured in a biological sample like blood or urine (5,6), is an excellent approach to obtain a global representation of the metabolic status of an organism with respect to a healthy status or a particular pathophysiological condition (7–9). The analysis of metabolomic profiles obtained from participants of different ages, performed using an integrative systems biology approach (10), allows the comprehensive description of the metabolic dynamics and can help to quantify and decipher the relationships between molecular features and aging process (4,11). Studies have been conducted in humans, highlighting how the metabolome is sex and age-dependent, indicating sex-specific association of certain genetic loci with several metabolites and lipid species: the levels of many metabolites (among them fatty acids, including 10 long-chain fatty acids, polyunsaturated fatty acids, glutamine, tyrosine, and histidine) and variation thereof are highly dependent on sex and age, and that sex differentially influences the levels and variation over time of many metabolites (12,13).

Correlations and ratios among molecules, and not only their levels, bear relevant biological information: Because molecules behave in an orchestrated way through metabolic pathways, changes in their association patterns, as represented by correlations and ratios (14,15), can provide information on the remodulation of

biochemical reaction networks and metabolic pathways associated with age or sexual dimorphism, thus suggesting mechanisms through which molecules may modify cell membranes and affect hormonal activities, mitochondrial metabolism, and cell responses to oxidative stress (11,16).

In this study, making use of publicly available data, we took a comprehensive system biology approach, focusing on the association of the blood circulating unconjugated metabolites and lipids with age and sex in a large population cohort with an age range between 48 and 98 years (17). We investigated how metabolite and lipid abundances correlate with age groups, but also how the correlation and the ratios between metabolites and lipids change in groups of participants of different (increasing) ages.

Material and Methods

Experimental Data

We used data from the TwinGene project (17) that includes a longitudinal cohort from the Swedish Twin Register and a matched subcontrol cohort stratified on age and sex. The cohort was selected by Ganna et al. (17). This data set is a valid representation of a population consisting of not related participants and with a wide age range. It contains 202 quantified blood metabolites and lipids measured on n = 2 139 participants (n_W = 921 women [43%] and n_M = 1 218 men [57%]) with an overall age range of 47.6–93.9 years (women age range = 48.4–93.9 years and men age range = 47.6–93.3 years) and with an overall average age of 68.8 (women average age = 68.8 years and men average age = 68.7 years). This data set was used to identify potential molecular features and metabolic pathways associated with the sex-related aging process. Data were downloaded from the MetaboLights database (https://www.ebi.ac.uk/metabolights/) with accession number MTBLS93. Briefly, metabolomic profiling was performed on ultra-performance liquid chromatography to quadrupole time-of-flight mass spectrometry with an atmospheric electrospray interface operating in positive ion mode. The first step was the detection, alignment, grouping, and assignment of metabolites, performed by Ganna et al. (17), using the XCSM software. For the metabolic annotation, 4 approaches were performed by the authors: (a) based on matching accurate mass, fragmentation pattern, and retention time with their in-house spectral library of authentic standards collected; (b) based on spectrum and/or m/z similarities, but not retention time, and the annotation relies on the information of public databases; (c) based on the combination of spectral data, accurate mass, and retention time to assign the metabolite to a specific chemical class; (d) the other approaches failed in the annotation of the metabolite and the metabolite was annotated as "unknown." Combining these approaches, m = 202 molecular features, divided into m_1 = 36 metabolites and m_2 = 166 lipids and lipid precursors, were assigned in the original publication (Supplementary Table 1).

For further details, we refer the reader to the original publication (17).

Data Preprocessing

Removal of outliers

To obtain a uniform study population, we removed those participants showing outlying blood metabolites and lipid profiles under the assumption of the presence of possibly undiagnosed pathophysiological conditions. Outliers were removed using a Principal Components Analysis (PCA) based approach. Hotelling's T^2 values were calculated from PCA scores; samples whose T^2 values exceeded the 95% confidence ellipsis were considered outliers and were removed from subsequent analysis. The optimal number of significant principal components to be retained (at the α = 0.05 level) was determined using a statistical test based on the Tracy–Widom distribution (18). A total of 117 women (18%) and 140 men (11%) were removed from the analysis. This left n = 1 882 (n_W = 804 women, 43%, n_M = 1 078 men, 57%) samples/participants available for further analysis.

Subject stratification

The n_W = 804 women and the n_M = 1 078 men were separately stratified by age in 20 groups, W_t (for women) and M_t (for men) with t = 1,2, …, 20 of size w_t and m_t by taking the 20 quantiles QT_1, QT_2, …, QT_{20} of the women and men age distributions, reflecting the 5th, 10th, …, 95th, and 100th percentiles of the sex-specific age distribution. Consequently, each W_t group and M_t group had approximately 5% of the sex-specific sample (≅40 for women and ≅54 for men). The age characteristics for each women and men group are given in Supplementary Table 2. A graphical illustration is shown in Figure 1. For each W_t and M_t group, we defined the corresponding data matrices $\mathbf{W}_t$ and $\mathbf{M}_t$ of size $w_t \times p$ and $m_t \times p$ containing the concentrations of the p = 202 metabolites and lipids measured on the w_t and m_t participants in the corresponding group. Each set of data matrices is associated with a 1 × 20 vector $\mathbf{t}_M$ (respectively $\mathbf{t}_W$) containing the average age of the $M_1, M_2, \ldots, M_{20}$ group (respectively $W_1, W_2, \ldots, W_{20}$).

Statistical Analysis

Estimation of the average concentration of molecular features specific to age groups

For each data set $\mathbf{W}_t$ and $\mathbf{M}_t$ we calculated the mean abundance m_t between each molecular feature x_i. As for the correlation case, we obtained thus 20 values for each metabolite–lipid, representing the changes of the average abundance of molecular feature $\mathbf{x}_i$ associated with the age groups (a graphical representation is shown in Figure 2A):

We considered the standard mean estimation:

$$a_i = \frac{1}{n}\sum_{k=1}^{n} x_i \quad (1)$$

For each feature, we thus obtained 20 mean values:

$$\mathbf{A}(x_i) = \{a_{i\ (t=1)},\ a_{i\ (t=2)},\ \ldots,\ a_{i\ (t=20)}\} \quad (2)$$

Estimation of correlations between molecular features specific to age groups

For each data set $\mathbf{W}_t$ and $\mathbf{M}_t$ of size $w_t \times p$ and $m_t \times p$, we calculated the correlation r_{ij} between each pair of molecular features x_i, x_j. For each pair, we obtained thus 20 correlation values, representing the evolution of the strength of the relationship between molecular features x_i, x_j associated with different age groups (Figure 2A):

$$\mathbf{C}(x_i,\ x_j) = \{r_{ij\ (t=1)},\ r_{ij\ (t=2)},\ \ldots,\ r_{ij\ (t=20)}\} \quad (3)$$

We used Winsorized correlation coefficients that are robust toward the shape, sample size, and outliers in the metabolite concentration distribution (19) to estimate the correlation r_{ij} within molecular features pairs. The Winsorized correlation coefficient is obtained by replacing the k smallest observations with the $(k + 1)$st smallest

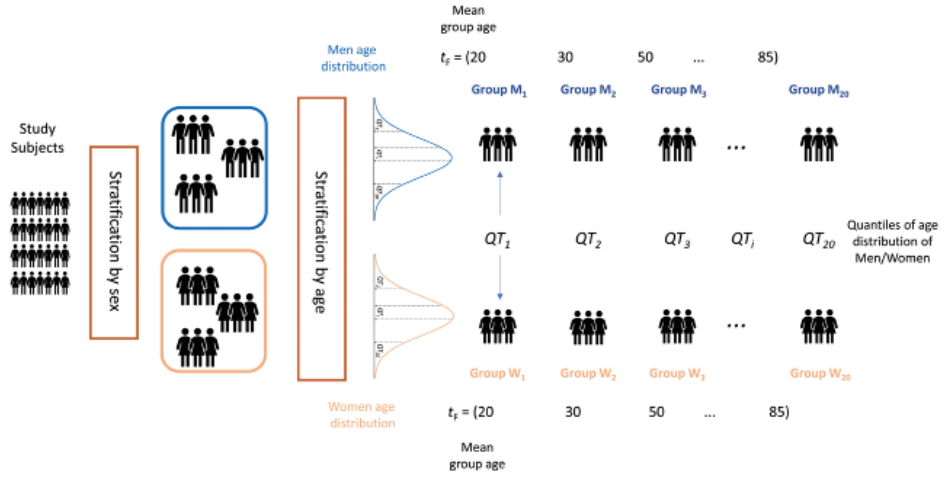


Figure 1. Overview of stratification of the study participants. Participants are first stratified by sex and then by age. Women and men are divided into 20 groups according to the 20 quantiles obtained from the age distribution of the 2 sex-specific groups.

observation, and the k largest observations with the $(k + 1)$st largest observation. In this way, the observations are Winsorized at each end of both x_i, and x_j. The Pearson's correlation coefficient is then calculated on the Winsorized variables (20). A 10% Winsorization was used. Among the $\frac{1}{2}p(p-1)$ possible correlations we retained for further analysis only those pairs of molecular features for which the correlation r_{ij} was found to be significant at the $\alpha = 0.01$ level in at least 10 of the 20 data sets $\mathbf{W}_t$ and $\mathbf{M}_t$.

Estimation of ratios between molecular features specific to age groups

For each data set $\mathbf{W}_t$ and $\mathbf{M}_t$, we calculated the ratio q_{ij} between each pair of molecular features x_i, x_j. As for the correlation case, we thus obtained 20 values for each pair, representing the evolution of the ratio magnitude of molecular features x_i and x_j (Figure 2A). We considered the unbiased ratio estimator proposed by van Kempen and van Vliet (21) which is defined as:

$$q_{ij} = \frac{\overline{x_i}}{\overline{x_j}} - \frac{1}{n}\left(\frac{\overline{x_i}}{\overline{x_j}}\mathrm{var}(x_i) - \frac{\mathrm{cov}(x_i,x_j)}{\overline{x_j}}\right) \quad (4)$$

where $\overline{x_i}$ is the mean of x_i, $\overline{x_j}$ is the mean of x_j, $\mathrm{var}(x_i)$ is the variance of x_i, $\mathrm{cov}(x_i,x_j)$ is the covariance between x_i and x_j, and n is the sample size. For each ratio, we thus obtained 20 ratio values:

$$Q(x_i, \quad x_j) = \{q_{ij\ (t=1)}, \quad q_{ij\ (t=2)}, \quad \cdots, \quad q_{ij\ (t=20)}\} \quad (5)$$

Because we were looking for ratio values varying over the 20 age groups, we retained for further analysis only those ratios q_{ij} for which the relative variation between $q_{ij}(t = 1)$ and $q_{ij}(t = 20)$ was larger than 10%.

Estimation of the association with average group age of the correlation and ratios among molecular features

The association $rc_A(x_i, x_j)$ of the correlation of each pair of molecular features x_i, x_j with the average group age t_M was estimated by taking the Winsorized correlation between the vectors of correlations $\mathbf{C}(x_i, x_j)$ defined in eqn (3) and the average group age vector t_M (respectively, t_F):

$$rc_A(x_i, \quad x_j) = \mathrm{corr}(C(x_i, \quad x_j), \ t_M) \quad (6)$$

The association $rq_A(x_i, x_j)$ of the ratio of each pair of molecular features x_i, x_j with the average group age t_M was estimated in a similar fashion:

$$rq_A(x_i, \quad x_j) = \mathrm{corr}(Q(x_i, \quad x_j), \ t_M) \quad (7)$$

The association $ra_A(x_i, x_j)$ of the mean abundance of each molecular features x_i with the average group age t_M was estimated as:

$$ra_A(x_i) = \mathrm{corr}(A(x_i), \ t_M) \quad (8)$$

We considered to be associated with age only the correlations, ratios, or mean abundances of those molecular features for which $|rc_A(x_i, x_j)| \geq 0.65$, $|rq_A(x_i, x_j)| \geq 0.65$ and $|ra_A(x_i)| \geq 0.65$ and $p < .01$ after correction (*fdr*) for multiple testing with the Benjamini–Hochberg method. Correction for multiple testing (Benjamini–Hochberg) was applied at all analysis stages. This choice is based on both statistical and biological considerations. There are 20 age groups, which means that the sample size available to estimate the correlation between metabolite concentrations and associations (correlations and ratios) is 20: With 20 observations, it is possible to assess significance at $\alpha = 0.01$ with 80% power only of correlations $|r| \geq 0.65$. In addition, there are ~20 participants per age group, thus metabolite–metabolite correlation $|r| \geq 0.65$ can be estimated. Biologically the 0.65 threshold is justified by considering that the majority of correlations observed in metabolomics studies are below 0.6 (22,23): Setting a higher threshold allows to focus on correlations that really stand out of the background correlation.

Validation of the results

To validate the results of the analysis described in the previous sections, that is, the existence of an association between average

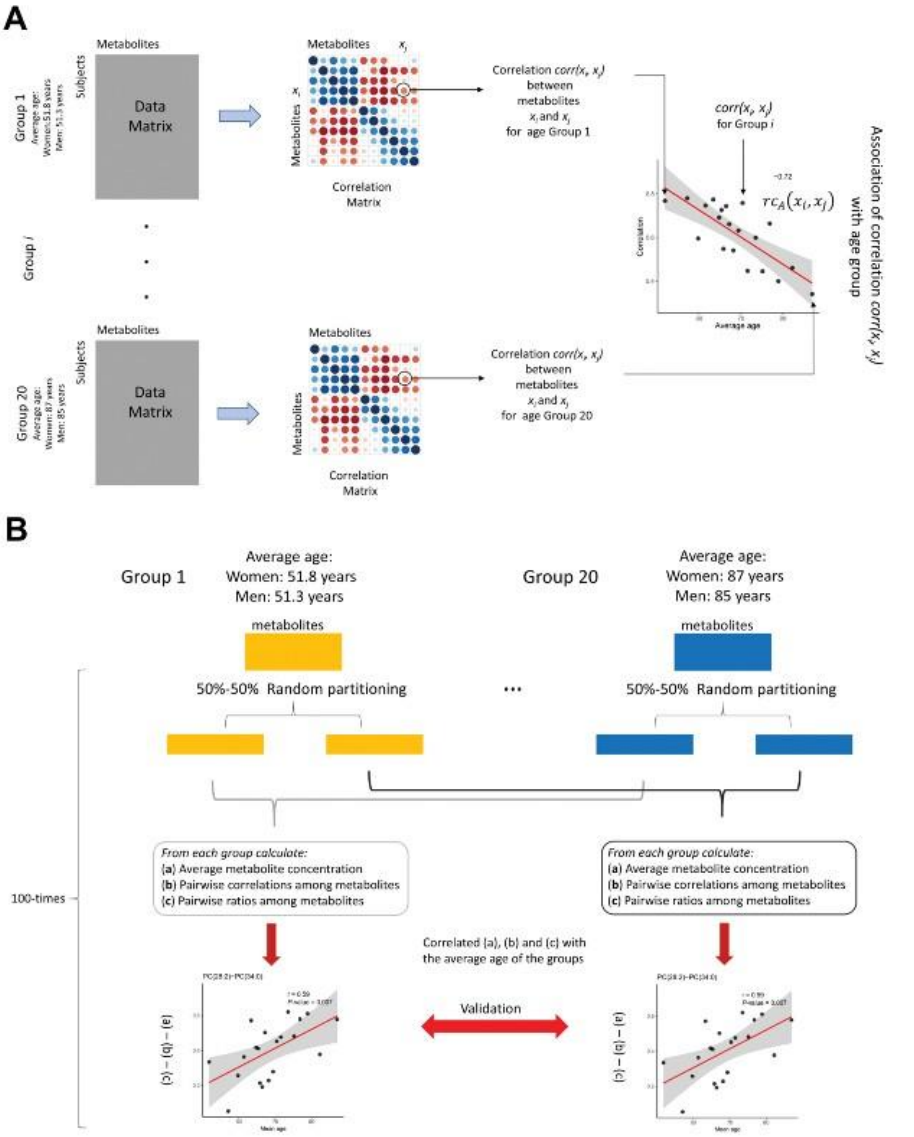


Figure 2. (**A**) Overview of the statistical procedure used to establish the association $rc_A(x_i, x_j)$ (eqn (6)) between the metabolite and lipids pairwise correlations $C(x_i, x_j)$ (eqn (3)), with the group age (Figure 1). In the case of ratios (eqn (7)), the correlation matrix is replaced with the matrix of pairwise ratios, for average abundances is replaced by the vectors of means (eqn (8)). (**B**). Overview of the data-splitting procedure used to validate the results. Each subject group is randomly split into 2 halves, obtaining 2 sets of 20 groups. The analysis is performed on the first set, while the second set is used for validation. The procedure is repeated 100 times: Only results validated more than 50% of the times are considered significant.

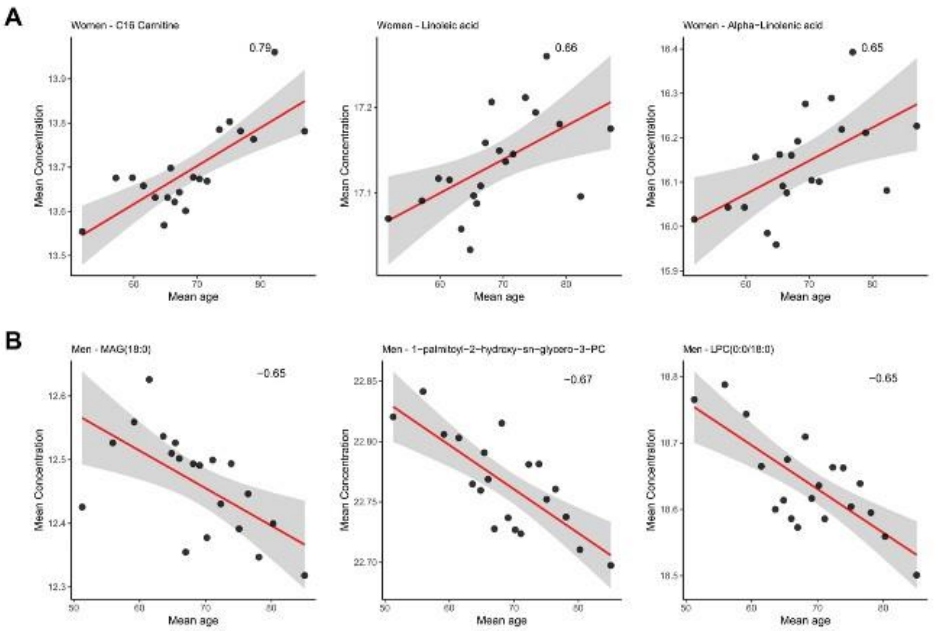


Figure 3. Correlations between average metabolites and lipids concentrations and the average age of the 20 subject groups: women (**A**) and men (**B**). LPC = lysophosphatidylcholine; PC = phosphatidylcholine; PE = phosphatidylethanolamine; MAG = monoacylglycerol. See Figure 2 for an overview of the statistical procedure.

age group and metabolite and lipid concentration (eqn (8)), correlations (eqn (6)), and ratios (eqn (7)), we implemented a data splitting approach (24,25). Basically, we randomly split each of the 20 age groups into 2 halves and performed the analysis independently on the 2 data splits to ascertain if the results could be reproduced. To consider the variability due to the random splitting, the overall procedure was repeated generating $k = 100$ different pairs of data splits: Analysis was repeated on the 100 pairs of data. We considered to be valid only those results that were confirmed in at least 50% of the splits (Figure 2B). In this way, we could obtain an estimation of the reproducibility and robustness of the results by mimicking validation in an external cohort: A portion of the data is used to suggest a hypothesis, and a second, independent portion is used to test it. Note that this approach can be rephrased in an inferential setting and implies that Type I error (ie, the risk of false positives) is controlled (conservatively) at the 0.01 level (26) after correction for multiple testing. The downside of such an approach is a potential loss of power, due to the reduction of the sample size used to estimate correlation. However, this approach is effective in giving valid inference after the selection of a hypothesis, estimating nuisance parameters, and avoiding overfitting (26).

Software

All calculations and plots were performed in R (version 3.3.2). The function "win.cor," implemented in WRS2 package, was used to calculate the Winsorized correlations.

Results

Association of Metabolite and Lipid Abundances With Age

Starting from a total of $n = 202$ metabolites and lipids, a total of $p_W = 3$ (women) and $p_M = 3$ (men) compounds were found statistically significant (adjusted $p \leq .01$ and absolute value of $ra_A \geq 0.65$, see eqn (8)) in more than 50% of splits obtained performing the validation method.

In particular, in the women cohort, we observed a positive correlation of the concentrations of carnitine with $ra_A = 0.79$ and an adjusted $p = .0009$ in the 79% of validation splits, linoleic acid with $ra_A = 0.66$ and an adjusted $p = .001$ in the 59% of validation splits, and α-linoleic acid with $ra_A = 0.65$ and an adjusted $p = .01$ in the 66% of validation splits with the average age of women group (Figure 3A).

The age groups that we used here are data-driven and are not physiologically informed. In particular, the first group of women (W1) corresponds to a 6-year age bin that likely represents perimenopausal women, given that the average age of menopause in women in the Western world is 51 years (27). Although this does not affect statistical analysis, we shall consider that menopausal transition aligns with age.

In the men cohort, we observed negative correlation with age of monoacylglycerol (MAG), especially MAG (18:0) with $ra_A = -0.65$ and an adjusted $p = .005$ in the 53% of validation splits, and lysophosphatidylcholines (LPCs), especially

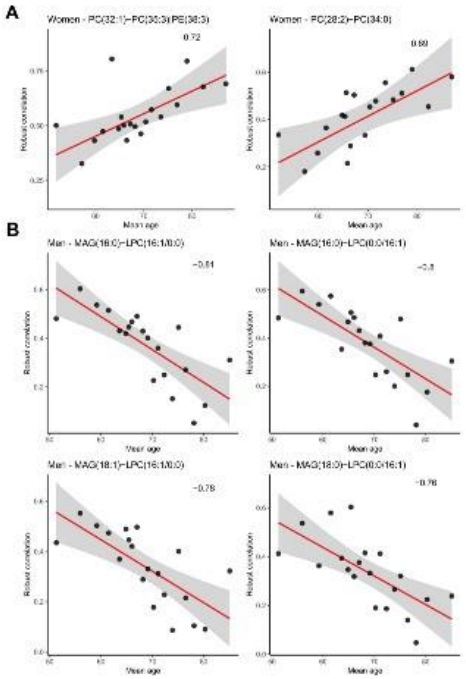


Figure 4. Correlations between metabolites and lipids correlations and the average age of the 20 subject groups: women (**A**) and men (**B**). LPC = lysophosphatidylcholine; PC = phosphatidylcholine; PE = phosphatidylethanolamine; MAG = monoacylglycerol. See Figure 2 for an overview of the statistical procedure.

1-palmitoyl-2-hydroxy-*sn*-glycero-3-phosphocholine (LPC [16:0]) with $ra_A = -0.67$ and an adjusted $p = .005$ in the 58% of validation splits, and LPC (0:0/18:0) with $ra_A = -0.65$ and an adjusted $p = .008$ in the 62% of validation splits (Figure 3B).

For a complete overview of the results for all metabolites see Supplementary Table 3.

Association of the Correlation Among Molecular Features With Age

Starting from a total of $n = 20\,301$ metabolites and lipids correlations, a total of $c_W = 2$ (women) and $c_M = 4$ (men) correlations among molecules result to be statistically significant (adjusted $p \leq .01$ and absolute value of $rc_A \geq 0.65$, see eqn (6)) in more than 50% of splits after the validation method.

In the women cohort, the correlations between phosphocholines (PCs), especially between (a) PC (28:2)–PC (32:1) with $rc_A = 0.69$ and an adjusted $p = .008$ in the 57% of validation splits and (b) PC (32:1)–PC (35:3) | PE (38:3) with $rc_A = 0.72$ and an adjusted $p = .008$ in the 54% of validation splits, tend to increase with age (Figure 4A).

In Figure 4B, correlations between MAG and LPC, especially (a) MAG (16:0)–LPC (16:1/0:0) with $rc_A = -0.81$ and an adjusted $p = .002$ in the 53% of validation splits, (b) MAG (16:0)–LPC (0:0/16:1) with $rc_A = -0.80$ and an adjusted $p = .002$ in the 59% of validation splits, (c) MAG (18:1)–LPC (16:1/0:0) with $rc_A = -0.78$ and an adjusted $p = .003$ in the 60% of validation splits, and (d) MAG (18:0)–LPC (0:0/16:1) with $rc_A = -0.76$ and an adjusted $p = .004$ in the 53% of validation splits, decrease with the average age of men-specific groups. The levels of these lipids vary in a similar fashion, decreasing with the age.

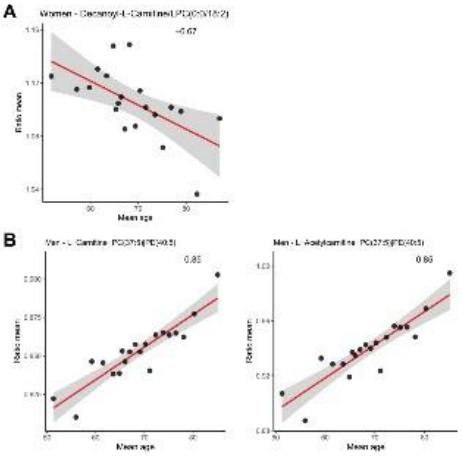


Figure 5. Correlations between average metabolites and lipids ratios and the average age of the 20 subject groups: women (**A**) and men (**B**). GCA = glycocholic acid; PC = phosphatidylcholine; PE = phosphatidylethanolamine; MAG = monoacylglycerol. See Figure 2 for an overview of the statistical procedure.

For a complete overview of the results for all metabolites see Supplementary Table 3.

Association of the Ratios Among Molecular Features With Age

Alterations in the ratios between 2 single lipids and/or metabolites may point at perturbations in pathways relevant for a certain specific phenotype and they could influence the physiological course of aging. In this light, pairwise ratios may serve as potential biomarkers of the aging process (28,29).

Starting from a total of $n = 20\,301$ metabolites and lipids ratios, after the validation method, we found only $q_W = 1$ (women) and $q_M = 2$ (men) ratios between molecules whose variation is significantly associated with the average age (adjusted $p \leq 0.01$ and absolute value of $rq_A \geq 0.65$, see eqn (7)). In particular, the ratio between decanoyl-L-carnitine/LPC (0:0/18:2) with $rq_A = -0.67$ and an adjusted $p = .002$ in the 56% of validation splits shows a negative association with the average age of the women cohort (Figure 5A).

In Figure 5B, the ratios between L-carnitine/PC (37:5) with $rq_A = 0.85$ and an adjusted $p = 1 \times 10^{-4}$ in the 55% of validation splits and L-acetylcarnitine/PC (37:5) with $rq_A = 0.85$ and an adjusted $p = 2 \times 10^{-4}$ in the 51% of validation splits tend to be positively correlated with the average age of men-specific groups.

For a complete overview of the results for all metabolites see Supplementary Table 3.

Discussion

To shed light on the molecular mechanisms possibly associated with age, we studied how the concentration, correlations, and ratios of and among circulating blood metabolites and lipids vary with subject age groups, considering men and women separately to highlight possible dependencies on sex. Using different approaches, in the original paper, Ganna et al. (17) demonstrated that LPC (18:1) and LPC (18:2) are not directly associated with coronary heart disease (CHD), but they found an age-dependent negative trend of these 2 lipids in association with CHD risk. Moreover, MAG (18:2) and sphingomyelin (28:1) have a positive correlation with the CHD risk. Our results support the usefulness of the metabolomic analysis conjugated with a system biology approach for the identification of age-related metabolites and their association patterns, providing additional information compared to what is already known from the literature.

In women, the levels of carnitine, linoleic acid, and α-linoleic acid show a positive correlation with (group) age. These significant correlations are of particular interest because previous studies showed that the age-dependent carnitine serum levels increase more with age in adult women than men (30,31), and the endogenous biosynthesis of carnitine depends on the production, by lysosomal protein degradation, of trimethyl-lysine (32) whose homeostasis is regulated by dietary intake, intestinal absorption, and renal reabsorption. Carnitine also plays an important role in carnitine-shuttle biochemical reactions and in the energy pool metabolism, inducing an expression of intramitochondrial alterations (30,33), fundamental in linoleic acid metabolism. Previous studies report that the reduction of estrogens activity and the increase of testosterone levels induce modification of the rate of conversion of linoleic acid and α-linolenic acid into $n - 3$ long-chain polyunsaturated fatty acids, inducing changes in cell membrane composition and in cell cycle mechanisms (34,35). Endogenous biosynthesis of carnitine depends on the production, by lysosomal protein degradation, of trimethyllysine (32). The homeostasis of this molecule is regulated by dietary intake, intestinal absorption, and renal reabsorption. Carnitine also plays an important role in carnitine-shuttle biochemical reactions and in the energy pool metabolism, inducing an expression of intramitochondrial alterations (30,33), fundamental in linoleic acid metabolisms, whose activity shows age-dependent dysregulation (36). In addition to the role of polyunsaturated fatty acids as energy sources, they have several functions, as cellular signaling pathways (37) and as structural components of cell membranes (38), inducing age-dependent changes (39).

The negative correlation of LPCs concentrations with age, molecularly associated with the reduction of MAGs levels by the MAG lipase enzyme activity (36,40), induces a skeletal muscle mitochondrial dysfunction (41); the decreasing of LPCs is, generally, also associated with the increase of body mass index (BMI) but, in an older population, this effect is associated, firstly, with the increasing of age-dependent inflammation, depending on an overall remodulation of cell membrane and mitochondrial dysfunction.

Because the pairwise correlations among molecules can be used as a proxy to describe the underlying metabolic network (10), here we consider the correlations observed as the result of the combination of all reactions and regulatory processes occurring in the metabolic network (18,42) at a given age.

In women, the correlations between PC (28:2)–PC (32:1) and PC (32:1)–PC(35:3) | PE(38:3) tend to increase with age. During the menopause period, a global dysregulation on liver enzymes is induced, causing the synthesis of PCs from choline (43,44). The interactions of PCs are associated with the remodulation of membranes integrity, promoting their conservation and directly affecting the membrane permeability, increasing the fluidity of the bilayer and protecting it from peroxidative damage (38,45), a frequent phenomenon in advanced age (46). Correlations between MAG (16:0)–LPC (16:1/0:0), MAG (16:0)–LPC (0:0/16:1), MAG (18:1)–LPC (16:1/0:0), and MAG (18:0)–LPC (0:0/16:1) decrease with age in men, and this has been related to the increase of the MAG lipase enzyme activity that determines the hydrolysis of MAG into glycerol and fatty acid alkyl ester (36,40) and to the impaired mitochondrial oxidative capacity associated with low levels of LPCs in advanced age (36,41).

The alterations in the ratios between 2 single lipids and/or metabolites may point at perturbations in pathways relevant for a certain specific phenotype. We considered the pairwise ratios as potential biomarkers (28,29) of the aging process. We found that only the ratio between decanoyl-L-carnitine/LPC(0:0/18:2) shows a negative association with the average age in women, and, at best of our knowledge, this association has never been reported. We can speculate that decreasing levels of LPCs and the increasing levels of decanoyl-L-carnitine induce, synergistically, a mitochondrial dysfunction (36,41), contributing to age-dependent metabolic changes and being an indirect result of aging (47). In contrast, the ratios between L-carnitine/PC(37:5) and L-acetylcarnitine/PC(37:5) tend to be positively correlated with the average age of men-specific groups.

Little is known about these molecular ratios. As said before, carnitine plays a role in carnitine-shuttle biochemical reactions: carnitine palmitoyltransferase 1 enzyme is involved in the reversible acylation of L-carnitine, producing L-acetylcarnitine, and this event is fundamental in fatty acid beta-oxidation, maintenance of acyl-coenzyme A pools, and energy metabolism (30). The carnitine-shuttle activity could generate a specific remodeling of mitochondrial fatty acids oxidation, promoting a modification in the mitochondrial membrane lipidome (48), increasing PCs fraction (49,50). Although, actually, the overall aging molecular mechanisms are unclear, our results show that lipids (ie, LPC, MAG, PC, PE, linoleic acid) and carnitine are fundamental in the age-related metabolic pathways.

Strengths and Limitations

One of the strengths of this study is a large number of patients with a very wide age range (47.6–93.9 years) whose metabolome was analyzed. We implemented a stringent validation of the results using a repeated data resampling to account for varaibility and to obtain robust estimate of metabolite concentrations, correlations, and ratios calculated at the age group level to eliminate subject-to-subject variability.

One limitation of this study is the lack of availability of the clinical data (ie, BMI, waist circumference, systolic and diastolic blood pressures) associated with the participants' metabolite data, publicly available on the MetaboLights public database, resulting in an incomplete representation of the pathophysiological conditions of the cohort, indicating that we could not correct at the individual level for such factors in the analysis.

Conclusions

In this study, we presented a comprehensive biology approach to highlight potential molecular features concentrations, associations, and ratios directly associated with the increase of the age of a sex- and age-matched population. We showed that linoleic acid, α-linoleic acid, and carnitine have, in the women cohort, a positive correlation trend with age, while MAGs and LPCs have, in the men cohort, a negative correlation trend with age. These results highlight, in women, the effect of the reduction of estrogens activity and the increase of testosterone levels on the linoleic acid metabolism and on the energy pool metabolism that induces the overall changes in cell membrane composition and cell cycle mechanisms. In men, low levels of LPCs concentrations are directly connected with the reduction of MAGs levels by the MAG lipase enzymatic activity that induces mitochondrial dysfunction.

Analyzing the pairwise correlations among molecules, we observed that PCs/PCs correlations tend to have a positive trend associated with the average ages of women, while MAGs/LPCs correlations tend to have a negative trend associated with men average ages. These results, in both cases, suggest an age-dependent remodeling of fatty acid metabolism that induces, overall, remodeling of cell and mitochondrial membranes and modification in terms of fluidity of membranes bilayers.

We studied the pairwise ratios as potential biomarkers of aging. In women, the decanoyl-L-carnitine/LPC ratio has a negative association with the increasing of the average ages, while in men the ratios between L-carnitine/PC and L-acetylcarnitine/PC have a positive association with the increase of age, suggesting, in both cases, a radical remodeling of the dynamic membrane fluidity and carnitine-shuttle activity.

This study brings forward the concept that correlation and ratios among molecular features, and not only abundances along, could be used to investigate the dynamic of molecular mechanisms and their association with age.

Supplementary Material

Supplementary data are available at *The Journals of Gerontology, Series A: Biological Sciences and Medical Sciences* online.

Funding

This research received no external funding.

Conflict of Interest

None declared.

Acknowledgments

The authors acknowledge the support and the use of resources of Instruct-ERIC, a Landmark ESFRI project, and specifically the CERM/CIRMMP Italy Centre.

References

1. Karasik D, Demissie S, Cupples LA, Kiel DP. Disentangling the genetic determinants of human aging: biological age as an alternative to the use of survival measures. *J Gerontol A Biol Sci Med Sci*. 2005;60(5):574–587. doi:10.1093/gerona/60.5.574
2. Kerber RA, O'Brien E, Cawthon RM. Gene expression profiles associated with aging and mortality in humans. *Aging Cell*. 2009;8(3):239–250. doi:10.1111/j.1474-9726.2009.00467.x
3. Hoffman JM, Lyu Y, Pletcher SD, Promislow DEL. Proteomics and metabolomics in ageing research: from biomarkers to systems biology. *Essays Biochem*. 2017;61(3):379–388. doi:10.1042/EBC20160083
4. Jové M, Maté I, Naudi A, et al. Human aging is a metabolome-related matter of gender. *J Gerontol A Biol Sci Med Sci*. 2016;71(5):578–585. doi:10.1093/gerona/glv074
5. Vignoli A, Ghini V, Meoni G, et al. High-throughput metabolomics by 1D NMR. *Angew Chem Int Ed Engl*. 2019;58(4):968–994. doi:10.1002/anie.201804736
6. Eckhart AD, Beebe K, Milburn M. Metabolomics as a key integrator for "omic" advancement of personalized medicine and future therapies. *Clin Transl Sci*. 2012;5(3):285–288. doi:10.1111/j.1752-8062.2011.00388.x
7. 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. doi:10.1021/acs.jproteome.7b00404
8. Vignoli A, Tenori L, Giusti B, et al. 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. doi:10.1186/s12916-018-1240-2
9. Vignoli A, Paciotti S, Tenori L, et al. Fingerprinting Alzheimer's disease by 1H nuclear magnetic resonance spectroscopy of cerebrospinal fluid. *J Proteome Res*. 2020;19(4):1696–1705. doi:10.1021/acs.jproteome.9b00850
10. Rosato A, Tenori L, Cascante M, De Atauri Carulla PR, Martins Dos Santos VAP, Saccenti E. From correlation to causation: analysis of metabolomics data using systems biology approaches. *Metabolomics*. 2018;14(4):37. doi:10.1007/s11306-018-1335-y
11. Yu Z, Zhai G, Singmann P, et al. Human serum metabolic profiles are age dependent. *Aging Cell*. 2012;11(6):960–967. doi:10.1111/j.1474-9726.2012.00865.x
12. Darst BF, Koscik RL, Hogan KJ, Johnson SC, Engelman CD. Longitudinal plasma metabolomics of aging and sex. *Aging (Albany NY)*. 2019;11(4):1262–1282. doi:10.18632/aging.101837
13. Mittelstrass K, Ried JS, Yu Z, et al. Discovery of sexual dimorphisms in metabolic and genetic biomarkers. *PLoS Genet*. 2011;7(8):e1002215. doi:10.1371/journal.pgen.1002215
14. Petersen AK, Krumsiek J, Wägele B, et al. On the hypothesis-free testing of metabolite ratios in genome-wide and metabolome-wide association studies. *BMC Bioinformatics*. 2012;13(1):120. doi:10.1186/1471-2105-13-120
15. Altmaier E, Ramsay SL, Graber A, Mewes HW, Weinberger KM, Suhre K. Bioinformatics analysis of targeted metabolomics—uncovering old and new tales of diabetic mice under medication. *Endocrinology*. 2008;149(7):3478–3489. doi:10.1210/en.2007-1747
16. Barbieri M, Boccardi V, Papa M, Paolisso G. Metabolic journey to healthy longevity. *Horm Res*. 2009;71(suppl 1):24–27. doi:10.1159/000178032
17. Ganna A, Salihovic S, Sundström J, et al. Large-scale metabolomic profiling identifies novel biomarkers for incident coronary heart disease. *PLoS Genet*. 2014;10(12):e1004801. doi:10.1371/journal.pgen.1004801
18. Saccenti E. Correlation patterns in experimental data are affected by normalization procedures: consequences for data analysis and network inference. *J Proteome Res*. 2017;16(2):619–634. doi:10.1021/acs.jproteome.6b00704
19. Tuğran E, Kocak M, Mirtagioğlu H, Yiğit S, Mendes M. A simulation based comparison of correlation coefficients with regard to type I error rate and power. *J Data Anal Inf Process*. 2015;3(3):87–101. doi:10.4236/jdaip.2015.33010
20. Wilcox RR. Some results on a Winsorized correlation coefficient. *Br J Math Stat Psychol*. 1993;46(2):339–349. doi:10.1111/j.2044-8317.1993.tb01020.x
21. van Kempen GM, van Vliet LJ. Mean and variance of ratio estimators used in fluorescence ratio imaging. *Cytometry*. 2000;39(4):300–305. doi:10.1002/(sici)1097-0320(20000401)39:4<300::aid-cyto8>3.0.co;2-o

22. Camacho D, Fuente ADL, Mendes P. The origins of correlations in metabolomics data. *Metabolomics*. 2005;1(1):53–63. doi:10.1007/s11306-005-1107-3
23. Jahagirdar S, Saccenti E. On the use of correlation and MI as a measure of metabolite–metabolite association for network differential connectivity analysis. *Metabolites*. 2020;10(4):171. doi:10.3390/metabo10040171
24. Cox DR. A note on data-splitting for the evaluation of significance levels. *Biometrika*. 1975;62(2):441–444. doi:10.1093/biomet/62.2.441
25. Rubin D, Dudoit S, van der Laan M. A method to increase the power of multiple testing procedures through sample splitting. *Stat Appl Genet Mol Biol*. 2006;5:Article19. doi:10.2202/1544-6115.1148
26. DiCiccio CJ, DiCiccio TJ, Romano JP. Exact tests via multiple data splitting. *Stat Probab Lett*. 2020;166:108865. doi:10.1016/j.spl.2020.108865
27. Lindh-Åstrand L, Hoffmann M, Järvstråt L, Fredriksson M, Hammar M, Spetz Holm AC. Hormone therapy might be underutilized in women with early menopause. *Hum Reprod*. 2015;30(4):848–852. doi:10.1093/humrep/dev017
28. Zelezniak A, Sheridan S, Patil KR. Contribution of network connectivity in determining the relationship between gene expression and metabolite concentration changes. *PLoS Comput Biol*. 2014;10(4):e1003572. doi:10.1371/journal.pcbi.1003572
29. Krumsiek J, Stückler F, Suhre K, et al. Network-based metabolite ratios for an improved functional characterization of genome-wide association study results. *bioRxiv*, doi:10.1101/048512, 13 April 2016, preprint: not peer reviewed.
30. Mitchell SL, Uppal K, Williamson SM, et al. The carnitine shuttle pathway is altered in patients with neovascular age-related macular degeneration. *Invest Ophthalmol Vis Sci*. 2018;59(12):4978–4985. doi:10.1167/iovs.18-25137
31. Malaguarnera G, Catania VE, Bonfiglio C, Bertino G, Vicari E, Malaguarnera M. Carnitine serum levels in frail older subjects. *Nutrients*. 2020;12(12):3887. doi:10.3390/nu12123887
32. Koves TR, Ussher JR, Noland RC, et al. Mitochondrial overload and incomplete fatty acid oxidation contribute to skeletal muscle insulin resistance. *Cell Metab*. 2008;7(1):45–56. doi:10.1016/j.cmet.2007.10.013
33. Judit B, Andras S, Katalin K, Bela M. Mass spectrometric analysis of L-carnitine and its esters: potential biomarkers of disturbances in carnitine homeostasis. *Curr Mol Med*. 2020;20(5):336–354. doi:10.2174/1566524019666191113120828
34. Janssen CI, Kiliaan AJ. Long-chain polyunsaturated fatty acids (LCPUFA) from genesis to senescence: the influence of LCPUFA on neural development, aging, and neurodegeneration. *Prog Lipid Res*. 2014;53:1–17. doi:10.1016/j.plipres.2013.10.002
35. Cybulska AM, Skonieczna-Żydecka K, Drozd A, et al. Fatty acid profile of postmenopausal women receiving, and not receiving, hormone replacement therapy. *Int J Environ Res Public Health*. 2019;16(21):4273. doi:10.3390/ijerph16214273
36. Johnson AA, Stolzing A. The role of lipid metabolism in aging, lifespan regulation, and age-related disease. *Aging Cell*. 2019;18(6):e13048. doi:10.1111/acel.13048
37. Sokoła-Wysoczańska E, Wysoczański T, Wagner J, et al. Polyunsaturated fatty acids and their potential therapeutic role in cardiovascular system disorders—a review. *Nutrients*. 2018;10(10):1561. doi:10.3390/nu10101561
38. Li Z, Agellon LB, Allen TM, et al. The ratio of phosphatidylcholine to phosphatidylethanolamine influences membrane integrity and steatohepatitis. *Cell Metab*. 2006;3(5):321–331. doi:10.1016/j.cmet.2006.03.007
39. Chung KW. Advances in understanding of the role of lipid metabolism in aging. *Cells*. 2021;10(4):880. doi:10.3390/cells10040880
40. Grabner GF, Zimmermann R, Schicho R, Taschler U. Monoglyceride lipase as a drug target: at the crossroads of arachidonic acid metabolism and endocannabinoid signaling. *Pharmacol Ther*. 2017;175:35–46. doi:10.1016/j.pharmthera.2017.02.033
41. Semba RD, Zhang P, Adelnia F, et al. Low plasma lysophosphatidylcholines are associated with impaired mitochondrial oxidative capacity in adults in the Baltimore Longitudinal Study of Aging. *Aging Cell*. 2019;18(2):e12915. doi:10.1111/acel.12915
42. Steuer R, Kurths J, Fiehn O, Weckwerth W. Observing and interpreting correlations in metabolomic networks. *Bioinformatics*. 2003;19(8):1019–1026. doi:10.1093/bioinformatics/btg120
43. Auro K, Joensuu A, Fischer K, et al. A metabolic view on menopause and ageing. *Nat Commun*. 2014;5:4708. doi:10.1038/ncomms5708
44. Cui X, Yu X, Sun G, et al. Differential metabolomics networks analysis of menopausal status. *PLoS One*. 2019;14(9):e0222353. doi:10.1371/journal.pone.0222353
45. Rabini RA, Moretti N, Staffolani R, et al. Reduced susceptibility to peroxidation of erythrocyte plasma membranes from centenarians. *Exp Gerontol*. 2002;37(5):657–663. doi:10.1016/s0531-5565(02)00006-2
46. Akila VP, Harishchandra H, D'souza V, D'souza B. Age related changes in lipid peroxidation and antioxidants in elderly people. *Indian J Clin Biochem*. 2007;22(1):131–134. doi:10.1007/BF02912896
47. Haas RH. Mitochondrial dysfunction in aging and diseases of aging. *Biology*. 2019;8(2):48. doi:10.3390/biology8020048
48. Lum H, Sloane R, Huffman KM, et al. Plasma acylcarnitines are associated with physical performance in elderly men. *J Gerontol A Biol Sci Med Sci*. 2011;66(5):548–553. doi:10.1093/gerona/glr006
49. Burstein MT, Titorenko VI. A mitochondrially targeted compound delays aging in yeast through a mechanism linking mitochondrial membrane lipid metabolism to mitochondrial redox biology. *Redox Biol*. 2014;2:305–307. doi:10.1016/j.redox.2014.01.011
50. Janikiewicz J, Szymański J, Malinska D, et al. Mitochondria-associated membranes in aging and senescence: structure, function, and dynamics. *Cell Death Dis*. 2018;9(3):332. doi:10.1038/s41419-017-0105-5

Supplementary Tables

Molecular features	**Compound CHEBI codes**
Choline	15354
Creatinine	16737
e-Caprolactam	28579
DL-2-Aminooctanoic acid	75145
Γ-Caprolactone	85235
L-Proline	17203
Betaine	17750
Salicylic acid	16914|15365
Theobromine	28946
Creatine	16919
L-Leucine	15603|18347
Hypoxanthine	17368
3-Pyridylacetic acid|trigonelline	86390|18123
Stachydrine	35280
Benzenebutanoic acid	41500
Indole-3-carbinol	24814
Acetaminophen	46195
2-Ketohexanoic acid	17308
S-(4,5-Dihydro-2-methyl-3-furanyl) ethanethioate	131456
L-Carnitine	16347
L-Phenylalanine	17295
Uric acid	17775
Indoleacetic acid	16411
Paraxanthine|Theophylline	25858|28177
L-Tyrosine|o-Tyrosine	17895|89461
Naproxen	7476
L-Tryptophan	16828
3-Indolepropionic acid	43580
1,3,7-Trimethyluric acid	691622
Caffeine	27732
Hippuric acid	18089
L-Acetylcarnitine	57589
Indolelactic acid	24813
Pantothenic acid	7916
Flavone	42491
Butyryl-L-carnitine|Isobutyryl-L-carnitine	21949|84838

Geranyl acetoacetate	85255
Dodecanedioic acid	4676
5a-Androst-3-en-17-one	86393
Propranolol	8499
γ-Glutamyl-leucine	68433
Myristic acid	28875
A-Linolenic acid	27432
L-Aspartyl-L-phenylalanine	73830
Pentadecanoic acid	42504
D-erythro-sphingosine	46964
Oleamide	116314
Piperine	28821
L-Octanoylcarnitine	18102
Palmitoleic acid	28716
2,6 dimethylheptanoyl carnitine	84095
4-Androsten-11Beta-ol-3,17-dione	27967|27967
Arachidonic acid	15843
Heptadecanoic aicd	32365
Phenylalanylphenylalanine	72723
Decanoyl-L-carnitine	28717
Fatty acid C20:5 methyl ester	91031
Linoleic acid	17351
Arachidonic acid methyl ester	78033
MAG(14:0)	87249
10-nitro-9E-octadecenoic acid|9-nitro-9E-octadecenoic acid	86285|86329
Fatty acid C22:6	36005
4,8 dimethylnonanoyl carnitine	63874
Arachidonic acid ethyl ester	84873
MAG(18:0)	87255
C12 Carnitine	73054
Corticosterone	16827
MAG(16:1)	87253
MAG(16:0)	87251
MAG(18:3)	87258
Treprostinil	50861
Prostaglandin J2	27485
Hyodeoxycholic acid	52023
Deoxycholic acid	28834
Cortisol	17650
6-hydroxy-5-cholestanol|cholesterol	|16113
cis-5-Tetradecenoylcarnitine	73060
Cholic acid	16359

MAG(18:2)	87257
MAG(18:1)	87256
Stearic acid	28842
MAG(20:5)	86397
C16 Carnitine	73067
7-Ketocholesterol	64294
17-phenyl trinor Prostaglandin E2|17-phenyl trinor Prostaglandin D2	87820|87821
sodium glycochenodeoxycholate	87818
Deoxycholic acid glycine conjugate	27471
Chenodeoxycholic acid	16755
Linoleyl carnitine	73072
Oleoyl-L-carnitine hydrochloride	91318
A-Tocopherol	22470
3a,6b,7b-Trihydroxy-5b-cholanoic acid	81298
Barogenin	86509
Γ-Tocopherol	18185
Dodecanoic acid	30805
dehydroepiandrosterone sulfate	91028
LPE(16:0)	90452
1-palmitoyl-2-hydroxy-sn-glycero-3-PE	73134
GCA	17687
1-O-1'-(Z)-octadecenyl-2-hydroxy-sn-glycero-3-PE	87823
LPE(18:2)	91296
PC(15:2/0:0)|PE(18:2/0:0)	131590|91296
LPE(18:1)	64575
1-oleoyl-2-hydroxy-sn-glycero-3-PE	75168
LPE(18:0)	64576
1-Stearoyl-2-Hydroxy-sn-Glycero-3-PE	83047
2a-(3-Hydroxypropyl)-1a,25-dihydroxy-19-norvitamin D3	91006
PC(37:5)|PE(40:5)	85767|71745
LPC(0:0/16:1)	91298
LPC(16:1/0:0)	91305
LPC(0:0/16:0)	91297
1-palmitoyl-2-hydroxy-sn-glycero-3-PC	72998
LPE(20:4)	64569
PC(O-18:1/0:0)|PC(P-18:0/0:0)	64591|88779
LPC(18e:0/0:0)	75216
Lyso-PAF C-18	91144
PS(18:0)	131443

LPC(18:3)	64565
LPC(0:0/18:2)	91302
LPC(18:2/0:0)	91309
LPC(18:1)	64566
1-oleoyl-2-hydroxy-sn-glycero-3-PC	28610
LPC(0:0/18:0)	91299
1-stearoyl-2-hydroxy-sn-glycero-3-PC	73858
LPC(0:0/20:4)	91303
LPC(20:4/0:0)	91310
LPC(20:3)	64481
LPC(20:4)	64481
LPC(20:2)	67056
LPC(20:1)	67057
3-cis-Hydroxy-b,e-Caroten-3'-one	HMDB02890
Palmitic acid	15756
1-arachidoyl-2-hydroxy-sn-glycero-3-PC	74968
LPC(0:0/20:5)	91304
LPC(20:5/0:0)	91311
LPC(22:7)	74349
LPC(22:6)	74349
LPC(22:5)	74349
LPC(22:4)	91312
Biliverdin hydrochloride b	91027
Biliverdin hydrochloride a	91027
Biliribin I	16990
Biliribin II	16990
D-Urobilinogen|I-Urobilin	4260|36378
cis/trans-Oleic acid	36021
1-linoleoyl-2-stearoyl-sn-glycerol	86337
1-vaccenoyl-2-palmitoyl-sn-glycerol	86346
Ceramide PE(33:1)|SPM(30:1)	86515|72505
Ceramide PE(34:1)	86517
Ceramide PE(35:2)|SPM(32:2)	86523|72510
PC(28:2)	65292
Ceramide PE(35:1)|SPM(32:1)	86519|64586
PC(28:1)	65293
Ceramide PE(36:2)	86525
PC(29:1)	131438
Ornithine	18257
Ceramide PE(37:2)|SPM(34:2)	86527|64587
PC(30:2)	65301
N-palmitoyl-D-erythro-sphingosylphosphorylcholine	78646
PC(30:1)	65302

Ceramide PE(38:2)	86968
SPM(d18:2/18:1)	105799
N-(9Z-octadecenoyl)-sphing-4-enine-1-PC	84487
1,2-dipalmitoleoyl-sn-glycero-3-PC	83717
N-(octadecanoyl)-sphing-4-enine-1-PC	83358
PC(32:1)	66849
1,2-dipalmitoyl-sn-glycero-3-PC	72999
1,2-dilinoleoyl-sn-glycero-3-PC	42027
PC(33:1)|PE(36:1)	86472|71727
PC(34:5)|PE(37:5)	66854|131584
PC(34:4)	64423
PC(34:3)	64424
PC(34:2)|PE(37:2)	64516|131583
1-oleoyl-2-palmitoyl-sn-glycero-3-PC	74667
PC(34:0)	66855
C16-20:5 PC	unknown
PC(35:4)|PE(38:4)	91322|71737
1-stearoyl-2-arachidonoyl-sn-glycero-3-PE|C15-20:4 PC	79110|86344
PC(35:3)|PE(38:3)	131439|71736
PC(35:2)|PE(38:2)	85766|71735
PC(34:3)|PE(37:3)	64424|131581
PC(36:6)	66856
PC(36:5)	64504
1,2-dilinoleoyl-sn-glycero-4-PC	42027
PC(36:3)|PE(39:3)	64523|131585
SPM(40:2)	72529
1,2-dioleoyl-sn-glycero-3-PC|1,2-dipetroselenoyl-sn-glycero-3-PC	74669|86330
PC(36:1)	66857
SPM(41:2)	85762
PC(38:7)	64498
1-palmitoyl-2-docosahexaenoyl-sn-glycero-3-PC	74963
PC(38:5)|PE(41:5)	64525|131586
PC(38:4)	64526
PC(38:3)	64446
PC(38:2)	66859
N-(15Z-tetracosenoyl)-sphinganine-1-PC	91146
PC(40:6)	64431
SPM(42:3)	72535
LacCer(d18:1/14:0)	91034
PC(40:5)	64524

PC(42:7)	131440
Lactosyl ceramide(d18:1/16:0)	84758

Table S1: List of metabolites and lipids tested and CHEBI codes. Abbreviation: LPC = Lysophosphatidylcholine; LPE = lysophosphatidylethenolamine; PC = phosphatidylcholine; PE = phosphatidylethanolamine; MAG = monoacylglycerol; GCA = glycocholic acid; SPM = sphingomyelin. See Figure 2 for an overview of the statistical procedure.

Women (n_1=804)				Men (n_2=1078)			
Group	Age range (min – max) years	Age mean±sd years	n. individuals per group	Group	Age range (min- max) years	Age mean±sd years	n. individuals per group
W_1	48.4-55.1	51.8±2.1	41	M_1	47.6-53.8	51.3±1.5	54
W_2	55.1-58.6	57.1±1.1	40	M_2	53.8-57.8	55.9±1.2	54
W_3	58.6-60.7	59.8±0.6	40	M_3	57.8-60.2	59.2±0.7	54
W_4	60.7-62.5	61.5±0.5	40	M_4	60.2-62.7	61.5±0.7	54
W_5	62.5-64.3	63.3±0.5	40	M_5	62.7-64.4	63.6±0.5	54
W_6	64.3-65.1	64.8±0.2	40	M_6	64.4-65.2	64.8±0.2	55
W_7	65.1-65.5	65.3±0.1	42	M_7	65.2-65.8	65.4±0.2	52
W_8	65.5-66.1	65.8±0.2	39	M_8	65.8-66.4	66.0±0.2	54
W_9	66.1-66.8	66.4±0.2	40	M_9	66.4-67.5	67.0±0.3	54
W_{10}	66.8-67.6	67.2±0.2	40	M_{10}	67.5-68.6	68.1±0.3	54
W_{11}	67.6-68.9	68.2±0.4	40	M_{11}	68.6-69.7	69.1±0.3	54
W_{12}	68.9-69.9	69.4±0.3	40	M_{12}	69.7-70.6	70.2±0.3	54
W_{13}	69.9-70.8	70.4±0.3	40	M_{13}	70.6-71.5	71.1±0.3	54
W_{14}	70.8-72.5	71.6±0.5	41	M_{14}	71.5-73.1	72.3±0.5	53
W_{15}	72.5-74.4	73.5±0.5	40	M_{15}	73.1-74.5	73.9±0.4	54

W_{16}	74.4-76.0	75.1±0.4	40	M_{16}	74.5-75.7	75.1±0.3	54
W_{17}	76.0-77.8	76.9±0.5	40	M_{17}	75.7-77.2	76.5±0.4	54
W_{18}	77.8-80.3	78.9±0.7	40	M_{18}	77.2-79.2	78.1±0.5	54
W_{19}	80.3-84.2	82.2±1.1	40	M_{19}	79.2-81.7	80.3±0.7	54
W_{20}	84.2-93.9	87.0±2.4	41	M_{20}	81.7-93.3	85.0±2.9	54

Table S2: Characteristics of the 20 groups resulting from the stratification of subjects by age using the 20 quantiles of the age distribution of women and men. For each group, the number of subjects per group, age-range, and age mean±standard deviation were reported. See figure 1 for a graphical illustration of the stratification procedure.

Association of abundance of molecular features a with age

	Molecular features	Correlation	*P*-value	*P*-value adjusted	Validation (>50%)
Women (n_W=804)	1. C-16 Carnitine	0.79	$7x10^{-5}$	$9x10^{-4}$	79
	2. Pentadecanoic acid	0.58	$7x10^{-5}$	$9x10^{-4}$	22
	3. Palmitic acid	0.55	$8x10^{-5}$	$9x10^{-4}$	41
	4. Oleoyl-L-carnitine hydrochloride	0.72	$8x10^{-5}$	$9x10^{-4}$	32
	5. Heptadecanoic acid	0.67	$8x10^{-5}$	$9x10^{-4}$	32
	6. Myristic acid	0.67	$9x10^{-5}$	0.001	34
	7. cis/trans-Oleic acid	0.67	$9x10^{-5}$	0.001	42
	8. cis-5-Tetradecenoylcarnitine	0.57	$9x10^{-5}$	0.001	20
	9. Propranolol	0.52	$1x10^{-4}$	0.001	10
	10. Choline	0.51	$1x10^{-4}$	0.001	15
	11. Linoleyl carnitine	0.50	$1x10^{-4}$	0.001	26
	12. Corticosterone	0.58	$2x10^{-4}$	0.003	44
	13. L-Acetylcarnitine	0.57	$2x10^{-4}$	0.003	36
	14. Dodecanoic acid	0.57	$2x10^{-4}$	0.003	29
	15. 10-nitro-9E-octadecenoic acid|9-nitro-9E-octadecenoic acid	0.55	$2x10^{-4}$	0.003	41
	16. Barogenin	0.35	$3x10^{-4}$	0.003	37
	17. Dodecanedioic acid	0.52	$3x10^{-4}$	0.003	40
	18. PC(15:2/0:0)|PE(18:2/0:0)	0.52	$4x10^{-4}$	0.003	21
	19. LPC(20:1)	0.52	$4x10^{-4}$	0.003	17
	20. 5a-Androst-3-en-17-one	-0.52	$4x10^{-4}$	0.003	30
	21. MAG(16:0)	0.42	$4x10^{-4}$	0.003	12
	22. 4,8 dimethylnonanoyl carnitine	0.51	$4x10^{-4}$	0.003	32
	23. C12 Carnitine	0.51	$4x10^{-4}$	0.003	39
	24. 2a-(3-Hydroxypropyl)-1a,25-dihydroxy-19-norvitamin D3	0.51	$4x10^{-4}$	0.003	22

	25. gamma-Glutamyl-leucine	0.50	$4x10^{-4}$	0.003	19
	26. 7-Ketocholesterol	0.49	$4x10^{-4}$	0.003	16
	27. 3-Indolepropionic acid	-0.50	$4x10^{-4}$	0.003	26
	28. Naproxen	-0.49	$4x10^{-4}$	0.003	23
	29. Ceramide PE(38:2)	0.57	$4x10^{-4}$	0.003	20
	30. Theobromine	0.47	$4x10^{-4}$	0.003	5
	31. Paraxanthine|Theophylline	-0.46	$5x10^{-4}$	0.003	6
	32. PC(28:1)	0.46	$5x10^{-4}$	0.003	11
	33. Ceramide PE(35:1)|SPM(32:1)	0.46	$5x10^{-4}$	0.003	25
	34. Deoxycholic acid glycine conjugate	0.45	$6x10^{-4}$	0.003	27
	35. Ceramide PE(36:2)	0.44	$6x10^{-4}$	0.003	36
	36. MAG(14:0)	0.48	$7x10^{-4}$	0.004	28
	37. Uric acid	0.43	$7x10^{-4}$	0.004	15
	38. 1-Stearoyl-2-Hydroxy-sn-Glycero-3-PE	0.43	$8x10^{-4}$	0.004	18
	39. sodium glycochenodeoxycholate	0.42	$8x10^{-4}$	0.004	22
	40. L-Proline	0.42	$9x10^{-4}$	0.004	36
	41. Caffeine	-0.41	$9x10^{-4}$	0.004	45
	42. Indole-3-carbinol	0.41	$9x10^{-4}$	0.004	42
	43. L-Octanoylcarnitine	0.40	0.001	0.008	31
	44. Geranyl acetoacetate	0.40	0.001	0.008	22
	45. L-Leucine|L-Norleucine	0.40	0.001	0.008	29
	46. SPM(41:2)	-0.39	0.002	0.01	37
	47. 1-palmitoyl-2-hydroxy-sn-glycero-3-PE	0.39	0.002	0.01	42
	48. Biliverdin hydrochloride a	-0.39	0.002	0.01	24
	49. Linoleic acid	0.66	0.002	0.01	59
	50. α-Linolenic acid	0.65	0.003	0.01	66

	51. Butyryl-L-carnitine|Isobutyryl-L-carnitine	0.38	0.004	0.02	23
	52. Cholic acid	0.38	0.004	0.02	4
	53. Deoxycholic acid	0.38	0.004	0.02	16
	54. LPC(18e:0/0:0)	0.38	0.005	0.02	21
	55. LPE(18:1)	0.38	0.005	0.02	12
	56. LPE(18:2)	0.38	0.005	0.02	10
	57. PC(30:1)	0.38	0.005	0.02	37
	58. Creatine	0.38	0.005	0.02	22
	59. 1,3,7-Trimethyluric acid	0.38	0.005	0.02	6
	60. PC(42:7)	0.37	0.006	0.02	22
	61. Decanoyl-L-carnitine	0.37	0.006	0.02	15
	62. Stearic acid	0.37	0.007	0.02	30
	63. Palmitoleic acid	0.37	0.007	0.02	17
	64. N-palmitoyl-D-erythro-sphingosylphosphorylcholine	0.36	0.007	0.02	8
	65. Lactosyl ceramide(d18:1/16:0)	0.36	0.008	0.02	19
	66. Treprostinil	0.36	0.008	0.02	22
	67. PC(29:1)	0.36	0.008	0.02	19
	68. 4-Androsten-11Beta-ol-3,17-dione|11-Hydroxy-4-androstene-3,17-dione	-0.36	0.008	0.02	20
	69. Ceramide PE(37:2)|SPM(34:2)	0.36	0.009	0.02	20
	70. N-(octadecanoyl)-sphing-4-enine-1-PC	0.36	0.009	0.02	2
	71. PC(37:5)|PE(40:5)	0.35	0.009	0.02	14
	72. Hyodeoxycholic acid	0.35	0.009	0.02	12
	73. Pantothenic acid	0.34	0.009	0.02	13
	74. 1,4-dipalmitoyl-sn-glycero-3-PC	0.34	0.009	0.02	12
	75. LPE(16:0)	0.34	0.009	0.02	5

	76. MAG(18:0)	0.33	0.01	0.02	10
	77. Chenodeoxycholic acid	0.33	0.01	0.02	11
	78. 2,6 dimethylheptanoyl carnitine	0.33	0.01	0.02	6
	79. SPM(42:3)	-0.33	0.01	0.02	2
	80. Fatty acid C22:6	0.33	0.01	0.02	2
	81. Ornithine	0.33	0.01	0.02	1
	82. PC(30:2)	0.33	0.01	0.02	16
	83. Ceramide PE(34:1)	0.31	0.01	0.02	14
	84. N-(15Z-tetracosenoyl)-sphinganine-1-PC	-0.31	0.02	0.04	13
	85. 1-stearoyl-2-arachidonoyl-sn-glycero-3-PE|C15-20:4 PC	-0.31	0.02	0.04	18
	86. LPE(18:0)	0.31	0.02	0.04	22
	87. MAG(18:1)	0.31	0.02	0.04	29
	88. Lyso-PAF C-18	0.30	0.02	0.04	22
	89. Prostaglandin J2	-0.30	0.02	0.04	12
	90. LPE(20:4)	0.30	0.02	0.04	36
	91. PC(34:3)	0.30	0.02	0.04	17
	92. 2-Ketohexanoic acid	-0.28	0.02	0.04	15
	93. DL-2-Aminooctanoic acid	0.28	0.02	0.04	10
	94. LacCer(d18:1/14:0)	0.28	0.02	0.04	4
	95. Piperine	-0.27	0.02	0.04	13
	96. 1,2-dilinoleoyl-sn-glycero-3-PC	-0.27	0.03	0.06	2
	97. D-Urobilinogen|I-Urobilin	0.27	0.03	0.06	11
	98. C16-20:5 PC	-0.27	0.03	0.06	2
	99. Ceramide PE(33:1)|SPM(30:1)	0.26	0.03	0.06	12
	100. D-erythro-sphingosine	0.26	0.03	0.06	26

	101. 3a,6b,7b-Trihydroxy-5b-cholanoic acid	0.26	0.03	0.06	29
	102. LPC(20:4/0:0)	-0.25	0.04	0.08	14
	103. PC(35:3)|PE(38:3)	0.25	0.04	0.08	17
	104. LPC(16:1/0:0)	-0.24	0.05	0.09	11
	105. PC(O-18:1/0:0)|PC(P-18:0/0:0)	0.24	0.05	0.09	11
	106. 3-cis-Hydroxy-b,e-Caroten-3'-one	-0.23	0.05	0.09	3
	107. PS(18:0)	0.23	0.05	0.09	5
	108. Acetaminophen	0.23	0.05	0.09	19
	109. LPC(0:0/16:0)	0.23	0.05	0.09	2
	110. Arachidonic acid	0.22	0.05	0.09	17
	111. Biliribin I	-0.22	0.05	0.09	25
	112. Cortisol	0.22	0.05	0.09	12
	113. MAG(18:3)	0.22	0.05	0.09	1
	114. Glycocholic acid	0.22	0.05	0.09	1
	115. LPC(20:3)	0.22	0.05	0.09	3
	116. Indolelactic acid	0.22	0.06	0.10	5
	117. PC(38:7)	-0.22	0.06	0.10	5
	118. 1-palmitoyl-2-hydroxy-sn-glycero-3-PC	0.21	0.06	0.10	1
	119. Hippuric acid	0.21	0.07	0.12	1
	120. MAG(20:5)	0.21	0.07	0.12	1
	121. 1,3-dipalmitoleoyl-sn-glycero-3-PC	0.21	0.08	0.13	15
	122. Biliribin II	-0.21	0.08	0.13	26
	123. 3-Pyridylacetic acid|trigonelline	-0.20	0.09	0.14	4
	124. PC(38:4)	-0.20	0.09	0.14	6
	125. Gamma-Tocopherol	0.20	0.09	0.14	12
	126. MAG(18:2)	0.20	0.09	0.14	14
	127. 1-arachidoyl-2-hydroxy-sn-glycero-3-PC	0.20	0.10	0.16	14

	128. PC(34:4)	-0.20	0.10	0.16	2
	129. PC(36:5)	-0.19	0.10	0.16	2
	130. Creatinine	0.19	0.10	0.16	3
	131. N-(9Z-octadecenoyl)-sphing-4-enine-1-PC	0.19	0.11	0.17	15
	132. S-(4,5-Dihydro-2-methyl-3-furanyl) ethanethioate	-0.18	0.11	0.17	2
	133. SPM(d18:2/18:1)	0.18	0.11	0.17	15
	134. LPC(0:0/16:1)	-0.17	0.12	0.18	9
	135. 6-hydroxy-5-cholestanol|cholesterol	-0.17	0.12	0.18	10
	136. LPC(20:3)	0.17	0.12	0.18	13
	137. Alpha-Tocopherol	-0.17	0.13	0.19	14
	138. SPM(40:2)	-0.17	0.13	0.19	16
	139. LPC(18:3)	0.16	0.13	0.19	2
	140. 1,2-dilinoleoyl-sn-glycero-3-PC	0.16	0.14	0.20	4
	141. PC(38:2)	-0.16	0.14	0.20	5
	142. LPC(0:0/20:4)	-0.16	0.15	0.21	13
	143. 1-palmitoyl-2-docosahexaenoyl-sn-glycero-3-PC	0.16	0.15	0.21	21
	144. PC(36:1)	-0.15	0.16	0.22	12
	145. Arachidonic acid ethyl ester	0.15	0.16	0.22	10
	146. 1-stearoyl-2-hydroxy-sn-glycero-3-PC	0.14	0.17	0.23	8
	147. Oleamide	0.14	0.17	0.23	9
	148. PC(35:2)|PE(38:2)	0.14	0.17	0.23	9
	149. 1-vaccenoyl-2-palmitoyl-sn-glycerol	0.14	0.18	0.24	7
	150. 1,2-dioleoyl-sn-glycero-3-PC|1,2-dipetroselenoyl-sn-glycero-3-PC	-0.13	0.19	0.25	5
	151. Arachidonic acid methyl ester	-0.13	0.19	0.25	10

	152. LPC(0:0/18:0)	0.13	0.20	0.26	10
	153. LPC(22:4)	-0.12	0.20	0.26	12
	154. PC(34:3)|PE(37:3)	-0.12	0.20	0.26	4
	155. PC(33:1)|PE(36:1)	0.12	0.21	0.27	2
	156. MAG(16:1)	0.11	0.21	0.27	1
	157. Indoleacetic acid	0.10	0.25	0.32	2
	158. PC(40:6)	0.10	0.25	0.32	15
	159. Hypoxanthine	-0.11	0.25	0.32	1
	160. 1-O-1'-(Z)-octadecenyl-2-hydroxy-sn-glycero-3-PE	0.20	0.27	0.34	1
	161. LPC(20:2)	0.20	0.27	0.34	14
	162. PC(38:5)|PE(41:5)	-0.19	0.33	0.41	6
	163. Flavone	-0.18	0.35	0.43	4
	164. L-Tryptophan	0.19	0.35	0.43	4
	165. 17-phenyl trinor Prostaglandin E2|17-phenyl trinor Prostaglandin D2	0.18	0.35	0.43	10
	166. PC(32:1)	0.18	0.35	0.43	5
	167. LPC (22:5)	-0.18	0.38	0.46	5
	168. e-Caprolactam	0.18	0.39	0.47	9
	169. L-Phenylalanine	-0.17	0.40	0.48	2
	170. Benzenebutanoic acid	0.16	0.40	0.48	10
	171. Salicylic acid|Aspirin	0.16	0.41	0.48	1
	172. Phenylalanine	0.16	0.43	0.51	15
	173. Stachydrine	-0.15	0.45	0.52	13
	174. dehydroepiandrosterone sulfate	0.10	0.45	0.52	2
	175. LPC(22:5)	0.17	0.50	0.58	1
	176. LPC(20:5/0:0)	0.10	0.55	0.63	0
	177. PC(36:6)	0.14	0.58	0.66	1
	178. 1-oleoyl-2-hydroxy-sn-glycero-3-PC	0.17	0.59	0.67	6

	179. γ-Caprolactone	0.14	0.62	0.70	6
	180. L-Carnitine	0.11	0.63	0.70	8
	181. PC(36:3)|PE(39:3)	-0.29	0.73	0.81	6
	182. 1-oleoyl-2-hydroxy-sn-glycero-3-PE	0.29	0.74	0.82	1
	183. LPC (18:1)	0.29	0.77	0.85	10
	184. 1-oleoyl-2-palmitoyl-sn-glycero-3-PC	-0.29	0.78	0.86	3
	185. L-Tyrosine|o-Tyrosine	0.28	0.79	0.86	2
	186. PC(34:2)|PE(37:2)	-0.11	0.80	0.87	1
	187. Betaine	-0.10	0.82	0.88	13
	188. PC(34:5)|PE(37:5)	0.17	0.84	0.90	12
	189. 1-linoleoyl-2-stearoyl-sn-glycerol	-0.15	0.85	0.90	14
	190. PC(34:0)	-0.15	0.85	0.90	7
	191. L-Aspartyl-L-phenylalanine	0.13	0.85	0.90	1
	192. LPC(0:0/20:5)	0.16	0.86	0.91	1
	193. LPC(0:0/18:2)	-0.14	0.8	0.91	12
	194. PC(35:4)|PE(38:4)	-0.13	0.87	0.91	1
	195. PC(40:5)	0.12	0.87	0.91	2
	196. LPC(22:5)	0.11	0.90	0.92	4
	197. LPC(18:2/0:0)	-0.09	0.90	0.92	10
	198. Fatty acid C20:5 methyl ester	0.08	0.91	0.93	1
	199. Biliverdin hydrochloride b	-0.06	0.91	0.93	1
	200. PC(28:2)	0.05	0.92	0.93	1
	201. PC(38:3)	0.05	0.92	0.93	1
	202. Ceramide PE(35:2)|SPM(32:2)	-0.10	0.95	0.97	2
Men (n_M=1078)	1. 1,2-dipalmitoleoyl-sn-glycero-3-PC	-0.78	$9x10^{-5}$	0.005	6
	2. PC(34:4)	-0.76	$9x10^{-5}$	0.005	10
	3. PC(34:3)|PE(37:3)	-0.76	$9x10^{-5}$	0.005	18

	4. Hippuric acid	0.74	$9x10^{-5}$	0.005	3
	5. MAG(18:0)	-0.65	$2x10^{-4}$	0.005	53
	6. 1-palmitoyl-2-hydroxy-sn-glycero-3-PC	-0.67	$2x10^{-4}$	0.005	58
	7. PC(38:7)	-0.75	$2x10^{-4}$	0.005	20
	8. 2-Ketohexanoic acid	-0.70	$3x10^{-4}$	0.006	27
	9. LPC(16:1/0:0)	-0.67	$3x10^{-4}$	0.006	2
	10. LPC(0:0/16:0)	-0.69	$3x10^{-4}$	0.006	1
	11. Indole-3-carbinol	0.67	$4x10^{-4}$	0.006	46
	12. PC(38:3)	0.63	$4x10^{-4}$	0.006	37
	13. PS(18:0)	-0.63	$4x10^{-4}$	0.006	39
	14. LPC(0:0/18:0)	-0.65	$5x10^{-4}$	0.008	62
	15. LPC(22:5)	-0.63	$5x10^{-4}$	0.008	44
	16. PC(36:6)	-0.62	$5x10^{-4}$	0.008	48
	17. LPC(0:0/16:1)	-0.59	$5x10^{-4}$	0.008	42
	18. LPC(18:1)	-0.62	$8x10^{-4}$	0.008	44
	19. LPC(20:3)	-0.62	$8x10^{-4}$	0.008	6
	20. LPC(20:1)	-0.61	$8x10^{-4}$	0.008	38
	21. PC(36:5)	-0.61	0.001	0.008	10
	22. 1,2-dilinoleoyl-sn-glycero-3-PC	-0.62	0.001	0.008	20
	23. PC(36:3)|PE(39:3)	-0.62	0.001	0.008	15
	24. Indoleacetic acid	0.60	0.001	0.008	19
	25. MAG(16:0)	-0.58	0.002	0.01	22
	26. LPC(22:4)	-0.59	0.002	0.01	19
	27. PC(34:5)|PE(37:5)	-0.60	0.002	0.01	22
	28. LPC(0:0/20:4)	-0.57	0.002	0.01	3
	29. Corticosterone	-0.56	0.002	0.01	20
	30. PC(38:4)	-0.53	0.002	0.01	19
	31. PC(32:1)	-0.53	0.002	0.01	25

	32. Decanoyl-L-carnitine	0.53	0.002	0.01	41
	33. LPC(20:4/0:0)	-0.53	0.002	0.01	32
	34. Biliverdin hydrochloride a	-0.53	0.002	0.01	6
	35. Geranyl acetoacetate	-0.52	0.002	0.01	9
	36. Arachidonic acid ethyl ester	-0.52	0.002	0.01	18
	37. PC(34:2)|PE(37:2)	-0.52	0.002	0.01	17
	38. MAG(18:1)	-0.52	0.002	0.01	25
	39. 4-Androsten-11Beta-ol-3,17-dione|11-Hydroxy-4-androstene-3,17-dione	0.52	0.002	0.01	23
	40. PC(40:5)	-0.51	0.002	0.01	21
	41. LPC(22:5)	-0.51	0.002	0.01	15
	42. MAG(14:0)	-0.51	0.003	0.01	14
	43. Cholic acid	0.50	0.003	0.01	12
	44. L-Octanoylcarnitine	0.50	0.004	0.02	6
	45. LPC(22:5)	-0.49	0.004	0.02	9
	46. PC(35:4)|PE(38:4)	-0.49	0.004	0.02	5
	47. Prostaglandin J2	0.49	0.005	0.02	5
	48. PC(38:5)|PE(41:5)	-0.48	0.005	0.02	2
	49. LPE(18:1)	-0.48	0.005	0.02	1
	50. Gamma-Tocopherol	-0.47	0.005	0.02	10
	51. LPC(18:1)	-0.47	0.005	0.02	12
	52. LPC(20:5/0:0)	-0.46	0.005	0.02	15
	53. 1-oleoyl-2-hydroxy-sn-glycero-3-PC	-0.46	0.005	0.02	12
	54. L-Aspartyl-L-phenylalanine	0.46	0.005	0.02	12
	55. 1,2-dioleoyl-sn-glycero-3-phosphocholine|1,2-dipetroselenoyl-sn-glycero-3-PC	-0.46	0.005	0.02	13
	56. Treprostinil	0.45	0.005	0.02	21

	57. PC(36:1)	-0.45	0.005	0.02	24
	58. C16-20:5 PC	-0.44	0.005	0.02	31
	59. 1-stearoyl-2-arachidonoyl-sn-glycero-3-PE|C15-20:4 PC	-0.44	0.005	0.02	9
	60. 2a-(3-Hydroxypropyl)-1a,25-dihydroxy-19-norvitamin D3	0.44	0.005	0.02	19
	61. Biliribin II	-0.43	0.005	0.02	5
	62. MAG(18:2)	-0.42	0.005	0.02	2
	63. Piperine	-0.42	0.005	0.02	2
	64. Betaine	-0.41	0.005	0.02	10
	65. PC(34:0)	-0.41	0.005	0.02	15
	66. PC(42:7)	0.41	0.005	0.02	14
	67. LPE(18:0)	-0.41	0.005	0.02	14
	68. L-Tryptophan	-0.41	0.006	0.02	12
	69. Ceramide PE(35:2)|SPM(32:2)	0.41	0.006	0.02	21
	70. Dodecanedioic acid	0.40	0.006	0.02	21
	71. PC(28:2)	0.40	0.006	0.02	25
	72. LPC(0:0/18:2)	-0.40	0.006	0.02	5
	73. Butyryl-L-carnitine|Isobutyryl-L-carnitine	0.40	0.006	0.02	5
	74. 1-Stearoyl-2-Hydroxy-sn-Glycero-3-PE	-0.40	0.008	0.02	9
	75. PC(33:1)|PE(36:1)	0.39	0.009	0.02	3
	76. 3a,6b,7b-Trihydroxy-5b-cholanoic acid	0.39	0.009	0.02	1
	77. LPC(18:2/0:0)	-0.39	0.009	0.02	1
	78. Linoleic acid	-0.39	0.009	0.02	1
	79. Propranolol	0.39	0.01	0.02	2
	80. 1-arachidoyl-2-hydroxy-sn-glycero-3-PC	-0.38	0.01	0.02	10
	81. Stearic acid	-0.38	0.01	0.02	12

	82. Biliribin I	-0.38	0.01	0.02	11
	83. LPC(0:0/20:5)	-0.38	0.01	0.02	10
	84. 1-oleoyl-2-palmitoyl-sn-glycero-3-PC	-0.38	0.01	0.02	19
	85. L-Proline	-0.37	0.01	0.02	14
	86. cis-5-Tetradecenoylcarnitine	0.37	0.01	0.02	7
	87. Arachidonic acid methyl ester	-0.37	0.01	0.02	14
	88. Alpha-Tocopherol	-0.37	0.01	0.02	5
	89. Arachidonic acid	-0.36	0.01	0.02	21
	90. Uric acid	-0.35	0.02	0.04	10
	91. PC(34:3)	-0.35	0.02	0.04	1
	92. PC(35:3)|PE(38:3)	-0.35	0.02	0.04	2
	93. Ornithine	-0.35	0.03	0.06	1
	94. MAG(16:1)	-0.35	0.03	0.06	1
	95. C12 Carnitine	0.35	0.03	0.06	2
	96. Palmitic acid	-0.34	0.03	0.06	6
	97. 3-Pyridylacetic acid|trigonelline	0.33	0.03	0.06	3
	98. 10-nitro-9E-octadecenoic acid|9-nitro-9E-octadecenoic acid	0.32	0.04	0.08	3
	99. MAG(20:5)	-0.32	0.04	0.08	4
	100. 1-linoleoyl-2-stearoyl-sn-glycerol	-0.32	0.04	0.08	14
	101. Indolelactic acid	0.32	0.05	0.10	14
	102. 3-cis-Hydroxy-b,e-Caroten-3'-one	-0.31	0.06	0.12	12
	103. LPC(20:2)	-0.31	0.06	0.12	15
	104. 1,2-dilinoleoyl-sn-glycero-3-PC	-0.31	0.07	0.13	2
	105. Hypoxanthine	-0.30	0.07	0.13	1
	106. Alpha-Linolenic acid	-0.30	0.08	0.15	2
	107. Barogenin	0.30	0.08	0.15	2

	108. Fatty acid C20:5 methyl ester	-0.29	0.08	0.15	2
	109. dehydroepiandrosterone sulfate (sodium salt)	0.28	0.08	0.15	1
	110. Theobromine	-0.27	0.09	0.16	1
	111. cis/trans-Oleic acid	-0.27	0.09	0.16	1
	112. Creatinine	0.27	0.09	0.16	10
	113. sodium glycochenodeoxycholate	0.26	0.1	0.18	1
	114. Oleoyl-L-carnitine hydrochloride	0.26	0.1	0.18	1
	115. L-Tyrosine|o-Tyrosine	-0.26	0.11	0.19	4
	116. 1,3,7-Trimethyluric acid	-0.25	0.11	0.19	6
	117. 1-O-1'-(Z)-octadecenyl-2-hydroxy-sn-glycero-3-PE	-0.25	0.11	0.19	4
	118. LPE(20:4)	-0.25	0.12	0.20	3
	119. Choline	0.25	0.12	0.20	3
	120. Ceramide PE(34:1)	0.25	0.12	0.20	3
	121. LPE(18:2)	-0.24	0.12	0.20	2
	122. LPC(18e:0/0:0)	-0.24	0.13	0.21	1
	123. C16 Carnitine	0.24	0.13	0.21	1
	124. Deoxycholic acid glycine conjugate	0.23	0.13	0.21	1
	125. SPM(42:3)	0.23	0.13	0.21	1
	126. N-(15Z-tetracosenoyl)-sphinganine-1-PC	-0.23	0.13	0.21	1
	127. PC(28:1)	-0.23	0.13	0.21	10
	128. Gamma-Caprolactone	0.22	0.14	0.21	1
	129. Ceramide PE(35:1)|SPM(32:1)	-0.22	0.14	0.21	9
	130. Biliverdin hydrochloride b	-0.22	0.14	0.21	5
	131. gamma-Glutamyl-leucine	-0.21	0.14	0.21	5
	132. DL-2-Aminooctanoic acid	-0.21	0.14	0.21	12
	133. PC(40:6)	0.21	0.14	0.21	12
	134. D-erythro-sphingosine	0.21	0.15	0.22	5

	135. Deoxycholic acid	0.21	0.15	0.22	2
	136. PC(38:2)	0.21	0.15	0.22	14
	137. Ceramide PE(33:1)|Sphingomyelin(30:1)	0.20	0.15	0.22	10
	138. Acetaminophen	0.20	0.17	0.25	11
	139. Lactosyl ceramide(d18:1/16:0)	0.20	0.17	0.25	12
	140. Dodecanoic acid	0.19	0.18	0.26	10
	141. 1-palmitoyl-2-docosahexaenoyl-sn-glycero-3-PC	-0.19	0.18	0.26	1
	142. Stachydrine	0.19	0.2	0.26	1
	143. PC(35:2)|PE(38:2)	-0.19	0.2	0.28	12
	144. Creatine	-0.19	0.21	0.29	1
	145. D-Urobilinogen|I-Urobilin	0.18	0.21	0.29	4
	146. Ceramide PE(36:2)	0.17	0.22	0.30	6
	147. 1-oleoyl-2-hydroxy-sn-glycero-3-PE	-0.16	0.22	0.30	19
	148. Chenodeoxycholic acid	0.16	0.25	0.34	8
	149. 5a-Androst-3-en-17-one	-0.16	0.25	0.34	7
	150. Flavone	0.15	0.25	0.34	12
	151. LacCer(d18:1/14:0)	0.15	0.27	0.36	10
	152. Lyso-PAF C-18	-0.15	0.27	0.36	1
	153. Linoleyl carnitine	0.15	0.3	0.40	2
	154. L-Acetylcarnitine	0.15	0.31	0.41	2
	155. MAG(18:3)	-0.14	0.32	0.41	2
	156. L-Leucine|L-Norleucine	-0.14	0.32	0.41	3
	157. LPC(18:3)	-0.14	0.32	0.41	4
	158. Oleamide	0.14	0.32	0.41	5
	159. 6-hydroxy-5-cholestanol|cholesterol	-0.14	0.33	0.41	6
	160. 7-Ketocholesterol	-0.14	0.33	0.41	3

	161. e-Caprolactam	-0.13	0.33	0.41	2
	162. 1,2-dipalmitoyl-sn-glycero-3-PC	0.13	0.34	0.42	2
	163. LPC(20:1)	-0.13	0.34	0.42	4
	164. 1-vaccenoyl-2-palmitoyl-sn-glycerol	-0.13	0.34	0.42	1
	165. PC(O-18:1/0:0)|PC(P-18:0/0:0)	0.12	0.35	0.43	1
	166. Palmitoleic acid	-0.12	0.35	0.43	10
	167. Cortisol	-0.12	0.36	0.44	1
	168. Hyodeoxycholic acid	0.12	0.37	0.44	1
	169. 17-phenyl trinor Prostaglandin E2|17-phenyl trinor Prostaglandin D2	0.12	0.37	0.44	10
	170. Paraxanthine|Theophylline	0.11	0.38	0.45	1
	171. SPM(41:2)	0.11	0.4	0.47	2
	172. Phenylalanine	0.11	0.45	0.53	1
	173. PC(30:2)	-0.11	0.46	0.53	10
	174. SPM(d18:2/18:1)	-0.11	0.46	0.53	1
	175. Naproxen	0.11	0.46	0.53	1
	176. SPM(40:2)	-0.10	0.52	0.60	1
	177. Caffeine	-0.10	0.55	0.63	1
	178. S-(4,5-Dihydro-2-methyl-3-furanyl) ethanethioate	0.10	0.57	0.65	12
	179. PC(29:1)	0.10	0.58	0.65	1
	180. Ceramide PE(37:2)|SPM(34:2)	-0.10	0.58	0.65	2
	181. 2,6 dimethylheptanoyl carnitine	-0.10	0.61	0.68	3
	182. N-(9Z-octadecenoyl)-sphing-4-enine-1-PC	-0.10	0.62	0.69	3
	183. L-Phenylalanine	0.10	0.63	0.69	1
	184. PC(37:5)|PE(40:5)	0.10	0.63	0.69	2
	185. 4,8 dimethylnonanoyl carnitine	0.10	0.65	0.71	2
	186. Fatty acid C22:6	0.10	0.68	0.74	1
	187. Myristic acid	-0.10	0.71	0.77	5

	188. N-(octadecanoyl)-sphing-4-enine-1-PC	-0.10	0.73	0.78	9
	189. Ceramide PE(38:2)	-0.10	0.75	0.80	6
	190. LPE(16:0)	-0.10	0.83	0.882	7
	191. L-Carnitine	-0.10	0.85	0.90	1
	192. PC(15:2/0:0)|PE(18:2/0:0)	0.10	0.86	0.90	1
	193. Pantothenic acid	-0,09	0.87	0.90	1
	194. Heptadecanoic acid	-0,09	0.87	0.90	2
	195. Pentadecanoic acid	-0,09	0.87	0.90	2
	196. Salicylic acid|Aspirin	0,09	0.88	0.90	2
	197. 3-Indolepropionic acid	0,09	0.88	0.90	3
	198. 1-palmitoyl-2-hydroxy-sn-glycero-3-PE	0,09	0.89	0.90	4
	199. Benzenebutanoic acid	-0,09	0.89	0.90	2
	200. PC(30:1)	-0,09	0.9	0.90	1
	201. N-palmitoyl-D-erythro-sphingosylphosphorylcholine	0,09	0.93	0.93	1
	202. Glycocholic acid	0,09	0.95	0.95	1

Association of correlation among molecular features a with age

Women (n_w=804)	Molecular features	Correlation	*P*-value	*P*-value adjusted	Validation (>50%)
	1. Palmitoleic acid – Stearic acid	-0.75	$3x10^{-4}$	0.008	2
	2. Linoleic acid – Stearic acid	-0.73	$4x10^{-4}$	0.008	1
	3. Stearic acid – Palmitic acid	-0.72	$4x10^{-4}$	0.008	34
	4. Stearic acid – cis/trans-Oleic acid	-0.70	$4x10^{-4}$	0.008	1
	5. PC(32:1) – PC(35:3)|PE(38:3)	0.72	$5x10^{-4}$	0.008	54
	6. Ceramide PE(35:2)|SPM(32:2) – PC(32:1)	0.70	$5x10^{-4}$	0.008	2

	7. 1-Stearoyl-2-Hydroxy-sn-Glycero-3-PE – LPE(20:4)	-0.65	$6x10^{-4}$	0.008	27
	8. 1-2-dipalmitoyl-sn-glycero-3-PC – N-(15Z-tetracosenoyl)-sphinganine-1-PC	-0.69	$6x10^{-4}$	0.008	28
	9. PC(28:2) – PC(32:1)	0.69	0.001	0.008	57
	10. SPM(40:2) – Lactosyl ceramide(d18:1/16:0)	-0.67	0.001	0.008	25
	11. Heptadecanoic acid – Stearic acid	-0.64	0.001	0.008	4
	12. Biliverdin hydrochloride b – Bilirubin II	-0.63	0.001	0.008	25
	13. LPC(20:4/0:0) – PC(38:4)	-0.65	0.001	0.008	1
	14. N-(15Z-tetracosenoyl)-sphinganine-1-PC – Lactosyl ceramide(d18:1/16:0)	-0.65	0.002	0.01	20
	15. Deoxycholic acid – Deoxycholic acid glycine conjugate	-0.63	0.002	0.01	26
	16. N-palmitoyl-D-erythro-sphingosylphosphorylcholine – SPM(40:2)	-0.62	0.002	0.01	26
	17. PC(34:4) – PC(38:7)	0.62	0.002	0.01	2
	18. 1-2-dilinoleoyl-sn-glycero-3-PC – PC(38:4)	-0.62	0.003	0.01	20
	19. Ceramide PE(35:2)|SPM(32:2) – 1-oleoyl-2-palmitoyl-sn-glycero-3-PC	0.61	0.003	0.01	2
	20. PC(34:4) – PC(35:3)|PE(38:3)	0.61	0.003	0.01	37
	21. PC(30:1) – SPM(40:2)	-0.61	0.003	0.01	36
	22. Deoxycholic acid glycine conjugate – Chenodeoxycholic acid	-0.60	0.004	0.02	23
	23. Fatty acid C20:5 methyl ester – PC(34:5)|PE(37:5)	-0.60	0.004	0.02	1
	24. PC(28:2) – 1-oleoyl-2-palmitoyl-sn-glycero-3-PC	0.60	0.004	0.02	2
	25. PC(28:2) – PC(34:0)	0.59	0.004	0.02	13

	26. 1-palmitoyl-2-hydroxy-sn-glycero-3-PE – LPE(20:4)	-0.59	0.005	0.02	13
	27. Biliverdin hydrochloride b – Biliverdin hydrochloride a	-0.59	0.006	0.02	29
	28. Biliverdin hydrochloride b – Bilirubin I	-0.59	0.006	0.02	16
	29. Myristic acid – Stearic acid	-0.58	0.007	0.02	1
	30. LPC(20:5/0:0) – PC(34:5)|PE(37:5)	-0.58	0.007	0.02	26
	31. Fatty acid C20:5 methyl ester – PC(38:7)	-0.58	0.007	0.02	1
	32. 1-stearoyl-2-hydroxy-sn-glycero-3-PC – PC(40:5)	-0.52	0.007	0.02	12
	33. D-Urobilinogen|I-Urobilin – PC(40:5)	-0.51	0.008	0.02	13
	34. PC(36:6) – PC(38:7)	0.51	0.008	0.02	5
	35. PC(36:5) – PC(38:7)	0.50	0.008	0.02	28
	36. 4-8 dimethylnonanoyl carnitine – PC(40:5)	0.48	0.009	0.02	17
	37. MAG(18:3) –PC(40:5)	-0.42	0.009	0.02	9
	38. MAG(18:0) –PC(38:5)|PE(41:5)	-0.41	0.009	0.02	41
	39. Arachidonic acid –PC(38:5)|PE(41:5)	0.42	0.009	0.02	21
	40. Heptadecanoic acid – PC(38:5)|PE(41:5)	-0.43	0.009	0.02	22
	41. Phenylalanine –PC(38:5)|PE(41:5)	0.40	0.01	0.02	13
	42. 7-Ketocholesterol –Alpha-Tocopherol	-0.30	0.01	0.02	2
	43. 17-phenyl trinor Prostaglandin E2|17 – phenyl trinor Prostaglandin D2-Alpha-Tocopherol	-0.33	0.01	0.02	10
	44. sodium glycochenodeoxycholate – Alpha-Tocopherol	-0.30	0.01	0.02	6

	45. Deoxycholic acid glycine conjugate – Alpha-Tocopherol	-0.31	0.01	0.02	12
	46. Chenodeoxycholic acid – Alpha-Tocopherol	-0.30	0.01	0.02	15
	47. Linoleyl carnitine –Alpha-Tocopherol	-0.30	0.02	0.04	5
	48. Choline – Arachidonic acid	0.40	0.02	0.04	8
	49. Creatinine – Arachidonic acid	0.42	0.02	0.04	18
	50. LPC(20:4/0:0) – LPC(22:5)	-0.40	0.03	0.06	7
	51. LPC(20:3) – LPC(22:5)	-0.42	0.03	0.06	23
	52. LPC(20:5/0:0) – LPC(22:5)	-0.41	0.03	0.06	19
	53. LPC(22:4) – N-(9Z-octadecenoyl)-sphing-4-enine-1-PC	-0.39	0.04	0.07	2
	54. Biliverdin hydrochloride b – N-(9Z-octadecenoyl)-sphing-4-enine-1-PC	-0.21	0.04	0.07	1
	55. Biliverdin hydrochloride a – N-(9Z-octadecenoyl)-sphing-4-enine-1-PC	-0.19	0.04	0.07	1
	56. Biliribin I – N-(9Z-octadecenoyl)-sphing-4-enine-1-PC	-0.20	0.05	0.09	1
	57. Flavone – N-(octadecanoyl)-sphing-4-enine-1-PC	-0.23	0.05	0.09	23
	58. Geranyl acetoacetate – N-(octadecanoyl)-sphing-4-enine-1-PC	-0.22	0.07	0.12	12
	59. Dodecanedioic acid – N-(octadecanoyl)-sphing-4-enine-1-PC	-0.25	0.07	0.12	1
	60. 1-3-7-Trimethyluric acid – N-palmitoyl-D-erythro-sphingosylphosphorylcholine	0.25	0.07	0.12	14
	61. L-Aspartyl-L-phenylalanine – LPC(16:1/0:0)	-0.30	0.09	0.15	4
	62. Heptadecanoic acid – LPC(16:1/0:0)	-0.32	0.09	0.15	4
	63. Choline – 10-nitro-9E-octadecenoic acid|9-nitro-9E-octadecenoic acid	-0.29	0.1	0.16	2
	64. 3-Pyridylacetic acid|trigonelline – Alpha-Linolenic acid	-0.28	0.12	0.19	3

	65. Stachydrine – Alpha-Linolenic acid	-0.29	0.12	0.19	5
	66. Benzenebutanoic acid – Alpha-Linolenic acid	-0.22	0.13	0.20	10
	67. Geranyl acetoacetate – Arachidonic acid	-0.30	0.15	0.22	1
	68. Dodecanedioic acid – Arachidonic acid	-0.28	0.15	0.22	1
	69. 5a-Androst-3-en-17-one – Arachidonic acid	0.25	0.15	0.22	6
	70. Caffeine – Barogenin	-0.26	0.19	0.27	7
	71. Hippuric acid – Barogenin	0.25	0.19	0.27	2
	72. L-Acetylcarnitine – Barogenin	-0.25	0.19	0.27	14
	73. Indolelactic acid – Barogenin	-0.28	0.20	0.28	11
	74. Geranyl acetoacetate – Biliribin I	0.25	0.20	0.28	10
	75. Dodecanedioic acid – Biliribin I	0.23	0.22	0.30	1
	76. 5a-Androst-3-en-17-one – Biliribin I	0.22	0.22	0.30	1
	77. Propranolol – Biliribin I	0.23	0.25	0.33	2
	78. L-Phenylalanine – Biliverdin hydrochloride a	-0.24	0.3	0.40	9
	79. 2-6 dimethylheptanoyl carnitine – Stearic acid	-0.22	0.35	0.45	8
	80. 4-Androsten-11Beta-ol-3-17-dione|11-Hydroxy-4-androstene-3-17-dione – Stearic acid	-0.13	0.35	0.45	5
	81. Arachidonic acid – Stearic acid	-0.20	0.37	0.47	11
	82. MAG(16:1) – Treprostinil	-0.23	0.4	0.50	15
	83. MAG(16:0) – Treprostinil	-0.25	0.42	0.52	10
	84. MAG(18:3) – Treprostinil	-0.23	0.45	0.55	1
	85. Choline – Uric acid	-0.22	0.47	0.56	1
	86. Betaine – Uric acid	-0.23	0.47	0.56	1

	87. Barogenin – LPC(22:4)	-0.23	0.49	0.57	1
	88. GCA – LPC(22:4)	-0.24	0.49	0.57	2
	89. Uric acid – D-Urobilinogen|I-Urobilin	0.15	0.5	0.57	5
	90. L-Leucine|L-Norleucine – 1-stearoyl-2-hydroxy-sn-glycero-3-PC	0.21	0.5	0.57	6
	91. Hypoxanthine – 1-stearoyl-2-hydroxy-sn-glycero-3-PC	0.22	0.55	0.62	4
	92. PC(O-18:1/0:0)|PE(P-18:0/0:0) – 1-stearoyl-2-hydroxy-sn-glycero-3-PC	0.22	0.58	0.64	2
	93. L-Tyrosine|o-Tyrosine – 1-vaccenoyl-2-palmitoyl-sn-glycerol	-0.23	0.58	0.64	2
	94. Naproxen – 1-vaccenoyl-2-palmitoyl-sn-glycerol	0.23	0.6	0.66	1
	95. Oleamide – 2-6 dimethylheptanoyl carnitine	-0.22	0.62	0.67	1
	96. Pentadecanoic acid – 2a-(3-Hydroxypropyl)-1a-25-dihydroxy-19-norvitamin D3	-0.23	0.62	0.67	1
	97. 3-Pyridylacetic acid|trigonelline – 7-Ketocholesterol	0.21	0.64	0.67	2
	98. Stachydrine – 7 -Ketocholesterol	0.22	0.64	0.67	1
	99. Flavone – 5a-Androst-3-en-17-one	0.25	0.64	0.67	1
	100. LPC(18:2/0:0) – 3-cis-Hydroxy-b-e-Caroten-3'-one	-0.20	0.69	0.71	2
	101. Indole-3-carbinol – 3a-6b-7b-Trihydroxy-5b-cholanoic acid	-0.15	0.72	0.73	3
	102. MAG(16:1) – LPC(0:0/16:1)	0.20	0.82	0.83	4
	103. LPC(18:3) – LPC(18:2/0:0)	0.25	0.93	0.93	2
Men (n_M=1078)	1. Palmitoleic acid – MAG(16:1)	-0.85	$1x10^{-5}$	0.001	21
	2. MAG(16:0) – LPC(16:1/0:0)	-0.81	$4x10^{-5}$	0.002	53
	3. PC(35:3)|PE(38:3) – PC(36:3)|PE(39:3)	0.81	$4x10^{-5}$	0.002	2

	4. MAG(16:0) – LPC(0:0/16:1)	-0.80	$6x10^{-5}$	0.002	59
	5. MAG(18:1) – LPC(16:1/0:0)	-0.78	$1x10^{-4}$	0.003	60
	6. MAG(18:0) – LPC(0:0/16:1)	-0.76	$2x10^{-4}$	0.004	53
	7. MAG(16:1) – LPC(16:1/0:0)	-0.76	$2x10^{-4}$	0.004	42
	8. MAG(16:1) – cis/trans-Oleic acid	-0.77	$2x10^{-4}$	0.004	35
	9. Ceramide PE(37:2)|SPM(34:2) – SPM(d18:2/18:1)	0.77	$2x10^{-4}$	0.004	2
	10. LPE(18:2) – LPC(18:3)	0.74	$4x10^{-4}$	0.004	13
	11. PC(30:2) – SPM(d18:2/18:1)	0.74	$4x10^{-4}$	0.004	9
	12. LPE(18:2) – LPC(0:0/18:2)	0.73	$5x10^{-4}$	0.005	18
	13. MAG(18:0) – PC(32:1)	-0.72	$6x10^{-4}$	0.005	49
	14. LPE(18:2) – LPC(18:2/0:0)	0.71	$8x10^{-4}$	0.005	45
	15. Palmitoleic acid – LPC(0:0/16:1)	-0.71	$8x10^{-4}$	0.005	16
	16. N-(9Z-octadecenoyl)-sphing-4-enine-1-PC – 1,2-dilinoleoyl-sn-glycero-3-PC	0.71	$9x10^{-4}$	0.005	1
	17. Palmitoleic acid – LPC(16:1/0:0)	-0.70	$9x10^{-4}$	0.005	1
	18. Arachidonic acid ethyl ester – LPC(16:1/0:0)	-0.71	$9x10^{-4}$	0.005	21
	19. LPC(0:0/16:1) – Palmitic acid	-0.70	$9x10^{-4}$	0.005	37
	20. MAG(16:1) – LPC(0:0/16:1)	-0.70	0.001	0.005	14
	21. LPC(18:3) – LPC(18:2/0:0)	0.70	0.001	0.005	42
	22. Salicylic acid|Aspirin – 10-nitro-9E-octadecenoic acid|9-nitro-9E-octadecenoic acid	0.69	0.001	0.005	12
	23. Theobromine – 10-nitro-9E-octadecenoic acid|9-nitro-9E-octadecenoic acid	0.69	0.002	0.009	32
	24. Creatine – 10-nitro-9E-octadecenoic acid|9-nitro-9E-octadecenoic acid	0.68	0.002	0.009	24

	25. L-Leucine|L-Norleucine – 10-nitro-9E-octadecenoic acid|9-nitro-9E-octadecenoic acid	0.68	0.002	0.009	28
	26. MAG(14:0) – LPE(18:0)	0.69	0.004	0.02	8
	27. 10-nitro-9E-octadecenoic acid|9-nitro-9E-octadecenoic acid – LPE(18:0)	0.69	0.004	0.02	16
	28. Fatty acid C22:6 – LPE(18:0)	0.67	0.004	0.02	17
	29. 4-8 dimethylnonanoyl carnitine – LPE(18:0)	0.65	0.006	0.02	9
	30. Indole-3-carbinol – 10-nitro-9E-octadecenoic acid|9-nitro-9E-octadecenoic acid	0.64	0.006	0.02	2
	31. Acetaminophen – 10-nitro-9E-octadecenoic acid|9-nitro-9E-octadecenoic acid	0.65	0.006	0.02	1
	32. 2-Ketohexanoic acid – 10-nitro-9E-octadecenoic acid|9-nitro-9E-octadecenoic acid	0.64	0.006	0.02	1
	33. S-(4-5-Dihydro-2-methyl-3-furanyl) ethanethioate – 10-nitro-9E-octadecenoic acid|9-nitro-9E-octadecenoic acid	0.63	0.009	0.03	32
	34. L-Carnitine – 10-nitro-9E-octadecenoic acid|9-nitro-9E-octadecenoic acid	0.63	0.009	0.03	31
	35. L-Phenylalanine – 10-nitro-9E-octadecenoic acid|9-nitro-9E-octadecenoic acid	0.62	0.009	0.03	26
	36. Uric acid – 10-nitro-9E-octadecenoic acid|9-nitro-9E-octadecenoic acid	0.62	0.009	0.03	21
	37. L-Tyrosine|o-Tyrosine – LPE(20:4)	0.60	0.01	0.03	20
	38. Naproxen – LPE(20:4)	0.60	0.01	0.03	15
	39. e-Caprolactam – DL-2-Aminooctanoic acid	-0.59	0.01	0.03	19
	40. Choline – Gamma-Caprolactone	-0.56	0.01	0.03	5

	41. Creatinine – Gamma-Caprolactone	-0.58	0.02	0.05	9
	42. e-Caprolactam – Gamma-Caprolactone	0.59	0.02	0.05	12
	43. DL-2-Aminooctanoic acid – Gamma-Caprolactone	-0.56	0.02	0.05	12
	44. Choline – L-Proline	0.53	0.03	0.07	5
	45. Creatinine – L-Proline	0.52	0.03	0.07	6
	46. e-Caprolactam – L-Proline	-0.52	0.03	0.07	13
	47. DL-2-Aminooctanoic acid – L-Proline	-0.51	0.05	0.12	21
	48. Gamma – Caprolactone-L-Proline	-0.50	0.05	0.12	13
	49. Benzenebutanoic acid – Alpha-Linolenic acid	0.49	0.06	0.13	21
	50. Geranyl acetoacetate – Arachidonic acid	0.48	0.06	0.13	5
	51. LPC(20:4/0:0 – PC(34:0)	-0.46	0.06	0.13	12
	52. LPC(20:2) – PC(34:0)	-0.45	0.07	0.15	11
	53. LPC(20:3) – PC(34:0)	-0.45	0.07	0.15	11
	54. 1-3-7-Trimethyluric acid – C16-20:5 PC	0.40	0.07	0.15	32
	55. Deoxycholic acid glycine conjugate – PC(35:3)|PE(38:3)	-0.39	0.09	0.18	18
	56. LPC(20:3) – Biliverdin hydrochloride a	-0.34	0.09	0.18	19
	57. Lyso-PAF C-18 – Biliverdin hydrochloride a	0.33	0.11	0.21	9
	58. LPC(20:2) – Biliverdin hydrochloride a	-0.32	0.11	0.21	8
	59. LPC(20:1) – Biliverdin hydrochloride a	-0.34	0.11	0.21	5
	60. PS(18:0) – Biliribin I	0.32	0.13	0.24	2

	61. L-Carnitine – LPC(22:5)	0.35	0.13	0.24	2
	62. Biliribin II – PC(34:2)|PE(37:2)	-0.33	0.15	0.27	2
	63. D-Urobilinogen|I-Urobilin – PC(34:2)|PE(37:2)	-0.31	0.15	0.27	1
	64. cis/trans-Oleic acid – PC(34:2)|PE(37:2)	0.30	0.15	0.27	1
	65. 1-linoleoyl-2-stearoyl-sn-glycerol – PC(34:2)|PE(37:2)	0.32	0.17	0.29	2
	66. 1-vaccenoyl-2-palmitoyl-sn-glycerol – PC(34:2)|PE(37:2)	-0.31	0.17	0.29	3
	67. Ceramide PE(33:1)|SPM(30:1) – 1-oleoyl-2-palmitoyl-sn-glycero-3-PC	0.31	0.18	0.30	5
	68. Ceramide PE(34:1) – 1-oleoyl-2-palmitoyl-sn-glycero-3-PC	0.31	0.18	0.30	6
	69. Stearic acid – cis/trans-Oleic acid	-0.30	0.19	0.31	5
	70. PC(32:1) – PC(35:3)|PE(38:3)	0.30	0.19	0.31	7
	71. Deoxycholic acid glycine conjugate – Chenodeoxycholic acid	0.35	0.21	0.33	10
	72. Fatty acid C20:5 methyl ester – PC(34:5)|PE(37:5)	0.34	0.21	0.33	1
	73. Cortisol – SPM(41:2)	0.35	0.21	0.33	2
	74. 6-hydroxy-5-cholestanol|cholesterol – SPM(41:2)	-0.36	0.25	0.39	2
	75. cis-5-Tetradecenoylcarnitine – SPM(41:2)	0.34	0.27	0.41	2
	76. 6-hydroxy-5-cholestanol|cholesterol – PC(38:7)	0.36	0.32	0.47	1
	77. cis-5-Tetradecenoylcarnitine – PC(38:7)	0.38	0.32	0.47	3
	78. Cholic acid – PC(38:7)	-0.37	0.34	0.48	14
	79. PC(36:5) – PC(38:7)	-0.38	0.34	0.48	12
	80. 1-2-dilinoleoyl-sn-glycero-3-PC – PC(38:7)	-0.39	0.34	0.48	11

	81. PC(36:3)|PE(39:3) – PC(38:7)	-0.35	0.35	0.48	11
	82. SPM(41:2) – N-(15Z-tetracosenoyl)-sphinganine-1-PC	0.38	0.35	0.48	10
	83. MAG(16:0) – C16-20:5 PC	-0.39	0.36	0.48	2
	84. MAG(18:3) – C16-20:5 PC	-0.40	0.36	0.48	2
	85. L-Leucine|L-Norleucine – 1-stearoyl-2-hydroxy-sn-glycero-3-PC	-0.40	0.36	0.48	1
	86. Hypoxanthine – 1-stearoyl-2-hydroxy-sn-glycero-3-PC	-0.41	0.41	0.53	3
	87. Caffeine – GCA	-0.40	0.41	0.53	1
	88. Creatinine – 1-Stearoyl-2-Hydroxy-sn-Glycero-3-PE	-0.40	0.41	0.53	1
	89. MAG(18:2) – LPC(0:0/18:2)	-0.41	0.45	0.57	2
	90. Deoxycholic acid – cis-5-Tetradecenoylcarnitine	0.42	0.45	0.57	2
	91. cis-5-Tetradecenoylcarnitine – 2a-(3-Hydroxypropyl)-1a,25-dihydroxy-19-norvitamin D3	0.39	0.46	0.57	3
	92. LPC(0:0/20:4) – Biliverdin hydrochloride b	-0.39	0.46	0.57	2
	93. LPC(20:4/0:0) – Biliverdin hydrochloride b	-0.38	0.52	0.64	4
	94. 4-Androsten-11Beta-ol-3-17-dione|11-Hydroxy-4-androstene-3-17-dione – PC(35:3)|PE(38:3)	0.37	0.55	0.67	10
	95. Arachidonic acid – PC(35:3)|PE(38:3)	0.36	0.57	0.68	1
	96. Myristic acid – PC(36:5)	0.33	0.58	0.68	10
	97. 1-vaccenoyl-2-palmitoyl-sn-glycerol – SPM(40:2)	0.30	0.58	0.68	12
	98. Ceramide PE(33:1)|SPM(30:1) – SPM(40:2)	-0.30	0.61	0.71	1

	99. Ceramide PE(34:1) – SPM(40:2)	-0.29	0.62	0.71	2
	100. Oleoyl-L-carnitine hydrochloride – PC(36:1)	0.28	0.63	0.71	1
	101. Alpha-Tocopherol – PC(36:1)	0.29	0.63	0.71	12
	102. 3a-6b-7b-Trihydroxy-5b-cholanoic acid – PC(36:1)	-0.26	0.65	0.72	10
	103. 1-O-1'-(Z)-octadecenyl-2-hydroxy-sn-glycero-3-phosphoethanolamine – PC(36:1)	-0.27	0.65	0.72	2
	104. D-erythro-sphingosine – PC(36:6)	0.25	0.68	0.75	5
	105. Oleamide – PC(36:6)	0.24	0.71	0.77	3
	106. Piperine – PC(36:6)	0.22	0.73	0.79	5
	107. dehydroepiandrosterone sulfate – PC(35:3)|PE(38:3)	0.22	0.75	0.80	3
	108. LPE(16:0) – PC(35:3)|PE(38:3)	0.20	0.83	0.88	2
	109. LPC(0:0/20:5) – 1,2-dipalmitoleoyl-sn-glycero-3-PC	-0.19	0.85	0.89	1
	110. DL-2-Aminooctanoic acid– LPE(18:1)	0.20	0.86	0.89	1
	111. L-Tyrosine|o-Tyrosine – Biliverdin hydrochloride b	-0.18	0.88	0.90	2
	112. Naproxen – Biliverdin hydrochloride b	-0.15	0.9	0.92	1
	113. L-Tryptophan – Biliverdin hydrochloride b	0.15	0.93	0.94	2
	114. 16. 1-O-1'-(Z)-octadecenyl-2-hydroxy-sn-glycero-3-PE – LPC(20:5/0:0)	0.12	0.97	0.97	2
Association of ratios among molecular features a with age					
Women (n_w=804)	Molecular features	Correlation	*P*-value	BH *P*-value adjusted	Validation (>50%)
	1. Acetaminophen/1-Stearoyl-2-Hydroxy-sn-Glycero-3-PE	-0.67	$2x10^{-4}$	0.002	15

	2. Arachidonic acid ethyl ester/1-vaccenoyl-2-palmitoyl-sn-glycerol	0.67	$2x10^{-4}$	0.002	21
	3. Deoxycholic acid glycine conjugate/N-palmitoyl-D-erythro-sphingosylphosphorylcholine	0.66	$2x10^{-4}$	0.002	23
	4. 10-nitro-9E-octadecenoic acid|9-nitro-9E-octadecenoic acid/PC(15:2/0:0)|PE(18:2/0:0)	-0.64	$2x10^{-4}$	0.002	11
	5. Paraxanthine|Theophylline/GCA	-0.64	$2x10^{-4}$	0.002	1
	6. Arachidonic acid/GCA	-0.63	$2x10^{-4}$	0.002	35
	7. Pantothenic acid/Ceramide PE(35:2)|SPM(32:2)	0.63	$2x10^{-4}$	0.002	3
	8. Acetaminophen/GCA	-0.62	$2x10^{-4}$	0.002	49
	9. Decanoyl-L-carnitine/LPC(0:0/18:2)	-0.67	$3x10^{-4}$	0.002	56
	10. LPC(0:0/20:5)/1,2-dipalmitoleoyl-sn-glycero-3-PC	0.62	$3x10^{-4}$	0.002	23
	11. DL-2-Aminooctanoic acid/LPE(18:1)	-0.61	$3x10^{-4}$	0.002	12
	12. 1-O-1'-(Z)-octadecenyl-2-hydroxy-sn-glycero-3-PE/ LPC(0:0/18:2)	-0.60	$3x10^{-4}$	0.002	2
	13. MAG(16:0)/PC(15:2/0:0)|PE(18:2/0:0)	-0.60	$4x10^{-4}$	0.002	1
	14. 6-hydroxy-5-cholestanol|cholesterol/ PC(15:2/0:0)|PE(18:2/0:0)	-0.60	$4x10^{-4}$	0.002	34
	15. L-Carnitine/1-Stearoyl-2-Hydroxy-sn-Glyce:ro-3-PE	-0.59	$4x10^{-4}$	0.002	33
	16. 1-O-1'-(Z)-octadecenyl-2-hydroxy-sn-glycero-3-PE/LPC(20:5/0:0)	-0.59	$4x10^{-4}$	0.002	25
	17. Arachidonic acid methyl ester/LPC(18:1)	0.60	$4x10^{-4}$	0.002	11
	18. Prostaglandin J2/1-palmitoyl-2-hydroxy-sn-glycero-3-phosphoethanolamine	-0.59	0.001	0.005	10

	19. Caffeine/GCA	-0.59	0.001	0.005	42
	20. Creatinine/1-Stearoyl-2-Hydroxy-sn-Glycero-3-PE	-0.59	0.001	0.005	15
	21. 2a-(3-Hydroxypropyl)-1a,25-dihydroxy-19-norvitamin D3/2a-(3-Hydroxypropyl)-1a,25-dihydroxy-19-norvitamin D3	-0.59	0.001	0.005	12
	22. Arachidonic acid methyl ester/PC(35:3)|PE(38:3)	0.66	0.002	0.009	2
	23. Indolelactic acid/GCA	-0.58	0.002	0.009	34
	24. Pantothenic acid/GCA	-0.58	0.002	0.009	63
	25. Dodecanedioic acid/GCA	-0.58	0.003	0.01	68
	26. Arachidonic acid methyl ester/PC(15:2/0:0)|PE(18:2/0:0)	-0.58	0.003	0.01	15
	27. MAG(20:5)/PC(15:2/0:0)|PE(18:2/0:0)	-0.58	0.004	0.01	2
	28. 7-Ketocholesterol/PC(15:2/0:0)|PE(18:2/0:0)	-0.58	0.004	0.01	2
	29. MAG(18:3)/2a-(3-Hydroxypropyl)-1a,25-dihydroxy-19-norvitamin D3	-0.58	0.004	0.01	1
	30. MAG(18:2)/ LPC(0:0/18:2)	-0.58	0.004	0.01	36
	31. Deoxycholic acid/cis-5-Tetradecenoylcarnitine	0.58	0.004	0.01	34
	32. cis-5-Tetradecenoylcarnitine/2a-(3-Hydroxypropyl)-1a,25-dihydroxy-19-norvitamin D3	-0.58	0.005	0.02	1
	33. LPE(16:0)/Palmitic acid	0.58	0.005	0.02	2
	34. Arachidonic acid ethyl ester/Biliverdin hydrochloride b	0.58	0.006	0.02	9
	35. Myristic acid/Biliribin II	-0.58	0.006	0.02	21
	36. L-Carnitine/LPC(22:5)	0.67	0.006	0.02	12
	37. Oleoyl-L-carnitine hydrochloride/PC(36:1)	0.68	0.006	0.02	32

	38. Alpha-Tocopherol/PC(36:1)	-0.67	0.006	0.02	21
	39. 3a-6b-7b-Trihydroxy-5b-cholanoic acid/PC(36:1)	-0.66	0.006	0.02	5
	40. 1-O-1'-(Z)-octadecenyl-2-hydroxy-sn-glycero-3-PE/PC(36:1)	-0.65	0.006	0.02	21
	41. Heptadecanoic acid/SPM(40:2)	0.64	0.007	0.02	22
	42. Phenylalanine/SPM(40:2)	0.63	0.007	0.02	31
	43. Ceramide PE(37:2)|SPM(34:2)/SPM(40:2)	0.62	0.007	0.02	15
	44. PC(30:2)/SPM(40:2)	0.62	0.007	0.02	26
	45. N-palmitoyl-D-erythro-sphingosylphosphorylcholine/SPM(40:2)	0.63	0.007	0.02	29
	46. Arachidonic acid/PC(35:3)|PE(38:3)	0.62	0.007	0.02	15
	47. Myristic acid/PC(36:5)	0.61	0.007	0.02	18
	48. 1-vaccenoyl-2-palmitoyl-sn-glycerol/SPM(40:2)	0.61	0.008	0.02	39
	49. Ceramide PE(33:1)|SPM(30:1)/SPM(40:2)	0.60	0.008	0.02	42
	50. Ceramide PE(34:1)/SPM(40:2)	0.60	0.008	0.02	12
	51. 3-cis-Hydroxy-b-e-Caroten-3'-one/N-(9Z-octadecenoyl)-sphing-4-enine-1-PC	-0.59	0.008	0.02	25
	52. Creatinine/1-Stearoyl-2-Hydroxy-sn-Glycero-3-PE	-0.58	0.009	0.02	26
	53. MAG(18:2)/LPC(0:0/18:2)	-0.58	0.009	0.02	15
	54. Deoxycholic acid/cis-5-Tetradecenoylcarnitine	-0.59	0.009	0.02	3
	55. cis-5-Tetradecenoylcarnitine/2a-(3-Hydroxypropyl)-1a,25-dihydroxy-19-norvitamin D3	-0.57	0.009	0.02	15

	56. LPC(0:0/20:4)/Biliverdin hydrochloride b	-0.56	0.009	0.02	19
	57. Lyso-PAF C-18/PC(34:3)	0.57	0.01	0.02	26
	58. PS(18:0)/PC(34:3)	0.56	0.01	0.02	2
	59. LPC(18:3)/PC(34:3)	0.56	0.01	0.02	4
	60. 6-hydroxy-5-cholestanol|cholesterol/Ceramide PE(37:2)|SPM(34:2)	-0.55	0.02	0.03	21
	61. cis-5-Tetradecenoylcarnitine/Ceramide PE(37:2)|SPM(34:2)	-0.55	0.02	0.03	10
	62. Cholic acid/Ceramide PE(37:2)|SPM(34:2)	-0.52	0.05	0.08	11
	63. MAG(18:2)/Ceramide PE(37:2)|SPM(34:2)	-0.53	0.07	0.13	11
	64. MAG(18:1)-Ceramide PE(37:2)|SPM(34:2)	-0.52	0.09	0.15	6
	65. Oleamide/PC(36:6)	0.51	0.10	0.16	9
	66. Piperine/PC(36:6)	0.52	0.10	0.16	15
	67. dehydroepiandrosterone sulfate/PC(35:3)|PE(38:3)	0.51	0.10	0.16	12
	68. Ceramide PE(35:2)|SPM(32:2)/PC(30:1)	0.50	0.14	0.21	3
	69. PC(28:2)/PC(30:1)	0.50	0.14	0.21	12
	70. Ceramide PE(35:1)|SPM(32:1)/PC(30:1)	0.49	0.15	0.22	1
	71. Hippuric acid – Barogenin	-0.48	0.25	0.25	5
	72. L-Acetylcarnitine – Barogenin	-0.47	0.17	0.25	4
	73. Ceramide PE(34:1)/PC(30:1)	0.46	0.19	0.27	7
	74. LPC(22:4)-Ornithine	-0.43	0.19	0.27	16
	75. Biliverdin hydrochloride b-Ornithine	-0.42	0.20	0.27	14
	76. Biliverdin hydrochloride a-Ornithine	-0.40	0.20	0.27	12

	77. LPE(18:2)/1-2-dioleoyl-sn-glycero-3-PC|1-2-dipetroselenoyl-sn-glycero-3-PC	0.41	0.21	0.28	2
	78. PC(15:2/0:0)|PE(18:2/0:0)/1-2-dioleoyl-sn-glycero-3-PC|1-2-dipetroselenoyl-sn-glycero-3-PC	0.40	0.22	0.29	1
	79. LPE(18:1)/1-2-dioleoyl-sn-glycero-3-PC|1-2-dipetroselenoyl-sn-glycero-3-PC	0.39	0.23	0.30	2
	80. 1-oleoyl-2-hydroxy-sn-glycero-3-PE/1-2-dioleoyl-sn-glycero-3-PC|1-2-dipetroselenoyl-sn-glycero-3-PC	-0.39	0.24	0.31	3
	81. Stearic acid/Ornithine	-0.38	0.25	0.32	4
	82. Heptadecanoic acid/MAG(20:5)	0.37	0.27	0.34	54
	83. MAG(16:0)/PC(35:4)|PE(38:4)	-0.36	0.30	0.38	1
	84. 4,8 dimethylnonanoyl carnitine/2a-(3-Hydroxypropyl)-1a,25-dihydroxy-19-norvitamin D3	0.35	0.35	0.43	12
	85. 3-Pyridylacetic acid|trigonelline/1-2-dioleoyl-sn-glycero-3-PC|1-2-dipetroselenoyl-sn-glycero-3-PC	-0.35	0.37	0.45	1
	86. Stachydrine/1-2-dioleoyl-sn-glycero-3-PC|1-2-dipetroselenoyl-sn-glycero-3-PC	-0.34	0.39	0.47	3
	87. Benzenebutanoic acid/1-2-dioleoyl-sn-glycero-3-PC|1-2-dipetroselenoyl-sn-glycero-3-PC	-0.34	0.40	0.49	5
	88. Myristic acid/SPM(d18:2/18:1)	-0.33	0.45	0.53	6
	89. LPC(0:0/16:1)/Ceramide PE(38:2)	-0.33	0.50	0.58	9
	90. LPC(0:0/16:0)/Ceramide PE(38:2)	-0.32	0.51	0.59	4
	91. LPC(18:3)/D-Urobilinogen|I-Urobilin	0.31	0.69	0.798	1
	92. LPC(0:0/18:2)/D-Urobilinogen|I-Urobilin	0.30	0.72	0.81	5

	93. LPC(18:2/0:0)/D-Urobilinogen|I-Urobilin	0.30	0.73	0.82	2
	94. 1-palmitoyl-2-hydroxy-sn-glycero-3-PC/cis/trans-Oleic acid	-0.29	0.75	0.83	1
	95. LPE(20:4)/cis/trans-Oleic acid	-0.29	0.77	0.84	3
	96. PC(O-18:1/0:0)|PC(P-18:0/0:0)/cis/trans-Oleic acid	-0.28	0.79	0.86	6
	97. LPC(18e:0/0:0)/cis/trans-Oleic acid	-0.28	0.81	0.87	5
	98. LPC(22:5)-D-Urobilinogen|I-Urobilin	-0.28	0.82	0.87	1
	99. 5a-Androst-3-en-17-one/PC(35:4)|PE(38:4)	-0.27	0.83	0.87	2
	100. Decanoyl-L-Carnitine/PC(35:4)|PE(38:4)	-0.24	0.91	0.93	1
	101. 4-Androsten-11Beta-ol-3,17-dione|11-Hydroxy-4-androstene-3,17-dione/2a−(3−Hydroxypropyl)−1a−25−dihydroxy−19−norvitamin D3	-0.23	0.91	0.93	1
	102. Gamma-Caprolactone/1-2-dilinoleoyl-sn-glycero-3-PC	-0.18	0.91	0.93	1
	103. Deoxycholic acid glycine conjugate/PC(32:1)	-0.16	0.92	0.93	2
	104. LPC(22:4)-Biliribin II	0.14	0.95	0.95	3
Men (n_M=1078)	1. Arachidonic acid ethyl ester/PC(35:4)|PE(38:4)	0.92	$1x10^{-7}$	$2x10^{-5}$	2
	2. Fatty acid C20:5 methyl ester/PC(35:4)|PE(38:4)	0.9	$4x10^{-7}$	$4x10^{-5}$	16
	3. Cortisol/PC(35:4)|PE(38:4)	0.9	$5x10^{-7}$	$4x10^{-5}$	19
	4. C16 Carnitine/PC(35:4)|PE(38:4)	0.9	$7x10^{-7}$	$4x10^{-5}$	6
	5. 4,8 dimethylnonanoyl carnitine/PC(35:4)|PE(38:4)	0.88	$2x10^{-6}$	$9x10^{-5}$	45
	6. Hippuric acid/PC(34:5)|PE(37:5)	0.87	$3x10^{-6}$	$1x10^{-4}$	43
	7. Deoxycholic acid glycine conjugate/PC(35:4)|PE(38:4)	0.86	$4x10^{-6}$	$1x10^{-4}$	12

	8. 1,3,7-Trimethyluric acid/PC(34:5)|PE(37:5)	0.86	$5x10^{-6}$	$1x10^{-4}$	10
	9. Barogenin/PC(35:4)|PE(38:4)	0.86	$5x10^{-6}$	$1x10^{-4}$	8
	10. 17-phenyl trinor Prostaglandin E2|17-phenyl trinor Prostaglandin D2/PC(35:4)|PE(38:4)	0.85	$6x10^{-6}$	$1x10^{-4}$	13
	11. L-Acetylcarnitine/PC(37:5)|PE(40:5)	0.85	$7x10^{-6}$	$1x10^{-4}$	55
	12. L-Carnitine/PC(37:5)|PE(40:5)	0.85	$1x10^{-5}$	$2x10^{-4}$	51
	13. 2-Ketohexanoic acid/PC(37:5)|PE(40:5)	0.84	$1x10^{-5}$	$2x10^{-4}$	0
	14. Naproxen/PC(37:5)|PE(40:5)	0.85	$1x10^{-5}$	$2x10^{-4}$	46
	15. sodium glycochenodeoxycholate/PC(35:4)|PE(38:4)	0.84	$1x10^{-5}$	$2x10^{-4}$	2
	16. Arachidonic acid ethyl ester/2a-(3-Hydroxypropyl)-1a,25-dihydroxy-19-norvitamin D3	0.83	$2x10^{-5}$	$2x10^{-4}$	35
	17. MAG(16:1)/2a-(3-Hydroxypropyl)-1a,25-dihydroxy-19-norvitamin D3	0.83	$2x10^{-5}$	$2x10^{-4}$	12
	18. 1-oleoyl-2-hydroxy-sn-glycero-3-PE/ 2a−(3−Hydroxypropyl)−1a−25−dihydroxy−19−norvitamin D3	0.83	$2x10^{-5}$	$2x10^{-4}$	2
	19. Indole-3-carbinol/PC(37:5)|PE(40:5)	0.82	$2x10^{-5}$	$2x10^{-4}$	41
	20. L-Phenylalanine/PC(37:5)|PE(40:5)	0.83	$2x10^{-5}$	$2x10^{-4}$	26
	21. L-Tryptophan/PC(37:5)|PE(40:5)	0.83	$2x10^{-5}$	$2x10^{-4}$	11
	22. Caffeine/PC(34:5)|PE(37:5)	0.84	$2x10^{-5}$	$2x10^{-4}$	2
	23. cis-5-Tetradecenoylcarnitine/PC(35:4)|PE(38:4)	0.83	$2x10^{-5}$	$2x10^{-4}$	13
	24. Cholic acid/PC(35:4)|PE(38:4)	0.83	$2x10^{-5}$	$2x10^{-4}$	27
	25. LPE(18:0)/ 2a-(3-Hydroxypropyl)-1a,25-dihydroxy-19-norvitamin D3	0.82	$3x10^{-5}$	$3x10^{-4}$	16

	26. S-(4,5-Dihydro-2-methyl-3-furanyl) ethanethioate/PC(37:5)|PE(40:5)	0.82	$3x10^{-5}$	$3x10^{-4}$	1
	27. Treprostinil/PC(35:4)|PE(38:4)	0.82	$3x10^{-5}$	$3x10^{-4}$	1
	28. Stearic acid/PC(35:4)|PE(38:4)	0.82	$3x10^{-5}$	$3x10^{-4}$	1
	29. Linoleic acid/2a-(3-Hydroxypropyl)-1a,25-dihydroxy-19-norvitamin D3	0.81	$4x10^{-5}$	$3x10^{-4}$	2
	30. Barogenin/2a-(3-Hydroxypropyl)-1a,25-dihydroxy-19-norvitamin D3	0.79	$4x10^{-5}$	$3x10^{-4}$	12
	31. MAG(18:3)/PC(35:4)|PE(38:4)	0.83	$4x10^{-5}$	$3x10^{-4}$	6
	32. Heptadecanoic acid/MAG(20:5)	-0.79	$5x10^{-5}$	$4x10^{-4}$	5
	33. MAG(16:0)/PC(35:4)|PE(38:4)	0.79	$5x10^{-5}$	$4x10^{-4}$	38
	34. 4,8 dimethylnonanoyl carnitine/2a-(3-Hydroxypropyl)-1a,25-dihydroxy-19-norvitamin D3	0.78	$5x10^{-5}$	$4x10^{-4}$	42
	35. 10-nitro-9E-octadecenoic acid|9-nitro-9E-octadecenoic acid/PE(38:4)	0.78	$6x10^{-5}$	$4x10^{-4}$	16
	36. Prostaglandin J2/PC(35:4)|PE(38:4)	0.78	$6x10^{-5}$	$4x10^{-4}$	11
	37. PC(O-18:1/0:0)|PC(P-18:0/0:0)/LPC(20:1)	-0.76	$6x10^{-5}$	$4x10^{-4}$	54
	38. Oleoyl-L-carnitine hydrochloride/2a-(3-Hydroxypropyl)-1a,25-dihydroxy-19-norvitamin D3	0.73	$7x10^{-5}$	$4x10^{-4}$	36
	39. Pantothenic acid/PC(35:4)|PE(38:4)	0.73	$7x10^{-5}$	$4x10^{-4}$	10
	40. 5a-Androst-3-en-17-one/PC(35:4)|PE(38:4)	0.73	$7x10^{-5}$	$4x10^{-4}$	12
	41. Decanoyl-L-Carnitine/PC(35:4)|PE(38:4)	0.73	$7x10^{-5}$	$4x10^{-4}$	5
	42. 4-Androsten-11Beta-ol-3,17-dione|11-Hydroxy-4-androstene-3,17-dione/2a−(3−Hydroxypropyl)−1a−25−dihydroxy−19−norvitamin D3	0.72	$8x10^{-5}$	$4x10^{-4}$	52
	43. MAG(18:2)/PC(35:4)|PE(38:4)	0.71	$8x10^{-5}$	$4x10^{-4}$	54
	44. L-Octanoylcarnitine/LPC(22:5)	0.70	$8x10^{-5}$	$4x10^{-4}$	10

	45. 1-linoleoyl-2-stearoyl-sn-glycerol/Ceramide PE(36:2)	0.70	$9x10^{-5}$	$5x10^{-4}$	12
	46. PC(15:2/0:0)|PE(18:2/0:0)/PC(30:1)	0.70	$1x10^{-4}$	$5x10^{-4}$	36
	47. Flavone/PC(35:4)|PE(38:4)	0.70	$1x10^{-4}$	$5x10^{-4}$	22
	48. Butyryl-L-carnitine|Isobutyryl-L-carnitine/PE(38:4)	0.70	$1x10^{-4}$	$5x10^{-4}$	24
	49. L-Carnitine/LPC(22:5)	0.70	$2x10^{-4}$	0.001	35
	50. 10-nitro-9E-octadecenoic acid|9-nitro-9E-octadecenoic acid/LPC(22:5)	0.70	$2x10^{-4}$	0.001	5
	51. LPC(18e:0/0:0)/PC(40:6)	0.69	$2x10^{-4}$	0.001	10
	52. Lyso-PAF C-18/PC(40:6)	0.69	$3x10^{-4}$	0.001	2
	53. PS(18:0)/PC(40:6)	0.68	$3x10^{-4}$	0.001	6
	54. LPC(18:3)/PC(40:6)	0.69	$3x10^{-4}$	0.001	0
	55. LPC(0:0/18:2)-PC(40:6)	0.68	$3x10^{-4}$	0.001	24
	56. Flavone/SPM(42:3)	-0.68	$4x10^{-4}$	0.002	23
	57. Butyryl-L-carnitine|Isobutyryl-L-carnitine/SPM(42:3)	-0.67	$5x10^{-4}$	0.002	14
	58. Geranyl acetoacetate/SPM(42:3)	-0.67	$5x10^{-4}$	0.002	6
	59. N-(9Z-octadecenoyl)-sphing-4-enine-1-PC/SPM(42:3)	-0.68	$5x10^{-4}$	0.002	1
	60. 1-2-dipalmitoleoyl-sn-glycero-3-PC/SPM(42:3)	-0.66	$5x10^{-4}$	0.002	3
	61. Biliribin I/LacCer(d18:1/14:0)	0.69	$6x10^{-4}$	0.002	12
	62. PC(32:1)-SPM(42:3)	-0.65	$8x10^{-4}$	0.003	17
	63. L-Carnitine/LPC(22:5)	-0.68	$9x10^{-4}$	0.003	9
	64. Biliribin II/PC(34:2)|PE(37:2)	0.69	$9x10^{-4}$	0.003	4
	65. S-(4-5-Dihydro-2-methyl-3-furanyl) ethanethioate/PC(40:5)	0.67	$9x10^{-4}$	0.003	0
	66. L-Carnitine/PC(40:5)	0.69	0.001	0.004	11

	67. L-Phenylalanine/PC(40:5)	0.68	0.001	0.004	1
	68. Treprostinil/ LPC(0:0/18:2)	-0.68	0.001	0.004	9
	69. LPC(0:0/20:5)/1,2-dipalmitoleoyl-sn-glycero-3-PC	-0.67	0.001	0.004	19
	70. DL-2-Aminooctanoic acid/LPE(18:1)	0.65	0.005	0.02	21
	71. 1-O-1'-(Z)-octadecenyl-2-hydroxy-sn-glycero-3-PE/ LPC(0:0/18:2)	-0.67	0.005	0.02	17
	72. 1-O-1'-(Z)-octadecenyl-2-hydroxy-sn-glycero-3-PE/1-2-dioleoyl-sn-glycero-3-PC|1-2-dipetroselenoyl-sn-glycero-3-PC	0.66	0.006	0.02	7
	73. LPE(18:2)/1-2-dioleoyl-sn-glycero-3-PC|1-2-dipetroselenoyl-sn-glycero-3-PC	-0.68	0.007	0.02	4
	74. PC(15:2/0:0)|PE(18:2/0:0)/1-2-dioleoyl-sn-glycero-3-PC|1-2-dipetroselenoyl-sn-glycero-3-PC	-0.66	0.007	0.02	2
	75. LPE(18:1)/1-2-dioleoyl-sn-glycero-3-PC|1-2-dipetroselenoyl-sn-glycero-3-PC	-0.65	0.007	0.02	3
	76. 1-oleoyl-2-hydroxy-sn-glycero-3-PE/1-2-dioleoyl-sn-glycero-3-PC|1-2-dipetroselenoyl-sn-glycero-3-PC	0.65	0.008	0.02	1
	77. Lyso-PAF C-18/PC(34:3)	-0.64	0.008	0.02	3
	78. PS(18:0)/PC(34:3)	-0.63	0.008	0.02	1
	79. LPC(18:3)/PC(34:3)	-0.62	0.008	0.02	10
	80. e-Caprolactam/1-2-dilinoleoyl-sn-glycero-3-PC	0.61	0.008	0.02	12
	81. DL-2-Aminooctanoic acid/1-2-dilinoleoyl-sn-glycero-3-PC	0.62	0.009	0.03	13
	82. Gamma-Caprolactone/1-2-dilinoleoyl-sn-glycero-3-PC	0.61	0.009	0.03	12
	83. Deoxycholic acid glycine conjugate/PC(32:1)	0.60	0.009	0.03	20
	84. Choline/PC(34:3)	-0.60	0.01	0.03	4

	85. Creatinine/PC(34:3)	-0.61	0.01	0.03	6
	86. e-Caprolactam/PC(34:3)	-0.59	0.02	0.06	0
	87. DL-2-Aminooctanoic acid/PC(34:3)	-0.62	0.02	0.06	1
	88. Oleoyl-L-carnitine hydrochloride/PC(36:1)	-0.58	0.02	0.06	22
	89. Alpha-Tocopherol/3a-6b-7b-Trihydroxy-5b-cholanoic acid	-0.62	0.03	0.08	12
	90. 3a-6b-7b-Trihydroxy-5b-cholanoic acid/PC(36:1)	-0.61	0.03	0.08	13
	91. 1-O-1'-(Z)-octadecenyl-2-hydroxy-sn-glycero-3-PE/PC(36:1)	-0.62	0.03	0.08	5
	92. MAG(20:5)/N-(octadecanoyl)-sphing-4-enine-1-PC	0.58	0.03	0.08	4
	93. Acetaminophen/PC(32:1)	0.57	0.03	0.08	2
	94. 2-Ketohexanoic acid/PC(32:1)	0.57	0.04	0.10	6
	95. S-(4-5-Dihydro-2-methyl-3-furanyl) ethanethioate/PC(32:1)	0.57	0.04	0.10	7
	96. L-Carnitine/PC(32:1)	0.56	0.04	0.10	1
	97. e-Caprolactam/Gamma-Caprolactone	-0.53	0.04	0.10	5
	98. DL-2-Aminooctanoic acid/Gamma-Caprolactone	-0.52	0.04	0.10	2
	99. e-Caprolactam/L-Proline	-0.50	0.05	0.12	3
	100. DL-2-Aminooctanoic acid/L-Proline	0.51	0.05	0.12	3
	101. gamma-Glutamyl-leucine/1-2-dipalmitoleoyl-sn-glycero-3-PC	-0.50	0.05	0.12	4
	102. 3-Pyridylacetic acid|trigonelline/1-2-dioleoyl-sn-glycero-3-PC|1-2-dipetroselenoyl-sn-glycero-3-PC	0.50	0.09	0.21	14
	103. Stachydrine/1-2-dioleoyl-sn-glycero-3-PC|1-2-dipetroselenoyl-sn-glycero-3-PC	0.51	0.10	0.23	6

	104. Benzenebutanoic acid/1-2-dioleoyl-sn-glycero-3-PC|1-2-dipetroselenoyl-sn-glycero-3-PC	0.52	0.11	0.25	8
	105. C12 Carnitine/1-2-dipalmitoleoyl-sn-glycero-3-PC	-0.51	0.11	0.25	15
	106. Corticosterone/1-2-dipalmitoleoyl-sn-glycero-3-PC	-0.52	0.12	0.27	0
	107. LPC(0:0/16:1)/N-(9Z-octadecenoyl)-sphing-4-enine-1-PC	0.52	0.12	0.27	11
	108. LPC(16:1/0:0)/N-(9Z-octadecenoyl)-sphing-4-enine-1-PC	0.51	0.12	0.27	2
	109. LPC(0:0/16:0)/N-(9Z-octadecenoyl)-sphing-4-enine-1-PC	0.49	0.13	0.28	13
	110. Heptadecanoic acid/SPM(40:2)	0.48	0.13	0.28	27
	111. Phenylalanine/SPM(40:2)	0.48	0.13	0.28	16
	112. Decanoyl-L-carnitine/SPM(40:2)	-0.48	0.13	0.28	1
	113. 5a-Androst-3-en-17-one/SPM(d18:2/18:1)	0.49	0.14	0.29	1
	114. Propranolol/SPM(d18:2/18:1)	0.49	0.14	0.29	1
	115. gamma-Glutamyl-leucine/SPM(d18:2/18:1)	-0.47	0.14	0.29	2
	116. Myristic acid/SPM(d18:2/18:1)	-0.47	0.14	0.29	12
	117. LPC(0:0/16:1)/Ceramide PE(38:2)	0.46	0.15	0.30	23
	118. LPC(16:1/0:0)/Ceramide PE(38:2)	0.46	0.15	0.30	6
	119. LPC(0:0/16:0)-Ceramide PE(38:2)	0.45	0.15	0.30	9
	120. Oleamide/PC(36:6)	-0.42	0.18	0.36	14
	121. Piperine/PC(36:6)	-0.43	0.22	0.44	2
	122. dehydroepiandrosterone sulfate/PC(35:3)|PE(38:3)	-0.45	0.25	0.50	5
	123. LPC(22:4)/Ornithine	-0.44	0.26	0.51	1
	124. Biliverdin hydrochloride b/Ornithine	-0.44	0.26	0.51	1
	125. Biliverdin hydrochloride a/Ornithine	-0.42	0.27	0.51	1

	126. LPC(22:4)-Biliribin II	-0.41	0.27	0.51	2
	127. 6-hydroxy-5-cholestanol|cholesterol/LPC(20:5/0:0)	-0.40	0.27	0.51	4
	128. cis-5-Tetradecenoylcarnitine/LPC(20:5/0:0)	-0.40	0.27	0.51	12
	129. cis/trans-Oleic acid/PC(30:2)	0.40	0.28	0.51	10
	130. 1-linoleoyl-2-stearoyl-sn-glycerol/PC(30:2)	0.39	0.28	0.51	2
	131. 1-vaccenoyl-2-palmitoyl-sn-glycerol/PC(30:2)	0.38	0.28	0.51	6
	132. Ceramide PE(33:1)|SPM(30:1)/PC(30:2)	-0.39	0.28	0.51	12
	133. Arachidonic acid/1-arachidoyl-2-hydroxy-sn-glycero-3-PC	-0.37	0.28	0.51	3
	134. Heptadecanoic acid/1-arachidoyl-2-hydroxy-sn-glycero-3-PC	-0.37	0.28	0.51	1
	135. SPM(30:1)/PC(30:2)/1-arachidoyl-2-hydroxy-sn-glycero-3-PC	-0.36	0.29	0.52	1
	136. Decanoyl-L-carnitine-1-arachidoyl-2-hydroxy-sn-glycero-3-phosphocholine	-0.35	0.29	0.52	1
	137. 2-6 dimethylheptanoyl carnitine/3-cis-Hydroxy-b-e-Caroten-3'-one	0.34	0.30	0.52	2
	138. 4-Androsten-11Beta-ol-3-17-dione|11-Hydroxy-4-androstene-3-17-dione/3-cis-Hydroxy-b-e-Caroten-3'-one	0.34	0.30	0.52	5
	139. Arachidonic acid/3-cis-Hydroxy-b-e-Caroten-3'-one	0.33	0.30	0.52	2
	140. Decanoyl-L-carnitine/LPC(0:0/18:2)	0.34	0.33	0.56	2
	141. LPC(0:0/20:5)/1,2-dipalmitoleoyl-sn-glycero-3-PC	-0.34	0.33	0.56	3

	142. DL-2-Aminooctanoic acid/LPE(18:1)	-0.33	0.33	0.56	1
	143. MAG(18:3)/3-cis-Hydroxy-b-e-Caroten-3'-one	0.31	0.33	0.56	1
	144. Cholic acid/Ornithine	-0.31	0.33	0.56	2
	145. MAG(18:2)/Ornithine	-0.31	0.34	0.57	4
	146. MAG(18:1)/Ornithine	-0.30	0.35	0.58	1
	147. Stearic acid/Ornithine	-0.30	0.42	0.69	1
	148. L-Proline/Lyso-PAF C-18	0.31	0.43	0.70	5
	149. Betaine/Lyso-PAF C-18	0.29	0.44	0.72	6
	150. Salicylic acid|Aspirin/Lyso-PAF C-18	0.30	0.45	0.72	8
	151. Prostaglandin J2/LPC(20:1)	-0.29	0.45	0.72	2
	152. Hyodeoxycholic acid/LPC(20:1)	-0.32	0.45	0.72	2
	153. Deoxycholic acid/LPC(20:1)	-0.34	0.48	0.75	1
	154. Cortisol/LPC(20:1)	-0.32	0.48	0.75	1
	155. 4-8 dimethylnonanoyl carnitine/LPC(0:0/16:0)	-0.29	0.48	0.75	1
	156. Arachidonic acid ethyl ester/LPC(0:0/16:0)	-0.29	0.49	0.76	1
	157. MAG(18:0)/LPC(0:0/16:0)	-0.28	0.50	0.76	1
	158. 17-phenyl trinor Prostaglandin E2|17-phenyl trinor Prostaglandin D2/LPC(16:1/0:0)	0.27	0.50	0.76	1
	159. Sodium glycochenodeoxycholate/LPC(16:1/0:0)	0.28	0.50	0.76	1
	160. Deoxycholic acid glycine conjugate/LPC(16:1/0:0)	0.29	0.55	0.83	3
	161. Chenodeoxycholic acid/LPC(16:1/0:0)	0.28	0.60	0.89	1
	162. LPC(22:5)/D-Urobilinogen|I-Urobilin	-0.26	0.60	0.89	2
	163. MAG(16:0)/LPC(16:1/0:0)	0.25	0.60	0.89	15

	164. Ceramide PE(35:2)|SPM(32:2)/PC(30:1)	0.27	0.62	0.89	1
	165. PC(28:2)/PC(30:1)	-0.26	0.62	0.91	2
	166. Ceramide PE(35:1)|SPM(32:1)/PC(30:1)	0.26	0.63	0.91	3
	167. PC(30:1)/SPM(40:2)	-0.26	0.63	0.92	1
	168. Deoxycholic acid glycine conjugate/Chenodeoxycholic acid	0.24	0.65	0.92	1
	169. Fatty acid C20:5 methyl ester/PC(34:5)|PE(37:5)	0.25	0.65	0.93	1
	170. PC(28:2)/ 1-oleoyl-2-palmitoyl-sn-glycero-3-PC	0.28	0.67	0.93	1
	171. PC(28:2)/PC(34:0)	-0.25	0.67	0.95	3
	172. L-Phenylalanine/LPC(0:0/16:1)	0.21	0.67	0.95	2
	173. Uric acid/LPC(0:0/16:1)	0.21	0.69	0.95	2
	174. PC(32:1)/PC(35:3)|PE(38:3)	-0.24	0.69	0.95	2
	175. Stearic acid/Palmitic acid	0.22	0.69	0.95	3
	176. Stearic acid/cis/trans-Oleic acid	0.22	0.70	0.95	4
	177. Ceramide PE(35:2)|SPM(32:2)/PC(32:1)	-0.24	0.70	0.95	4
	178. PC(28:2)/PC(32:1)	-0.24	0.70	0.95	1
	179. Deoxycholic acid glycine conjugate/2a-(3-Hydroxypropyl)-1a-25-dihydroxy-19-norvitamin D3	-0.22	0.70	0.95	12
	180. Chenodeoxycholic acid/2a-(3-Hydroxypropyl)-1a-25-dihydroxy-19-norvitamin D3	-0.22	0.71	0.95	1
	181. Linoleyl carnitine/2a-(3-Hydroxypropyl)-1a-25-dihydroxy-19-norvitamin D3	-0.22	0.71	0.95	2

	182. Oleoyl-L-carnitine hydrochloride/2a-(3-Hydroxypropyl)-1a-25-dihydroxy-19-norvitamin D3	-0.23	0.71	0.95	2
	183. Alpha-Tocopherol/2a-(3-Hydroxypropyl)-1a-25-dihydroxy-19-norvitamin D3	-0.22	0.72	0.95	8
	184. cis-5-Tetradecenoylcarnitine/LPC(0:0/16:1)	0.21	0.72	0.95	7
	185. Cholic acid/LPC(0:0/16:1)	0.22	0.72	0.95	11
	186. MAG(18:2)/LPC(0:0/16:1)	0.22	0.73	0.95	1
	187. MAG(18:1)/LPC(0:0/16:1)	0.22	0.74	0.96	2
	188. LPE(18:0)/1-Stearoyl-2-Hydroxy-sn-Glycero-3-PE	0.21	0.74	0.96	3
	189. Choline/2a-(3-Hydroxypropyl)-1a-25-dihydroxy-19-norvitamin D3	-0.21	0.75	0.96	1
	190. Creatinine/2a-(3-Hydroxypropyl)-1a-25-dihydroxy-19-norvitamin D3	-0.21	0.75	0.96	1
	191. Fatty acid C22:6/PC(15:2/0:0)|PE(18:2/0:0)	0.20	0.77	0.97	1
	192. 4-8 dimethylnonanoyl carnitine/PC(15:2/0:0)|PE(18:2/0:0)	0.21	0.77	0.97	5
	193. Arachidonic acid ethyl ester/PC(15:2/0:0)|PE(18:2/0:0)	0.21	0.78	0.97	3
	194. MAG(18:0)/PC(15:2/0:0)|PE(18:2/0:0)	0.20	0.78	0.97	1
	195. Cholic acid/PC(15:2/0:0)|PE(18:2/0:0)	0.20	0.80	0.97	12
	196. 1-palmitoyl-2-hydroxy-sn-glycero-3-PC/cis/trans-Oleic acid	-0.20	0.80	0.97	4
	197. LPE(20:4)/cis/trans-Oleic acid	-0.20	0.81	0.97	2
	198. PC(O-18:1/0:0)|PC(P-18:0/0:0)/cis/trans-Oleic acid	-0.22	0.81	0.97	3
	199. LPC(18e:0/0:0)/cis/trans-Oleic acid	-0.22	0.81	0.97	8

	200. LPE(16:0)/1-palmitoyl-2-hydroxy-sn-glycero-3-PE	0.22	0.82	0.97	7
	201. Choline/GCA	-0.20	0.82	0.97	7
	202. Creatinine/GCA	-0.19	0.82	0.97	6
	203. e-Caprolactam/GCA	-0.19	0.84	0.97	2
	204. DL-2-Aminooctanoic acid/GCA	-0.18	0.84	0.97	2
	205. Gamma-Caprolactone/GCA	-0.18	0.84	0.97	2
	206. C12 Carnitine/Gamma-Tocopherol	0.20	0.85	0.97	1
	207. Corticosterone/Gamma-Tocopherol	0.21	0.85	0.97	2
	208. MAG(16:1)/Gamma-Tocopherol	0.22	0.85	0.97	1
	209. MAG(16:0)/Gamma-Tocopherol	0.22	0.85	0.97	1
	210. Myristic acid/LPE(18:0)	-0.23	0.87	0.97	1
	211. Alpha-Linolenic acid/LPE(18:0)	-0.24	0.87	0.97	3
	212. L-Aspartyl-L-phenylalanine/LPE(18:0)	-0.24	0.87	0.97	2
	213. Propranolol/3a-6b-7b-Trihydroxy-5b-cholanoic acid	0.22	0.90	0.97	1
	214. gamma-Glutamyl-leucine/3a-6b-7b-Trihydroxy-5b-cholanoic acid	0.21	0.90	0.97	1
	215. Myristic acid/3a-6b-7b-Trihydroxy-5b-cholanoic acid	0.20	0.90	0.97	1
	216. Alpha-Linolenic acid/3a-6b-7b-Trihydroxy-5b-cholanoic acid	0.22	0.90	0.97	1
	217. 2-Ketohexanoic acid/Alpha-Tocopherol	0.20	0.91	0.97	1
	218. C12 Carnitine/PC(30:1)	-0.20	0.91	0.97	2
	219. Corticosterone/PC(30:1)	-0.21	0.91	0.97	2
	220. MAG(16:1)/PC(30:1)	-0.21	0.91	0.97	2
	221. MAG(16:0)/PC(30:1)	-0.22	0.92	0.97	1
	222. L-Carnitine/Linoleyl carnitine	-0.20	0.92	0.97	1

	223. L-Phenylalanine/Linoleyl carnitine	-0.18	0.92	0.97	1
	224. Uric acid/Linoleyl carnitine	-0.19	0.92	0.97	1
	225. Indoleacetic acid/Linoleyl carnitine	-0.18	0.92	0.97	2
	226. Paraxanthine|Theophylline/Linoleyl carnitine	-0.17	0.93	0.97	1
	227. MAG(18:0)/3a-6b-7b-Trihydroxy-5b-cholanoic acid	0.18	0.93	0.97	3
	228. PS(18:0)/D-Urobilinogen|I-Urobilin	0.19	0.93	0.97	1
	229. LPC(18:3)/D-Urobilinogen|I-Urobilin	0.19	0.93	0.97	3
	230. LPC(0:0/18:2)/D-Urobilinogen|I-Urobilin	0.19	0.94	0.97	5
	231. LPC(18:2/0:0)/D-Urobilinogen|I-Urobilin	0.18	0.94	0.97	5
	232. Caffeine/LPC(22:5)	-0.16	0.94	0.97	4
	233. Hippuric acid/LPC(22:5)	-0.19	0.94	0.97	1
	234. L-Acetylcarnitine/LPC(22:5)	-0.18	0.95	0.97	6
	235. Indolelactic acid/LPC(22:5)	-0.18	0.95	0.97	1
	236. 1-2-dioleoyl-sn-glycero-3-phosphocholine|1-2-dipetroselenoyl-sn-glycero-3-PC/PC(36:1)	0.17	0.95	0.97	4
	237. Indoleacetic acid/PC(2:7)	0.16	0.95	0.97	1
	238. Paraxanthine|Theophylline/PC(42:7)	0.16	0.96	0.97	3
	239. L-Tyrosine|o-Tyrosine/PC (42:7)	0.15	0.96	0.97	2
	240. Naproxen/PC(42:7)	0.13	0.97	0.97	1
	241. Prostaglandin J2/1-palmitoyl-2-hydroxy-sn-glycero-3-PE	0.12	0.97	0.97	1
	242. Caffeine/GCA	-0.10	0.97	0.97	1
	243. Creatinine/1-Stearoyl-2-Hydroxy-sn-Glycero-3-PE	0.15	0.97	0.97	1

Table S3: Association with age of molecular feature abundances, correlation and ratios for women and men. Compound names, Winsorized Pearson's correlation coefficient, the *P*-values, the Benjamini-Hochberg adjusted *P*-values, and the number of time that the association with age was validated are given. The correlation with age of

abundances, correlation and ratios is given by Equations (6) – (8). An overview of the statistical procedure used to establish such association is given in Figure 2. Abbreviations: LPC = Lysophosphatidylcholine; LPE = lysophosphatidylethenolamine; PC = phosphatidylcholine; PE = phosphatidylethanolamine; MAG = monoacylglycerol; GCA = glycocholic acid; SPM = sphingomyelin.

4.1.4 ^{1}H-NMR metabolomics to investigate the overall and dosage-dependent effect of probiotics on human urine and serum metabolome

Di Cesare F. [1], Ghini V. [1], Calgaro M. [2], Tenori L. [3], Vitulo N. [2], De Prisco A.[4], Pane M. [4], Amoruso A.[4], Squarzanti D. [4], Azzimonti B. [5,6], Luchinat C. [1,3,7]

[1]Magnetic Resonance Center (CERM), University of Florence, 50019 Sesto Fiorentino, Italy;

[2] Department of Biotechnology, Università di Verona, Italy;

[3] Department of Chemistry, University of Florence, 50019 Sesto Fiorentino, Italy;

[4] Probiotical Research., 28100 Novara, Italy;

[5] Department of Health Sciences (DiSS), University of Piemonte Orientale (UPO), Via Solaroli 17, 28100 Novara, Italy;

[6] Center for Translational Research on Autoimmune and Allergic Diseases (CAAD), DiSS, UPO, Corso Trieste 15/A, 28100 Novara;

[7] Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine (CIRMMP), 50019 Sesto Fiorentino, Italy

In preparation

Candidate's contribuitions: Sample preparation, acquisition and pre-processing of NMR data, statistical analysis, interpretation of results, and writing the manuscript.

Abstract

The human gut hosts around one thousand different species of commensal microorganisms that play a crucial role in health promotion, implementing the host's physiology and metabolism. In this perspective, probiotics are increasingly used, with the final aim of manipulating the composition of the gut microbiota, and with the aim of improving balanced microbial communities. In this context, the challenge is to characterize and understand potential metabolic changes that could determine and affect the dynamic relationship between host and microbiome. Nuclear Magnetic Resonance (NMR)-based metabolomics offers the possibility to investigate hundreds of various metabolites and lipids detectable in biological fluids (*i.e.* serum, urine, etc.), providing a global representation of the molecular mechanisms and potential effects of the dynamic and evolving interactions between the microflora and host, and of the response to probiotic assumption. In this study, using an NMR-based metabolomic approach, we highlighted the molecular effects obtained by microorganisms modulation through probiotic treatment, on human urine and serum metabolome. Twenty-one healthy volunteers were enrolled in the study and administered two different dosages of probiotics (high and low) for a total of 8 weeks. 20 urine samples per subject and 1 serum sample per subject were collected before and during the probiotic assumption. Univariate and multivariate statistical analyses were used to evaluate the ^{1}H-NMR urine and serum spectra acquired, and to characterize the individual effects of the treatment.

1 Introduction

Several numbers of microorganisms, approximately 1.3 times more than host cells, exist and coexist in the human gastrointestinal tract, and they directly maintain and modulate the metabolic and molecular balance of the gut environment[1–4]. It is demonstrated that the highly complex net of microorganisms that compose the gut microbiota, thanks to the production of specific antimicrobial proteins and the change of redox status, pH, and nutrient distribution, prevent the adhesion and proliferation of exogenous organisms, determining the fortification of the host's gut immunity barrier[5–8]. Moreover, gut microorganisms are responsible for the regulation of many important human physiological pathways, including those involved in the synthesis of proinflammatory cytokines[9,10], reactive oxygen compounds[11,12], enzymes able to digest polysaccharides[13,14], and the production of vitamin K, and most of the water-soluble B vitamins, such as biotin, cobalamin, folates, nicotinic acid, pantothenic acid, pyridoxine, riboflavin, and thiamine[15–17] in humans. The human microbiota, in this way, contributes to the host's metabolism and physiology.

In this scenario, it is possible to define the constant host-microbiome interactions as generative, from a genetic and a metabolic point of view, of a superorganism, called the holobiont[18,19] and the individual phenotypes are seen as direct results of this complex and dynamic interactions[20–23]. The gut-microbiota composition regularly experiences changes in terms of structure and function. These changes could be dependent on physiological aspects (*i.e.* age, sex, BMI, etc.) and lifestyle and clinical aspects (*i.e.* diet, medical condition, drug treatments, etc.)[24,25]. When the microflora balance is definitely altered, human well-being could be compromised, driving several pathophysiological alterations[26–29]. In this light, to preserve and promote the healthy interactions between host and microbiota, probiotics, defined as "living microorganisms, which when administered in adequate amounts confer health benefits on the host"[30–32], are increasingly used as dietary supplements or functional food for improving balanced microbial communities, for the suppression of potential pathogens, for the immunomodulation, and the stimulation of epithelial cells proliferation and fortification[33]. Currently, high-throughput metagenomic studies have been widely conducted to achieve deep characterization of the gut microbiome strain-level composition after the probiotics treatments[34–37], but the metabolic interaction between probiotics and host remains only partially understood and investigated. Therefore, in order to better understand the holobiont metabolic interactions in health, changes in the functional and metabolic composition of the gut microbiota should be deeply explored.

In this context, nuclear magnetic resonance (NMR)-based metabolomics represents a powerful technique to bring to light the complex molecular mechanisms and the highly interconnected dynamics between the host and the associated microbiota, taking into account the response to

probiotics administration, providing crucial information about several metabolites detectable in biological fluids (*i.e.* serum, plasma, urine), shedding a deeper light on the metabolic function that the microbiota exerts on human health.

Several metabolomics studies on the administration of probiotics in patients in the state of health and disease demonstrated that the microbiota is intrinsically associated with overall health, including gut pathologies in both adults and children, clarifying how much probiotics can influence the healthy community and the physiological functions[38–42]. In this work, we analysed and characterised by NMR spectroscopy the metabolic concentration changes in urine and plasma samples of 21 healthy adult volunteers, regularly administered with two different probiotic posologies. The probiotic blends were composed of strains of the *Lactobacillus plantarum*, *Lactobacillus rhamnosus*, *Lactobacillus fermentum* and *Bifidobacterium longum* species and the period of the administration was characterised by a total of 8 weeks. As previously demonstrated [38], in order to have a more robust image of the metabolic behaviour of the holobiont, a strong point that we propose in this work is the use of different biofluids and 20 samples per subject for urine, considered much more related to inherent variability than serum. Univariate and multivariate analyses were used to evaluate the unidimensional NMR urine and serum spectra, and to evaluate the direct effect of the probiotics on the human metabolome.

2 Materials and Methods

2.1 Study population

The trial was registered at clinical.trials.gov with the registration number NCT04506385. The study group consists of twenty-one adult healthy volunteers with an overall age range from 24 to 64 years (6 men with a mean age of 52.8±10.8 years and 15 women with a mean age of 40.7±10.8 years), whose demographic characteristics are reported in Table 1. The exclusion criteria were: previous chronic and common diseases, previous surgery on the intestinal tract, probiotic/prebiotic and antibiotic treatments within 2 months, before the beginning of the study. In addition, each subject declared to keep their diet and lifestyle stable during the entire duration of the study.

2.2 Ethical issues

The study was conducted in accordance with the Declaration of Helsinki (1964) and its later amendments. Informed, written consent was obtained from all participants. Ethical approval (protocol n◦ 294/CE, study n◦ CE 14/20, International Ethics Committee A.O.U. "Maggiore della Carità", Novara, Italy) was obtained.

2.3 Sample collection

Forty-two serum and 840 urine samples were collected during the entire course of the study. All the serum samples were collected under overnight fasting conditions. For urine, the midstream of the first urine of the morning was collected. The pre-analytical treatment of all the samples followed standard operating procedures (SOPs) to obtain high-quality specimens for metabolomic analysis[43–45].

Blood samples were collected in serum blood collection tubes without anticoagulants at room temperature. The samples were processed within two hours of the blood draw. The samples were centrifuged at room temperature for 10 min at 1500 g, then serum was collected, and the aliquots were transferred into pre-labeled cryovials. Urine samples were collected in sterile plastic cups. All the samples were processed within two hours from the collection; centrifugation at 1000 to 3000 g for 5 min at + 4 °C was followed by a filtration using 0.20 μm cut-off filters. All the processing procedures are detailed in the paper by Takis et *al.*[46].

After the processing, both serum and urine samples were stored at -80 °C.

2.4 Probiotic formulations

The commercial probiotic formulation (2.5 g) administered during the study was a blend of four strains belonging to Probiotical S.P.A. collection. The formulation was made up of

Lactobacillus plantarum LP01 (LMG P-21021), Lactobacillus fermentum LF16 (DSM 26956), Lactobacillus rhamnosus LR06 (DSM 21981) and Bifidobacterium longum 04 (DSM 23233) blended with maltodextrin to obtain a cell load of 1x109 CFU/dose or 10x109 CFU/dose (Biolab Research Method 014-06). The probiotic formulations were referred to as Dosage A and Dosage B for the blend carrying 1x109 CFU/dose and 10x109 CFU/dose, respectively.

2.5 NMR sample preparation

NMR samples were prepared according to SOPs for urine and serum. Frozen samples were thawed at room temperature and shaken before use. A total of 300 µL of each plasma sample was added to 300 µL of a phosphate sodium buffer (70 mM Na2HPO4; 20% (v/v) 2H2O; 0.025% (v/v) NaN3; 0.8% (w/v) sodium trimethylsilyl [2,2,3,3–2H4] propionate (TSP) pH 7.4); a total of 750 µL of each urine sample was centrifuged at 14000 g for 5 min, and 630 µL of the supernatant was added to 70 µL of a potassium phosphate buffer (1.5 M K2HPO4, 100% (v/v) 2H2O, 10 mM sodium trimethylsilyl [2,2,3,3−2H4] propionate (TMSP) pH 7.4).The mixtures were homogenised by vortexing for 30 s and a total of 600 µL of each mixture was transferred into a 5.00 mm NMR tube (Bruker BioSpin, Rheinstetten, Germany) for analysis.

2.6 NMR analysis and processing

Monodimensional (1D) ^{1}H-NMR spectra were acquired using a Bruker 600 MHz spectrometer (Bruker BioSpin s.r.l., Germany) optimised for metabolomics samples, operating at 600.13 MHz and equipped with a 5 mm cryoprobe, an automatic tuning-matching (ATM), and an automatic sample changer. In the NMR probe, the samples were kept for 3 minutes ahead for temperature equilibration and maintenance. The acquisition temperature used was 300 K for urine and 310 K for serum samples.

According to standard procedures, for each serum sample, three monodimensional ^{1}H NMR spectra were acquired with water peak suppression and different pulse sequences: (i) a standard NOESY (Nuclear Overhauser Effect Spectroscopy) 1Dpresat (noesygppr1d.comp; Bruker BioSpin) pulse sequence, using 32 scans, 98304 data points, a spectral width of 18028.846 Hz, an acquisition time of 2.7 s, a relaxation delay of 4 s and a mixing time of 0.1 s. (ii) a standard CPMG [43] (cpmgpr1d.comp; Bruker BioSpin) pulse sequence, using 32 scans, 73728 data points, a spectral width of 12019.230 Hz, and a relaxation delay of 4 s. (iii) a standard diffusion-edited (ledbgppr2s1d.comp; Bruker BioSpin) pulse sequence, using 32 scans, 98304 data points, a spectral width of 18028.846 Hz, and a relaxation delay of 4s.

For each urine sample, only monodimensional ^{1}H NMR spectra were acquired with water peak suppression and a standard NOESY[47] pulse sequence using 64 scans, 65536 data points, a

spectral width of 12019.230 Hz, an acquisition time of 2.7 s, a relaxation delay of 4 s and a mixing time of 0.1 s. Samples collected from the different subjects were mixed and acquired in a totally random order to avoid any batch effects. All the NMR spectra were automatically corrected for phase and baseline distortions and calibrated to the reference signal of TMSP at δ 0.00 ppm, and to the glucose doublet at δ 5.24 ppm, for urine and serum respectively, using TopSpin 3.6.2 (Bruker BioSpin Gmbh, Germany). Each spectrum in the range 0.2 -10.0 ppm was segmented into 0.02 ppm chemical shift bins and the corresponding areas were integrated using AssureNMR software (Bruker BioSpin s.r.l., Germany); the region between 6.0 - 4.5 ppm containing residual water signal was excluded. For urine samples, normalisation was applied to the obtained bins to minimise dilution effects caused, for example, by variation in fluid intake; the area of each bin was normalised using Probabilistic Quotient Normalisation (PQN)[48]. Unlike urine, the serum is not affected by dilution effects, and solute concentrations are tightly controlled; thus, serum spectra and any normalisation method has been applied.

2.7 Metabolite assignment and quantification

Twenty-eight metabolites in serum samples and 40 metabolites in urine samples were correctly assigned in all spectra using a ^{1}H-NMR spectra library of pure organic compounds (BBIOREFCODE, Bruker BioSpin), public databases, as Human Metabolome Database[49], storing reference ^{1}H-NMR spectra of metabolites, and using information available in the literature. Matching between new NMR data and databases was performed using the AssureNMR and AMIX software (Bruker BioSpin s.r.l., Germany). The quantification of the assigned metabolites was performed directly by integrating the signals in the spectra in a defined spectral range, using a house-developed tool.

For completeness, the metabolites correctly assigned and quantified in both urine and serum samples are presented in Table 2.

2.8 Statistical analysis

2.8.1 Multivariate analysis

Firstly, the principal component analysis (PCA)-canonical analysis (CA)[50,51] was performed to increase the supervised data visualization, data space reduction, cluster detections, and analyzed group discrimination.

To obtain pairwise comparisons, before and after the treatment, multilevel Partial Least Square Discriminant Analysis (mPLS-DA)[52] was employed. For all classification models, accuracy, sensitivity, and specificity were calculated according to the standard definition. Moreover, the results were validated using the Monte Carlo cross-validation algorithm (MCCV)[53]. Using this approach, the original dataset was randomly split into a training set, characterised by 80% of

the data, which was used to test the test set, characterised by the remaining 20% of data. This procedure was repeated k = 100-times.

2.8.2 Univariate analysis

To study the metabolites trends and their relationship with the treatment, a mixed-effects linear regression framework was employed for each metabolite. Using a simplified notation, the full model for the log-quantification of a generic serum metabolite was specified as it follows:

$$\begin{aligned}\log(Q) = {} & Subject + Subject * Day + \beta_0 + \beta_1 Day + \beta_2(Time = POST) \\ & + \beta_3(Treatment = high) + \beta_4\, Day{:}\,(Time = POST) \\ & + \beta_5(Time = POST){:}\,(Treatment = high) \\ & + \beta_6 Day{:}\,(Treatment = high) \\ & + \beta_7 Day{:}\,(Time = POST){:}\,(Treatment = high) + \beta_8 Age \\ & + \beta_9(Gender = Female) + \beta_{10} BMI\end{aligned}$$

where:

- $log(Q)$ is the dependent variable of the model, *i.e.* the log-quantification of a generic metabolite;
- $Subject + Subject \cdot Day$ is the random part of the model: each subject has a random intercept and slope, hence, for each subject, the temporal trend defined by the variable Day could be different;
- β_0 is the fixed intercept;
- β_1 to β_7 are the coefficients for the day of the measurement (numerical, from 1 to 40, one for each subsequent subject's measurement), treatment dosage (categorical, low or high dosage), time (categorical, PRE if the measurement belongs to the phase I, POST otherwise), and their interactions;
- β_8 to β_{10} are the coefficients for age, gender, and BMI which were included in the models as they are considered possible confounding variables;
- the reference level for this model is represented by a male subject, belonging to the low dosage group, before the treatment intake starts.

To obtain a reduced model, more parsimonious than the full model, but with a comparable ability to describe the data variability, a stepwise model selection procedure was used. In detail, a log-likelihood ratio test is used to determine which variables are not significant for a

given metabolite (p-value threshold = 0.1). In addition, two different models are estimated in each step of the selection: the first one with an undefined intra-group correlation structure between observations and a second model with an autoregressive correlation structure of the first order. In the latter model a relationship between consecutive observations is hypothesised (*e.g.*, day 1 and day 2, day 2 and day 3, etc.). The Akaike Information Criterion (AIC) is used to choose between the two models.

Similarly, the full model for the log-quantification of a generic urine metabolite was specified as it follows:

$$\log(Q) = Subject + \beta_0 + \beta_1(Time = POST) + \beta_2(Treatment = high) + \beta_3(Time = POST) : (Treatment = high) + \beta_4 Age + \beta_5(Gender = Female) + \beta_6 BMI$$

where:

- $log(Q)$ is the dependent variable of the model, *i.e.*, the log-quantification of a generic metabolite;
- $Subject$ is the random intercept for each subject;
- β_0 is the fixed intercept;
- $\beta_1$to β_3 are the coefficients for the time of the measurement (categorical, PRE if the measurement belongs to the pre-treatment period, POST otherwise), treatment dosage (categorical, low or high dosage), and their interactions;
- $\beta_4$to β_6 are the coefficients for age, gender, and BMI which were included in the models as they are considered possible confounding variables;
- the reference level for this model is represented by a male subject, belonging to the low dosage group, before the treatment intake starts.

Once the models had been estimated, linear combinations of the parameters were used, and 95% confidence intervals were computed to describe trends (*i.e.* average differences between subsequent measurements) and average differences between phases for each metabolic log-quantification.

2.9 Software

All calculations were performed in the R (version 4.0.3) statistical environment. All plots were obtained using the ggplot2[54] R package. The multivariate analyses were carried out using a R

software developed in-house. The mixed-effects models were estimated using the “nlme” [55,56] R package.

3 Results

3.1 Study design

The study was based on two different phases (Figure 1A and 1B):

I. Phase I: as performed by Ghini et *al.*[38], the first phase of the study is characterised by a period of 4 weeks during which the healthy volunteers did not take any supplementation of probiotics. During this phase, each volunteer collected twenty urine samples (1 sample per day, excluding the weekend and the menstrual cycle days) and proceeded with their usual diet and lifestyle. At the beginning of Phase I, a serum sample from each subject was also collected.

II. Phase II: the second phase of the study is a period of 8 weeks during which the volunteers were administered a daily dose of probiotics. In this phase, the subjects were randomly divided into two groups, named "High-dosage" and "Low-dosage". The "high-dosage" group is characterise by $n = 10$ subjects; the "low-dosage" group is characterised by $n = 11$ subjects. Starting from day 28 of probiotic assumption, each volunteer collected twenty urine samples (1 sample per day, excluding the weekend and the menstrual cycle days) and proceeded with their usual diet and lifestyle. In the 4^{th} week of probiotic assumption of Phase II, a serum sample from each subject was also collected.

3.2 The effect of dose-dependent probiotics administration on urinary metabolic human phenotype

To characterise the urinary individual metabolic phenotype of the healthy subjects, and to investigate the effect of the probiotics and the dose-specific probiotic assumption on the metabolic profile, the Principal Component Analysis - Canonical Analysis - K-Nearest Neighbours (PCA-CA-KNN) statistical analysis, also used in previous studies conducted by our research group, was performed[50].

As expected[21,38], considering all urine samples collected before the administration of the probiotics at the baseline reference (Phase I), the individual discrimination was almost perfect, with an accuracy value of 99% (**Figure 2A).** During the probiotic treatment, individual discrimination decreases by 1% passing from 99% in Phase I to 98% in Phase II (**Figure 2B**).

Performing the same analysis on the dose-specific groups separately, we observed the same behaviour. In particular, the subjects treated with a low dose of probiotics, showed, in Phase I (**Figure 3A**) accuracy discrimination of 99% and, after the treatment, accuracy discrimination of 98% (**Figure 3B**). The subjects treated with a high dose of probiotics pass from an individual discrimination accuracy of 98% before the treatment (**Figure 3C**) to 97% after treatment (**Figure 3D**).

The decrease of the individual discrimination accuracy among subjects, also considering the dose-effect, could be related to the probiotic administration, but we cannot consider this effect as generating a solid structural change.

3.3 Effect of different doses of probiotics on the urinary metabolome

To highlight a potential global effect and a potential dose-dependent effect of the probiotic assumption, minimising the intra-individual variability, the entire urine spectra collected during the two treatment phases were compared using Multilevel - Partial Least Squares (M-PLS) analysis (**Figure 4**). Using this statistical approach, we evaluated how much the urinary profile changes after the introduction of an exogenous set of microorganisms.

We observed moderate discrimination (80%) and good separation between urine metabolome before and after treatment, considering the entire cohort of healthy volunteers (**Figure 4A**).

Investigating a dose-dependent effect, we interestingly and unexpectedly observed that the subjects treated with a low dose of probiotics tended to have a discrimination accuracy higher than the subjects treated with a high dose of probiotics (**Figure 4B, 4C**); more precisely 79% accuracy for the first group and 61% for the second group. This result suggests that, after 8 weeks of treatment, the metabolome reacts more efficiently to low doses than high doses of probiotics.

To describe metabolic variations, a mixed-effects regression model was implemented for each urinary metabolite (see Methods).

Firstly, log-quantification levels were tested for differences between the 10-th measure of phase II and the 10-th measure of phase I (net of other variables). 3-hydroxyisobutyric acid, 4-hydroxyphenylacetate, and sugar-unknown (range ppm = 5.218-5.200) were decreased in phase II for both dosage groups, valine and isoleucine were significantly decreased only for the low dosage group, allantoin was decreased only for the high dosage group, while tartrate was significantly increased for the low dosage group (Table 3).

3.4 Effect of different doses of probiotics on serum metabolome

To evaluate the overall effect of probiotics on serum samples, ^{1}H-NMR serum spectra, collected during both phases, were compared using the same statistical approach performed on urine: M-PLS analysis. As reported before, a multilevel approach could be useful for reducing intra-individual variability.

We observed fair discrimination (77%) and good separation between serum metabolic profiles before and after treatment, considering the entire cohort of healthy volunteers (**Figure 5A**).

Investigating a dose-dependent effect, as expected, but in contrast with what we assessed in urine samples, we observed that the subjects treated with a low dose of probiotics tend to have a discrimination accuracy (76%) comparable to that evaluated in the subjects treated with a high dose of probiotics (78%) (**Figure 5B, 5C**).

Similar to the analysis of metabolic variations in urine, a simple mixed-effects regression model was implemented for each serum metabolite (see Methods).

Log-quantification levels were tested for differences between phase II and phase I (net of other variables). Acetone and pyruvate were significantly increased in phase II for both dosage groups, while histidine, glutamine, creatinine, acetate, and citrate were significantly decreased (Table 4).

4 Discussion

This study demonstrates that a probiotic administration can lead to changes in metabolites at urinary and serum system levels, without altering significantly the individual metabotypes. Our study also demonstrated the paramount role of having access to multiple and prolonged collections of samples in the pre-treatment condition to define a reliable baseline.

It is currently well recognised that gut microbiota can produce a wide range of metabolites by human endogenous or exogenous factors (*e.g.* food compounds) and some of them are exclusive of bacterial origin with a key role in host-microbiota cross-talk[57]. These metabolic changes driven by bacteria have been documented at faecal, urine, and serum level. Dietary interventions are emerging as a strategy to re-shaping and modulate not only the gut microbiota composition but even their metabolomes with concomitant positive effects on the hosts[58]. Similarly, it is demonstrated that the administration of exogenous beneficial bacteria in a close pre-existent ecosystem generates metabolic mutualistic benefits for both microbial communities and host[59,60].

In this work we observed that the administration of probiotics reduces the urinary individual discrimination accuracy by 1%, suggesting that probiotic leads to greater similarity in the metabolic host-microbiome cross-talk[38]. Although present, we are not able to consider this highlighted effect statistically robust.

Reducing the intra-individual variability, we determined the overall effect of probiotics on urine subject-specific metabolomic profile. In particular, the urine profiles related to the baseline period were discriminated from the urine profiles collected during the administration of the probiotics with an accuracy of 80%. The same approach was also conducted by dividing our cohort into dosage-dependent groups. Interestingly, a greater overall effect of the non-invasive treatment on the low-dosage group (accuracy = 79%) compared with the high-dosage

group (accuracy = 61%) was recorded. The same approach was also performed on serum samples. The overall effect of probiotics assumption on serum metabolomic profiles was about 77%, considering the entire cohort of study and the two dosage-dependent sub-groups. The molecular mechanisms by which these different dosage-dependent effects are generated in urine metabolomes are still not clear and need to be more deeply investigated; probably the indigenous microbiome suffered increased gut selection pressure with the invasion of higher-dosage of probiotics, generating, in this way, an initial antagonistic relationship[61,62].

Several differences at the metabolic level, derived from the probiotic treatment, were highlighted in both biofluids considered in this study. The metabolic variations in urine samples were evaluated taking into account the differences between the 10-th measure of phase II and the 10-th measure of phase I. In particular, in urine samples, we observed that 3-hydroxyisobutyric acid, 4-hydroxyphenylacetate, and resonance of a sugar detectable in ranges from 5.218 to 5.200 significantly decreased after the probiotic treatment in both dosage-dependent groups. Interesting is the role played by 4-hydroxyphenylacetate, an intermediate of tyrosine metabolism, both in humans and in microbes. In the human gut, the amino acids that are not digested and absorbed can be metabolized by the gut microbiota to form the 3- and 4-hydroxyphenylacetate organic acids. Higher levels of these compounds are considered markers to reflect protein malabsorption or dysbiosis[63–66]. The reduction of urinary 4-hydroxyphenylacetate confirms, in our study, the potentiality of probiotics to re-balance the pre-existent gut microflora, ensuring an improvement in the absorption and bioavailability of molecules, AAs, necessary for human well-being. Although we are not able to perform a precise assignment of the resonance in 5.218-5.200 ppm, we can affirm that, due to its spectroscopic characteristics, this signal belongs to a sugar. It is well known that the presence of sugars in the urine indicates an alteration of their metabolism[67]. Although the concentration of this sugar in phase one is in the normal range, as a result of the probiotic we notice a decrease, suggesting that the overall rebalancing of the intestinal microbiota also plays a fundamental role in improving the metabolism of sugars[68].

In the low dosage-dependent group only, we observed a significant decrease in valine and isoleucine concentration, while, in the high-dosage-dependent group only, we observed a decrease in allantoin. The branched-chain amino acids (BCAAs), in particular valine and isoleucine, are essential nutrients with important roles in protein synthesis in humans. The gut microbiota is a major source of circulating BCAAs through biosynthesis and absorption modification but elevated circulating BCAAs are associated with metabolic disorders (i.e., Type 1 and 2 diabetes)[69]. The same behaviour is observed for 3-hydroxyisobutyric acid. A

global reduction of BCAAs and of 3-hydroxyisobutyric acid indicates a potential role of these probiotics to promote both correct reabsorption and/or modification in the gut but also a potential protective role in pathologies that involve the alteration of the metabolism of BCAAs[69,70].

Urinary allantoin, an end product of purine metabolism, is normally present in urine and is formed from uric acid through reactions with oxidative species. The increase in terms of concentration of this molecule is directly associated with a systemic increase in oxidative stress[71,72]. In this perspective, the probiotics assumption reveal to be efficient in rebalancing the allantoin concentration in urine, promoting a reduction of this oxidative stress biomarker[73].

Lastly, tartaric acid significantly increases in the low-dosage dependent group. The biological and molecular significance linked to the increase of this compound is still to be discovered[74].

Considering the serum samples, the levels of the log-quantification metabolites were tested for differences between phase II and phase I. In this case, acetone and pyruvate were significantly increased in phase II, while histidine, glutamine, creatinine, acetate, and citrate were significantly decreased, for both dosage-dependent groups. Pyruvate performs several functions in human metabolism, revesting important roles in ATP production. We can speculate that the use of these probiotic strains could be useful for counteracting pathologies and/or dysbiosis phenomena that lead to a malfunction and a reduction in the activity of the glycolytic processes. It is demonstrated that the glycolysis impairment can cause a lowering of pyruvate and, also, lactate levels and an increase of glucose in serum, in particular in diabetic and coeliac patients[38,68,75]. A revertant effect could be stimulated by the use of probiotics, resulting in a remodeling of the gut microflora.

The decrease of histidine could be directly related to bacterial fermentation; in fact, the gut microbiota converts histidine into an immunoregulatory signal, histamine, able to suppress pro-inflammatory TNF production that could generate several local and/or systemic diseases[70,76].

It is also known that the gut microbiota utilizes glutamine as a nitrogen source for optimal survival and growth. Alteration in microbiota composition can profoundly influence glutamine metabolism, determining metabolic alteration in pathologies such as fibromyalgia[77,78].

Probiotic treatments, especially mixtures of Lactobacilli, are also used in promoting healthy kidney function; in particular, it was observed in the literature that exogenous microorganisms reduce the overall blood creatinine concentration, which is one of the most relevant biomarkers of chronic kidney disease[79,80].

5 References

(1) Fava, F.; Danese, S. Intestinal Microbiota in Inflammatory Bowel Disease: Friend of Foe? *World Journal of Gastroenterology* **2011**, *17* (5), 557–566. https://doi.org/10.3748/wjg.v17.i5.557.

(2) Mohri-Shiomi, A.; Garsin, D. A. Insulin Signaling and the Heat Shock Response Modulate Protein Homeostasis in the Caenorhabditis Elegans Intestine during Infection*. *Journal of Biological Chemistry* **2008**, *283* (1), 194–201. https://doi.org/10.1074/jbc.M707956200.

(3) Leser, T. D.; Mølbak, L. Better Living through Microbial Action: The Benefits of the Mammalian Gastrointestinal Microbiota on the Host. *Environmental Microbiology* **2009**, *11* (9), 2194–2206. https://doi.org/10.1111/j.1462-2920.2009.01941.x.

(4) Hooper, L. V.; Gordon, J. I. Commensal Host-Bacterial Relationships in the Gut. *Science* **2001**, *292* (5519), 1115–1118. https://doi.org/10.1126/science.1058709.

(5) Garrett, W. S.; Gordon, J. I.; Glimcher, L. H. Homeostasis and Inflammation in the Intestine. *Cell* **2010**, *140* (6), 859–870. https://doi.org/10.1016/j.cell.2010.01.023.

(6) Chung, K. W. Advances in Understanding of the Role of Lipid Metabolism in Aging. *Cells* **2021**, *10* (4), 880. https://doi.org/10.3390/cells10040880.

(7) Jones, R. M.; Neish, A. S. Redox Signaling Mediated by the Gut Microbiota. *Free Radic Res* **2013**, *47* (11), 950–957. https://doi.org/10.3109/10715762.2013.833331.

(8) Circu, M. L.; Aw, T. Y. Redox Biology of the Intestine. *Free Radic Res* **2011**, *45* (11–12), 1245–1266. https://doi.org/10.3109/10715762.2011.611509.

(9) Schirmer, M.; Smeekens, S. P.; Vlamakis, H.; Jaeger, M.; Oosting, M.; Franzosa, E. A.; ter Horst, R.; Jansen, T.; Jacobs, L.; Bonder, M. J.; Kurilshikov, A.; Fu, J.; Joosten, L. A. B.; Zhernakova, A.; Huttenhower, C.; Wijmenga, C.; Netea, M. G.; Xavier, R. J. Linking the Human Gut Microbiome to Inflammatory Cytokine Production Capacity. *Cell* **2016**, *167* (4), 1125-1136.e8. https://doi.org/10.1016/j.cell.2016.10.020.

(10) Li, H.-L.; Lu, L.; Wang, X.-S.; Qin, L.-Y.; Wang, P.; Qiu, S.-P.; Wu, H.; Huang, F.; Zhang, B.-B.; Shi, H.-L.; Wu, X.-J. Alteration of Gut Microbiota and Inflammatory Cytokine/Chemokine Profiles in 5-Fluorouracil Induced Intestinal Mucositis. *Front Cell Infect Microbiol* **2017**, *7*, 455. https://doi.org/10.3389/fcimb.2017.00455.

(11) Jones, R. M.; Mercante, J. W.; Neish, A. S. Reactive Oxygen Production Induced by the Gut Microbiota: Pharmacotherapeutic Implications. *Current Medicinal Chemistry 19* (10), 1519–1529.

(12) Shandilya, S.; Kumar, S.; Kumar Jha, N.; Kumar Kesari, K.; Ruokolainen, J. Interplay of Gut Microbiota and Oxidative Stress: Perspective on Neurodegeneration and Neuroprotection. *Journal of Advanced Research* **2022**, *38*, 223–244. https://doi.org/10.1016/j.jare.2021.09.005.

(13) Zhang, T.; Yang, Y.; Liang, Y.; Jiao, X.; Zhao, C. Beneficial Effect of Intestinal Fermentation of Natural Polysaccharides. *Nutrients* **2018**, *10* (8), 1055. https://doi.org/10.3390/nu10081055.

(14) Flint, H. J.; Scott, K. P.; Duncan, S. H.; Louis, P.; Forano, E. Microbial Degradation of Complex Carbohydrates in the Gut. *Gut Microbes* **2012**, *3* (4), 289–306. https://doi.org/10.4161/gmic.19897.

(15) Clarke, G.; Stilling, R. M.; Kennedy, P. J.; Stanton, C.; Cryan, J. F.; Dinan, T. G. Minireview: Gut Microbiota: The Neglected Endocrine Organ. *Mol Endocrinol* **2014**, *28* (8), 1221–1238. https://doi.org/10.1210/me.2014-1108.

(16) Martin, A. M.; Sun, E. W.; Rogers, G. B.; Keating, D. J. The Influence of the Gut Microbiome on Host Metabolism Through the Regulation of Gut Hormone Release. *Frontiers in Physiology* **2019**, *10*, 428. https://doi.org/10.3389/fphys.2019.00428.

(17) LeBlanc, J. G.; Chain, F.; Martín, R.; Bermúdez-Humarán, L. G.; Courau, S.; Langella, P. Beneficial Effects on Host Energy Metabolism of Short-Chain Fatty Acids and Vitamins Produced by Commensal and Probiotic Bacteria. *Microb Cell Fact* **2017**, *16* (1), 79. https://doi.org/10.1186/s12934-017-0691-z.

(18) Bordenstein, S. R.; Theis, K. R. Host Biology in Light of the Microbiome: Ten Principles of Holobionts and Hologenomes. *PLOS Biology* **2015**, *13* (8), e1002226. https://doi.org/10.1371/journal.pbio.1002226.

(19) Simon, J.-C.; Marchesi, J. R.; Mougel, C.; Selosse, M.-A. Host-Microbiota Interactions: From Holobiont Theory to Analysis. *Microbiome* **2019**, *7* (1), 5. https://doi.org/10.1186/s40168-019-0619-4.
(20) 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.
(21) 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.
(22) 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.
(23) 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.
(24) Odamaki, T.; Kato, K.; Sugahara, H.; Hashikura, N.; Takahashi, S.; Xiao, J.; Abe, F.; Osawa, R. Age-Related Changes in Gut Microbiota Composition from Newborn to Centenarian: A Cross-Sectional Study. *BMC Microbiology* **2016**, *16* (1), 90. https://doi.org/10.1186/s12866-016-0708-5.
(25) Rinninella, E.; Raoul, P.; Cintoni, M.; Franceschi, F.; Miggiano, G. A. D.; Gasbarrini, A.; Mele, M. C. What Is the Healthy Gut Microbiota Composition? A Changing Ecosystem across Age, Environment, Diet, and Diseases. *Microorganisms* **2019**, *7* (1), 14. https://doi.org/10.3390/microorganisms7010014.
(26) Durack, J.; Lynch, S. V. The Gut Microbiome: Relationships with Disease and Opportunities for Therapy. *J Exp Med* **2019**, *216* (1), 20–40. https://doi.org/10.1084/jem.20180448.
(27) Davis, C. D. The Gut Microbiome and Its Role in Obesity. *Nutr Today* **2016**, *51* (4), 167–174. https://doi.org/10.1097/NT.0000000000000167.
(28) Shreiner, A. B.; Kao, J. Y.; Young, V. B. The Gut Microbiome in Health and in Disease. *Curr Opin Gastroenterol* **2015**, *31* (1), 69–75. https://doi.org/10.1097/MOG.0000000000000139.
(29) Vijay, A.; Valdes, A. M. Role of the Gut Microbiome in Chronic Diseases: A Narrative Review. *Eur J Clin Nutr* **2022**, *76* (4), 489–501. https://doi.org/10.1038/s41430-021-00991-6.
(30) Hemarajata, P.; Versalovic, J. Effects of Probiotics on Gut Microbiota: Mechanisms of Intestinal Immunomodulation and Neuromodulation. *Therap Adv Gastroenterol* **2013**, *6* (1), 39–51. https://doi.org/10.1177/1756283X12459294.
(31) Plaza-Diaz, J.; Ruiz-Ojeda, F. J.; Gil-Campos, M.; Gil, A. Mechanisms of Action of Probiotics. *Adv Nutr* **2019**, *10* (Suppl 1), S49–S66. https://doi.org/10.1093/advances/nmy063.
(32) O'Connell, T. M. The Application of Metabolomics to Probiotic and Prebiotic Interventions in Human Clinical Studies. *Metabolites* **2020**, *10* (3). https://doi.org/10.3390/metabo10030120.
(33) Thomas, S.; Izard, J.; Walsh, E.; Batich, K.; Chongsathidkiet, P.; Clarke, G.; Sela, D. A.; Muller, A. J.; Mullin, J. M.; Albert, K.; Gilligan, J. P.; DiGuilio, K.; Dilbarova, R.; Alexander, W.; Prendergast, G. C. The Host Microbiome Regulates and Maintains Human Health: A Primer and Perspective for Non-Microbiologists. *Cancer Res* **2017**, *77* (8), 1783–1812. https://doi.org/10.1158/0008-5472.CAN-16-2929.
(34) Kristensen, N. B.; Bryrup, T.; Allin, K. H.; Nielsen, T.; Hansen, T. H.; Pedersen, O. Alterations in Fecal Microbiota Composition by Probiotic Supplementation in Healthy Adults: A Systematic Review of Randomized Controlled Trials. *Genome Medicine* **2016**, *8* (1), 52. https://doi.org/10.1186/s13073-016-0300-5.
(35) Westaway, J. A. F.; Huerlimann, R.; Kandasamy, Y.; Miller, C. M.; Norton, R.; Watson, D.; Infante-Vilamil, S.; Rudd, D. To Probiotic or Not to Probiotic: A Metagenomic Comparison of the Discharge Gut Microbiome of Infants Supplemented With Probiotics in NICU and Those Who Are Not. *Frontiers in Pediatrics* **2022**, *10*.
(36) Gueimonde, M.; Collado, M. C. Metagenomics and Probiotics. *Clinical Microbiology and Infection* **2012**, *18*, 32–34. https://doi.org/10.1111/j.1469-0691.2012.03873.x.
(37) Salminen, S.; Nurmi, J.; Gueimonde, M. The Genomics of Probiotic Intestinal Microorganisms. *Genome Biology* **2005**, *6* (7), 225. https://doi.org/10.1186/gb-2005-6-7-225.
(38) Ghini, V.; Tenori, L.; Pane, M.; Amoruso, A.; Marroncini, G.; Squarzanti, D. F.; Azzimonti, B.; Rolla, R.; Savoia, P.; Tarocchi, M.; Galli, A.; Luchinat, C. Effects of Probiotics Administration on

Human Metabolic Phenotype. *Metabolites* **2020**, *10* (10). https://doi.org/10.3390/metabo10100396.

(39) Baldassarre, M. E.; Di Mauro, A.; Tafuri, S.; Rizzo, V.; Gallone, M. S.; Mastromarino, P.; Capobianco, D.; Laghi, L.; Zhu, C.; Capozza, M.; Laforgia, N. Effectiveness and Safety of a Probiotic-Mixture for the Treatment of Infantile Colic: A Double-Blind, Randomized, Placebo-Controlled Clinical Trial with Fecal Real-Time PCR and NMR-Based Metabolomics Analysis. *Nutrients* **2018**, *10* (2), 195. https://doi.org/10.3390/nu10020195.

(40) Fan, Y.; Pedersen, O. Gut Microbiota in Human Metabolic Health and Disease. *Nat Rev Microbiol* **2021**, *19* (1), 55–71. https://doi.org/10.1038/s41579-020-0433-9.

(41) Ouald Chaib, A.; Levy, E. I.; Ouald Chaib, M.; Vandenplas, Y. The Influence of the Gastrointestinal Microbiome on Infant Colic. *Expert Review of Gastroenterology & Hepatology* **2020**, *14* (10), 919–932. https://doi.org/10.1080/17474124.2020.1791702.

(42) Hong, Y.-S.; Hong, K. S.; Park, M.-H.; Ahn, Y.-T.; Lee, J.-H.; Huh, C.-S.; Lee, J.; Kim, I.-K.; Hwang, G.-S.; Kim, J. S. Metabonomic Understanding of Probiotic Effects in Humans With Irritable Bowel Syndrome. *Journal of Clinical Gastroenterology* **2011**, *45* (5), 415–425. https://doi.org/10.1097/MCG.0b013e318207f76c.

(43) 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.

(44) Ghini, V.; Quaglio, D.; Luchinat, C.; Turano, P. NMR for Sample Quality Assessment in Metabolomics. *N Biotechnol* **2019**, *52*, 25–34. https://doi.org/10.1016/j.nbt.2019.04.004.

(45) Ghini, V.; Abuja, P. M.; Polasek, O.; Kozera, L.; Laiho, P.; Anton, G.; Zins, M.; Klovins, J.; Metspalu, A.; Wichmann, H.-E.; Gieger, C.; Luchinat, C.; Zatloukal, K.; Turano, P. Impact of the Pre-Examination Phase on Multicenter Metabolomic Studies. *New Biotechnology* **2022**, *68*, 37–47. https://doi.org/10.1016/j.nbt.2022.01.006.

(46) 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.

(47) Meiboom, S.; Gill, D. Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. *Review of Scientific Instruments* **1958**, *29*, 688–691. https://doi.org/10.1063/1.1716296.

(48) Dieterle, F.; Ross, A.; Schlotterbeck, G.; Senn, H. Probabilistic Quotient Normalization as Robust Method to Account for Dilution of Complex Biological Mixtures. Application in 1H NMR Metabonomics. *Anal. Chem.* **2006**, *78* (13), 4281–4290. https://doi.org/10.1021/ac051632c.

(49) Wishart, D. S.; Feunang, Y. D.; Marcu, A.; Guo, A. C.; Liang, K.; Vázquez-Fresno, R.; Sajed, T.; Johnson, D.; Li, C.; Karu, N.; Sayeeda, Z.; Lo, E.; Assempour, N.; Berjanskii, M.; Singhal, S.; Arndt, D.; Liang, Y.; Badran, H.; Grant, J.; Serra-Cayuela, A.; Liu, Y.; Mandal, R.; Neveu, V.; Pon, A.; Knox, C.; Wilson, M.; Manach, C.; Scalbert, A. HMDB 4.0: The Human Metabolome Database for 2018. *Nucleic Acids Res.* **2018**, *46* (D1), D608–D617. https://doi.org/10.1093/nar/gkx1089.

(50) Hotelling, H. Analysis of a Complex of Statistical Variables into Principal Components. *Journal of Educational Psychology* **1933**, *24* (6), 417–441. https://doi.org/10.1037/h0071325.

(51) Eisa, D. A.; Taloba, A. I.; Ismail, S. S. I. A Comparative Study on Using Principle Component Analysis with Different Text Classifiers. *International Journal of Computer Applications* **2018**, *180* (31), 1–6.

(52) Westerhuis, J. A.; van Velzen, E. J. J.; Hoefsloot, H. C. J.; Smilde, A. K. Multivariate Paired Data Analysis: Multilevel PLSDA versus OPLSDA. *Metabolomics* **2010**, *6* (1), 119–128. https://doi.org/10.1007/s11306-009-0185-z.

(53) Xu, Q.-S.; Liang, Y.-Z.; Du, Y.-P. Monte Carlo Cross-Validation for Selecting a Model and Estimating the Prediction Error in Multivariate Calibration. *Journal of Chemometrics* **2004**, *18* (2), 112–120. https://doi.org/10.1002/cem.858.

(54) Villanueva, R. A. M.; Chen, Z. J. Ggplot2: Elegant Graphics for Data Analysis (2nd Ed.). *Measurement: Interdisciplinary Research and Perspectives* **2019**, *17* (3), 160–167. https://doi.org/10.1080/15366367.2019.1565254.

(55) *Mixed-Effects Models in S and S-PLUS*; Statistics and Computing; Springer-Verlag: New York, 2000. https://doi.org/10.1007/b98882.
(56) Pinheiro, J., Bates, D., DebRoy, S., Sarkar. Nlme: Linear and Nonlinear Mixed Effects Models., 2021.
(57) Agus, A.; Clément, K.; Sokol, H. Gut Microbiota-Derived Metabolites as Central Regulators in Metabolic Disorders. *Gut* **2021**, *70* (6), 1174–1182. https://doi.org/10.1136/gutjnl-2020-323071.
(58) Zhang, C.; Yin, A.; Li, H.; Wang, R.; Wu, G.; Shen, J.; Zhang, M.; Wang, L.; Hou, Y.; Ouyang, H.; Zhang, Y.; Zheng, Y.; Wang, J.; Lv, X.; Wang, Y.; Zhang, F.; Zeng, B.; Li, W.; Yan, F.; Zhao, Y.; Pang, X.; Zhang, X.; Fu, H.; Chen, F.; Zhao, N.; Hamaker, B. R.; Bridgewater, L. C.; Weinkove, D.; Clement, K.; Dore, J.; Holmes, E.; Xiao, H.; Zhao, G.; Yang, S.; Bork, P.; Nicholson, J. K.; Wei, H.; Tang, H.; Zhang, X.; Zhao, L. Dietary Modulation of Gut Microbiota Contributes to Alleviation of Both Genetic and Simple Obesity in Children. *eBioMedicine* **2015**, *2* (8), 968–984. https://doi.org/10.1016/j.ebiom.2015.07.007.
(59) Cunningham, M.; Azcarate-Peril, M. A.; Barnard, A.; Benoit, V.; Grimaldi, R.; Guyonnet, D.; Holscher, H. D.; Hunter, K.; Manurung, S.; Obis, D.; Petrova, M. I.; Steinert, R. E.; Swanson, K. S.; van Sinderen, D.; Vulevic, J.; Gibson, G. R. Shaping the Future of Probiotics and Prebiotics. *Trends in Microbiology* **2021**, *29* (8), 667–685. https://doi.org/10.1016/j.tim.2021.01.003.
(60) Dobbins, J. J. Prescott's Microbiology, Eighth Edition. *J Microbiol Biol Educ* **2010**, *11* (1), 64–65. https://doi.org/10.1128/jmbe.v11.i1.154.
(61) Ma, C.; Zhang, C.; Chen, D.; Jiang, S.; Shen, S.; Huo, D.; Huang, S.; Zhai, Q.; Zhang, J. Probiotic Consumption Influences Universal Adaptive Mutations in Indigenous Human and Mouse Gut Microbiota. *Commun Biol* **2021**, *4* (1), 1–12. https://doi.org/10.1038/s42003-021-02724-8.
(62) Huang, S.; Jiang, S.; Huo, D.; Allaband, C.; Estaki, M.; Cantu, V.; Belda-Ferre, P.; Vázquez-Baeza, Y.; Zhu, Q.; Ma, C.; Li, C.; Zarrinpar, A.; Liu, Y.-Y.; Knight, R.; Zhang, J. Candidate Probiotic Lactiplantibacillus Plantarum HNU082 Rapidly and Convergently Evolves within Human, Mice, and Zebrafish Gut but Differentially Influences the Resident Microbiome. *Microbiome* **2021**, *9* (1), 151. https://doi.org/10.1186/s40168-021-01102-0.
(63) Mora Brugués, J.; González Sastre, F. Influence of Intestinal Flora on the Elimination of Phenylacetic Acid in Urine. *Clinical Chemistry* **1986**, *32* (1), 223. https://doi.org/10.1093/clinchem/32.1.223a.
(64) Blaut, M.; Clavel, T. Metabolic Diversity of the Intestinal Microbiota: Implications for Health and Disease. *The Journal of Nutrition* **2007**, *137* (3), 751S-755S. https://doi.org/10.1093/jn/137.3.751S.
(65) McGrath, K. H.; Pitt, J.; Bines, J. E. Small Intestinal Bacterial Overgrowth in Children with Intestinal Failure on Home Parenteral Nutrition. *JGH Open* **2019**, *3* (5), 394–399. https://doi.org/10.1002/jgh3.12174.
(66) Hertel, J.; Fässler, D.; Heinken, A.; Weiß, F. U.; Rühlemann, M.; Bang, C.; Franke, A.; Budde, K.; Henning, A.-K.; Petersmann, A.; Völker, U.; Völzke, H.; Thiele, I.; Grabe, H.-J.; Lerch, M. M.; Nauck, M.; Friedrich, N.; Frost, F. NMR Metabolomics Reveal Urine Markers of Microbiome Diversity and Identify Benzoate Metabolism as a Mediator between High Microbial Alpha Diversity and Metabolic Health. *Metabolites* **2022**, *12* (4), 308. https://doi.org/10.3390/metabo12040308.
(67) Tasevska, N. Urinary Sugars—A Biomarker of Total Sugars Intake. *Nutrients* **2015**, *7* (7), 5816–5833. https://doi.org/10.3390/nu7075255.
(68) Dai, Y.; Quan, J.; Xiong, L.; Luo, Y.; Yi, B. Probiotics Improve Renal Function, Glucose, Lipids, Inflammation and Oxidative Stress in Diabetic Kidney Disease: A Systematic Review and Meta-Analysis. *Renal Failure* **2022**, *44* (1), 862–880. https://doi.org/10.1080/0886022X.2022.2079522.
(69) Andersson-Hall, U.; Gustavsson, C.; Pedersen, A.; Malmodin, D.; Joelsson, L.; Holmäng, A. Higher Concentrations of BCAAs and 3-HIB Are Associated with Insulin Resistance in the Transition from Gestational Diabetes to Type 2 Diabetes. *Journal of Diabetes Research* **2018**, *2018*, e4207067. https://doi.org/10.1155/2018/4207067.
(70) Zhai, L.; Wu, J.; Lam, Y. Y.; Kwan, H. Y.; Bian, Z.-X.; Wong, H. L. X. Gut-Microbial Metabolites, Probiotics and Their Roles in Type 2 Diabetes. *Int J Mol Sci* **2021**, *22* (23), 12846. https://doi.org/10.3390/ijms222312846.

(71) Peron, G.; Santarossa, D.; Voinovich, D.; Dall'Acqua, S.; Sut, S. Urine Metabolomics Shows an Induction of Fatty Acids Metabolism in Healthy Adult Volunteers after Supplementation with Green Coffee (Coffea Robusta L.) Bean Extract. *Phytomedicine* **2018**, *38*, 74–83. https://doi.org/10.1016/j.phymed.2017.11.002.

(72) Tolun, A. A.; Zhang, H.; Il'yasova, D.; Sztáray, J.; Young, S. P.; Millington, D. S. Allantoin in Human Urine Quantified by Ultra-Performance Liquid Chromatography-Tandem Mass Spectrometry. *Anal Biochem* **2010**, *402* (2), 191–193. https://doi.org/10.1016/j.ab.2010.03.033.

(73) Pan, L.; Han, P.; Ma, S.; Peng, R.; Wang, C.; Kong, W.; Cong, L.; Fu, J.; Zhang, Z.; Yu, H.; Wang, Y.; Jiang, J. Abnormal Metabolism of Gut Microbiota Reveals the Possible 00Molecular Mechanism of Nephropathy Induced by Hyperuricemia. *Acta Pharmaceutica Sinica B* **2020**, *10* (2), 249–261. https://doi.org/10.1016/j.apsb.2019.10.007.

(74) Kieffer, D. A.; Piccolo, B. D.; Vaziri, N. D.; Liu, S.; Lau, W. L.; Khazaeli, M.; Nazertehrani, S.; Moore, M. E.; Marco, M. L.; Martin, R. J.; Adams, S. H. Resistant Starch Alters Gut Microbiome and Metabolomic Profiles Concurrent with Amelioration of Chronic Kidney Disease in Rats. *American Journal of Physiology-Renal Physiology* **2016**, *310* (9), F857–F871. https://doi.org/10.1152/ajprenal.00513.2015.

(75) Tabasi, M.; Eybpoosh, S.; Sadeghpour Heravi, F.; Siadat, S. D.; Mousavian, G.; Elyasinia, F.; Soroush, A.; Bouzari, S. Gut Microbiota and Serum Biomarker Analyses in Obese Patients Diagnosed with Diabetes and Hypothyroid Disorder. *Metab Syndr Relat Disord* **2021**, *19* (3), 144–151. https://doi.org/10.1089/met.2020.0119.

(76) Histamine Derived from Probiotic Lactobacillus Reuteri Suppresses TNF via Modulation of PKA and ERK Signaling. *PLOS ONE* **2012**, *7* (2), e31951. https://doi.org/10.1371/journal.pone.0031951.

(77) Clos-Garcia, M.; Andrés-Marin, N.; Fernández-Eulate, G.; Abecia, L.; Lavín, J. L.; van Liempd, S.; Cabrera, D.; Royo, F.; Valero, A.; Errazquin, N.; Vega, M. C. G.; Govillard, L.; Tackett, M. R.; Tejada, G.; Gónzalez, E.; Anguita, J.; Bujanda, L.; Orcasitas, A. M. C.; Aransay, A. M.; Maíz, O.; López de Munain, A.; Falcón-Pérez, J. M. Gut Microbiome and Serum Metabolome Analyses Identify Molecular Biomarkers and Altered Glutamate Metabolism in Fibromyalgia. *EBioMedicine* **2019**, *46*, 499–511. https://doi.org/10.1016/j.ebiom.2019.07.031.

(78) Yang, Y.; Bin, P.; Tao, S.; Zhu, G.; Wu, Z.; Cheng, W.; Ren, W.; Wei, H. Evaluation of the Mechanisms Underlying Amino Acid and Microbiota Interactions in Intestinal Infections Using Germ-Free Animals. *Infectious Microbes & Diseases* **2021**, *3* (2), 79–86. https://doi.org/10.1097/IM9.0000000000000060.

(79) Ranganathan, N.; Ranganathan, P.; Friedman, E. A.; Joseph, A.; Delano, B.; Goldfarb, D. S.; Tam, P.; Venketeshwer Rao, A.; Anteyi, E.; Guido Musso, C. Pilot Study of Probiotic Dietary Supplementation for Promoting Healthy Kidney Function in Patients with Chronic Kidney Disease. *Adv Therapy* **2010**, *27* (9), 634–647. https://doi.org/10.1007/s12325-010-0059-9.

(80) Koppe, L.; Mafra, D.; Fouque, D. Probiotics and Chronic Kidney Disease. *Kidney Int* **2015**, *88* (5), 958–966. https://doi.org/10.1038/ki.2015.255.

Table

	All	**Women**	**Men**
n (n, %)	21	15 (71.4)	6 (28.6)
Age (yrs ± SD)	45.9 ± 11.8	40.7 ± 10.8	52.8 ± 10.8
Heigh (m ± SD)	1.70 ± 0.09	1.65 ± 0.05	1.82 ± 0.07
Weight (kg ± SD)	66.5 ± 14.7	59.4 ± 8.3	84.3 ± 11.9
BMI (kg/m^2 ± SD)	22.7 ± 3.1	21.7 ± 2.8	25.4 ± 2.3

Table 1: Demographic characteristics of healthy adult volunteers enrolled in the study.

Serum Metabolites			
Acetate	Creatinine	Histidine	Phenylalanine
Acetone	Formate	Isoleucine	Proline
Alanine	Glycoprotein	Lactate	Pyruvate
Betaine	Glucose	Leucine	Threonine
Citrate	Glutamine	Lysine	Tyrosine
Creatine	Glycine	Mannose	Valine
Urine Metabolites			
Acetoacetic Acid	dTTP	Leucine	3-hydroxyisobutyric acid
Acetone	Formate	Lysine	3-hydroxymandelate
Alanine	Fumarate	Phenylacetylglutamine	3-hydroxypropanoic acid (HPHPA)
Allantoin	Glycine	Tartrate	4-hydroxyphenylacetate
Asparagine	Glucose	Trimethylamine-N-oxide (TMAO)	Unk (ppm range = 0.940 - 0.932, singlet)
Betaine	Hippurate	Trigonelline	Unk2 (ppm range = 2.910 - 2.886, singlet)
Citrate	Histidine	Tyrosine	Unk3 (ppm range = 2.790 - 2.781, singlet)

Creatine + Creatinine	Isoleucine	Valine	Unk4 (ppm range = 5.401 - 5.387, singlet)
Creatinine	Indoxyl-sulphate	1-methylnicotinamide	Sugar_unk (ppm range = 5.218 - 5.200)
Dimethylamine	Lactic acid + Threonine	2-hydroxyisobutyric acid	

Table 2: List of metabolites assigned and quantified in both serum and urine samples.

Metabolites	**Estimate**	**Lower**	**Upper**	**Dosage**
Tartrate	0.402	0.210	0.594	low
Valine	-0.036	-0.071	-0.001	low
3-hydroxyisobutyric acid	-0.040	-0.075	-0.004	low and high
4-hydroxyphenylacetate	-0.050	-0.090	-0.010	low and high
Isoleucine	-0.053	-0.088	-0.018	low
Allantoin	-0.087	-0.150	-0.025	high
Sugar_unk (ppm range = 5.218 - 5.200)	-0.127	-0.183	-0.071	low and high

Table 3: Average differences between the 30-th (middle Phase II) and the 10-th (middle Phase I) measurements for urinary metabolites log-quantifications (95% confidence intervals), net of other variables.

Metabolites	**Estimate**	**Lower**	**Upper**
Pyruvate	0.144	0.022	0.266

Histidine	-0.038	-0.070	-0.005
Glutamine	-0.048	-0.094	-0.002
Creatinine	-0.066	-0.123	-0.010

Table 4: Average differences between Phase II and Phase I for serum metabolites log-quantifications (with 95% confidence intervals), net of other variables.

Figure

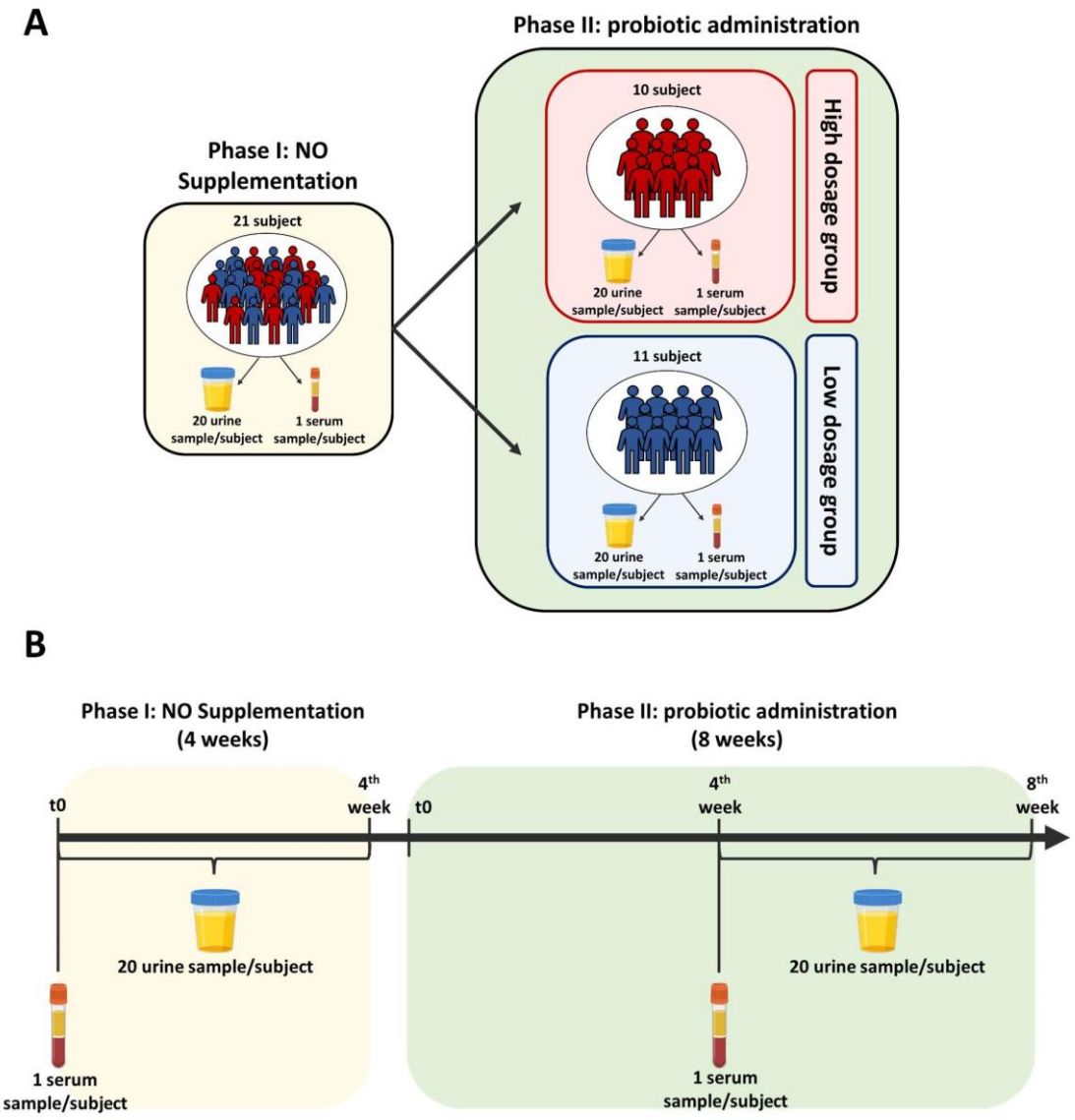


Figure 1: **Study design.** A) Experimental scheme of the two phases of the project; B) Temporal scheme of urine and serum sampling in Phase I and Phase II.

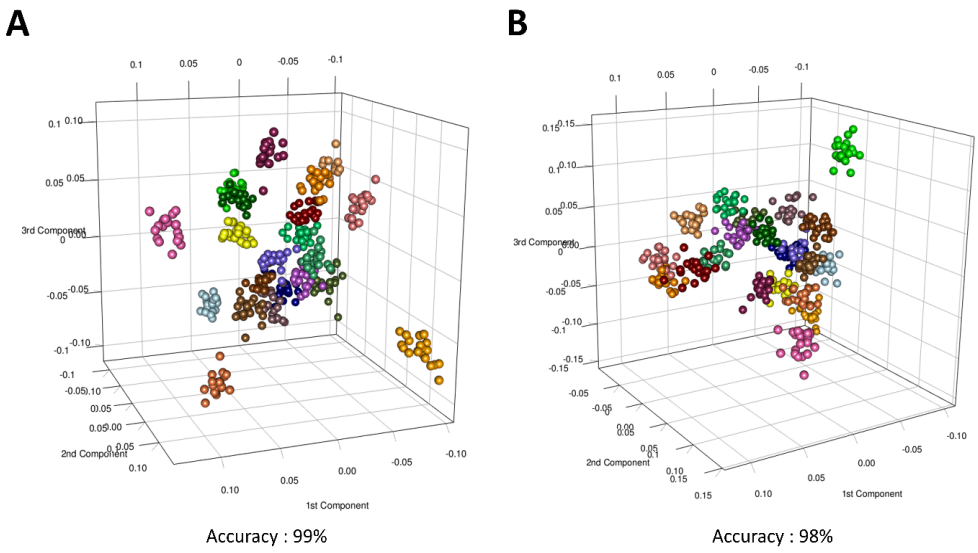


Figure 2: **Urinary subject-specific metabolic phenotype discrimination** in (A) Phase I; (B) Phase II. Each colour in the PCA-CA score plot represents a different healthy subject. At the bottom of the score plot the Accuracy of the model, expressed in percentage, is also reported.

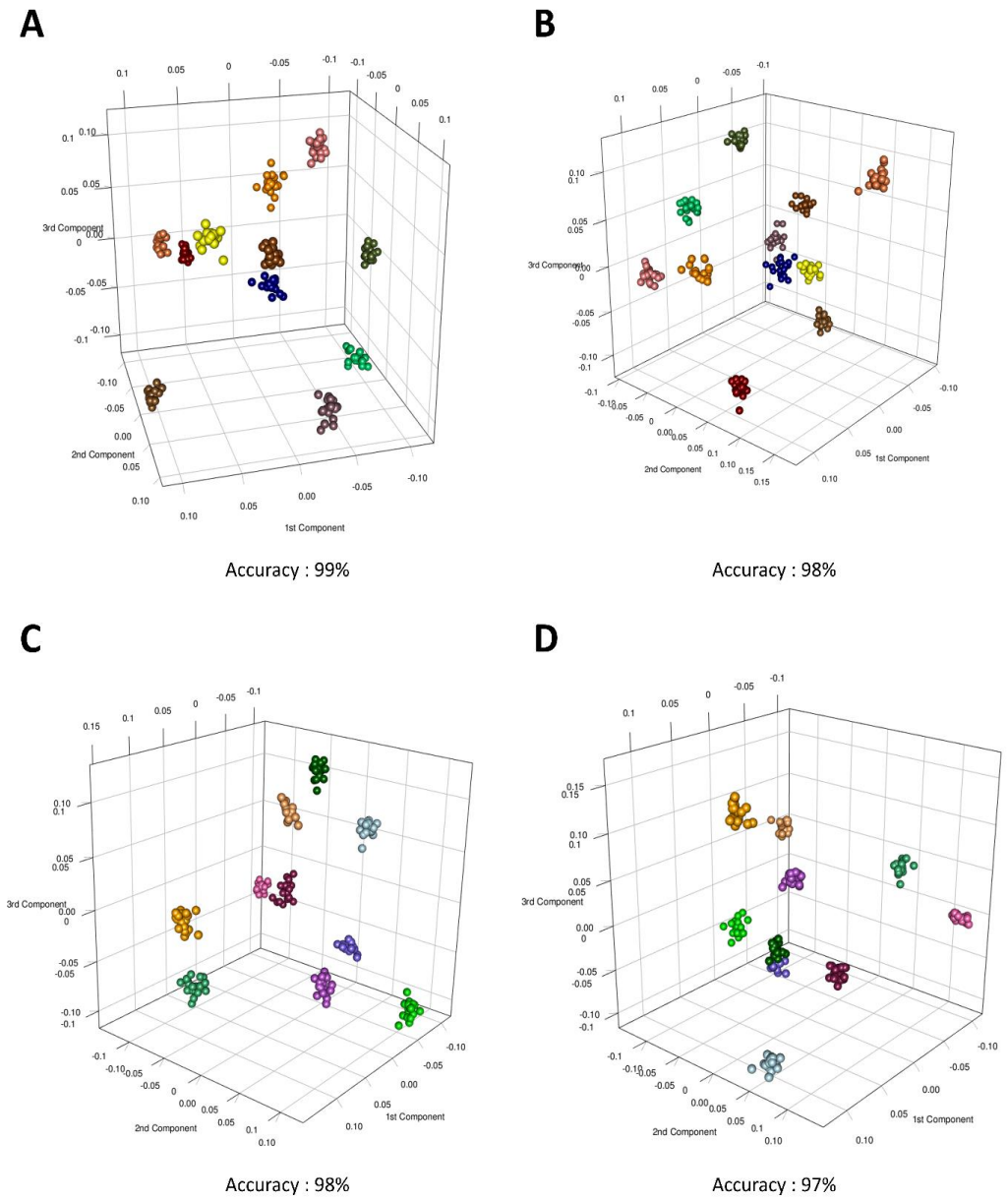


Figure 3: **Discrimination of urinary dose-dependent subject-specific metabolic phenotype** in (A) Phase I in subjects administered with a low dose of probiotics; (B) Phase II in subjects administrated with a low dose of probiotics; (C) Phase I in subjects administrated with a high dose of probiotics; (D) in subjects administered with a high dose of probiotics. Each colour in the PCA-CA score plot represents a different healthy subject. At the bottom of the score plot the Accuracy of the model, expressed in percentage, is also reported.

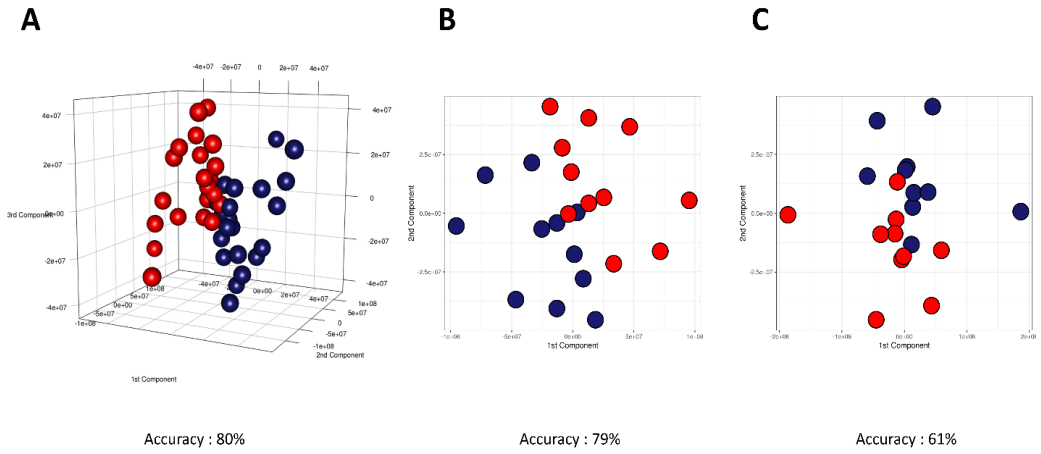


Figure 4: **Score plots of M-PLS discrimination between urine samples** collected (A) for all subjects during Phase I (blue dots) and Phase II (red dots); (B) for subjects administered with a low dose of probiotics during Phase I (blue dots) and Phase II (red dots); (C) for subjects administered with a high dose of probiotics during Phase I (blue dots) and Phase II (red dots). Discrimination accuracy values for the three pairwise comparisons were also reported. The median spectrum of each subject at every phase was calculated and used to build the MPLS models.

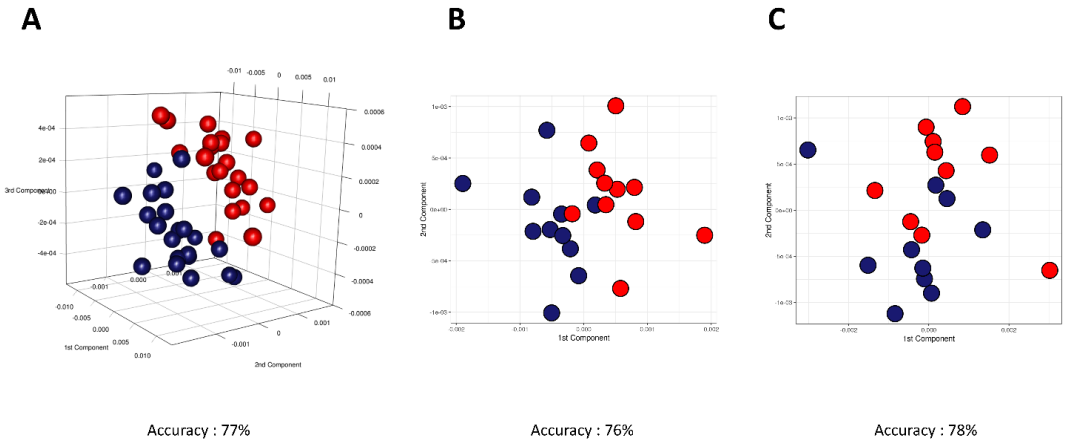


Figure 5: **Score plots of M-PLS discrimination between serum samples** collected (A) for all subjects during Phase I (blue dots) and Phase II (red dots); (B) for subjects administered with a low dose of probiotics during Phase I (blue dots) and Phase II (red dots); (C) for subjects administered with a high dose of probiotics during Phase I (blue dots) and Phase II (red dots). Discrimination accuracy values for the three pairwise comparisons were also reported. The

median spectrum of each subject in each phase was calculated and used to build the MPLS models.

4.1.5 Associations of plasma metabolites and lipoproteins with Rh and ABO blood systems in healthy subjects

Francesca Di Cesare[1], Leonardo Tenori[1,2], Claudio Luchinat[1,2], Edoardo Saccenti[3]

[1] Magnetic Resonance Center (CERM), and Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine (CIRMMP), University of Florence, Via Luigi Sacconi 6, 50019, Sesto Fiorentino, Firenze, Italy

[2] Department of Chemistry "Ugo Schiff", University of Florence, Via della Lastruccia 3, 50019, Sesto Fiorentino, Italy

[3] Laboratory of Systems and Synthetic Biology, Wageningen University & Research, Stippeneng 4, 6708 WE, Wageningen, the Netherlands.

Published

J. Proteome Res., **2022**,

doi: 10.1021/acs.jproteome.2c00375

Candidate's contribuitions: NMR data pre-processing, statistical analysis, interpretation of results, and writing the manuscript.

Journal of proteome research

pubs.acs.org/jpr

Article

Association of Plasma Metabolites and Lipoproteins with Rh and ABO Blood Systems in Healthy Subjects

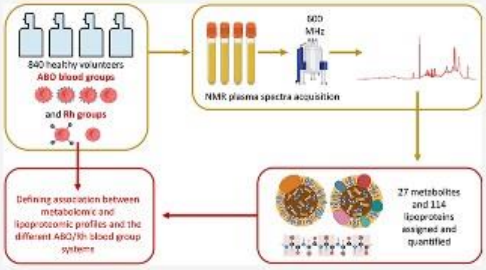
Francesca Di Cesare, Leonardo Tenori, Claudio Luchinat, and Edoardo Saccenti*

Cite This: https://doi.org/10.1021/acs.jproteome.2c00375

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: This study investigated the associations between the levels of 27 plasma metabolites, 114 lipoprotein parameters, determined using nuclear magnetic resonance spectroscopy, and the ABO blood groups and the Rhesus (Rh) blood system in a cohort of $n = 840$ Italian healthy blood donors of both sexes. We observed good multivariate discrimination between the metabolomic and lipoproteomic profiles of subjects with positive and negative Rh. In contrast, we did not observe significant discrimination for the ABO blood group pairwise comparisons, suggesting only slight metabolic differences between these group-specific metabolic profiles. We report univariate associations (P-value < 0.05) between the subfraction HDL1 related to Apo A1, the subfraction HDL2 related to cholesterol and phospholipids, and the particle number of LDL2 related to free cholesterol, cholesterol, phospholipids, and Apo B and the ABO blood groups; we observed association of the lipid main fraction LDL4 related to free cholesterol, triglycerides, and Apo B; creatine; the particle number of LDL5; the subfraction LDL5 related to Apo B; the particle number of LDL4; and the subfraction LDL4 related to Apo B with Rh blood factors. These results suggest blood group-dependent (re)shaping of lipoprotein metabolism in healthy subjects, which may provide relevant information to explain the differential susceptibility to certain diseases observed in different blood groups.



KEYWORDS: *metabolomics, lipoproteomics, robust linear models, LDL, HDL*

1. INTRODUCTION

The most important blood group systems in humans are the ABO and Rh (Rhesus) groups, which consist of carbohydrate moieties, also named human histo-blood group antigens, at the extracellular surface of red blood cell (RBC) membranes.[1] The human ABO locus is located on chromosome 9 (9q34.2) and has three main allelic forms: (1) allele A that encodes a glycosyltransferase which catalyzes the conversion of the H antigen precursor to the A antigen characterized by the covalent linkage of *N*-acetylgalactosamine to the O antigen; (2) allele B encoding a glycosyltransferase which catalyzes the conversion of the H antigen precursor to the B antigen characterized by the covalent linkage of D-galactose to the O antigen; and (3) allele O that encodes an enzyme with no function, leaving the underlying H antigen precursor structurally unchanged.[2−4]

The Rh system, also known as the D antigen, is based, like the ABO system, on the absence or presence of an antigen on the RBC membrane surfaces; if the antigen D is present, the individual is recognized as Rh positive (Rh^+), and, if absent, the individual is identified as Rh negative (Rh^-).[5]

The ABO and Rh systems, which play a fundamental role in transfusion medicine and hematopoietic transplantation, have been associated to the pathogenesis and pathophysiology of various human diseases, such as cardiovascular[6−8] and oncological diseases,[6,9] and also have a role in the susceptibility to microorganism, viral, and parasitic infections.[10−13]

While there has been renewed interest on potential associations between ABO/Rh blood groups and the development of specific pathologies,[6,7,14,15] little is known about the association between the blood metabolomic and lipoproteomic profiles and ABO and Rh blood group systems. The investigation of the existence of blood metabolomic and/or lipoproteomic profiles specific to certain ABO and Rh groups could provide relevant information in the quest for potential blood-specific fingerprints associated with the predispositions to common and chronic pathologies. In this study, we explored the association between the ABO and Rh blood systems and the levels of 27 free circulating plasma metabolites and 114

Received: June 28, 2022

Table 1. Demographic and Clinic Characteristics of the Study Population Stratified by ABO Blood Groups

		ABO blood group				
	overall (n = 840)	non-O (n = 449, 53.5%)	O (n = 391, 46.6%)	A (n = 330, 39.3%)	AB (n = 41, 4.9%)	B (n = 78, 9.3%)
Demographic Parameters						
age (years)	40.87 ± 11.0	40.9 ± 10.6	40.9 ± 11.4	41.0 ± 10.6	41.1 ± 10.4	40.2 ± 11.1
men (n, %)	658, 78.3%	357, 54.3%	301, 45.7%	264, 40.1%	36, 5.5%	57, 8.7%
women (n, %)	182, 21.7%	92, 50.6%	90, 49.5%	66, 36.3%	5, 2.8%	21, 11.5%
Clinical Parameters						
maximum pressure (mmHg)	123.2 ± 10.7	123.37 ± 10.9	123.1 ± 10.5	123.1 ± 10.6	125.4 ± 13.8	112.9 ± 10.5
minimum pressure (mmHg)	80.6 ± 7.0	80.8 ± 6.9	80.4 ± 7.2	80.5 ± 6.7	82.8 ± 7.5	80.8 ± 7.3
heart rate (bpm)	70.1 ± 5.9	70.2 ± 5.5	70.0 ± 6.3	70.2 ± 5.6	69.3 ± 6.3	71.0 ± 4.7
glycemia (mg/dL)	89.7 ± 11.7	89.8 ± 11.9	89.6 ± 11.5	89.5 ± 12.0	93.3 ± 13.6	89.3 ± 10.5
Chol (mg/dL)	204.1 ± 35.2	205.3 ± 35.2	202.8 ± 35.2	206.0 ± 35.5	208.7 ± 32.9	200.5 ± 35.3
triglycerides (TGs) (mg/dL)	102.3 ± 55.4	104.9 ± 61.6	99.4 ± 47.3	103.8 ± 58.5	111.8 ± 63.3	105.8 ± 73.0
alanine aminotransferase (Unit/L)	23.6 ± 12.1	23.2 ± 13.0	23.1 ± 11.1	23.7 ± 13.3	24.5 ± 12.3	24.5 ± 11.7
hematocrit (HCT) (dL/dL(%))	43.5 ± 3.2	43.3 ± 2.6	43.6 ± 3.7	43.2 ± 2.6	44.0 ± 2.3	43.6 ± 2.9
white blood cells (WBCs)($10^3/\mu$L)	6.2 ± 4.1	6.3 ± 5.2	6.1 ± 2.1	6.1 ± 1.5	5.6 ± 1.2	7.5 ± 12.2
red blood cells (RBCs) ($10^6/\mu$L)	5.0 ± 0.5	5.0 ± 0.4	5.1 ± 0.6	5.0 ± 0.4	5.1 ± 0.3	5.1 ± 0.4
hemoglobin (g/dL)	15.0 ± 1.4	14.9 ± 1.1	15.0 ± 1.8	14.9 ± 1.1	15.3 ± 1.0	15.1 ± 1.2
mean corpuscular volume (MCV) (fL)	86.8 ± 4.0	86.9 ± 3.9	86.6 ± 4.0	87.1 ± 3.8	86.5 ± 3.3	86.2 ± 4.4
platelets ($10^3/\mu$L)	227.7 ± 48.3	227.8 ± 49.3	227.5 ± 47.2	228.3 ± 49.9	221.2 ± 42.5	229.2 ± 50.2

Table 2. Demographic and Clinic Characteristics of the Study Population Stratified by the Rh Blood Group System

		Rh blood group system	
	overall (n = 840)	Rh^+ (n = 710, 84.5%)	Rh^- (n = 130, 15.5%)
Demographic Parameters			
age (years)	40.9 ± 11.0	40.7 ± 10.3	41.7 ± 11.1
men (n, %)	658, 78.3%	560, 85.1%	98, 14.9%
women (n, %)	182, 21.7%	150, 82.4%	32, 17.6%
Clinic Parameters			
maximum pressure (mmHg)	123.2 ± 10.7	123.0 ± 10.9	124.4 ± 10.7
minimum pressure (mmHg)	80.6 ± 7.0	80.5 ± 7.1	81.0 ± 7.0
heart rate (bpm)	70.1 ± 5.9	70.4 ± 6.6	69.0 ± 5.7
glycemia (mg/dL)	89.7 ± 11.7	90.1 ± 10.9	87.9 ± 11.8
Chol (mg/dL)	204.1 ± 35.2	204.3 ± 32.3	203.0 ± 35.8
triglycerides (TGs) (mg/dL)	102.3 ± 55.4	102.3 ± 47.6	102.5 ± 56.8
alanine aminotransferase (Unit/L)	23.6 ± 12.1	23.8 ± 12.1	22.2 ± 12.1
hematocrit (HCT) (dL/dL(%))	43.5 ± 3.2	43.5 ± 2.8	43.2 ± 3.2
white blood cells (WBCs)($10^3/\mu$L)	6.2 ± 4.1	6.3 ± 1.4	5.8 ± 4.4
red blood cells (RBCs) ($10^6/\mu$L)	5.0 ± 0.5	5.0 ± 0.4	5.0 ± 0.5
hemoglobin (g/dL)	15.0 ± 1.4	15.0 ± 1.2	14.8 ± 1.5
mean corpuscular volume (MCV) (fL)	86.8 ± 4.0	86.9 ± 4.0	86.0 ± 3.9
platelets ($10^3/\mu$L)	227.7 ± 48.3	229.4 ± 44.1	218.1 ± 48.8

lipoprotein concentrations and associated parameters, measured using nuclear magnetic resonance (NMR) spectroscopy, in a cohort of 840 healthy blood donor volunteers, using both multivariate and univariate approaches. We report the existence of some weak but significant associations mostly concerning high-density lipoproteins (HDL), low-density lipoproteins (LDL), and apolipoproteins, with blood group systems.

2. MATERIALS AND METHODS

2.1. Study Population

The study group consists of 840 adult healthy blood donors, with an overall age range from 19 to 65 years (658 men, with average age of 40.6 ± 10.7 years, and 182 women, with average age of 41.9 ± 12.0 years). Demographic and clinic characteristics are reported in Tables 1 and 2, respectively. Blood donors were enrolled in collaboration with the Tuscany section of the Italian Association of Blood Donors (AVIS) in the Transfusion Service of the Pistoia Hospital (Ospedale del Ceppo, AUSL 3—Pistoia, Tuscany, Italy).

According to the Italian guidelines for blood donation (Annex III of the Decree of the Italian Ministry of Health dated 2 November 2015),[16] blood donors must not have (had) infectious, chronic, and/or common diseases before donation, surgery within 3 months before donation, endoscopic exams within 4 months before donation, current menstruation, pregnancy within 12 months before donation, and abortion within 4 months before donation; they had not participated in sport activity within 24 h before donation; and they had not taken drugs within 1 week before donation. For further details, see previous publications.[17−19]

https://doi.org/10.1021/acs.jproteome.2c00375
J. Proteome Res. XXXX, XXX, XXX−XXX

Table 3. List of Metabolites and Lipoprotein Fractions and Subfractions Analyzed[a]

metabolites	lipid fractions and subfractions				
3-hydroxybutyrate	MP TG	LMF Free Chol−IDL	Subfr PL−VLDL 2	Subfr Apo B−LDL 1	Subfr Apo A2−HDL 3
acetate	MP Chol	LMF Free Chol−LDL	Subfr PL−VLDL 3	Subfr Apo B−LDL 2	Subfr Apo A2−HDL 4
alanine	MP LDL−Chol	LMF Free Chol−HDL	Subfr PL−VLDL 4	Subfr Apo B−LDL 3	
arginine + lysine	MP HDL−Chol	LMF PL−VLDL	Subfr PL−VLDL 5	Subfr Apo B−LDL 4	
adenosine nucleotide + inosine monophosphate	MP Apo A1	LMF PL−IDL	Subfr TG−LDL 1	Subfr Apo B−LDL 5	
citrate	MP Apo A2	LMF PL−LDL	Subfr TG−LDL 2	Subfr Apo B−LDL 6	
creatine	MP Apo B100	LMF PL−HDL	Subfr TG−LDL 3	Subfr TG−HDL 1	
creatinine	CFLDL−Chol/HDL−Chol	LMF Apo A1−HDL	Subfr TG−LDL 4	Subfr TG−HDL 2	
formate	CFApo A1/Apo B100	LMF Apo A2−HDL	Subfr TG−LDL 5	Subfr TG−HDL 3	
fumarate	PN Total PN	LMF Apo B−VLDL	Subfr TG−LDL 6	Subfr TG−HDL 4	
glucose	PN VLDL	LMF Apo B−IDL	Subfr Chol−LDL 1	Subfr Chol−HDL 1	
glutamate	PN IDL	LMF Apo B−LDL	Subfr Chol−LDL 2	Subfr Chol−HDL 2	
glutamine	PN LDL	Subfr TG−VLDL 1	Subfr Chol−LDL 3	Subfr Chol−HDL 3	
glycine	PN LDL 1	Subfr TG−VLDL 2	Subfr Chol−LDL 4	Subfr Chol−HDL 4	
histidine	PN LDL 2	Subfr TG−VLDL 3	Subfr Chol−LDL 5	Subfr Free Chol−HDL 1	
isoleucine	PN LDL 3	Subfr TG−VLDL 4	Subfr Chol−LDL 6	Subfr Free Chol−HDL 2	
lactate	PN LDL 4	Subfr TG−VLDL 5	Subfr Free Chol−LDL 1	Subfr Free Chol−HDL 3	
leucine	PN LDL 5	Subfr Chol−VLDL 1	Subfr Free Chol−LDL 2	Subfr Free Chol−HDL 4	
mannose	PN LDL 6	Subfr Chol−VLDL 2	Subfr Free Chol−LDL 3	Subfr PL−HDL 1	
methionine	LMF TG−VLDL	Subfr Chol−VLDL 3	Subfr Free Chol−LDL 4	Subfr PL−HDL 2	
phenylalanine	LMF TG−IDL	Subfr Chol−VLDL 4	Subfr Free Chol−LDL 5	Subfr PL−HDL 3	
proline	LMF TG−LDL	Subfr Chol−VLDL 5	Subfr Free Chol−LDL 6	Subfr PL−HDL 4	
pyruvate	LMF TG−HDL	Subfr Free Chol−VLDL 1	Subfr PL−LDL 1	Subfr Apo A1−HDL 1	
tyrosine	LMF Chol−VLDL	Subfr Free Chol−VLDL 2	Subfr PL−LDL 2	Subfr Apo A1−HDL 2	
unknown 1	LMF Chol−IDL	Subfr Free Chol−VLDL 3	Subfr PL−LDL 3	Subfr Apo A1−HDL 3	
unknown 2	LMF Chol−LDL	Subfr Free Chol−VLDL 4	Subfr PL−LDL 4	Subfr Apo A1−HDL 4	
valine	LMF Chol−HDL	Subfr Free Chol−VLDL 5	Subfr PL−LDL 5	Subfr Apo A2−HDL 1	
	LMF Free Chol−VLDL	Subfr PL−VLDL 1	Subfr PL−LDL 6	Subfr Apo A2−HDL 2	

[a]Abbreviations used: Subfr, subfraction; Chol, cholesterol; MP, main parameter; CF, calculated figure; TG, triglycerides; LMF, lipoprotein main fraction; PL, phospholipids; and PN, particles number.

2.2. Ethics Statement

The study adheres to the directives of the Declaration of Helsinki (1964).

2.3. Sample Preparation and ABO and Rh Determination

Plasma samples were obtained after overnight fasting and, after collection, were stored at −80 °C and handled according to standard operating procedures.[20] The ABO and Rh blood groups were determined by standard procedures using agglutination techniques.[21]

2.4. NMR Analysis

The mono-dimensional nuclear Overhauser effect spectroscopy (NOESY) ^{1}H spectra of plasma samples were acquired using a Bruker 600 MHz spectrometer (Bruker BioSpin s.r.l., Germany), operating at 600.13 MHz. For more details about NMR sample preparation, acquisition, and spectral processing, we refer the reader to the original publications.[17,22]

Twenty-seven (27) metabolites were assigned and identified using in-house software developed based on standard line-shape analysis methods and with the help of matching routines of AMIX 7.3.2 (Bruker BioSpin) in combination with the BBIOREFCODE (Version 2−0−0; Bruker BioSpin) reference database and published literature (when available). In the Supporting Information (Figure S1), the assignment of a plasma spectrum is shown. The relative concentration of each metabolite was calculated by integrating the signals in the spectra; 114 lipoprotein fractions and subfractions were assigned and quantified using the AVANCE IVDr [Clinical Screening and In Vitro Diagnostics (IVD) research with B.I. Methods, Bruker BioSpin].[23] The direct integration of NMR

https://doi.org/10.1021/acs.jproteome.2c00375
J. Proteome Res. XXXX, XXX, XXX−XXX

signals was carried out. A list of metabolites and lipoprotein fractions and subfractions and lipids is given in Table 3.

2.4.1. Data Pre-processing. Only clinical variables with less than 20% missing data were considered; missing data were imputed using a Random Forest (RF) approach as implemented in R package missForest,[24] using the default parameters. Variables that have been imputed and the percentage of imputation for that specific variable are glycemia (19.6%), maximum pressure (3.7%), minimum pressure (3.8%), heart rate (4.3%), cholesterol (Chol, 17%), triglycerides (TG, 17.8%), alanine aminotransferase (ALT, 0.1%), and hematocrit (HCT, 0.1%).

All metabolite concentrations and lipoproteomic parameters were square-root-transformed before analysis to adjust for heteroscedasticity.[25]

2.5. Statistical Analysis

2.5.1. Two-Proportion Z-Test. The observed percentages of ABO and Rh blood groups in the study population were compared with those observed in the general Italian population[26,27] using a two-proportion *Z*-test.[28] Results are reported with 95% confidence intervals (CIs).

2.5.2. Exploratory Data Analysis. Principal component analysis (PCA)[29,30] was used to explore data patterns. Data was scaled to unit variance before analysis.

2.5.3. Predictive Modeling. The Random Forest (RF) algorithm[31−33] was employed for pairwise classification of metabolite and lipoproteomic profiles of subjects with different ABO and Rh blood groups. The following comparisons were performed: A *versus* AB, AB *versus* B, A *versus* B, A *versus* O, AB *versus* O, B *versus* O blood groups, and Rh^- *versus* Rh^+ groups.

To reduce the potential bias due to the unbalanced number of subjects per group, we implemented a resampling scheme with $K = 100$ resampling, considering the sex distribution. In this procedure, we selected, for each pairwise comparison (*i.e.*, Rh^+ *vs* Rh^-, A *vs* AB, etc), an equal number of individuals, stratified by sex; in more detail, 85% of the subjects per balanced group were randomly selected at each iteration step; basically, for each comparison, 100 different RF models were built.

The model quality statistics, including the accuracy, sensitivity, specificity, and area under the curve (AUC), are given as average values over the $K = 100$ models with the corresponding 95%. Quality statistics were calculated according to standard definitions.[34]

2.5.4. Permutation Test. The statistical significance of the RF classification models was determined with a permutation test using $M = 1000$ permutation. *P*-values were calculated by comparing the value $model_0$, obtained from the original, and non-permuted data with the values $model_1$, $model_2$, ..., $model_M$ obtained from the *M*-times permutation-test. The *P*-value for a specific quality measure is calculated as follows[35]

$$P - \text{value} = \frac{1 + (|D_{\text{perm}}| \geq |\text{measure model}_0|)}{M} \quad (1)$$

where $|D_{\text{perm}}|$ is the number of permuted models for which a given quality measure is larger or equal to the quality measure from the original (non-permuted) model ($model_0$).

2.5.5. Robust Linear Regression. The association between plasma metabolites, lipids, and lipoproteins and the ABO and Rh groups was determined by linear regression models,[36,37] according to the following formula

$$y_i = b_0 + b_1(\text{group}) + b_2(\text{age}) + b_3(\text{sex}) + \varepsilon_i \quad (2)$$

where y_i is the abundance/concentration of the *i*th molecular feature (metabolite or lipid/lipoprotein), b_i is the estimated regression coefficient that quantifies the association between the ABO/Rh blood groups, adjusted for age and sex, and the molecular features, and ε_i is the residual part not accounted for by the other terms. To reduce the influence of outliers and high leverage points on the regression solutions, we used a robust version for the linear regression, where the fitting is done by iterated re-weighted least squares.[36,37] For each model, post-hoc pairwise comparisons among ABO blood group systems were also performed using the estimated marginal means.[38]

The Benjamini−Hochberg method[39] was used to correct for multiple testing.

2.6. Software

All calculations were performed in R (version 4.0.3).[40] The function "missForest", implemented in the missForest package, was used to impute the missing data.[24] RF models were built using the "randomForest" function, implemented in the R package RF,[31,41] growing a decision forest composed of 1000 trees, using default parameters. To estimate the significance of importance metrics for the RF models by the 1000-times permutation of the response variable, the "importance" function, implemented in the R package rfPermute, was used.[42] The function "rlm", implemented in the MASS R package, was used to perform the robust fitting of linear models.[43] Default parameters were used. The function "emmeans", implemented in the emmeans R package,[38,44] was used to compute comparisons among specified factors and/or factor combinations in the linear models.

2.7. Data Availability

NMR spectra and associated clinical data are available in the MetaboLights repository[45] (http://www.ebi.ac.uk/metabolights) with accession number MTBLS147. Data on lipoprotein fractions and subfractions are available at the NIH Common Fund's National Metabolomics Data Repository (NMDR) the Metabolomics Workbench (https://www.metabolomicsworkbench.org), where it has been assigned Project ID ST001785. The data can be accessed directly *via* its Project DOI: http://dx.doi.org/10.21228/M8470J.

3. RESULTS AND DISCUSSION

3.1. Distribution of ABO and Rh Blood Group Systems in Tuscany

Clinical and demographic characteristics of the study subjects, divided by ABO and Rh groups, are given in Tables 1 and 2, respectively. The list of the metabolites and lipoproteins assigned and quantified is reported in Table 3.

The distribution of the ABO blood groups among subjects (all original from the Pistoia area in Tuscany, Italy), as shown in Table 4, is in line with the distribution of the general population living in Italy.[27] In particular, the distribution of A, B, and O groups in the study cohort is similar to the distribution observed in Italy, except for the AB group (*P*-value = 0.006).

The proportion of Rh^+ and Rh^- subjects is not different from the general Italian population (see Table 4).[26]

https://doi.org/10.1021/acs.jproteome.2c00375
J. Proteome Res. XXXX, XXX, XXX−XXX

Table 4. Results of the Two-Proportion Z-Test of ABO Blood Groups and Rh Blood Group System[a]

ABO blood groups	number of individuals (*n*)	ABO distribution (%)	95% CI	ABO distribution in Italy (%)	adjusted *P*-value
A	330	39.3	36.0−43.0	41.0	0.62
AB	41	4.9	4.0−7.0	3.0	0.006
B	78	9.3	7.0−11.0	11.0	0.34
O	391	46.6	43.0−50.0	46.0	0.75
Rh blood system	number of individuals (*n*)	Rh distribution (%)	95% CI	Rh distribution in Italy (%)	adjusted *P*-Value
Rh^+	710	84.5	82.0−87.0	85.0	0.74
Rh^-	130	15.5	13.0−18.0	15.0	0.74

[a]In the table, the number of individuals per blood group, the ABO distribution (%) and the Rh blood group system distribution (%) of our population, the 95% CI, the ABO distribution (%), and the Rh blood group system distribution (%) in Italy, and the adjusted *P*-value of the two-proportion *Z*-test are reported.

3.2. Exploration and Discrimination of Metabolites, Lipids, and Lipoproteins Associated with ABO and Rh Blood Groups

To explore comprehensively the metabolomic and lipoproteomic profiles (consisting of 27 metabolites and 114 lipoproteins) associated with the ABO and Rh blood group systems, we applied PCA on the $n = 840$ plasma samples. The PCA score plot in Figure 1a shows no clear separation among the A, AB, B, and O groups, suggesting that metabolic differences are too subtle to be resolved using an unsupervised multivariate approach. A similar lack of separation can be observed in the case of Rh^+ and Rh^- profiles (Figure 1b).

RF classification was applied to investigate whether subjects with different ABO and Rh blood group systems could be discriminated from the metabolic and lipoproteomic profiles. The results of the RF classification for ABO and Rh blood groups are given in Tables 5 and 6, respectively. Overall, we obtained extremely weak classification models to discriminate between the different ABO groups: A *versus* AB groups (AUC = 0.562, *P*-value = 0.04), A *versus* B groups (AUC = 0.514, *P*-value = 0.03), AB *versus* B groups (AUC = 0.535, *P*-value = 0.03), A *versus* O groups (AUC = 0.557, *P*-value = 0.01), and AB *versus* O groups (AUC = 0.568, *P*-value = 0.03). A non-significant predictive model (AUC = 0.503, *P*-value = 0.09) was obtained for the discrimination of B *versus* O groups (see Table 5). These results suggest that the metabolic and lipoproteomic profiles, as a whole, may be weakly associated with the ABO groups; why this is the case is not clear: a lack of association may be a consequence of the limited sample size, or relevant associations may be limited to a few metabolites and/or lipoprotein/lipid features.

In contrast, subjects with different Rh blood groups can be easily and accurately discriminated on the basis of their metabolite and lipoproteomic profiles (AUC = 0.808, *P*-value = 0.02) (see Table 6 and Figure 2). By evaluating the RF important variables, as reported in Figure 3, we observed that the subfraction LDL5 related to Apo B (Subfr. Apo B−LDL5), the subfraction LDL4 related to Apo B (Subfr. Apo B−LDL4), the subfraction HDL4 related to Apo A1 (Subfr. Apo A1−HDL4) and Apo A2 (Subfr. Apo A2−HDL4), the subfraction HDL2 related to cholesterol (Subfr. Chol−HDL2), and the lipid main fraction LDL related to Apo B (LMF Apo B−LDL)

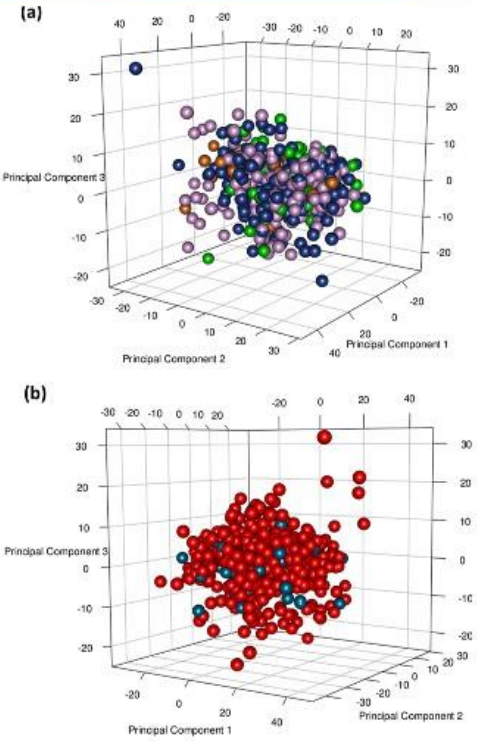


Figure 1. (a) PCA model score plot [PC1 (11.4%) *versus* PC2 (9.2%) *versus* PC3 (5.9%)]. Each dot represents a single metabolic profile with colors denoting different groups of patients: $n = 330$, A blood group subjects (violet dots); $n = 41$, AB blood group subjects (dark orange dots); $n = 78$, B blood group subjects (green dots); and $n = 391$, O blood group subjects (dark blue dots). (b) PCA model score plot [PC1 (11.4%) *versus* PC2 (9.2%) *versus* PC3 (5.9%)]. Each dot represents a single metabolic profile colored by the different groups of patients: $n = 710$, Rh^+ blood group subjects (red dots); and $n = 130$, Rh^- blood group subjects (light blue dots).

are the most relevant and significant variables in the model discriminating Rh^+ with respect to Rh^- blood groups.

3.3. Association of ABO Blood Groups to Metabolites and Lipids

The association between the levels of circulating plasma metabolites, lipoproteins, and ABO and Rh groups was assessed by means of robust linear regression, correcting for sex and age. We observed significant association (*P*-value < 0.05) for 8 out of 114 lipoprotein fractions and subfractions with ABO groups; no significant association remains after correction for multiple testing (see Table 7). Since correction for multiple testing increases the risk of false-negative, especially in the case where (possibly) weak associations are tested on a large number of variables, we took an honest and pragmatic approach, presenting both corrected and uncorrected *P*-values and discussing the biological implications of the results for which *P*-values were significant before

https://doi.org/10.1021/acs.jproteome.2c00375
J. Proteome Res. XXXX, XXX, XXX−XXX

Table 5. Mean Values of Accuracy, Specificity, Sensitivity, and AUC of RF Models Built Comparing the ABO Blood Groups, A *versus* AB Subjects, A *versus* B Subjects, A *versus* O Subjects, AB *versus* O Subjects, and B *versus* O Subjects

	mean accuracy % (95% CI and *P*-value)	mean specificity % (95% CI and *P*-value)	mean sensitivity % (95% CI and *P*-value)	AUC (95% CI and *P*-value)
A *vs* AB	55.1 (54.9−55.4 and 0.01)	52.1 (51.1−52.9 and 0.03)	55.5 (56.3−55.7 and 0.04)	0.562 (0.558−0.565 and 0.04)
A *vs* B	53.82 (53.0−53.5 and 0.03)	49.9 (49.7−50.4 and 0.03)	54.0 (53.3−54.3 and 0.01)	0.514 (0.513−0.519 and 0.03)
AB *vs* B	50.2 (49.7−50.7 and 0.04)	49.4 (48.8−50.1 and 0.03)	51.7 (51.2−52.6 and 0.04)	0.535 (0.530−0.537 and 0.03)
A *vs* O	54.9 (54.6−55.1 and 0.01)	55.9 (55.5−56.2 and 0.02)	53.6 (53.2−54.1 and 0.01)	0.557 (0.555−0.559 and 0.01)
AB *vs* O	55.2 (55.3−55.7 and 0.02)	55.8 (55.6−56.0 and 0.02)	53.1 (52.2−54.0 and 0.05)	0.568 (0.564−0.572 and 0.03)
B *vs* O	45.1 (44.9−45.3 and 0.10)	45.6 (45.4−45.8 and 0.09)	42.4 (41.6−43.1 and 0.12)	0.503 (0.500−0.507 and 0.09)

Table 6. Mean Values of Accuracy, Specificity, Sensitivity, and AUC of RF Models Built Comparing the Rh Blood Groups

	mean accuracy % (95% CI and *P*-value)	mean specificity % (95% CI and *P*-value)	mean sensitivity % (95% CI and *P*-value)	AUC (95% CI and *P*-value)
Rh^- *vs* Rh^+ blood groups	77.3 (77.1−77.5 and 0.01)	76.3 (76.2−76.5 and 0.02)	82.6 (82.1−83.6 and 0.01)	0.808 (0.808−0.810 and 0.02)

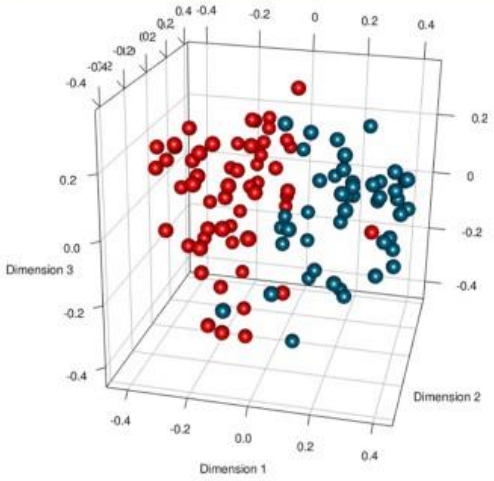


Figure 2. Balanced RF model score plot, taking into account the sexual dimorphism distribution. Each dot represents a single metabolic profile with colors denoting different groups of subjects. n = 51, randomly selected Rh^+ blood group subjects stratified by sex (red dots); and n = 51, Rh^- blood group subjects stratified by sex (light blue dots).

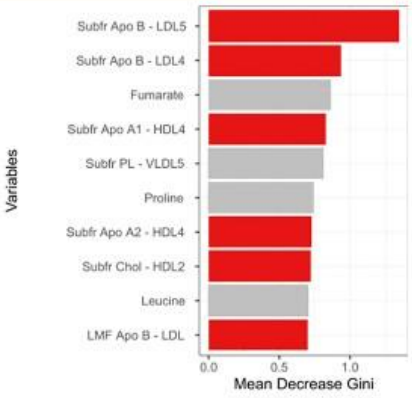


Figure 3. Importance of metabolites and lipoproteins in the RF model for the classification of Rh^+ and Rh^- subjects calculated using the mean decrease Gini index. Gray bars correspond to no significant variables. Red bars correspond to statistically significant (false discovery rate (FDR)-adjusted P-value < 0.05) important variables more representative in the Rh^+ and Rh^- comparison obtained with the permutation test. Only variables with a mean decrease Gini score >0.7 are shown.

correction. Using robust linear modeling, we observed that the subfractions of HDL (in particular, HDL1 with a density of 1.063−1.100 kg/L and HDL2 with a density of 1.100−1.112 kg/L) and the subfraction of LDL (in particular, LDL2 with a density of 1.031−1.034 kg/L) turned out to be relevant in ABO lipidic and lipoproteomic differences. We observed non-statistically significant associations between the particle number of LDL2 (PN LDL2), the subfraction HDL1 related to Apo A1 (Subfr. Apo A1−HDL1), the subfraction HDL2 related to cholesterol (Subfr. Chol−HDL2), phospholipids (Subfr. PL−HDL2), the subfraction LDL2 related to free cholesterol (Subfr. Free Chol−LDL2), cholesterol (Subfr. Chol−LDL2), phospholipids (Subfr. PL−LDL2), and Apo B (Subfr. Apo B−LDL2) and the ABO groups. A post-hoc test on the 8 lipids and lipoproteins poorly associated with the ABO groups was also performed (see Supporting Information Table 1), highlighting that the non-statistically differences exist mainly between the A and AB, and the AB and O groups. No differences were observed between the B and O, and the A and O groups.

It has been shown in both experimental and clinical studies that higher plasma levels of LDL and cholesterol in non-O blood groups (A, AB, B) influence the susceptibility of these groups to cardiovascular diseases, while in the O blood group, higher levels of HDL tend to play a protective role in these systemic pathologies,[46−50] but the molecular mechanisms by

https://doi.org/10.1021/acs.jproteome.2c00375
J. Proteome Res. XXXX, XXX, XXX−XXX

Table 7. Robust Linear Regression Models Were Performed on Serum Metabolites and Lipids of the ABO Blood Groups[a]

compound name	*P*-value	FDR *P*-value
Subfr Apo A1−HDL1	0.008	0.11
Subfr Chol−HDL2	0.01	0.11
Subfr PL−HDL2	0.02	0.11
Subfr Free Chol−LDL2	0.04	0.16
Subfr Chol−LDL2	0.04	0.16
Subfr PL−LDL2	0.04	0.16
Subfr ApoB−LDL2	0.04	0.16
PN LDL2	0.04	0.33

[a]Only molecular compounds with a *P*-value < 0.05 were reported; the FDR method was used for multiple testing correction. Models were adjusted for age and sex. Abbreviations used: Subfr, subfractions; Chol, cholesterol; PL, phospholipids; and PN, particle number.

which these group-specific pathologies are influenced have not yet been elucidated.

Concerning the role played by apolipoproteins (Apo), especially Apo B, there is evidence that higher numbers of RBC-bound Apo B in the O blood group compared with the non-O blood groups are associated with an atheroprotective effect and the reduction of the risk of developing vascular diseases (*i.e.*, venous thromboembolism, ischemic stroke, peripheral vascular thrombosis, and so forth).[46,51,52]

3.4. Association of Rh Blood Groups to Metabolites and Lipids

We observed significant associations with the Rh groups of 7 out of 114 lipoprotein fractions and subfractions and of 1 out of 27 metabolites (*P*-value < 0.05); after the correction for multiple testing, the particle number of LDL5 (PN LDL5) and the subfraction LDL5 related to Apo B (Subfr. Apo B−LDL5) were still significantly associated with Rh blood groups, as shown in Table 8. We observed associations (*P*-value < 0.05)

Table 8. Robust Linear Regression Models Were Performed on Serum Metabolites and Lipids of the Rh Blood Group System[a]

compound name	*P*-value	FDR *P*-value
PN LDL5	0.0008	0.01
Subfr Apo B−LDL5	0.0008	0.02
PN LDL4	0.008	0.05
Subfr Apo B−LDL4	0.008	0.05
Subfr Free Chol−LDL4	0.01	0.09
LMF TG−LDL	0.02	0.33
LMF Apo B−LDL	0.04	0.41
creatine	0.04	0.51

[a]Only molecular compounds with a *P*-value < 0.05 were reported; the FDR method was used for multiple testing correction. For each model built, the adjustment for age and sex was performed. Abbreviations used: Subfr, subfractions; Chol, cholesterol; LMF, lipoprotein main fractions; PN, particle number; and TG, triglycerides.

of the lipid main fraction LDL related to triglycerides (LMF TG−LDL), Apo B (LMF Apo B−LDL), and creatine, the particle number of LDL4 (PN LDL4), and the subfraction LDL4 related to Apo B (Subfr. Apo B−LDL4) and free cholesterol (Subfr. Free Chol−LDL4) with Rh blood factors.

The molecular roles played by LDL4 and LDL5 in Rh^+ and Rh^- group subjects have, at the best of our knowledge, never been investigated in full. One interesting observation is that the presence of the D antigen on the RBCs membrane was found to be significantly associated with lower HDL, higher triglycerides, and, in particular, higher LDL levels than the Rh^- group; this metabolic behavior could determine the major predisposition of Rh^+ to develop CVDs and lipidic metabolic syndromes.[53,54]

4. CONCLUSIONS

The clinical significance of the ABO and the Rh blood group systems has grown beyond its use in blood transfusion and organ transplantation, and their association and correlation with various physiological and pathophysiological mechanisms have started receiving attention. In this context, to the best of our knowledge, we first present results showing the existence of specific associations between circulating levels of some plasma metabolites, lipoproteins, and the ABO/Rh blood group system in a healthy population. Using a supervised multivariate statistical approach, we were able to very weakly discriminate, using the metabolomic and lipoproteomic information, the ABO groups; in contrast, using the same approach, we were able to discriminate very well the Rh groups. We also obtained univariate associations, applying robust linear regression, between lipoproteins (especially the HDL1, HDL2, and LDL2 subfractions) and the ABO blood groups. Moreover, the LDL5 and LDL4 subfractions and creatine turned out to be significantly associated with Rh blood factors. All results highlighted how the blood groups (ABO and Rh) could be directly associated with a specific remodeling of lipoproteomic metabolism in a healthy population and can provide relevant information for further studies about the association between blood groups and disease susceptibility.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jproteome.2c00375.

Post-hoc test for the regression models of ABO groups and metabolites and lipoproteins; assignment of NMR plasma spectra (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Edoardo Saccenti − *Laboratory of Systems and Synthetic Biology, Wageningen University & Research, Wageningen 6708 WE, The Netherlands;* orcid.org/0000-0001-8284-4829; Email: edoardo.saccenti@wur.nl

Authors

Francesca Di Cesare − *Magnetic Resonance Center (CERM) and Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine (CIRMMP), University of Florence, Firenze 50019, Italy*

Leonardo Tenori − *Magnetic Resonance Center (CERM) and Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine (CIRMMP), University of Florence, Firenze 50019, Italy; Department of Chemistry "Ugo Schiff", University of Florence, Sesto Fiorentino 50019, Italy;* orcid.org/0000-0001-6438-059X

Claudio Luchinat − *Magnetic Resonance Center (CERM) and Consorzio Interuniversitario Risonanze Magnetiche di*

https://doi.org/10.1021/acs.jproteome.2c00375
J. Proteome Res. XXXX, XXX, XXX−XXX

Metallo Proteine (CIRMMP), University of Florence, Firenze 50019, Italy; Department of Chemistry "Ugo Schiff", University of Florence, Sesto Fiorentino 50019, Italy

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jproteome.2c00375

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors acknowledge the support and resources provided by Instruct-ERIC, a Landmark ESFRI project, and specifically the CERM/CIRMMP Italy Centre.

■ ABBREVIATIONS

Apo, apolipoprotein; Chol, cholesterol; CF, calculated figures; CVDs, cardiovascular diseases; HDL, high-density lipoprotein; LDL, low-density lipoprotein; LMF, lipoprotein main fractions; NMR, nuclear magnetic resonance; PL, phospholipid; PN, particle number; RBCs, red blood cells; Rh, Rhesus; Subfr, subfractions; TG, triglycerides

■ REFERENCES

(1) Yamamoto, F.; Clausen, H.; White, T.; Marken, J.; Hakomori, S. Molecular Genetic Basis of the Histo-Blood Group ABO System. *Nature* **1990**, *345*, 229−233.

(2) Lowe, J. B. 7 The blood group-specific human glycosyltransferases. *Baillière's Clin. Haematol.* **1993**, *6*, 465.

(3) Chester, M. A.; Olsson, M. L. The ABO Blood Group Gene: A Locus of Considerable Genetic Diversity. *Transfus. Med. Rev.* **2001**, *15*, 177−200.

(4) Melzer, D.; Perry, J. R. B.; Hernandez, D.; Corsi, A.-M.; Stevens, K.; Rafferty, I.; Lauretani, F.; Murray, A.; Gibbs, J. R.; Paolisso, G.; Rafiq, S.; Simon-Sanchez, J.; Lango, H.; Scholz, S.; Weedon, M. N.; Arepalli, S.; Rice, N.; Washecka, N.; Hurst, A.; Britton, A.; Henley, W.; van de Leemput, J.; Li, R.; Newman, A. B.; Tranah, G.; Harris, T.; Panicker, V.; Dayan, C.; Bennett, A.; McCarthy, M. I.; Ruokonen, A.; Jarvelin, M.-R.; Guralnik, J.; Bandinelli, S.; Frayling, T. M.; Singleton, A.; Ferrucci, L. A Genome-Wide Association Study Identifies Protein Quantitative Trait Loci (PQTLs). *PLoS Genet.* **2008**, *4*, No. e1000072.

(5) Avent, N. D.; Reid, M. E. The Rh Blood Group System: A Review. *Blood* **2000**, *95*, 375−387.

(6) Franchini, M.; Favaloro, E. J.; Targher, G.; Lippi, G. ABO Blood Group, Hypercoagulability, and Cardiovascular and Cancer Risk. *Crit. Rev. Clin. Lab. Sci.* **2012**, *49*, 137−149.

(7) Vasan, S. K.; Rostgaard, K.; Majeed, A.; Ullum, H.; Titlestad, K.-E.; Pedersen, O. B. V.; Erikstrup, C.; Nielsen, K. R.; Melbye, M.; Nyrén, O.; Hjalgrim, H.; Edgren, G. ABO Blood Group and Risk of Thromboembolic and Arterial Disease. *Circulation* **2016**, *133*, 1449−1457.

(8) Lindström, S.; Wang, L.; Smith, E. N.; Gordon, W.; van Hylckama Vlieg, A.; de Andrade, M.; Brody, J. A.; Pattee, J. W.; Haessler, J.; Brumpton, B. M.; Chasman, D. I.; Suchon, P.; Chen, M.-H.; Turman, C.; Germain, M.; Wiggins, K. L.; MacDonald, J.; Braekkan, S. K.; Armasu, S. M.; Pankratz, N.; Jackson, R. D.; Nielsen, J. B.; Giulianini, F.; Puurunen, M. K.; Ibrahim, M.; Heckbert, S. R.; Damrauer, S. M.; Natarajan, P.; Klarin, D.; de Vries, P. S.; Sabater-Lleal, M.; Huffman, J. E.; Bammler, T. K.; Frazer, K. A.; McCauley, B. M.; Taylor, K.; Pankow, J. S.; Reiner, A. P.; Gabrielsen, M. E.; Deleuze, J.-F.; O'Donnell, C. J.; Kim, J.; McKnight, B.; Kraft, P.; Hansen, J.-B.; Rosendaal, F. R.; Heit, J. A.; Psaty, B. M.; Tang, W.; Kooperberg, C.; Hveem, K.; Ridker, P. M.; Morange, P.-E.; Johnson, A. D.; Kabrhel, C.; Trégouët, D.-A.; Smith, N. L.; The Million Veteran Program; The CHARGE Hemostasis Working Group. Genomic and transcriptomic association studies identify 16 novel susceptibility loci for venous thromboembolism. *Blood* **2019**, *134*, 1645−1657.

(9) Franchini, M.; Lippi, G. The Intriguing Relationship between the ABO Blood Group, Cardiovascular Disease, and Cancer. *BMC Med.* **2015**, *13*, 7.

(10) Franchini, M.; Bonfanti, C. Evolutionary Aspects of ABO Blood Group in Humans. *Clin. Chim. Acta* **2015**, *444*, 66−71.

(11) Abegaz, S. B. Human ABO Blood Groups and Their Associations with Different Diseases. *BioMed Res. Int.* **2021**, *2021*, 1.

(12) Fan, Q.; Zhang, W.; Li, B.; Li, D.-J.; Zhang, J.; Zhao, F. Association Between ABO Blood Group System and COVID-19 Susceptibility in Wuhan. *Front. Cell. Infect. Microbiol.* **2020**, *10*, 404.

(13) Monaco, A.; Pantaleo, E.; Amoroso, N.; Bellantuono, L.; Stella, A.; Bellotti, R. Country-Level Factors Dynamics and ABO/Rh Blood Groups Contribution To COVID-19 Mortality. *Sci. Rep.* **2022**, *11*, 24527.

(14) Cildag, S.; Kara, Y.; Senturk, T. ABO Blood Groups and Rheumatic Diseases. *Eur. J. Rheumatol.* **2017**, *4*, 250−253.

(15) Anstee, D. J. The Relationship between Blood Groups and Disease. *Blood* **2010**, *115*, 4635−4643.

(16) EC. Mapping of More Stringent Blood Donor Testing Requirements. https://ec.europa.eu/health/blood-tissues-cells-and-organs/blood/mapping-more-stringent-blood-donor-testing-requirements_it (accessed May 19, 2022).

(17) 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*, 97−107.

(18) 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*, 949−961.

(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. *J. Proteome Res.* **2015**, *14*, 1101−1111.

(20) 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*, 231−243.

(21) Mujahid, A.; Dickert, F. L. Blood Group Typing: From Classical Strategies to the Application of Synthetic Antibodies Generated by Molecular Imprinting. *Sensors* **2015**, *16*, 51.

(22) Bernini, P.; Bertini, I.; Luchinat, C.; Tenori, L.; Tognaccini, A. The Cardiovascular Risk of Healthy Individuals Studied by NMR Metabonomics of Plasma Samples. *J. Proteome Res.* **2011**, *10*, 4983−4992.

(23) Bruker. AVANCE IVDr - Clinical Screening and In Vitro Diagnostics (IVD). https://www.bruker.com/products/mr/nmr/avance-ivdr.html (accessed Mar 16, 2020).

(24) Stekhoven, D. J.; Buhlmann, P. MissForest−non-parametric missing value imputation for mixed-type data. *Bioinformatics* **2012**, *28*, 112−118.

(25) van den Berg, R. A.; Hoefsloot, H. C.; Westerhuis, J. A.; Smilde, A. K.; van der Werf, M. J. Centering, Scaling, and Transformations: Improving the Biological Information Content of Metabolomics Data. *BMC Genomics* **2006**, *7*, 142.

(26) Wikipedia. Blood Type Distribution by Country. 2022, https://en.wikipedia.org/wiki/Blood_type_distribution_by_country (accessed Aug 25, 2022).

(27) Blood Information for Life. Racial and Ethnic Distribution of ABO Blood Types - BloodBook.com. http://www.bloodbook.com/world-abo.html (accessed May 18, 2022).

(28) Pearson. Basic Statistical Analysis, 9th Edition. https://www.pearson.com/content/one-dot-com/one-dot-com/us/en/higher-education/program.html (accessed Apr 08, 2022).

(29) Wall, M. E.; Rechtsteiner, A.; Rocha, L. M. Singular Value Decomposition and Principal Component Analysis. In *A Practical Approach to Microarray Data Analysis*; Berrar, D. P., Dubitzky, W., Granzow, M., Eds.; Springer US: Boston, MA, 2003; pp 91−109.
(30) Hotelling, H. Analysis of a Complex of Statistical Variables into Principal Components. *J. Educ. Psychol.* **1933**, *24*, 417−441.
(31) Fratello, M.; Tagliaferri, R. Decision Trees and Random Forests. In *Encyclopedia of Bioinformatics and Computational Biology*; Ranganathan, S., Gribskov, M., Nakai, K., Schönbach, C., Eds.; Academic Press: Oxford, 2019; pp 374−383.
(32) 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 Complementary Altern. Med.* **2013**, *2013*, 1.
(33) Amaratunga, D.; Cabrera, J.; Lee, Y.-S. Enriched Random Forests. *Bioinformatics* **2008**, *24*, 2010−2014.
(34) Baratloo, A.; Hosseini, M.; Negida, A.; El Ashal, G. Part 1: Simple Definition and Calculation of Accuracy, Sensitivity and Specificity. *Emerg* **2015**, *3*, 48−49.
(35) Szymańska, E.; Saccenti, E.; Smilde, A. K.; Westerhuis, J. A. Double-Check: Validation of Diagnostic Statistics for PLS-DA Models in Metabolomics Studies. *Metabolomics* **2012**, *8*, 3−16.
(36) Fox, J. *Applied Regression Analysis, Linear Models, and Related Methods*; Sage Publications, Inc: Thousand Oaks, CA, US, 1997; pp xxi, 597.
(37) Hoaglin, D. C.; Mosteller, F.; Tukey, J. W. *Exploring Data Tables, Trends, and Shapes*; Wiley: New York, 1985.
(38) Searle, S. R.; Speed, F. M.; Milliken, G. A. Population Marginal Means in the Linear Model: An Alternative to Least Squares Means. *Am. Stat.* **1980**, *34*, 216−221.
(39) 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.* **1995**, *57*, 289−300.
(40) Tippmann, S. Programming Tools: Adventures with R. *Nature* **2015**, *517*, 109−110.
(41) Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5−32.
(42) Nicodemus, K. K.; Malley, J. D.; Strobl, C.; Ziegler, A. The Behaviour of Random Forest Permutation-Based Variable Importance Measures under Predictor Correlation. *BMC Bioinf.* **2010**, *11*, 110.
(43) Venables, W. N.; Ripley, B. D. *Modern Applied Statistics with S*; Chambers, J., Eddy, W., Härdle, W., Sheather, S., Tierney, L., Eds.; Statistics and Computing; Springer: New York, NY, 2002.
(44) Lenth, R. V.; Buerkner, P.; Herve, M.; Love, J.; Miguez, F.; Riebl, H.; Singmann, H. *Emmeans: Estimated Marginal Means, Aka Least-Squares Means*; Emmeans, 2022.
(45) 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. Bioinf.* **2016**, *53*, 14.13.1−14.13.18.
(46) Mengoli, M.; Capuzzo, C.; Terenziani, E.; Bonfanti, I.; Lippi, C.; Franchini, G. Correlation between ABO Blood Group, and Conventional Hematological and Metabolic Parameters in Blood Donors. *Semin. Thromb. Hemostasis* **2016**, *42*, 075−086.
(47) Chen, Y.; Chen, C.; Ke, X.; Xiong, L.; Shi, Y.; Li, J.; Tan, X.; Ye, S. Analysis of Circulating Cholesterol Levels as a Mediator of an Association Between ABO Blood Group and Coronary Heart Disease. *Circ.: Cardiovasc. Genet.* **2014**, *7*, 43−48.
(48) Bartimaeus, E. S.; Waribo, H. A. Relationship between ABO Blood Groups and Lipid Profile Level in Healthy Adult Residents in Port Harcourt Metropolis, Nigeria. *J. Appl. Sci. Environ. Manage.* **2017**, *21*, 1003−1011.
(49) Gillum, R. F. Blood Groups, Serum Cholesterol, Serum Uric Acid, Blood Pressure, and Obesity in Adolescents. *J. Natl. Med. Assoc.* **1991**, *83*, 682−688.
(50) Liu, N.; Zhang, T.; Ma, L.; Wang, H.; Li, H. Association between ABO Blood Groups and Risk of Coronavirus Disease 2019. *Medicine* **2020**, *99*, No. e21709.
(51) Klop, B.; van de Geijn, G.-J. M.; Bovenberg, S. A.; van der Meulen, N.; Elte, J. W. F.; Birnie, E.; Njo, T. L.; Janssen, H. W.; van Miltenburg, A.; Jukema, J. W.; Cabezas, M. C. Erythrocyte-Bound Apolipoprotein B in Relation to Atherosclerosis, Serum Lipids and ABO Blood Group. *PLoS One* **2013**, *8*, No. e75573.
(52) Wu, O.; Bayoumi, N.; Vickers, M. A.; Clark, P. ABO(H) Blood Groups and Vascular Disease: A Systematic Review and Meta-Analysis. *J. Thromb. Haemostasis* **2008**, *6*, 62−69.
(53) Kanbay, M.; Yildirir, A.; Ulus, T.; Bilgi, M.; Kucuk, A.; Muderrisoglu, H. Rhesus Positivity and Low High-Density Lipoprotein Cholesterol: A New Link? *Asian Cardiovasc. Thorac. Ann.* **2006**, *14*, 119−122.
(54) Mashahit, M.; Khattab, D.; Zaki, O.; Owis, A. Relation between ABO/Rh Blood Groups and Metabolic Syndrome. *Fayoum Univ. Med. J.* **2020**, *5*, 15−31.

https://doi.org/10.1021/acs.jproteome.2c00375
J. Proteome Res. XXXX, XXX, XXX−XXX

Supplementary Information

SUPPLEMENTARY MATERIAL: ASSOCIATION OF PLASMA METABOLITES AND LIPOPROTEINS WITH RH AND ABO BLOOD SYSTEMS IN HEALTHY SUBJECTS

Francesca Di Cesare[1,2], Leonardo Tenori[1,2,3], Claudio Luchinat[1,2,3], Edoardo Saccenti[4,].*

[1] Magnetic Resonance Center (CERM), , University of Florence, Via Luigi Sacconi 6, 50019, Sesto Fiorentino, Firenze, Italy

[2] Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine (CIRMMP), University of Florence, Via Luigi Sacconi 6, 50019, Sesto Fiorentino, Firenze, Italy

[3] Department of Chemistry "Ugo Schiff", University of Florence, Via della Lastruccia 3, 50019, Sesto Fiorentino, Italy

[4] Laboratory of Systems and Synthetic Biology, Wageningen University & Research, Stippeneng 4, 6708 WE, Wageningen, the Netherlands.

*To whom correspondence should be addressed: Edoardo Saccenti, Laboratory of Systems and Synthetic Biology, Wageningen University & Research, Stippeneng 4, 6708 WE, Wageningen, the Netherlands. E-mail: edoardo.saccenti@wur.nl

Content:

Supplementary Tables

Table S1: Robust linear regression models performed on serum metabolites and lipids of the ABO blood groups post-hoc tests. Only molecular compounds with a *P*-value < 0.05 were reported. For each comparison, estimate value, Standard Deviation (SD), *P*-value, the FDR adjusted *P*-values, and model name are also reported.

ABO blood group comparison	Estimate	SD	*P*-value	FDR *P*-value	Model name
A - AB	3.18	1.08	0.02	0.08	PN LDL2
AB - O	0.82	0.81	0.02	0.08	PN LDL2
A - AB	0.15	0.35	0.02	0.08	Subfr Chol – LDL2
AB - O	-0.99	0.34	0.02	0.08	Subfr Chol – LDL2
A - AB	0.55	0.19	0.02	0.08	Subfr Free Chol – LDL2
AB - O	-0.52	0.18	0.02	0.08	Subfr Free Chol – LDL2
A - AB	0.68	0.24	0.02	0.08	Subfr PL – LDL2
AB - O	-0.65	0.23	0.03	0.10	Subfr PL – LDL2
A - AB	0.75	0.25	0.02	0.08	Subfr ApoB – LDL2
AB - O	-0.71	0.25	0.02	0.08	Subfr ApoB – LDL2
A - AB	0.33	0.09	0.003	0.05	Subfr Chol – HDL2
AB - O	-0.31	0.09	0.007	0.07	Subfr Chol – HDL2
A - AB	0.36	0.11	0.004	0.05	Subfr PL – HDL2
AB - O	-0.32	0.11	0.01	0.08	Subfr PL – HDL2
A - AB	0.91	0.25	0.002	0.04	Subfr ApoA1 – HDL1
AB - B	-0.77	0.29	0.04	0.11	Subfr ApoA1 – HDL1
AB - O	-0.91	0.25	0.001	0.04	Subfr ApoA1 – HDL1

Abbreviation used: Subfr, Subfractions; Chol, Cholesterol; LMF, Lipoprotein Main Fractions; PN, Particle Number; TG, Triglycerides

Supplementary Figures

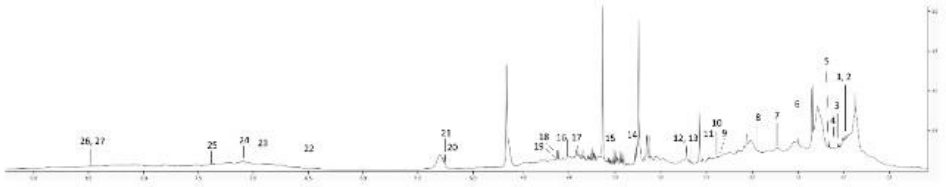


Figure S1: Assignment of a plasma ^{1}H-NOESY spectrum. 1, leucine; 2, isoleucine; 3, valine; 4, unknown 1; 5, 3-hydroxybutyrate; 6, alanine; 7, arginine+lysine; 8, acetate; 9, glutamate; 10, pyruvate; 11, glutamine; 12, methionine; 13, citrate; 14, unknown 2; 15, glycine; 16, creatine; 17, creatinine; 18, lactate; 19, proline; 20, mannose; 21, glucose; 22, fumarate; 23, tyrosine; 24, histidine; 25, phenylalanine; 26, formate; 27, adenosine nucleotide+inosine monophosphate

4.1.6 NMR-based metabolomics to predict early and late adverse outcomes in ischemic stroke treated with intravenous thrombolysis

Cristina Licari[1*], Leonardo Tenori[1,2*], Francesca Di Cesare[1], Claudio Luchinat [1,2,3], Betti Giusti[4,5,6], Ada Kura[4,5], Rosina De Cario[4], Domenico Inzitari[7,8], Benedetta Piccardi[7], Mascia Nesi[7], Cristina Sarti[9], Francesco Arba[10], Vanessa Palumbo[7], PatriziaNencini[7], Rossella Marcucci[4,5,6], Anna Maria Gori4,5,6, Elena Sticchi [4]

[*] contributed equally

[1] Magnetic Resonance Center (CERM), University of Florence, Via Luigi Sacconi 6, 50019, Sesto Fiorentino, Firenze, Italy 2

[2] Department of Chemistry "Ugo Schiff", University of Florence, Via della Lastruccia 3–13, 50019, Sesto Fiorentino, Florence, Italy.

[3] C.I.R.M.M.P., Via Luigi Sacconi 6, 50019 Sesto Fiorentino, Florence, Italy

[4] Department of Experimental and Clinical Medicine, University of Florence, Largo Brambilla 3, Florence 50134, Italy.

[5]Atherothrombotic Diseases Center, Careggi Hospital, Florence, Largo Brambilla 3, Florence 50134, Italy.

[6] Excellence Centre for Research, Transfer and High Education for the Development of DE NOVO Therapies (DENOTHE), University of Florence, Viale Pieraccini 6, Firenze 50139, Italy.

[7] Stroke Unit, Careggi University Hospital, Florence 50134, Italy.

[8] Institute of Neuroscience, Italian National Research Council (CNR), Via Madonna del Piano, 10, Sesto Fiorentino, Florence 50019, Italy.

[9] NEUROFARBA Department, Neuroscience Section, University of Florence, Largo Brambilla 3, Florence 50134, Italy.

[10] Department of Neurology, Careggi University Hospital, Largo Brambilla 3, Florence 50134, Italy.

Awaiting the second round of revision

Candidate's contribuitions: statistical analysis, interpretation of results, and writing the manuscript**.**

Journal of Proteome Research

NMR-BASED METABOLOMICS TO PREDICT EARLY AND LATE ADVERSE OUTCOMES IN ISCHEMIC STROKE TREATED WITH INTRAVENOUS THROMBOLYSIS

Journal:	*Journal of Proteome Research*
Manuscript ID	pr-2022-00333t.R1
Manuscript Type:	Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Licari, Cristina; University of Florence, Magnetic Resonance Center (CERM) Tenori, Leonardo; University of Florence, Magnetic Resonance Center (CERM); University of Florence, Department of Chemistry "Ugo Schiff" Di Cesare, Francesca; University of Florence, Magnetic Resonance Center (CERM) Luchinat, Claudio; University of Florence, Magnetic Resonance Center (CERM); University of Florence, Department of Chemistry "Ugo Schiff"; CIRMMP Giusti , Betti ; University of Florence, Department of Experimental and Clinical Medicine; University Hospital Careggi, Atherothrombotic Diseases Center Kura, Ada; University of Florence, Department of Experimental and Clinical Medicine De Cario, Rosina; University of Florence, Department of Experimental and Clinical Medicine Inzitari, Domenico; University Hospital Careggi, Stroke Unit; Italian National Research Council (CNR), Institute of Neuroscience Piccardi, Benedetta; University Hospital Careggi, Stroke Unit Nesi, Mascia; University Hospital Careggi, Stroke Unit Sarti, Cristina; University of Florence, NEUROFARBA Department, Neuroscience Section Arba, Francesco; University Hospital Careggi, Department of Neurology Palumbo, Vanessa; University Hospital Careggi, Stroke Unit Nencini, Patrizia; University Hospital Careggi, Stroke Unit Marcucci, Rossella; University of Florence, Department of Experimental and Clinical Medicine; University Hospital Careggi, Atherothrombotic Diseases Center Gori, Anna Maria; University of Florence, Department of Experimental and Clinical Medicine; University Hospital Careggi, Atherothrombotic Diseases Center Sticchi, Elena; University of Florence, Experimental and Clinical Medicine

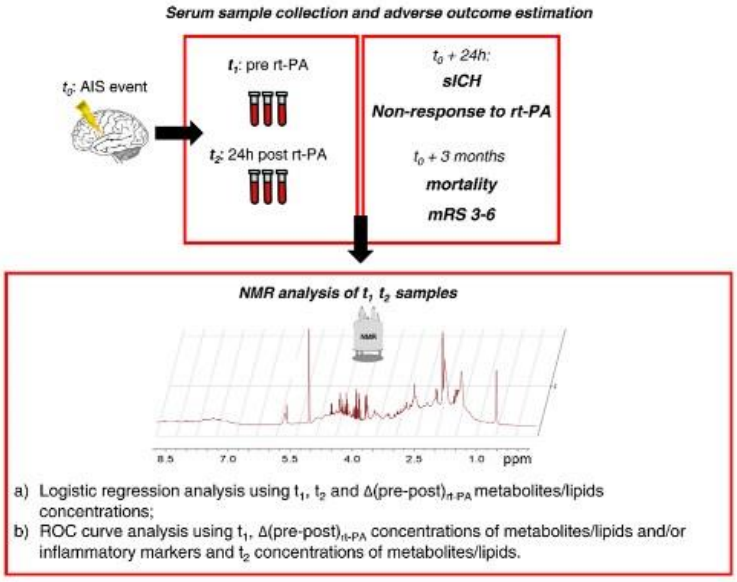


Figure 1: Graphical representation of the analysis followed to identify, in serum samples, possible predictors of early and late adverse outcomes of acute ischemic stroke (AIS) treated with intravenous thrombolysis with recombinant tissue plasminogen activator (rt-PA). Blood samples were collected before (t1) and 24h after (t2) the administration of rt-PA. Early adverse outcomes were defined 24h after the event, as follows: development of symptomatic intracerebral haemorrhage (sICH) and non-response to the intravenous thrombolysis. Late outcomes (mortality and disability (mRS 3-6)) were defined three-months after the transient ischemia. For each time-point, 1D NMR spectra have been acquired and therefore used to estimate metabolites and lipids concentrations; t1, t2, and Δ(pre-post)rt-PA concentrations of metabolites/lipoprotein.

127x106mm (220 x 220 DPI)

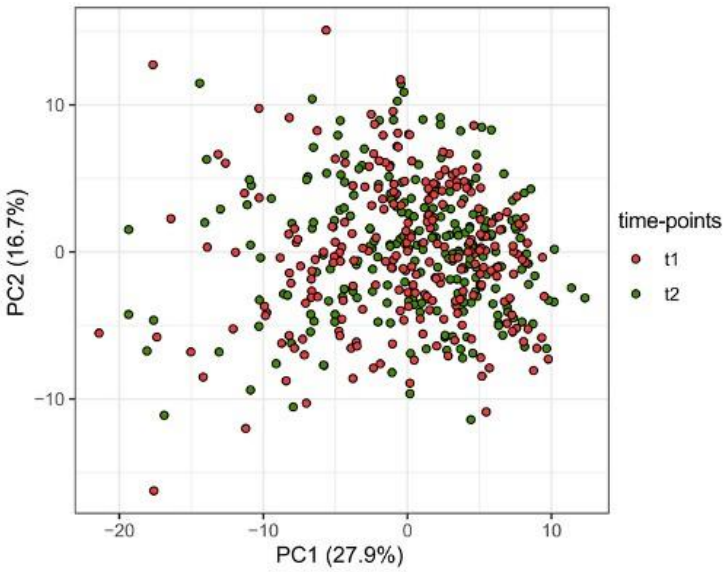
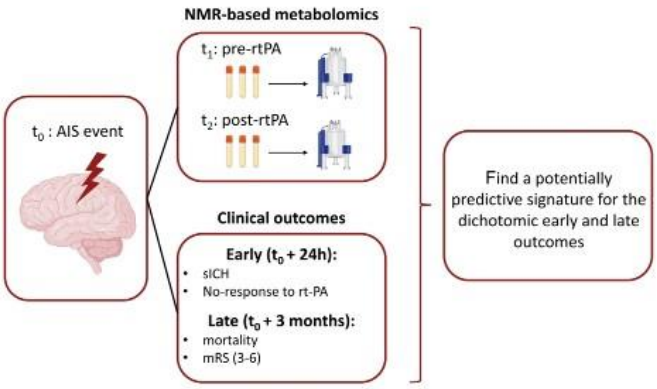


Figure 2: Scatter plots of Principal component analysis (PCA), taking into account the first two principal components (PC). The red dots represent the serum metabolic profiles collected before (t1) the administration of rt-PA and the green dots represent the serum metabolic profiles collected 24h after (t2) the administration of rt-PA.

31x24mm (600 x 600 DPI)



For TOC only

338x190mm (300 x 300 DPI)

NMR-BASED METABOLOMICS TO PREDICT EARLY AND LATE ADVERSE OUTCOMES IN ISCHEMIC STROKE TREATED WITH INTRAVENOUS THROMBOLYSIS

Cristina Licari[1]#, Leonardo Tenori[1,2]#, Francesca Di Cesare[1], Claudio Luchinat [1,2,3], Betti Giusti[4,5,6], Ada Kura[4,5], Rosina De Cario[4], Domenico Inzitari[7,8], Benedetta Piccardi[7], Mascia Nesi[7], Cristina Sarti[9], Francesco Arba[10], Vanessa Palumbo[7], PatriziaNencini[7], Rossella Marcucci[4,5,6], Anna Maria Gori[4,5,6], Elena Sticchi (Corresponding)[4]*

 # contributed equally

Corresponding author email: elena.sticchi@unifi.it

[1] Magnetic Resonance Center (CERM), University of Florence, Via Luigi Sacconi 6, 50019, Sesto Fiorentino, Firenze, Italy 2

[2] Department of Chemistry "Ugo Schiff", University of Florence, Via della Lastruccia 3–13, 50019, Sesto Fiorentino, Florence, Italy.

[3] C.I.R.M.M.P., Via Luigi Sacconi 6, 50019 Sesto Fiorentino, Florence, Italy

[4] Department of Experimental and Clinical Medicine, University of Florence, Largo Brambilla 3, Florence 50134, Italy.

[5]Atherothrombotic Diseases Center, Careggi Hospital, Florence, Largo Brambilla 3, Florence 50134, Italy.

[6] Excellence Centre for Research, Transfer and High Education for the Development of DE NOVO Therapies (DENOTHE), University of Florence, Viale Pieraccini 6, Firenze 50139, Italy.

[7] Stroke Unit, Careggi University Hospital, Florence 50134, Italy.

[8] Institute of Neuroscience, Italian National Research Council (CNR), Via Madonna del Piano, 10, Sesto Fiorentino, Florence 50019, Italy.

[9] NEUROFARBA Department, Neuroscience Section, University of Florence, Largo Brambilla 3, Florence 50134, Italy.

[10] Department of Neurology, Careggi University Hospital, Largo Brambilla 3, Florence 50134, Italy.

KEYWORDS: ischemic stroke, metabolomics, lipoproteomics, Nuclear Magnetic Resonance

ABSTRACT

Metabolic perturbations and inflammatory mediators play a fundamental role in both early and late adverse post-acute ischemic stroke (AIS) outcomes. Using data from the observational MAGIC (MArker bioloGici nell'Ictus Cerebrale) study, we evaluated the effect of 130 serum metabolic features, using a Nuclear Magnetic Spectroscopy (NMR) approach, on: haemorrhagic transformation at 24 hours after stroke, non-response to intravenous thrombolytic treatment with recombinant tissue plasminogen activator (rt-PA), and three-month functional outcome. Blood circulating metabolites, lipoproteins, and inflammatory markers were assessed at baseline and 24 hours after rt-PA treatment. Adjusting for the major determinants for unfavourable outcomes (*i.e.* age, sex, time onset-to-treatment, etc.), we found that acetone and 3-hydroxybutyrate were associated with symptomatic haemorrhagic transformation and with non-response to rt-PA; while 24 hours after rt-PA levels of triglycerides-HDL and triglycerides-LDL were associated with three-month mortality. Cholesterol and phospholipids levels, mainly related to smaller and denser VLDL and LDL subfractions were associated with three-month poor functional outcomes. We also reported associations between baseline-24 hours relative variation (Δ) in VLDL subfractions and ΔC-reactive protein, Δinterleukin-10 levels with haemorrhagic transformation. All observed metabolic changes reflect a general condition of energy failure, oxidative stress, and systemic inflammation that characterize the development of adverse outcomes.

1. INTRODUCTION

Ischemic stroke is a leading cause of death and disability continuously increasing [1] and contributing significantly to health costs. There is an urgent need to find biomarkers useful for clinical practice and to better understand metabolic dysregulation on the basis of the pathophysiological mechanisms of the disease. In this light, metabolic perturbations are fundamental events that contribute to ischemic stroke, its progression, and the development of unfavourable outcomes.[2–5] Regarding the knowledge about biomarkers associated with poor prognosis in the setting of stroke patients treated with thrombolysis, little evidence are reported (*i.e.* glucose).[6] Dyslipidaemia is a risk factor contributing to the onset of ischemic stroke; high levels of total cholesterol and low-density lipoprotein (LDL) cholesterol increase the risk for cerebral ischemia.[7,8] However, the effects of lipid levels on clinical outcomes after the ischemic attack are controversial: high total cholesterol and LDL levels have been associated with better functional and vital outcomes after stroke[9,10], while low LDL levels increased the risk of early symptomatic intracranial heamorrhage[11] and low total cholesterol was related to worse functional outcome in ischemic stroke patients after the thrombolytic treatment.[12] Therefore, the effective contribution of lipid levels to stroke outcomes, particularly after thrombolysis, needs to be further investigated.

In this framework, Nuclear Magnetic Resonance (NMR)-based metabolomics can provide crucial information, allowing a high-throughput analysis of various types of samples (*i.e.* blood and urine) and giving information on various molecular features present in biological matrices.[13,14] Here, using data of patients enrolled in the MAGIC (MArker bioloGici nell'Ictus Cerebrale) study,[15–17] we aimed at providing metabolic insights underlying susceptibility to early and late post-acute ischemic stroke adverse outcomes.

We also studied the interplay between statistically significant metabolomic features and circulating inflammatory markers, such as Δ of metalloproteinase 9 (MMP9), alpha2 macroglubulin (A2M),

serum amyloid protein (SAP), C reactive Protein (CRP), interleukins (ILs), tumour necrosis factor alpha (TNFα) and monocyte chemo-attractant protein I (MCPI), resulted to be statistically associated to secondary intracerebral hemorrhage (sICH), three-month mortality and three-month poor functional outcome, as previously described in the publication of Gori *et al.* [16]

An overview of the present study design is reported in **Figure 1**.

2. MATERIALS AND METHODS

2.1 Study Population and Outcomes

The subjects and the study samples considered are from the MAGIC study[15,16] in which 327 patients were enrolled; only subjects for whom serum aliquots were available for metabolomics analysis are considered in this study (n=243).

The study population considered here is characterized by patients who had an acute ischemic stroke (AIS) 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,[18] in the frame of the national, observational and multicentric MAGIC study.[15,16]

The whole study focuses on the analysis of serum samples collected at two different time points: before (t_1) and 24h after (t_2) the administration of rt-PA. Adverse outcomes were defined as follows: *i*) early outcomes (24h): development of symptomatic intracerebral haemorrhage (sICH), according to the National Institute of Neurological Disorders and Stroke criteria,[19] and absence of clinical response to systemic thrombolysis (<4 point decrease on 24 h NIHSS); *ii*) late outcomes (three-months): death and disability defined as modified Rankin scale (mRS), dichotomized into good (mRS, 0-2) or poor (mRS, 3-6) functional status.

A description of the clinical and demographic characteristics of the patients is reported in **Table 1**.

2.2 Ethical Issues

The study protocol was approved by the local Ethical Committee of each participating Centre complies with the Declaration of Helsinki. All patients gave informed consent.

2.3 NMR sample collection, 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.[13]

2.4 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 ^{1}H-^{13}C-^{15}N and ^{2}H 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, three one-dimensional proton NMR spectra (*i.e.* 1D NOESY, 1D CPMG, and 1D Diffusion-Edited) were acquired with different pulse sequences,[20–22] allowing the selective detection of different molecular components. Detailed procedures on parameters of NMR experiments are reported elsewhere.[13]

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).

2.5 Metabolite and lipoprotein identification and quantification

18 metabolites and 112 lipoproteins were identified and estimated from ^{1}H-1D NOESY NMR spectra according to Bruker's B.I.-LISA protocols.[23] A complete list of the molecular features analysed in this study is presented in Supplementary **Table S1**.

2.6 Laboratory measurements

Levels of different inflammatory markers, *i.e.* IL, TNFα, CRP, A2M, SAP, and MCPI, have been measured as previously specified in Gori *et al.* [16]

2.7 Statistical analysis

2.7.1 Demographic and clinics

For demographic, clinical characteristics, and risk factors, a t-test and chi-square test were applied, respectively, for comparisons including continuous variables and categorical variables.

2.7.2 Exploratory analysis

Principal Component Analysis (PCA) [24,25] was performed to explore data patterns. Data were scaled to unit variance before analysis.

2.7.3 Correlation analysis

Pearson correlation analysis[26,27] was used to study the potential association between inflammatory markers and metabolomics data at t_1 and t_2. The significance threshold has been imposed at P-value<0.05. For completeness, adjusted P-values, using the Benjamini- Hochberg method,[28] were also reported.

2.7.4 Penalized 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 and post rt-PA variation ($\Delta(\text{pre-post})_{\text{rt-PA}}$) of metabolites and lipids concentrations.

For each metabolite or lipid x, the variation in terms of concentration between the time-points t_1 and t_2 was calculated as follows:

$$x_{\Delta(\text{pre}-\text{post})_{\text{rtPA}}} = \frac{x_{t_2} - x_{t_1}}{\frac{x_{t_1} + x_{t_2}}{2}} \#(1)$$

The statistically robust effect of each variable at t_1, t_2, or considering $\Delta(\text{pre-post})_{\text{rtPA}}$ on the four adverse outcomes was estimated by penalized logistic regression analysis,[29] 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, 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 metabolite or lipoprotein analysed. For completeness, adjusted *P*-values, using the Benjamini- Hochberg method,[28] were also reported. Since correction for multiple testing increases the risk of false-negative, especially in the case where (possibly) weak associations are tested on a large number of variables, we presented both corrected and uncorrected *P*-values. We considered significant *P*-values<0.05.

2.7.5 Receiver Operating Characteristic curves (ROC) analysis

Adverse outcomes were evaluated also by applying ROC curve analysis 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 of the evaluated outcomes, we estimated the values of the area under the ROC curve (AUC-ROC) for two different penalized logistic regression models. Firstly, we calculated AUC values for models (hereunder referred as "Bas1" for t_1 and $\Delta(\text{pre-post})_{\text{rtPA}}$, and "Bas2" for t_2) that

included only clinical and risks factors known to affect the outcomes (*i.e.* age, gender, blood collection center, time onset-to-treatment, recent infections or inflammations, glycemia, National Institutes of Health Stroke Scale (NIHSS), history of atrial fibrillation, and congestive heart failure). Secondly, we quantify how much the addition of a combination of metabolomic features increases the prediction for events when added to each "Bas" ROC curve model. The metabolites and/or lipoproteins to be included are chosen among the top three statistically significant features based on the results of the penalized logistic regression models.

Thirdly, for t_1 and $\Delta(\text{pre-post})_{\text{rtPA}}$ ROC models, we combined statistically significant metabolic features with specific blood circulating inflammatory markers statistically associated with sICH, three-month mortality, and three-month poor functional outcome (mRS = 3-6), selecting them based on the results reported in the original publication of Gori *et al.*[16]

For all ROC models, 95% Cis of the AUC values have been calculated, together with a *P*-value, to highlight any significant changes in the prediction of the adverse outcome after considering the association of metabolomics and clinical features.

To avoid overfitting, before performing any ROC analysis, all penalized logistic regression models were cross-validated using the leave-one-out scheme.

2.7.6 Software

All statistical analyses were performed using R (version 3.5.3), open-source software for the statistical management of data.[30] Penalized logistic regression models were built using the *logistf* function of the R package "logistf".[29,31] To perform ROC analysis *roc* function of the R package "pROC" was used.[32] To perform correlation analysis *rcorr* function of the R package "Hmisc" was used. The plot was generated using the R package "ggplot2".[33]

3. RESULTS

3.1 Exploration analysis and association between molecular features and inflammatory markers

All analyses have been performed using data from 243 patients of the original cohort belonging to the MAGIC study.[15,16] Notwithstanding this, all demographic, clinical characteristics, and risk factors do not statistically change (P-value>0.05) between the original cohort and the sub-cohorts analysed in this study (see Supplementary **Table S2**). Firstly, as a multivariate exploratory approach, PCA was performed using all 18 metabolites and 112 lipoproteins detected on n=243 serum NMR spectra in order to obtain an overview of the variation in the data and to check the presence of metabolic signatures among the different evaluated condition. As shown in **Figure 2**, there is no separation among the samples collected before (t_1) and 24h after (t_2) the administration of rt-PA, and no outliers are highlighted, confirming the good quality of the data.

To highlight potential associations between molecular features and inflammatory markers before (t_1) and 24h after (t_2) rt-PA, Pearson correlation analysis was also performed. In Supplementary **Table S3** and **S4** only the significant correlations (P-value<0.05) were reported, respectively, for t_1 and t_2 time-points. Before the administration of rt-PA (t_1) we observed, in particular, correlations between liporpoteins and IL-8 and CRP; after the administration of rt-PA (t_2) we observed, in particular, correlations between liporpoteins and IL-6, IL-12 and CRP. It is important to underline that in no case we observed strong correlations, which present a Pearson's r> |0.6|.

3.2 Molecular features associated with early adverse outcomes

3.2.1 Development of sICH

Before starting the thrombolytic therapy (t_1), acetone and 3-hydroxybutyrate resulted to be the only metabolites significantly (*P*-values<0.05) associated with the development of symptomatic intracranial haemorrhage. At t_2, 17 lipids out of 112 were associated with the evaluated outcome, especially lipids related to bigger and less dense VLDL particles. Considering $\Delta(\text{pre-post})_{\text{rt-PA}}$ metabolites/lipids concentrations, we reported phenylalanine, pyruvate, and glucose as the most statistically associated metabolites, while among lipid parameters, phospholipids related to VLDL particles are the most statistically associated analytes (see **Table 2**). To view the complete lists of all metabolites and lipids effects at t_1, t_2, and $\Delta(\text{pre-post})_{\text{rt-PA}}$ we refer the reader to Supplementary Information (**Tables S5**, **S6,** and **S7)**.

Considering these results, as completely reported in **Table 3**, for the t_1 time-point, the addition of the acetone and 3-hydroxybutyrate to the Bas1 ROC curve model did not statistically improve the AUC value (model Bas1: AUC=0.58, model Bas1+G: AUC=0.59, *P*-value=0.349). Adding baseline levels of interleukin I receptor antagonist (IL-IRa) and IL-10 to Bas1+G model, the resulting model got worse (model Bas1+G+P: AUC=0.51, *P*-value=0.87).

For the t_2 time point, the addition of VLDL-2 cholesterol, phospholipids, and triglycerides to the Bas2 model statistically increased the baseline AUC value of 0.556 to a value of 0.636 (*P*-value=0.039, model Bas2+H).

Finally, considering $\Delta(\text{pre-post})_{\text{rt-PA}}$ metabolic features concentrations, adding the values related to VLDL-phospholipids main fraction and the ones related to the VLDL-2 cholesterol and phospholipids subfractions to the Bas 1 ROC model, we obtained an improvement of the related AUC value, passing from 0.58 to a value of 0.65 (*P*-value = 0.063, model Bas1+I). Moreover, ROC analysis demonstrated that the addition of ΔCRP and ΔIL-10 to the Bas1+I model, significantly improved the AUC for the prediction of sICH in AIS patients treated with thrombolysis (model Bas1+I+Q: AUC = 0.75, *P*-value = $2.07*10^{-5}$).

3.2.2 Non-response to the intravenous thrombolytic therapy

At t_1 time-point, acetone and 3-hydroxybutyrate result to be associated with the non-response to the intravenous thrombolysis. At t_2, we estimated the association for phenylalanine, VLDL-5 phospholipids, HDL-4 triglycerides, HDL-2 free cholesterol, acetone, and 3-hydroxybutyrate. About $\Delta(\text{pre-post})_{\text{rt-PA}}$, alanine, acetone, total particle number of Apo-B100, related LDL-3 sub-particles, and HDL-4 triglycerides were the significant predictors of non-response to the therapeutic thrombolytic treatment (see **Table 2, and Supplementary Tables S5, S6, and S7**).

As reported in **Table 3**, at t_1 the addiction of acetone and 3-hydroxybutyrate to the baseline Bas1 ROC curve model did not statistically improve the AUC (model Bas1: AUC=0.54, model Bas1+J: AUC=0.55). At t_2 adding the same two ketone bodies, also reporting FDR values<0.05, to the Bas2 ROC curve model, the AUC significantly increased to a value of 0.62 (model Bas2+K). Regarding

Δ(pre-post)$_{\pi\text{-PA}}$, the addition of acetone, LDL-3 particle number, and LDL-3 Apolipoprotein-B to the respective Bas1 ROC curve model led to a slight and not significant improvement of the AUC, passing from 0.54 to 0.58 (model Bas1+L).

3.3 Molecular features associated with late adverse outcomes

3.3.1 Three-month mortality

As reported in **Table 2**, triglycerides related to the HDL-3 subfraction are the only analytes presenting an association with the AIS patients' three-month mortality at t_1. For t_2, HDL triglycerides, LDL-6 related cholesterol, triglycerides and phospholipids, HDL-3 and HDL-4 triglycerides resulted associated with this late outcome. Considering Δ(pre-post)$_{\pi\text{-PA}}$, triglycerides related to LDL, HDL particles, and LDL-1, LDL-2, and HDL-4 sub-particles resulted associated with three-month mortality (for more information see **Supplementary Tables S5, S6, and S7**).

At t_1, adding triglycerides related to the HDL-3 subfraction to the Bas1 ROC model, the AUC did not improve (model Bas1+A). Adding to this model levels of ΔMMP9, SAP, and A2M, a non-statistically significant improvement of the AUC was obtained (model Bas+A+M: AUC=0.75, *P*-value=0.28). At t_2, adding LDL-6, HDL-triglycerides, and LDL-6 cholesterol to the related baseline model, the AUC increased, without statistical significance, from 0.76 to 0.80 (Bas2+B). Lastly, considering Δ(pre-post)$_{\pi\text{-PA}}$, the addition of HDL, LDL, and LDL-2 triglycerides to Bas1 ROC curve model increased, without significance, the AUC value from 0.72 to 0.74 (Bas1+C). The Δ(pre-post)

levels of IL-IRa, IL-10, and TNFα were added to the model Bas1+C; the Bas1+C+N model resulted to be worse (AUC=69.5, *P*-value=0.29) (see **Table 3**).

3.3.2 Three-month mRS 3-6

As reported in **Table 2**, at t_1, 15 different lipids, mainly related to small dense VLDL, LDL, and HDL particles, resulted being statistically associated with a three-month poor functional outcome (mRS = 3-6). Considering t_2 samples, acetic acid, 3-hydroxybutyrate, and 16 different lipids resulted to be statistically related to the development of poor functional outcomes (mRS = 3-6). Among the 16 lipids: triglycerides, total cholesterol and free cholesterol estimated for LDL-6 subfractions appeared as the most statistically significant associated metabolic features. When relating $\Delta(\text{pre-post})_{\text{rt-PA}}$ metabolites and lipids concentrations to the three-month poor functional outcome (mRS = 3-6), we found that glutamate, acetate, free cholesterol, phospholipids, and Apo A2 related mainly to HDL-1 and HDL-2 were statistically associated with the outcome (for more information see **Supplementary Tables S5, S6, and S7**).

As reported in **Table 3**, at t_1, the addition of the top three statistically associated lipid fractions to the related baseline ROC model built at t_1, did not statistically improve the AUC (model Bas1: AUC = 0.81; model Bas1+D AUC = 0.83, *P*-value = 0.193). For the ROC curve analysis, at t_2, the addition of 24h post rt-PA values of VLDL-5 cholesterol, LDL-6 free cholesterol, LDL-6 triglycerides, reporting also FDR values <0.05 in the penalized logistic regression, to the Bas2 ROC model, led to

a statically significant improvement in the already good value of baseline AUC (model Bas2: AUC = 0.807, model Bas2+E: AUC = 0.844, *P*-value = 0.037). Considering $\Delta(\text{pre-post})_{\text{rt-PA}}$, adding glutamate, ApoA2 related to both HDL-1 and HDL-2 to the Bas1 ROC curve model, the AUC remained the same (model Bas1: AUC = 0.81; model Bas1+F: AUC = 0.81, *P*-value = 0.57). Lastly, adding to Bas1+F model, Δ values of IL-6, IL-8, IL-10, IL-12, MCPI, TNFα and CRP, the resulting Bas1+F+O model, reported a similar value of the AUC (model Bas1+F+O: AUC=0.80, *P*-value = 0.83).

4. DISCUSSION

We observed that various metabolomic features (especially lipids related to HDL, LDL, VLDL particles, and ketone bodies) resulted in statistically associated (P-value<0.05) with each of the assessed post-AIS adverse outcomes. Since correction for multiple testing increases the risk of false-negative, especially in the case where (possibly) weak associations are tested on a large number of variables, in this work we present both corrected and uncorrected P-values and discussing the biological implications of the results for which P-values were significant before correction, improving the value of baseline AUC models for each specific outcome. Combining metabolomic features with inflammatory markers resulted in statistically significant associations among Δ levels of VLDL-2 cholesterol, VLDL-2 phospholipids, main fraction of VLDL phospholipids, and ΔCRP, ΔIL-10 for the prediction of sICH.

We observed that acetone and 3-hydroxybutyrate appear involved in the symptomatic development of intracranial haemorrhage and in the non-response to the thrombolytic therapy; while $t_2$triglycerides levels, mainly associated with HDL and LDL particles, seem to be related to three-month death. Moreover, 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.

Many studies evidenced associations between lipids and ischemic stroke,[2,4] demonstrating how these molecules are involved in the aetiology and progression of AIS. Serum cholesterol is an independent predictor for long-term functional outcomes and higher serum total cholesterol levels have been associated with better prognosis.[34] Triglycerides have been significantly associated with the risk of stroke and carotid atherosclerosis,[35] but the biological mechanisms by which they could affect the three-month death or the survival of AIS patients need further investigations. Since triglycerides are hydrolyzed to fatty acids to furnish alternative energy sources during stress conditions, we can hypothesize a situation of energy failure, thus leading to an increase in demand for energy and an enhanced transition from aerobic to anaerobic glycolysis. As a consequence, levels of pyruvate and lactate may change, altering their role in providing substitute energy fuel and in metabolic pathways of neuroprotection where lactate is normally largely involved.[36] Moreover, citrate and ketone bodies levels can change to restore energy homeostasis. It is 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.[37]

We reported associations of LDL estimated at t_2, with three-month death and three-month poor functional outcome. 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).[38] Our results also evidence the role of low LDL and VLDL cholesterol, estimated at t_2, in increasing the risk of symptomatic intracranial haemorrhage development at 24 hours.

Moreover, statistically significant associations between HDL-related parameters and adverse outcomes could reflect a general condition of inflammation that characterizes the post-stroke course. Inflammation may alter the lipoprotein profile by modulating the HDL function[39] which contributes to the development of adverse outcomes, linked to the activity of rt-PA itself. 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.

TABLES

Table 1. Demographic and clinical characteristics of the 243 patients selected for this study.

Demographics	*n* = 243
Age, years, mean and SD	68.8 ± 11.9
Sex (male), n (%)	137 (56.4%)
Onset to treatment time, minutes, mean and SD	163.4 ± 83.7
NIHSS, mean and SD	11.9 ± 6.1
Baseline systolic blood pressure, mmHg, mean and SD	147.5 ± 21.3
Baseline diastolic blood pressure, mmHg, mean and SD	79.7 ± 12.7
Blood glucose, mg/dL, mean and SD	130.2 ± 49.5
Risk factors	
Hypertension, n (%)	143 (58.8%)
Diabetes, n (%)	36 (14.8%)
Hyperlipidaemia, n (%)	56 (23%)
Current smoking, n (%)	35 (14.4%)
Atrial fibrillation, n (%)	56 (23%)
Congestive heart failure, n (%)	26 (10.7%)

* Main abbreviations: SD: Standard Deviation; NIHSS: National Institutes of Health Stroke Scale

Table 2. Effect of statistically significantly (*P*-values < 0.05) associated pre- (t_1), 24h post- (t_2), and Δ(pre-post) rt-PA molecular features concentrations on early (sICH, non-response to thrombolysis) and late (three-month mortality and three-month poor functional outcome) adverse outcomes, 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.

	PRE rt-PA (t_1)				24h POST rt-PA (t_2)				Δ(pre-post)$_{rt\text{-}PA}$			
	Analytes	**OR (95% CI)**	**P**	**FDR**	**Analytes**	**OR (95% CI)**	**P**	**FDR**	**Analytes**	**OR (95% CI)**	**P**	**FDR**
Early outcomes	**sICH**											
	3-HB	1.457 (1.062-1.998)	0.025	0.455	SubPhosp_VLDL-2	0.364 (0.193-0.687)	0.001	0.097	SubPhosp_VLDL-2	0.501 (0.323-0.777)	0.004	0.301
	Acetone	1.393 (1.016-1.909)	0.046	0.502	SubTrigl_VLDL-2	0.372 (0.193-0.717)	0.003	0.097	LMF_Phosp_VLDL	0.525 (0.339-0.814)	0.006	0.301
					SubChol_VLDL-2	0.383 (0.203-0.719)	0.003	0.097	SubChol_VLDL-2	0.554 (0.366-0.838)	0.009	0.301
					LMF_Phosp_VLDL	0.382 (0.194-0.753)	0.006	0.128	SubPhosp_VLDL-3	0.534 (0.332-0.86)	0.014	0.301
					SubTrigl_VLDL-3	0.44 (0.242-0.803)	0.008	0.128	Phe	1.736 (1.157-2.605)	0.015	0.274
					LMF_FreeChol_VLDL	0.419 (0.222-0.793)	0.009	0.128	SubTrigl_HDL-3	0.626 (0.43-0.912)	0.019	0.301
					LMF_Chol_VLDL	0.452 (0.251-0.812)	0.010	0.128	SubFreeChol_VLDL-1	0.57 (0.363-0.896)	0.020	0.301
					SubPhosp_VLDL-3	0.46 (0.257-0.822)	0.010	0.128	LMF_FreeChol_VLDL	0.602 (0.399-0.91)	0.023	0.301
					SubChol_VLDL-3	0.465 (0.262-0.825)	0.010	0.128	VLDL_PN	0.597 (0.39-0.914)	0.024	0.301
					SubFreeChol_VLDL-2	0.46 (0.249-0.852)	0.014	0.164	LMF_ApoB_VLDL	0.597 (0.39-0.914)	0.024	0.301
					SubPhosp_VLDL-1	0.361 (0.157-0.832)	0.017	0.171	LMF_Chol_VLDL	0.611 (0.405-0.921)	0.026	0.301
					SubFreeChol_VLDL-3	0.468 (0.253-0.868)	0.018	0.171	Pyruvic acid	1.692 (1.062-2.695)	0.035	0.297
					LMF_Trigl_VLDL	0.405 (0.19-0.867)	0.020	0.178	SubTrigl_HDL-4	0.656 (0.449-0.959)	0.037	0.339
					SubChol_VLDL-1	0.395 (0.179-0.869)	0.024	0.194	SubPhosp_VLDL-4	0.627 (0.412-0.952)	0.039	0.339
					SubFreeChol_VLDL-1	0.424 (0.195-0.923)	0.031	0.237	SubTrigl_VLDL-2	0.608 (0.392-0.945)	0.041	0.339
					SubChol_VLDL-4	0.575 (0.349-0.947)	0.038	0.270	LMF_Trigl_IDL	0.609 (0.383-0.97)	0.047	0.339
					LMF_Trigl_IDL	0.487 (0.247-0.962)	0.040	0.271	SubTrigl_VLDL-4	0.602 (0.372-0.973)	0.047	0.339

					Trigl	0.483 (0.241-0.968)	0.044	0.276	Glucose	1.68 (1.042-2.707)	0.050	0.297
	Non-response to thrombolysis											
	Acetone	0.696 (0.509-0.95)	0.017	0.219	3-HB	0.571 (0.413-0.791)	0.0004	0.007	Acetone	0.684 (0.516-0.906)	0.007	0.129
	3-HB	0.718 (0.531-0.97)	0.024	0.219	Acetone	0.609 (0.43-0.863)	0.003	0.026	SubTrigl_HDL-4	1.386 (1.026-1.872)	0.028	0.238
	LMF_Trigl_HDL	1.375 (1.016-1.862)	0.036	0.956	SubTrigl_HDL-4	1.559 (1.135-2.143)	0.005	0.586	3-HB	0.725 (0.543-0.969)	0.028	0.238
	SubTrigl HDL-3	1.343 (0.998-1.806)	0.049	0.956	Phe	0.722 (0.542-0.962)	0.024	0.143	Creatinine	0.755 (0.568-1.003)	0.047	0.238
	Three-month mortality											
	SubTrigl_HDL-3	0.61 (0.372-1.002	0.045	0.994	SubTrigl LDL-6	1.751 (1.182-2.594)	0.010	0.982	LMF_Trigl LDL	0.609 (0.414-0.897)	0.023	0.934
					SubChol LDL-6	1.804 (1.084-3.003)	0.042	0.982	LMF_Trigl HDL	0.597 (0.404-0.882)	0.031	0.934
					LMF_Trigl_HDL	0.526 (0.299-0.927)	0.048	0.982	SubTrigl_LDL-2	0.567 (0.353-0.912)	0.039	0.934
									SubTrigl_HDL-4	0.634 (0.429-0.937)	0.040	0.934
									SubTrigl_LDL-1	0.576 (0.357-0.927)	0.041	0.934
Late outcomesoutoutcomes	**Three-month poor functional outcome**											
	SubPhosp_HDL-3	0.549 (0.365-0.826)	0.004	0.209	SubChol_VLDL-5	0.497 (0.334-0.738)	0.0005	0.040	Glu	1.565 (1.069-2.29)	0.023	0.412
	Apo-B100-Apo-A1	1.63 (1.131-2.349)	0.007	0.209	SubTrigl_LDL-6	1.745 (1.213-2.508)	0.001	0.040	SubApoA2_HDL-2	1.419 (1.004-2.006)	0.042	0.973
	LMF_Phosp_HDL	0.584 (0.393-0.867)	0.008	0.209	SubFreeChol_LDL-6	1.844 (1.266-2.687)	0.001	0.040	SubApoA2_HDL-1	1.38 (0.978-1.946)	0.044	0.973
	SubChol_VLDL-5	0.618 (0.427-0.893)	0.011	0.209	LDL6_PN	1.79 (1.233-2.597)	0.002	0.040				
	SubFreeChol_LDL-6	1.575 (1.116-2.221)	0.011	0.209	SubApoB_LDL-6	1.79 (1.233-2.597)	0.002	0.040				
	SubPhosp_VLDL-5	0.641 (0.444-0.926)	0.019	0.268	SubChol_LDL-6	1.743 (1.207-2.516)	0.003	0.049				
	SubPhosp_HDL-2	0.633 (0.429-0.934)	0.022	0.268	SubPhosp_LDL-6	1.701 (1.184-2.445)	0.004	0.061				
	SubTrigl LDL-6	1.464 (1.059-2.025)	0.024	0.268	SubPhosp_VLDL-5	0.585 (0.401-0.852)	0.006	0.076				
	SubApoA1_HDL-3	0.637 (0.43-0.944)	0.025	0.268	Apo-B100-Apo-A1	1.738 (1.17-2.583)	0.006	0.076				
	SubApoA1_HDL-2	0.644 (0.438-0.946)	0.026	0.268	Acetic acid	1.651 (1.133-2.407)	0.010	0.173				
	SubChol LDL-6	1.461 (1.049-2.035)	0.029	0.280	Apo-B100	1.595 (1.087-2.339)	0.018	0.175				

	LMF_ApoA1_HDL	0.674 (0.465-0.978)	0.040	0.331	LMF_ApoB_LDL	1.524 (1.044-2.226)	0.031	0.246				
	SubApoB_LDL-6	1.418 (1.022-1.967)	0.044	0.331	LDL_PN	1.524 (1.044-2.226)	0.031	0.246				
	LDL6_PN	1.418 (1.022-1.967)	0.044	0.331	LMF_FreeChol_LDL	1.532 (1.043-2.251)	0.033	0.246				
					LMF_Chol_IDL	1.5 (1.034-2.177)	0.037	0.246				
					SubTrigl_VLDL-1	1.533 (1.044-2.251)	0.037	0.246				
					LMF_Phosp_HDL	0.669 (0.458-0.977)	0.041	0.257				
					3-HB	1.434 (1.013-2.032)	0.045	0.403				

* Main abbreviations: 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 3. ROC curve models for t_1, t_2, and Δ(pre-post)$_{rt\text{-}PA}$ metabolites/lipids values considering early (sICH, non-response to thrombolysis) and late (three-month mortality and three-month poor functional outcome) outcomes. AUC values are reported both for models built considering only clinical characteristics, risk factors, and comorbidities (referred to as Bas1 or Bas2 for t_1, Δ(pre-post)$_{rt\text{-}PA}$, and t_2 samples, respectively) and for models obtained after adding selected combinations among the top three statistically significant metabolomic features, selected from previous penalized logistic regression analysis. Models obtained after combining metabolic features and circulating inflammatory markers statistically associated with sICH, mortality, and poor functional outcome are also described. 95% CIs and *P*-values (*P*) are also reported.

	Early outcomes								**Late outcomes**							
	sICH				**non-response to thrombolysis**				**three-month mortality**				**three-month poor functional outcome**			
	model	**AUC**	**95% CI**	***P***	**model**	**AUC**	**95% CI**	***P***	**Model**	**AUC**	**95% CI**	***P***	**model**	**AUC**	**95% CI**	***P***
t_1	Bas1	0.58	0.454-0.700	/	Bas1	0.54	0.467-0.614	/	Bas1	0.72	0.612-0.836	/	Bas1	0.81	0.758-0.871	/
	Bas1+A	0.72	0.609-0.826	0.825	Bas1+D	0.83	0.774-0.880	0.193	Bas1+G	0.59	0.459-0.713	0.349	Bas1+J	0.55	0.479-0.625	0.258
	Bas1+A+M	0.75	0.651-0.855	0.280					Bas1+G+P	0.51	0.371-0.646	0.87				
t_2	Bas2	0.56	0.438-0.673	/	Bas2	0.56	0.486-0.635	/	Bas2	0.76	0.662-0.854	/	Bas2	0.81	0.748-0.866	/
	Bas2+H	0.64	0.536-0.736	**0.039**	Bas2+K	0.62	0.545-0.69	**0.015**	Bas2+B	0.80	0.730-0.880	0.155	Bas2+E	0.84	0.791-0.895	**0.037**
Δ(pre-post)$_{rt\text{-}PA}$	Bas1	0.72	0.612-0.836	/	Bas1	0.81	0.758-0.871	/	Bas1	0.58	0.454-0.700	/	Bas1	0.54	0.467-0.614	/
	Bas1+I	0.65	0.531-0.776	0.063	Bas1+L	0.58	0.509-0.654	0.062	Bas1+C	0.74	0.636-0.85	0.29	Bas1+F	0.81	0.756-0.87	0.57
	Bas1+I+Q	0.75	0.645-0.854	**$2.07^{*}10^{-5}$**					Bas1+C+N	0.69	0.571-0.818	0.29	Bas1+F+O	0.80	0.739-0.859	0.83

Abbreviations used: Bas1: correction for age, gender, blood collection center, time onset-to-treatment, recent infections or inflammations, baseline values of glycemia, NIHSS, history of atrial fibrillation, and congestive heart failure;

Bas2: correction for 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-6 triglycerides; 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; M: ΔMMP9, SAP and A2M; N: ΔIL-IRa, ΔIL-10, ΔTNFα; O: ΔIL-6, ΔIL-8, ΔIL-10, ΔIL-12, ΔMCPI, ΔTNFα, ΔCRP; P: IL-Ira, IL-10; Q: ΔCRP, ΔIL-10.

FIGURES LEGENDS

Figure 1: Graphical representation of the analysis followed to identify, in serum samples, possible predictors of early and late adverse outcomes of acute ischemic stroke (AIS) treated with intravenous thrombolysis with recombinant tissue plasminogen activator (rt-PA). Blood samples were collected before (t_1) and 24h after (t_2) the administration of rt-PA. Early adverse outcomes were defined 24h after the event, as follows: development of symptomatic intracerebral haemorrhage (sICH) and non-response to the intravenous thrombolysis. Late outcomes (mortality and disability (mRS 3-6)) were defined three-months after the transient ischemia. For each time-point, 1D NMR spectra have been acquired and therefore used to estimate metabolites and lipids concentrations; t_1, t_2, and $\Delta(\text{pre-post})_{\text{rt-PA}}$ concentrations of metabolites/lipoprotein.

Figure 2: Scatter plots of Principal component analysis (PCA), taking into account the first two principal components (PC). The red dots represent the serum metabolic profiles collected before (t_1) the administration of rt-PA and the green dots represent the serum metabolic profiles collected 24h after (t_2) the administration of rt-PA.

SUPPORTING INFORMATION

The manuscript contains supporting information. Table S1: Complete list of metabolites and lipoproteins correctly assigned and quantified in serum NMR spectra. **Table S2:** Comparison between demographic and clinical characteristics of patients enrolled in the original study ($n = 327$) and in the presented metabolomic study ($n = 243$). **Table S3**: Pearson correlation analysis performed to find association between metabolites and lipids levels at pre (t_1) rt-PA and inflammatory markers. **Table S4**: Pearson correlation analysis performed to find association between metabolites and lipids levels at pre (t_2) rt-PA and inflammatory markers. **Table S5**. Complete list of the effects of pre (t_1) rt-PA metabolites and lipids levels on early (*i.e.* symptomatic intracerebral haemorrhage (sICH) and non-response to intravenous thrombolysis intervention) and late (*i.e.* three-month mortality and three-month poor functional outcome) outcomes. **Table S6**: Complete list of the effects of 24h post (t_2) rt-PA metabolites and lipids levels on each evaluated outcome. **Table S7**: Complete list of the effects of $\Delta(\text{pre-post})_{\text{rt-PA}}$ metabolites and lipids levels on each evaluated outcome.

AUTHOR INFORMATION

Corresponding Author

*Elena Sticchi (elena.sticchi@unifi.it)

Author Contributions

B.P., V.P., M.N., P.N., A.M.G, B.G., E.S., and D.I. designed the study. B.G., C.S., A.K., R.D.C, F.A., R.M., A.M.G, E.S., recruited patients for the study and collected serum samples. C.Li. collected NMR data. C.Li., and F.D.C. performed statistical analyses. C.Li., L.T., F.D.C., and C.Lu. interpreted the data and wrote the manuscript. All authors read, amended, and approved the final manuscript.

Funding Sources

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: MAGIC Study was funded by grants from Italian Ministry of Health,

2006 Finalized Research Programmes (RFPS-2006-1-336520) and Ente Cassa di Risparmio di Firenze (2010.06.03)

ACKNOWLEDGMENT

The authors acknowledge the support and the use of resources of Instruct-ERIC, a Landmark ESFRI project, and specifically the CERM/CIRMMP Italy Centre

ABBREVIATIONS

A2M, alpha2 macroglubulin; Apo, Apolipoprotein; AUC, Area Under the Curve; CRP, C reactive Protein; HDL, High Density Lipoprotein; ILs, interleukins; LDL, Low Density Lipoprotein; MAD, Median Absolute Deviation; MAGIC, MArker bioloGici nell'Ictus Cerebrale; MCPI, monocyte chemo-attractant protein I; MMP9, metalloproteinase 9; mRS, ; NIHSS, ; ROC, 2.7.3 Receiver Operating Characteristic curves: rT-PA, ; SAP, serum amyloid protein; sICH, secondary intracerebral hemorrhage; TNFα, tumour necrosis factor alpha; VLDL, Very Low Density Lipoprotein.

REFERENCE

(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) 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.

(3) 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.

(4) 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.

(5) 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.

(6) 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.

(7) 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.

(8) 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.

(9) 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.

(10) 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.

(11) 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.

(12) 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.

(13) 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.

(14) 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.

(15) 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.

(16) 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.

(17) Licari, C.; Tenori, L.; Giusti, B.; Sticchi, E.; Kura, A.; De Cario, R.; Inzitari, D.; Piccardi, B.; Nesi, M.; Sarti, C.; Arba, F.; Palumbo, V.; Nencini, P.; Marcucci, R.; Gori, A. M.; Luchinat, C.; Saccenti, E. Analysis of Metabolite and Lipid Association Networks Reveals Molecular Mechanisms Associated with 3-Month Mortality and Poor Functional Outcomes in Patients with Acute Ischemic Stroke after Thrombolytic Treatment with Recombinant Tissue Plasminogen Activator. *J Proteome Res* **2021**, *20* (10), 4758–4770. https://doi.org/10.1021/acs.jproteome.1c00406.

(18) 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.; Soinne, 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.

(19) Larrue Vincent; von Kummer Rüdiger; Müller Achim; Bluhmki Erich. Risk Factors for Severe Hemorrhagic Transformation in Ischemic Stroke Patients Treated With Recombinant Tissue Plasminogen Activator. *Stroke* **2001**, *32* (2), 438–441. https://doi.org/10.1161/01.STR.32.2.438.

(20) 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.* **2011**, *38A* (5), 197–220. https://doi.org/10.1002/cmr.a.20223.

(21) Meiboom, S.; Gill, D. Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. *Review of Scientific Instruments* **1958**, *29* (8), 688–691.

(22) 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.

(23) 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.

(24) Wall, M. E.; Rechtsteiner, A.; Rocha, L. M. Singular Value Decomposition and Principal Component Analysis. In *A Practical Approach to Microarray Data Analysis*; Berrar, D. P., Dubitzky, W., Granzow, M., Eds.; Springer US: Boston, MA, 2003; pp 91–109. https://doi.org/10.1007/0-306-47815-3_5.

(25) Hotelling, H. Analysis of a Complex of Statistical Variables into Principal Components. *Journal of Educational Psychology* **1933**, *24* (6), 417–441. https://doi.org/10.1037/h0071325.

(26) Pearson's Correlation Coefficient. In *Encyclopedia of Public Health*; Kirch, W., Ed.; Springer Netherlands: Dordrecht, 2008; pp 1090–1091. https://doi.org/10.1007/978-1-4020-5614-7_2569.

(27) Hauke, J.; Kossowski, T. Comparison of Values of Pearson's and Spearman's Correlation Coefficients on the Same Sets of Data. *Quaestiones Geographicae* **2011**, *30* (2), 87–93. https://doi.org/10.2478/v10117-011-0021-1.

(28) 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.

(29) Araveeporn, A. The Penalized Regression and Penalized Logistic Regression of Lasso and Elastic Net Methods for High- Dimensional Data: A Modelling Approach. *Innovations in Science and Technology Vol. 3* **2022**, 28–48. https://doi.org/10.9734/bpi/ist/v3/1695B.

(30) Ihaka, R.; Gentleman, R. R: A Language for Data Analysis and Graphics. *J Comput Stat Graph* **1996**, *5*, 299–314.

(31) Puhr, R.; Heinze, G.; Nold, M.; Lusa, L.; Geroldinger, A. Firth's Logistic Regression with Rare Events: Accurate Effect Estimates and Predictions? *Statistics in Medicine* **2017**, *36* (14), 2302–2317. https://doi.org/10.1002/sim.7273.

(32) Robin, X.; Turck, N.; Hainard, A.; Tiberti, N.; Lisacek, F.; Sanchez, J.-C.; Müller, M. PROC: An Open-Source Package for R and S+ to Analyze and Compare ROC Curves. *BMC Bioinformatics* **2011**, *12* (1), 77. https://doi.org/10.1186/1471-2105-12-77.

(33) Villanueva, R. A. M.; Chen, Z. J. Ggplot2: Elegant Graphics for Data Analysis (2nd Ed.). *Measurement: Interdisciplinary Research and Perspectives* **2019**, *17* (3), 160–167. https://doi.org/10.1080/15366367.2019.1565254.

(34) 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.

(35) *Association between change in plasma triglyceride levels and risk of stroke and carotid atherosclerosis: systematic review and meta-regression anal... - PubMed - NCBI.* https://www.ncbi.nlm.nih.gov/pubmed/20457452 (accessed 2020-05-12).

(36) 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.

(37) 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.

(38) 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.

(39) 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.

Supplementary tables

Table S1: Complete list of metabolites and lipoproteins correctly assigned and quantified in serum NMR spectra

Metabolites	Lipoproteins						
Creatinine	Trigl	LMF_Trigl_IDL	LMF_ApoB_IDL	SubPhosp_VLDL-2	SubFreeChol_LDL-3	SubTrigl_HDL-3	SubApoA2_HDL-1
Ala	Chol	LMF_Trigl_LDL	LMF_ApoB_LDL	SubPhosp_VLDL-3	SubFreeChol_LDL-4	SubTrigl_HDL-4	SubApoA2_HDL-2
Glu	LDL-Chol	LMF_Trigl_HDL	SubTrigl_VLDL-1	SubPhosp_VLDL-4	SubFreeChol_LDL-5	SubChol_HDL-1	SubApoA2_HDL-3
Gln	HDL-Chol	LMF_Chol_VLDL	SubTrigl_VLDL-2	SubPhosp_VLDL-5	SubFreeChol_LDL-6	SubChol_HDL-2	SubApoA2_HDL-4
Gly	Apo-A1	LMF_Chol_IDL	SubTrigl_VLDL-3	SubTrigl_LDL-1	SubPhosp_LDL-1	SubChol_HDL-3	
His	Apo-A2	LMF_Chol_LDL	SubTrigl_VLDL-4	SubTrigl_LDL-2	SubPhosp_LDL-2	SubChol_HDL-4	
Ile	Apo-B100	LMF_Chol_HDL	SubTrigl_VLDL-5	SubTrigl_LDL-3	SubPhosp_LDL-3	SubFreeChol_HDL-1	
Leu	Apo-B100-Apo-A1	LMF_FreeChol_VLDL	SubChol_VLDL-1	SubTrigl_LDL-4	SubPhosp_LDL-4	SubFreeChol_HDL-2	
Phe	VLDL_PN	LMF_FreeChol_IDL	SubChol_VLDL-2	SubTrigl_LDL-5	SubPhosp_LDL-5	SubFreeChol_HDL-3	
Tyr	IDL_PN	LMF_FreeChol_LDL	SubChol_VLDL-3	SubTrigl_LDL-6	SubPhosp_LDL-6	SubFreeChol_HDL-4	
Val	LDL_PN	LMF_FreeChol_HDL	SubChol_VLDL-4	SubChol_LDL-1	SubApoB_LDL-1	SubPhosp_HDL-1	
Acetic acid	LDL1_PN	LMF_Phosp_VLDL	SubChol_VLDL-5	SubChol_LDL-2	SubApoB_LDL-2	SubPhosp_HDL-2	
Citric acid	LDL2_PN	LMF_Phosp_IDL	SubFreeChol_VLDL-1	SubChol_LDL-3	SubApoB_LDL-3	SubPhosp_HDL-3	
Lactic acid	LDL3_PN	LMF_Phosp_LDL	SubFreeChol_VLDL-2	SubChol_LDL-4	SubApoB_LDL-4	SubPhosp_HDL-4	
3-HB	LDL4_PN	LMF_Phosp_HDL	SubFreeChol_VLDL-3	SubChol_LDL-5	SubApoB_LDL-5	SubApoA1_HDL-1	
Acetone	LDL5_PN	LMF_ApoA1_HDL	SubFreeChol_VLDL-4	SubChol_LDL-6	SubApoB_LDL-5	SubApoA1_HDL-2	
Pyruvic acid	LDL6_PN	LMF_ApoA2_HDL	SubFreeChol_VLDL-5	SubFreeChol_LDL-1	SubTrigl_HDL-1	SubApoA1_HDL-3	
Glucose	LMF_Trigl_VLDL	LMF_ApoB_VLDL	SubPhosp_VLDL-1	SubFreeChol_LDL-2	SubTrigl_HDL-2	SubApoA1_HDL-4	

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

Table S2: Comparison between demographic and clinical characteristics of patients enrolled in the original study (*n* = 327) and the presented metabolomic study (*n* = 243).

Demographics	Patients cohort of the presented study (*n* = 243)	Patients cohort of the original study (*n* = 327)	*P*-value
Age, years, mean, and SD	68.8 ± 11.9	68.9 ± 12	0.88
Sex (male), n (%)	137/243 (56.4%)	190/327 (58.1%)	0.68
Onset to treatment time, minutes, mean, and SD	163.4 ± 83.7	163.5 ± 75.7	0.86
Baseline NIHSS, mean, and SD	11.9 ± 6.1	11.9 ± 6.0	0.94
Baseline systolic blood pressure, mmHg, mean, and SD	147.5 ± 21.3	148.2 ± 21.7	0.69
Baseline diastolic blood pressure, mmHg, mean, and SD	79.7 ± 12.7	80.1 ± 12.7	0.71
Blood glucose, mg/dL, mean, and SD	130.2 ± 49.5	130.2 ± 47.9	0.99
Risk factors			
Hypertension, n (%)	143/243 (58.8%)	197/327 (61.0%)	0.77
Diabetes, n (%)	36/243 (14.8%)	50/327 (15.4%)	0.89
Hyperlipidemia, n (%)	56/243 (23%)	81/327 (25.8%)	0.68
Current smoking, n (%)	35/243 (14.4%)	51/327 (15.9%)	0.70
Atrial fibrillation, n (%)	56/243 (23%)	73/327 (22.7%)	0.83
Congestive heart failure, n (%)	26/243 (10.7%)	35/327 (10.9%)	0.98

Abbreviation used: SD = Standard Deviation; NIHSS = National Institute of Health Stroke Scale

Table S3: Pearson correlation analysis bewteen molecular features and inflammatory markers evalutated in t_1. Only stastistically significant correlation were reported (*P*-value<0.05). For completeness FDR *P*-value was also reported

Molecular features	**Inflammatory markers**	***P*-value**	**FDR *P*-value**	**Correlation coefficient**
SubApoA2_HDL.4	CRP	2.82E-05	9.43E-05	0.265
SubApoA1_HDL.4	CRP	6.85E-05	0.0002	0.253
SubChol_HDL.4	CRP	0.0001	0.0004	0.244
SubFreeChol_HDL.4	CRP	0.0002	0.0007	0.235
SubPhosp_HDL.4	CRP	0.0003	0.001	0.228
SubApoA2_HDL.4	IL6	0.003	0.007	0.191
SubFreeChol_HDL.4	IL8	0.003	0.008	0.189
SubPhosp_HDL.4	IL6	0.004	0.009	0.185
SubChol_HDL.4	IL6	0.004	0.010	0.183
SubPhosp_HDL.4	IL8	0.008	0.019	0.169
SubChol_HDL.4	IL8	0.010	0.023	0.165
SubPhosp_LDL.4	CRP	0.012	0.027	0.162
LDL4_PN	CRP	0.012	0.028	0.160
SubApoB_LDL.4	CRP	0.012	0.028	0.160
SubTrigl_VLDL.3	IL12	0.013	0.028	0.160
SubChol_LDL.4	CRP	0.014	0.030	0.158
SubApoA1_HDL.4	IL6	0.014	0.031	0.158
LMF_FreeChol_LDL	CRP	0.016	0.036	0.154
Creatinine	MCPI	0.017	0.036	0.153
SubFreeChol_HDL.3	CRP	0.018	0.039	0.152
Glucose	TNFalpha	0.019	0.041	0.150
SubFreeChol_LDL.4	CRP	0.019	0.042	0.150
Apo.A2	CRP	0.022	0.046	0.147
SubFreeChol_LDL.5	CRP	0.023	0.049	0.146
SubTrigl_VLDL.4	IL12	0.024	0.051	0.145
SubChol_LDL.3	CRP	0.024	0.051	0.145
SubFreeChol_HDL.4	IL6	0.030	0.061	0.140
Glucose	IL8	0.030	0.063	0.139
SubTrigl_HDL.4	CRP	0.034	0.070	0.136
SubFreeChol_LDL.3	CRP	0.035	0.072	0.135
SubPhosp_VLDL.3	IL12	0.037	0.074	0.134
SubApoA2_HDL.4	IL8	0.037	0.074	0.134
SubFreeChol_HDL.4	IL10	0.039	0.078	0.133
SubPhosp_LDL.3	CRP	0.039	0.079	0.132
SubPhosp_VLDL.2	IL12	0.043	0.086	0.130
LMF_ApoA2_HDL	CRP	0.044	0.086	0.129
SubPhosp_LDL.5	CRP	0.047	0.093	0.127
SubApoA2_HDL.4	MCPI	0.048	0.093	0.127
SubTrigl_LDL.4	IL10	0.048	0.094	0.127
SubChol_HDL.3	IL8	0.048	0.095	-0.127
LMF_FreeChol_HDL	A2M	0.047	0.091	-0.128
LMF_FreeChol_HDL	A2M.1	0.047	0.091	-0.128

Val	IL10	0.046	0.091	-0.128
His	IL8	0.046	0.090	-0.128
Apo.A1	IL10	0.045	0.089	-0.129
SubTrigl_HDL.2	TNFalpha	0.044	0.087	-0.129
SubApoA1_HDL.2	IL12	0.044	0.086	-0.129
SubApoA1_HDL.1	IL1ra	0.043	0.084	-0.130
Val	TNFalpha	0.042	0.084	-0.130
SubChol_VLDL.1	IL10	0.042	0.084	-0.130
SubChol_HDL.3	IL10	0.041	0.082	-0.131
LMF_Phosp_HDL	MCPI	0.041	0.082	-0.131
SubPhosp_LDL.6	IL10	0.039	0.078	-0.132
SubChol_HDL.1	MMP9	0.039	0.078	-0.133
SubApoA2_HDL.1	IL12	0.038	0.076	-0.133
SubApoA1_HDL.3	IL12	0.038	0.076	-0.133
SubTrigl_VLDL.5	IL8	0.037	0.075	-0.134
Glu	IL8	0.037	0.074	-0.134
SubTrigl_LDL.6	IL10	0.037	0.074	-0.134
SubFreeChol_LDL.2	IL8	0.036	0.073	-0.135
Apo.A1	IL8	0.035	0.071	-0.135
SubFreeChol_LDL.1	IL8	0.034	0.069	-0.136
SubPhosp_HDL.1	IL12	0.034	0.069	-0.136
Gly	IL1ra	0.034	0.069	-0.136
Pyruvicacid	IL10	0.034	0.069	-0.136
SubFreeChol_HDL.1	IL12	0.033	0.067	-0.137
Glu	IL10	0.031	0.064	-0.138
Gly	IL10	0.031	0.063	-0.139
SubChol_HDL.2	IL1ra	0.030	0.063	-0.139
SubChol_VLDL.5	MCPI	0.030	0.062	-0.139
SubFreeChol_VLDL.5	IL10	0.030	0.062	-0.139
Val	IL12	0.029	0.060	-0.140
SubTrigl_HDL.1	MCPI	0.029	0.060	-0.140
Pyruvicacid	IL1ra	0.029	0.060	-0.140
Glu	IL1ra	0.028	0.059	-0.141
Ala	IL12	0.027	0.057	-0.142
SubTrigl_HDL.1	IL1ra	0.026	0.055	-0.143
SubApoB_LDL.6	IL10	0.026	0.054	-0.143
LDL6_PN	IL10	0.026	0.054	-0.143
SubTrigl_HDL.1	IL6	0.024	0.051	-0.145
LMF_Phosp_HDL	IL8	0.024	0.050	-0.145
SubChol_HDL.1	MCPI	0.023	0.050	-0.145
SubPhosp_VLDL.4	IL6	0.023	0.050	-0.145
LMF_Phosp_HDL	IL10	0.023	0.049	-0.146
SubChol_HDL.2	CRP	0.023	0.049	-0.146
SubPhosp_HDL.2	IL6	0.022	0.047	-0.147

LMF_ApoA1_HDL	IL10	0.022	0.047	-0.147
Pyruvicacid	MCPI	0.020	0.044	-0.149
SubPhosp_HDL.2	IL1ra	0.020	0.043	-0.149
Phe	TNFalpha	0.020	0.042	-0.150
Gly	CRP	0.020	0.042	-0.150
LMF_FreeChol_HDL	MCPI	0.019	0.042	-0.150
LMF_FreeChol_HDL	IL8	0.019	0.042	-0.150
LMF_FreeChol_IDL	IL6	0.019	0.041	-0.150
SubTrigl_VLDL.5	IL10	0.019	0.040	-0.151
Val	CRP	0.018	0.039	-0.152
SubApoA1_HDL.1	IL12	0.018	0.039	-0.152
SubTrigl_VLDL.5	IL1ra	0.017	0.037	-0.153
Ile	TNFalpha	0.017	0.037	-0.153
SubTrigl_HDL.3	IL1ra	0.017	0.037	-0.153
SubFreeChol_HDL.2	IL12	0.016	0.036	-0.154
SubFreeChol_HDL.2	IL10	0.016	0.035	-0.155
LMF_ApoA1_HDL	IL8	0.016	0.035	-0.155
SubPhosp_HDL.1	MCPI	0.016	0.035	-0.155
Lactic acid	MCPI	0.016	0.035	-0.155
Glu	IL12	0.016	0.034	-0.155
SubPhosp_LDL.1	IL8	0.016	0.034	-0.155
LMF_Trigl_HDL	IL1ra	0.015	0.033	-0.156
SubPhosp_HDL.2	CRP	0.015	0.033	-0.156
SubFreeChol_HDL.2	IL8	0.015	0.033	-0.156
SubChol_VLDL.1	IL8	0.015	0.032	-0.156
Ile	IL10	0.014	0.032	-0.157
SubTrigl_HDL.2	MCPI	0.014	0.032	-0.157
Ile	IL1ra	0.014	0.032	-0.157
Phe	IL6	0.014	0.031	-0.157
Val	IL8	0.014	0.031	-0.158
LMF_ApoB_IDL	IL6	0.013	0.030	-0.158
Ile	IL8	0.013	0.030	-0.159
IDL_PN	IL6	0.013	0.030	-0.159
His	IL10	0.013	0.030	-0.159
SubPhosp_VLDL.5	MCPI	0.013	0.030	-0.159
LMF_Chol_IDL	IL6	0.013	0.030	-0.159
SubFreeChol_HDL.1	MMP9	0.013	0.029	-0.160
SubChol_LDL.1	IL8	0.013	0.029	-0.160
SubApoA1_HDL.1	MCPI	0.012	0.028	-0.160
LMF_FreeChol_HDL	IL12	0.012	0.028	-0.160
SubPhosp_VLDL.5	CRP	0.012	0.028	-0.160
LMF_Trigl_HDL	IL6	0.012	0.027	-0.161
HDL.Chol	IL8	0.012	0.027	-0.161
LMF_Chol_HDL	IL8	0.012	0.027	-0.161
His	IL12	0.012	0.027	-0.162
HDL.Chol	IL10	0.012	0.027	-0.162

LMF_Chol_HDL	IL10	0.012	0.027	-0.162
Lactic acid	TNFalpha	0.011	0.026	-0.162
SubPhosp_VLDL.5	IL1ra	0.010	0.023	-0.165
SubTrigl_HDL.3	CRP	0.010	0.023	-0.165
Phe	IL1ra	0.009	0.022	-0.166
His	IL1ra	0.009	0.021	-0.168
SubTrigl_HDL.3	IL10	0.009	0.020	-0.168
SubChol_HDL.2	IL6	0.008	0.019	-0.169
SubFreeChol_HDL.1	IL8	0.008	0.019	-0.169
SubTrigl_HDL.2	IL12	0.008	0.019	-0.170
SubApoA2_HDL.2	IL10	0.008	0.019	-0.170
SubFreeChol_VLDL.5	IL8	0.008	0.018	-0.171
LMF_Trigl_HDL	IL10	0.008	0.018	-0.171
Gly	IL6	0.008	0.018	-0.171
SubPhosp_VLDL.5	IL6	0.008	0.018	-0.171
LDL1_PN	IL8	0.007	0.018	-0.171
SubApoB_LDL.1	IL8	0.007	0.018	-0.171
Lactic acid	IL1ra	0.007	0.018	-0.171
SubChol_VLDL.5	CRP	0.007	0.017	-0.172
SubApoA1_HDL.3	IL10	0.007	0.016	-0.173
Glu	MCPI	0.007	0.016	-0.173
TPN	IL6	0.007	0.016	-0.174
Apo.B100	IL6	0.007	0.016	-0.174
Tyr	IL12	0.006	0.014	-0.176
Leu	IL6	0.006	0.014	-0.177
Ile	IL12	0.006	0.014	-0.177
LMF_FreeChol_HDL	MMP9	0.005	0.013	-0.178
SubTrigl_VLDL.5	MCPI	0.005	0.013	-0.178
SubTrigl_LDL.2	IL6	0.005	0.012	-0.179
Leu	CRP	0.005	0.012	-0.179
SubApoA1_HDL.3	IL8	0.005	0.012	-0.180
SubTrigl_LDL.3	MCPI	0.005	0.012	-0.180
SubChol_VLDL.5	IL6	0.005	0.012	-0.180
Citricacid	CRP	0.005	0.012	-0.180
SubApoA2_HDL.2	IL8	0.005	0.012	-0.181
SubApoA1_HDL.2	IL8	0.005	0.011	-0.181
SubChol_VLDL.4	IL6	0.004	0.011	-0.182
Leu	TNFalpha	0.004	0.011	-0.182
SubChol_HDL.2	MCPI	0.004	0.010	-0.184
LMF_FreeChol_HDL	IL10	0.004	0.010	-0.184
Leu	MCPI	0.004	0.010	-0.184
SubFreeChol_VLDL.4	IL6	0.004	0.009	-0.186
SubTrigl_HDL.3	IL8	0.003	0.008	-0.189
SubTrigl_LDL.1	IL6	0.003	0.007	-0.190

LMF_Trigl_LDL	IL6	0.003	0.007	-0.190
SubPhosp_HDL.2	MCPI	0.003	0.007	-0.190
SubTrigl_HDL.2	IL1ra	0.003	0.007	-0.191
SubChol_HDL.1	IL10	0.003	0.007	-0.192
SubFreeChol_HDL.1	IL10	0.003	0.007	-0.192
SubTrigl_HDL.1	IL10	0.003	0.007	-0.192
SubApoA1_HDL.2	IL10	0.003	0.007	-0.192
SubApoA2_HDL.1	IL8	0.003	0.007	-0.192
Phe	IL10	0.002	0.006	-0.194
SubTrigl_LDL.3	IL6	0.002	0.006	-0.196
Apo.B100.Apo.A1	IL6	0.002	0.005	-0.198
SubPhosp_HDL.1	IL10	0.002	0.005	-0.199
SubApoA2_HDL.1	IL10	0.002	0.005	-0.199
SubChol_HDL.2	IL12	0.002	0.005	-0.200
SubTrigl_HDL.2	IL10	0.001	0.004	-0.204
Gly	IL8	0.001	0.004	-0.204
SubChol_HDL.1	IL8	0.001	0.004	-0.204
SubTrigl_VLDL.5	IL6	0.001	0.004	-0.205
SubTrigl_HDL.2	CRP	0.001	0.003	-0.205
Leu	IL8	0.001	0.003	-0.208
SubTrigl_VLDL.5	CRP	0.001	0.003	-0.208
Lactic acid	IL10	0.001	0.003	-0.208
Leu	IL1ra	0.001	0.003	-0.208
SubPhosp_HDL.2	IL12	0.001	0.003	-0.209
Gly	IL12	0.0010	0.003	-0.211
LMF_Trigl_HDL	IL8	0.0009	0.003	-0.211
SubPhosp_HDL.1	IL8	0.0009	0.003	-0.211
SubChol_LDL.1	IL6	0.0009	0.002	-0.212
SubApoA1_HDL.1	IL10	0.0008	0.002	-0.214
SubChol_LDL.2	IL6	0.0006	0.002	-0.218
Pyruvicacid	IL8	0.0005	0.002	-0.221
SubPhosp_LDL.2	IL6	0.0005	0.001	-0.223
SubPhosp_LDL.1	IL6	0.0004	0.001	-0.224
SubFreeChol_LDL.1	IL6	0.0004	0.001	-0.224
Gly	TNFalpha	0.0003	0.0008	-0.231
Leu	IL10	0.0003	0.0008	-0.231
SubChol_HDL.2	IL10	0.0003	0.0008	-0.233
SubPhosp_HDL.2	IL10	0.0002	0.0007	-0.235
SubApoA1_HDL.1	IL8	0.0002	0.0006	-0.236
SubTrigl_HDL.1	IL8	0.0002	0.0005	-0.239
SubApoB_LDL.1	IL6	0.0002	0.0005	-0.239
LDL1_PN	IL6	0.0002	0.0005	-0.239
SubTrigl_HDL.2	IL8	0.0002	0.0005	-0.240
LDL2_PN	IL6	0.0001	0.0003	-0.246
SubApoB_LDL.2	IL6	0.0001	0.0003	-0.246
SubTrigl_HDL.3	IL6	6.31E-05	0.0002	-0.254

SubFreeChol_LDL.2	IL6	5.89E-05	0.0002	-0.255
Leu	IL12	5.49E-05	0.0002	-0.256
Phe	IL12	5.22E-05	0.0002	-0.256
SubPhosp_HDL.2	IL8	3.15E-05	0.0001	-0.264
SubChol_HDL.2	IL8	2.55E-05	8.57E-05	-0.267
SubTrigl_HDL.2	IL6	1.54E-05	5.32E-05	-0.273
Phe	CRP	1.52E-05	5.23E-05	-0.274
Lactic acid	IL8	4.67E-10	2.29E-09	-0.386

Table S4: Pearson correlation analysis bewteen molecular features and inflammatory markers evalutated in t_2. Only stastistically significant correlation were reported (*P*-value<0.05). For completeness FDR *P*-value was also reported

Molecular features	**Metalloproteins**	***P*-value**	**FDR *P*-value**	**Correlation coefficient**
SubTrigl_HDL.2	CRP	2.46E-07	9.45E-07	-0.324
SubTrigl_HDL.3	CRP	9.19E-07	3.36E-06	-0.309
SubTrigl_LDL.1	IL6	2.65E-06	9.31E-06	-0.296
SubTrigl_HDL.1	CRP	3.55E-06	1.23E-05	-0.292
LMF_Trigl_HDL	CRP	4.75E-06	1.63E-05	-0.289
SubTrigl_HDL.3	IL6	2.97E-05	9.47E-05	-0.264
SubTrigl_LDL.2	IL6	4.67E-05	0.0001	-0.258
Gly	CRP	6.05E-05	0.0002	-0.254
LMF_Trigl_LDL	IL6	6.11E-05	0.0002	-0.254
SubApoB_LDL.1	IL6	0.0001	0.0003	-0.245
LDL1_PN	IL6	0.0001	0.0003	-0.245
SubPhosp_LDL.1	IL6	0.0002	0.0005	-0.240
SubTrigl_HDL.2	IL6	0.0004	0.001	-0.225
LMF_Trigl_HDL	IL6	0.0004	0.001	-0.224
SubTrigl_LDL.3	IL6	0.0006	0.002	-0.218
SubChol_LDL.1	IL6	0.0007	0.002	-0.217
SubFreeChol_LDL.1	IL6	0.0007	0.002	-0.216
IDL_PN	CRP	0.001	0.003	-0.208
LMF_ApoB_IDL	CRP	0.001	0.003	-0.208
IDL_PN	IL6	0.001	0.003	-0.208
LMF_ApoB_IDL	IL6	0.001	0.003	-0.208
SubFreeChol_VLDL.5	CRP	0.002	0.004	-0.202
SubTrigl_HDL.4	IL8	0.002	0.004	-0.200
Gly	IL6	0.002	0.005	-0.199
SubApoA1_HDL.2	IL10	0.002	0.005	-0.199
SubTrigl_LDL.1	CRP	0.002	0.005	-0.198
Apo.A1	IL8	0.002	0.005	-0.196
LMF_Phosp_HDL	IL10	0.002	0.006	-0.194
LMF_ApoA1_HDL	IL8	0.003	0.006	-0.192
SubPhosp_HDL.3	IL10	0.003	0.007	-0.191

LMF_FreeChol_IDL	IL6	0.003	0.008	-0.189
SubTrigl_HDL.1	IL6	0.003	0.008	-0.187
SubPhosp_HDL.3	MCPI	0.004	0.010	-0.183
SubChol_VLDL.4	IL6	0.004	0.011	-0.182
SubApoA1_HDL.3	IL8	0.005	0.011	-0.182
SubFreeChol_VLDL.4	CRP	0.005	0.011	-0.181
SubPhosp_HDL.2	IL10	0.005	0.012	-0.179
LDL2_PN	IL6	0.005	0.012	-0.178
SubApoB_LDL.2	IL6	0.005	0.012	-0.178
LMF_Chol_IDL	IL6	0.005	0.013	-0.178
SubFreeChol_VLDL.4	IL6	0.006	0.014	-0.176
Apo.B100.Apo.A1	IL6	0.006	0.015	-0.175
LMF_Phosp_HDL	MCPI	0.006	0.015	-0.175
Ile	CRP	0.006	0.015	-0.174
SubChol_VLDL.1	CRP	0.007	0.015	-0.174
LMF_FreeChol_IDL	CRP	0.007	0.016	-0.173
LMF_Chol_IDL	CRP	0.008	0.018	-0.171
Pyruvicacid	MCPI	0.008	0.018	-0.171
SubTrigl_LDL.3	MCPI	0.008	0.019	-0.169
SubChol_VLDL.5	IL6	0.010	0.022	-0.166
VLDL_PN	IL6	0.010	0.022	-0.166
LMF_ApoB_VLDL	IL6	0.010	0.022	-0.166
SubPhosp_HDL.3	IL8	0.010	0.022	-0.165
SubApoA2_HDL.2	CRP	0.010	0.023	-0.165
SubPhosp_LDL.1	MCPI	0.010	0.023	-0.165
SubTrigl_LDL.2	CRP	0.011	0.025	-0.162
SubPhosp_VLDL.4	IL6	0.011	0.025	-0.162
Glucose	MCPI	0.012	0.027	-0.160
SubChol_VLDL.5	IL1ra	0.013	0.028	-0.160
SubPhosp_HDL.2	MCPI	0.013	0.028	-0.160
LMF_Trigl_LDL	CRP	0.013	0.029	-0.159
SubTrigl_VLDL.5	CRP	0.014	0.030	-0.158
Trigl	CRP	0.014	0.031	-0.157
Gln	IL8	0.014	0.031	-0.157
SubTrigl_HDL.4	IL6	0.014	0.031	-0.157
SubFreeChol_VLDL.1	CRP	0.014	0.031	-0.157
LMF_ApoA1_HDL	IL10	0.016	0.034	-0.155
Apo.A1	IL10	0.016	0.035	-0.154
SubApoA1_HDL.2	MCPI	0.016	0.035	-0.154
SubPhosp_LDL.1	TNFalpha	0.017	0.037	-0.153
LMF_ApoB_VLDL	CRP	0.017	0.037	-0.152
VLDL_PN	CRP	0.017	0.037	-0.152
SubChol_VLDL.5	A2M	0.018	0.038	-0.152
SubChol_VLDL.5	A2M.1	0.018	0.038	-0.152
SubFreeChol_LDL.2	IL6	0.018	0.039	-0.151
SubApoB_LDL.1	MCPI	0.019	0.040	-0.151

LDL1_PN	MCPI	0.019	0.040	-0.150
SubPhosp_LDL.2	IL6	0.019	0.040	-0.150
SubChol_HDL.2	IL10	0.020	0.042	-0.149
SubApoA1_HDL.3	IL10	0.020	0.042	-0.149
SubChol_VLDL.5	TNFalpha	0.020	0.043	-0.149
SubPhosp_LDL.1	IL8	0.021	0.044	-0.148
HDL.Chol	MMP9	0.021	0.044	-0.148
LMF_Chol_HDL	MMP9	0.021	0.044	-0.148
LDL1_PN	CRP	0.021	0.045	-0.147
SubChol_LDL.1	MCPI	0.021	0.045	-0.147
SubApoB_LDL.1	CRP	0.021	0.045	-0.147
SubTrigl_VLDL.5	A2M	0.022	0.045	-0.147
SubTrigl_VLDL.5	A2M.1	0.022	0.045	-0.147
SubChol_LDL.1	IL8	0.022	0.045	-0.147
SubFreeChol_LDL.1	IL8	0.022	0.046	-0.147
HDL.Chol	IL10	0.022	0.046	-0.147
LMF_Chol_HDL	IL10	0.022	0.046	-0.147
SubPhosp_HDL.2	CRP	0.022	0.047	-0.146
Gln	IL6	0.023	0.047	-0.146
LMF_Chol_VLDL	CRP	0.023	0.047	-0.146
Val	CRP	0.023	0.048	-0.146
SubPhosp_VLDL.5	IL6	0.023	0.048	-0.145
SubPhosp_VLDL.1	CRP	0.023	0.049	-0.145
LMF_Chol_VLDL	IL6	0.024	0.049	-0.145
LDL1_PN	TNFalpha	0.024	0.050	-0.145
TPN	IL6	0.024	0.050	-0.145
Apo.B100	IL6	0.024	0.050	-0.145
SubApoB_LDL.1	TNFalpha	0.024	0.050	-0.145
SubChol_VLDL.5	IL12	0.025	0.051	-0.144
SubChol_LDL.1	TNFalpha	0.025	0.052	-0.144
SubApoA1_HDL.1	IL10	0.026	0.053	-0.143
SubPhosp_VLDL.5	IL8	0.027	0.055	-0.142
SubTrigl_LDL.2	MCPI	0.027	0.055	-0.142
SubApoA1_HDL.4	IL8	0.027	0.056	-0.141
SubTrigl_LDL.1	MCPI	0.028	0.056	-0.141
SubChol_VLDL.4	CRP	0.028	0.056	-0.141
SubTrigl_VLDL.5	IL6	0.029	0.058	-0.140
SubChol_HDL.2	MCPI	0.030	0.061	-0.139
SubTrigl_HDL.4	IL1ra	0.030	0.061	-0.139
SubChol_VLDL.5	IL8	0.030	0.061	-0.139
Gln	MCPI	0.031	0.061	-0.139
SubFreeChol_VLDL.5	IL6	0.031	0.062	-0.139
SubChol_LDL.2	IL6	0.031	0.063	-0.138
LMF_Trigl_HDL	IL8	0.033	0.065	-0.137

SubChol_VLDL.5	MCPI	0.033	0.066	-0.137
SubPhosp_HDL.1	A2M	0.033	0.066	-0.137
SubPhosp_HDL.1	A2M.1	0.033	0.066	-0.137
SubFreeChol_LDL.1	TNFalpha	0.034	0.067	-0.136
SubChol_VLDL.2	IL6	0.034	0.068	-0.136
SubTrigl_HDL.1	A2M	0.034	0.068	-0.136
SubTrigl_HDL.1	A2M.1	0.034	0.068	-0.136
SubTrigl_LDL.2	TNFalpha	0.036	0.071	-0.135
SubTrigl_LDL.1	IL1ra	0.036	0.071	-0.134
SubApoA1_HDL.1	A2M	0.036	0.071	-0.134
SubApoA1_HDL.1	A2M.1	0.036	0.071	-0.134
LMF_Phosp_IDL	IL6	0.038	0.075	-0.133
SubPhosp_VLDL.4	CRP	0.039	0.076	-0.133
SubFreeChol_LDL.1	MCPI	0.039	0.076	-0.132
SubPhosp_VLDL.5	IL1ra	0.040	0.078	-0.132
LMF_Phosp_HDL	IL8	0.040	0.078	-0.132
Gly	TNFalpha	0.041	0.079	-0.131
SubTrigl_VLDL.1	CRP	0.041	0.079	-0.131
Leu	CRP	0.041	0.080	-0.131
SubChol_HDL.3	IL8	0.042	0.081	-0.131
SubApoB_LDL.1	IL8	0.042	0.082	-0.130
LDL1_PN	IL8	0.042	0.082	-0.130
SubPhosp_LDL.1	CRP	0.044	0.084	-0.129
Gln	IL1ra	0.044	0.085	-0.129
LMF_ApoA1_HDL	MMP9	0.044	0.085	-0.129
SubTrigl_LDL.5	IL6	0.045	0.086	-0.129
Trigl	IL6	0.045	0.086	-0.129
SubTrigl_LDL.4	IL6	0.045	0.087	-0.129
SubFreeChol_VLDL.3	CRP	0.045	0.087	-0.128
SubApoA1_HDL.1	CRP	0.046	0.088	-0.128
LMF_ApoA1_HDL	MCPI	0.046	0.089	-0.128
SubChol_HDL.1	A2M	0.047	0.089	-0.128
SubChol_HDL.1	A2M.1	0.047	0.089	-0.128
His	IL10	0.048	0.092	-0.127
SubFreeChol_LDL.6	IL12	0.050	0.094	0.126
SubFreeChol_VLDL.2	IL12	0.048	0.091	0.127
SubFreeChol_LDL.6	IL1ra	0.044	0.085	0.129
Glu	TNFalpha	0.041	0.080	0.131
SubChol_LDL.6	TNFalpha	0.041	0.079	0.131
SubTrigl_VLDL.3	IL12	0.041	0.079	0.131
SubFreeChol_VLDL.4	IL12	0.039	0.076	0.132
LMF_FreeChol_IDL	IL12	0.037	0.072	0.134
SubChol_VLDL.4	IL12	0.034	0.067	0.136
SubPhosp_VLDL.2	IL12	0.034	0.067	0.136
SubChol_LDL.6	IL12	0.033	0.066	0.137
Lactic acid	IL1ra	0.030	0.061	0.139

Citricacid	IL8	0.028	0.057	0.141
LDL4_PN	CRP	0.024	0.050	0.144
SubApoB_LDL.4	CRP	0.024	0.050	0.144
LMF_Chol_IDL	IL12	0.024	0.049	0.145
SubApoA2_HDL.4	IL6	0.022	0.046	0.147
Citricacid	IL12	0.022	0.045	0.147
Lactic acid	MCPI	0.020	0.042	0.149
Citricacid	IL10	0.019	0.040	0.150
SubFreeChol_HDL.4	IL6	0.019	0.040	0.151
Glucose	MMP9	0.019	0.039	0.151
SubFreeChol_VLDL.3	IL12	0.018	0.039	0.151
X3.HB	IL10	0.017	0.036	0.153
LDL3_PN	CRP	0.016	0.035	0.154
SubApoB_LDL.3	CRP	0.016	0.035	0.154
Phe	IL8	0.016	0.035	0.154
Lactic acid	IL8	0.015	0.032	0.156
SubTrigl_LDL.6	IL1ra	0.015	0.032	0.156
SubChol_HDL.4	IL6	0.014	0.031	0.157
SubChol_VLDL.3	IL12	0.011	0.025	0.163
SubPhosp_VLDL.3	IL12	0.009	0.020	0.168
SubChol_LDL.4	CRP	0.008	0.018	0.170
SubPhosp_LDL.4	CRP	0.008	0.018	0.171
Val	MMP9	0.008	0.018	0.171
Tyr	IL8	0.007	0.017	0.172
SubFreeChol_LDL.4	CRP	0.007	0.016	0.174
SubFreeChol_HDL.3	CRP	0.007	0.015	0.174
SubPhosp_LDL.3	CRP	0.003	0.008	0.187
SubFreeChol_LDL.3	CRP	0.002	0.005	0.199
SubApoA1_HDL.4	CRP	0.002	0.005	0.199
Lactic acid	IL12	0.002	0.005	0.199
SubChol_LDL.3	CRP	0.001	0.004	0.203
SubApoA2_HDL.4	CRP	0.0006	0.002	0.218
SubPhosp_HDL.4	CRP	0.0004	0.001	0.225
SubChol_HDL.4	CRP	6.43E-05	0.0002	0.253
SubFreeChol_HDL.4	CRP	3.10E-05	9.86E-05	0.264

Table S5: Effect of pre (t_1) rt-PA metabolites and lipids levels on early (*i.e.* sICH, non-response to thrombolysis) and late (*i.e.* three-month mortality and three-month mRS 3-6) adverse outcomes, adjusting for the major determinants for unfavourable outcomes.

	Early outcomes						**Late outcomes**					
	sICH			**Non-response to thrombolysis**			**Three-month mortality**			**Three-month mRS 3-6**		
Molecular features	**OR (95% CI)**	**P**	**FDR**	**OR (95% CI)**	**P**	**FDR**	**OR (95% CI)**	**P**	**FDR**	**OR (95% CI)**	**P**	**FDR**
Creatinine	1.072 (0.763-1.506)	0.725	0.998	1.272 (0.932-1.737)	0.102	0.615	1.512 (1.042-2.193)	0.100	0.777	0.908 (0.633-1.302)	0.619	0.880

Ala	0.844 (0.511-1.393)	0.549	0.998	0.935 (0.699-1.25)	0.650	0.909	1.144 (0.754-1.734)	0.562	0.869	1.235 (0.865-1.765)	0.262	0.880
Glu	0.955 (0.606-1.507)	0.860	0.998	1.086 (0.818-1.443)	0.570	0.909	1.263 (0.844-1.89)	0.300	0.869	0.837 (0.586-1.196)	0.334	0.880
Gln	1.188 (0.707-1.996)	0.557	0.998	0.931 (0.684-1.267)	0.650	0.909	1.043 (0.627-1.736)	0.884	0.936	0.879 (0.617-1.251)	0.484	0.880
Gly	0.819 (0.487-1.378)	0.489	0.998	0.906 (0.679-1.21)	0.507	0.909	1.158 (0.747-1.794)	0.549	0.869	1.047 (0.734-1.495)	0.806	0.931
His	1.075 (0.827-1.396)	0.635	0.998	0.92 (0.748-1.131)	0.424	0.909	1.149 (0.918-1.439)	0.220	0.793	1.103 (0.891-1.365)	0.367	0.880
Ile	0.985 (0.632-1.535)	0.951	0.998	0.981 (0.75-1.282)	0.886	0.940	1.003 (0.66-1.524)	0.990	0.990	1.004 (0.711-1.417)	0.984	0.984
Leu	0.977 (0.635-1.503)	0.925	0.998	0.948 (0.728-1.236)	0.697	0.909	1.144 (0.772-1.695)	0.552	0.869	1.087 (0.785-1.506)	0.636	0.880
Phe	0.977 (0.636-1.502)	0.925	0.998	0.877 (0.669-1.148)	0.341	0.909	1.095 (0.742-1.615)	0.676	0.869	1.232 (0.897-1.69)	0.216	0.880
Tyr	0.883 (0.561-1.392)	0.626	0.998	0.947 (0.715-1.255)	0.707	0.909	1.137 (0.748-1.728)	0.578	0.869	1.296 (0.92-1.826)	0.152	0.880
Val	0.737 (0.454-1.194)	0.252	0.998	1.011 (0.76-1.346)	0.940	0.940	1.598 (0.989-2.58)	0.074	0.777	1.099 (0.781-1.547)	0.598	0.880
Acetic acid	1.001 (0.621-1.615)	0.998	0.998	1.142 (0.853-1.528)	0.354	0.909	1.09 (0.781-1.52)	0.673	0.869	0.976 (0.715-1.333)	0.882	0.934
Citric acid	1.291 (0.867-1.922)	0.450	0.998	1.14 (0.786-1.655)	0.498	0.909	1.075 (0.738-1.566)	0.737	0.884	0.938 (0.611-1.44)	0.828	0.931
Lactic acid	0.845 (0.453-1.578)	0.642	0.998	1.017 (0.763-1.357)	0.908	0.940	1.431 (0.898-2.281)	0.197	0.793	1.093 (0.769-1.555)	0.634	0.880
3-HB	1.457 (1.062-1.998)	0.025	0.455	0.718 (0.531-0.97)	0.024	0.219	0.897 (0.605-1.33)	0.604	0.869	1.121 (0.833-1.508)	0.465	0.880
Acetone	1.393 (1.016-1.909)	0.046	0.502	0.696 (0.509-0.95)	0.017	0.219	0.963 (0.634-1.465)	0.874	0.936	1.148 (0.836-1.576)	0.407	0.880
Pyruvic acid	1.124 (0.713-1.77)	0.738	0.998	1.09 (0.827-1.435)	0.537	0.909	1.456 (1.021-2.077)	0.129	0.777	1.077 (0.76-1.525)	0.723	0.929
Glucose	0.828 (0.343-1.998)	0.712	0.998	0.96 (0.679-1.357)	0.818	0.940	1.293 (0.564-2.965)	0.578	0.869	1.364 (0.721-2.58)	0.354	0.880
Trigl	1.053 (0.684-1.62)	0.832	0.999	1.131 (0.862-1.484)	0.375	0.956	0.839 (0.497-1.417)	0.549	0.994	1.09 (0.768-1.547)	0.641	0.812
Chol	0.864 (0.541-1.38)	0.578	0.999	0.96 (0.721-1.278)	0.780	0.998	0.947 (0.568-1.58)	0.851	0.994	1.207 (0.841-1.734)	0.320	0.629
LDL-Chol	0.933 (0.593-1.468)	0.787	0.999	0.916 (0.688-1.221)	0.551	0.966	1.034 (0.632-1.69)	0.905	0.994	1.377 (0.965-1.965)	0.083	0.392
HDL-Chol	0.997 (0.638-1.558)	0.991	0.999	0.908 (0.679-1.216)	0.521	0.956	0.746 (0.457-1.216)	0.273	0.994	0.741 (0.516-1.063)	0.111	0.448
Apo-A1	0.936 (0.591-1.483)	0.795	0.999	1.035 (0.767-1.396)	0.822	0.998	0.704 (0.428-1.157)	0.195	0.994	0.698 (0.478-1.018)	0.066	0.387
Apo-A2	0.854 (0.54-1.35)	0.535	0.999	0.946 (0.7-1.277)	0.717	0.998	0.861 (0.523-1.416)	0.589	0.994	0.825 (0.574-1.187)	0.311	0.623
Apo-B100	0.967 (0.609-1.536)	0.898	0.999	1.033 (0.783-1.362)	0.820	0.998	0.94 (0.556-1.589)	0.836	0.994	1.403 (0.982-2.004)	0.068	0.387
Apo-B100-Apo-A1	0.996 (0.663-1.497)	0.987	0.999	0.988 (0.751-1.299)	0.932	0.998	1.212 (0.803-1.831)	0.395	0.994	1.63 (1.131-2.349)	0.007	0.209
VLDL_PN	1.073 (0.708-1.626)	0.762	0.999	1.203 (0.915-1.581)	0.183	0.956	0.857 (0.521-1.411)	0.579	0.994	1.068 (0.752-1.516)	0.722	0.858
IDL_PN	1.006 (0.651-1.553)	0.981	0.999	1.14 (0.868-1.499)	0.347	0.956	1.009 (0.624-1.63)	0.975	0.994	1.144 (0.807-1.621)	0.465	0.680
LDL_PN	0.975 (0.622-1.529)	0.920	0.999	0.942 (0.713-1.244)	0.674	0.998	1.095 (0.664-1.806)	0.747	0.994	1.392 (0.974-1.989)	0.075	0.387
LDL1_PN	1.165 (0.745-1.823)	0.542	0.999	1 (0.745-1.344)	0.998	0.998	0.936 (0.572-1.533)	0.812	0.994	1.006 (0.705-1.434)	0.975	0.984
LDL2_PN	1.131 (0.721-1.775)	0.623	0.999	0.998 (0.74-1.346)	0.991	0.998	0.77 (0.465-1.275)	0.353	0.994	1.153 (0.822-1.618)	0.419	0.680
LDL3_PN	1.01 (0.643-1.587)	0.968	0.999	0.879 (0.658-1.175)	0.384	0.956	1.271 (0.774-2.088)	0.394	0.994	1.335 (0.934-1.907)	0.121	0.448
LDL4_PN	0.957 (0.638-1.436)	0.842	0.999	0.933 (0.704-1.238)	0.632	0.998	1.401 (0.924-2.124)	0.164	0.994	1.109 (0.799-1.541)	0.546	0.731
LDL5_PN	0.884 (0.557-1.401)	0.626	0.999	1.025 (0.773-1.359)	0.866	0.998	1.081 (0.654-1.784)	0.784	0.994	1.25 (0.885-1.764)	0.216	0.564
LDL6_PN	0.827 (0.492-1.389)	0.515	0.999	0.978 (0.745-1.285)	0.874	0.998	0.949 (0.533-1.689)	0.875	0.994	1.418 (1.022-1.967)	0.044	0.331
LMF_Trigl_VLDL	1.007 (0.643-1.578)	0.978	0.999	1.111 (0.845-1.461)	0.452	0.956	0.871 (0.515-1.475)	0.644	0.994	1.154 (0.811-1.641)	0.442	0.680
LMF_Trigl_IDL	1.09 (0.717-1.656)	0.717	0.999	1.125 (0.86-1.471)	0.391	0.956	0.825 (0.489-1.392)	0.508	0.994	1.047 (0.741-1.479)	0.801	0.897
LMF_Trigl_LDL	1.103 (0.741-1.641)	0.659	0.999	1.156 (0.876-1.524)	0.305	0.956	1.034 (0.671-1.594)	0.887	0.994	1.146 (0.814-1.614)	0.449	0.680
LMF_Trigl_HDL	1.06 (0.691-1.625)	0.806	0.999	1.375 (1.016-1.862)	0.036	0.956	0.711 (0.43-1.177)	0.220	0.994	0.791 (0.554-1.13)	0.204	0.564
LMF_Chol_VLDL	0.906 (0.575-1.428)	0.698	0.999	1.209 (0.92-1.587)	0.171	0.956	0.856 (0.497-1.473)	0.614	0.994	1.069 (0.754-1.516)	0.718	0.858
LMF_Chol_IDL	0.892 (0.56-1.42)	0.660	0.999	1.14 (0.865-1.502)	0.354	0.956	1.039 (0.62-1.742)	0.896	0.994	1.239 (0.869-1.765)	0.251	0.584
LMF_Chol_LDL	0.933 (0.593-1.468)	0.787	0.999	0.916 (0.688-1.221)	0.551	0.966	1.034 (0.632-1.69)	0.905	0.994	1.377 (0.965-1.965)	0.083	0.392
LMF_Chol_HDL	0.997 (0.638-1.558)	0.991	0.999	0.908 (0.679-1.216)	0.521	0.956	0.746 (0.457-1.216)	0.273	0.994	0.741 (0.516-1.063)	0.111	0.448
LMF_FreeChol_VLDL	0.964 (0.611-1.52)	0.885	0.999	1.166 (0.887-1.534)	0.271	0.956	0.796 (0.458-1.383)	0.462	0.994	1.065 (0.748-1.517)	0.735	0.864
LMF_FreeChol_IDL	0.93 (0.588-1.471)	0.778	0.999	1.147 (0.871-1.511)	0.329	0.956	0.991 (0.594-1.655)	0.976	0.994	1.192 (0.835-1.702)	0.347	0.671
LMF_FreeChol_LDL	0.873 (0.551-1.382)	0.596	0.999	0.895 (0.669-1.198)	0.458	0.956	1.189 (0.713-1.985)	0.551	0.994	1.363 (0.947-1.961)	0.101	0.445
LMF_FreeChol_HDL	0.763 (0.486-1.197)	0.275	0.999	0.901 (0.67-1.211)	0.491	0.956	0.865 (0.532-1.407)	0.592	0.994	0.872 (0.591-1.287)	0.506	0.714

LMF_Phosp_VLDL	1 (0.637-1.572)	0.999	0.999	1.131 (0.858-1.491)	0.383	0.956	0.852 (0.505-1.438)	0.586	0.994	1.04 (0.729-1.484)	0.832	0.912
LMF_Phosp_IDL	0.967 (0.608-1.538)	0.898	0.999	1.118 (0.85-1.47)	0.425	0.956	0.945 (0.565-1.581)	0.846	0.994	1.155 (0.812-1.643)	0.438	0.680
LMF_Phosp_LDL	0.974 (0.618-1.537)	0.920	0.999	0.905 (0.678-1.207)	0.497	0.956	1.083 (0.659-1.779)	0.776	0.994	1.303 (0.911-1.866)	0.156	0.528
LMF_Phosp_HDL	1.079 (0.676-1.723)	0.772	0.999	0.955 (0.711-1.283)	0.762	0.998	0.793 (0.477-1.319)	0.409	0.994	0.584 (0.393-0.867)	0.008	0.209
LMF_ApoA1_HDL	0.989 (0.633-1.544)	0.964	0.999	1.072 (0.796-1.444)	0.649	0.998	0.627 (0.379-1.038)	0.086	0.994	0.674 (0.465-0.978)	0.040	0.331
LMF_ApoA2_HDL	0.882 (0.556-1.398)	0.625	0.999	0.951 (0.706-1.28)	0.740	0.998	0.841 (0.509-1.389)	0.535	0.994	0.821 (0.571-1.18)	0.298	0.623
LMF_ApoB_VLDL	1.073 (0.708-1.626)	0.762	0.999	1.203 (0.915-1.581)	0.183	0.956	0.857 (0.521-1.412)	0.580	0.994	1.068 (0.752-1.516)	0.722	0.858
LMF_ApoB_IDL	1.006 (0.652-1.553)	0.981	0.999	1.141 (0.868-1.499)	0.346	0.956	1.009 (0.624-1.63)	0.975	0.994	1.144 (0.807-1.621)	0.465	0.680
LMF_ApoB_LDL	0.975 (0.622-1.529)	0.920	0.999	0.942 (0.714-1.244)	0.675	0.998	1.095 (0.664-1.806)	0.747	0.994	1.392 (0.974-1.989)	0.075	0.387
SubTrigl_VLDL-1	1.01 (0.642-1.59)	0.968	0.999	1.048 (0.794-1.383)	0.743	0.998	0.855 (0.508-1.441)	0.592	0.994	1.16 (0.811-1.659)	0.433	0.680
SubTrigl_VLDL-2	0.87 (0.553-1.37)	0.576	0.999	1.152 (0.873-1.52)	0.317	0.956	1.072 (0.647-1.777)	0.807	0.994	1.246 (0.881-1.763)	0.233	0.564
SubTrigl_VLDL-3	0.894 (0.565-1.415)	0.660	0.999	1.199 (0.906-1.588)	0.203	0.956	0.995 (0.593-1.668)	0.985	0.994	1.227 (0.858-1.755)	0.278	0.610
SubTrigl_VLDL-4	1.038 (0.672-1.602)	0.878	0.999	1.221 (0.925-1.613)	0.157	0.956	0.904 (0.551-1.484)	0.716	0.994	1.102 (0.777-1.562)	0.596	0.781
SubTrigl_VLDL-5	1.169 (0.764-1.79)	0.510	0.999	1.099 (0.839-1.439)	0.496	0.956	0.758 (0.473-1.215)	0.298	0.994	0.801 (0.568-1.131)	0.219	0.564
SubChol_VLDL-1	0.993 (0.632-1.56)	0.977	0.999	1.113 (0.843-1.469)	0.451	0.956	0.819 (0.477-1.407)	0.510	0.994	1.074 (0.751-1.535)	0.706	0.858
SubChol_VLDL-2	0.866 (0.552-1.356)	0.563	0.999	1.224 (0.932-1.608)	0.143	0.956	0.976 (0.577-1.65)	0.935	0.994	1.06 (0.75-1.498)	0.750	0.872
SubChol_VLDL-3	0.832 (0.52-1.332)	0.482	0.999	1.222 (0.927-1.611)	0.153	0.956	1 (0.588-1.702)	0.999	0.999	1.184 (0.835-1.679)	0.360	0.680
SubChol_VLDL-4	0.855 (0.54-1.353)	0.536	0.999	1.249 (0.946-1.65)	0.115	0.956	0.928 (0.554-1.554)	0.796	0.994	1.042 (0.735-1.476)	0.824	0.912
SubChol_VLDL-5	1.293 (0.84-1.99)	0.283	0.999	1.113 (0.84-1.475)	0.456	0.956	0.701 (0.431-1.141)	0.186	0.994	0.618 (0.427-0.893)	0.011	0.209
SubFreeChol_VLDL-1	1.04 (0.664-1.63)	0.876	0.999	1.11 (0.842-1.463)	0.460	0.956	0.806 (0.467-1.392)	0.479	0.994	1.074 (0.752-1.535)	0.704	0.858
SubFreeChol_VLDL-2	0.951 (0.611-1.481)	0.839	0.999	1.207 (0.92-1.583)	0.174	0.956	0.928 (0.547-1.577)	0.806	0.994	1.149 (0.811-1.628)	0.451	0.680
SubFreeChol_VLDL-3	0.942 (0.599-1.481)	0.813	0.999	1.188 (0.903-1.562)	0.217	0.956	0.922 (0.545-1.562)	0.787	0.994	1.176 (0.829-1.669)	0.380	0.680
SubFreeChol_VLDL-4	0.904 (0.574-1.424)	0.691	0.999	1.234 (0.938-1.623)	0.133	0.956	0.966 (0.577-1.62)	0.907	0.994	1.08 (0.764-1.528)	0.671	0.841
SubFreeChol_VLDL-5	1.201 (0.778-1.854)	0.457	0.999	1.101 (0.828-1.464)	0.508	0.956	0.673 (0.391-1.158)	0.177	0.994	0.816 (0.56-1.189)	0.303	0.623
SubPhosp_VLDL-1	0.983 (0.625-1.548)	0.948	0.999	1.066 (0.809-1.404)	0.653	0.998	0.832 (0.489-1.416)	0.537	0.994	1.16 (0.811-1.659)	0.431	0.680
SubPhosp_VLDL-2	0.873 (0.559-1.362)	0.578	0.999	1.177 (0.895-1.547)	0.243	0.956	1.055 (0.633-1.758)	0.851	0.994	1.223 (0.865-1.729)	0.274	0.610
SubPhosp_VLDL-3	0.937 (0.599-1.467)	0.795	0.999	1.229 (0.933-1.618)	0.140	0.956	0.971 (0.583-1.619)	0.920	0.994	1.178 (0.829-1.673)	0.375	0.680
SubPhosp_VLDL-4	0.964 (0.621-1.497)	0.881	0.999	1.262 (0.957-1.663)	0.097	0.956	0.951 (0.578-1.565)	0.856	0.994	1.058 (0.747-1.497)	0.758	0.873
SubPhosp_VLDL-5	1.164 (0.756-1.792)	0.527	0.999	1.147 (0.869-1.514)	0.333	0.956	0.67 (0.409-1.095)	0.142	0.994	0.641 (0.444-0.926)	0.019	0.268
SubTrigl_LDL-1	1.193 (0.821-1.734)	0.392	0.999	1.145 (0.869-1.51)	0.336	0.956	0.923 (0.602-1.415)	0.732	0.994	0.992 (0.704-1.399)	0.966	0.984
SubTrigl_LDL-2	1.252 (0.856-1.832)	0.285	0.999	1.116 (0.842-1.48)	0.445	0.956	1.018 (0.675-1.536)	0.937	0.994	1.024 (0.724-1.448)	0.896	0.946
SubTrigl_LDL-3	1.205 (0.809-1.794)	0.407	0.999	1.004 (0.762-1.323)	0.977	0.998	1.045 (0.681-1.605)	0.853	0.994	0.998 (0.707-1.409)	0.992	0.992
SubTrigl_LDL-4	1.054 (0.694-1.602)	0.820	0.999	1.096 (0.825-1.457)	0.526	0.956	1.286 (0.831-1.99)	0.300	0.994	1.048 (0.74-1.483)	0.799	0.897
SubTrigl_LDL-5	0.951 (0.616-1.469)	0.835	0.999	1.093 (0.834-1.432)	0.522	0.956	1.285 (0.807-2.048)	0.336	0.994	1.092 (0.778-1.532)	0.621	0.795
SubTrigl_LDL-6	0.792 (0.476-1.318)	0.412	0.999	0.917 (0.699-1.201)	0.528	0.956	1.224 (0.758-1.975)	0.465	0.994	1.464 (1.059-2.025)	0.024	0.268
SubChol_LDL-1	1.125 (0.709-1.786)	0.651	0.999	0.971 (0.722-1.306)	0.845	0.998	0.928 (0.563-1.531)	0.792	0.994	0.99 (0.697-1.407)	0.956	0.984
SubChol_LDL-2	1.101 (0.703-1.726)	0.699	0.999	0.96 (0.714-1.291)	0.790	0.998	0.765 (0.463-1.265)	0.340	0.994	1.163 (0.831-1.628)	0.389	0.680
SubChol_LDL-3	0.959 (0.618-1.489)	0.864	0.999	0.853 (0.64-1.138)	0.279	0.956	1.23 (0.758-1.994)	0.450	0.994	1.326 (0.934-1.883)	0.122	0.448
SubChol_LDL-4	0.917 (0.606-1.389)	0.699	0.999	0.892 (0.671-1.187)	0.431	0.956	1.406 (0.932-2.12)	0.153	0.994	1.122 (0.81-1.554)	0.499	0.714
SubChol_LDL-5	0.898 (0.568-1.417)	0.667	0.999	0.993 (0.746-1.321)	0.961	0.998	1.055 (0.639-1.743)	0.850	0.994	1.242 (0.877-1.757)	0.233	0.564
SubChol_LDL-6	0.826 (0.496-1.377)	0.505	0.999	0.935 (0.71-1.232)	0.634	0.998	0.96 (0.547-1.687)	0.900	0.994	1.461 (1.049-2.035)	0.029	0.280
SubFreeChol_LDL-1	1.032 (0.643-1.655)	0.906	0.999	1.001 (0.742-1.35)	0.993	0.998	0.926 (0.555-1.543)	0.788	0.994	1.033 (0.726-1.471)	0.859	0.925
SubFreeChol_LDL-2	1.013 (0.646-1.586)	0.960	0.999	1.004 (0.742-1.36)	0.977	0.998	0.695 (0.414-1.167)	0.202	0.994	1.162 (0.826-1.634)	0.399	0.680

SubFreeChol_LDL-3	0.987 (0.623-1.564)	0.959	0.999	0.867 (0.648-1.16)	0.338	0.956	1.142 (0.696-1.875)	0.635	0.994	1.264 (0.883-1.808)	0.210	0.564
SubFreeChol_LDL-4	0.953 (0.621-1.461)	0.837	0.999	0.873 (0.653-1.169)	0.361	0.956	1.312 (0.817-2.108)	0.325	0.994	1.228 (0.868-1.737)	0.258	0.588
SubFreeChol_LDL-5	0.883 (0.556-1.401)	0.624	0.999	0.964 (0.723-1.285)	0.802	0.998	1.109 (0.665-1.849)	0.721	0.994	1.29 (0.909-1.831)	0.162	0.528
SubFreeChol_LDL-6	0.741 (0.457-1.201)	0.257	0.999	0.93 (0.708-1.222)	0.605	0.998	1.016 (0.603-1.71)	0.958	0.994	1.575 (1.116-2.221)	0.011	0.209
SubPhosp_LDL-1	1.172 (0.74-1.855)	0.539	0.999	0.99 (0.735-1.333)	0.947	0.998	0.908 (0.55-1.499)	0.732	0.994	0.955 (0.669-1.362)	0.803	0.897
SubPhosp_LDL-2	1.136 (0.722-1.788)	0.612	0.999	0.973 (0.723-1.309)	0.859	0.998	0.786 (0.476-1.299)	0.392	0.994	1.122 (0.8-1.573)	0.513	0.714
SubPhosp_LDL-3	0.99 (0.634-1.545)	0.967	0.999	0.853 (0.638-1.14)	0.282	0.956	1.237 (0.759-2.015)	0.442	0.994	1.294 (0.909-1.843)	0.162	0.528
SubPhosp_LDL-4	0.946 (0.631-1.419)	0.801	0.999	0.883 (0.662-1.177)	0.392	0.956	1.418 (0.94-2.14)	0.142	0.994	1.135 (0.817-1.576)	0.462	0.680
SubPhosp_LDL-5	0.906 (0.576-1.426)	0.694	0.999	1 (0.752-1.33)	0.998	0.998	1.067 (0.649-1.753)	0.817	0.994	1.207 (0.853-1.709)	0.300	0.623
SubPhosp_LDL-6	0.836 (0.503-1.388)	0.531	0.999	0.929 (0.706-1.223)	0.601	0.998	0.956 (0.546-1.672)	0.888	0.994	1.392 (1.003-1.931)	0.056	0.379
SubApoB_LDL-1	1.165 (0.745-1.823)	0.541	0.999	1 (0.745-1.344)	0.998	0.998	0.936 (0.572-1.533)	0.813	0.994	1.006 (0.705-1.435)	0.975	0.984
SubApoB_LDL-2	1.132 (0.721-1.776)	0.623	0.999	0.998 (0.74-1.346)	0.991	0.998	0.77 (0.465-1.275)	0.353	0.994	1.153 (0.822-1.618)	0.419	0.680
SubApoB_LDL-3	1.01 (0.643-1.587)	0.969	0.999	0.879 (0.658-1.175)	0.384	0.956	1.271 (0.774-2.088)	0.394	0.994	1.335 (0.935-1.907)	0.120	0.448
SubApoB_LDL-4	0.957 (0.638-1.436)	0.842	0.999	0.933 (0.704-1.238)	0.632	0.998	1.401 (0.924-2.124)	0.164	0.994	1.109 (0.799-1.541)	0.546	0.731
SubApoB_LDL-5	0.884 (0.557-1.401)	0.627	0.999	1.025 (0.773-1.359)	0.866	0.998	1.081 (0.654-1.784)	0.784	0.994	1.25 (0.885-1.764)	0.216	0.564
SubApoB_LDL-6	0.827 (0.492-1.389)	0.515	0.999	0.978 (0.745-1.285)	0.874	0.998	0.948 (0.533-1.688)	0.874	0.994	1.418 (1.022-1.967)	0.044	0.331
SubTrigl_HDL-1	1.071 (0.694-1.652)	0.779	0.999	1.249 (0.925-1.687)	0.144	0.956	0.808 (0.494-1.321)	0.429	0.994	0.873 (0.612-1.246)	0.465	0.680
SubTrigl_HDL-2	1.008 (0.656-1.548)	0.974	0.999	1.327 (0.979-1.797)	0.066	0.956	0.735 (0.452-1.194)	0.245	0.994	0.861 (0.605-1.226)	0.418	0.680
SubTrigl_HDL-3	1.071 (0.711-1.615)	0.761	0.999	1.343 (0.998-1.806)	0.049	0.956	0.61 (0.372-1.002)	0.045	0.994	0.805 (0.569-1.139)	0.231	0.564
SubTrigl_HDL-4	1.057 (0.673-1.661)	0.826	0.999	1.242 (0.929-1.661)	0.142	0.956	0.609 (0.353-1.051)	0.097	0.994	0.746 (0.508-1.094)	0.144	0.511
SubChol_HDL-1	1.021 (0.636-1.64)	0.937	0.999	0.956 (0.712-1.283)	0.766	0.998	0.845 (0.504-1.418)	0.558	0.994	0.892 (0.617-1.289)	0.552	0.731
SubChol_HDL-2	1.093 (0.698-1.711)	0.721	0.999	1.004 (0.747-1.349)	0.979	0.998	0.781 (0.48-1.269)	0.354	0.994	0.725 (0.502-1.046)	0.092	0.420
SubChol_HDL-3	1.044 (0.675-1.613)	0.860	0.999	0.87 (0.644-1.176)	0.369	0.956	0.79 (0.478-1.305)	0.396	0.994	0.689 (0.472-1.004)	0.056	0.379
SubChol_HDL-4	0.9 (0.585-1.384)	0.656	0.999	0.806 (0.598-1.088)	0.159	0.956	0.942 (0.568-1.563)	0.836	0.994	0.908 (0.635-1.299)	0.610	0.790
SubFreeChol_HDL-1	0.8 (0.496-1.291)	0.401	0.999	0.933 (0.699-1.247)	0.642	0.998	1.019 (0.627-1.656)	0.943	0.994	0.976 (0.672-1.417)	0.902	0.946
SubFreeChol_HDL-2	0.769 (0.478-1.238)	0.317	0.999	0.941 (0.698-1.27)	0.694	0.998	1.248 (0.748-2.082)	0.447	0.994	0.877 (0.6-1.283)	0.512	0.714
SubFreeChol_HDL-3	0.785 (0.488-1.264)	0.351	0.999	0.892 (0.654-1.217)	0.472	0.956	0.979 (0.572-1.676)	0.945	0.994	0.886 (0.602-1.303)	0.548	0.731
SubFreeChol_HDL-4	0.836 (0.529-1.32)	0.469	0.999	0.847 (0.625-1.149)	0.285	0.956	1.138 (0.692-1.871)	0.650	0.994	1.023 (0.709-1.477)	0.904	0.946
SubPhosp_HDL-1	1.053 (0.654-1.697)	0.848	0.999	1.013 (0.754-1.361)	0.932	0.998	0.869 (0.516-1.463)	0.627	0.994	0.791 (0.539-1.16)	0.237	0.564
SubPhosp_HDL-2	1.167 (0.735-1.854)	0.552	0.999	1.043 (0.777-1.4)	0.782	0.998	0.797 (0.483-1.316)	0.412	0.994	0.633 (0.429-0.934)	0.022	0.268
SubPhosp_HDL-3	1.174 (0.755-1.825)	0.514	0.999	0.944 (0.706-1.264)	0.703	0.998	0.763 (0.454-1.28)	0.351	0.994	0.549 (0.365-0.826)	0.004	0.209
SubPhosp_HDL-4	0.964 (0.633-1.469)	0.875	0.999	0.842 (0.625-1.135)	0.259	0.956	0.967 (0.58-1.613)	0.908	0.994	0.775 (0.536-1.121)	0.188	0.564
SubApoA1_HDL-1	1.033 (0.651-1.638)	0.901	0.999	1.116 (0.829-1.503)	0.471	0.956	0.722 (0.423-1.235)	0.267	0.994	0.768 (0.523-1.129)	0.185	0.564
SubApoA1_HDL-2	1.135 (0.725-1.777)	0.618	0.999	1.016 (0.765-1.348)	0.913	0.998	0.765 (0.469-1.249)	0.319	0.994	0.644 (0.438-0.946)	0.026	0.268
SubApoA1_HDL-3	1.094 (0.707-1.692)	0.713	0.999	0.977 (0.73-1.309)	0.880	0.998	0.687 (0.41-1.152)	0.185	0.994	0.637 (0.43-0.944)	0.025	0.268
SubApoA1_HDL-4	0.862 (0.549-1.354)	0.551	0.999	0.891 (0.657-1.208)	0.458	0.956	0.784 (0.468-1.314)	0.404	0.994	0.86 (0.593-1.249)	0.443	0.680
SubApoA2_HDL-1	1.043 (0.662-1.643)	0.871	0.999	1.067 (0.807-1.41)	0.651	0.998	0.956 (0.593-1.541)	0.865	0.994	0.794 (0.55-1.145)	0.223	0.564
SubApoA2_HDL-2	1.126 (0.73-1.735)	0.631	0.999	1.052 (0.799-1.387)	0.717	0.998	0.955 (0.607-1.504)	0.855	0.994	0.829 (0.582-1.181)	0.308	0.623
SubApoA2_HDL-3	1.043 (0.667-1.632)	0.867	0.999	1.007 (0.756-1.34)	0.963	0.998	0.907 (0.556-1.48)	0.721	0.994	0.781 (0.538-1.133)	0.201	0.564
SubApoA2_HDL-4	0.83 (0.53-1.301)	0.449	0.999	0.834 (0.614-1.134)	0.246	0.956	1.047 (0.622-1.762)	0.879	0.994	0.967 (0.673-1.389)	0.860	0.925

Table S6: Effect of 24h post (t_2) rt-PA metabolites and lipids levels on early (*i.e.* sICH, non-response to thrombolysis) and late (*i.e.* three-month mortality and three-month mRS 3-6) adverse outcomes, adjusting for the major determinants for unfavourable outcomes.

	Early outcomes		Late outcomes	
	sICH	Non-response to thrombolysis	Three-month mortality	Three-month mRS 3-6

Molecular features	OR (95% CI)	P	FDR	OR (95% CI)	P	FDR	OR (95% CI)	P	FDR	OR (95% CI)	P	FDR
Creatinine	1.46 (1.003-2.123)	0.062	0.460	0.806 (0.599-1.083)	0.147	0.469	1.346 (0.837-2.164)	0.280	0.630	1.177 (0.83-1.67)	0.379	0.745
Ala	0.806 (0.522-1.243)	0.362	0.811	1.204 (0.909-1.595)	0.196	0.469	1.211 (0.779-1.883)	0.435	0.766	1.061 (0.75-1.502)	0.745	0.818
Glu	0.844 (0.529-1.345)	0.506	0.886	0.791 (0.579-1.081)	0.142	0.469	1.289 (0.773-2.15)	0.372	0.743	1.068 (0.734-1.554)	0.736	0.818
Gln	1.009 (0.652-1.563)	0.970	0.976	1.168 (0.867-1.573)	0.309	0.469	1.09 (0.694-1.713)	0.729	0.782	0.705 (0.482-1.032)	0.074	0.445
Gly	0.956 (0.592-1.545)	0.868	0.976	0.979 (0.735-1.304)	0.886	0.886	1.183 (0.73-1.919)	0.546	0.766	0.936 (0.66-1.327)	0.717	0.818
His	1.072 (0.828-1.388)	0.640	0.886	0.9 (0.728-1.112)	0.313	0.469	1.137 (0.898-1.439)	0.274	0.630	1.091 (0.882-1.349)	0.414	0.745
Ile	0.815 (0.524-1.267)	0.406	0.811	0.914 (0.703-1.189)	0.506	0.569	0.702 (0.408-1.207)	0.257	0.630	0.949 (0.677-1.331)	0.773	0.818
Leu	1.141 (0.765-1.703)	0.559	0.886	0.854 (0.651-1.121)	0.258	0.469	1.183 (0.742-1.886)	0.553	0.766	0.916 (0.643-1.306)	0.644	0.818
Phe	1.389 (0.944-2.042)	0.128	0.460	0.722 (0.542-0.962)	0.024	0.143	1.396 (0.917-2.126)	0.147	0.604	1.243 (0.889-1.739)	0.212	0.637
Tyr	1.009 (0.691-1.473)	0.966	0.976	0.9 (0.687-1.18)	0.446	0.535	1.079 (0.716-1.626)	0.738	0.782	1.319 (0.929-1.873)	0.132	0.476
Val	0.993 (0.66-1.494)	0.976	0.976	0.894 (0.675-1.184)	0.434	0.535	1.384 (0.906-2.112)	0.168	0.604	0.94 (0.675-1.31)	0.724	0.818
Acetic acid	0.877 (0.561-1.372)	0.591	0.886	1.181 (0.883-1.579)	0.262	0.469	0.851 (0.559-1.294)	0.472	0.766	1.651 (1.133-2.407)	0.010	0.173
Citric acid	1.019 (0.688-1.508)	0.931	0.976	0.911 (0.669-1.241)	0.555	0.587	1.408 (0.956-2.073)	0.115	0.604	1.237 (0.858-1.783)	0.265	0.682
Lactic acid	1.245 (0.83-1.869)	0.352	0.811	0.842 (0.642-1.103)	0.213	0.469	1.468 (1.001-2.153)	0.065	0.604	1.302 (0.934-1.814)	0.121	0.476
3-HB	1.382 (0.953-2.004)	0.115	0.460	0.571 (0.413-0.791)	0.000	0.007	1.108 (0.756-1.625)	0.634	0.782	1.434 (1.013-2.032)	0.045	0.403
Acetone	1.223 (0.886-1.688)	0.246	0.738	0.609 (0.43-0.863)	0.003	0.026	1.042 (0.757-1.433)	0.813	0.813	1.177 (0.831-1.666)	0.377	0.745
Pyruvic acid	1.453 (0.962-2.194)	0.105	0.460	0.83 (0.633-1.088)	0.177	0.469	1.426 (0.927-2.196)	0.145	0.604	0.882 (0.626-1.243)	0.486	0.796
Glucose	1.589 (1.002-2.519)	0.071	0.460	0.846 (0.582-1.23)	0.385	0.533	0.867 (0.477-1.574)	0.674	0.782	1.051 (0.677-1.633)	0.829	0.829
Trigl	0.483 (0.241-0.968)	0.044	0.276	1.16 (0.858-1.568)	0.334	0.686	1.3 (0.708-2.385)	0.483	0.987	1.363 (0.924-2.012)	0.133	0.353
Chol	0.813 (0.511-1.294)	0.421	0.819	0.899 (0.672-1.204)	0.477	0.724	1.151 (0.67-1.978)	0.651	0.987	1.376 (0.939-2.018)	0.111	0.327
LDL-Chol	1.025 (0.655-1.605)	0.920	0.973	0.841 (0.629-1.125)	0.243	0.686	1.092 (0.653-1.825)	0.766	0.987	1.461 (0.997-2.14)	0.056	0.280
HDL-Chol	1.075 (0.707-1.636)	0.754	0.961	0.861 (0.644-1.151)	0.313	0.686	1.018 (0.61-1.701)	0.950	0.987	0.837 (0.584-1.199)	0.344	0.560
Apo-A1	0.861 (0.549-1.35)	0.547	0.890	1.035 (0.762-1.405)	0.829	0.976	0.713 (0.424-1.2)	0.248	0.987	0.765 (0.519-1.127)	0.186	0.436
Apo-A2	0.879 (0.555-1.393)	0.616	0.936	0.99 (0.728-1.344)	0.946	0.976	0.927 (0.554-1.551)	0.795	0.987	0.993 (0.68-1.452)	0.973	0.990
Apo-B100	0.897 (0.572-1.407)	0.663	0.946	0.889 (0.67-1.179)	0.415	0.696	1.143 (0.689-1.896)	0.643	0.987	1.595 (1.087-2.339)	0.018	0.175
Apo-B100-Apo-A1	1.054 (0.714-1.556)	0.807	0.973	0.975 (0.74-1.285)	0.857	0.976	1.261 (0.818-1.945)	0.332	0.987	1.738 (1.17-2.583)	0.006	0.076
VLDL_PN	0.636 (0.371-1.092)	0.114	0.511	0.901 (0.682-1.19)	0.462	0.714	1.236 (0.711-2.147)	0.512	0.987	1.198 (0.813-1.766)	0.378	0.580
IDL_PN	0.829 (0.531-1.295)	0.438	0.819	0.889 (0.67-1.179)	0.415	0.696	1.095 (0.684-1.752)	0.731	0.987	1.403 (0.966-2.037)	0.084	0.280
LDL_PN	1.001 (0.646-1.552)	0.997	0.997	1.235 (0.92-1.659)	0.159	0.686	1.13 (0.688-1.858)	0.664	0.987	1.524 (1.044-2.226)	0.031	0.246
LDL1_PN	1.178 (0.77-1.801)	0.489	0.870	1.013 (0.769-1.335)	0.926	0.976	1.005 (0.625-1.616)	0.985	0.987	1.11 (0.765-1.61)	0.593	0.727
LDL2_PN	1.426 (0.883-2.303)	0.176	0.558	0.846 (0.637-1.123)	0.246	0.686	0.751 (0.428-1.317)	0.374	0.987	1.164 (0.799-1.696)	0.441	0.606
LDL3_PN	1.515 (0.937-2.449)	0.123	0.511	0.858 (0.635-1.16)	0.320	0.686	0.971 (0.562-1.679)	0.928	0.987	1.448 (0.993-2.112)	0.061	0.280
LDL4_PN	0.98 (0.647-1.486)	0.930	0.973	0.857 (0.631-1.163)	0.322	0.686	1.189 (0.681-2.075)	0.597	0.987	1.17 (0.817-1.674)	0.410	0.592
LDL5_PN	0.786 (0.491-1.257)	0.347	0.791	0.78 (0.582-1.044)	0.094	0.686	1.035 (0.601-1.783)	0.912	0.987	1.289 (0.882-1.884)	0.203	0.436
LDL6_PN	0.658 (0.379-1.142)	0.161	0.540	0.962 (0.723-1.281)	0.791	0.976	1.797 (1.056-3.055)	0.069	0.982	1.79 (1.233-2.597)	0.002	0.040
LMF_Trigl_VLDL	0.405 (0.19-0.867)	0.020	0.178	0.994 (0.741-1.333)	0.968	0.976	1.604 (0.873-2.948)	0.205	0.987	1.488 (1.004-2.206)	0.055	0.280
LMF_Trigl_IDL	0.487 (0.247-0.962)	0.040	0.271	0.899 (0.676-1.196)	0.463	0.714	1.123 (0.588-2.145)	0.773	0.987	1.198 (0.809-1.776)	0.383	0.580
LMF_Trigl_LDL	1.024 (0.713-1.471)	0.905	0.973	1.163 (0.855-1.582)	0.336	0.686	0.983 (0.668-1.447)	0.935	0.987	1.278 (0.884-1.848)	0.195	0.436
LMF_Trigl_HDL	0.743 (0.477-1.156)	0.223	0.604	1.194 (0.878-1.623)	0.257	0.686	0.526 (0.299-0.927)	0.048	0.982	0.797 (0.553-1.147)	0.235	0.490
LMF_Chol_VLDL	0.452 (0.251-0.812)	0.010	0.128	0.942 (0.721-1.231)	0.665	0.960	1.093 (0.61-1.959)	0.792	0.987	1.161 (0.798-1.688)	0.450	0.611
LMF_Chol_IDL	0.711 (0.441-1.146)	0.187	0.563	1.316 (0.971-1.785)	0.074	0.686	1.085 (0.655-1.796)	0.755	0.987	1.5 (1.034-2.177)	0.037	0.246
LMF_Chol_LDL	1.025 (0.655-1.605)	0.920	0.973	1.247 (0.929-1.674)	0.139	0.686	1.092 (0.653-1.825)	0.766	0.987	1.461 (0.997-2.14)	0.056	0.280

LMF_Chol_HDL	1.075 (0.707-1.636)	0.754	0.961	1.017 (0.768-1.345)	0.909	0.976	1.018 (0.61-1.701)	0.950	0.987	0.837 (0.584-1.199)	0.344	0.560
LMF_FreeChol_VLDL	0.419 (0.222-0.793)	0.009	0.128	0.841 (0.629-1.125)	0.243	0.686	1.394 (0.755-2.575)	0.365	0.987	1.244 (0.844-1.834)	0.286	0.515
LMF_FreeChol_IDL	0.725 (0.45-1.167)	0.212	0.604	0.861 (0.644-1.151)	0.313	0.686	1.014 (0.606-1.695)	0.963	0.987	1.414 (0.974-2.054)	0.076	0.280
LMF_FreeChol_LDL	1.039 (0.657-1.644)	0.882	0.973	1.23 (0.908-1.667)	0.179	0.686	1.159 (0.678-1.98)	0.636	0.987	1.532 (1.043-2.251)	0.033	0.246
LMF_FreeChol_HDL	0.904 (0.59-1.385)	0.670	0.946	1.044 (0.788-1.384)	0.763	0.976	1.18 (0.705-1.975)	0.586	0.987	1.045 (0.729-1.498)	0.818	0.897
LMF_Phosp_VLDL	0.382 (0.194-0.753)	0.006	0.128	0.801 (0.597-1.073)	0.136	0.686	1.575 (0.845-2.934)	0.227	0.987	1.207 (0.809-1.8)	0.373	0.580
LMF_Phosp_IDL	0.647 (0.384-1.09)	0.120	0.511	0.853 (0.641-1.134)	0.272	0.686	1.065 (0.607-1.869)	0.846	0.987	1.336 (0.915-1.951)	0.147	0.372
LMF_Phosp_LDL	1.067 (0.683-1.669)	0.795	0.973	1.293 (0.946-1.767)	0.104	0.686	1.06 (0.635-1.769)	0.843	0.987	1.369 (0.937-1.998)	0.112	0.327
LMF_Phosp_HDL	1.111 (0.708-1.744)	0.673	0.946	1.058 (0.793-1.411)	0.705	0.976	0.747 (0.436-1.279)	0.337	0.987	0.669 (0.458-0.977)	0.041	0.257
LMF_ApoA1_HDL	0.894 (0.575-1.39)	0.644	0.946	0.836 (0.625-1.119)	0.228	0.686	0.675 (0.396-1.152)	0.189	0.987	0.718 (0.489-1.055)	0.098	0.311
LMF_ApoA2_HDL	0.902 (0.566-1.436)	0.691	0.949	0.898 (0.662-1.219)	0.493	0.740	0.952 (0.561-1.616)	0.871	0.987	1 (0.683-1.467)	0.998	0.998
LMF_ApoB_VLDL	0.636 (0.371-1.092)	0.114	0.511	1.021 (0.752-1.385)	0.894	0.976	1.236 (0.712-2.148)	0.511	0.987	1.198 (0.813-1.766)	0.378	0.580
LMF_ApoB_IDL	0.829 (0.53-1.295)	0.438	0.819	0.985 (0.725-1.337)	0.921	0.976	1.095 (0.684-1.752)	0.731	0.987	1.403 (0.966-2.038)	0.083	0.280
LMF_ApoB_LDL	1.001 (0.646-1.552)	0.997	0.997	1.236 (0.92-1.659)	0.159	0.686	1.13 (0.688-1.858)	0.664	0.987	1.524 (1.044-2.226)	0.031	0.246
SubTrigl_VLDL-1	0.458 (0.198-1.057)	0.075	0.426	1.013 (0.769-1.335)	0.926	0.976	1.685 (0.905-3.138)	0.183	0.987	1.533 (1.044-2.251)	0.037	0.246
SubTrigl_VLDL-2	0.372 (0.193-0.717)	0.003	0.097	0.846 (0.637-1.123)	0.246	0.686	1.574 (0.866-2.861)	0.205	0.987	1.275 (0.861-1.888)	0.242	0.490
SubTrigl_VLDL-3	0.44 (0.242-0.803)	0.008	0.128	1.085 (0.797-1.476)	0.604	0.883	1.476 (0.817-2.667)	0.263	0.987	1.315 (0.886-1.953)	0.188	0.436
SubTrigl_VLDL-4	0.656 (0.393-1.096)	0.126	0.511	1.233 (0.908-1.675)	0.179	0.686	1.415 (0.808-2.477)	0.287	0.987	1.239 (0.843-1.821)	0.289	0.515
SubTrigl_VLDL-5	0.917 (0.594-1.416)	0.720	0.961	1.271 (0.939-1.72)	0.119	0.686	1.341 (0.776-2.319)	0.361	0.987	0.833 (0.584-1.189)	0.331	0.559
SubChol_VLDL-1	0.395 (0.179-0.869)	0.024	0.194	1.213 (0.906-1.622)	0.194	0.686	1.244 (0.655-2.365)	0.583	0.987	1.364 (0.932-1.996)	0.128	0.353
SubChol_VLDL-2	0.383 (0.203-0.719)	0.003	0.097	1.121 (0.845-1.488)	0.429	0.696	1.044 (0.577-1.889)	0.900	0.987	1.106 (0.764-1.6)	0.607	0.729
SubChol_VLDL-3	0.465 (0.262-0.825)	0.010	0.128	1.137 (0.839-1.54)	0.408	0.696	1.151 (0.657-2.016)	0.662	0.987	1.341 (0.925-1.945)	0.133	0.353
SubChol_VLDL-4	0.575 (0.349-0.947)	0.038	0.270	1.252 (0.936-1.674)	0.127	0.686	1.014 (0.604-1.7)	0.963	0.987	1.075 (0.75-1.542)	0.704	0.810
SubChol_VLDL-5	1.005 (0.649-1.556)	0.984	0.997	1.188 (0.89-1.586)	0.242	0.686	0.771 (0.444-1.342)	0.417	0.987	0.497 (0.334-0.738)	0.000	0.040
SubFreeChol_VLDL-1	0.424 (0.195-0.923)	0.031	0.237	1.214 (0.915-1.612)	0.178	0.686	1.131 (0.584-2.191)	0.763	0.987	1.305 (0.886-1.92)	0.198	0.436
SubFreeChol_VLDL-2	0.46 (0.249-0.852)	0.014	0.164	1.226 (0.92-1.633)	0.163	0.686	1.283 (0.718-2.293)	0.471	0.987	1.326 (0.906-1.942)	0.159	0.394
SubFreeChol_VLDL-3	0.468 (0.253-0.868)	0.018	0.171	1.173 (0.869-1.584)	0.296	0.686	1.31 (0.735-2.336)	0.432	0.987	1.441 (0.984-2.11)	0.068	0.280
SubFreeChol_VLDL-4	0.681 (0.425-1.09)	0.130	0.511	1.239 (0.922-1.665)	0.153	0.686	1.071 (0.648-1.77)	0.808	0.987	1.202 (0.837-1.726)	0.334	0.559
SubFreeChol_VLDL-5	0.732 (0.411-1.304)	0.336	0.791	1.183 (0.882-1.586)	0.262	0.686	1.362 (0.743-2.496)	0.393	0.987	0.975 (0.645-1.473)	0.907	0.949
SubPhosp_VLDL-1	0.361 (0.157-0.832)	0.017	0.171	1.145 (0.864-1.517)	0.346	0.686	1.389 (0.735-2.624)	0.405	0.987	1.457 (0.984-2.157)	0.071	0.280
SubPhosp_VLDL-2	0.364 (0.193-0.687)	0.001	0.097	1.157 (0.849-1.577)	0.354	0.686	1.346 (0.74-2.446)	0.406	0.987	1.251 (0.853-1.834)	0.268	0.501
SubPhosp_VLDL-3	0.46 (0.257-0.822)	0.010	0.128	1.14 (0.838-1.551)	0.406	0.696	1.292 (0.729-2.289)	0.447	0.987	1.31 (0.891-1.925)	0.181	0.436
SubPhosp_VLDL-4	0.646 (0.396-1.052)	0.096	0.511	1.275 (0.942-1.726)	0.113	0.686	1.132 (0.667-1.922)	0.682	0.987	1.088 (0.752-1.575)	0.664	0.780
SubPhosp_VLDL-5	0.812 (0.525-1.255)	0.389	0.819	1.292 (0.96-1.739)	0.089	0.686	0.841 (0.493-1.435)	0.577	0.987	0.585 (0.401-0.852)	0.006	0.076
SubTrigl_LDL-1	1.082 (0.765-1.53)	0.680	0.946	1.224 (0.921-1.627)	0.164	0.686	0.97 (0.652-1.443)	0.887	0.987	1.131 (0.787-1.625)	0.518	0.667
SubTrigl_LDL-2	1.155 (0.825-1.617)	0.437	0.819	1.323 (0.992-1.766)	0.055	0.686	1.026 (0.711-1.48)	0.897	0.987	1.099 (0.772-1.564)	0.614	0.729
SubTrigl_LDL-3	1.166 (0.818-1.661)	0.437	0.819	0.975 (0.742-1.281)	0.855	0.976	1.063 (0.727-1.553)	0.770	0.987	1.057 (0.755-1.479)	0.754	0.857
SubTrigl_LDL-4	1.01 (0.695-1.469)	0.960	0.995	0.953 (0.723-1.257)	0.734	0.976	1.016 (0.681-1.517)	0.941	0.987	1.055 (0.744-1.495)	0.773	0.864
SubTrigl_LDL-5	0.803 (0.515-1.25)	0.358	0.801	0.883 (0.669-1.166)	0.379	0.696	1.041 (0.665-1.629)	0.873	0.987	1.159 (0.819-1.642)	0.420	0.598
SubTrigl_LDL-6	0.706 (0.405-1.231)	0.249	0.661	0.992 (0.756-1.302)	0.954	0.976	1.751 (1.182-2.594)	0.010	0.982	1.745 (1.213-2.508)	0.001	0.040
SubChol_LDL-1	1.162 (0.728-1.856)	0.565	0.895	1.024 (0.782-1.342)	0.864	0.976	0.936 (0.551-1.592)	0.827	0.987	1.043 (0.715-1.52)	0.834	0.905
SubChol_LDL-2	1.405 (0.874-2.259)	0.192	0.563	0.782 (0.582-1.051)	0.092	0.686	0.802 (0.457-1.406)	0.498	0.987	1.169 (0.806-1.695)	0.425	0.598
SubChol_LDL-3	1.451 (0.909-2.317)	0.154	0.540	0.879 (0.648-1.194)	0.413	0.696	1.036 (0.601-1.786)	0.911	0.987	1.406 (0.968-2.04)	0.081	0.280
SubChol_LDL-4	1.004 (0.665-1.514)	0.987	0.997	0.845 (0.624-1.144)	0.277	0.686	1.188 (0.682-2.069)	0.596	0.987	1.146 (0.799-1.645)	0.475	0.630
SubChol_LDL-5	0.811 (0.504-1.306)	0.420	0.819	0.803 (0.602-1.07)	0.133	0.686	1.071 (0.613-1.872)	0.830	0.987	1.258 (0.854-1.851)	0.260	0.493
SubChol_LDL-6	0.691 (0.41-1.165)	0.190	0.563	0.959 (0.718-1.28)	0.775	0.976	1.804 (1.084-3.003)	0.042	0.982	1.743 (1.207-2.516)	0.003	0.049

SubFreeChol_LDL-1	1.082 (0.682-1.718)	0.759	0.961	1.006 (0.746-1.358)	0.967	0.976	0.92 (0.546-1.552)	0.777	0.987	1.127 (0.773-1.643)	0.546	0.691
SubFreeChol_LDL-2	1.213 (0.778-1.894)	0.428	0.819	0.888 (0.667-1.184)	0.418	0.696	0.873 (0.508-1.499)	0.662	0.987	1.253 (0.864-1.818)	0.246	0.490
SubFreeChol_LDL-3	1.538 (0.939-2.52)	0.120	0.511	0.874 (0.644-1.185)	0.388	0.696	1.077 (0.622-1.865)	0.817	0.987	1.427 (0.975-2.086)	0.073	0.280
SubFreeChol_LDL-4	1.114 (0.723-1.717)	0.652	0.946	0.852 (0.63-1.153)	0.301	0.686	1.099 (0.633-1.909)	0.769	0.987	1.337 (0.925-1.931)	0.134	0.353
SubFreeChol_LDL-5	0.879 (0.553-1.399)	0.615	0.936	0.755 (0.562-1.014)	0.060	0.686	1.075 (0.628-1.839)	0.814	0.987	1.444 (0.978-2.13)	0.071	0.280
SubFreeChol_LDL-6	0.767 (0.487-1.207)	0.281	0.728	0.855 (0.639-1.144)	0.291	0.686	1.267 (0.771-2.081)	0.400	0.987	1.844 (1.266-2.687)	0.001	0.040
SubPhosp_LDL-1	1.181 (0.756-1.843)	0.504	0.870	0.906 (0.673-1.221)	0.519	0.768	0.935 (0.562-1.557)	0.814	0.987	1.007 (0.693-1.465)	0.971	0.990
SubPhosp_LDL-2	1.483 (0.903-2.434)	0.147	0.540	0.84 (0.631-1.118)	0.230	0.686	0.781 (0.439-1.389)	0.458	0.987	1.121 (0.768-1.636)	0.567	0.710
SubPhosp_LDL-3	1.492 (0.919-2.422)	0.140	0.533	0.885 (0.652-1.2)	0.434	0.696	0.962 (0.553-1.672)	0.904	0.987	1.342 (0.921-1.954)	0.136	0.353
SubPhosp_LDL-4	1.025 (0.682-1.541)	0.912	0.973	0.841 (0.618-1.145)	0.272	0.686	1.171 (0.675-2.033)	0.624	0.987	1.138 (0.792-1.635)	0.501	0.657
SubPhosp_LDL-5	0.817 (0.511-1.306)	0.430	0.819	0.796 (0.595-1.065)	0.123	0.686	1.009 (0.579-1.759)	0.978	0.987	1.19 (0.813-1.742)	0.387	0.580
SubPhosp_LDL-6	0.712 (0.429-1.184)	0.218	0.604	0.959 (0.718-1.281)	0.778	0.976	1.764 (1.056-2.947)	0.063	0.982	1.701 (1.184-2.445)	0.004	0.061
SubApoB_LDL-1	1.178 (0.77-1.801)	0.489	0.870	1.015 (0.754-1.366)	0.921	0.976	1.005 (0.625-1.616)	0.985	0.987	1.11 (0.765-1.61)	0.593	0.727
SubApoB_LDL-2	1.425 (0.882-2.302)	0.176	0.558	0.875 (0.657-1.164)	0.356	0.686	0.751 (0.428-1.316)	0.374	0.987	1.164 (0.799-1.695)	0.441	0.606
SubApoB_LDL-3	1.515 (0.937-2.449)	0.123	0.511	0.858 (0.635-1.16)	0.320	0.686	0.972 (0.562-1.68)	0.928	0.987	1.448 (0.992-2.112)	0.061	0.280
SubApoB_LDL-4	0.98 (0.647-1.486)	0.930	0.973	0.857 (0.632-1.163)	0.323	0.686	1.189 (0.681-2.076)	0.597	0.987	1.17 (0.817-1.674)	0.410	0.592
SubApoB_LDL-5	0.786 (0.491-1.258)	0.347	0.791	0.78 (0.583-1.044)	0.094	0.686	1.036 (0.601-1.784)	0.911	0.987	1.289 (0.882-1.884)	0.203	0.436
SubApoB_LDL-6	0.658 (0.379-1.142)	0.161	0.540	0.962 (0.722-1.281)	0.791	0.976	1.797 (1.057-3.056)	0.069	0.982	1.79 (1.233-2.597)	0.002	0.040
SubTrigl_HDL-1	0.839 (0.526-1.338)	0.497	0.870	0.994 (0.741-1.333)	0.967	0.976	0.692 (0.39-1.227)	0.255	0.987	0.948 (0.649-1.383)	0.787	0.871
SubTrigl_HDL-2	0.83 (0.542-1.271)	0.426	0.819	0.899 (0.676-1.196)	0.463	0.714	0.601 (0.347-1.041)	0.099	0.987	0.909 (0.635-1.302)	0.613	0.729
SubTrigl_HDL-3	0.784 (0.512-1.202)	0.299	0.757	1.161 (0.85-1.586)	0.348	0.686	0.543 (0.309-0.952)	0.052	0.982	0.868 (0.606-1.244)	0.456	0.611
SubTrigl_HDL-4	0.619 (0.385-0.994)	0.060	0.363	1.193 (0.884-1.61)	0.249	0.686	0.527 (0.296-0.936)	0.057	0.982	0.681 (0.455-1.018)	0.066	0.280
SubChol_HDL-1	1.022 (0.655-1.594)	0.930	0.973	1.325 (0.981-1.79)	0.064	0.686	1.213 (0.704-2.089)	0.535	0.987	1.08 (0.743-1.569)	0.696	0.810
SubChol_HDL-2	1.276 (0.826-1.974)	0.307	0.761	1.559 (1.135-2.143)	0.005	0.586	0.974 (0.575-1.65)	0.929	0.987	0.806 (0.561-1.16)	0.257	0.493
SubChol_HDL-3	1.167 (0.761-1.788)	0.516	0.878	0.853 (0.633-1.149)	0.297	0.686	0.917 (0.558-1.506)	0.759	0.987	0.889 (0.628-1.26)	0.520	0.667
SubChol_HDL-4	0.972 (0.622-1.52)	0.910	0.973	0.814 (0.6-1.105)	0.186	0.686	1.036 (0.608-1.765)	0.907	0.987	0.836 (0.567-1.233)	0.379	0.580
SubFreeChol_HDL-1	0.962 (0.62-1.494)	0.875	0.973	0.865 (0.646-1.159)	0.334	0.686	1.148 (0.686-1.923)	0.640	0.987	1.256 (0.873-1.808)	0.238	0.490
SubFreeChol_HDL-2	1.092 (0.69-1.728)	0.735	0.961	0.978 (0.719-1.33)	0.886	0.976	1.101 (0.631-1.919)	0.768	0.987	1.247 (0.866-1.794)	0.249	0.490
SubFreeChol_HDL-3	0.852 (0.535-1.358)	0.539	0.890	0.817 (0.612-1.092)	0.171	0.686	0.811 (0.473-1.388)	0.494	0.987	1.014 (0.693-1.483)	0.946	0.980
SubFreeChol_HDL-4	0.921 (0.582-1.457)	0.745	0.961	0.741 (0.546-1.005)	0.051	0.686	0.985 (0.563-1.722)	0.962	0.987	1.027 (0.701-1.505)	0.894	0.944
SubPhosp_HDL-1	1.045 (0.656-1.663)	0.865	0.973	0.942 (0.697-1.273)	0.699	0.976	0.992 (0.565-1.741)	0.981	0.987	0.94 (0.638-1.383)	0.760	0.857
SubPhosp_HDL-2	1.283 (0.815-2.02)	0.318	0.771	0.947 (0.699-1.282)	0.723	0.976	0.804 (0.462-1.4)	0.487	0.987	0.718 (0.492-1.046)	0.090	0.294
SubPhosp_HDL-3	1.153 (0.741-1.795)	0.561	0.895	0.883 (0.65-1.2)	0.429	0.696	0.689 (0.412-1.153)	0.203	0.987	0.701 (0.486-1.011)	0.062	0.280
SubPhosp_HDL-4	0.954 (0.613-1.483)	0.845	0.973	0.864 (0.632-1.181)	0.361	0.686	0.831 (0.484-1.427)	0.547	0.987	0.693 (0.465-1.033)	0.076	0.280
SubApoA1_HDL-1	0.959 (0.605-1.52)	0.869	0.973	0.949 (0.705-1.277)	0.729	0.976	0.731 (0.413-1.292)	0.327	0.987	0.808 (0.544-1.199)	0.301	0.527
SubApoA1_HDL-2	1.041 (0.665-1.628)	0.872	0.973	1.044 (0.768-1.419)	0.786	0.976	0.747 (0.45-1.241)	0.311	0.987	0.726 (0.499-1.055)	0.101	0.311
SubApoA1_HDL-3	1.033 (0.66-1.618)	0.895	0.973	0.977 (0.721-1.323)	0.881	0.976	0.817 (0.475-1.404)	0.515	0.987	0.813 (0.564-1.17)	0.278	0.511
SubApoA1_HDL-4	0.825 (0.52-1.308)	0.448	0.824	1.003 (0.746-1.349)	0.984	0.984	0.887 (0.515-1.528)	0.699	0.987	0.808 (0.539-1.21)	0.313	0.541
SubApoA2_HDL-1	0.928 (0.577-1.493)	0.779	0.973	0.967 (0.717-1.303)	0.826	0.976	0.904 (0.513-1.591)	0.755	0.987	0.999 (0.685-1.456)	0.994	0.998
SubApoA2_HDL-2	0.964 (0.599-1.55)	0.891	0.973	1.063 (0.775-1.457)	0.707	0.976	1.005 (0.57-1.774)	0.987	0.987	1.176 (0.815-1.698)	0.404	0.592
SubApoA2_HDL-3	0.909 (0.566-1.459)	0.720	0.961	0.979 (0.729-1.315)	0.889	0.976	0.818 (0.48-1.392)	0.509	0.987	1.033 (0.712-1.498)	0.870	0.926
SubApoA2_HDL-4	0.869 (0.549-1.376)	0.581	0.907	0.988 (0.737-1.325)	0.937	0.976	1.047 (0.612-1.793)	0.881	0.987	0.963 (0.651-1.424)	0.853	0.917

Table S7: Effect of pre-post-rtPA variations of metabolites and lipids levels on early (*i.e.* sICH, non-response to thrombolysis) and late (*i.e.* three-month mortality and three-month mRS 3-6) adverse outcomes, adjusting for the major determinants for unfavourable outcomes.

	Early outcomes						Late outcomes					
	sICH			Non-response to thrombolysis			Three-month mortality			Three-month mRS 3-6		
Molecular features	**OR (95% CI)**	**P**	**FDR**	**OR (95% CI)**	**P**	**FDR**	**OR (95% CI)**	**P**	**FDR**	**OR (95% CI)**	**P**	**FDR**
Creatinine	1.285 (0.887-1.861)	0.229	0.796	0.755 (0.568-1.003)	0.047	0.238	0.691 (0.429-1.114)	0.166	0.819	1.185 (0.837-1.677)	0.365	0.766
Ala	0.919 (0.603-1.401)	0.719	0.846	1.327 (0.994-1.771)	0.053	0.238	1.031 (0.667-1.595)	0.898	0.957	0.909 (0.645-1.282)	0.596	0.766
Glu	0.925 (0.593-1.442)	0.752	0.846	0.822 (0.605-1.116)	0.208	0.558	1.042 (0.654-1.66)	0.876	0.957	1.565 (1.069-2.29)	0.023	0.412
Gln	0.889 (0.596-1.325)	0.619	0.796	1.285 (0.922-1.791)	0.132	0.474	1.083 (0.763-1.535)	0.680	0.957	0.864 (0.612-1.221)	0.413	0.766
Gly	1.245 (0.814-1.904)	0.351	0.796	1.04 (0.789-1.371)	0.783	0.888	0.786 (0.489-1.263)	0.362	0.930	0.903 (0.639-1.276)	0.576	0.766
His	0.996 (0.666-1.488)	0.984	0.984	0.938 (0.711-1.237)	0.652	0.888	1.165 (0.784-1.731)	0.481	0.957	1.127 (0.804-1.579)	0.508	0.766
Ile	0.803 (0.549-1.174)	0.292	0.796	0.986 (0.756-1.287)	0.919	0.919	0.678 (0.448-1.024)	0.084	0.819	0.911 (0.657-1.262)	0.585	0.766
Leu	1.176 (0.791-1.749)	0.457	0.796	0.946 (0.727-1.232)	0.683	0.888	0.832 (0.55-1.257)	0.422	0.949	0.833 (0.596-1.165)	0.302	0.766
Phe	1.736 (1.157-2.605)	0.015	0.274	0.86 (0.654-1.131)	0.280	0.558	1.133 (0.735-1.746)	0.610	0.957	1.023 (0.73-1.433)	0.898	0.898
Tyr	1.125 (0.759-1.668)	0.596	0.796	0.972 (0.742-1.274)	0.839	0.888	0.937 (0.63-1.392)	0.764	0.957	0.948 (0.69-1.301)	0.746	0.790
Val	1.119 (0.749-1.671)	0.614	0.796	0.92 (0.703-1.204)	0.543	0.888	0.969 (0.637-1.474)	0.893	0.957	0.894 (0.649-1.231)	0.503	0.766
Acetic acid	0.993 (0.611-1.615)	0.980	0.984	0.962 (0.719-1.286)	0.793	0.888	0.965 (0.575-1.62)	0.904	0.957	1.45 (0.972-2.165)	0.073	0.527
Citric acid	1.239 (0.808-1.901)	0.362	0.796	0.832 (0.614-1.13)	0.239	0.558	1.265 (0.829-1.931)	0.302	0.907	1.399 (0.958-2.044)	0.088	0.527
Lactic acid	1.55 (0.969-2.479)	0.093	0.417	0.853 (0.642-1.134)	0.275	0.558	1.461 (0.886-2.41)	0.182	0.819	1.272 (0.895-1.808)	0.192	0.766
3-HB	1.214 (0.783-1.883)	0.423	0.796	0.725 (0.543-0.969)	0.028	0.238	1.394 (0.823-2.361)	0.259	0.907	1.238 (0.867-1.768)	0.250	0.766
Acetone	1.162 (0.753-1.794)	0.531	0.796	0.684 (0.516-0.906)	0.007	0.129	1.073 (0.668-1.724)	0.790	0.957	1.226 (0.865-1.738)	0.261	0.766
Pyruvic acid	1.692 (1.062-2.695)	0.035	0.297	0.86 (0.643-1.15)	0.310	0.558	1.512 (0.943-2.425)	0.107	0.819	0.927 (0.646-1.33)	0.689	0.775
Glucose	1.68 (1.042-2.707)	0.050	0.297	0.955 (0.717-1.272)	0.757	0.888	1.006 (0.618-1.636)	0.984	0.984	0.926 (0.671-1.279)	0.649	0.775
Trigl	0.651 (0.431-0.985)	0.057	0.339	0.961 (0.72-1.283)	0.788	0.990	1.179 (0.741-1.877)	0.513	0.976	1.148 (0.791-1.665)	0.479	0.973
Chol	0.91 (0.599-1.382)	0.681	0.955	0.948 (0.718-1.253)	0.711	0.990	1.15 (0.761-1.737)	0.543	0.976	1.172 (0.827-1.662)	0.387	0.973
LDL-Chol	0.871 (0.586-1.295)	0.528	0.955	0.976 (0.735-1.297)	0.868	0.990	0.964 (0.648-1.433)	0.865	0.976	0.906 (0.63-1.305)	0.610	0.973
HDL-Chol	1.121 (0.738-1.703)	0.621	0.955	0.938 (0.708-1.244)	0.659	0.990	1.242 (0.776-1.99)	0.411	0.976	1.209 (0.841-1.739)	0.320	0.973
Apo-A1	0.869 (0.575-1.313)	0.538	0.955	0.997 (0.756-1.315)	0.981	0.990	1.002 (0.659-1.525)	0.992	0.992	1.221 (0.85-1.755)	0.295	0.973
Apo-A2	1.044 (0.695-1.568)	0.851	0.980	1.037 (0.79-1.362)	0.793	0.990	0.926 (0.613-1.397)	0.734	0.976	1.244 (0.882-1.756)	0.225	0.973
Apo-B100	0.909 (0.606-1.363)	0.669	0.955	0.785 (0.592-1.041)	0.090	0.990	1.08 (0.715-1.632)	0.734	0.976	1.104 (0.778-1.565)	0.593	0.973
Apo-B100-Apo-A1	1.015 (0.656-1.571)	0.952	0.997	1.016 (0.767-1.346)	0.914	0.990	1.088 (0.697-1.7)	0.737	0.976	0.933 (0.652-1.335)	0.720	0.979
VLDL_PN	0.597 (0.39-0.914)	0.024	0.301	0.776 (0.58-1.038)	0.081	0.990	1.095 (0.687-1.747)	0.723	0.976	1.044 (0.724-1.505)	0.822	0.979

IDL_PN	0.87 (0.591-1.283)	0.517	0.955	0.785 (0.592-1.041)	0.090	0.990	0.917 (0.603-1.395)	0.710	0.976	1.268 (0.904-1.779)	0.177	0.973
LDL_PN	1.054 (0.709-1.567)	0.811	0.955	0.943 (0.704-1.265)	0.698	0.990	0.918 (0.601-1.402)	0.714	0.976	1.025 (0.709-1.483)	0.898	0.985
LDL1_PN	1.092 (0.728-1.638)	0.698	0.955	0.842 (0.642-1.106)	0.216	0.990	0.864 (0.525-1.421)	0.612	0.976	1.161 (0.819-1.646)	0.420	0.973
LDL2_PN	1.023 (0.687-1.525)	0.917	0.987	0.839 (0.632-1.114)	0.225	0.990	0.764 (0.468-1.249)	0.334	0.976	0.956 (0.68-1.343)	0.800	0.979
LDL3_PN	1.237 (0.8-1.913)	0.388	0.955	0.875 (0.668-1.146)	0.330	0.990	0.725 (0.458-1.149)	0.193	0.976	1.034 (0.712-1.503)	0.866	0.979
LDL4_PN	0.769 (0.506-1.171)	0.257	0.826	0.873 (0.646-1.18)	0.376	0.990	1.192 (0.744-1.91)	0.506	0.976	1.093 (0.764-1.563)	0.637	0.973
LDL5_PN	0.827 (0.564-1.214)	0.370	0.955	0.765 (0.567-1.032)	0.076	0.990	0.752 (0.51-1.108)	0.170	0.976	0.881 (0.637-1.219)	0.454	0.973
LDL6_PN	0.97 (0.659-1.428)	0.887	0.982	1.01 (0.766-1.331)	0.944	0.990	1.31 (0.83-2.069)	0.275	0.976	1.197 (0.846-1.694)	0.332	0.973
LMF_Trigl_VLDL	0.639 (0.418-0.977)	0.051	0.339	0.969 (0.733-1.28)	0.824	0.990	1.21 (0.764-1.918)	0.443	0.976	1.188 (0.821-1.721)	0.373	0.973
LMF_Trigl_IDL	0.609 (0.383-0.97)	0.047	0.339	0.87 (0.652-1.162)	0.346	0.990	0.951 (0.615-1.471)	0.834	0.976	1.267 (0.887-1.808)	0.203	0.973
LMF_Trigl_LDL	0.944 (0.656-1.358)	0.775	0.955	0.976 (0.729-1.306)	0.870	0.990	0.609 (0.414-0.897)	0.023	0.934	1.035 (0.734-1.458)	0.854	0.979
LMF_Trigl_HDL	0.803 (0.574-1.124)	0.245	0.826	0.972 (0.727-1.3)	0.851	0.990	0.597 (0.404-0.882)	0.031	0.934	0.885 (0.664-1.18)	0.417	0.973
LMF_Chol_VLDL	0.611 (0.405-0.921)	0.026	0.301	0.808 (0.611-1.067)	0.123	0.990	1.066 (0.685-1.66)	0.790	0.976	1.119 (0.783-1.599)	0.548	0.973
LMF_Chol_IDL	0.893 (0.607-1.313)	0.596	0.955	0.969 (0.75-1.252)	0.811	0.990	1.049 (0.701-1.569)	0.827	0.976	1.337 (0.96-1.863)	0.092	0.973
LMF_Chol_LDL	0.871 (0.586-1.295)	0.528	0.955	1.032 (0.777-1.372)	0.827	0.990	0.964 (0.648-1.433)	0.865	0.976	0.906 (0.63-1.305)	0.610	0.973
LMF_Chol_HDL	1.121 (0.738-1.703)	0.621	0.955	0.875 (0.665-1.152)	0.342	0.990	1.242 (0.776-1.99)	0.411	0.976	1.209 (0.841-1.739)	0.320	0.973
LMF_FreeChol_VLDL	0.602 (0.399-0.91)	0.023	0.301	0.976 (0.735-1.297)	0.868	0.990	1.221 (0.781-1.909)	0.404	0.976	1.141 (0.791-1.646)	0.490	0.973
LMF_FreeChol_IDL	0.873 (0.598-1.274)	0.520	0.955	0.938 (0.708-1.244)	0.659	0.990	0.976 (0.659-1.446)	0.910	0.985	1.277 (0.921-1.772)	0.150	0.973
LMF_FreeChol_LDL	1.29 (0.843-1.971)	0.283	0.872	1.01 (0.755-1.35)	0.947	0.990	0.991 (0.649-1.515)	0.970	0.992	1.106 (0.764-1.601)	0.607	0.973
LMF_FreeChol_HDL	1.352 (0.863-2.12)	0.226	0.822	0.905 (0.688-1.191)	0.479	0.990	1.606 (0.977-2.639)	0.088	0.976	1.308 (0.893-1.914)	0.179	0.973
LMF_Phosp_VLDL	0.525 (0.339-0.814)	0.006	0.301	0.858 (0.639-1.151)	0.307	0.990	1.159 (0.735-1.827)	0.548	0.976	1.126 (0.778-1.628)	0.540	0.973
LMF_Phosp_IDL	0.799 (0.528-1.21)	0.331	0.920	0.918 (0.69-1.223)	0.561	0.990	0.898 (0.582-1.386)	0.660	0.976	1.05 (0.75-1.47)	0.783	0.979
LMF_Phosp_LDL	1.068 (0.714-1.598)	0.767	0.955	1.11 (0.825-1.493)	0.491	0.990	0.907 (0.597-1.377)	0.668	0.976	0.965 (0.668-1.394)	0.855	0.979
LMF_Phosp_HDL	1.023 (0.662-1.581)	0.926	0.987	1.019 (0.779-1.333)	0.894	0.990	0.932 (0.581-1.494)	0.790	0.976	1.216 (0.849-1.742)	0.302	0.973
LMF_ApoA1_HDL	0.851 (0.575-1.261)	0.452	0.955	0.918 (0.691-1.218)	0.553	0.990	1.037 (0.691-1.558)	0.871	0.976	1.171 (0.834-1.645)	0.370	0.973
LMF_ApoA2_HDL	1.058 (0.703-1.592)	0.807	0.955	0.947 (0.713-1.258)	0.709	0.990	0.938 (0.617-1.427)	0.783	0.976	1.249 (0.883-1.766)	0.222	0.973
LMF_ApoB_VLDL	0.597 (0.39-0.914)	0.024	0.301	0.905 (0.688-1.191)	0.478	0.990	1.095 (0.687-1.747)	0.723	0.976	1.044 (0.724-1.505)	0.822	0.979
LMF_ApoB_IDL	0.87 (0.591-1.282)	0.516	0.955	1.03 (0.784-1.354)	0.831	0.990	0.917 (0.603-1.395)	0.708	0.976	1.269 (0.904-1.78)	0.177	0.973
LMF_ApoB_LDL	1.054 (0.709-1.566)	0.811	0.955	0.944 (0.704-1.265)	0.698	0.990	0.918 (0.601-1.402)	0.714	0.976	1.025 (0.709-1.483)	0.898	0.985

SubTrigl_VLDL-1	0.727 (0.468-1.129)	0.186	0.772	0.842 (0.642-1.106)	0.216	0.990	1.113 (0.716-1.729)	0.657	0.976	1.045 (0.733-1.49)	0.815	0.979
SubTrigl_VLDL-2	0.608 (0.392-0.945)	0.041	0.339	0.839 (0.632-1.114)	0.224	0.990	1.064 (0.659-1.718)	0.816	0.976	0.986 (0.68-1.431)	0.944	0.985
SubTrigl_VLDL-3	0.643 (0.406-1.02)	0.082	0.435	1.012 (0.762-1.344)	0.935	0.990	1.114 (0.677-1.834)	0.705	0.976	1.042 (0.719-1.51)	0.836	0.979
SubTrigl_VLDL-4	0.602 (0.372-0.973)	0.047	0.339	1.236 (0.928-1.648)	0.146	0.990	1.159 (0.735-1.828)	0.571	0.976	1.06 (0.755-1.488)	0.741	0.979
SubTrigl_VLDL-5	0.734 (0.478-1.126)	0.190	0.772	1.194 (0.902-1.579)	0.214	0.990	1.359 (0.922-2.002)	0.165	0.976	0.952 (0.691-1.312)	0.768	0.979
SubChol_VLDL-1	0.704 (0.464-1.068)	0.121	0.573	0.921 (0.691-1.228)	0.575	0.990	1.101 (0.712-1.704)	0.684	0.976	1.186 (0.824-1.706)	0.368	0.973
SubChol_VLDL-2	0.554 (0.366-0.838)	0.009	0.301	0.914 (0.692-1.207)	0.524	0.990	1.018 (0.626-1.654)	0.949	0.985	1.075 (0.746-1.55)	0.708	0.979
SubChol_VLDL-3	0.631 (0.406-0.981)	0.055	0.339	0.989 (0.745-1.314)	0.942	0.990	1.119 (0.707-1.771)	0.657	0.976	1.224 (0.857-1.747)	0.278	0.973
SubChol_VLDL-4	0.667 (0.453-0.982)	0.056	0.339	1.119 (0.849-1.475)	0.425	0.990	0.966 (0.647-1.441)	0.873	0.976	1.085 (0.779-1.511)	0.636	0.973
SubChol_VLDL-5	0.841 (0.556-1.271)	0.448	0.955	1.051 (0.793-1.392)	0.733	0.990	0.775 (0.483-1.243)	0.327	0.976	0.885 (0.618-1.266)	0.518	0.973
SubFreeChol_VLDL-1	0.57 (0.363-0.896)	0.020	0.301	1.006 (0.76-1.331)	0.965	0.990	1.025 (0.647-1.623)	0.924	0.985	1.02 (0.72-1.445)	0.912	0.985
SubFreeChol_VLDL-2	0.676 (0.443-1.031)	0.090	0.446	0.972 (0.735-1.285)	0.842	0.990	1.08 (0.683-1.71)	0.764	0.976	1.1 (0.773-1.564)	0.607	0.973
SubFreeChol_VLDL-3	0.661 (0.425-1.029)	0.084	0.435	1.006 (0.754-1.343)	0.969	0.990	1.156 (0.715-1.87)	0.582	0.976	1.277 (0.882-1.85)	0.208	0.973
SubFreeChol_VLDL-4	0.777 (0.538-1.124)	0.218	0.822	1.069 (0.814-1.404)	0.631	0.990	1.054 (0.711-1.563)	0.804	0.976	1.236 (0.89-1.717)	0.212	0.973
SubFreeChol_VLDL-5	0.719 (0.471-1.097)	0.157	0.714	0.986 (0.742-1.311)	0.925	0.990	1.323 (0.861-2.034)	0.242	0.976	0.897 (0.641-1.254)	0.533	0.973
SubPhosp_VLDL-1	0.63 (0.401-0.989)	0.061	0.345	0.898 (0.677-1.19)	0.455	0.990	1.131 (0.704-1.816)	0.637	0.976	1.13 (0.781-1.636)	0.530	0.973
SubPhosp_VLDL-2	0.501 (0.323-0.777)	0.004	0.301	1.008 (0.765-1.328)	0.957	0.990	1.125 (0.664-1.906)	0.691	0.976	1.068 (0.731-1.561)	0.740	0.979
SubPhosp_VLDL-3	0.534 (0.332-0.86)	0.014	0.301	1.027 (0.768-1.372)	0.858	0.990	1.059 (0.645-1.738)	0.838	0.976	1.149 (0.798-1.653)	0.471	0.973
SubPhosp_VLDL-4	0.627 (0.412-0.952)	0.039	0.339	1.179 (0.878-1.584)	0.272	0.990	0.961 (0.623-1.484)	0.869	0.976	1.043 (0.737-1.475)	0.817	0.979
SubPhosp_VLDL-5	0.74 (0.502-1.091)	0.165	0.723	1.058 (0.797-1.405)	0.698	0.990	0.957 (0.625-1.467)	0.853	0.976	0.862 (0.614-1.209)	0.403	0.973
SubTrigl_LDL-1	0.939 (0.633-1.392)	0.776	0.955	0.962 (0.721-1.284)	0.794	0.990	0.576 (0.357-0.927)	0.041	0.934	1.017 (0.716-1.444)	0.929	0.985
SubTrigl_LDL-2	1.007 (0.668-1.518)	0.977	0.997	1.069 (0.81-1.41)	0.638	0.990	0.567 (0.353-0.912)	0.039	0.934	1.048 (0.73-1.504)	0.813	0.979
SubTrigl_LDL-3	1 (0.65-1.538)	0.999	0.999	0.832 (0.631-1.097)	0.187	0.990	0.859 (0.533-1.383)	0.567	0.976	1.105 (0.774-1.578)	0.609	0.973
SubTrigl_LDL-4	0.799 (0.502-1.272)	0.393	0.955	0.866 (0.661-1.135)	0.294	0.990	0.755 (0.439-1.297)	0.370	0.976	0.998 (0.694-1.436)	0.991	0.991
SubTrigl_LDL-5	0.783 (0.531-1.154)	0.261	0.826	0.787 (0.589-1.054)	0.102	0.990	0.717 (0.455-1.131)	0.207	0.976	1.095 (0.788-1.522)	0.603	0.973
SubTrigl_LDL-6	0.999 (0.651-1.531)	0.995	0.999	0.962 (0.732-1.263)	0.779	0.990	1.478 (0.93-2.35)	0.151	0.976	1.075 (0.752-1.537)	0.704	0.979
SubChol_LDL-1	1.107 (0.731-1.676)	0.662	0.955	0.934 (0.717-1.215)	0.609	0.990	0.826 (0.504-1.351)	0.490	0.976	1.07 (0.752-1.521)	0.719	0.979
SubChol_LDL-2	1.083 (0.729-1.61)	0.716	0.955	0.85 (0.637-1.136)	0.271	0.990	0.953 (0.616-1.473)	0.841	0.976	0.951 (0.682-1.325)	0.771	0.979
SubChol_LDL-3	1.081 (0.695-1.68)	0.754	0.955	0.971 (0.743-1.27)	0.833	0.990	0.84 (0.525-1.343)	0.501	0.976	0.926 (0.633-1.354)	0.703	0.979
SubChol_LDL-4	0.803 (0.537-1.201)	0.321	0.914	0.933 (0.698-1.248)	0.642	0.990	0.97 (0.633-1.484)	0.896	0.985	0.934 (0.664-1.313)	0.702	0.979
SubChol_LDL-5	0.849 (0.579-1.246)	0.437	0.955	0.97 (0.73-1.289)	0.834	0.990	0.819 (0.56-1.197)	0.335	0.976	0.791 (0.569-1.099)	0.165	0.973

SubChol_LDL-6	0.989 (0.677-1.445)	0.958	0.997	1.156 (0.882-1.514)	0.295	0.990	1.209 (0.786-1.86)	0.417	0.976	1.136 (0.809-1.594)	0.484	0.973
SubFreeChol_LDL-1	1.159 (0.775-1.735)	0.512	0.955	1.006 (0.764-1.323)	0.969	0.990	0.816 (0.508-1.311)	0.458	0.976	1.1 (0.782-1.548)	0.595	0.973
SubFreeChol_LDL-2	1.099 (0.763-1.583)	0.635	0.955	0.913 (0.688-1.212)	0.530	0.990	1.08 (0.701-1.665)	0.751	0.976	0.993 (0.717-1.376)	0.967	0.985
SubFreeChol_LDL-3	1.385 (0.87-2.205)	0.231	0.822	0.906 (0.693-1.183)	0.467	0.990	0.928 (0.585-1.471)	0.776	0.976	1.009 (0.699-1.454)	0.965	0.985
SubFreeChol_LDL-4	0.94 (0.606-1.457)	0.801	0.955	0.907 (0.677-1.216)	0.515	0.990	0.945 (0.614-1.455)	0.812	0.976	0.894 (0.625-1.279)	0.553	0.973
SubFreeChol_LDL-5	0.919 (0.621-1.36)	0.701	0.955	0.913 (0.688-1.211)	0.529	0.990	0.851 (0.588-1.231)	0.424	0.976	0.901 (0.657-1.237)	0.528	0.973
SubFreeChol_LDL-6	1.067 (0.703-1.619)	0.778	0.955	0.977 (0.735-1.299)	0.875	0.990	0.901 (0.598-1.356)	0.640	0.976	0.918 (0.65-1.298)	0.640	0.973
SubPhosp_LDL-1	1.095 (0.723-1.66)	0.697	0.955	0.944 (0.722-1.235)	0.677	0.990	0.795 (0.483-1.307)	0.413	0.976	1.077 (0.759-1.529)	0.691	0.979
SubPhosp_LDL-2	1.074 (0.713-1.618)	0.756	0.955	0.979 (0.743-1.29)	0.882	0.990	0.743 (0.457-1.207)	0.276	0.976	1.007 (0.714-1.421)	0.968	0.985
SubPhosp_LDL-3	1.266 (0.788-2.033)	0.381	0.955	0.945 (0.723-1.235)	0.677	0.990	0.793 (0.502-1.253)	0.354	0.976	1.015 (0.696-1.481)	0.941	0.985
SubPhosp_LDL-4	0.872 (0.568-1.339)	0.566	0.955	0.852 (0.627-1.158)	0.302	0.990	1.025 (0.64-1.643)	0.924	0.985	1.071 (0.737-1.555)	0.729	0.979
SubPhosp_LDL-5	0.872 (0.586-1.297)	0.535	0.955	0.836 (0.626-1.116)	0.224	0.990	0.804 (0.542-1.193)	0.314	0.976	0.84 (0.602-1.172)	0.311	0.973
SubPhosp_LDL-6	0.978 (0.656-1.456)	0.917	0.987	1.133 (0.855-1.5)	0.386	0.990	1.286 (0.819-2.02)	0.311	0.976	1.144 (0.806-1.622)	0.469	0.973
SubApoB_LDL-1	1.092 (0.728-1.638)	0.698	0.955	0.964 (0.726-1.28)	0.801	0.990	0.864 (0.525-1.421)	0.612	0.976	1.161 (0.819-1.646)	0.420	0.973
SubApoB_LDL-2	1.023 (0.687-1.525)	0.917	0.987	0.917 (0.688-1.223)	0.558	0.990	0.764 (0.468-1.249)	0.335	0.976	0.956 (0.681-1.344)	0.802	0.979
SubApoB_LDL-3	1.236 (0.8-1.912)	0.389	0.955	0.875 (0.668-1.146)	0.330	0.990	0.726 (0.458-1.149)	0.193	0.976	1.034 (0.711-1.502)	0.867	0.979
SubApoB_LDL-4	0.769 (0.505-1.171)	0.256	0.826	0.873 (0.646-1.18)	0.376	0.990	1.192 (0.743-1.91)	0.506	0.976	1.093 (0.764-1.563)	0.637	0.973
SubApoB_LDL-5	0.837 (0.57-1.228)	0.400	0.955	0.765 (0.567-1.032)	0.076	0.990	0.762 (0.519-1.12)	0.190	0.976	0.846 (0.606-1.182)	0.335	0.973
SubApoB_LDL-6	0.971 (0.659-1.428)	0.887	0.982	1.01 (0.766-1.331)	0.944	0.990	1.311 (0.83-2.069)	0.274	0.976	1.197 (0.846-1.694)	0.332	0.973
SubTrigl_HDL-1	0.88 (0.594-1.302)	0.562	0.955	0.996 (0.752-1.318)	0.975	0.990	0.887 (0.54-1.457)	0.683	0.976	1.193 (0.857-1.661)	0.307	0.973
SubTrigl_HDL-2	0.773 (0.535-1.116)	0.204	0.802	0.87 (0.652-1.162)	0.347	0.990	0.821 (0.533-1.264)	0.426	0.976	1.023 (0.747-1.401)	0.889	0.985
SubTrigl_HDL-3	0.626 (0.43-0.912)	0.019	0.301	0.877 (0.668-1.15)	0.340	0.990	0.737 (0.481-1.13)	0.220	0.976	1.003 (0.736-1.367)	0.986	0.991
SubTrigl_HDL-4	0.656 (0.449-0.959)	0.037	0.339	0.874 (0.659-1.158)	0.344	0.990	0.634 (0.429-0.937)	0.040	0.934	0.823 (0.59-1.147)	0.267	0.973
SubChol_HDL-1	0.992 (0.661-1.487)	0.971	0.997	0.972 (0.742-1.275)	0.839	0.990	1.497 (0.909-2.465)	0.176	0.976	1.315 (0.936-1.848)	0.130	0.973
SubChol_HDL-2	1.094 (0.767-1.56)	0.658	0.955	1.386 (1.026-1.872)	0.028	0.990	1.207 (0.75-1.943)	0.523	0.976	1.276 (0.929-1.754)	0.141	0.973
SubChol_HDL-3	1.149 (0.743-1.776)	0.570	0.955	0.915 (0.699-1.198)	0.518	0.990	0.983 (0.613-1.577)	0.950	0.985	1.265 (0.879-1.821)	0.215	0.973
SubChol_HDL-4	1.095 (0.748-1.602)	0.672	0.955	0.771 (0.57-1.042)	0.074	0.990	1.112 (0.754-1.642)	0.613	0.976	1.165 (0.836-1.624)	0.409	0.973
SubFreeChol_HDL-1	1.148 (0.735-1.793)	0.579	0.955	1.009 (0.761-1.338)	0.951	0.990	1.234 (0.767-1.985)	0.420	0.976	1.252 (0.858-1.828)	0.257	0.973
SubFreeChol_HDL-2	1.273 (0.84-1.928)	0.295	0.884	1.182 (0.902-1.548)	0.219	0.990	0.997 (0.639-1.556)	0.991	0.992	1.405 (0.973-2.027)	0.073	0.973

SubFreeChol_HDL-3	1.144 (0.72-1.819)	0.625	0.955	0.885 (0.664-1.18)	0.406	0.990	0.87 (0.569-1.331)	0.578	0.976	1.16 (0.82-1.642)	0.411	0.973
SubFreeChol_HDL-4	1.077 (0.707-1.64)	0.757	0.955	0.815 (0.615-1.081)	0.156	0.990	1.107 (0.758-1.616)	0.628	0.976	1.097 (0.797-1.508)	0.581	0.973
SubPhosp_HDL-1	0.994 (0.659-1.5)	0.980	0.997	1.058 (0.806-1.388)	0.686	0.990	1.396 (0.832-2.342)	0.290	0.976	1.367 (0.979-1.91)	0.076	0.973
SubPhosp_HDL-2	1.084 (0.736-1.596)	0.719	0.955	1.117 (0.852-1.463)	0.423	0.990	1.024 (0.617-1.702)	0.936	0.985	1.292 (0.943-1.769)	0.121	0.973
SubPhosp_HDL-3	0.964 (0.618-1.502)	0.882	0.982	0.861 (0.654-1.135)	0.283	0.990	0.738 (0.441-1.236)	0.297	0.976	1.224 (0.845-1.772)	0.300	0.973
SubPhosp_HDL-4	0.927 (0.6-1.432)	0.758	0.955	0.799 (0.6-1.063)	0.115	0.990	0.927 (0.576-1.493)	0.776	0.976	1.042 (0.719-1.509)	0.839	0.979
SubApoA1_HDL-1	0.96 (0.682-1.352)	0.828	0.963	1.024 (0.771-1.36)	0.872	0.990	1.072 (0.714-1.611)	0.762	0.976	1.113 (0.82-1.509)	0.504	0.973
SubApoA1_HDL-2	0.935 (0.606-1.442)	0.782	0.955	1.26 (0.956-1.661)	0.095	0.990	0.997 (0.646-1.538)	0.989	0.992	1.207 (0.852-1.711)	0.306	0.973
SubApoA1_HDL-3	0.94 (0.612-1.444)	0.796	0.955	0.937 (0.724-1.212)	0.620	0.990	1.095 (0.696-1.722)	0.717	0.976	1.269 (0.888-1.812)	0.200	0.973
SubApoA1_HDL-4	0.904 (0.604-1.353)	0.649	0.955	1.002 (0.757-1.325)	0.990	0.990	1.089 (0.734-1.617)	0.689	0.976	1.018 (0.712-1.455)	0.925	0.985
SubApoA2_HDL-1	0.942 (0.603-1.47)	0.813	0.955	1.012 (0.762-1.342)	0.937	0.990	1.151 (0.713-1.859)	0.606	0.976	1.38 (0.978-1.946)	0.044	0.973
SubApoA2_HDL-2	0.96 (0.624-1.475)	0.865	0.982	1.237 (0.935-1.636)	0.135	0.990	1.056 (0.666-1.674)	0.832	0.976	1.419 (1.004-2.006)	0.042	0.973
SubApoA2_HDL-3	0.918 (0.605-1.394)	0.716	0.955	0.892 (0.675-1.177)	0.419	0.990	0.774 (0.491-1.22)	0.311	0.976	1.277 (0.899-1.815)	0.184	0.973
SubApoA2_HDL-4	1.112 (0.747-1.656)	0.631	0.955	0.914 (0.692-1.208)	0.530	0.990	1.026 (0.703-1.497)	0.900	0.985	1.108 (0.796-1.542)	0.560	0.973

4.1.7 Exploration of blood metabolite signature of colorectal cancer and polyposis through integrated statistical and network analysis

Francesca Di Cesare[1,#]; Alessia Vignoli[1,2,3,#]; Claudio Luchinat[1,2,3], Prof; Leonardo Tenori[1,2,3]; and Edoardo Saccenti[4,*]

[#] Co-first authors

[1] Magnetic Resonance Center (CERM), University of Florence, 50019 Sesto Fiorentino, Italy

[2] Department of Chemistry "Ugo Schiff", University of Florence, 50019 Sesto Fiorentino, Italy

[3] Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine (C.I.R.M.M.P.), 50019 Sesto Fiorentino, Italy

[4] Laboratory of Systems and Synthetic Biology, Wageningen University & Research, 6708 WE 643 Wageningen, The Netherlands

In preparation

Candidate's contribuitions: statistical analysis, interpretation of results, and writing the manuscript.

ABSTRACT

BACKGROUND: Colorectal cancer (CRC), one of the most prevalent and deadly cancers worldwide, generally evolves from adenomatous polyps. The understanding of the molecular mechanisms underlying this pathological evolution is crucial for diagnostic and prognostic purposes. Integrative systems biology approaches offer an optimal point of view to analyze CRC and polyp patients.

METHODS: Here we present the study of the association networks constructed from a publicly available array of 113 serum metabolites measured on a cohort of 234 subjects from three groups (66 CRC patients, 76 polyp patients, and 92 healthy controls). The concentrations of serum metabolites were obtained via targeted liquid chromatography-tandem mass spectrometry.

RESULTS: In terms of architecture, topology, and connectivity, the metabolite-metabolite association network of CRC patients appears to be completely different with respect to polyp patients and healthy controls. The most relevant nodes in the CRC network are those related to energy metabolism. Interestingly, phenylalanine, tyrosine, and tryptophan metabolism are found to be involved in both CRC and polyposis.

CONCLUSIONS: Our results demonstrate that the characterization of metabolite–metabolite association networks is a promising and powerful tool to investigate molecular aspects of CRC.

1. INTRODUCTION

Colorectal cancer is the third most commonly diagnosed cancer and the second leading cause of cancer-related death worldwide[1]. The major risk factors for developing CRC include increased age, male sex, inflammatory bowel disease, alcohol intake, smoking, obesity, a diet rich in red meat, and family history[2,3]. Early-stage diagnosis is associated with a good prognosis (~90% 5-year survival), however, survival declines substantially when the tumor is identified later and is already metastasized[4]. In its early stages, CRC remains often asymptomatic; thus, organized screening programs aimed at increasing CRC early diagnosis are needed to decrease its morbidity and mortality[3]. Currently, occult blood in the feces is the most commonly used marker in clinical practice; however, it lacks the sensitivity and specificity needed for unambiguous early diagnosis[5]. Colonoscopy is currently recognized as the gold-standard diagnostic method for CRC: it offers high sensitivity, specificity, and accuracy; unfortunately, it is a costly and extremely invasive procedure[6]. Therefore, the development of new methodologies able to identify CRC with minimal invasiveness and high accuracy at an early stage is highly desirable.

It is well known from the clinical literature that about 95% of CRC begin as colonic adenomatous polyps[7]. A series of still not completely characterized molecular alterations induce CRC growth. Broadly speaking, the development of CRC can be attributable to a complex multistep process, ignited by the growth of a benign polyp with the potential to evolve into an in-situ carcinoma by the accumulation of additional somatic mutations[8,9]. When detected, polyps' excision and adequate treatment[10] can prevent further tumor development. To develop therapeutic strategies able to prevent the transition from health epithelium to polyps to CRC it is necessary to fully understand the puzzle of the underlying biochemical mechanisms. The biochemistry of early carcinogenesis is controlled by an interplay between genomic susceptibility, metabolic reprogramming, and the local microenvironment. The development of adenomatous polyps probably passes through the impact of oxidative stress on the metabolic pathways involved in the renewal of the colon epithelium. The immune system also plays a role because chronic inflammation promotes cell proliferation and differentiation. Last, but not least, microbiota dysregulation may influence microenvironmental homeostasis altering the integrity of the colon mucosa[8].

Considering that all these entangled effects have an impact on the local and systemic patient metabolism, metabolomics represents a valid instrument to provide further insight into the CRC metabolic mechanisms. Indeed, metabolomics has already proved to be an excellent tool for biomedical investigations[11–17], and it has been opportunely applied to the study of CRC adopting different analytical strategies and kind of samples[5]. To expand the vision from the

particular to the general, it is necessary to perform a further step considering the shape of the metabolite-metabolite association networks[18]. Changes in metabolite association patterns are associated with changes in the pathophysiological conditions, and the resulting networks can be compared across conditions under the postulate that the differences and commonalities observed in the parameters of the reconstructed networks[19–23] faithfully mirror the reshape of the underlying biological networks.

With this premise, we decided to apply a metabolite-metabolite association network approach (**Figure 1**) to the openly available data published by Zhu *et al.*[24]. In that study, the authors applied a mass spectrometry-based metabolic profiling approach to discover candidate biomarkers for CRC detection using human serum samples. The authors found a robust metabolic signature able to discriminate between healthy controls (CTR), CRC patients, and polyp patients (PP). From our analysis, it emerges that the network calculated on serum samples of CRC patients has a different architecture with respect to that of PP and CTR patients, while no relevant differences emerge comparing colon and rectus cancers.

2. MATERIALS AND METHODS

2.1. Dataset description

We re-analyzed the metabolomics data set collected by Zhu *et al.*[24]. Data and sample information were downloaded from the UC San-Diego Metabolomics Workbench public repository[25] (https://www.metabolomicsworkbench.org/) with the following Project ID number PR000226. For further details on the clinical samples, sample collection and preparation, Mass Spectrometry experiments, and metabolite annotation and quantification we refer the readers to the original paper[24].

The original data set contains 113 metabolites measured using Mass Spectroscopy on n=234 serum samples from three different groups. Patients with colorectal cancer (CRC) which included patients with colon and rectal cancer, patient with polyposis (PP), and healthy controls (CTR). The patients enrolled were age- and gender-matched in each group.

We removed n = 8 outliers (see Section 2.2.1), leaving n=226 samples/subjects for analysis, (see Supplementary Material, Figure S1) divided into n=65 CRC patients (n_1=35 men, n_2=30 women), n=74 PP patients (n_1=35 men, n_2=39 women), and n=87 CTR (n_1=42 men, n_2=45 women).

The mean age ± standard deviation (SD) of each group is 58.4±13.3 yrs, 55.5±7.0 yrs, and 54.2±13.6 yrs, respectively. The n=65 CRC patients were also divided into n=39 colon cancer patients (n_1=20 men, n_2=19 women) and n=26 rectal cancer patients (n_1=15 men, n_2=11 women), with a mean of age ± standard deviation (SD) of 58.7±14.2 yrs and 58.0±12.0 yrs, respectively.

2.2. Data pre-processing

2.2.1. Data overview and normalization

Outliers were determined as those samples/subjects falling outside the 95% confidence ellipses on a two-dimensional space reduced with Principal component analysis (PCA)[26,27]; a total of 8 patients (3.4%) (1 CRC patient, forming part of the rectal cancer sub-group, 2 PP patients, and 5 CTR) were removed from the analysis (see Supplementary Material Figure 1S).

Metabolites' abundances were normalized using "RankNorm" function before analysis. The offset k=(3/8) default parameter, corresponding to the Blom's transformation[28], was used. This method applies the rank-based inverse normal transform (INT) in two-steps. Firstly, the observations were transformed into the scale of probabilities using the empirical cumulative distribution function (eCDF)[29]. Subsequently, the observations were transformed into Z-scores, using the probit function[30].

2.3. Univariate analysis

Univariate Student's t-test[31] was used to compare normalized metabolite concentrations between patient groups (CRC vs CTR, PP vs CTR, and CRC vs PP). Benjamini- Hochberg method was used to correct for multiple testing[32] and FDR adjusted *P*-values < 0.05 were considered statistically significant. Variables were transformed by taking the square root of the values to correct for heteroscedasticity[33,34]

2.4. Multivariate analysis

Principal Component Analysis (PCA)[26,27] was applied to data scaled to unit variance to explore data patterns.

The Random Forest (RF) algorithm[35–37] was employed for pairwise classification comparing CRC and healthy patients, CRC and PP patients, and PP and CTR patients. To reduce the potential bias due to an unbalanced number of subjects/samples per group, we imposed a number of k=100 resampling, retaining 85% of data for each group to be compared. Accuracy, sensitivity, specificity, and corresponding 95% CI were calculated according to the standard definitions and given as average over the 100 resamplings. The model quality statistics (accuracy, sensitivity, specificity, and the area under the ROC, AUROC) were calculated according to standard definitions. The significance of the model and the importance metrics were determined with a permutation-test using k=1000 times permutations. In particular, the permutation-test yields a null distribution D_{perm} of permutated model quality measures. The significance of each measure was calculated as a *P*-value by comparing the

value $model_0$ obtained from the original and non-permuted data with the values $model_1$, $model_2$, ..., $model_K$ obtained from the *k*-times permutation-test. The P-value for a specific model measure is calculated as follows:

$$P - value(measure) = \frac{1 + \left(\left|D_{perm|measure}\right| > |measure\ model_0|\right)}{k} \quad (1)$$

2.5. Network analysis

2.5.1. Reconstruction of the metabolite association network

The Probabilistic Context Likelihood of Relatedness based on Correlation (PCLRC)[38] algorithm was used to infer metabolite-metabolite association networks. In order to remove non-significant background correlations, this algorithm provides a robust evaluation of the correlation using a resampling strategy in combination with the previously published Context Likelihood of Relatedness (CLR)[39] approach. Spearman correlation was used as a measure of association between metabolites[40]. The PCLRC algorithm outputs a probability matrix **P** giving the likelihood p_{ij} for each Spearman correlation r_{ij} between metabolites *i* and *j*. A detailed description of the PCLRC approach is provided in the Supplementary Methods (Section 1).

2.5.2. Pathway enrichment analysis

Pathway enrichment analysis on the set of statistically differential connected metabolites obtained comparing CRC vs CTR and PP vs CTR was performed using the tool available on MetaboAnalyst 4.0[41] (www.metaboanalyst.ca), implementing a hypergeometric test. The pathway impact score is calculated by this tool as the sum of the importance measures of the matched metabolites normalized by the sum of the importance measures of all metabolites in each pathway. Only enriched pathways with an impact score > 0.01 were considered.

2.5.3. Measures for network topology

Network topology and node (metabolite) characteristics were evaluated using several standard metrics in addition to connectivity (node degree). We used *Average Shortest Path Length*, *Betweenness Centrality*, *Closeness Centrality*, *Clustering Coefficient*, *Degree*, *Eccentricity*, *Neighborhood Connectivity*, *Radiality*, *Stress*, and *Topological Coefficient*. These measures were calculated using Network Analyzer[42], a Java plugin available for the Cytoscape platform[43]. A brief overview of the measures is given in the Supplementary Methods (Section 2).

2.6. Software

Calculations were performed using R statistical software (version 3.3.2). The R implementation of PCRLC algorithm and the code to perform differential connectivity analysis are available at the following link: semantics.systemsbiology.nl. The RNOmi R package[44] was used to normalize metabolic abundance. To generate RF-classification models, the "randomForest" function, implemented in the R package Random Forest[45], was used to grow a decision forest composed of 1000 trees, using default parameters. To estimate the significance of importance metrics for the RF models by 1000-times permuting the response variable, the "rp.importance" function, implemented in the R package rfPermute, was used. The network visualization and the estimation of network topology and statistics were performed using Cytoscape platform (version 3.8.2)[43], integrated with the NetworkAnalyzer plugin[42].

3. RESULTS

3.1. Exploratory analysis of serum metabolomic profiles of CRC, polyposis, and healthy patients

To analyze comprehensively the metabolic profiles of the subjects from the three analyzed groups (CRC, PP, and CTR patients), a Principal Component Analysis (PCA) model was performed on the *n*=226 serum samples.

The PCA score plot (**Figure 2a**) shows that the CRC, PP patients, and CTR subjects are not clearly separated, suggesting that metabolic differences are too subtle to be resolved using an unsupervised multivariate approach.

Moreover, to evaluate the metabolic differences among the three groups, a univariate Student's t-test was performed on normalized metabolite concentration. As reported in Table 1, 24 metabolites showed statistically significant (FDR adjusted *P*-value < 0.05) differences comparing CRC with CTR subjects. Instead, no significantly different metabolites were found comparing PP with CTR subjects. 23 metabolites showed statistically significant (FDR adjusted *P*-value < 0.05) differences comparing CRC with PP subjects (Table 1).

3.2. Classification of CRC, polyposis, and healthy patients

We used Random Forest classification to investigate whether the metabolomic profiles could be used to discriminate in a predictive way CRC, PP patients, and CTR subjects. Classification results are given in **Table 2**. Overall, we obtained good classification models to discriminate between CRC patients and healthy control (78.2% mean accuracy) (**Figure 2b**) and between CRC patients and PP patients (79.5% mean accuracy) (**Figure 2c**). A weaker predictive model (62.2% mean accuracy) (**Figure 2d**) was obtained for the discrimination of

CTR subjects and PP patients. Hippuric acid, malonic acid/3-hydroxybutyric acid, linolenic acid, histidine, glycochenodeoxycholate, adenylsuccinate, phosphoenolpyruvic acid (PEP), glyceraldehyde, fructose 1,6 bisphosphate/ fructose 2,6 bisphosphate (F16BP/F26BP), linolenic acid, maleic acid, adipic acid, glycocholate, and gamma-aminobutyrate are the most relevant and significant variables in the model discriminating CRC with respect to CTR (**Figure S3a**). PEP, adenosine, glyceraldehyde, methionine, hippuric acid, 2-deoxyuridine, linolenic acid, creatinine, xanthurenate, linoleic acid, uridine, glycocholate, aspartic acid, glycochenodeoxycholate, dimethylglycine, adipic acid, glutaric acid, lysine, and histidine are the metabolites that contribute significantly in the discrimination between CRC and PP (**Figure S3b**). Only a few variables, namely F16BP/F26BP, adenosine, tryptophan, xanthurenate, salicylurate, G16BP, oxaloacetate, glyceraldehyde, and histidine, significantly discriminate PP and CTR subjects. Taken together these results indicate the presence of a strong metabolic signature specific to CRC patients and a weaker signature specific to polyposis (**Figure S3c**).

3.1. Analysis of metabolite-metabolite association networks specific to colorectal cancer, polyposis, and healthy subjects

The metabolite-metabolite association networks of CRC, PP, and CTR patients are given in **Figure 3a-c**, respectively. Comparing the three metabolic structures, we observed that the CRC-specific network has a different topology than polyposis and healthy control-specific networks. In particular, the CRC network (**Figure 3a**) tends to be poorly connected but establishes robust connections ($|r_{ij}|>0.6$) between glucose, lactate, oxalic acid, aconitate, citraconic acid, leucine/isoleucine, valine, and guanidinoacetate. In contrast, the polyposis (**Figure 3b**) and the healthy control networks (**Figure 3c**) are more interconnected, showing a very similar topology.

3.1. Comparison of the topological properties of metabolite-metabolite association network

To compare comprehensively CRC, PP, and CTR metabolite-metabolite association networks, we examined the topological and statistical network parameters via PCA (**Figure 4**). Analyzing the PCA biplot, the network specific to the healthy control and the polyposis patients are characterized by different Topological Coefficient and Eccentricity. This means that the nodes in the two networks tend to have different shared neighbors. Eccentricity is a measure of centrality thus indicating the importance of a node (*i.e.* metabolite) in the networks and in this case the two networks tend to have different important nodes. The CRC network differs from the PP and CTR network on node characteristics like Betweenness, Centrality, and Average Shortest Path Length, which describe how the different nodes, *i.e.* metabolites,

are interconnected reflecting the low connectedness of the CRC network. Moreover, the metabolite-metabolite association networks for the two CRC sub-groups (colon and rectal cancer) show no differences in terms of topological parameters (Supplementary Material Figure S3).

3.1. Differential Connectivity analysis

To quantify the difference in metabolite association patterns and to highlight possible underlying metabolic differences within the three groups as well as within the two CRC sub-groups (colon and rectal cancer), a differential connectivity analysis was performed comparing the CRC and the polyposis networks against the control (healthy) network, the CRC against polyposis, and the colon cancer against rectal cancer: results are shown in **Figure 5**.

We observed a large number of metabolites (amino acids and metabolites involved in carbohydrates metabolism) that are differentially connected in the CRC and CTR networks (**Figure 5a**) and in the CRC and PP networks (**Figure 5b**). In contrast, we observed a limited number of differentially connected metabolites when comparing the CTR and the PP networks (**Figure 5c**), in complementary agreement with the Random Forest predictive models (**Table 1**). When comparing the networks of the two sub-groups of colon and rectal cancer, we observed differential connectivity only for 4-pyridoxic acid, L-kynurenine, and phenylalanine (Figure 4(d)), which indicates similarity among the two networks.

Metabolite pathway enrichment analysis was performed on the set of statistically differential metabolites found comparing CRC vs CTR, PP vs CTR group: results are shown in **Figure 6.** We found a total of 22 unique enriched pathways (impact score > 0.01) in CRC compared with CTR network and in PP compared with CTR network. The most impacted and differential enriched pathways are: i) the phenylalanine, tyrosine, and tryptophan biosynthesis and phenylalanine metabolism in both CRC and PP compared to the CTR group, ii) the synthesis and degradation of ketone bodies in polyposis compared to CTR, and iii) the glycine, serine, and threonine metabolism in CRC compared to CTR. Moreover, the aminoacyl−tRNA biosynthesis pathway plays a statistically significant role (FDR adjusted *P*-value < 0.05) with a moderate impact position in CRC with respect to CTR.

4. DISCUSSION

Colorectal carcinomas principally evolve from adenomatous polyps. The adenoma–carcinoma sequence is a central tenet for the early diagnosis and prevention of CRC. Fortunately, more than 90% of adenomas do not progress to cancer; however, it is currently not possible to reliably identify those ones that will progress and those not, thus the complete

resection of polyps during colonoscopy represents the only available option to eliminate the risk of cancer derived from that adenomas[46]. Malignant tumors are characterized by peculiar metabolic alterations such as increased gluconeogenesis, glycolysis and fat mobilization, and decreased protein synthesis. These alterations as well as the individual immune-metabolic response induced by the presence of cancer constitute the characteristic metabolic signature of CRC patients[24,47]. In the analyses here proposed, we present an evaluation of the serum metabolomic profiles of CRC and PP patients with respect to healthy controls using a metabolite–metabolite association networks approach to investigate and explore the existence of molecular mechanisms underlying these different clinical profiles.

Supervised balanced RF models show good discriminations for the comparisons between CRC and CTR and between CRC and PP obtaining an AUC of 0.875 and of 0.871 respectively. Conversely, PP and CTR present only slight differences (AUC of 0.661). Despite the different statistical approaches employed, our results are in line with those reported in the original publication of Zhu and coauthors[24]. These data confirm that CRC patients develop systemic metabolic alterations that are not present, not only in healthy controls but also in PP patients.

We compared the metabolite-metabolite association networks of CRC, PP, and CTR to explore the magnitude, topology, and architecture of metabolite connections and their variability. The rationale of this systems-based approach is that metabolites behave in an orchestrated fashion; perturbations of the systems induced by different pathophysiological conditions cause modifications in the relationships among metabolites which are reflected not only in their levels but also in their connectivity patterns[22].

From our analysis, it emerges that the network calculated on serum samples of CRC patients has an architecture completely different from the ones of PP and CTR, whereas no relevant difference emerges comparing colon and rectus cancers. The CRC network is characterized by a relatively low number of strong connections. This difference is corroborated also by the differential connectivity analysis and by the analysis of the topological parameters: CRC differs from PP and CTR for betweenness, centrality, average shortest path length, and for the number of differential connections reflecting the less interconnection of this network. Of note, nodes present in the CRC network are all directly or indirectly related to the different pathways involved in the energetic metabolism. Cancer cells need to meet a high energy demand to support cell proliferation and migration, thus they acquire molecular substrates and energy through unusual metabolic pathways. This need induces a profound rewiring of their metabolic network that extends beyond the Warburg effect and alterations of individual metabolic flux[48,49].

We can speculate that the defragmentation of the CRC network could be the result of the multiple activations of several metabolic pathways and that these profound alterations are

reflected at a systemic level in the sera of CRC patients. On the other hand, PP and CTR seem to have a similar architecture with more interconnections. However, they are characterized by different topological coefficients and eccentricity; this means that the same node (metabolite) in the two networks has different importance and is connected with diverse metabolites.

Pathway analysis of the differentially connected metabolites reveals the involvement of phenylalanine, tyrosine, and tryptophan metabolism in both colorectal cancer and polyposis. Tryptophan and phenylalanine are essential amino acids: tryptophan metabolism is linked to the production of serotonin while phenylalanine is required for the production of tyrosine which is catalyzed by phenylalanine hydroxylase. It has been shown that phenylalanine hydroxylase activity can be altered in inflammation or malignancy[50,51]. Tyrosine is converted into several metabolites, including L-DOPA, pyruvate, fumarate, and phenol, the latter conversion is mediated by the enzyme tyrosine phenol-lyase (β-tyrosinase)[52]. Alterations in the profiles of these metabolites have been reported in colorectal cancer[53] and other cancer types[54,55] but, to the best of our knowledge, alterations in these pathways have never been reported in association with polyposis.

The landscape of metabolic alterations associated with polyposis appears to be more heterogenous than the ones associated with colorectal cancer. We observed the involvement of glycolysis and glucogenesis and glycine, serine, and threonine metabolism as a signature for polyposis, together with fructose and mannose metabolism. However, none of these alterations are statistically significant after FDR correction.

The alteration of fructose and mannose metabolism is of particular interest since a previous article demonstrated that higher levels of D-Mannose are associated with increased risks of PP and CRC[56]. Mannose is a monosaccharide that is a key metabolite in human metabolism, involved in the glycosylation of proteins[57]: mannose and glucose are involved in the production of mannose-6-phosphate, which is then converted to mannose-1-phosphate and guanosine diphosphate mannose for the process of N-linked glycosylation[58]. Increased concentrations of mannose have been observed in metastatic breast cancer in comparison with localized early disease[59] suggesting the altered anabolic requirements of cancer cells[60,61] and incompletion of the glycosylation process[62].

5. CONCLUSIONS

In this study we-reanalyzed a publicly available data set[24] concerning a metabolomic investigation of blood metabolite profiles in patients with colorectal cancer, polyposis, and healthy controls. We expanded the original analysis by integrating classical univariate analysis of metabolite abundances with predictive multivariate modelling and the analysis of

metabolite-metabolite association patterns. We observed that both colorectal cancer and polyposis possess specific metabolic signatures that distinguish them from healthy control profiles. In both cases, we observed deregulation of phenylalanine, tyrosine, and tryptophan metabolism. However, the metabolic landscape associated with polyposis appears to be more variegated than that of colorectal cancer, with several affected pathways, although these dysregulations are subtle, as indicated by the relatively weaker discriminating models between healthy and polyposis. Noteworthy, metabolite network analysis indicates the involvement of mannose and mannose metabolism in polyposis, a new result not found in the original study, suggesting the need for follow-up studies in this direction.

6. REFERENCES

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*. n/a(n/a). doi:https://doi.org/10.3322/caac.21660
2. Ren Z, Rajani C, Jia W. The Distinctive Serum Metabolomes of Gastric, Esophageal and Colorectal Cancers. *Cancers (Basel)*. 2021;13(4). doi:10.3390/cancers13040720
3. Dekker E, Tanis PJ, Vleugels JLA, Kasi PM, Wallace MB. Colorectal cancer. *Lancet*. 2019;394(10207):1467-1480. doi:10.1016/S0140-6736(19)32319-0
4. Siegel R, Desantis C, Jemal A. Colorectal cancer statistics, 2014. *CA Cancer J Clin*. 2014;64(2):104-117. doi:10.3322/caac.21220
5. Nannini G, Meoni G, Amedei A, Tenori L. Metabolomics profile in gastrointestinal cancers: Update and future perspectives. *World journal of gastroenterology*. 2020;26(20):2514-2532. doi:10.3748/wjg.v26.i20.2514
6. Issa IA, Noureddine M. Colorectal cancer screening: An updated review of the available options. *World journal of gastroenterology*. 2017;23(28):5086-5096. doi:10.3748/wjg.v23.i28.5086
7. Manne U, Shanmugam C, Katkoori VR, Bumpers HL, Grizzle WE. Development and progression of colorectal neoplasia. *Cancer Biomarkers*. 2011;9:235-265. doi:10.3233/CBM-2011-0160
8. Aceto GM, Catalano T, Curia MC. Molecular Aspects of Colorectal Adenomas: The Interplay among Microenvironment, Oxidative Stress, and Predisposition. *Biomed Res Int*. 2020;2020. doi:10.1155/2020/1726309
9. Haggar FA, Preen DB, Pereira G, Holman CDJ, Einarsdottir K. Cancer incidence and mortality trends in Australian adolescents and young adults, 1982-2007. *BMC Cancer*. 2012;12:151. doi:10.1186/1471-2407-12-151
10. Yang L, Wang S, Lee JJ-K, et al. An enhanced genetic model of colorectal cancer progression history. *Genome Biology*. 2019;20(1):168. doi:10.1186/s13059-019-1782-4
11. Vignoli A, Muraro E, Miolo G, et al. Effect of Estrogen Receptor Status on Circulatory Immune and Metabolomics Profiles of HER2-Positive Breast Cancer Patients Enrolled for Neoadjuvant Targeted Chemotherapy. *Cancers (Basel)*. 2020;12(2). doi:10.3390/cancers12020314
12. Vignoli A, Paciotti S, Tenori L, et al. Fingerprinting Alzheimer's Disease by 1H Nuclear Magnetic Resonance Spectroscopy of Cerebrospinal Fluid. *J Proteome Res*. 2020;19(4):1696-1705. doi:10.1021/acs.jproteome.9b00850
13. Dani C, Bresci C, Berti E, et al. Metabolomic profile of term infants of gestational diabetic mothers. *Journal of Maternal-Fetal & Neonatal Medicine*. 2014;27(6):537-542. doi:10.3109/14767058.2013.823941
14. Pirazzini C, Azevedo T, Baldelli L, et al. A geroscience approach for Parkinson's disease: Conceptual framework and design of PROPAG-AGEING project. *Mechanisms of Ageing and Development*. 2021;194:111426. doi:10.1016/j.mad.2020.111426
15. Basoglu A, Baspinar N, Tenori L, Vignoli A, Yildiz R. Plasma metabolomics in calves with acute bronchopneumonia. *Metabolomics*. 2016;12(8):128. doi:10.1007/s11306-016-1074-x
16. Takis PG, Ghini V, Tenori L, Turano P, Luchinat C. Uniqueness of the NMR approach to metabolomics. *TrAC Trends in Analytical Chemistry*. 2019;120:115300. doi:10.1016/j.trac.2018.10.036
17. Vignoli A, Ghini V, Meoni G, et al. High-Throughput Metabolomics by 1D NMR. *Angew Chem Int Ed Engl*. 2019;58(4):968-994. doi:10.1002/anie.201804736
18. Rosato A, Tenori L, Cascante M, De Atauri Carulla PR, Martins dos Santos VAP, Saccenti E. From correlation to causation: analysis of metabolomics data using systems biology approaches. *Metabolomics*. 2018;14(4):37. doi:10.1007/s11306-018-1335-y
19. 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. doi:10.1021/acs.jproteome.6b00454
20. 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. doi:10.1021/pr501075r
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. doi:10.1021/acs.jproteome.7b00404

22. Vignoli A, Tenori L, Luchinat C, Saccenti E. Differential Network Analysis Reveals Molecular Determinants Associated with Blood Pressure and Heart Rate in Healthy Subjects. *J Proteome Res*. 2021;20(1):1040-1051. doi:10.1021/acs.jproteome.0c00882
23. Vignoli A, Tenori L, Giusti B, et al. 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. doi:10.1021/acs.jproteome.9b00779
24. Zhu J, Djukovic D, Deng L, et al. Colorectal Cancer Detection Using Targeted Serum Metabolic Profiling. *J Proteome Res*. 2014;13(9):4120-4130. doi:10.1021/pr500494u
25. Sud M, Fahy E, Cotter D, et al. Metabolomics Workbench: An international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. *Nucleic Acids Res*. 2016;44(Database issue):D463-D470. doi:10.1093/nar/gkv1042
26. Hotelling H. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*. 1933;24(6):417-441. doi:10.1037/h0071325
27. Wall ME, Rechtsteiner A, Rocha LM. Singular Value Decomposition and Principal Component Analysis. In: Berrar DP, Dubitzky W, Granzow M, eds. *A Practical Approach to Microarray Data Analysis*. Springer US; 2003:91-109. doi:10.1007/0-306-47815-3_5
28. Ludwig O. Blom, Gunnar: Statistical estimates and transformed beta-variables. Wiley/New York, Almquist und Wiksell/Stockholm 1958; 176 S., Kr. 20,—. *Biometrische Zeitschrift*. 1961;3(4):285-285. doi:https://doi.org/10.1002/bimj.19610030410
29. Vaart AW van der. *Asymptotic Statistics*. Cambridge University Press; 2000. Accessed March 2, 2021. https://econpapers.repec.org/bookchap/cupcbooks/9780521784504.htm
30. Bliss CI. The Method of Probits. *Science*. 1934;79(2037):38-39. doi:10.1126/science.79.2037.38
31. Student. The Probable Error of a Mean. *Biometrika*. 1908;6(1):1-25. doi:10.2307/2331554
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;57(1):289-300.
33. van den Berg RA, Hoefsloot HC, Westerhuis JA, Smilde AK, van der Werf MJ. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics*. 2006;7(1):142. doi:10.1186/1471-2164-7-142
34. Bartlett MS. The Square Root Transformation in Analysis of Variance. *Supplement to the Journal of the Royal Statistical Society*. 1936;3(1):68-78. doi:https://doi.org/10.2307/2983678
35. Ali J, Khan R, Ahmad N, Maqsood I. Random Forests and Decision Trees. *International Journal of Computer Science Issues(IJCSI)*. 2012;9.
36. Chen T, Cao Y, Zhang Y, et al. Random Forest in Clinical Metabolomics for Phenotypic Discrimination and Biomarker Selection. *Evidence-Based Complementary and Alternative Medicine*. 2013;2013:e298183. doi:10.1155/2013/298183
37. Amaratunga D, Cabrera J, Lee Y-S. Enriched random forests. *Bioinformatics*. 2008;24(18):2010-2014. doi:10.1093/bioinformatics/btn356
38. Suarez-Diez M, Saccenti E. Effects of Sample Size and Dimensionality on the Performance of Four Algorithms for Inference of Association Networks in Metabonomics. *J Proteome Res*. 2015;14(12):5119-5130. doi:10.1021/acs.jproteome.5b00344
39. Akhand MAH, Nandi RN, Amran SM, Murase K. Context likelihood of relatedness with maximal information coefficient for Gene Regulatory Network inference. In: *2015 18th International Conference on Computer and Information Technology (ICCIT)*. ; 2015:312-316. doi:10.1109/ICCITechn.2015.7488088
40. Jahagirdar S, Saccenti E. On the Use of Correlation and MI as a Measure of Metabolite—Metabolite Association for Network Differential Connectivity Analysis. *Metabolites*. 2020;10(4):171. doi:10.3390/metabo10040171
41. Chong J, Soufan O, Li C, et al. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res*. 2018;46(W1):W486-W494. doi:10.1093/nar/gky310
42. Doncheva NT, Assenov Y, Domingues FS, Albrecht M. Topological analysis and interactive visualization of biological networks and protein structures. *Nature Protocols*. 2012;7(4):670-685. doi:10.1038/nprot.2012.004
43. Shannon P, Markiel A, Ozier O, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res*. 2003;13(11):2498-2504. doi:10.1101/gr.1239303

44. McCaw ZR, Lane JM, Saxena R, Redline S, Lin X. Operating characteristics of the rank-based inverse normal transformation for quantitative trait analysis in genome-wide association studies. *Biometrics*. 2020;76(4):1262-1272. doi:https://doi.org/10.1111/biom.13214
45. Liaw A, Wiener M. Classification and Regression by randomForest. *R news*. 2002;2(3):18-22.
46. Levine JS, Ahnen DJ. Clinical practice. Adenomatous polyps of the colon. *N Engl J Med*. 2006;355(24):2551-2557. doi:10.1056/NEJMcp063038
47. Nannini G, Meoni G, Amedei A, Tenori L. Metabolomics profile in gastrointestinal cancers: Update and future perspectives. *World J Gastroenterol*. 2020;26(20):2514-2532. doi:10.3748/wjg.v26.i20.2514
48. Tang Z, Xu Z, Zhu X, Zhang J. New insights into molecules and pathways of cancer metabolism and therapeutic implications. *Cancer Communications*. 2021;41(1):16-36. doi:https://doi.org/10.1002/cac2.12112
49. La Vecchia S, Sebastián C. Metabolic pathways regulating colorectal cancer initiation and progression. *Seminars in Cell & Developmental Biology*. 2020;98:63-70. doi:10.1016/j.semcdb.2019.05.018
50. Neurauter G, Grahmann AV, Klieber M, et al. Serum phenylalanine concentrations in patients with ovarian carcinoma correlate with concentrations of immune activation markers and of isoprostane-8. *Cancer Lett*. 2008;272(1):141-147. doi:10.1016/j.canlet.2008.07.002
51. Ploder M, Neurauter G, Spittler A, Schroecksnadel K, Roth E, Fuchs D. Serum phenylalanine in patients post trauma and with sepsis correlate to neopterin concentrations. *Amino Acids*. 2008;35(2):303-307. doi:10.1007/s00726-007-0625-x
52. Yamada H, Kumagai H. Synthesis of L-Tyrosine-Related Amino Acids by β-Tyrosinase11Systematic name: Tyrosine phenol-lyase (deaminating) (EC 4.1.99.2). In: Perlman D, ed. *Advances in Applied Microbiology*. Vol 19. Academic Press; 1975:249-288. doi:10.1016/S0065-2164(08)70431-3
53. Leichtle AB, Nuoffer J-M, Ceglarek U, et al. Serum amino acid profiles and their alterations in colorectal cancer. *Metabolomics*. 2012;8(4):643-653. doi:10.1007/s11306-011-0357-5
54. Wiggins T, Kumar S, Markar SR, Antonowicz S, Hanna GB. Tyrosine, Phenylalanine, and Tryptophan in Gastroesophageal Malignancy: A Systematic Review. *Cancer Epidemiol Biomarkers Prev*. 2015;24(1):32-38. doi:10.1158/1055-9965.EPI-14-0980
55. Contorno S, Darienzo RE, Tannenbaum R. Evaluation of aromatic amino acids as potential biomarkers in breast cancer by Raman spectroscopy analysis. *Scientific Reports*. 2021;11(1):1698. doi:10.1038/s41598-021-81296-3
56. Long Y, Sanchez-Espiridion B, Lin M, et al. Global and targeted serum metabolic profiling of colorectal cancer progression. *Cancer*. 2017;123(20):4066-4074. doi:10.1002/cncr.30829
57. Alton G, Hasilik M, Niehues R, et al. Direct utilization of mannose for mammalian glycoprotein biosynthesis. *Glycobiology*. 1998;8(3):285-295. doi:10.1093/glycob/8.3.285
58. Ichikawa M, Scott DA, Losfeld M-E, Freeze HH. The metabolic origins of mannose in glycoproteins. *J Biol Chem*. 2014;289(10):6751-6761. doi:10.1074/jbc.M113.544064
59. Jobard E, Pontoizeau C, Blaise BJ, Bachelot T, Elena-Herrmann B, Trédan O. A serum nuclear magnetic resonance-based metabolomic signature of advanced metastatic human breast cancer. *Cancer Lett*. 2014;343(1):33-41. doi:10.1016/j.canlet.2013.09.011
60. Sleat DE, Chen TL, Raska K, Lobel P. Increased levels of glycoproteins containing mannose 6-phosphate in human breast carcinomas. *Cancer Res*. 1995;55(15):3424-3430.
61. Goetz JA, Mechref Y, Kang P, Jeng M-H, Novotny MV. Glycomic profiling of invasive and non-invasive breast cancer cells. *Glycoconj J*. 2009;26(2):117-131. doi:10.1007/s10719-008-9170-4
62. de Leoz MLA, Young LJT, An HJ, et al. High-mannose glycans are elevated during breast cancer progression. *Mol Cell Proteomics*. 2011;10(1):M110.002717. doi:10.1074/mcp.M110.002717

FIGURES

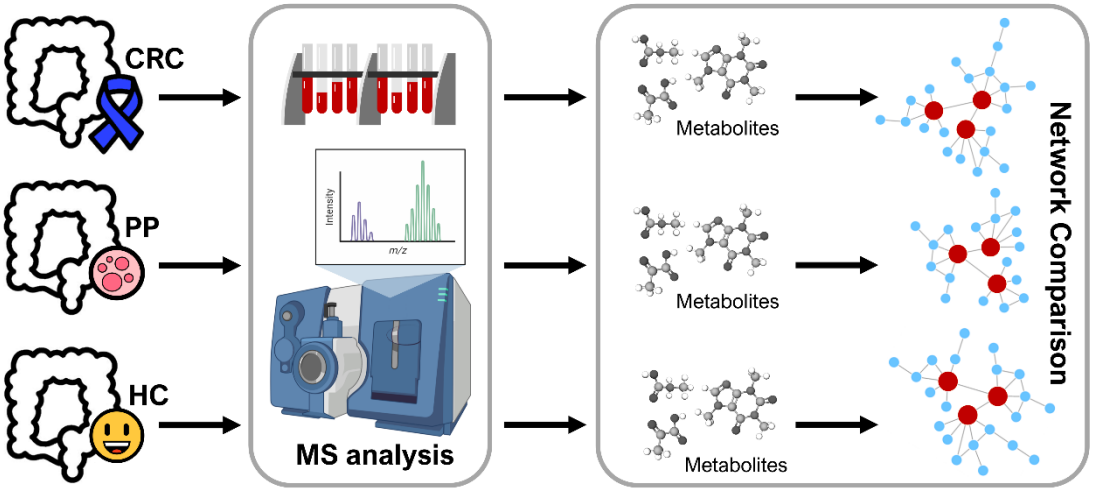


Figure 1. Study design.

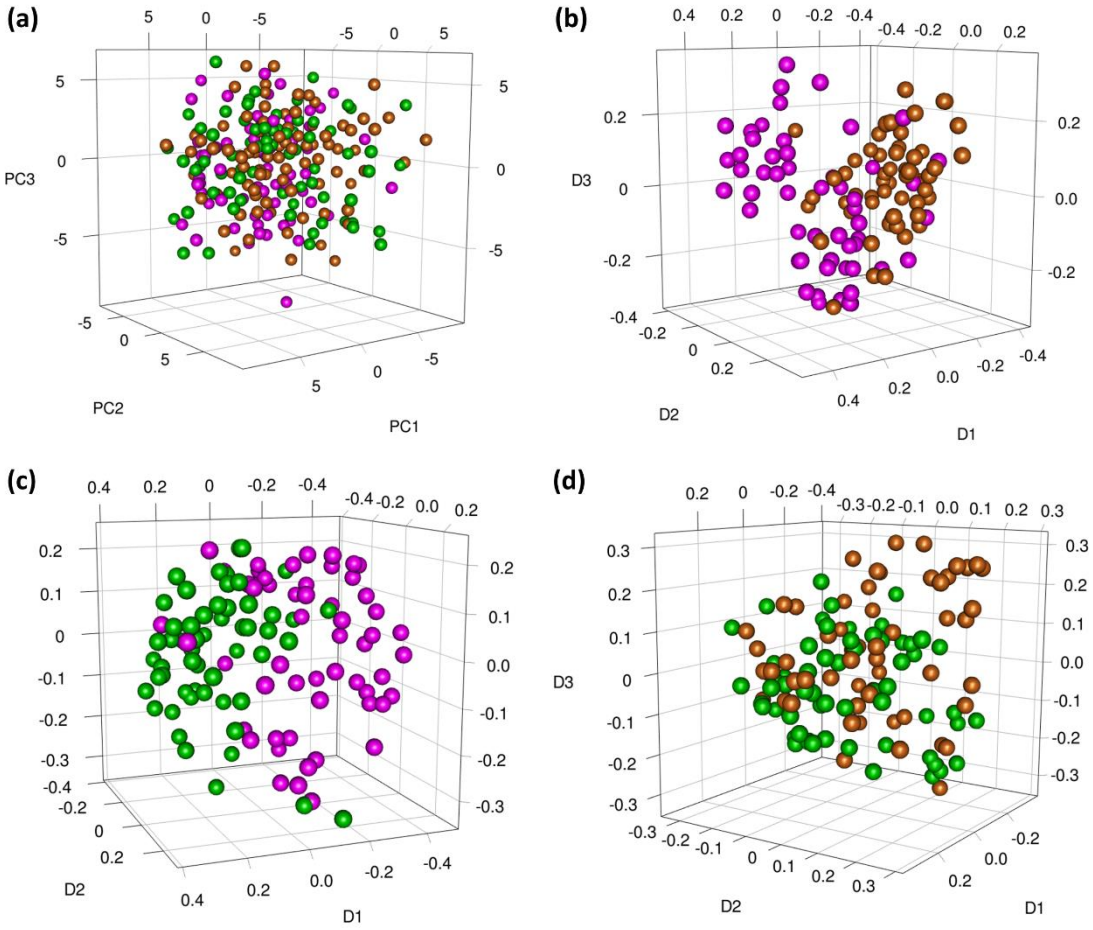


Figure 2. (a) Principal Component Analysis (PCA) model score plot (PC1 (13.4%) vs PC2 (8.0%) vs PC3 (7.3%)). Each dot represents a single metabolic profile colored by the different groups of patients: n=65 CRC patients (magenta dots), n=74 PP patients (green dots), and n=87 CTR patients (dark orange dots). Balanced Random Forest score plot and overall accuracy of the model discriminating: (b) n=55 randomly selected CRC (magenta dots) and n=55 randomly selected CTR patients (dark orange dots); (c) n =55 randomly selected CRC (magenta dots) and n=55 randomly selected PP patients (green dots); (d) n=62 randomly selected PP (green dots) and n=62 randomly

selected CTR patients (dark orange dots). 55 patients were randomly selected to guarantee unbiased RF predictive models.

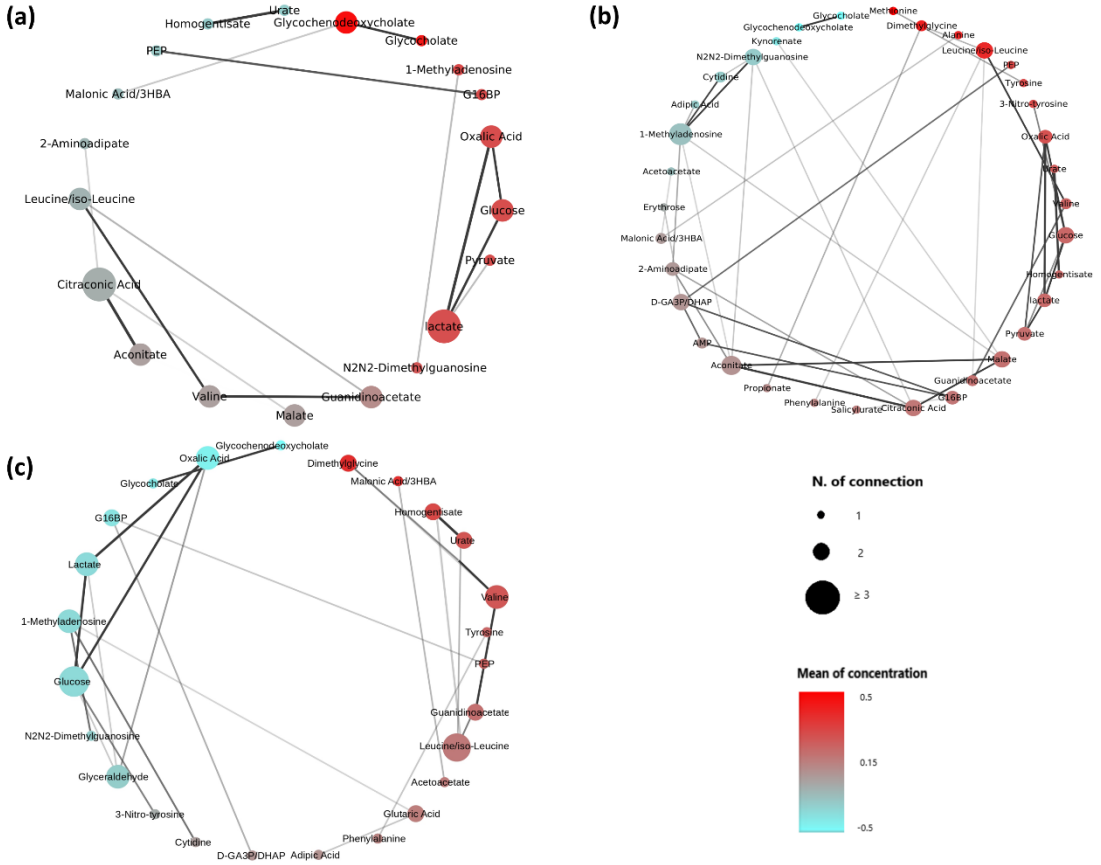


Figure 3. Metabolite-metabolite association network of (a) CRC patients; (b) PP patients; (c) CTR. Only the connections of statistically significant (Benjamini-Hochberg adjusted P-value ≤ 0.05) differentially connected metabolites are reported. Nodes are colored according to their mean of a normalized metabolite concentration (from light blue to red) and their dimension is proportional to the increasing metabolite-metabolite degree of connectivity. Edges represent correlation with $|r| \geq 0.6$ and their transparency depends on the likelihood of the metabolic connections.

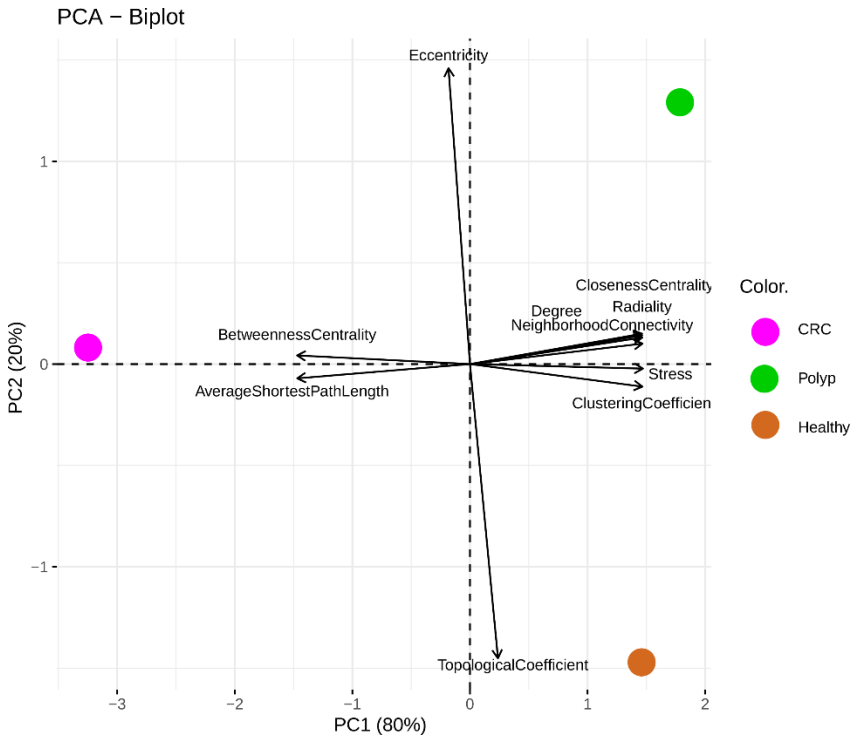


Figure 4. Principal component analysis (PCA) biplot performed on topological metabolite-metabolite association network parameter. PCA dots are colored according to the CRC (magenta dot), PP (green dot), and CTR (dark orange dot) group. PCA loadings represent the following network statistical parameters: Average Shortest Path Length, Betweenness Centrality, Closeness Centrality, Clustering Coefficient, Degree, Eccentricity, Neighborhood Connectivity, Radiality, Stress, and Topological Coefficient.

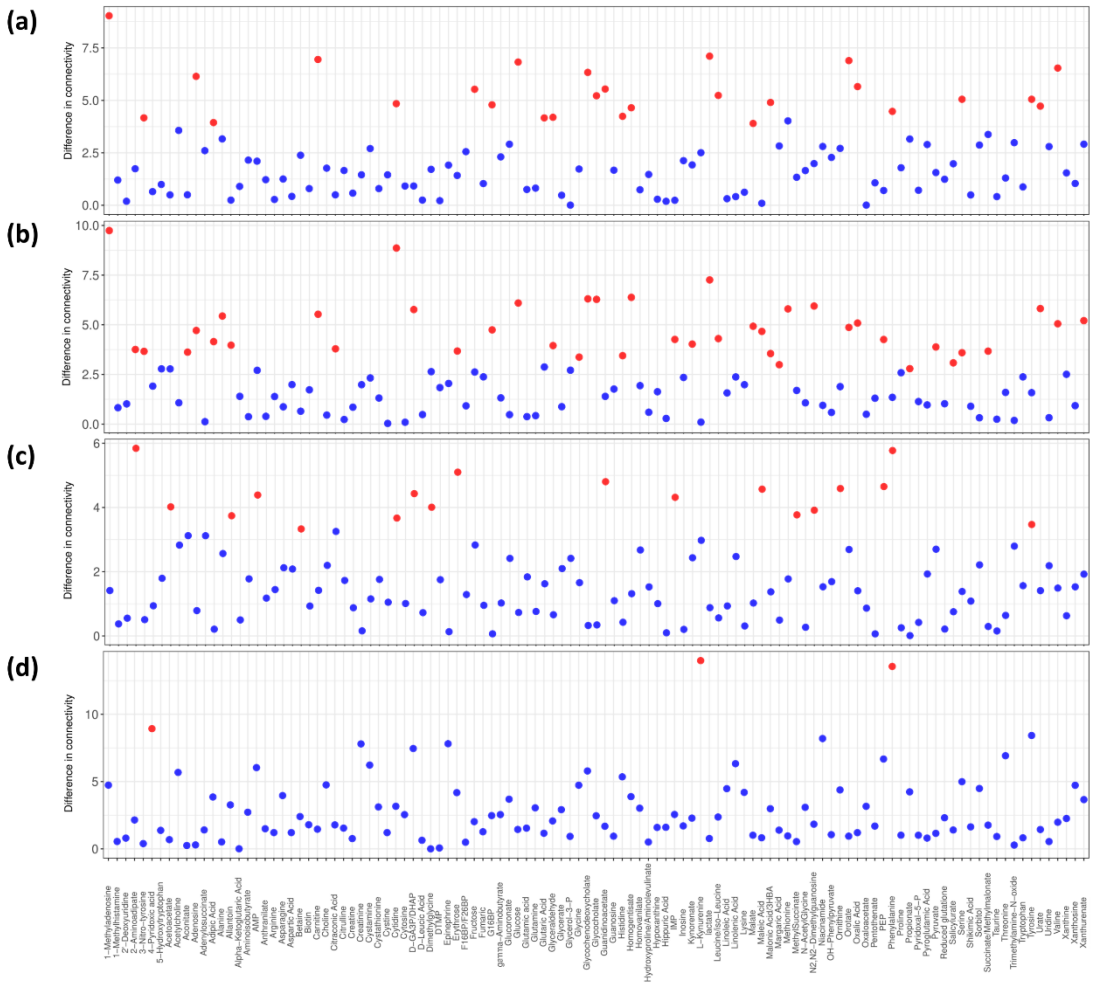
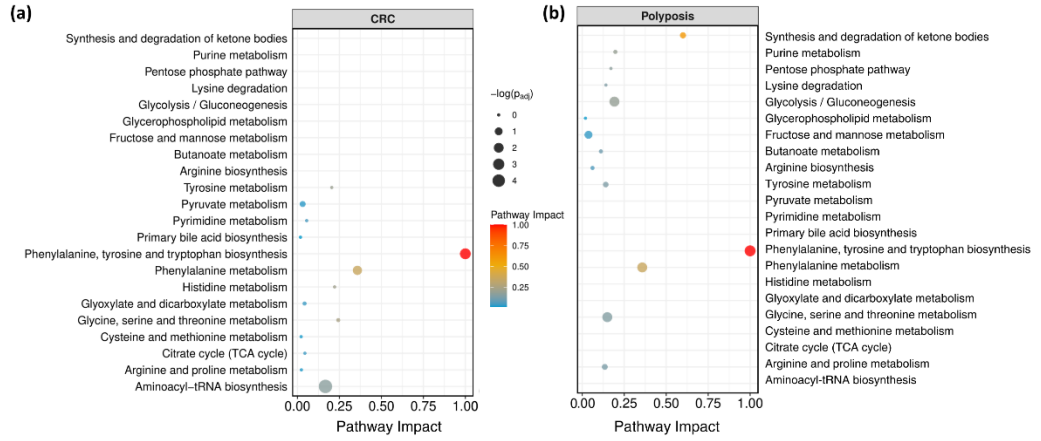


Figure 5. Differences in terms of connectivity in metabolite-metabolite association networks of (a) CRC and healthy control group, (b) CRC and PP group, (c) PP and CTR group, and (d) colon cancer sub-group and rectal cancer sub-group are reported against each specific node. Blue dots correspond to a no statistically significant metabolite and red dots correspond to a statistically significant metabolite. The threshold for significance of corrected P-values (using Benjamini-Hochberg correction) ≤ 0.05 was im-posed.



Pathways	Raw *P*	FDR	Impact
Phenylalanine, tyrosine and tryptophan biosynthesis	0.002	0.07	1
Primary bile acid biosynthesis	0.18	1	0.02
Cysteine and methionine metabolism	0.43	1	0.02
Arginine and proline metabolism	0.48	1	0.02
Pyruvate metabolism	0.05	0.71	0.03
Glyoxylate and dicarboxylate metabolism	0.10	0.97	0.04
Citrate cycle (TCA cycle)	0.29	1	0.04
Pyrimidine metabolism	0.14	1	0.06
Aminoacyl-tRNA biosynthesis	0.0001	0.008	0.17
Tyrosine metabolism	0.16	1	0.20
Histidine metabolism	0.24	1	0.22
Glycine, serine and threonine metabolism	0.10	0.97	0.24
Phenylalanine metabolism	0.01	0.19	0.36

Pathways	Raw *P*	FDR	Impact
Phenylalanine, tyrosine and tryptophan biosynthesis	0.0008	0.06	1
Synthesis and degradation of ketone bodies	0.06	0.68	0.60
Phenylalanine metabolism	0.005	0.12	0.36
Purine metabolism	0.17	0.97	0.20
Glycolysis / Gluconeogenesis	0.003	0.12	0.19
Pentose phosphate pathway	0.23	1	0.17
Glycine, serine and threonine metabolism	0.01	0.12	0.15
Lysine degradation	0.26	1	0.14
Tyrosine metabolism	0.08	0.78	0.14
Arginine and proline metabolism	0.07	0.74	0.13
Butanoate metabolism	0.16	0.97	0.11
Arginine biosynthesis	0.15	0.97	0.06
Fructose and mannose metabolism	0.02	0.36	0.04
Glycerophospholipid metabolism	0.35	1	0.02

Figure 6. Pathway analysis plot according to the statistically differential connected metabolites (Benjamini-Hochberg adjusted P-value ≤ 0.05) in (a) CRC compared with CTR, and (b) PP compared with CTR. Dots are colored according to the metabolic pathway impact score and their dimensions are proportional to their significance. Only enriched pathways with an impact score per cohort > 0.1 were considered. For each pathway, raw P-value, FDR adjusted P-value, and the impact score were also reported.

TABLES

Table 1: Statistically significant (FDR adjusted P-value < 0.05) metabolic differences obtained using univariate Student's t-test comparing CRC vs PP and CRC vs CTR). No significance (FDR adjusted P-value < 0.05) was observed comparing PP vs CTR.

Metabolite	CRC vs. PP	CRC vs. CTR
1-Methyladenosine	> 0.05	0.04
2'-Deoxyuridine	> 0.05	0.003
Adenylosuccinate	> 0.05	0.03
Alanine	0.03	> 0.05
Allantoin	> 0.05	0.005
Alpha-Ketoglutaric Acid	0.01	0.01
Aspartic Acid	0.01	0.01
Biotin	0.01	> 0.05
Cystathionine	> 0.05	0.01
Dimethylglycine	0.002	0.01
gama-Aminobutyrate	0.02	> 0.05
Glutamic acid	0.03	0.03
Glutamine	0.001	0.003
Glyceraldehyde	0.0003	0.002
Glycochenodeoxycholate	0.004	0.01
Glycocholate	0.01	0.02
Histidine	0.0002	0.00002
Hydroxyproline/Aminolevulinate	> 0.05	0.04
Hyppuric Acid	0.0009	0.004
Kynorenate	> 0.05	0.01
Linoleic Acid	0.01	> 0.05
Linolenic Acid	0.0004	0.002
Lysine	0.00007	0.0005
Maleic Acid	> 0.05	0.01
Margaric Acid	0.04	> 0.05
Methionine	0.00007	0.0005
N-AcetylGlycine	0.01	0.03
Oxalic Acid	> 0.05	0.04
PEP	0.002	0.04
Urate	0.04	> 0.05

Uridine	0.0008	> 0.05
Xanthurenate	0.02	> 0.05

Table 2: Mean values of Accuracy. Specificity. Sensitivity. and Area under the curve (AUC) of RF models built comparing CRC and CTR patients. CRC and PP patients. and PP and CTR patients.

	Mean Accuracy % (95% CI)	**Mean Specificity % (95% CI)**	**Mean Sensitivity % (95% CI)**	**AUC (95% CI)**
CRC vs CTR groups	78.2 (77.8 – 78.5)	79.0 (78.6 – 79.3)	77.1 (76.6 – 77.7)	0.875 (0.873 – 0.877)
CRC vs PP groups	79.5 (79.2 – 79.9)	77.7 (77.1 – 78.3)	81.2 (70.8 – 81.7)	0.871 (0.869 – 0.872)
PP vs CTR groups	62.2 (61.8 – 62.6)	63.1 (62.3 – 63.8)	61.5 (60.9 – 63.8)	0.661 (0.658 – 0.665)

SUPPLEMENTARY METHODS

1. Reconstruction of metabolite association network

The Probabilistic Context Likelihood of Relatedness based on Correlation (PCLRC)[1] algorithm was used to infer metabolite-metabolite association networks. In order to remove non-significant background correlations, this algorithm provides a robust evaluation of the correlation using a resampling strategy in combination with the previously published Context Likelihood of Relatedness (CLR)[2] approach. The PCLRC algorithm gives like output a probability matrix **P** showing the likelihood p_{ij} for each revealed Spearman correlation r_{ij} between two metabolites *i* and *j*. We considered correlation for which the probabilistic value p_{ij} was > 97% and we set to 0 all remaining correlations:

$$r_{ij} = \begin{cases} r_{ij} & if\, p_{ij} \geq 0.97 \\ 0 & if\, p_{ij} < 0.97 \end{cases} \quad (2)$$

The 0.97 probability threshold was chosen as the best compromise between network complexity (*i.e.* number of nodes) and interpretability. PCRLC was used with default parameters, with 1000 resampling iterations, and 75% of samples were kept in each iteration and the top 30% correlation retained for each iteration.

2.6.1. Differential connectivity analysis

Given a specific network *a*, the connectivity χ_i^a for each metabolite *i* is described as:

$$\chi_i^a = \left(\sum_{j=1}^{J} |r_{ij}| \right) - 1 \quad (3)$$

Moreover, the differential connectivity $\Delta_i^{a,b}$ for each metabolite *i* among two networks *a* and *b* is calculated as follows:

$$\Delta_i^{a,b} = \chi_i^a - \chi_i^b \quad (4)$$

The statistical significance of the differentially connected metabolites was determined by means of a permutation-test. In order to eliminate the relationship between variables and in order to maintain their variance, the columns of each input matrix were independently permuted defining a permutated matrix $\mathbf{X}_{(k)}$. The overall network estimation is performed on permutated data matrix, generating the related Spearman correlation $R_{(k)}$ analysis. These estimations were subsequentially used to assess, for each metabolite contained in the permutated matrix $\mathbf{X}_{(k)}$, the permutated connectivity (Equation (4)), and the permutated differential connectivity (Equation (5)):

$$\chi_{i,k}^{a} = \left(\sum_{j=1}^{J} \left| r_{ij}^{k} \right| \right) - 1 \quad (5)$$

$$\Delta_{i,k}^{a,b} = \chi_{i,k}^{a} - \chi_{i,k}^{b} \quad (6)$$

The permutation step was repeated k = 1000 yielding a null distribution D_i of permutated differential connectivity values. The significance of a given differential connectivity value $\Delta_i^{a,b}$ (estimated from the non-permutated original data) was calculated as a *P*-value, according to the following formula:

$$P - value = \frac{1 + \left(|D_i| > \left| \Delta_i^{a,b} \right| \right)}{k} \quad (7)$$

All *P*-values were corrected for multiple test comparisons using the Benjamini-Hochberg approach[3].

2. Measures for network topology

Network topology and node (metabolite) characteristics were evaluated using several standard metrics in addition to connectivity (node degree). We used *Average Shortest Path Length*, *Betweenness Centrality*, *Closeness Centrality*, *Clustering Coefficient*, *Degree*, *Eccentricity*, *Neighborhood Connectivity*, *Radiality*, *Stress*, and *Topological Coefficient*. These measures were calculated using Network Analyzer[4], a Java plugin available for the Cytoscape platform[5]. A brief overview of the measure is given in the following.

The *Average Shortest Path Length* [6] , also known as the characteristic path length, indicates the expected distance between two connected nodes. The shortest path length is considered the shortest distance between two nodes *i* and *j*, denoted by L(*i*,*j*). The shortest path length distribution gives the number of node pairs (*i*,*j*) with L(*i*,*j*). = *k* for *k* = 1,2, …, and it indicates small-word properties of a network.

The *Eccentricity* of a metabolite *a* is the maximum noninfinite length of the shortest path between a metabolite *i* and another metabolite in the network.

The *Betweenness Centrality* Bc_i [7] of a node *i* is calculated as follows:

$$Bc_i = \sum_{s \neq i \neq j,} \frac{\sigma_{sij}}{\sigma_{sj}} \quad (8)$$

where *s* and *j* are nodes in the network different from node *i*, σ_{sj} denotes the number of shortest paths from nodes *s* and *j*, and σ_{sij} is considered the number of shortest paths from s to j passing through node *i*. The Betweenness Centrality could be normalized by dividing the number of node pairs excluding *i*:

$$\frac{Bc_i}{\frac{(N-1)(N-2)}{2}} \quad (9)$$

where N is the total number of nodes in the connected component that *i* belongs to. This parameter reflects the degree of control that given node exercises over the interactions of the other nodes in the network.

The *Closeness Centrality* Cc_i [8] of a node *i* is the reciprocal of the average shortest path length and it is a number between 0 and 1. Cc_i measures the distance of a node *i* to all other nodes and it is calculated as follows:

$$Cc_i = \frac{1}{\frac{\sum_j d_{ij}}{N-1}} \tag{10}$$

where d_{ij} is the distance between node *i* and *j*.

In the undirected networks, the *Clustering* C_i[6,9,10] of a node *i* is defined as a ratio $\frac{N}{M}$, where N is the number of edges between the neighbors of *i*, and M is the maximum number of edges that could possibly exist between the neighbors of *i*. This coefficient is always a number between 0 and 1. The network clustering coefficient is the average of the clustering coefficients of all nodes in the network and it is used to highlight a modular organization of metabolic networks[10].

The *Neighborhood Connectivity*[11] of node *i* is the average connectivity of all neighbors of *i*.

The *Radiality*[9] is a node centrality index and gives high centralities to nodes that have a small distance to any other node in their reachable neighborhood compared to their diameter.
The *Stress*[12] of a node *i* is the number of shortest paths passing through *i*.
The *Topological Coefficient* T_i[13] of a node *i* with k_n as the number of neighbors of node *i* is calculated as follows:

$$T_i = \frac{average\big(J(i,j)\big)}{k_i} \tag{11}$$

where the value $J(i,j)$ is the number of neighbors shared between the nodes *i* and *j*. This coefficient indicates the tendency of the nodes in the network to have shared neighbors.
In order to compare the metabolite-metabolite association networks topology, we evaluate the average of node characteristics and connectivity among metabolite *i* and *j*.

REFERENCES:

1. Suarez-Diez M, Saccenti E. Effects of Sample Size and Dimensionality on the Performance of Four Algorithms for Inference of Association Networks in Metabonomics. *J Proteome Res*. 2015;14(12):5119-5130. doi:10.1021/acs.jproteome.5b00344

2. Akhand MAH, Nandi RN, Amran SM, Murase K. Context likelihood of relatedness with maximal information coefficient for Gene Regulatory Network inference. In: *2015 18th International Conference on Computer and Information Technology (ICCIT)*. ; 2015:312-316. doi:10.1109/ICCITechn.2015.7488088
3. 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;57(1):289-300.
4. Doncheva NT, Assenov Y, Domingues FS, Albrecht M. Topological analysis and interactive visualization of biological networks and protein structures. *Nature Protocols*. 2012;7(4):670-685. doi:10.1038/nprot.2012.004
5. Shannon P, Markiel A, Ozier O, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res*. 2003;13(11):2498-2504. doi:10.1101/gr.1239303
6. Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. *Nature*. 1998;393(6684):440-442. doi:10.1038/30918
7. Yoon J, Blumer A, Lee K. An algorithm for modularity analysis of directed and weighted biological networks based on edge-betweenness centrality. *Bioinformatics*. 2006;22(24):3106-3108. doi:10.1093/bioinformatics/btl533
8. Freeman LC. Centrality in social networks conceptual clarification. *Social Networks*. 1978;1(3):215-239. doi:10.1016/0378-8733(78)90021-7
9. Barabási A-L, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*. 2004;5(2):101-113. doi:10.1038/nrg1272
10. Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabási A-L. Hierarchical Organization of Modularity in Metabolic Networks. *Science*. 2002;297(5586):1551-1555. doi:10.1126/science.1073374
11. Maslov S, Sneppen K. Specificity and stability in topology of protein networks. *Science*. 2002;296(5569):910-913. doi:10.1126/science.1065103
12. Brandes U. A faster algorithm for betweenness centrality. *The Journal of Mathematical Sociology*. 2001;25(2):163-177. doi:10.1080/0022250X.2001.9990249
13. Stelzl U, Worm U, Lalowski M, et al. A Human Protein-Protein Interaction Network: A Resource for Annotating the Proteome. *Cell*. 2005;122(6):957-968. doi:10.1016/j.cell.2005.08.029

4.1.8 NMR-based metabolomics to evaluate individual response to treatments

Alessia Vignoli[1,2,#], Gaia Meoni[1,2,#], Veronica Ghini[1,2], Francesca Di Cesare[1,2], Leonardo Tenori[1,2,3], Claudio Luchinat[1,2,3], Paola Turano[1,2,3,*]

[#] These authors contributed equally.

[1] Magnetic Resonance Center (CERM), University of Florence, 50019, Sesto Fiorentino, Italy;

[2] Department of Chemistry "Ugo Schiff", University of Florence, 50019, Sesto Fiorentino, Italy;

[3] Consorzio Interuniversitario Risonanze Magnetiche MetalloProteine (CIRMMP), 50019, Sesto Fiorentino, Italy

Accepted for publication

Candidate's contribuitions: figure and writing manuscript.

Abstract

The aim of this chapter is to highlight the various aspects of metabolomics in relation to health and diseases, starting from the definition of metabolic space and of how individuals tend to maintain their own position in this space. Physio-pathological stimuli may cause individuals to lose their position and then regain it, or move irreversibly to other positions. By way of examples, mostly selected from our own work using ^{1}H NMR on biological fluids, we describe the effects on the individual metabolomic fingerprint of mild external interventions, such as diet or probiotic administration. Then we move to pathologies (such as celiac disease, various types of cancer, viral infections, and other diseases), each characterized by a well-defined metabolomic fingerprint. We describe the effects of drugs on the disease fingerprint and on its reversal to a healthy metabolomic status. Drug toxicity can be also monitored by metabolomics. We also show how the individual metabolomic fingerprint at the onset of disease may discriminate responders from non-responders to a given drug, or how it may be prognostic of e.g., cancer recurrence after many years. In parallel with fingerprinting, profiling (i.e., the identification and quantification of many metabolites and, in the case of selected biofluids, of the lipoprotein components that contribute to the ^{1}H NMR spectral features) can provide hints on the metabolic pathways that are altered by disease and assess their restoration after treatment.

1 INTRODUCTION

The use of metabolomics in personalized medicine originates from two basic facts

i) the existence of an individual metabolic phenotype characteristic for each individual, *i.e.* of an invariant part of the metabolome that allows each subject to be discriminated from all others;

ii) the existence of a signature of the disease, which may range from very weak to very strong depending on the pathology or its severity. The disease transiently alters the individual metabolic phenotype, but this alteration disappears when the individual reverts to the "healthy" status following medical or pharmaceutical interventions.

Despite ^{1}H NMR features a lower sensitivity (detection limit in the order of μM) with respect to MS analysis, which downsizes the total number of measurable metabolites, it is exquisitely amenable to the **untargeted fingerprinting** of the sample metabolome due to its very high reproducibility and capability for high throughput analysis (Griffiths, 2008; Takis et al., 2019; Vignoli et al., 2019a).

With the exception of inorganic ions, almost all small molecules contain hydrogen atoms that can be measured simultaneously, providing rapid and distinct global spectral patterns (or fingerprints) of the samples under investigation (Klupczyńska et al., 2015; Vignoli et al., 2019a). Thus, the fingerprint of a sample is a recapitulation of its current metabolome, independently of the identification of the metabolites. Conversely, **metabolomic profiling** is a global evaluation of the metabolite contents of all the samples in a comparative fashion. The final aim is to identify and accurately quantify as many compounds as possible. This approach enables the detection of changes in the concentration of the measurable metabolites related to specific physiological conditions.

Here, we first provide a series of examples demonstrating how untargeted ^{1}H NMR metabolomic fingerprinting can be used to identify the individual metabolic phenotype and assess its stability over a long timescale in the absence of important physio-pathological alterations. The existence of this individual fingerprint has been observed in several biosamples (such as urine, saliva, blood, and breath condensate) that are commonly used in metabolomics. Then we provide a series of examples spanning from celiac disease to cancer and viral infections. These pathologies are very different in their etiology and clinical manifestation, and involve different organs. Likewise, in our lab we have observed metabolomic alterations that are clearly characteristic of each of them. Additionally, we have

monitored the response to pharmaceutical treatments, and in some cases we have been able to identify features of the individual metabolome before treatment that are prognostics to discriminate responders from non-responders.

To help readers assess the potentiality of the NMR-based approach, we describe a series of case studies drawn on the many years of experience of our group.

1.1 The individual metabolic phenotype

In 2008 our research group (Assfalg et al., 2008) demonstrated, for the first time, that the NMR detectable part of the metabolome of urine contains an invariant part, which can be considered as the chemical signature of each individual. Multivariate statistical analysis of multiple urine samples of different individuals enables the definition of their "metabolic space", where the metabolomic fingerprint of each subject can be visualized and discriminated from that of the other subjects with an accuracy close to 100%. Over a timescale of 10 years (Ghini et al., 2015a), in the absence of important stressful perturbations, each individual still occupies its specific region of the metabolic space; daily intra-individual variability leads to small fluctuations inside the individual metabolic space; conversely, shifts to other distinct regions are associated to significant changes of the individual metabolomic phenotype as a consequence of the occurrence of important physio-pathological conditions (Figure 1).

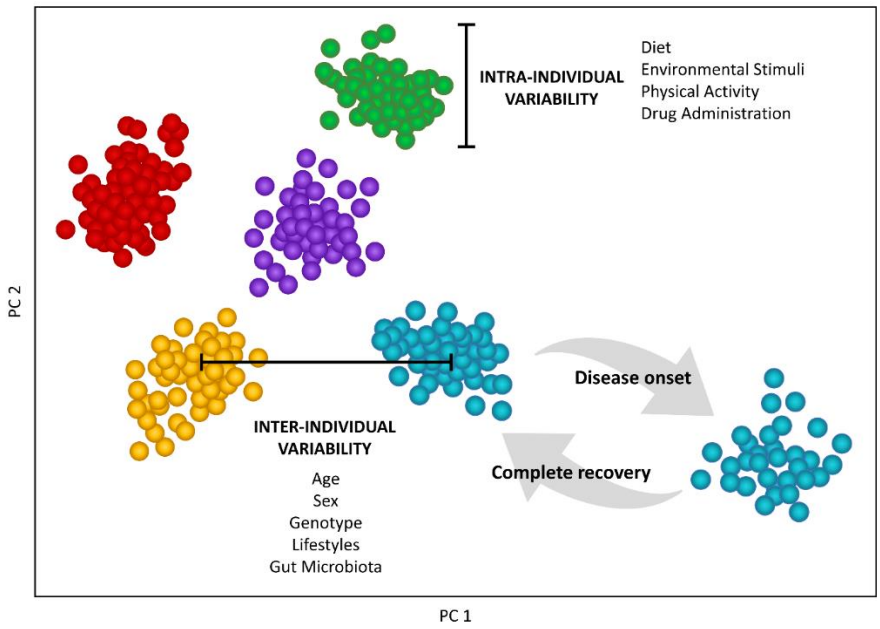


Fig. 1 A schematic representation of the individual metabolic phenotype. Each individual (represented by a different color code in the picture) has a stable phenotype, which distinguishes him/her from the other subjects (inter-individual variability). This phenotype is flexible enough to compensate for day-by-day changes and external

stimuli (intra-individual variability). This capacity is lost at the onset of a disease but can be restored after a complete recovery.

A strong individual phenotype also exists in saliva (Wallner-Liebmann et al., 2016) and blood (Holmes et al., 2008). For the latter two biofluids, their intrinsic nature causes daily intra-individual fluctuations in metabolome composition that are smaller than what occurs in urine; this is particularly true for blood, the composition of which is tightly regulated by homeostasis. Additionally, at variance with urine and saliva that have individual metabolomic fingerprints dominated by low-molecular weight molecules, in both plasma and serum the presence of lipids and lipoproteins significantly contributes to the individual fingerprint.

For serum/plasma stability of the individual phenotype has been reported over a period of 7 years (but is likely to be much longer), while no information about the long-term stability of the metabolic phenotype in the saliva is yet available.

1.2 Modulation of the individual metabolic phenotype

Besides the genetic components, all endogenous and exogenous metabolites derived from extrinsic factors such as diet, drugs, gut microflora, stressors, and pollutants contribute to the definition of the human metabolic phenotype. Thus, several studies have been performed to understand the effects of mild dietary interventions on the metabolomic profile (Andersen et al., 2014; O'Sullivan et al., 2011). Importantly, despite the levels of a large set of metabolites have been reported to be modulated by the different dietary habits and to become more homogeneous in individuals undergoing diet standardization for a short period of time, diet regimes do not significantly change the individual signatures, which remain equally evident in all types of biofluids (Lenz et al., 2003; Marianne C Walsh, 2006; Wallner-Liebmann et al., 2015; Winnike et al., 2009). The same effects are reported for probiotics assumption in 22 healthy volunteers for 8 weeks (Ghini et al., 2020b). Because of the probiotic treatment, significant modulations in the levels of a few urine and serum metabolites are observed, but the observed changes are not so strong as to hamper individual recognition in the metabolic space. Another example is represented by the evaluation of the metabolomic effects induced by the administration of bioactive molecules in the sera profiles of volunteers at risk of metabolic syndrome. This study was performed in the context of the EC-funded project PATHWAY-27 (FP-KBBE # 311876). In this framework, beneficial effects of DHA administration on serum metabolomic profiles could be assessed. However, this was only possible by suppressing the intra-individual variations using paired approaches that compare

the serum paired samples collected at the beginning and at the end of the trial for each subject (Ghini et al., 2017).

The above examples demonstrate that little or no alteration of the individual phenotype is brought about by mild treatments, whereas it can be profoundly affected by the presence of pathologies and major surgical interventions, such as bariatric surgery. Currently, bariatric surgery is the only available treatment to provide sustained weight loss (Buchwald et al., 2009), improve glucose regulation, and to even promote complete remission of type 2 diabetes in severely obese patients (Meijer et al., 2011). All three different bariatric procedures, i.e. sleeve gastrectomy, proximal Roux-en-Y gastric bypass, and distal bypass, are associated with a strong alteration of the serum metabolomic fingerprints of the patients (Gralka et al., 2015). Within the common strong alterations, distal bypass patients could be discriminated from the other two groups of patients, suggesting a stronger impact of this procedure on the metabolic fingerprint. Different short term and long-term alterations are also observed. The serum metabolomic profiles of severe obese patients are characterized by high levels of aromatic and branched-chain amino acids (AAA and BCAA, respectively), and of metabolites related to energy metabolism (pyruvate and citrate). Elevated levels of some metabolites related to the gut microbiota such as formate, methanol and isopropanol are also associated with obesity. Interestingly, after bariatric surgery, independently of the type of procedure used, a significant reduction of AAA, BCAA, pyruvate, methanol and isopropanol concentrations, along with an increase in arginine and glutamine levels are observed, indicating that surgically induced weight loss can, at least in part, normalize the alterations in amino acid metabolism associated with obesity (Gralka et al., 2015).

2 Disease fingerprinting and individual response to pharmacological therapy

As anticipated, several pathologies have a strong metabolomic fingerprint at the systemic level. Below we provide some examples from our lab of the metabolomic characterization of different types of diseases. In these examples, NMR-based metabolomics successfully provided a medium-to-strong fingerprint of the disease and a definition of the molecular profile of the pathology in terms of small molecules and lipoprotein parameters. Additionally, NMR was used to monitor the evolution of the metabolomic fingerprint and profile along the disease evolution and in particular following pharmacological treatments. Indeed, pharmacometabolomics is defined as the determination of the individual metabolic phenotype to characterize signatures, both before and after drug administration, that might inform treatment outcomes, map the effects of drugs on metabolism and identify molecular pathways contributing to drug-response and drug-toxicity phenotypes (Kaddurah-Daouk et al., 2014).

2.1 Celiac disease

The celiac disease (CD) originates from an aberrant adaptive immune response against gluten. The presence of a characteristic metabolomic fingerprint of CD has been demonstrated by multivariate analysis of the NMR spectra of serum and urine samples from affected subjects and healthy controls (Bernini et al., 2011a; Bertini et al., 2009; Vignoli et al., 2019b). For example, when comparing adult patients with sex and age matched controls, a discrimination accuracy in the range of 75-83% was obtained in urine, which rises to 81-94% in serum, depending on the study cohorts. The differences originate from variations in the levels of metabolites related to three main mechanisms: malabsorption, altered energy metabolism, and altered gut microflora (Bertini et al., 2009).

In serum samples, the main reported differences between CD patients and controls consist in lower levels of several amino acids (asparagine, isoleucine, methionine, proline, and valine) and of pyruvate, lactate, and lipids, and in higher levels of glucose and 3-hydroxybutyric acid. The decreased levels of pyruvate and lactate along with higher levels of glucose are indicative of an impaired glycolysis process. Enhanced beta-oxidation and malabsorption can instead explain the lower levels of amino acids and lipids observed. A possible increase of the use of ketonic bodies as a source of energy in coeliac patients is consistent with the high levels of 3-hydroxybutyric acid. Energy conversion from lipids and catabolism of ketonic bodies is far less efficient than that from glucids; consistently, coeliac subjects often report symptoms of fatigue. Further, urine samples of CD patients are characterized by high levels of some metabolites related to gut microbiota, *i.e.* indoxyl sulphate, meta-[hydroxyphenyl] propionic acid, phenylacetylglycine and p-cresolsulfate. All these findings are consistent with the hypothesis that in CD patients the gut microflora of the small bowel is altered or presents peculiar microbial species with their own metabolome.

Notably, the so-called potential coeliac patients i.e., those individuals who tested positive for the antibodies but have no evidence of intestinal damage, have an NMR fingerprint that is similar to that of the patients, suggesting that the dysmetabolism precedes the intestinal damage (Bernini et al., 2011a).

Celiac disease represents one example where dietary intervention induces relevant changes in the metabolomic fingerprint (Bernini et al., 2011a). The treatment of choice for this disorder is to follow a gluten-free diet. One of the most interesting findings was that the metabolic fingerprint of CD patients reverts to normality after 12 months of a strict gluten-free diet. After

the gluten-free diet normal levels of glucose and 3-hydroxy-butyric acids are restored, and the sense of fatigue tends to be reduced.

2.2 Cancer

Metabolomics can be used to derive information on cancer at various levels. Several studies concern the characterization of the metabolome of cancer cell lines as a complementary approach with respect to classical biochemical analyses and to other omics, with the aim to derive information on altered metabolic pathways (Cuperlovic-Culf et al., 2012; Li et al., 2017) and to conduct preclinical tests of new anticancer agents (D'Alessandro et al., 2019; Ghini et al., 2021; Li et al., 2019; Resendiz-Acevedo et al., 2021). NMR is particularly suitable for this purpose, as it allows for a fast untargeted characterization of the endo and exometabolome (i.e., the intracellular and extracellular composition in terms of metabolites), also at multiple post-drug treatment times. Additionally, isotopically labelled substrates (e.g., ^{13}C labelled glucose) can be used to define metabolic fluxes by following the metabolism of labelled substrates into their pathway products at specific time points (Antoniewicz, 2018; Saborano et al., 2019).

Tumors can also be characterized at the level of the tissue metabolome; this approach provides a specific view of cancer cells and their cross-talk with the tumor microenvironment (Márquez and Matés, 2021). Cancer tissues represent the localized site of the disease. ^{1}H NMR, in its HR-MAS version, is applicable to acquire spectra of small quantities of cryo-preserved tissues (of the order of 10 mg), provided they have not undergone any further sample manipulation (Vignoli et al., 2019a). Indeed, the use of tissues for ex-vivo metabolomics raises a number of criticalities, ranging from ethical to technical issues. Surgical specimens collected at different times of intraoperative (warm) or post-operative (cold) ischemia undergo significant molecular degradation, so that the measured "metabolome" no longer reflects the original physiological state of the tissues before intervention (Cacciatore et al., 2013). The need of suitable frozen material collected and handled under strict preanalytical conditions conflicts with ethics that assigns precedence, during intraoperative procedures, to the best surgical performances and afterwards to the diagnostic needs of the pathologist.

Instead, the metabolomic analysis of biofluids, such as serum or plasma, or urine, presents several advantages. Sample collection is only minimally invasive, and multiple collection at different times points can be easily obtained to establish a metabolomic signature both before and after a given drug therapy. The preanalytical SOPs are simple and easily satisfied (Ghini et al., 2019; "ISO/DIS 23118," 2021; Vignoli et al., 2022). The resulting analyses might inform

on the presence of specific features of prognostic value and on treatment outcomes, respectively.

Below we provide examples that have seen a significant contribution from our group and that regard the characterization of three of the most common types of cancer, namely breast cancer, colorectal cancer, and lung cancer; all of them are based on patients' serum/plasma (Figure 2).

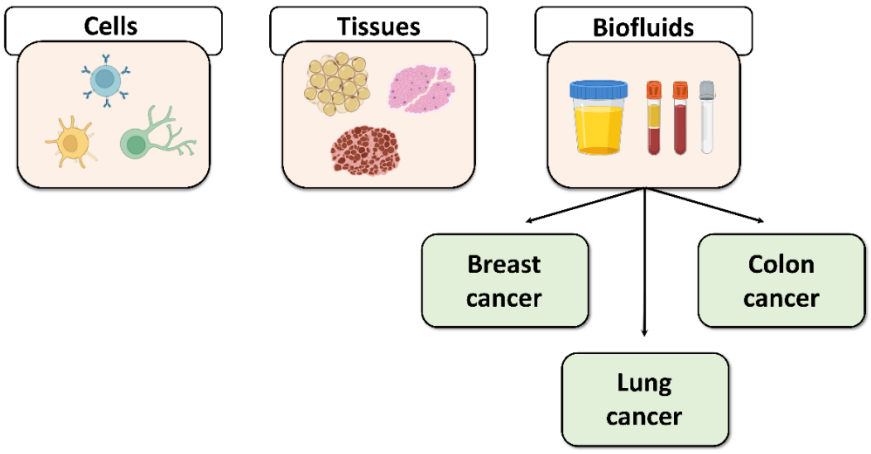


Fig. 2 Types of samples that can be analyzed via NMR-based metabolomics: cells, tissues, and biofluids. In particular, in section 2.2 we will focus on the analysis of biofluids (i.e. serum, plasma, and urine) to investigate pathophysiological conditions associated with breast cancer, lung cancer, and colon cancer.

2.2.1 Breast cancer

Female breast cancer (BC) is the most commonly diagnosed cancer, with an estimated 2.3 million new cases per year (11.7%), and is the leading cause of cancer death in women (Sung et al., 2021). For many years BC has been considered a unique clinical entity and treated with only one general approach. However, it has become extremely clear that there is a high degree of diversity between and within the BC subclasses (Polyak, 2011). Based on their molecular characteristics BC is classified into three major subtypes: luminal (positive for estrogen and progesterone receptors), human epidermal growth factor receptor 2+ (HER2+) enriched, and basal (the majority of the latter tumors are also called triple-negative BC). Currently, these different tumor subtypes are treated with specific therapies improving patient survival: hormone receptor-positive disease is generally treated with endocrine therapy (Early Breast Cancer Trialists' Collaborative Group (EBCTCG), 2005; Lerner et al., 1976; Wiggans et al., 1979); HER2+ disease is now cured with targeted agents, which have significantly improved outcomes in both the (neo)adjuvant (Gianni et al., 2012; Martin et al., 2017; Slamon et al.,

2011; von Minckwitz et al., 2019, 2017) and the metastatic setting (Swain et al., 2013; Verma et al., 2012).

Metabolomics can distinguish patients with BC with respect to healthy controls with high predictive accuracies: PLS models with $Q^2 > 0.6$ for studies on serum/plasma (Cala et al., 2018; Lécuyer et al., 2018; Silva et al., 2019; Singh et al., 2017; Sitter et al., 2006; Slupsky et al., 2010; Suman et al., 2018; Tayyari et al., 2018; Wojtowicz et al., 2020; Zhou et al., 2017), and models with $Q^2 > 0.5$ for urinary studies (Silva et al., 2019; Slupsky et al., 2010; Zhou et al., 2017) were obtained.

Metabolomics proved useful in discriminating the plasma profiles of patients with different BC molecular subtypes as compared to controls (Díaz-Beltrán et al., 2021): as compared to HER2- patients, the HER2+ group showed elevated aerobic glycolysis, gluconeogenesis, and increased fatty acid biosynthesis with reduced Krebs cycle. ER+ patients, as compared to ER- ones, showed elevated alanine, aspartate and glutamate metabolism, decreased glycerol-lipid catabolism, and enhanced purine metabolism (Fan et al., 2016).

Given these premises, in the framework of BC precision oncology, it is crucial to identify, in each BC subtype, patients at higher risk of cancer recurrence and drug-response profiles able to guide patients' management (Vignoli et al., 2021b; Bendinelli et al., 2021). The NMR-based metabolomics has shown to be a valuable prognostic instrument and examples of its application with the abovementioned objectives are provided below.

2.2.1.1 Response to Chemotherapy

The possible association between different metabolic profiles and response to chemotherapy has been extensively investigated via NMR-based metabolomics of blood derivatives.

Pharmacometabolomics on blood plasma/serum with the aim of predicting which patients will benefit most from a specific treatment has provided significant results. First in 2012, our group demonstrated that NMR-based metabolomics may play a role in identifying patients with metastatic breast cancer (MBC) with HER2+ disease with greater sensitivity to paclitaxel plus lapatinib (Tenori et al., 2012). With a similar approach, Jiang and colleagues predict response to gemcitabine/carboplatin chemotherapy via NMR metabolomics (Jiang et al., 2018).

More recently, the general research attention has focused on neo-adjuvant chemotherapy (NAC), the preferred treatment strategy for patients with large primary tumors and/or locally advanced disease (Thompson and Moulder-Thompson, 2012). NAC offers the unquestionable benefit of downstaging disease and reducing the tumor size prior to surgery, thus making patients with inoperable tumors candidates for surgical resection (Debik et al., 2019;

Thompson and Moulder-Thompson, 2012). Nonetheless, only less than 30% of patients overall exhibit pathological complete response (disappearance of all invasive cancer in the breast) to NAC (Wei et al., 2013), with lower rates of response observed in endocrine receptor-positive, HER2- disease.

In 2020 our research group, for the first time, has investigated the capability of predicting pathological complete response, using the baseline host immune cytokines and nuclear magnetic resonance (^{1}H NMR) metabolomic fingerprints in a highly homogeneous population of HER2+ BC patients enrolled for NAC treatment (Vignoli et al., 2020a). For this study 43 HER2+ BC patients, stratified according to their ER status in 22 ER+ and 21 ER-, were enrolled at baseline prior to any interventions. The pathological complete response was obtained in 13 out of 22 ER+ patients and in 11 out of 21 ER- patients. The plasma metabolomic fingerprint can distinguish ER+ and ER- patients with 74.4% discrimination accuracy, suggesting a differential host–cancer interaction in these two subtypes of BC. Moreover, in the ER+ BC patients the baseline metabolomic fingerprint can be predictive of pathological complete response (72.7% accuracy). The good responders, as compared to poor responders, are characterized by lower concentrations of branched-chain amino acids, isoleucine, and valine, as well as ethanol, several phospholipids and cholesterol associated to almost all classes of lipoproteins assigned by ^{1}H NMR. In the ER+ subgroup the combination of a cytokine (TNF-α) and a metabolite (valine) was found to significantly enhance discrimination between complete and partial response to NAC, yielding an area under the receiver operating characteristic (ROC) curve of 0.92, and an accuracy of 90.9%. Conversely, no predictivity was observed in ER- BC patients. In their pilot study, Wei et al. had already shown that metabolic profiling, performed by combining NMR and LC–MS method, can distinguish groups of BC patients with no, partial, or complete response; our study confirms this evidence and enriches this scenario by coupling NMR-based metabolomics with the analysis of a panel of 10 different cytokines.

Conversely, from blood serum analyses, NMR-based metabolomics studies performed on breast biopsies did not reveal significant metabolic differences between complete pathological response and pathological non-responders before treatment (Choi et al., 2013; Euceda et al., 2017), implying that the host-cancer interaction at a systemic level plays a crucial role in the response to treatment. As a confirmation, a metabolomic study performed via MS on both blood serum and tissue biopsies (collected before, during and after NAC) showed that only marginally correlations are present between the two biospecimens, and that only the serum profile is predictive of NAC response (Debik et al., 2019).

Other relevant serum/plasma metabolomic studies have been targeted at characterizing the impact of NAC on the patient metabolism. The NAC induces relevant alterations in patient metabolism during and after treatment (Corona et al., 2021; Debik et al., 2019; Jobard et al., 2017). In particular, Jobard *et al* (Jobard et al., 2017) have shown that the administration of trastuzumab and everolimus in combination induced systemic effects by altering lipid, glucose and ketone bodies metabolisms. These alterations are observable on the metabolic profiles of patients even several weeks after the end of the drug intervention.

2.2.1.2 Disease Recurrence

Despite all efforts aimed at better stratifying BC patients, there is still a significant proportion of early breast cancer patients who are overtreated. Clinicians are looking for prognostic tools able to distinguish early BC patients at high risk of disease recurrence, who need to be treated with aggressive adjuvant therapies, with respect to low risk patients, who may be cured by locoregional therapy alone (McCartney et al., 2018, 2017). NMR-based metabolomics of blood derivatives has shown to play a role in this scenario.

The first evidence supporting the application of metabolomics as potential prognostic tool for recurrence prediction was published by Asiago *et al*. in 2010 (Asiago et al., 2010). They utilized metabolic profiling approach, obtained by combining metabolites detected by both NMR and MS, to identify breast cancer relapse before it occurs. Over the past decade our group has pursued this research line using a fingerprinting approach. We have established a reproducible method, based on serum NMR metabolomic analysis, able to distinguish early and metastatic breast cancer patients with high discrimination accuracy. Furthermore, we demonstrated that this model can be used to predict cancer relapse: early BC patients "misclassified" as metastatic on the basis of their metabolomic fingerprint showed indeed high risk of recurrence, whereas early BC patients correctly classified as early BC can be considered at low risk. The recurrence prediction with this approach has been validated and reproduced in monocentric and multicentric cohorts of patients (**Figure 3**) providing successful results (Tenori et al., 2015; Hart et al., 2017; McCartney et al., 2019). In our 2015 and 2019 studies serum samples of early BC patients were collected before surgery, thus when the tumor was still in place, whereas in the multicentric study of 2017 samples were collected after surgery but before starting (when indicated) adjuvant chemotherapy or radiotherapy. Moreover, the early BC patients enrolled for the studies of 2017 and 2019 had ER+ breast cancer, whereas those of the study of 2015 were diagnosed with ER- breast cancer. The reproducibility of our approach, despite these differences in the study design, reinforces the evidence that NMR-based metabolomics is really a promising instrument for the stratification of patients with early breast cancer.

From our multicentric study (Hart et al., 2017) it emerged that, as compared with metastatic BC patients, patients with early ER+ BC are characterized by lower serum levels of citrate, choline, acetate, formate, lactate, glutamate, 3-hydroxybutyrate, phenylalanine, glycine, leucine, alanine, proline, tyrosine, isoleucine, creatine, creatinine and methionine, and higher serum levels of glucose and glutamine. Interestingly, in the subgroup of early BC patients with either relapse or no-relapse (with a follow up of at least 6 years), the patients who relapsed showed significantly higher serum levels of choline, phenylalanine, leucine, histidine, glutamate, glycine, tyrosine, valine, lactate and isoleucine, thus resembling a "micro-metastatic" profile already at diagnosis. The decrease of tyrosine and lactate in early BC patients confirmed what already seen in the ER- negative cohort examined in 2015, whereas glucose and histidine showed opposite trends in ER- BC patients.

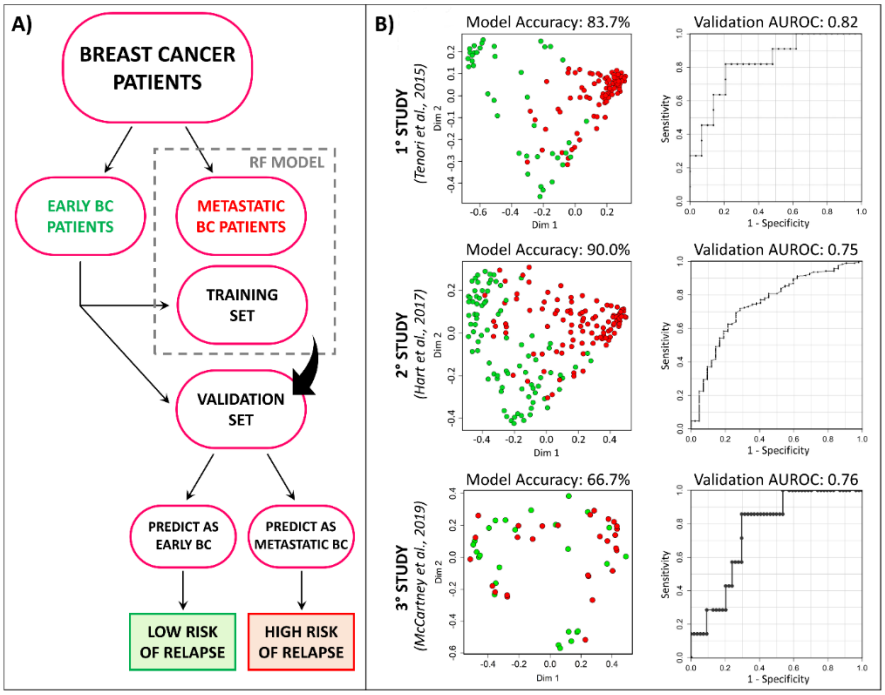


Fig. 3 A) The study design of the three metabolomic studies performed in our laboratory (Tenori et al., 2015; Hart et al., 2017; McCartney et al., 2019) on serum samples of BC patients. Early BC patients were randomly divided into a training and a validation set. Using the early BC training set and the metastatic BC group, a random forest model was calculated on the serum NMR data. Then the early BC validation set was used to test the hypothesis that early BC patients at high risk of cancer recurrence could be classified as metastatic by the RF model. B) The results of the three studies are summarized reporting for each of them the RF proximity plot (early BC green dots, metastatic BC red dots) along with the RF model accuracy, and the area under the ROC curve of the validation set used for the recurrence prediction.

2.2.2 Lung cancer

Lung cancer (LC) is the second most common cancer in both men and women and is by far the leading cause of cancer death worldwide, making up almost 25% of all cancer deaths. Most of the patients (>84%) with LC are affected by non-small cell lung cancer (NSCLC), with the majority of patients presenting with advanced, unresectable disease at the time of diagnosis (Siegel et al., 2021). Currently used treatments for advanced NSCLC include chemotherapy, targeted drug therapy, immunotherapy, or chemo-immunotherapy (Hirsch et al., 2017). Treatment options are based mainly on the tumor histology but other factors, such as certain cancer traits like PDL-1 TPS (tumor proportion score) and the presence of specific genomic mutations, are also important.

2.2.2.1 Response to Immunotherapy

Among all the therapeutic strategies, immunotherapy has become an attractive treatment modality for chemo-refractory solid tumors (Postow et al., 2015). NSCLC cells have the ability to evade the immune system by expressing on their surfaces certain "immune checkpoint" molecules that normally protect against autoimmunity and inflammation, such as cytotoxic T-lymphocyte antigen-4 (CTLA-4), programmed cell death protein 1 and its ligand (PD-1 and PDL-1, respectively). Immune Checkpoint Inhibitor agents (ICIs), such as the monoclonal antibodies nivolumab and pembrolizumab, reactivate T lymphocyte-mediated immune response against the cancer cells by blocking the immune checkpoint molecules (Brahmer, 2014; Hamada et al., 2018). Several ICIs have shown outstanding early success in many tumor types and have established an important role in the first line of treatment of advanced lung cancer as a monotherapy or in combination with chemotherapy, as well as in second line after standard treatment (Borghaei et al., 2015; Gandhi et al., 2018; Gettinger et al., 2015; Herbst et al., 2016; Rittmeyer et al., 2017).

Unfortunately, not all patients respond to ICIs; the response rates are modest (approximately 30% or less in LC), the associated costs are high, and true predictive markers of treatment efficacy do not exist. Thus, the identification of biomarkers able to identify the patients that are most likely to respond to, and benefit from, ICIs treatment is of pivotal importance (Brahmer, 2014). In this framework, metabolomic fingerprinting of biofluids may represent a timely tool to define metabolomic signatures that might inform on treatment outcomes.

In 2020 our research group conducted a pilot study based on ^{1}H NMR metabolomic investigation of sera samples from NSCLC patients treated with immune checkpoint inhibitors (Ghini et al., 2020a). The experimental scheme of the study is reported in Figure 4. A total of

53 patients with advanced NSCLC were enrolled; thirty-four patients were treated with nivolumab (monoclonal antibody directed against PD-1) and 19 patients were treated with pembrolizumab (monoclonal antibody against PDL-1). All the analyzed samples were collected before the beginning of the treatment (T0) with the aim to a-priori identify responder and non-responder subjects. Significantly, we could show that the metabolomic fingerprint of T0 serum acts as a predictive biomarker of immune checkpoint inhibitors response, being able to predict individual therapy outcome with > 80% accuracy (Ghini et al., 2020a). In the serum samples of non-responder subjects, we detected significantly higher levels of pyruvate and alanine along with, even if not statistically significant, higher lactate and glycine levels and lower citrate levels. All these changes are evocative of increased glycolysis and decreased TCA pathway in non-responders. It is important to underline that the significance of the univariate analysis performed on single metabolite levels strongly relies on the number of subjects included in the study. Thus, further investigations, enrolling much higher numbers of subjects are necessary to confirm these findings.

To our knowledge, this represents the first study using NMR-based metabolomic fingerprinting of serum/plasma samples to predict the individual response to anti PD-1 therapy in NSCLC. Instead, a few other examples based on MS metabolomic profiling can be found in the literature. In 2020, Hatae et al. analysed by GC-MS the plasma samples of 55 NSCLC patients treated with nivolumab and found that a combination of 4 plasma metabolites and several T cell markers could be used as a good biomarker for responder identification (area under the ROC curve = 0.96). The four selected metabolites include molecules related to i) gut microbiota (hippuric acid), ii) fatty acid oxidation (butyryl-carnitine), and iii) redox-related metabolites (cystine and glutathione disulfide) (Hatae et al., 2020). One year later, in 2021 Nie X. et al. used LC-MS profiling of early on-treatment serum to explore predictors of clinical outcomes of anti-PD-1 treatment in 74 Chinese NSCLC patients. Serum samples were collected 2–3 weeks after the first infusion of PD-1 inhibitor. A small metabolite panel consisting of hypoxanthine and histidine was identified and validated as a predictor of treatment response, and high levels of both hypoxanthine and histidine were associated with improved progression-free survival and overall survival (Nie et al., 2021). The completely different metabolites observed in the reported investigations and proposed as treatment-efficacy biomarkers derive from different analytical platforms, which allow for the observation of different panels of metabolites.

The strength of our study relies on the uniqueness of the ^{1}H NMR fingerprinting, which takes advantage of its intrinsically untargeted nature and high reproducibility. This approach

allowed us to identify a metabolomic signature associated to ICIs response that is independent of metabolite assignment and acts as a stronger "collective" biomarker with respect to a single molecule or to a panel of a few molecules.

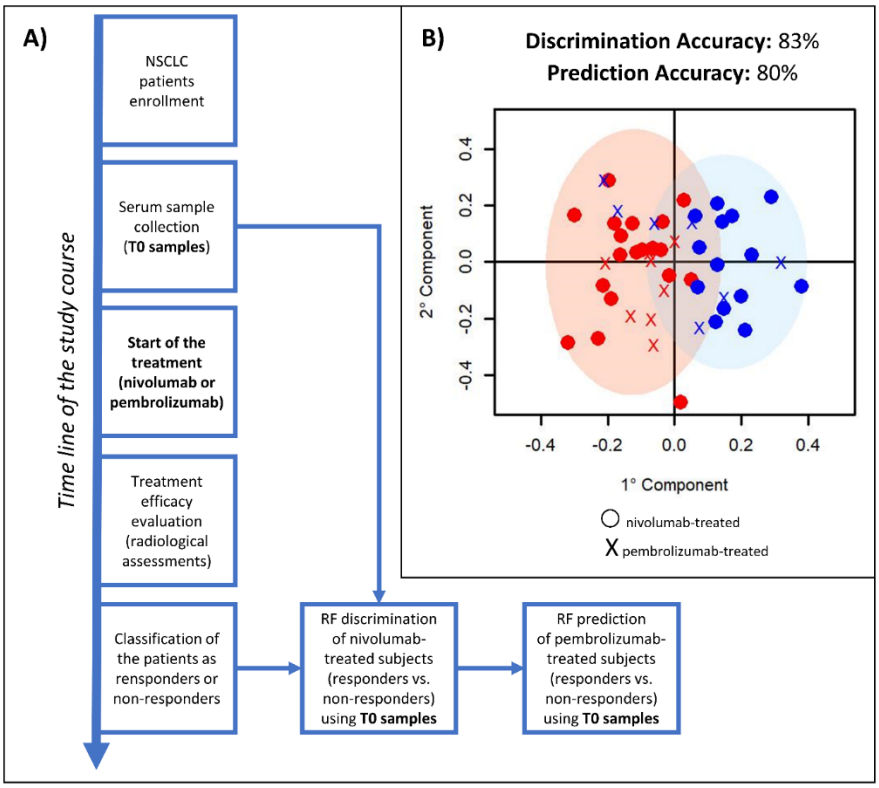


Fig. 4 A) Experimental scheme to evaluate individual response to immunotherapy. B) O-PLS discrimination between NSCLC responders and non-responders to immune check points inhibitors, adapted from (Ghini et al., 2020a). Score plot, PC1 vs. PC2. Each symbol in the O-PLS score plot represents the fingerprint of the NMR spectrum of each patient recruited for the study. Red dots: nivolumab responder subjects; blues dots: nivolumab non-responder subjects. Red crosses: pembrolizumab subjects predicted as responders; blue crosses: pembrolizumab subjects predicted as non-responders.

2.2.3 Colon Cancer

Colorectal cancer (CRC) is the third most prevalent malignancy after breast and lung cancer, and the second most lethal disease in the world, with an anticipated 1.9 million new cases and 0.9 million deaths in 2020 (Xi and Xu, 2021). The stage of CRC at diagnosis is the most important predictor of survival: if colon cancer is detected early, a relative 5-year survival rate of around 90% has been demonstrated for patients diagnosed with localized-stage disease, declining to around 71% and 14% for those diagnosed with regional and distant stages, respectively (Salmerón et al., 2022). Despite that, over the last 20 years, breakthroughs in CRC treatment have resulted in a steady increase in median overall survival (Dekker et al., 2019), CRC is still one of the most lethal diseases. CRC is known for its significant variety in clinical

presentation and underlying tumor biology, as well as its relationship with numerous types of etiological variables (Cunningham et al., 2010). Metabolomics of several biofluids is also increasingly used for successful patient classification in CRC, determining a strong signature of the disease. This allows for discrimination of CRC patients from healthy subjects (Nannini et al., 2020; Turano, 2014), and permits the prediction of the overall survival (OS) within a set of metastatic patients using serum samples (Bertini et al., 2012).

NMR based metabolomics of minimally invasive biospecimens such as feces can correctly classify not only CRC patient from healthy subjects, as it was already shown by several studies (Le Gall et al., 2018; Lin et al., 2019, 2016; Monleón et al., 2009), but can also discriminate CRC from patients with adenomatous polyps (AP) (Nannini et al., 2021). Some of the identified metabolites suggest that metabolic changes in CRC and adenoma are associated with different pathways, mainly involving amino acids metabolism. It is well known that 95% of CRCs begin as colonic AP or adenomas, and the possibility to correctly differentiate these two forms is of primary importance for the early detection of the tumor.

2.2.3.1 Disease Recurrence

Using metabolomics to determine the distinct profiles of CRCs might allow for more personalized or informed cancer therapy adjustments, contributing to precision medicine (Wishart, 2015). Indeed, CRC shows different characteristics of clinical onset and individual response, even at the same pathological stage. Despite 80% of CRCs are diagnosed at early stage and immediately treated with surgery, 35% of these treated patients develop cancer relapse within 2-3 years after surgery (Guraya, 2019). The assessment of recurrence risk in colon cancer primarily relies on the pathological stage defined by the TNM system, based on the depth of tumor invasion (T), lymph nodes involvement (N), and distant metastases (M). Risk stratification is of fundamental importance for the choice of adjuvant treatment; however, only a small portion of patients benefits from it, with the majority being already cured by primary surgery, and other experiencing disease relapse despite having received adjuvant therapy. Improved identification of individuals who would benefit most from adjuvant chemotherapy is a critical aim, particularly in older patients who are more susceptible to treatment-related toxicity.

In 2021 our group demonstrated that NMR based metabolomics on serum samples can improve risk stratification in elderly patients with early CRC (Di Donato et al., 2021). For this study 169 serum samples, taken from three distinct clinical trials, were collected before

treatment from elderly CRC patients. Of these, 94 were patients with early CRC (65 relapse free and 29 relapsed) and 75 from patients with metastatic CRC (Figure 5). The model built using a supervised algorithm to discriminate the serum fingerprint of the relapse-free patients in the early CRC cohort and the patients in the metastatic CRC cohort, yielded 70% accuracy, 71% sensitivity and 69% specificity. Then, with the hypothesis that the metabolomic fingerprint of relapsed early CRC patients would be more similar to that of patients with metastatic CRC, the model was used as training to predict the remaining early CRC who had disease recurrence. Among the early CRC patients, 69% were correctly predicted as metastatic (and therefore considered at high relapse risk). This suggests that even in the absence of clinically visible metastatic spread, the metabolomic fingerprint of individuals with early CRC, who may have cancer recurrence, has a potential "metastatic signature."

Furthermore, when the metabolomic classification of all patients with early CRC was analyzed using Kaplan-Meier curves, a strong prognostic effect was observed, with patients with a metabolomic profile similar to that of patients metastatic CRC having a significantly higher probability of disease relapse than those with a low risk metabolomic score (High vs. Low risk: Hazard Ratio= 3.68, p-value=0.001). Histidine and glutamine were shown to be significantly decreased in the serum of metastatic CRC patients. Previous evidences already suggested an association of glutamine with cancer progression and poor cancer-specific survival (Bertini, 2012; Sirniö et al., 2019). Also, the downregulation of histidine was already observed in other studies (Tan et al., 2013; Zhu et al., 2014) and its alteration was correlated with an increased activity of histidine decarboxylase.

Our research group, in another recent paper (Vignoli et al., 2021a), investigated the CRC relapse in the serum samples of a group of pre-chemotherapy CRC patients undergoing surgery. We demonstrated that several differences between post- and pre-operative serum samples are associated with cancer recurrence, in particular an increment of HDL-Chol subfractions coupled with a decrement of VLDL-Chol subfractions. This may corroborate the hypothesis that the development of CRC disrupts the physiological equilibrium of lipids and lipoproteins, leading to lipid metabolic disorders (Zhang et al., 2014).

Urine samples have been also used for an NMR based metabolomic approach to predict relapse in a group of 62 CRC patients, yielding an area under the ROC curve of 0.650 for cancer recurrence (Dykstra et al., 2017). Interestingly, here the authors use the NMR data to predict treatment delay which could depend on reaction to chemotherapy, reaching an area under the ROC curve values of 0.750.

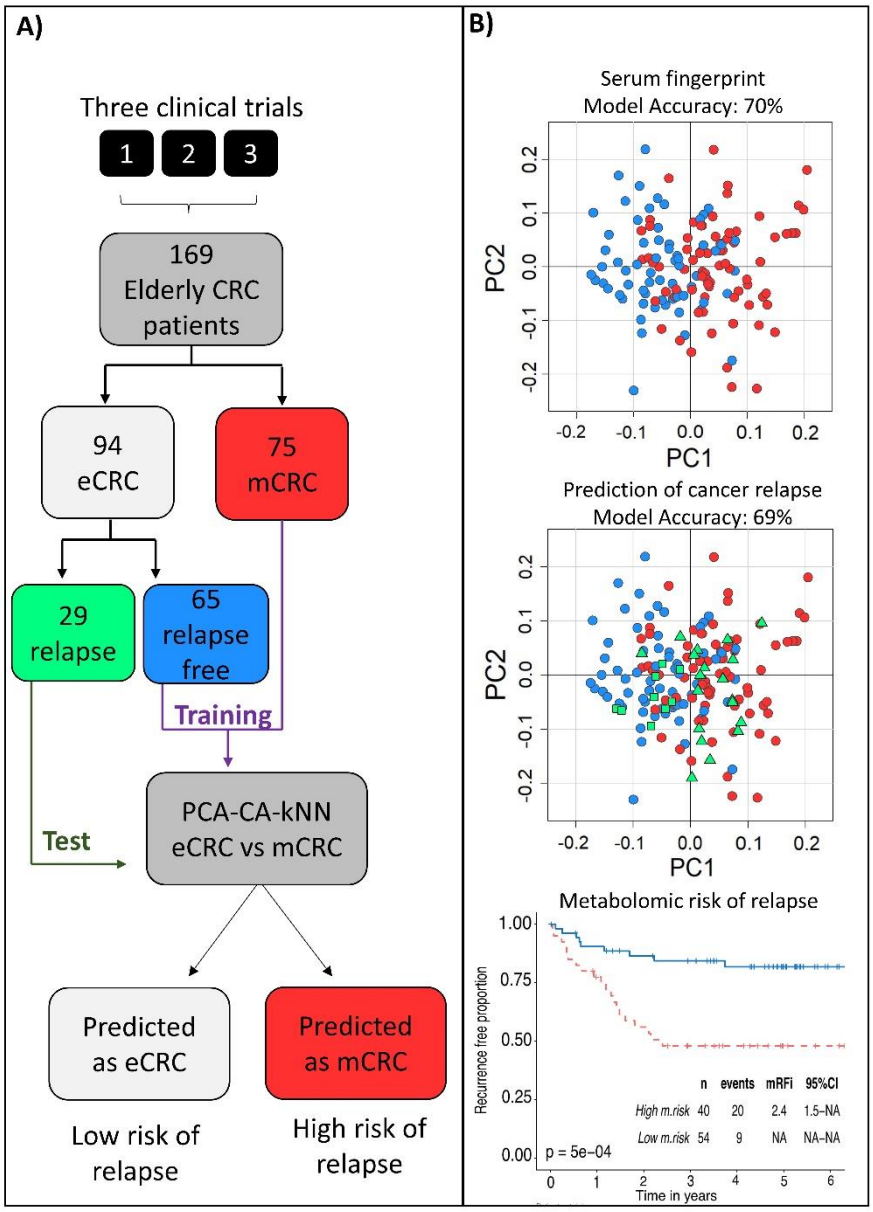


Fig. 5 Experimental scheme (A) and results (B) extracted from (Di Donato et al., 2021). Dots in the score plots of the multivariate supervised PCA-CA kNN model represent the fingerprint of the NMR spectra of each patient recruited for the study. Red dots code for metastatic CRC (mCRC) patients, blue dots for early CRC (eCRC) relapse free patients, green triangles for eCRC relapsed patients predicted as "metastatic", while green squares represent the CRC relapsed predicted as early CRC free from relapse. The Kaplan-Meier plot on the bottom right was used by the authors to estimate the outcome distribution.

2.2.3.2 Pre-surgical effects of anesthesia

The biofluids used by the various studies have been shown to contain important information detectable with metabolomics for the characterization of CRC patients. However, it is important to use the right pre-analytical procedures on the samples to obtain reliable results. We have highlighted (Ghini et al., 2015b) that the moment of blood sample collection can

strongly influence the plasma metabolomic profile. Blood samples were taken from 70 CRC patients (40 non-metastatic and 30 with liver metastasis) preoperatively, both prior and after anesthesia administration. Anesthesia depresses the metabolisms in uneven ways, thus reducing the information content of the metabolic profile and hence reducing the discrimination capability of the method. Consequently, post anesthesia samples are not very suitable for standard metabolomics studies.

2.3 Viral infections

Viruses utilize and/or rewire the host metabolism. Therefore, metabolomics is an excellent tool to study the effect of viral infection, either in vitro using infected cell cultures, or ex-vivo in biosamples from infected animals or humans. In terms of practical applicability, obvious limitations arise from biosecurity, which requires the evaluation of the viral load and infectious risk of the different biological matrices as a function of their nature and of the nature of the infective agent, as well as of the biocontainment level of the laboratory.

On the other hand, identifying the metabolic pathways utilized by a virus has the potential to help revealing drug targets, to monitor the response to antiviral agents but also to evaluate the effect of vaccine administration, although very little is available on this aspect.

Here, we report examples from our laboratory on the metabolomic characterization of viral hepatitis and of SARS-CoV-2 infection.

2.3.1 Hepatitis

Viral hepatitis is a global health issue that affect millions of individuals and is associated with a high fatality rate. Except for the hepatitis A virus (HAV), all hepatotropic viruses, including hepatitis B, C, D, and E viruses (HBV, HCV, HDV, and HEV), can induce chronic infections. HBV and HCV are the most common viral causes of liver disease. According to the updated estimate of the WHO, there are about 300 million people suffering from chronic HBV and 6 million people suffering from chronic HCV (World Health Organization, 2021).

Viral hepatitis chronic infection can cause progressive liver damage leading to fibrosis and cirrhosis. Cirrhosis is the end-stage of every chronic liver disease and is the major risk factor of hepatocellular carcinoma (HCC). HBV and HCV are the leading cause of HCC worldwide, accounting for a significant mortality of more than 1.3 million death per year (Ringehan et al., 2017).

Prevention campaigns are one of the main weapons to limit the incidence of these viruses, especially for HCV for which a vaccine is not yet available. Direct-acting antivirals (DAAs) were approved in 2014, revolutionizing HCV therapy and allowing almost all patients to be

cured. DAAs are very effective and well-tolerated, and they constitute the gold standard for the treatment of HCV chronic infection in patients at all stages of liver disease.

Metabolomics analysis is being utilized to better understand host–pathogen interactions and screen host biospecimen for biomarkers that are characteristic of the viral infection. In 2019 an interesting picture emerged of the metabolomic fingerprint of HCV infection compared to both healthy subjects (HS) and HBV-infected patients (Meoni et al., 2019), suggesting that the two viruses exert a different impact on human metabolism. Indeed, by the comparison of the ^{1}H-NMR serum profiles of HCV- and HBV-infected patients, we identified characteristic metabolomic fingerprints of the two viral infections, obtaining an overall discrimination accuracy of 86% (OPLS-DA algorithm). As expected, the serum fingerprint of HCV- and HBV- infected patients resulted to be extremely different also from the serum fingerprint of HS, with a classification accuracy of 98.7% in the model built to discriminate HCV vs. HS, and a classification accuracy of 80% in the model built of HBV vs. HS (**Figure 6**). Similarly, Godoy et al., 2010, using the ^{1}H-NMR fingerprinting approach, were able to accurately discriminate (95% predictive accuracy) the urine samples of HCV-infected patients from those of HCV-negative subjects, corroborating the potential of ^{1}H-NMR fingerprinting for the fast, non-invasive diagnosis of HCV infection using a urine sample.

The common changes we detected in the serum metabolomic profile of HBV- and HCV-infected patients when compared to HS (e.g. increased levels of 3-hydroxybutyrate, acetate, lactate, and pyruvate) support the hypothesis that these viruses preferentially stimulate glycolysis over oxidative phosphorylation, analogously to the Warburg effect in cancer (Okuda et al., 2002). Instead, in the comparison of HCV- and HBV-infected patients we noted a different behavior of several metabolites, suggesting that the perturbation could be attributable to a direct action of the two types of viruses rather than to the host response. Interestingly, the higher levels of 2-oxoglutarate and 3-hydroxybutyrate in HCV patients compared to HBV, also identified as biomarkers of cardiovascular disease and ketoacidosis, could explain why some extrahepatic manifestation, such as cardiovascular diseases and diabetes, are common in patients with chronic HCV and not in patients with HBV (Bernini et al., 2011b; Chen et al., 2014; Du et al., 2014).

Metabolomics proved useful also for disease staging and for characterizing the response to treatment. Anti-HCV treatment has advanced significantly in recent years, with direct-acting antivirals (DAAs) replacing pegylated interferon and ribavirin and providing effective treatment and less adverse effects. In our work we characterized also the metabolomic fingerprint and the profile of HCV patients before and after effective DAA treatment. In this

case we identified a major contribution of the low molecular weight molecules in characterizing the changes introduced in the individual metabolomic profile by the therapy, suggesting also that the perturbation in lipid metabolism induced by the infection persists after viral eradication.

According to other studies (Cano et al., 2017; Nguyen et al., 2021; Sarfaraz et al., 2016) tyrosine and formate levels increase on passing from no/mild fibrosis to severe fibrosis. Furthermore, differences in metabolite levels between patients with higher and lower fibrosis scores were reduced after DAAs therapy, confirming that altered metabolites are restored, most likely due to liver damage regression after viral eradication.

Regarding the effect of treatment on sera of HBV infected patients, Nguyen et al., 2021, recently published a study confirming that NMR based metabolomics is capable of revealing in serum samples a gradual metabolic transition from pretreatment to early treatment and then to a longer treatment period, as well as of accurately distinguishing the serum of patients who needed medical treatment (patients who would commence treatment within 6 months from sampling) from those who did not.

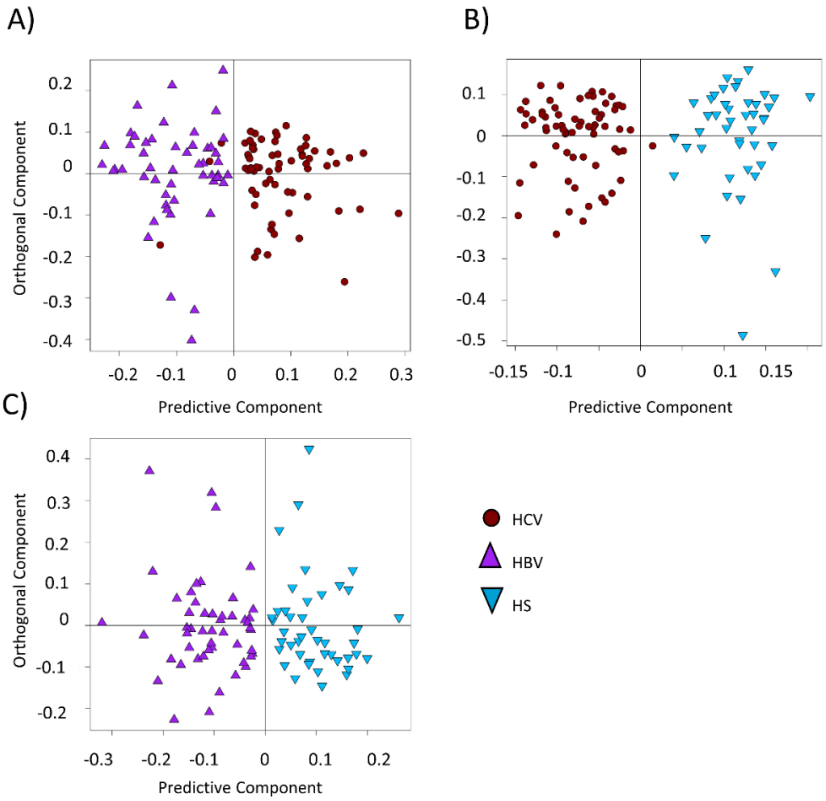


Fig. 6 Fingerprinting of the Hepatitis C and B viruses in serum samples resulting from OPLS-DA models. A) HCV- (red dots) vs HBV- (purple triangles) infected patients; B) HCV-infected patients vs HS (sky-blue triangles); C) HBV-infected patients vs HS. Adapted from (Meoni et al., 2019).

2.3.2 COVID-19

The COVID-19 pandemic has heavily reshaped research activities world-wide. Along with an incredibly fast development of vaccines that could bring the pandemic under control, efforts were also directed towards the development of antiviral drugs and monoclonal antibodies. Additionally, the pandemic has led to a rush to repurpose existing drugs. The pandemic has also stimulated researchers operating in fields complementary to pharmacology and vaccinology to contribute to the understanding of the physiopathology of the disease and to the characterization of risk factors and response to treatments. Coordinated efforts have been searched via the creation of networks that facilitate discussion among participants and communication of the key results prior to publication, together with establishment of strategic transnational collaborations. Worth mentioning in this context are: i) the Covid19-NMR project (https://covid19-nmr.de/), dealing with the determination of the structures of RNA and proteins of SARS-CoV-2 to investigate their drugability by small molecules; ii) the NMR international COVID-19 Research Network (CV19 Research Network) a metabolomics-based initiative consisting of several institutions that collaborate, using standardized NMR procedures, to detect the infection, predict outcomes during hospitalization, and direct efforts towards Long COVID.

Suitable biofluids from COVID-19 patients for metabolomics are serum, plasma, and urine, thanks to their low viral load, albeit collection of research samples during the worse phases of the COVID-19 pandemic was a further burden for the clinicians. In this frame, we present here some activities from our lab directed towards the use of NMR-based metabolomics and lipoproteomics to characterize the COVID-19 metabolomic fingerprints, to monitor the effect of repurposed drugs and vaccination follow-up (Ghini et al., 2022a; Meoni et al., 2021).

We and others (Ballout et al., 2021; Baranovicova et al., 2021; Bizkarguenaga et al., 2021; Bruzzone et al., 2020; Julkunen et al., 2021; Kimhofer et al., 2020; Lodge et al., 2021; Masuda et al., 2021) have shown that SARS-CoV-2 infection induces profound changes in the metabolic phenotype of the patients (Millet, “Prospective metabolomic studies in precision medicine. The AKRIBEA project.”; Rogers, “The Metabolomics of Critical Illness”). Accordingly, ^{1}H NMR spectra of plasma samples of COVID-19 patients could be strongly discriminated from the spectra of both healthy subjects and COVID-19-recovered subjects, with a discrimination accuracy higher than 90% in both cases. The differences originate from significant alterations in the concentrations of several metabolites and of a panel of lipoprotein components. The metabolites and lipoprotein parameters that are significantly dysregulated in COVID-19 acute subjects are listed in Figure 7.

Characteristic trends in metabolite and lipoprotein levels are also observed as a function of the disease severity (Ghini et al., 2022b). The analysis of the specific changes and correlations with clinical data enabled the identification of potential biochemical determinants of the disease fingerprint, which found confirmation in other studies performed at different centers worldwide and in some cases dealing with much larger cohorts. The parameters that are found altered in COVID-19 patients with respect to recovered individuals overlap with the acute infection biomarkers identified in the comparison with healthy subjects, indicating the substantial metabolic healing of COVID-19-recovered subjects. During the healing process, the metabolome and lipoproteome revert back to the "healthy" state with different rates; during either spontaneous healing or pharmacological treatments the metabolites are reverted faster than lipoproteins. Notably, several other metabolomics papers have been published, which identify common molecular features as characteristic for the COVID-19 profile. High convergence on common biomarkers from different metabolomics studies is not so common, and denotes the presence of a strong profile, independent of confounding factors like place of origin, sex, age, comorbidities.

Regarding the pharmacological treatments, we had the chance to analyze the effect of tocilizumab on a very small cohort (8 patients) treated at the Florence University hospital during the first wave of COVID-19 in spring 2020. Tocilizumab is a monoclonal antibody that attaches to the receptor of the cytokine interleukin-6, whose levels are elevated in response to systemic inflammation, and plays an important role in severe COVID-19 disease and associated respiratory failure. On December 6th 2021, the European Medicines Agency (EMA) recommended extending the indication of RoActemra (tocilizumab) to include the treatment of adult COVID-19 patients who are receiving systemic treatment with corticosteroids and require supplemental oxygen or mechanical ventilation.

When measuring the post-treatment levels of metabolites and lipoproteins that are significantly altered by the infection, we found that eight metabolites (namely, acetone, citrate, glutamine, glycine, lactate, mannose, phenylalanine, and pyruvate) partially or completely revert towards the levels of CTR subjects. Some lipoprotein main- and sub-fractions were also significantly affected by the tocilizumab treatment, but, they did not revert back to "healthy" values, in line with the above-reported observation that the recovery of the lipoproteome is slower than that of the metabolome along the healing process (Meoni et al., 2021).

Finally, ^{1}H NMR spectra of sera have also been used to define the changes induced by vaccination with Pfizer-BioNTech vaccine in a cohort of 20 healthcare workers, 10 COVID-19 naïve and 10 with a previous history of COVID-19 infection (infected in the period March-April 2020 with the Wuhan strain and recovered from the disease 208-280 days before

vaccination). All of them received two doses of vaccine, 21 days apart, and their serum samples were collected at 6 different time points to monitor time-dependent changes induced by the vaccination. Importantly, the vaccination does not induce a major modification of the metabolic phenotype of the subjects; the intra-individual differences remain smaller than the inter-individual ones during all the course of the study, with an individual discrimination accuracy >85% (considering the six samples collected for each subject). Nevertheless, in response to vaccination we could observe some common changes that are consistently occurring in all subjects within each group. While vaccination does not induce any significant variation in the metabolome, it causes changes at the level of lipoproteins that are smaller for COVID-19-recovered subjects with respect to naïve subjects, suggesting that a previous infection reduces the vaccine modulation of the lipoproteome composition. The differences between the two groups involve the nature and number of affected lipoprotein parameters. Additionally, the effect of the second dose is essentially negligible for the COVID-19-recovered subjects (Ghini et al., 2022a).

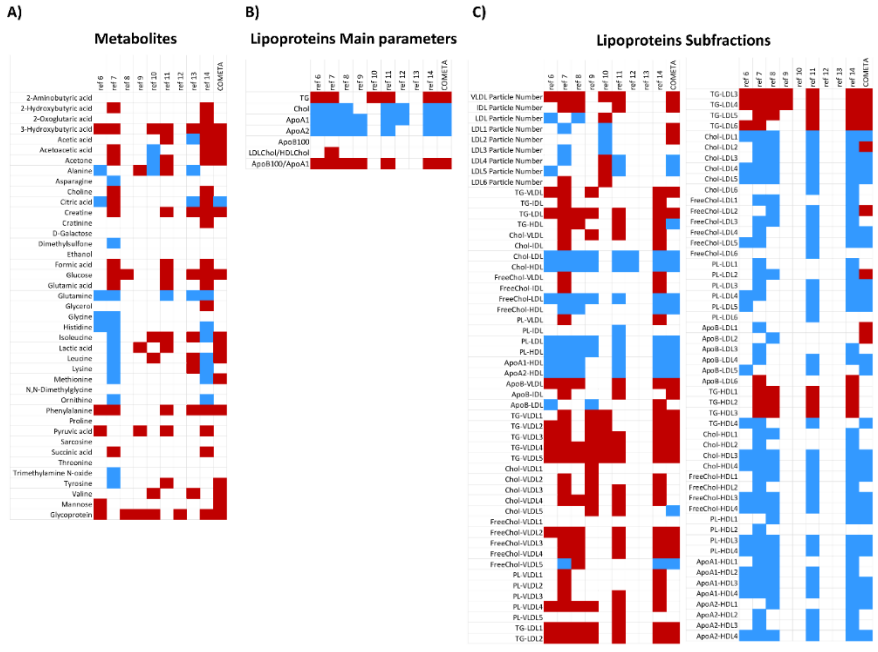


Fig. 7 Metabolites and lipoprotein parameters that are found significantly up- or down- regulated in COVID-19 patients with respect to healthy controls according to different Metabolomics papers [ref 6 (Meoni et al., 2021); ref 7 (Bruzzone et al., 2020); ref 8 (Kimhofer et al., 2020); ref 9 (Lodge et al., 2021); ref 10 (Ballout et al., 2021); ref 11 (Masuda et al., 2021); ref 12 (Julkunen et al., 2021); ref 13 (Baranovicova et al., 2021); ref 14 (Bruzzone et al., 2020) and COMETA project (Ghini et al., 2022b). Up-/down- regulated features are indicated by red/blue cells. A) Metabolites; B) Lipoprotein Main Parameters; C) Lipoprotein Subfractions. Adapted from (Ghini et al., 2022b).

2.4 Chronic Obstructive Pulmonary Disease

As highlighted in Chapter 12 (Lacy, "Metabolomics of Respiratory Diseases"), chronic obstructive pulmonary disease (COPD) is a pathological condition characterized by the chronic, poorly reversible and progressive development of airflow limitation often associated with parenchymal destruction and emphysema (Barnes, 2000).

Although abnormal respiratory inflammation is crucial for COPD development, the complex COPD pathophysiology is not yet fully understood. Currently, no validated biomarker is accepted for disease prognosis or COPD therapy monitoring (Ghosh et al., 2016).

Several studies have identified serum, exhaled breath condensate (EBC) and urine ^{1}H NMR-based metabolomic fingerprints of COPD patients, also showing the ability of EBC metabolomics to assess airway inflammation (Airoldi et al., 2016; Laurentiis et al., 2008; Motta et al., 2012; Wang et al., 2013; Ząbek et al., 2015). In this frame, a study from our group (Bertini et al., 2014) showed that NMR-metabolomics of EBC could discriminate COPD patients from controls with an overall accuracy of 86%. As compared to controls, EBC from COPD patients featured significantly lower levels of acetone, valine and lysine, and significantly higher levels of lactate, acetate, propionate, serine, proline, and tyrosine. Lower levels of valine and lysine (two essential amino acids) appear consistent with muscle wasting and weight loss that are known to occur in advanced COPD (Agusti et al., 2002). The hypothesis of a possible presence of a subclinical malnutrition in COPD is also discussed in Chapter 12, where the reader can find a more comprehensive overview of the metabolic alterations observed in sera of COPD patients.

The application of pharmacometabolomics in the context of COPD is aimed at identifying the appropriate treatment for each individual COPD patient, predicting his/her response to therapy. Our group applied an original and holistic approach, dubbed "breathomics", to monitor the effects of treatment with and withdrawal from inhaled beclomethasone/formoterol in patients with COPD (Montuschi et al., 2018). In our application, breathomics combined two electronic noses (carbon polymer sensor e-nose, quartz crystal sensor e-nose), EBC NMR-based metabolomics, sputum cell counts, sputum supernatant and EBC prostaglandin E2 (PGE2) and 15-F2t-isoprostane, fraction of exhaled nitric oxide, and spirometry data. Breathomics improves the identification of pharmacological treatment-induced effects as compared with standard spirometry. Furthermore, this approach provides insights into the anti-inflammatory effects of inhaled corticosteroids in COPD patients as reflected by reduced levels of sputum PGE2 and EBC acetate during treatment with formoterol alone. This research line was further explored with the analysis of urine, serum and sputum supernatant, demonstrating that different biological matrices provide complementary information on the

effects of beclomethasone/formoterol administration in COPD patients, and thus their integration could be useful for elucidating the metabolic mechanism of action of inhaled corticosteroids (Vignoli et al., 2020b). An overview of the metabolite correlation patterns among the different biofluids is presented in Figure 8.

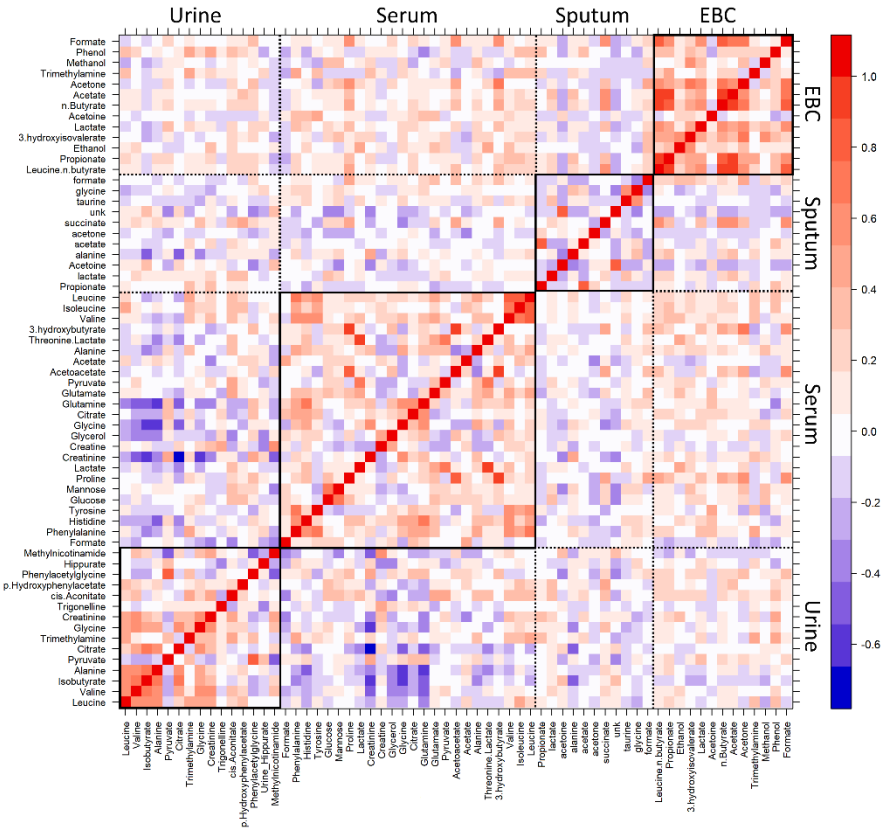


Fig. 8 Heatmap showing correlations among metabolites quantified in the urine, serum, exhaled breath condensate (EBC), and sputum supernatant samples of 14 patients with COPD. Correlation values (R) are reported as different degrees of color intensity (blue, negative correlation; red, positive correlations). The figure is adapted and reprinted with permission from (Vignoli et al., 2020b). Copyright 2020 American Chemical Society.

2.5 Drug Toxicity

Pharmacometabolomics includes drug safety evaluation, that is an important step in the drug pipeline and a main concern for regulatory agencies. Safety evaluation is required across the whole process of drug development, from preclinical studies to clinical trials, as well as in post-approval safety surveillance. Biochemical and histological analyses are the major approaches used for drug safety evaluation. These approaches are effective in most cases to determine the safety profile of drug candidates (Wang et al., 2017). However, these methods can neither provide detailed information nor explore the mechanisms of drug toxicity (Nicholson et al., 2002). Metabolomics is ideally positioned to address the challenges of drug

toxicity (Hertz, “Chemotherapy-induced Peripheral Neuropathy”). It represents a powerful tool for collecting mechanistic information, indicating not only the extent of a toxic insult but also its underlying mechanisms (Ramirez et al., 2013). Several examples exist in the literature of NMR-based metabolomic applications in toxicology and drug safety evaluation. Just to cite few of them, drugs and toxicants like flutamide (Choucha Snouber et al., 2013), hydrazine (Garrod et al., 2005; Lindon et al., 2003), and gentamicin (Lenz et al., 2005) have been assessed.

A typical example of a clinically useful class of molecules that presents high rates of therapy discontinuation due to acute side effects is represented by V-phosphodiesterase (PDE5) inhibitors (PDE5i) (Corona et al., 2016). These molecules are the first-line therapy for erectile dysfunction (ED), a widespread health problem in the general population of middle-aged men (Ayta et al., 1999; Yafi et al., 2016). Administration of PDE5i proved beneficial in 60–70% of patients with varying etiologies of sexual dysfunction (Yafi et al., 2016). However adverse effects are a common drawback. In a study from our research group, we retrospectively evaluated serum and urine NMR-based metabolomic profiles to identify prognostic biomarkers of unfavorable efficacy/safety profile of PDE5i before drug administration. To the best of our knowledge this is the first and only NMR-based metabolomic study focused on PDE5i toxicity. Patients who are likely to experience adverse effects can be identified with an accuracy of 77% using pre-treatment serum samples. Adverse drug reactions showed to be associated with high levels of LDL-lipoprotein subfractions at baseline (Rocca et al., 2020). The results of this pilot study underline how metabolomics may help in identifying the metabolic bases underlying efficacy/safety profile of the PDE5i therapy.

REFERENCES

Agusti, A.G., Sauleda, J., Miralles, C., Gomez, C., Togores, B., Sala, E., Batle, S., Busquets, X., 2002. Skeletal muscle apoptosis and weight loss in chronic obstructive pulmonary disease. Am.J Respir.Crit Care Med 166, 485–489. https://doi.org/10.1164/rccm.2108013

Airoldi, C., Ciaramelli, C., Fumagalli, M., Bussei, R., Mazzoni, V., Viglio, S., Iadarola, P., Stolk, J., 2016. 1H NMR To Explore the Metabolome of Exhaled Breath Condensate in α1-Antitrypsin Deficient Patients: A Pilot Study. J. Proteome Res. 15, 4569–4578. https://doi.org/10.1021/acs.jproteome.6b00648

Andersen, M.-B.S., Rinnan, Å., Manach, C., Poulsen, S.K., Pujos-Guillot, E., Larsen, T.M., Astrup, A., Dragsted, L.O., 2014. Untargeted metabolomics as a screening tool for estimating compliance to a dietary pattern. J. Proteome Res. 13, 1405–1418. https://doi.org/10.1021/pr400964s

Antoniewicz, M.R., 2018. A guide to 13C metabolic flux analysis for the cancer biologist. Exp Mol Med 50, 1–13. https://doi.org/10.1038/s12276-018-0060-y

Asiago, V.M., Alvarado, L.Z., Shanaiah, N., Gowda, G.A.N., Owusu-Sarfo, K., Ballas, R.A., Raftery, D., 2010. Early detection of recurrent breast cancer using metabolite profiling. Cancer Res. 70, 8309–8318. https://doi.org/10.1158/0008-5472.CAN-10-1319

Assfalg, M., Bertini, I., Colangiuli, D., Luchinat, C., Schäfer, H., Schütz, B., Spraul, M., 2008. Evidence of different metabolic phenotypes in humans. Proc Natl Acad Sci U S A 105, 1420–1424. https://doi.org/10.1073/pnas.0705685105

Ayta, I.A., McKinlay, J.B., Krane, R.J., 1999. The likely worldwide increase in erectile dysfunction between 1995 and 2025 and some possible policy consequences. BJU Int 84, 50–56. https://doi.org/10.1046/j.1464-410x.1999.00142.x

Ballout, R.A., Kong, H., Sampson, M., Otvos, J.D., Cox, A.L., Agbor-Enoh, S., Remaley, A.T., 2021. The NIH Lipo-COVID Study: A Pilot NMR Investigation of Lipoprotein Subfractions and Other Metabolites in Patients with Severe COVID-19. Biomedicines 9, 1090. https://doi.org/10.3390/biomedicines9091090

Baranovicova, E., Bobcakova, A., Vysehradsky, R., Dankova, Z., Halasova, E., Nosal, V., Lehotsky, J., 2021. The Ability to Normalise Energy Metabolism in Advanced COVID-19 Disease Seems to Be One of the Key Factors Determining the Disease Progression—A Metabolomic NMR Study on Blood Plasma. Applied Sciences 11, 4231. https://doi.org/10.3390/app11094231

Barnes, P.J., 2000. Chronic obstructive pulmonary disease. N. Engl. J. Med. 343, 269–280. https://doi.org/10.1056/NEJM200007273430407

Bendinelli, B., Vignoli, A., Palli, D., Assedi, M., Ambrogetti, D., Luchinat, C., Caini, S., Saieva, C., Turano, P., Masala, G., 2021. Prediagnostic circulating metabolites in female breast cancer cases with low and high mammographic breast density. Sci Rep 11, 13025. https://doi.org/10.1038/s41598-021-92508-1

Bernini, P., Bertini, I., Calabrò, A., la Marca, G., Lami, G., Luchinat, C., Renzi, D., Tenori, L., 2011a. Are Patients with Potential Celiac Disease Really Potential? The Answer of Metabonomics. J. Proteome Res. 10, 714–721. https://doi.org/10.1021/pr100896s

Bernini, P., Bertini, I., Luchinat, C., Tenori, L., Tognaccini, A., 2011b. The Cardiovascular Risk of Healthy Individuals Studied by NMR Metabonomics of Plasma Samples. J. Proteome Res. 10, 4983–4992. https://doi.org/10.1021/pr200452j

Bertini, I., 2012. Metabolomic NMR Fingerprinting to Identify and Predict Survival of Patients with Metastatic Colorectal Cancer. CANCER RESEARCH 10 1158 0008-5472-11–1543.

Bertini, I., Cacciatore, S., Jensen, B.V., Schou, J.V., Johansen, J.S., Kruhøffer, M., Luchinat, C., Nielsen, D.L., Turano, P., 2012. Metabolomic NMR Fingerprinting to Identify and Predict Survival of Patients with Metastatic Colorectal Cancer. Cancer Res 72, 356–364. https://doi.org/10.1158/0008-5472.CAN-11-1543

Bertini, I., Calabrò, A., De Carli, V., Luchinat, C., Nepi, S., Porfirio, B., Renzi, D., Saccenti, E., Tenori, L., 2009. The Metabonomic Signature of Celiac Disease. J. Proteome Res. 8, 170–177. https://doi.org/10.1021/pr800548z

Bertini, I., Luchinat, C., Miniati, M., Monti, S., Tenori, L., 2014. Phenotyping COPD by H-1 NMR metabolomics of exhaled breath condensate. Metabolomics 10, 302–311. https://doi.org/10.1007/s11306-013-0572-3

Bizkarguenaga, M., Bruzzone, C., Gil-Redondo, R., SanJuan, I., Martin-Ruiz, I., Barriales, D., Palacios, A., Pasco, S.T., González-Valle, B., Laín, A., Herrera, L., Azkarate, A., Vesga, M.A., Eguizabal,

C., Anguita, J., Embade, N., Mato, J.M., Millet, O., 2021. Uneven metabolic and lipidomic profiles in recovered COVID-19 patients as investigated by plasma NMR metabolomics. NMR Biomed e4637. https://doi.org/10.1002/nbm.4637

Borghaei, H., Paz-Ares, L., Horn, L., Spigel, D.R., Steins, M., Ready, N.E., Chow, L.Q., Vokes, E.E., Felip, E., Holgado, E., Barlesi, F., Kohlhäufl, M., Arrieta, O., Burgio, M.A., Fayette, J., Lena, H., Poddubskaya, E., Gerber, D.E., Gettinger, S.N., Rudin, C.M., Rizvi, N., Crinò, L., Blumenschein, G.R., Antonia, S.J., Dorange, C., Harbison, C.T., Graf Finckenstein, F., Brahmer, J.R., 2015. Nivolumab versus Docetaxel in Advanced Nonsquamous Non–Small-Cell Lung Cancer. New England Journal of Medicine 373, 1627–1639. https://doi.org/10.1056/NEJMoa1507643

Brahmer, J.R., 2014. Immune checkpoint blockade: the hope for immunotherapy as a treatment of lung cancer? Semin Oncol 41, 126–132. https://doi.org/10.1053/j.seminoncol.2013.12.014

Bruzzone, C., Bizkarguenaga, M., Gil-Redondo, R., Diercks, T., Arana, E., García de Vicuña, A., Seco, M., Bosch, A., Palazón, A., San Juan, I., Laín, A., Gil-Martínez, J., Bernardo-Seisdedos, G., Fernández-Ramos, D., Lopitz-Otsoa, F., Embade, N., Lu, S., Mato, J.M., Millet, O., 2020. SARS-CoV-2 Infection Dysregulates the Metabolomic and Lipidomic Profiles of Serum. iScience 23, 101645. https://doi.org/10.1016/j.isci.2020.101645

Buchwald, H., Estok, R., Fahrbach, K., Banel, D., Jensen, M.D., Pories, W.J., Bantle, J.P., Sledge, I., 2009. Weight and type 2 diabetes after bariatric surgery: systematic review and meta-analysis. Am J Med 122, 248-256.e5. https://doi.org/10.1016/j.amjmed.2008.09.041

Cacciatore, S., Hu, X., Viertler, C., Kap, M., Bernhardt, G.A., Mischinger, H.-J., Riegman, P., Zatloukal, K., Luchinat, C., Turano, P., 2013. Effects of Intra- and Post-Operative Ischemia on the Metabolic Profile of Clinical Liver Tissue Specimens Monitored by NMR. J. Proteome Res. 12, 5723–5729. https://doi.org/10.1021/pr400702d

Cala, M.P., Aldana, J., Medina, J., Sánchez, J., Guio, J., Wist, J., Meesters, R.J.W., 2018. Multiplatform plasma metabolic and lipid fingerprinting of breast cancer: A pilot control-case study in Colombian Hispanic women. PLoS One 13, e0190958. https://doi.org/10.1371/journal.pone.0190958

Cano, A., Mariño, Z., Millet, O., Martínez-Arranz, I., Navasa, M., Falcón-Pérez, J.M., Pérez-Cormenzana, M., Caballería, J., Embade, N., Forns, X., Bosch, J., Castro, A., Mato, J.M., 2017. A Metabolomics Signature Linked To Liver Fibrosis In The Serum Of Transplanted Hepatitis C Patients. Sci Rep 7. https://doi.org/10.1038/s41598-017-10807-y

Chen, P.-A., Xu, Z.-H., Huang, Y.-L., Luo, Y., Zhu, D.-J., Wang, P., Du, Z.-Y., Yang, Y., Wu, D.-H., Lai, W.-Y., Ren, H., Xu, D.-L., 2014. Increased serum 2-oxoglutarate associated with high myocardial energy expenditure and poor prognosis in chronic heart failure patients. Biochim. Biophys. Acta 1842, 2120–2125. https://doi.org/10.1016/j.bbadis.2014.07.018

Choi, J.S., Baek, H.-M., Kim, S., Kim, M.J., Youk, J.H., Moon, H.J., Kim, E.-K., Nam, Y.K., 2013. Magnetic resonance metabolic profiling of breast cancer tissue obtained with core needle biopsy for predicting pathologic response to neoadjuvant chemotherapy. PLoS One 8, e83866. https://doi.org/10.1371/journal.pone.0083866

Choucha Snouber, L., Bunescu, A., Naudot, M., Legallais, C., Brochot, C., Dumas, M.E., Elena-Herrmann, B., Leclerc, E., 2013. Metabolomics-on-a-Chip of Hepatotoxicity Induced by Anticancer Drug Flutamide and Its Active Metabolite Hydroxyflutamide Using HepG2/C3a Microfluidic Biochips. Toxicological Sciences 132, 8–20. https://doi.org/10.1093/toxsci/kfs230

Corona, G., Di Gregorio, E., Vignoli, A., Muraro, E., Steffan, A., Miolo, G., 2021. 1H-NMR Plasma Lipoproteins Profile Analysis Reveals Lipid Metabolism Alterations in HER2-Positive Breast Cancer Patients. Cancers 13, 5845. https://doi.org/10.3390/cancers13225845

Corona, G., Rastrelli, G., Burri, A., Serra, E., Gianfrilli, D., Mannucci, E., Jannini, E.A., Maggi, M., 2016. First-generation phosphodiesterase type 5 inhibitors dropout: a comprehensive review and meta-analysis. Andrology 4, 1002–1009. https://doi.org/10.1111/andr.12255

Cunningham, D., Atkin, W., Lenz, H.-J., Lynch, H.T., Minsky, B., Nordlinger, B., Starling, N., 2010. Colorectal cancer. Lancet 375, 1030–1047. https://doi.org/10.1016/S0140-6736(10)60353-4

Cuperlovic-Culf, M., Ferguson, D., Culf, A., Morin, P., Touaibia, M., 2012. 1H NMR Metabolomics Analysis of Glioblastoma Subtypes CORRELATION BETWEEN METABOLOMICS AND GENE EXPRESSION CHARACTERISTICS. J. Biol. Chem. 287, 20164–20175. https://doi.org/10.1074/jbc.M111.337196

D'Alessandro, G., Quaglio, D., Monaco, L., Lauro, C., Ghirga, F., Ingallina, C., De Martino, M., Fucile, S., Porzia, A., Di Castro, M.A., Bellato, F., Mastrotto, F., Mori, M., Infante, P., Turano, P., Salmaso, S., Caliceti, P., Di Marcotullio, L., Botta, B., Ghini, V., Limatola, C., 2019. 1H-NMR metabolomics reveals the Glabrescione B exacerbation of glycolytic metabolism beside the cell

growth inhibitory effect in glioma. Cell Commun. Signal 17, 108. https://doi.org/10.1186/s12964-019-0421-8

Debik, J., Euceda, L.R., Lundgren, S., Gythfeldt, H. von der L., Garred, Ø., Borgen, E., Engebraaten, O., Bathen, T.F., Giskeødegård, G.F., 2019. Assessing Treatment Response and Prognosis by Serum and Tissue Metabolomics in Breast Cancer Patients. J Proteome Res 18, 3649–3660. https://doi.org/10.1021/acs.jproteome.9b00316

Dekker, E., Tanis, P.J., Vleugels, J.L.A., Kasi, P.M., Wallace, M.B., 2019. Colorectal cancer. Lancet 394, 1467–1480. https://doi.org/10.1016/S0140-6736(19)32319-0

Di Donato, S., Vignoli, A., Biagioni, C., Malorni, L., Mori, E., Tenori, L., Calamai, V., Parnofiello, A., Di Pierro, G., Migliaccio, I., Cantafio, S., Baraghini, M., Mottino, G., Becheri, D., Del Monte, F., Miceli, E., McCartney, A., Di Leo, A., Luchinat, C., Biganzoli, L., 2021. A Serum Metabolomics Classifier Derived from Elderly Patients with Metastatic Colorectal Cancer Predicts Relapse in the Adjuvant Setting. Cancers 13, 2762. https://doi.org/10.3390/cancers13112762

Díaz-Beltrán, L., González-Olmedo, C., Luque-Caro, N., Díaz, C., Martín-Blázquez, A., Fernández-Navarro, M., Ortega-Granados, A.L., Gálvez-Montosa, F., Vicente, F., Pérez del Palacio, J., Sánchez-Rovira, P., 2021. Human Plasma Metabolomics for Biomarker Discovery: Targeting the Molecular Subtypes in Breast Cancer. Cancers 13, 147. https://doi.org/10.3390/cancers13010147

Du, Z., Shen, A., Huang, Y., Su, L., Lai, W., Wang, P., Xie, Zhibing, Xie, Zhiquan, Zeng, Q., Ren, H., Xu, D., 2014. 1H-NMR-based metabolic analysis of human serum reveals novel markers of myocardial energy expenditure in heart failure patients. PLoS One 9, e88102. https://doi.org/10.1371/journal.pone.0088102

Dykstra, M.A., Switzer, N., Eisner, R., Tso, V., Foshaug, R., Ismond, K., Fedorak, R., Wang, H., 2017. Urine metabolomics as a predictor of patient tolerance and response to adjuvant chemotherapy in colorectal cancer. Mol Clin Oncol 7, 767–770. https://doi.org/10.3892/mco.2017.1407

Early Breast Cancer Trialists' Collaborative Group (EBCTCG), 2005. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet 365, 1687–1717. https://doi.org/10.1016/S0140-6736(05)66544-0

Euceda, L.R., Haukaas, T.H., Giskeødegård, G.F., Vettukattil, M.R., Engel, J., Silwal-Pandit, L., Lundgren, S., Borgen, E., Garred, Ø., Postma, G., Buydens, L.M.C., Børresen-Dale, A.-L., Engebråten, O., Bathen, T.F., 2017. Evaluation of metabolomic changes during neoadjuvant chemotherapy combined with bevacizumab in breast cancer using MR spectroscopy. 13. https://doi.org/10.1007/s11306-017-1168-0

Fan, Y., Zhou, X., Xia, T.-S., Chen, Z., Li, J., Liu, Q., Alolga, R.N., Chen, Y., Lai, M.-D., Li, P., Zhu, W., Qi, L.-W., 2016. Human plasma metabolomics for identifying differential metabolites and predicting molecular subtypes of breast cancer. Oncotarget 7, 9925–9938. https://doi.org/10.18632/oncotarget.7155

Gandhi, L., Rodríguez-Abreu, D., Gadgeel, S., Esteban, E., Felip, E., De Angelis, F., Domine, M., Clingan, P., Hochmair, M.J., Powell, S.F., Cheng, S.Y.-S., Bischoff, H.G., Peled, N., Grossi, F., Jennens, R.R., Reck, M., Hui, R., Garon, E.B., Boyer, M., Rubio-Viqueira, B., Novello, S., Kurata, T., Gray, J.E., Vida, J., Wei, Z., Yang, J., Raftopoulos, H., Pietanza, M.C., Garassino, M.C., KEYNOTE-189 Investigators, 2018. Pembrolizumab plus Chemotherapy in Metastatic Non-Small-Cell Lung Cancer. N Engl J Med 378, 2078–2092. https://doi.org/10.1056/NEJMoa1801005

Garrod, S., Bollard, M.E., Nicholls, A.W., Connor, S.C., Connelly, J., Nicholson, J.K., Holmes, E., 2005. Integrated Metabonomic Analysis of the Multiorgan Effects of Hydrazine Toxicity in the Rat. Chem. Res. Toxicol. 18, 115–122. https://doi.org/10.1021/tx0498915

Gettinger, S.N., Horn, L., Gandhi, L., Spigel, D.R., Antonia, S.J., Rizvi, N.A., Powderly, J.D., Heist, R.S., Carvajal, R.D., Jackman, D.M., Sequist, L.V., Smith, D.C., Leming, P., Carbone, D.P., Pinder-Schenck, M.C., Topalian, S.L., Hodi, F.S., Sosman, J.A., Sznol, M., McDermott, D.F., Pardoll, D.M., Sankar, V., Ahlers, C.M., Salvati, M., Wigginton, J.M., Hellmann, M.D., Kollia, G.D., Gupta, A.K., Brahmer, J.R., 2015. Overall Survival and Long-Term Safety of Nivolumab (Anti-Programmed Death 1 Antibody, BMS-936558, ONO-4538) in Patients With Previously Treated Advanced Non-Small-Cell Lung Cancer. J Clin Oncol 33, 2004–2012. https://doi.org/10.1200/JCO.2014.58.3708

Ghini, V., Di Nunzio, M., Tenori, L., Valli, V., Danesi, F., Capozzi, F., Luchinat, C., Bordoni, A., 2017. Evidence of a DHA Signature in the Lipidome and Metabolome of Human Hepatocytes. International Journal of Molecular Sciences 18, 359. https://doi.org/10.3390/ijms18020359

Ghini, V., Laera, L., Fantechi, B., del Monte, F., Benelli, M., McCartney, A., Tenori, L., Luchinat, C., Pozzessere, D., 2020a. Metabolomics to Assess Response to Immune Checkpoint Inhibitors in Patients with Non-Small-Cell Lung Cancer. Cancers 12, 3574. https://doi.org/10.3390/cancers12123574

Ghini, V., Maggi, L., Mazzoni, A., Spinicci, M., Zammarchi, L., Bartoloni, A., Annunziato, F., Turano, P., 2022a. Serum NMR Profiling Reveals Differential Alterations in the Lipoproteome Induced by Pfizer-BioNTech Vaccine in COVID-19 Recovered Subjects and Naïve Subjects. Frontiers in Molecular Biosciences 9. https://doi.org/10.3389/fmolb.2022.839809

Ghini, V., Meoni, G., Pelagatti, L., Celli, T., Veneziani, F., Petrucci, F., Vannucchi, V., Bertini, L., Luchinat, C., Landini, G., Turano, P., 2022b. Profiling metabolites and lipoproteins in COMETA, an Italian cohort of COVID-19 patients. PLoS Pathog 18, e1010443. https://doi.org/doi.org/10.1371/journal. ppat.1010443

Ghini, V., Quaglio, D., Luchinat, C., Turano, P., 2019. NMR for sample quality assessment in metabolomics. N Biotechnol 52, 25–34. https://doi.org/10.1016/j.nbt.2019.04.004

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

Ghini, V., Senzacqua, T., Massai, L., Gamberi, T., Messori, L., Turano, P., 2021. NMR reveals the metabolic changes induced by auranofin in A2780 cancer cells: evidence for glutathione dysregulation. Dalton Trans. 50, 6349–6355. https://doi.org/10.1039/D1DT00750E

Ghini, V., Tenori, L., Pane, M., Amoruso, A., Marroncini, G., Squarzanti, D.F., Azzimonti, B., Rolla, R., Savoia, P., Tarocchi, M., Galli, A., Luchinat, C., 2020b. Effects of Probiotics Administration on Human Metabolic Phenotype. Metabolites 10, 396. https://doi.org/10.3390/metabo10100396

Ghini, V., Unger, F.T., Tenori, L., Turano, P., Juhl, H., David, K.A., 2015b. Metabolomics profiling of pre-and post-anesthesia plasma samples of colorectal patients obtained via Ficoll separation. Metabolomics 11, 1769–1778. https://doi.org/10.1007/s11306-015-0832-5

Ghosh, N., Dutta, M., Singh, B., Banerjee, R., Bhattacharyya, P., Chaudhury, K., 2016. Transcriptomics, proteomics and metabolomics driven biomarker discovery in COPD: an update. Expert Rev Mol Diagn 16, 897–913. https://doi.org/10.1080/14737159.2016.1198258

Gianni, L., Pienkowski, T., Im, Y.-H., Roman, L., Tseng, L.-M., Liu, M.-C., Lluch, A., Staroslawska, E., de la Haba-Rodriguez, J., Im, S.-A., Pedrini, J.L., Poirier, B., Morandi, P., Semiglazov, V., Srimuninnimit, V., Bianchi, G., Szado, T., Ratnayake, J., Ross, G., Valagussa, P., 2012. Efficacy and safety of neoadjuvant pertuzumab and trastuzumab in women with locally advanced, inflammatory, or early HER2-positive breast cancer (NeoSphere): a randomised multicentre, open-label, phase 2 trial. Lancet Oncol 13, 25–32. https://doi.org/10.1016/S1470-2045(11)70336-9

Gralka, E., Luchinat, C., Tenori, L., Ernst, B., Thurnheer, M., Schultes, B., 2015. Metabolomic fingerprint of severe obesity is dynamically affected by bariatric surgery in a procedure-dependent manner. American Journal of Clinical Nutrition 102, 1313–1322. https://doi.org/10.3945/ajcn.115.110536

Griffiths, W.J., 2008. Metabolomics, Metabonomics and Metabolite Profiling. Royal Society of Chemistry.

Guraya, S.Y., 2019. Pattern, Stage, and Time of Recurrent Colorectal Cancer After Curative Surgery. Clin Colorectal Cancer 18, e223–e228. https://doi.org/10.1016/j.clcc.2019.01.003

Hamada, T., Kosumi, K., Nakai, Y., Koike, K., 2018. Surrogate study endpoints in the era of cancer immunotherapy. Ann Transl Med 6, S27. https://doi.org/10.21037/atm.2018.09.31

Hart, C.D., Vignoli, A., Tenori, L., Uy, G.L., To, T.V., Adebamowo, C., Hossain, S.M., Biganzoli, L., Risi, E., Love, R.R., Luchinat, C., Di Leo, A., 2017. Serum Metabolomic Profiles Identify ER-Positive Early Breast Cancer Patients at Increased Risk of Disease Recurrence in a Multicenter Population. Clinical Cancer Research 23, 1422–1431. https://doi.org/10.1158/1078-0432.CCR-16-1153

Hatae, R., Chamoto, K., Kim, Y.H., Sonomura, K., Taneishi, K., Kawaguchi, S., Yoshida, H., Ozasa, H., Sakamori, Y., Akrami, M., Fagarasan, S., Masuda, I., Okuno, Y., Matsuda, F., Hirai, T., Honjo, T., 2020. Combination of host immune metabolic biomarkers for the PD-1 blockade cancer immunotherapy. JCI Insight 5, 133501. https://doi.org/10.1172/jci.insight.133501

Herbst, R.S., Baas, P., Kim, D.-W., Felip, E., Pérez-Gracia, J.L., Han, J.-Y., Molina, J., Kim, J.-H., Arvis, C.D., Ahn, M.-J., Majem, M., Fidler, M.J., de Castro, G., Garrido, M., Lubiniecki, G.M.,

Shentu, Y., Im, E., Dolled-Filhart, M., Garon, E.B., 2016. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. Lancet 387, 1540–1550. https://doi.org/10.1016/S0140-6736(15)01281-7

Hirsch, F.R., Scagliotti, G.V., Mulshine, J.L., Kwon, R., Curran, W.J., Wu, Y.-L., Paz-Ares, L., 2017. Lung cancer: current therapies and new targeted treatments. Lancet 389, 299–311. https://doi.org/10.1016/S0140-6736(16)30958-8

Holmes, E., Loo, R.L., Stamler, J., Bictash, M., Yap, I.K.S., Chan, Q., Ebbels, T., De Iorio, M., Brown, I.J., Veselkov, K.A., Daviglus, M.L., Kesteloot, H., Ueshima, H., Zhao, L., Nicholson, J.K., Elliott, P., 2008. Human metabolic phenotype diversity and its association with diet and blood pressure. Nature 453, 396–400. https://doi.org/10.1038/nature06882

ISO/DIS 23118 Molecular in vitro diagnostic examinations — Specifications for pre-examination processes in metabolomics in urine, venous blood serum and plasma [WWW Document], 2021. URL https://www.iso.org/obp/ui/#iso:std:iso:23118:ed-1:v1:en

Jiang, L., Lee, S.C., Ng, T.C., 2018. Pharmacometabonomics Analysis Reveals Serum Formate and Acetate Potentially Associated with Varying Response to Gemcitabine-Carboplatin Chemotherapy in Metastatic Breast Cancer Patients. J Proteome Res 17, 1248–1257. https://doi.org/10.1021/acs.jproteome.7b00859

Jobard, E., Trédan, O., Bachelot, T., Vigneron, A.M., Aït-Oukhatar, C.M., Arnedos, M., Rios, M., Bonneterre, J., Diéras, V., Jimenez, M., Merlin, J.-L., Campone, M., Elena-Herrmann, B., 2017. Longitudinal serum metabolomics evaluation of trastuzumab and everolimus combination as pre-operative treatment for HER-2 positive breast cancer patients. Oncotarget 8, 83570–83584. https://doi.org/10.18632/oncotarget.18784

Julkunen, H., Cichońska, A., Slagboom, P.E., Würtz, P., 2021. Metabolic biomarker profiling for identification of susceptibility to severe pneumonia and COVID-19 in the general population. eLife 10, e63033. https://doi.org/10.7554/eLife.63033

Kaddurah-Daouk, R., Weinshilboum, R.M., Pharmacometabolomics Research Network, 2014. Pharmacometabolomics: implications for clinical pharmacology and systems pharmacology. Clin Pharmacol Ther 95, 154–167. https://doi.org/10.1038/clpt.2013.217

Kimhofer, T., Lodge, S., Whiley, L., Gray, N., Loo, R.L., Lawler, N.G., Nitschke, P., Bong, S.-H., Morrison, D.L., Begum, S., Richards, T., Yeap, B.B., Smith, C., Smith, K.C.G., Holmes, E., Nicholson, J.K., 2020. Integrative Modelling of Quantitative Plasma Lipoprotein, Metabolic and Amino Acid Data Reveals a Multi-organ Pathological Signature of SARS-CoV-2 Infection. J. Proteome Res. https://doi.org/10.1021/acs.jproteome.0c00519

Klupczyńska, A., Dereziński, P., Kokot, Z.J., 2015. METABOLOMICS IN MEDICAL SCIENCES--TRENDS, CHALLENGES AND PERSPECTIVES. Acta Pol Pharm 72, 629–641.

Laurentiis, G. de, Paris, D., Melck, D., Maniscalco, M., Marsico, S., Corso, G., Motta, A., Sofia, M., 2008. Metabonomic analysis of exhaled breath condensate in adults by nuclear magnetic resonance spectroscopy. Eur Respir J 32, 1175–1183. https://doi.org/10.1183/09031936.00072408

Le Gall, G., Guttula, K., Kellingray, L., Tett, A.J., ten Hoopen, R., Kemsley, K.E., Savva, G.M., Ibrahim, A., Narbad, A., 2018. Metabolite quantification of faecal extracts from colorectal cancer patients and healthy controls. Oncotarget 9, 33278–33289. https://doi.org/10.18632/oncotarget.26022

Lécuyer, L., Victor Bala, A., Deschasaux, M., Bouchemal, N., Nawfal Triba, M., Vasson, M.-P., Rossary, A., Demidem, A., Galan, P., Hercberg, S., Partula, V., Le Moyec, L., Srour, B., Fiolet, T., Latino-Martel, P., Kesse-Guyot, E., Savarin, P., Touvier, M., 2018. NMR metabolomic signatures reveal predictive plasma metabolites associated with long-term risk of developing breast cancer. Int J Epidemiol. https://doi.org/10.1093/ije/dyx271

Lenz, E.M., Bright, J., Knight, R., Westwood, F.R., Davies, D., Major, H., Wilson, I.D., 2005. Metabonomics with 1H-NMR spectroscopy and liquid chromatography-mass spectrometry applied to the investigation of metabolic changes caused by gentamicin-induced nephrotoxicity in the rat. Biomarkers 10, 173–187. https://doi.org/10.1080/13547500500094034

Lenz, E.M., Bright, J., Wilson, I.D., Morgan, S.R., Nash, A.F.P., 2003. A 1H NMR-based metabonomic study of urine and plasma samples obtained from healthy human subjects. Journal of Pharmaceutical and Biomedical Analysis 33, 1103–1115. https://doi.org/10.1016/S0731-7085(03)00410-2

Lerner, H.J., Band, P.R., Israel, L., Leung, B.S., 1976. Phase II study of tamoxifen: report of 74 patients with stage IV breast cancer. Cancer Treat Rep 60, 1431–1435.

Li, C., Li, Z., Zhang, T., Wei, P., Li, N., Zhang, W., Ding, X., Li, J., 2019. 1H NMR-Based Metabolomics Reveals the Antitumor Mechanisms of Triptolide in BALB/c Mice Bearing CT26 Tumors. Frontiers in Pharmacology 10. https://doi.org/10.3389/fphar.2019.01175

Li, Y., Wang, C., Li, D., Deng, P., Shao, X., Hu, J., Liu, C., Jie, H., Lin, Y., Li, Z., Qian, X., Zhang, H., Zhao, Y., 2017. 1H-NMR-based metabolic profiling of a colorectal cancer CT-26 lung metastasis model in mice. Oncology Reports 38, 3044–3054. https://doi.org/10.3892/or.2017.5954

Lin, Y., Ma, C., Bezabeh, T., Wang, Z., Liang, J., Huang, Y., Zhao, J., Liu, X., Ye, W., Tang, W., Ouyang, T., Wu, R., 2019. 1 H NMR-based metabolomics reveal overlapping discriminatory metabolites and metabolic pathway disturbances between colorectal tumor tissues and fecal samples. Int J Cancer 145, 1679–1689. https://doi.org/10.1002/ijc.32190

Lin, Y., Ma, C., Liu, C., Wang, Z., Yang, J., Liu, X., Shen, Z., Wu, R., 2016. NMR-based fecal metabolomics fingerprinting as predictors of earlier diagnosis in patients with colorectal cancer. Oncotarget 7, 29454–29464. https://doi.org/10.18632/oncotarget.8762

Lindon, J.C., Nicholson, J.K., Holmes, E., Antti, H., Bollard, M.E., Keun, H., Beckonert, O., Ebbels, T.M., Reily, M.D., Robertson, D., Stevens, G.J., Luke, P., Breau, A.P., Cantor, G.H., Bible, R.H., Niederhauser, U., Senn, H., Schlotterbeck, G., Sidelmann, U.G., Laursen, S.M., Tymiak, A., Car, B.D., Lehman-McKeeman, L., Colet, J.-M., Loukaci, A., Thomas, C., 2003. Contemporary issues in toxicology the role of metabonomics in toxicology and its evaluation by the COMET project. Toxicology and Applied Pharmacology 187, 137–146. https://doi.org/10.1016/S0041-008X(02)00079-0

Lodge, S., Nitschke, P., Kimhofer, T., Coudert, J.D., Begum, S., Bong, S.-H., Richards, T., Edgar, D., Raby, E., Spraul, M., Schaefer, H., Lindon, J.C., Loo, R.L., Holmes, E., Nicholson, J.K., 2021. NMR Spectroscopic Windows on the Systemic Effects of SARS-CoV-2 Infection on Plasma Lipoproteins and Metabolites in Relation to Circulating Cytokines. J. Proteome Res. https://doi.org/10.1021/acs.jproteome.0c00876

Marianne C Walsh, L.B., 2006. Effect of acute dietary standardization on the urinary, plasma, and salivary metabolomic profiles of healthy humans. The American journal of clinical nutrition 84, 531–9. https://doi.org/10.1093/ajcn/84.3.531

Márquez, J., Matés, J.M., 2021. Tumor Metabolome: Therapeutic Opportunities Targeting Cancer Metabolic Reprogramming. Cancers (Basel) 13, 314. https://doi.org/10.3390/cancers13020314

Martin, M., Holmes, F.A., Ejlertsen, B., Delaloge, S., Moy, B., Iwata, H., von Minckwitz, G., Chia, S.K.L., Mansi, J., Barrios, C.H., Gnant, M., Tomašević, Z., Denduluri, N., Šeparović, R., Gokmen, E., Bashford, A., Ruiz Borrego, M., Kim, S.-B., Jakobsen, E.H., Ciceniene, A., Inoue, K., Overkamp, F., Heijns, J.B., Armstrong, A.C., Link, J.S., Joy, A.A., Bryce, R., Wong, A., Moran, S., Yao, B., Xu, F., Auerbach, A., Buyse, M., Chan, A., ExteNET Study Group, 2017. Neratinib after trastuzumab-based adjuvant therapy in HER2-positive breast cancer (ExteNET): 5-year analysis of a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet Oncol 18, 1688–1700. https://doi.org/10.1016/S1470-2045(17)30717-9

Masuda, R., Lodge, S., Nitschke, P., Spraul, M., Schaefer, H., Bong, S.-H., Kimhofer, T., Hall, D., Loo, R.L., Bizkarguenaga, M., Bruzzone, C., Gil-Redondo, R., Embade, N., Mato, J.M., Holmes, E., Wist, J., Millet, O., Nicholson, J.K., 2021. Integrative Modeling of Plasma Metabolic and Lipoprotein Biomarkers of SARS-CoV-2 Infection in Spanish and Australian COVID-19 Patient Cohorts. J. Proteome Res. 20, 4139–4152. https://doi.org/10.1021/acs.jproteome.1c00458

McCartney, A., Vignoli, A., Biganzoli, L., Love, R., Tenori, L., Luchinat, C., Di Leo, A., 2018. Metabolomics in breast cancer: A decade in review. Cancer Treat. Rev. 67, 88–96. https://doi.org/10.1016/j.ctrv.2018.04.012

McCartney, A., Vignoli, A., Hart, C., Tenori, L., Luchinat, C., Biganzoli, L., Di Leo, A., 2017. De-escalating and escalating treatment beyond endocrine therapy in patients with luminal breast cancer. Breast 34, S13–S18. https://doi.org/10.1016/j.breast.2017.06.021

McCartney, A., Vignoli, A., Tenori, L., Fornier, M., Rossi, L., Risi, E., Luchinat, C., Biganzoli, L., Di Leo, A., 2019. Metabolomic analysis of serum may refine 21-gene expression assay risk recurrence stratification. NPJ Breast Cancer 5, 26. https://doi.org/10.1038/s41523-019-0123-9

Meijer, R.I., van Wagensveld, B.A., Siegert, C.E., Eringa, E.C., Smulders, Y.M., 2011. Bariatric Surgery as a Novel Treatment for Type 2 Diabetes Mellitus: A Systematic Review. Archives of Surgery 146, 744–750. https://doi.org/10.1001/archsurg.2011.134

Meoni, G., Ghini, V., Maggi, L., Vignoli, A., Mazzoni, A., Salvati, L., Capone, M., Vanni, A., Tenori, L., Fontanari, P., Lavorini, F., Peris, A., Bartoloni, A., Liotta, F., Cosmi, L., Luchinat, C., Annunziato, F., Turano, P., 2021. Metabolomic/lipidomic profiling of COVID-19 and individual

response to tocilizumab. PLOS Pathogens 17, e1009243. https://doi.org/10.1371/journal.ppat.1009243

Meoni, G., Lorini, S., Monti, M., Madia, F., Corti, G., Luchinat, C., Zignego, A.L., Tenori, L., Gragnani, L., 2019. The metabolic fingerprints of HCV and HBV infections studied by Nuclear Magnetic Resonance Spectroscopy. Scientific Reports 9, 4128. https://doi.org/10.1038/s41598-019-40028-4

Monleón, D., Morales, J.M., Barrasa, A., López, J.A., Vázquez, C., Celda, B., 2009. Metabolite profiling of fecal water extracts from human colorectal cancer. NMR Biomed 22, 342–348. https://doi.org/10.1002/nbm.1345

Montuschi, P., Santini, G., Mores, N., Vignoli, A., Macagno, F., Shohreh, R., Tenori, L., Zini, G., Fuso, L., Mondino, C., Di Natale, C., D'Amico, A., Barnes, P.J., Higenbottam, T., 2018. BREATHOMICS FOR ASSESSING THE EFFECTS OF TREATMENT AND WITHDRAWAL WITH INHALED BECLOMETHASONE/FORMOTEROL IN PATIENTS WITH COPD. Front. Pharmacol. 9. https://doi.org/10.3389/fphar.2018.00258

Motta, A., Paris, D., Melck, D., de Laurentiis, G., Maniscalco, M., Sofia, M., Montuschi, P., 2012. Nuclear magnetic resonance-based metabolomics of exhaled breath condensate: methodological aspects. Eur. Respir. J. 39, 498–500. https://doi.org/10.1183/09031936.00036411

Nannini, G., Meoni, G., Amedei, A., Tenori, L., 2020. Metabolomics profile in gastrointestinal cancers: Update and future perspectives. World J Gastroenterol 26, 2514–2532. https://doi.org/10.3748/wjg.v26.i20.2514

Nannini, G., Meoni, G., Tenori, L., Ringressi, M.N., Taddei, A., Niccolai, E., Baldi, S., Russo, E., Luchinat, C., Amedei, A., 2021. Fecal metabolomic profiles: A comparative study of patients with colorectal cancer vs adenomatous polyps. World Journal of Gastroenterology 27, 6430–6441. https://doi.org/10.3748/wjg.v27.i38.6430

Nguyen, H.T.T., Wimmer, R., Le, V.Q., Krarup, H.B., 2021. Metabolic fingerprint of progression of chronic hepatitis B: changes in the metabolome and novel diagnostic possibilities. Metabolomics 17, 16. https://doi.org/10.1007/s11306-020-01767-y

Nicholson, J.K., Connelly, J., Lindon, J.C., Holmes, E., 2002. Metabonomics: a platform for studying drug toxicity and gene function. Nat Rev Drug Discov 1, 153–161. https://doi.org/10.1038/nrd728

Nie, X., Xia, L., Gao, F., Liu, L., Yang, Y., Chen, Y., Duan, H., Yao, Y., Chen, Z., Lu, S., Wang, Y., Yang, C., 2021. Serum Metabolite Biomarkers Predictive of Response to PD-1 Blockade Therapy in Non-Small Cell Lung Cancer. Frontiers in Molecular Biosciences 8, 472. https://doi.org/10.3389/fmolb.2021.678753

Okuda, M., Li, K., Beard, M.R., Showalter, L.A., Scholle, F., Lemon, S.M., Weinman, S.A., 2002. Mitochondrial injury, oxidative stress, and antioxidant gene expression are induced by hepatitis C virus core protein. Gastroenterology 122, 366–375. https://doi.org/10.1053/gast.2002.30983

O'Sullivan, A., Gibney, M.J., Brennan, L., 2011. Dietary intake patterns are reflected in metabolomic profiles: potential role in dietary assessment studies. Am. J. Clin. Nutr. 93, 314–321. https://doi.org/10.3945/ajcn.110.000950

Polyak, K., 2011. Heterogeneity in breast cancer. J Clin Invest 121, 3786–3788. https://doi.org/10.1172/JCI60534

Postow, M.A., Callahan, M.K., Wolchok, J.D., 2015. Immune Checkpoint Blockade in Cancer Therapy. J Clin Oncol 33, 1974–1982. https://doi.org/10.1200/JCO.2014.59.4358

Ramirez, T., Daneshian, M., Kamp, H., Bois, F.Y., Clench, M.R., Coen, M., Donley, B., Fischer, S.M., Ekman, D.R., Fabian, E., Guillou, C., Heuer, J., Hogberg, H.T., Jungnickel, H., Keun, H.C., Krennrich, G., Krupp, E., Luch, A., Noor, F., Peter, E., Riefke, B., Seymour, M., Skinner, N., Smirnova, L., Verheij, E., Wagner, S., Hartung, T., van Ravenzwaay, B., Leist, M., 2013. Metabolomics in Toxicology and Preclinical Research. ALTEX 30, 209–225. https://doi.org/10.14573/altex.2013.2.209

Resendiz-Acevedo, K., García-Aguilera, M.E., Esturau-Escofet, N., Ruiz-Azuara, L., 2021. 1H -NMR Metabolomics Study of the Effect of Cisplatin and Casiopeina IIgly on MDA-MB-231 Breast Tumor Cells. Frontiers in Molecular Biosciences 8, 1135. https://doi.org/10.3389/fmolb.2021.742859

Ringehan, M., McKeating, J.A., Protzer, U., 2017. Viral hepatitis and liver cancer. Philos Trans R Soc Lond B Biol Sci 372, 20160274. https://doi.org/10.1098/rstb.2016.0274

Rittmeyer, A., Barlesi, F., Waterkamp, D., Park, K., Ciardiello, F., von Pawel, J., Gadgeel, S.M., Hida, T., Kowalski, D.M., Dols, M.C., Cortinovis, D.L., Leach, J., Polikoff, J., Barrios, C., Kabbinavar, F., Frontera, O.A., De Marinis, F., Turna, H., Lee, J.-S., Ballinger, M., Kowanetz, M., He, P., Chen, D.S., Sandler, A., Gandara, D.R., OAK Study Group, 2017. Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. Lancet 389, 255–265. https://doi.org/10.1016/S0140-6736(16)32517-X

Rocca, M.S., Vignoli, A., Tenori, L., Ghezzi, M., De Rocco Ponce, M., Vatsellas, G., Thanos, D., Padrini, R., Foresta, C., De Toni, L., 2020. Evaluation of Serum/Urine Genomic and Metabolomic Profiles to Improve the Adherence to Sildenafil Therapy in Patients with Erectile Dysfunction. Front. Pharmacol. 11, 602369. https://doi.org/10.3389/fphar.2020.602369

Saborano, R., Eraslan, Z., Roberts, J., Khanim, F.L., Lalor, P.F., Reed, M.A.C., Günther, U.L., 2019. A framework for tracer-based metabolism in mammalian cells by NMR. Sci Rep 9, 2520. https://doi.org/10.1038/s41598-018-37525-3

Salmerón, A.M., Tristán, A.I., Abreu, A.C., Fernández, I., 2022. Serum Colorectal Cancer Biomarkers Unraveled by NMR Metabolomics: Past, Present, and Future. Anal. Chem. 94, 417–430. https://doi.org/10.1021/acs.analchem.1c04360

Sarfaraz, M.O., Myers, R.P., Coffin, C.S., Gao, Z.-H., Shaheen, A.A.M., Crotty, P.M., Zhang, P., Vogel, H.J., Weljie, A.M., 2016. A quantitative metabolomics profiling approach for the noninvasive assessment of liver histology in patients with chronic hepatitis C. Clin Transl Med 5, 33. https://doi.org/10.1186/s40169-016-0109-2

Siegel, R.L., Miller, K.D., Fuchs, H.E., Jemal, A., 2021. Cancer Statistics, 2021. CA: A Cancer Journal for Clinicians 71, 7–33. https://doi.org/10.3322/caac.21654

Silva, C.L., Olival, A., Perestrelo, R., Silva, P., Tomás, H., Câmara, J.S., 2019. Untargeted Urinary 1H NMR-Based Metabolomic Pattern as a Potential Platform in Breast Cancer Detection. Metabolites 9. https://doi.org/10.3390/metabo9110269

Singh, A., Sharma, R.K., Chagtoo, M., Agarwal, G., George, N., Sinha, N., Godbole, M.M., 2017. 1H NMR Metabolomics Reveals Association of High Expression of Inositol 1, 4, 5 Trisphosphate Receptor and Metabolites in Breast Cancer Patients. PLoS One 12. https://doi.org/10.1371/journal.pone.0169330

Sirniö, P., Väyrynen, J.P., Klintrup, K., Mäkelä, J., Karhu, T., Herzig, K.-H., Minkkinen, I., Mäkinen, M.J., Karttunen, T.J., Tuomisto, A., 2019. Alterations in serum amino-acid profile in the progression of colorectal cancer: associations with systemic inflammation, tumour stage and patient survival. British Journal of Cancer 120, 238–246. https://doi.org/10.1038/s41416-018-0357-6

Sitter, B., Lundgren, S., Bathen, T.F., Halgunset, J., Fjosne, H.E., Gribbestad, I.S., 2006. Comparison of HR MAS MR spectroscopic profiles of breast cancer tissue with clinical parameters. NMR Biomed 19, 30–40. https://doi.org/10.1002/nbm.992

Slamon, D., Eiermann, W., Robert, N., Pienkowski, T., Martin, M., Press, M., Mackey, J., Glaspy, J., Chan, A., Pawlicki, M., Pinter, T., Valero, V., Liu, M.-C., Sauter, G., von Minckwitz, G., Visco, F., Bee, V., Buyse, M., Bendahmane, B., Tabah-Fisch, I., Lindsay, M.-A., Riva, A., Crown, J., Breast Cancer International Research Group, 2011. Adjuvant trastuzumab in HER2-positive breast cancer. N Engl J Med 365, 1273–1283. https://doi.org/10.1056/NEJMoa0910383

Slupsky, C.M., Steed, H., Wells, T., Dabbs, K., Schepansky, A., Capstick, V., Faught, W., Sawyer, M.B., 2010. Urine metabolite analysis offers potential early diagnosis of ovarian and breast cancers. Clin Cancer Res clincanres.1434.2010. https://doi.org/10.1158/1078-0432.CCR-10-1434

Suman, S., Sharma, R.K., Kumar, V., Sinha, N., Shukla, Y., 2018. Metabolic fingerprinting in breast cancer stages through 1H NMR spectroscopy-based metabolomic analysis of plasma. J Pharm Biomed Anal 160, 38–45. https://doi.org/10.1016/j.jpba.2018.07.024

Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F., 2021. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 71, 209–249. https://doi.org/10.3322/caac.21660

Swain, S.M., Kim, S.-B., Cortés, J., Ro, J., Semiglazov, V., Campone, M., Ciruelos, E., Ferrero, J.-M., Schneeweiss, A., Knott, A., Clark, E., Ross, G., Benyunes, M.C., Baselga, J., 2013. Pertuzumab, trastuzumab, and docetaxel for HER2-positive metastatic breast cancer (CLEOPATRA study): overall survival results from a randomised, double-blind, placebo-controlled, phase 3 study. Lancet Oncol 14, 461–471. https://doi.org/10.1016/S1470-2045(13)70130-X

Takis, P.G., Ghini, V., Tenori, L., Turano, P., Luchinat, C., 2019. Uniqueness of the NMR approach to metabolomics. TrAC Trends in Analytical Chemistry 120, 115300. https://doi.org/10.1016/j.trac.2018.10.036

Tan, B., Qiu, Y., Zou, X., Chen, T., Xie, G., Cheng, Y., Dong, T., Zhao, L., Feng, B., Hu, X., Xu, L.X., Zhao, A., Zhang, M., Cai, G., Cai, S., Zhou, Z., Zheng, M., Zhang, Y., Jia, W., 2013. Metabonomics Identifies Serum Metabolite Markers of Colorectal Cancer. J. Proteome Res. 12, 3000–3009. https://doi.org/10.1021/pr400337b

Tayyari, F., Gowda, G.A.N., Olopade, O.F., Berg, R., Yang, H.H., Lee, M.P., Ngwa, W.F., Mittal, S.K., Raftery, D., Mohammed, S.I., 2018. Metabolic profiles of triple-negative and luminal A breast cancer subtypes in African-American identify key metabolic differences. Oncotarget 9, 11677–11690. https://doi.org/10.18632/oncotarget.24433

Tenori, L., Oakman, C., Claudino, W.M., Bernini, P., Cappadona, S., Nepi, S., Biganzoli, L., Arbushites, M.C., Luchinat, C., Bertini, I., Di Leo, A., 2012. Exploration of serum metabolomic profiles and outcomes in women with metastatic breast cancer: A pilot study. Molecular Oncology 6, 437–444. https://doi.org/10.1016/j.molonc.2012.05.003

Tenori, L., Oakman, C., Morris, P.G., Gralka, E., Turner, N., Cappadona, S., Fornier, M., Hudis, C., Norton, L., Luchinat, C., Di Leo, A., 2015. 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. 9, 128–139. https://doi.org/10.1016/j.molonc.2014.07.012

Thompson, A.M., Moulder-Thompson, S.L., 2012. Neoadjuvant treatment of breast cancer. Ann Oncol 23, x231–x236. https://doi.org/10.1093/annonc/mds324

Turano, P., 2014. Colorectal cancer: the potential of metabolic fingerprinting. Expert Review of Gastroenterology & Hepatology 8, 847–849. https://doi.org/10.1586/17474124.2014.945912

Verma, S., Miles, D., Gianni, L., Krop, I.E., Welslau, M., Baselga, J., Pegram, M., Oh, D.-Y., Diéras, V., Guardino, E., Fang, L., Lu, M.W., Olsen, S., Blackwell, K., EMILIA Study Group, 2012. Trastuzumab emtansine for HER2-positive advanced breast cancer. N Engl J Med 367, 1783–1791. https://doi.org/10.1056/NEJMoa1209124

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

Vignoli, A., Mori, E., Di Donato, S., Malorni, L., Biagioni, C., Benelli, M., Calamai, V., Cantafio, S., Parnofiello, A., Baraghini, M., Garzi, A., Monte, F.D., Romagnoli, D., Migliaccio, I., Luchinat, C., Tenori, L., Biganzoli, L., 2021a. Exploring Serum NMR-Based Metabolomic Fingerprint of Colorectal Cancer Patients: Effects of Surgery and Possible Associations with Cancer Relapse. Applied Sciences 11, 11120. https://doi.org/10.3390/app112311120

Vignoli, A., Muraro, E., Miolo, G., Tenori, L., Turano, P., Di Gregorio, E., Steffan, A., Luchinat, C., Corona, G., 2020a. Effect of Estrogen Receptor Status on Circulatory Immune and Metabolomics Profiles of HER2-Positive Breast Cancer Patients Enrolled for Neoadjuvant Targeted Chemotherapy. Cancers 12, 314. https://doi.org/10.3390/cancers12020314

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

Vignoli, A., Risi, E., McCartney, A., Migliaccio, I., Moretti, E., Malorni, L., Luchinat, C., Biganzoli, L., Tenori, L., 2021b. Precision Oncology via NMR-Based Metabolomics: A Review on Breast Cancer. IJMS 22, 4687. https://doi.org/10.3390/ijms22094687

Vignoli, A., Santini, G., Tenori, L., Macis, G., Mores, N., Macagno, F., Pagano, F., Higenbottam, T., Luchinat, C., Montuschi, P., 2020b. NMR-Based Metabolomics for the Assessment of Inhaled Pharmacotherapy in Chronic Obstructive Pulmonary Disease Patients. J. Proteome Res. 19, 64–74. https://doi.org/10.1021/acs.jproteome.9b00345

Vignoli, A., Tenori, L., Morsiani, C., Turano, P., Capri, M., Luchinat, C., 2022. Serum or Plasma (and Which Plasma), That Is the Question. J. Proteome Res. 21, 1061–1072. https://doi.org/10.1021/acs.jproteome.1c00935

von Minckwitz, G., Huang, C.-S., Mano, M.S., Loibl, S., Mamounas, E.P., Untch, M., Wolmark, N., Rastogi, P., Schneeweiss, A., Redondo, A., Fischer, H.H., Jacot, W., Conlin, A.K., Arce-Salinas, C., Wapnir, I.L., Jackisch, C., DiGiovanna, M.P., Fasching, P.A., Crown, J.P., Wülfing, P., Shao, Z., Rota Caremoli, E., Wu, H., Lam, L.H., Tesarowski, D., Smitt, M., Douthwaite, H., Singel,

S.M., Geyer, C.E., KATHERINE Investigators, 2019. Trastuzumab Emtansine for Residual Invasive HER2-Positive Breast Cancer. N Engl J Med 380, 617–628. https://doi.org/10.1056/NEJMoa1814017

von Minckwitz, G., Procter, M., de Azambuja, E., Zardavas, D., Benyunes, M., Viale, G., Suter, T., Arahmani, A., Rouchet, N., Clark, E., Knott, A., Lang, I., Levy, C., Yardley, D.A., Bines, J., Gelber, R.D., Piccart, M., Baselga, J., APHINITY Steering Committee and Investigators, 2017. Adjuvant Pertuzumab and Trastuzumab in Early HER2-Positive Breast Cancer. N Engl J Med 377, 122–131. https://doi.org/10.1056/NEJMoa1703643

Wallner-Liebmann, S., Gralka, E., Tenori, L., Konrad, M., Hofmann, P., Dieber-Rotheneder, M., Turano, P., Luchinat, C., Zatloukal, K., 2015. The impact of free or standardized lifestyle and urine sampling protocol on metabolome recognition accuracy. Genes Nutr 10, 441. https://doi.org/10.1007/s12263-014-0441-3

Wallner-Liebmann, S., Tenori, L., Mazzoleni, A., Dieber-Rotheneder, M., Konrad, M., Hofmann, P., Luchinat, C., Turano, P., Zatloukal, K., 2016. Individual Human Metabolic Phenotype Analyzed by (1)H NMR of Saliva Samples. J. Proteome Res. 15, 1787–1793. https://doi.org/10.1021/acs.jproteome.5b01060

Wang, L., Tang, Y., Liu, S., Mao, S., Ling, Y., Liu, D., He, X., Wang, X., 2013. Metabonomic Profiling of Serum and Urine by 1H NMR-Based Spectroscopy Discriminates Patients with Chronic Obstructive Pulmonary Disease and Healthy Individuals. PLoS ONE 8, e65675. https://doi.org/10.1371/journal.pone.0065675

Wang, P., Shehu, A.I., Ma, X., 2017. The Opportunities of Metabolomics in Drug Safety Evaluation. Curr Pharmacol Rep 3, 10–15. https://doi.org/10.1007/s40495-016-0079-5

Wei, S., Liu, L., Zhang, J., Bowers, J., Gowda, G.A.N., Seeger, H., Fehm, T., Neubauer, H.J., Vogel, U., Clare, S.E., Raftery, D., 2013. Metabolomics approach for predicting response to neoadjuvant chemotherapy for breast cancer. Mol Oncol 7, 297–307. https://doi.org/10.1016/j.molonc.2012.10.003

Wiggans, R.G., Woolley, P.V., Smythe, T., Hoth, D., Macdonald, J.S., Green, L., Schein, P.S., 1979. Phase-II trial of tamoxifen in advanced breat cancer. Cancer Chemother Pharmacol 3, 45–48. https://doi.org/10.1007/BF00254419

Winnike, J.H., Busby, M.G., Watkins, P.B., O'Connell, T.M., 2009. Effects of a prolonged standardized diet on normalizing the human metabolome. Am. J. Clin. Nutr. 90, 1496–1501. https://doi.org/10.3945/ajcn.2009.28234

Wojtowicz, W., Wróbel, A., Pyziak, K., Tarkowski, R., Balcerzak, A., Bębenek, M., Młynarz, P., 2020. Evaluation of MDA-MB-468 Cell Culture Media Analysis in Predicting Triple-Negative Breast Cancer Patient Sera Metabolic Profiles. Metabolites 10. https://doi.org/10.3390/metabo10050173

World Health Organization, 2021. Global progress report on HIV, viral hepatitis and sexually transmitted infections, 2021: accountability for the global health sector strategies 2016–2021: actions for impact: web annex 1: key data at a glance. World Health Organization, Geneva.

Xi, Y., Xu, P., 2021. Global colorectal cancer burden in 2020 and projections to 2040. Translational Oncology 14, 101174. https://doi.org/10.1016/j.tranon.2021.101174

Yafi, F.A., Jenkins, L., Albersen, M., Corona, G., Isidori, A.M., Goldfarb, S., Maggi, M., Nelson, C.J., Parish, S., Salonia, A., Tan, R., Mulhall, J.P., Hellstrom, W.J.G., 2016. Erectile dysfunction. Nat Rev Dis Primers 2, 16003. https://doi.org/10.1038/nrdp.2016.3

Ząbek, A., Stanimirova, I., Deja, S., Barg, W., Kowal, A., Korzeniewska, A., Orczyk-Pawiłowicz, M., Baranowski, D., Gdaniec, Z., Jankowska, R., Młynarz, P., 2015. Fusion of the 1H NMR data of serum, urine and exhaled breath condensate in order to discriminate chronic obstructive pulmonary disease and obstructive sleep apnea syndrome. Metabolomics 11, 1563–1574. https://doi.org/10.1007/s11306-015-0808-5

Zhang, X., Zhao, X.-W., Liu, D.-B., Han, C.-Z., Du, L.-L., Jing, J.-X., Wang, Y., 2014. Lipid levels in serum and cancerous tissues of colorectal cancer patients. World J Gastroenterol 20, 8646–8652. https://doi.org/10.3748/wjg.v20.i26.8646

Zhou, J., Wang, Y., Zhang, X., 2017. Metabonomics studies on serum and urine of patients with breast cancer using 1 H-NMR spectroscopy. Oncotarget 5. https://doi.org/10.18632/oncotarget.16210

Zhu, J., Djukovic, D., Deng, L., Gu, H., Himmati, F., Chiorean, E.G., Raftery, D., 2014. Colorectal Cancer Detection Using Targeted Serum Metabolic Profiling. J. Proteome Res. 13, 4120–4130. https://doi.org/10.1021/pr500494u

4.2 New strategies for NMR-based metabolomics data analysis

4.2.1 NMR-based metabolomics to indirectly quantify the chemical and sensioral profiles of olive oils

Meoni G.[1,2], Di Cesare F.[1], Tenori L. [1,2,3], Brizzolara S.[4], Tonutti P. [4], Luchinat C. [1,2,3]

[1]Magnetic Resonance Center (CERM), University of Florence, 50019 Sesto Fiorentino, Italy;

[2] Department of Chemistry, University of Florence, 50019 Sesto Fiorentino, Italy;

[3] Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine (CIRMMP), 50019 Sesto Fiorentino, Italy

[4] Institute of Life Sciences, Scuola Superiore Sant'Anna, Pisa, Italy, Piazza Martiri della Libertà 33,56127 Pisa, Italy

In preparation

Candidate's contribuitions: Sample preparation, acquisition and pre-processing of NMR data, statistical analysis, interpretation of results, and writing the manuscript.

Abstract

Olive oil quality is assessed by determining several parameters. Multivariate analysis was used to streamline the task due to the huge number of tests required for this verification as well as the drawbacks of using toxic solvents, the use of a lot of oil samples, waste generation, and a longer execution time. In this work, 49 significant olive oil quality parameters (chemical and sensorial analyses) were simultaneously determined using ^{1}H-NMR spectra to build multivariate models using Random Forest linear regressions. The models were also used to forecast all parameters for external samples, allowing us to ascertain how these values alter as olive oil ages, how these values change according to two of the most common Tuscany cultivars, Moraiolo and Leccino, and finally as a result of post-harvest practices like filtering. These results corroborate the goodness of the model prediction and demonstrate that ^{1}H-NMR spectra can also be used to simultaneously detect information about olive oil quality, significantly reducing analysis time, reagent, olive oil and solvent consumption, and waste generation.

1 Introduction

Olive oil (*Olea europea L.*) has been consumed by humans since antiquity and it remains a highly valued food today[1,2]. Due to the sensory and nutritional quality, there is a growing interest in extra virgin olive oils in the world market. Since 1991 international regulations have established analytical criteria to define olive oil genuineness and quality grade (European communities 1991, International olive oil council 1995, Characterization of Italian extra virgin olive oils using ^{1}H-NMR spectroscopy, 1998).

Generally, the assessment of olive oil grade is carried out by standard analytical methods[3,4]. These methods are often time-consuming, elaborate, and expensive, and require a large number of samples and solvents. As a fast and viable alternative is represented by ^{1}H-NMR spectroscopy combined with chemometrics[5–8] and this approach has been proposed in this study. Chemometric linear regression technique as Random Forest (RF) is used to highlight the correlations between the NMR spectra and the parameters of interest. We recorded the proton NMR spectra of olive oil with a 400 MHz spectrometer. RF linear regression models for prediction were built using the following analytical and sensorial data: UV spectrophotometric indices (K232, K268), acidity, peroxide value, the composition of fatty acids, tocopherols, the composition of biophenols, total biophenols, bitter, sweet, spicy, gustatory intense, olfactory fruity, olfactory intense, and overall pleasantness.

Therefore, in this work ^{1}H-NMR coupled with the RF regression technique is used to perform the simultaneous determination of 49 important olive oil quality parameters. In this context, considering a large number of analyses, as well as the use of toxic solvents and waste generation, our approach is suggested to reduce the time of the analyses, the number of samples, and the waste production.

Despite the considerable cost of an NMR spectrometer, many advantages are using this technique[9,10]. In particular, measurements can be performed quickly and simultaneously for multiple analytes present in a complex mixture.

Moreover, the models were also used to forecast chemical and sensorial parameters for external olive oil samples.

Using these predicted values, we attempted to describe: i) what happens in olive oil during one year of aging; ii) how different in terms of composition two of the most common Italian olive oil cultivars (Moraiolo cv and Leccino cv), iii) how different are filtered and not-filtered olive oils in terms of the predicted parameters.

2 Materials and methods

2.1Olive oil samples

n=221 bottles of olive oil were purchased from different Italian productores. All olive oil bottles were stored in the original dark green glass bottles at the same temperatures and under typical house-lighting conditions. In parallel, to evaluate the olive oil aging effect, a total of 49 samples were followed for one year. In particular, three spectra were recorded for each of these samples at different time points: t_0 at the opening of the bottle; t_1 after 4 months; t_2 after 12 months.

2.2 Chemical analyses

The free acidity, peroxide value, UV spectrophotometric indices (K232, K268), and the determination of composition of fatty acids of the olive oil samples were determined according to the analytical methods (EEC Reg. 2568/1991). The tocopherols of the olive oil samples were determined according to ISO 9936:2006/Corr. 1:2008 analytical method (ISO 9936/2008).

2.2.1 Determination of biophenols in olive oils by High-performance liquid chromatography (HPLC)

Two grams of olive oil were added to a plastic test tube together with 1 mL of the internal standard (syringic acid, 1.504 mg mL−1 in a MeOH/H2O 80:20 solution) and 5 mL of an EtOH/H2O 80:20 solution. The mixture was homogenized by whirl for 1 min, then sonicated at room temperature for 15 min in the ultrasonic bath. The mixture was then centrifuged at 5000 rev/min for 25 min. An aliquot of the supernatant was then taken through a 5mL plastic syringe and filtered using 0.45 μm PVDF filter into a 1.5 mL vial. The solution thus obtained was immediately used for the chromatographic analysis. The chromatographic analysis was performed by an HP 1100 Liquid Chromatograph equipped with 1100 Autosampler, column heater module, quaternary pump, coupled with DAD and MS detectors, and an HP 1100 MSD API-electrospray as an interface (all from Agilent Technologies, Palo Alto, AC, USA). The column was a Hypersil Gold QRP-18 (4.6 mm internal diameter, 250 mm length; particle size 3 μm, Thermo Electron Corporation, Austin, Texas), equipped with a 10 × 4 mm pre-column of the same phase and maintained at 30 °C. Elution was performed at a flow rate of 0.8 mL min−1 using H_2O at pH 3.2 by formic acid (A), acetonitrile (B), and methanol (C). The three-step linear gradient of both B and C changed as follows: from 2.5% to 27.5% in 45 min and increased to 50% in 10 min. Isocratic elution with 50% of B and 50% of C was then maintained for 5 min. All solvents used were of HPLC grade.

Total phenolic compounds were quantified according to the COI/T.20/Doc. no. 29 analytical method (COI/T.20/Doc.29, 2009).

2.2.2 Sensory evaluation

Sensory evaluation of olive oil was performed according to the European Union (EU) official method[11]. However, the number of descriptors on the official profile sheet was increased to obtain a detailed description of perceptions. Six trained tasters were requested to evaluate olive oil samples. A total of *n*=8 characteristic tastes were considered in this study: bitter, sweet, spicy, gustatory intense, olfactory fruity, olfactory intense, and overall pleasantness.

2.3 ^{1}H NMR analyses

For sample preparation, approximately 60 mg of olive oil was dissolved in 600 μL of deuterated chloroform (CDCl3). Into a 5 mm NMR tube 600 μL of the prepared mixture was transferred. NMR spectra acquisition was performed using an AVANCE III 400 MHz Bruker spectrometer working at 300K. For each olive oil sample, the following ^{1}H-NMR experiments were performed: i) ZG1H: a standard single pulse experiment, *i.e.* RD-P(90°)-acquisition of the free induction decay (FID); ii) NOESYGPPS: a one-dimensional 1H-NMR pulse sequence with strong saturating signals suppression. For more information about the experimental parameters used for both NMR experiments, we refer the reader to Dourou et *al*.[5].

The phase and baseline corrections were automatically performed using the NMR processing software Topspin (version 3.5 pl 7 Bruker BioSpin Srl). Spectra were aligned to 0.7 ppm, in correspondence of β-sitosterol singlet. Before proceeding with the statistical analysis, each 1D-NMR spectrum was segmented into 0.02 chemical bins and the integration of the corresponding spectral areas were performed using AssureNMR software (version 3.8.4, Bruker BioSpin Srl). This procedure, called binning or bucketing, is an NMR pre-processing method for the reduction of the number of total variables and for the compensation of the small shifts in the signals, necessary to generate more robust and reproducible NMR data. The binned matrix of ZG1H spectra was obtained considering whole spectra (from 0.2 to 11.9 ppm) except for chloroform residual signal (7.24 ppm) resulting in a system of 580 bins. Binning of NOESYGPPS spectra (from 0.2 to 11.9 ppm) was obtained by removing all suppressed saturating signals and the dimension of the system was reduced to 515 bins. To obtain a complete informative spectrum, only the bins from 11.90 to 5.62 of the NOESYGPPS experiment and the bins from 6.20 to 0.30 of the ZG1H experiment were considered and merged. Each resulting binned spectrum was scaled according to the oil weight measured

during the sample preparation. Probabilistic Quotient Normalization (PQN)[12] was applied concluding the 1D spectral processing phase.

2.4 NMR molecular assignment and quantification

41 molecular features in olive oil samples were correctly assigned in all spectra performed using a library of NMR spectra of pure organic compounds, public databases[13] (*i.e.* FooDB, PhytoHub, PhenolExplorer, etc.) storing reference, and literature data. Each NMR-identified metabolite was aligned to a reference value of chemical shift, obtaining a perfect alignment among all the spectra. The quantification of the assigned metabolites was performed directly by integrating the signals in the spectra in a defined spectral range, using a house-developed tool. For completeness, the 41 metabolites assigned and quantified are presented in Supplementary Table 1, reporting the spectral region of the assignment.

2.5 Statistical analysis

To correctly and robustly predict n=42 analytical variables and n=7 sensorial characteristics, using the ^{1}H-NMR information, 100-times Random Forest (RF) regression models[14–17], using a total of 1000 trees per model, were performed. For each analytical and sensorial variable predicted (precisely, 100 values per olive oil spectrum), the final prediction value was calculated by performing an average among the k=100 prediction obtained for each specific model. Pearson R^2 [18] was calculated by comparing the real value and the 100-times average predicted values. The same rationale was applied to identify which ^{1}H-NMR variables mostly contributed to correctly predicting that specific analytical and/or sensorial characteristic. In particular, the variable importance was calculated by extracting the mean decrease Gini coefficients of all ^{1}H-NMR variables per k=100 RF model. Subsequently, the mean decrease Gini coefficients were ranked where the highest coefficients corresponded with the highest ranks coefficients and the lowest coefficients with the lowest ranks. To obtain a unique and univocal value representing the most incident NMR variables, the 100-times ranks obtained for each model performed were averaged.

To correctly classify the different cultivars (Moraiolo *vs* Leccino) and the different post-harvest procedures (filtered *vs* non-filtered), classification RF models[14–17], using a total of 1000 trees per model, were performed on both bucketed olive oil spectra and on the matrix of previously correctly predicted (using 100-times RF regression models) analytical and sensorial profiles.

The univariate analysis was performed to evaluate metabolic (both predicted analytical and direct NMR integrated metabolic profiles) and sensorial differences in: i) olive oils collected in three-distinct timepoints, to evaluate the effect of natural oxidation in the olive oil aging process; ii) olive oils of two different cultivars: Moraiolo and Leccino; iii) olive oils

treated with two different post-harvest procedure: filtering and non-filtering. The Friedman test followed by post-hoc Nemenyi analysis[19,20] was chosen to infer differences between more than two classes of samples (case i) described above) and the Wilcoxon test[21] was chosen to infer pairwise differences between two classes of samples (case ii) and iii) described above). Benjamini- Hochberg method was used to correct for multiple testing[22] and FDR-adjusted *P*-values < 0.05 were considered statistically significant.

2.6 Software

All analyses were performed with R (version 4.1.0) software[23]. The randomForest package [14,15] was used to build Random Forest models. To perform The Friedman test followed by post-hoc Nemenyi analysis the "frdAllPairsNemenyiTest" function of PMCMRplus[24] R package was used. All plots were obtained using the ggplot2[25] R package.

3 Results and discussion

3.1 Prediction of chemical analyses

It is well accepted that olive oil (OO), the principal product of olive (*Olea europaea L.*, *Oleaceae*) and a crucial component of the Mediterranean diet[1], has significant nutritional and health benefits. Its importance is ascribed to its high proportion of monounsaturated and polyunsaturated fatty acids (MUFA and PUFA, respectively)[26], as well as minor polar compounds with a strong antioxidant profile[27,28]. In this scenario, the aim was to evaluate if, using the information contained in the entire olive oil NMR spectra, we were able to correctly indirectly quantify these relevant compounds that characterize the olive oil chemical profile.

The 100-times RF regression models were used to relate the ^{1}H-NMR bucketed spectra with the analytical results of several important chemical parameters of olive oil. Models were constructed for each of the n=42 responses separately. As described in the materials and methods section, for each olive oil spectrum, the average predicted value of chemical parameters was obtained by the k=100 independent RF regression models built.

For each analytical parameter, the mean of spectrum-specific real and correctly average predicted values, the standard deviation (sd), the Pearson's R^2, the *P*-values, the regression equation, the Root-mean-square deviation (RMSE), and the Error ((RMSE/mean$_{\text{analtycal parameter}}$)*100) expressed in percentage are reported in Table 1.

The regulated physiochemical quality parameters (free acidity, peroxide value), and UV absorption characteristics (K268 and K232) were predicted from ^{1}H-NMR bucketed spectra with a resulting R^2 of 0.98, 0.91, 0.89, and 0.89, respectively, while the %Error, excepting for the peroxide index, is minor of 15% (Table 1, SF1). Each fatty acid component has been also analysed. Oleic acid (C18:1), the most abundant fatty acid in olive oil (mean of 73.03% m/m in the olive oil samples collected for this study), fits with an R^2 of 0.99 and a %Error of 0.38, indicating that this value can be optimally predicted using NMR spectra. As well, linoleic acid (C18:2), the second most abundant fatty acid (8.18%m/m), shows an R^2 of 0.99 and a %Error of 1.82. Optimal data can be obtained by the RF linear regression built to predict the concentration of palmitic acid (C16:0) resulting in a regression coefficient value of 0.99 and a %Error value of 1.67. The other less abundant fatty acids present in olive oil show, in any case, good fitting, resulting in an R^2>0.80 and a %Error value < 5%, except for the Lignoceric acid (C24:0) and Margaric acid (C17:0), with a %Error of 18.75 and 12.20 respectively (Table 1).

RF statistics for total biophenols and for the different biophenol components are also examined (see Table 1, SF1). The sum of biophenols detectable in olive oil is composed mainly of simple phenolic acids, phenyl alcohols, secoiridoids, and flavonoids. The main phenolic alcohols identified in virgin olive oil are hydroxytyrosol and tyrosol. Their concentrations are

usually low in fresh samples but proportionally increase with advanced storage time due to the hydrolysis of secoiridoids (oleuropein and ligstroside aglycones)[3]. The phenyl alcohols tyrosol and hydroxytyrosol, together with the secoiridoids, constitute approximately 90% of the total biophenols of olive oil[28]. Hydroxytyrosol and tyrosol models show an optimal R^2 of 0.99 with a %Error of 30.93 and 28.25 respectively. Amongst secoiridoids, several compounds are listed and are described generally as oxidized and/or hydroxylic forms (*i.e.*, decarboxymethyl oleuropein aglycone, oxidized dialdehyde form; decarboxymethyl ligstroside aglycone, oxidized dialdehyde form; oleuropein aglycone, oxidized aldehyde, and hydroxylic form; ligstroside aglycone, oxidized aldehyde, and hydroxylic form). In general, all of them show excellent R^2, however, the predicted values of the aldehyde hydroxylic form of oleuropein and the oxidized aldehydic form of hydroxylic ligstroside (R^2 0.94 and 0.92, respectively) are characterized by high %Error values (>20). The total biophenol fraction RF model shows an R^2 of 0.99 and a %Error of 4.38. The total tocopherol fraction shows the perfect fitting (R^2 1.00), with a %Error of 3.28.

Other polyphenols considered were: vanilline, *p*- and *o*- coumaric acid, hydroxytyrosol acetate, ferulic acid, pinoresinol, cinnamic acid, luteolin, apigenin, and methyl luteolin. In this case, variables are very well predicted (R^2>0.90) but the %Error values were about 20/25.

To evaluate the spectral regions that mainly determine the good RF regression models prediction, the variable importance, considering the mean ranks based on the mean decrease Gini index, was also considered. In figure 1, for each analytical variable predicted, the top 5 average ranks NMR spectra regions were reported.

3.2 Prediction of sensory analysis

The same approach was also used to relate the ^{1}H-NMR bucketed spectra with the sensory profiles.

In Table 2 are reported the RF linear regression results for each model and the corresponding plots are reported in SF2. Independent models were built on 42 different olive oil samples for each sensory descriptor: bitter, sweet, and spicy taste, gustatory intensity, fruity olfactory intensity, and overall pleasantness. Other sensory attributes (rancid, fruity, heated, winey, metallic, and frostbitten olive) were excluded because of missing information (only 22 samples out of 42). Interestingly, using the NMR-based information we observed that the regression coefficients are optimal (R^2>0.97) and the calculated %Error values are about 15, suggesting that, using the entire olive oil spectra, we can correctly predict the sensorial profile.

The same rationale applied in §3.1 was also considered here. To highlight the spectral regions that mainly determine the good RF regression models prediction, the variable importance, considering the average ranks based on the mean decrease Gini index, was also considered. In Figure 2 for each sensorial variable predicted the top 5 average ranks NMR spectra regions were reported.

3.3 Characterization of the olive oil aging process using predicted values and molecules identified using the 1H-NMR spectra

With the aim to evaluate the effect of natural olive oil aging on the metabolic and sensorial profiles of olive oils, a total of 49 samples were followed for one year. In particular, three spectra were recorded for each of these samples at different time points: the first at t_0 at the opening of the bottle; the second after 4 months, t_1; the third after 12 months, t_2. The Friedman test followed by post-hoc Nemenyi analysis was chosen to ascertain pairwise differences among three different times collections of olive oil samples. Figure 1 shows the boxplot of the statistically significant (adjusted *P*-value<0.05) molecules, directly and indirectly, quantified using NMR spectra, and the olive oil sensory values.

Monitoring molecules' trends during a time course, suggests an increase in production of the following parameters: K 232, K 268, lignoceric acid (C24:0), margaroleic acid (C17:1), peroxide index (Figure 3 A) – all of them indirectly quantified (samples predicted on the best performing RF linear regression models) –, and hydroperoxides derivates (1), (2), and (3) – directly quantified (Figure 1 B). Indeed, a statistically significant decrement in terms of molecular features is significantly statistically detectable for methyl luteolin, oleuropein – indirectly quantified –, 1,2 diglycerides, alkenals, (E)-2-alkenals, and aldehyde – directly quantified.

The aging process is also connected to an alteration of sensorial profile. In particular, we observed an increase in a sweet taste and a significant decrease in olfactory and gustatory intensity and bitter taste (Figure 3 C). As known, olive oil is highly subject, as a result of oxidation phenomena, to alterations of the entire organoleptic profile and, interestingly, the age-dependent chemical and sensorial profile alterations observed in this study are totally in line with the results reported in the literature[29–31].

3.4 Characterization of Moraiolo and Leccino cultivars

The RF classification model was performed to ascertain if it was possible to distinguish and classify Moraiolo and Leccino cultivars using the entire NMR spectra (fingerprinting approach) or using the matrix of the quantified and predicted molecules and the predicted sensorial profiles. This analysis aims to verify if we can rapidly determine the olive oil cultivar using our data. It might be interesting to be able to guarantee, with a single analysis, using a

whole spectrum or a pool of molecules, the exact cultivar of a monovarietal olive oil. To the best of our knowledge, we are the first that used this approach to identify the chemical and sensorial differences between the two important Italian olive oil cultivars[5,32–34].

The RF models have been built starting from a group of 221 olive oil samples (83 different olive oil samples: 69x3 independent replicates and 14 single samples) for which the cultivar has been confirmed by genetic analysis harvested and acquired in three different years: 2019 (26 samples: 14 Leccino cv., and 12 Moraiolo cv.); 2020 (19 samples: 10 Leccino cv., and 9 Moraiolo cv.); and 2021 (38 samples: 17 Leccino cv., and 21 Moraiolo cv.). Therefore, the models have been built only on these two cultivars: Leccino (*n*=109) and Moraiolo (*n*=112). ^{1}H-NMR bucketed spectra consisting of a matrix of 551x221 dimensions were used to build a fingerprint model with the aim of correctly classifying ^{1}H-NMR olive oil spectra according to the cultivar. The resulting overall predictive accuracy is 75.9% (Table 3). Instead, using the matrix (dim: 221x91) of identified, and predicted molecules and sensory results, provides an overall predictive accuracy of 80.4% (Table 4). Both RF classification models have been built by iteratively and randomly removing from the training set all the 2 replicates belonging to the same olive oil samples, to avoid a classification bias due to the similarity of the samples.

The Wilcoxon univariate analysis was then performed to determine the discriminant molecules between the two cultivars (Figure 4). We observed that Moraiolo tends to have higher concentrations of free acidity, arachic acid (C20:0), behenic acid (C22:0), dialdehyde aglicone ligstroside, eicosanoic acid (C20:1), lignoceric acid (C24:0), luteolin, oleic acid (C18:1), oxidized aldehydic form of hydroxylic ligstoside aglycon, oxidized aldehydic form of hydroxylic oleuropein aglycon – indirectly quantified (Figure 4 A) –,1,2 diglycerides/triolein, and squalene – directly quantified (Figure 4 B).

In contrast, Leccino cultivar is characterized by higher levels of dialdehyde decarboxymethyl ligstroside aglycon, dialdehyde decarboxymethyl oleuropein aglycon, linoleic acid (C18:2), palmitic acid (C16:0), palmitoleic acid (C16:1), stearic acid (C18:0), total tocopherols, vanilline – indirectly quantified (Figure 4 A)–, 4-hydroxy-trans-alkenals, aldehydic forms of oleuropein and ligstroside, beta-sitosterol, cycloartenol, cyclopropane, hydroperoxides derivates (1), linoleic acid, linoneyl/linolenyl, terpene (1), terpene (2), and terpene (3), – directly quantified (Figure 4 B)[35].

Moraiolo, compared to Leccino, results to be fruity, and less spicy, and is characterized by a more bitter taste (Figure 4 C)[36,37].

3.5 Characterization of filtered and not-filtred olive oil samples.

The same approach of §3.4 was also applied to filtered and not-filtered olive oil samples, with the aim of correctly classifying samples according to these two different post-harvest procedures. RF classification models have been built on bucketed spectra and on identified (in direct and indirect ways) molecules and predicted sensorial profile of 12 not-filtered olive oil samples and 10 filtered olive oil samples. In this case, triplicates were not present in the dataset, and therefore all samples have been used to build the RF models. The resulting accuracy of the RF models built on bucketed spectra is 84.8% (Table 5) while the RF models built on olive oil's overall metabolic and sensorial profile is 75.8% (Table 6).

The Wilcoxon univariate analysis on olive oil molecules was performed to determine the change detectable in the olive oils according to the filtration step. As reported in violin plots of Figure 3, filtered olive oils tend to have higher levels of cinnamic acid, dialdehyde decarboxymthyl ligstroside aglycon – indirectly quantified (Figure 5 A) –, (E)-2-alkenals, aldehydic forms of oleuropein and ligstroside, linolenyl, and aldehydes – directly quantified (Figure 5 B). In contrast, non-filtered olive oils tend to have higher free acidity and K268 and higher levels of hydroxytyrosol, hydroxytyrosol acetate, luteoline, and tyrosol – indirectly quantified (Figure 5 A).

Regarding the sensorial profile, filtered olive oils tend to have a more spicy, bitter, and olfactory fruity taste than non-filtered oils (Figure 5 B)[38–40].

4 Conclusions

In this study we have developed a robust statistical approach to indirectly derive, starting only from the information contained in the ^{1}H-NMR spectrum, both the specific chemical and sensory parameters of olive oil. The Random Forest regression predictive models, iterated 100 times for each variable considered, determine optimal ($R^2 > 0.98$) and good ($R^2 < 0.98$) prediction of 49 important olive oil quality parameters. Using these predicted values, combined with the directly assigned and quantified compounds in the ^{1}H-NMR spectra, we highlighted statistically significant differences in terms of both chemical and sensorial profiles considering the olive oil aging phenomenon, considering two different commonly used Italian olive oil cultivars (Leccino cv and Moraiolo), and, lastly, considering two different post-harvest procedures (filtration and non-filtration).

References

(1) Grigg, D. Olive Oil, the Mediterranean and the World. *GeoJournal* **2001**, *53* (2), 163–172. https://doi.org/10.1023/A:1015702327546.

(2) Uylaşer, V.; Yildiz, G. The Historical Development and Nutritional Importance of Olive and Olive Oil Constituted an Important Part of the Mediterranean Diet. *Critical Reviews in Food Science and Nutrition* **2014**, *54* (8), 1092–1101. https://doi.org/10.1080/10408398.2011.626874.

(3) Di Lecce, G.; Piochi, M.; Pacetti, D.; Frega, N. G.; Bartolucci, E.; Scortichini, S.; Fiorini, D. Eleven Monovarietal Extra Virgin Olive Oils from Olives Grown and Processed under the Same Conditions: Effect of the Cultivar on the Chemical Composition and Sensory Traits. *Foods* **2020**, *9* (7), 904. https://doi.org/10.3390/foods9070904.

(4) Tena, N.; Wang, S. C.; Aparicio-Ruiz, R.; García-González, D. L.; Aparicio, R. In-Depth Assessment of Analytical Methods for Olive Oil Purity, Safety, and Quality Characterization. *J. Agric. Food Chem.* **2015**, *63* (18), 4509–4526. https://doi.org/10.1021/jf5062265.

(5) Dourou, A.-M.; Brizzolara, S.; Meoni, G.; Tenori, L.; Famiani, F.; Luchinat, C.; Tonutti, P. The Inner Temperature of the Olives (Cv. Leccino) before Processing Affects the Volatile Profile and the Composition of the Oil. *Food Research International* **2020**, *129*, 108861. https://doi.org/10.1016/j.foodres.2019.108861.

(6) Longobardi, F.; Ventrella, A.; Napoli, C.; Humpfer, E.; Schütz, B.; Schäfer, H.; Kontominas, M. G.; Sacco, A. Classification of Olive Oils According to Geographical Origin by Using 1H NMR Fingerprinting Combined with Multivariate Analysis. *Food Chemistry* **2012**, *130* (1), 177–183. https://doi.org/10.1016/j.foodchem.2011.06.045.

(7) Maestrello, V.; Solovyev, P.; Bontempo, L.; Mannina, L.; Camin, F. Nuclear Magnetic Resonance Spectroscopy in Extra Virgin Olive Oil Authentication. *Comprehensive Reviews in Food Science and Food Safety n/a* (n/a). https://doi.org/10.1111/1541-4337.13005.

(8) Mannina, L.; Sobolev, A. P. High Resolution NMR Characterization of Olive Oils in Terms of Quality, Authenticity and Geographical Origin. *Magnetic Resonance in Chemistry* **2011**, *49* (S1), S3–S11. https://doi.org/10.1002/mrc.2856.

(9) Vignoli, A.; Ghini, V.; Meoni, G.; Licari, C.; Takis, P. G.; Tenori, L.; Turano, P.; Luchinat, C. High-Throughput Metabolomics by 1D NMR. *Angewandte Chemie International Edition* **2019**, *58* (4), 968–994. https://doi.org/10.1002/anie.201804736.

(10) 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.

(11) *Commission Delegated Regulation (EU) 2015/1830 of 8 July 2015 Amending Regulation (EEC) No 2568/91 on the Characteristics of Olive Oil and Olive-Residue Oil and on the Relevant Methods of Analysis*; 2015; Vol. 266.

(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. *Anal. Chem.* **2006**, *78* (13), 4281–4290. https://doi.org/10.1021/ac051632c.

(13) Naveja, J. J.; Rico-Hidalgo, M. P.; Medina-Franco, J. L. Analysis of a Large Food Chemical Database: Chemical Space, Diversity, and Complexity. *F1000Res* **2018**, *7*, Chem Inf Sci-993. https://doi.org/10.12688/f1000research.15440.2.

(14) Ali, J.; Khan, R.; Ahmad, N.; Maqsood, I. Random Forests and Decision Trees. *International Journal of Computer Science Issues(IJCSI)* **2012**, *9*.

(15) Breiman, L. Random Forests. *Machine Learning* **2001**, *45* (1), 5–32. https://doi.org/10.1023/A:1010933404324.

(16) Amaratunga, D.; Cabrera, J.; Lee, Y.-S. Enriched Random Forests. *Bioinformatics* **2008**, *24* (18), 2010–2014. https://doi.org/10.1093/bioinformatics/btn356.

(17) Nguyen, J.-M.; Jézéquel, P.; Gillois, P.; Silva, L.; Azzouz, F. B.; Lambert-Lacroix, S.; Juin, P.; Campone, M.; Gaultier, A.; Moreau-Gaudry, A.; Antonioli, D. Random Forest of Perfect Trees: Concept, Performance, Applications and Perspectives. *Bioinformatics* **2021**, *37* (15), 2165. https://doi.org/10.1093/BIOINFORMATICS/BTAB074.

(18) Hauke, J.; Kossowski, T. Comparison of Values of Pearson's and Spearman's Correlation Coefficients on the Same Sets of Data. *Quaestiones Geographicae* **2011**, *30* (2), 87–93. https://doi.org/10.2478/v10117-011-0021-1.

(19) Friedman, M. A Comparison of Alternative Tests of Significance for the Problem of m Rankings. *The Annals of Mathematical Statistics* **1940**, *11* (1), 86–92. https://doi.org/10.1214/aoms/1177731944.

(20) García, S.; Herrera, F. An Extension on "Statistical Comparisons of Classifiers over Multiple Data Sets" for All Pairwise Comparisons. *Journal of Machine Learning Research* **2008**, *9* (89), 2677–2694.

(21) Wilcoxon, F. Individual Comparisons by Ranking Methods. *Biometrics Bulletin* **1945**, *1* (6), 80–83. https://doi.org/10.2307/3001968.

(22) 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**, *57* (1), 289–300.
(23) *RStudio | Open source & professional software for data science teams*. https://www.rstudio.com/ (accessed 2022-06-14).
(24) Pohlert, T. PMCMRplus: Calculate Pairwise Multiple Comparisons of Mean Rank Sums Extended, 2022.
(25) Villanueva, R. A. M.; Chen, Z. J. Ggplot2: Elegant Graphics for Data Analysis (2nd Ed.). *Measurement: Interdisciplinary Research and Perspectives* **2019**, *17* (3), 160–167. https://doi.org/10.1080/15366367.2019.1565254.
(26) Dais, P.; Hatzakis, E. Quality Assessment and Authentication of Virgin Olive Oil by NMR Spectroscopy: A Critical Review. *Analytica Chimica Acta* **2013**, *765*, 1–27. https://doi.org/10.1016/j.aca.2012.12.003.
(27) Ray, N. B.; Hilsabeck, K. D.; Pitsillou, E.; Mann, A.; Karagiannis, T. C.; McCord, D. E. Chapter 13 - Olive Fruit and Olive Oil Bioactive Polyphenols in the Promotion of Health. In *Functional Foods and Nutraceuticals in Metabolic and Non-Communicable Diseases*; Singh, R. B., Watanabe, S., Isaza, A. A., Eds.; Academic Press, 2022; pp 203–220. https://doi.org/10.1016/B978-0-12-819815-5.00014-8.
(28) Antoniadi, L.; Angelis, A.; Stathopoulos, P.; Bata, E.-M.; Papoutsaki, Z.; Halabalaki, M.; Skaltsounis, L. A. Oxidized Forms of Olive Oil Secoiridoids: Semisynthesis, Identification and Correlation with Quality Parameters. *Planta Med* **2022**, *88* (9/10), 805–813. https://doi.org/10.1055/a-1806-7815.
(29) Rodrigues, N.; Oliveira, L.; Mendanha, L.; Sebti, M.; Dias, L. G.; Oueslati, S.; Veloso, A. C. A.; Pereira, J. A.; Peres, A. M. Olive Oil Quality and Sensory Changes During House-Use Simulation and Temporal Assessment Using an Electronic Tongue. *Journal of the American Oil Chemists' Society* **2018**, *95* (9), 1121–1137. https://doi.org/10.1002/aocs.12093.
(30) Mousavi, S.; Mariotti, R.; Stanzione, V.; Pandolfi, S.; Mastio, V.; Baldoni, L.; Cultrera, N. G. M. Evolution of Extra Virgin Olive Oil Quality under Different Storage Conditions. *Foods* **2021**, *10* (8), 1945. https://doi.org/10.3390/foods10081945.
(31) Shendi, E. G.; Ozay, D. S.; Ozkaya, M. T.; Ustunel, N. F. Changes Occurring in Chemical Composition and Oxidative Stability of Virgin Olive Oil during Storage. *OCL* **2018**, *25* (6), A602. https://doi.org/10.1051/ocl/2018052.
(32) Migliorini, M.; Cherubini, C.; Mugelli, M.; Gianni, G.; Trapani, S.; Zanoni, B. Relationship between the Oil and Sugar Content in Olive Oil Fruits from Moraiolo and Leccino Cultivars during Ripening. *Scientia Horticulturae* **2011**, *129* (4), 919–921. https://doi.org/10.1016/j.scienta.2011.05.023.
(33) Xie, P.; Cecchi, L.; Bellumori, M.; Balli, D.; Giovannelli, L.; Huang, L.; Mulinacci, N. Phenolic Compounds and Triterpenes in Different Olive Tissues and Olive Oil By-Products, and Cytotoxicity on Human Colorectal Cancer Cells: The Case of Frantoio, Moraiolo and Leccino Cultivars (Olea Europaea L.). *Foods* **2021**, *10* (11), 2823. https://doi.org/10.3390/foods10112823.
(34) Astolfi, M. L.; Marini, F.; Frezzini, M. A.; Massimi, L.; Capriotti, A. L.; Montone, C. M.; Canepari, S. Multielement Characterization and Antioxidant Activity of Italian Extra-Virgin Olive Oils. *Frontiers in Chemistry* **2021**, *9*.
(35) Jukić Špika, M.; Perica, S.; Žanetić, M.; Škevin, D. Virgin Olive Oil Phenols, Fatty Acid Composition and Sensory Profile: Can Cultivar Overpower Environmental and Ripening Effect? *Antioxidants (Basel)* **2021**, *10* (5), 689. https://doi.org/10.3390/antiox10050689.
(36) Monteleone, E.; Langstaff, S. *Olive Oil Sensory Science*; John Wiley & Sons, 2013.
(37) Rotondi, A.; Magli, M.; Morrone, L.; Alfei, B.; Pannelli, G. *Italian National Database of Monovarietal Extra Virgin Olive Oils*; IntechOpen, 2013. https://doi.org/10.5772/51772.
(38) Jukić Špika, M.; Žanetić, M.; Kraljić, K.; Soldo, B.; Ljubenkov, I.; Politeo, O.; Škevin, D. Differentiation Between Unfiltered and Filtered Oblica and Leccino Cv. Virgin Olive Oils. *Journal of Food Science* **2019**, *84* (4), 877–885. https://doi.org/10.1111/1750-3841.14494.
(39) Lozano-Sánchez, J.; Cerretani, L.; Bendini, A.; Segura-Carretero, A.; Fernández-Gutiérrez, A. Filtration Process of Extra Virgin Olive Oil: Effect on Minor Components, Oxidative Stability and Sensorial and Physicochemical Characteristics. *Trends in Food Science & Technology* **2010**, *21* (4), 201–211. https://doi.org/10.1016/j.tifs.2009.12.004.
(40) Zullo, B. A.; Venditti, G.; Ciafardini, G. Effects of the Filtration on the Biotic Fraction of Extra Virgin Olive Oil. *Foods* **2021**, *10* (8), 1677. https://doi.org/10.3390/foods10081677.

Tables

Parameters	unit	mean±sd (analytic.)	mean±sd (pred.)	R^2	*P*-value	RMSE	Error %
Free acidity	%oleic acid	0.331±0.169	0.345±0.144	0.98	<0.001	0.04	12.08
peroxide index	meq/kg	11.539±13.709	13.301±10.942	0.91	<0.001	4.787	41.49
uv							
K 232	nm	2.108±0.765	2.205±0.699	0.89	<0.001	0.184	8.73
K 268	nm	0.162±0.132	0.174±0.128	0.89	<0.001	0.024	14.81
Methyl esters of fatty acids	**%m/m**						
Palmitic acid (C16:0)		13.855±1.276	13.865±1.127	0.99	<0.001	0.232	1.67
Palmitoleic acid (C16:1)		1.031±0.22	1.038±0.176	0.88	<0.001	0.087	8.44
Margaric acid (C17:0)		0.041±0.012	0.042±0.008	0.92	<0.001	0.005	12.20
Margaroleic acid (C17:1)		0.082±0.021	0.082±0.015	0.96	<0.001	0.008	9.76
Stearic acid (C18:0)		2.167±0.26	2.155±0.218	0.99	<0.001	0.059	2.72
Oleic acid (C18:1)		73.034±1.756	73.036±1.614	0.99	<0.001	0.278	0.38
Linoleic acid (C18:2)		8.181±1.042	8.186±0.982	0.99	<0.001	0.149	1.82
Arachic acid (C20:0)		0.355±0.028	0.355±0.022	0.99	<0.001	0.008	2.25
Linolenic acid (C18:3)		0.813±0.143	0.822±0.121	0.98	<0.001	0.029	3.57
Eicosanoic acid (C20:1)		0.289±0.03	0.289±0.026	0.98	<0.001	0.007	2.42
Behenic acid (C22:0)		0.102±0.016	0.102±0.013	0.97	<0.001	0.005	4.90
Lignoceric acid (C24:0)		0.048±0.023	0.054±0.017	0.98	<0.001	0.009	18.75
Profile of biophenols	**mg/kg**						
Hydroxytyrosol		9.652±15.00	9.561±12.583	**0.99**	<0.001	2.985	30.93
Tyrosol		10.147±13.958	10.989±11.712	**0.99**	<0.001	2.867	28.25
Vanillic acid+ Caffeic acid		1.462±0.94	1.48±0.657	0.97	<0.001	0.338	23.12
Vanilline		2.654±1.166	2.626±0.847	0.97	<0.001	0.397	14.96
p-Coumaric acid		1.691±0.968	1.809±0.741	0.98	<0.001	0.311	18.39
Hydroxytyrosol acetate		1.781±1.289	12.07±0.799	0.95	<0.001	0.654	36.72
Ferulic acid		2.271±1.363	2.416±0.934	0.98	<0.001	0.509	22.41
o-Coumaric acdid		1.869±1.219	1.737±0.899	0.95	<0.001	0.488	26.11
Oxidized Dialdehyde form of Decarboxymethyloleuropein Aglycon		5.899±6.691	5.864±5.521	0.97	<0.001	1.678	28.45

Dialdehyde Decarboxymethyloleuropein Aglycon		65.208±62.009	64.646±60.034	**0.99**	<0.001	5.709	11.82
Oleuropein		9.091±10.519	8.817±9.342	0.98	<0.001	1.901	20.91
Dialdehyde Oleuropein Aglycon		9.214±11.35	9.115±10.334	0.98	<0.001	1.684	18.28
Oxidized Dialdehyde form of Decarboxymethyl ligstroside Aglycon		10.539±9.458	10.908±7.671	0.91	<0.001	2.854	27.08
Dialdehyde Decarboxymethyl ligstroside Aglycon		49.665±23.265	49.428±18.764	0.98	<0.001	6.059	12.20
Pinoresinol, 1-acetoxypinoresinol		23.148±17.976	23.078±15.665	0.98	<0.001	3.569	15.42
Cinnamic acid		7.146±8.465	7.249±7.416	0.96	<0.001	1.797	25.15
Dialdehyde Aglicone Ligstroside		9.143±13.165	9.723±10.992	0.92	<0.001	3.276	35.83
Oxidized Aldehydic form of Hydroxylic Oleuropein Aglycon		16.265±7.524	16.347±6.186	0.98	<0.001	1.926	11.84
Luteolin		6.545±5.547	6.835±4.433	0.97	<0.001	1.559	23.82
Aldehydic Hydroxylic Oleuropein Aglycon		34.316±40.796	32.829±32.018	0.94	<0.001	11.672	34.01
Oxidized Aldehydic form of Hydroxylic Ligstroside Aglycon		48.186±48.014	48.076±40.185	0.92	<0.001	10.812	22.44
Apigenin		4.977±2.937	4.957±1.977	0.97	<0.001	1.064	21.38
Methyl luteolin		8.543±5.317	8.348±3.904	0.95	<0.001	1.731	20.26
Aldehydic Hydroxylic Ligstroside Aglycon		13.535±14.571	13.578±12.603	0.93	<0.001	2.954	21.82
Total biophenols	**mg/kg**	355.069±163.529	353.699±156.928	**0.99**	<0.001	15.568	4.38
Total Tocopherols	**mg/kg**	307.049±120.319	301.417±111.924	0.99	<0.001	10.083	3.28

Table 1: RF regression models performed on analytical chemical compounds. For each analytical parameter the mean of spectrum-specific real and correctly average predicted values ± the standard deviation (sd), the Pearson's R^2, the *P*-values, the Root-mean-square deviation (RMSE), and the %Error ($RMSE/mean_{analtycal\ parameter}*100$) are reported.

Parameters	**mean±sd (analytic.)**	**mean±sd (pred.)**	**R^2**	***P*-value**	**RMSE**	**Error %**
Bitter	3.629±1.707	3.653±1.465	0.98	<0.001	0.393	10.83
Sweet	1.501±0.896	1.509±0.67	0.98	<0.001	0.274	18.25
Spicy	3.446±1.488	3.429±1.121	0.98	<0.001	0.459	13.32
Gustatory intense	3.963±0.917	3.951±0.697	0.97	<0.001	0.283	7.14
Olfactory Fruity	3.119±1.433	3.098±1.019	0.97	<0.001	0.486	15.58
Olfactory intense	4.272±0.973	4.248±0.73	0.97	<0.001	0.316	7.40

Overall pleasantness	4.226±1.791	4.355±1.299	0.97	<0.001	0.609	14.41

Table 2: RF regression models performed on sensorial profiles. For each sensorial characteristic the mean of spectrum-specific real and correctly average predicted values ± the standard deviation (sd), the Pearson's R^2, the *P*-values, the Root-mean-square deviation (RMSE) and the %Error (RMSE/$mean_{analtycal\ parameter}$*100) are reported.

	Moraiolo	**Leccino**
Moraiolo	77.6	22.4
Leccino	25.8	74.2
Overall accuracy: 75.9%		

Table 3: RF classification confusion matrix performed on bucketed spectra to correctly classify the two main olive oil cultivars (Moraiolo vs Leccino). The overall accuracy of the RF model was also reported.

	Moraiolo	**Leccino**
Moraiolo	72.3	27.7
Leccino	11.5	88.5
Overall accuracy: 80.4%		

Table 4: RF classification confusion matrix performed on indirectly and directly quantified molecular feature, and sensory profiles to correctly classify the two main olive oil cultivars (Moraiolo *vs* Leccino). The overall accuracy of the RF model was also reported.

	Filtered	**Non-filtered**
Filtered	83.3	16.7
Non-filtered	33.5	66.7
Overall accuracy: 75.8%		

Table 5: RF classification confusion matrix performed on bucketed spectra to correctly classify the two main post-harvest olive oil procedures (filtration *vs* non-filtration). The overall accuracy of the RF model was also reported

	Filtered	**Non-filtered**
Filtered	100	0
Non-filtered	33.3	66.7
Overall accuracy: 84.8%		

Table 6: RF classification confusion matrix performed on indirectly and directly quantified molecular feature, and sensory profiles to correctly classify the two main post-harvest olive oil procedures (filtration *vs* non-filtration). The overall accuracy of the RF model was also reported

Figures

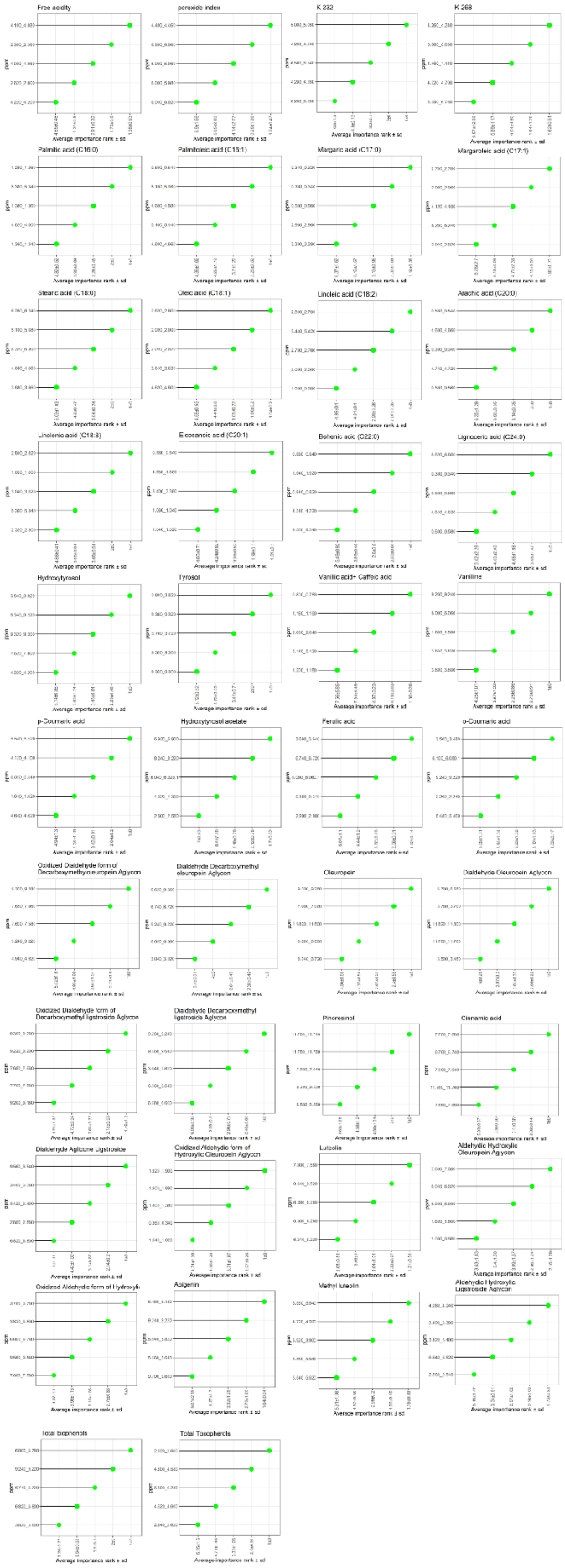

Free acidity
peroxide index
K 232
K 268
Palmitic acid (C16:0)
Palmitoleic acid (C16:1)
Margaric acid (C17:0)
Margaroleic acid (C17:1)
Stearic acid (C18:0)
Oleic acid (C18:1)
Linoleic acid (C18:2)
Arachic acid (C20:0)
Linolenic acid (C18:3)
Eicosenoic acid (C20:1)
Behenic acid (C22:0)
Lignoceric acid (C24:0)
Hydroxytyrosol
Tyrosol
Vanillic acid+ Caffeic acid
Vanillina
p-Coumaric acid
Hydroxytyrosol acetate
Ferulic acid
o-Coumaric acid
Oxidized Dialdehyde form of Decarboxymethyloleuropein Aglycon
Dialdehyde Decarboxymethyl oleuropein Aglycon
Oleuropein
Dialdehyde Oleuropein Aglycon
Oxidized Dialdehyde form of Decarboxymethyl ligstroside Aglycon
Dialdehyde Decarboxymethyl ligstroside Aglycon
Pinoresinol
Cinnamic acid
Dialdehyde Aglicone Ligstroside
Oxidized Aldehydic form of Hydroxylic Oleuropein Aglycon
Luteolin
Aldehydic Hydroxylic Oleuropein Aglycon
Oxidized Aldehydic form of Hydroxylic
Apigenin
Methyl luteolin
Aldehydic Hydroxylic Ligstroside Aglycon
Total biophenols
Total Tocopherols
ppm
Average importance rank ± sd

Figure 1: Spectra regions (ppm) RF importance plot. The top 5 average importances ranks for each analytical parameter are reported

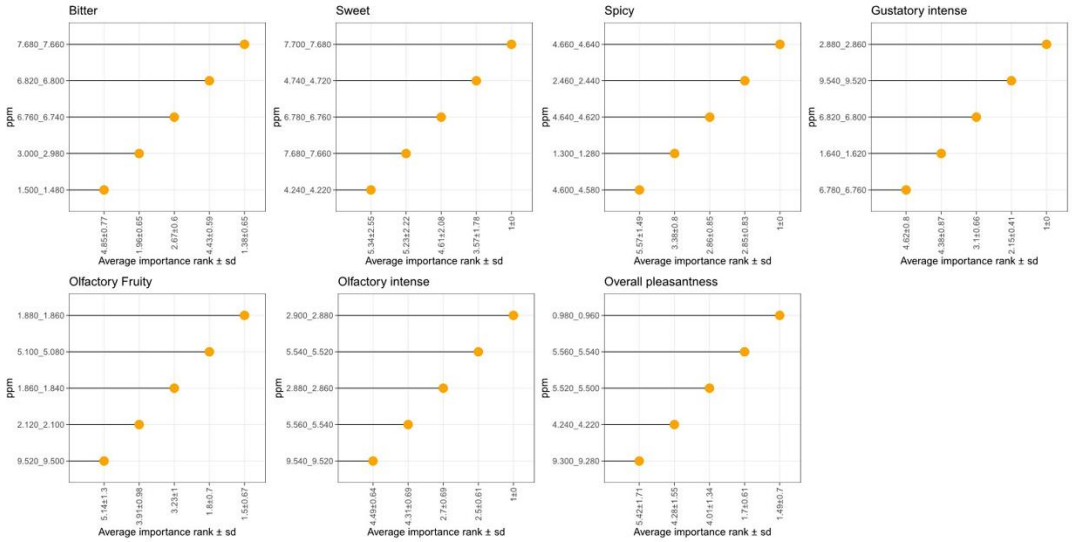


Figure 2: Spectra regions (ppm) RF importance plot. The top 5 average importances ranks for each sensorial parameter are reported

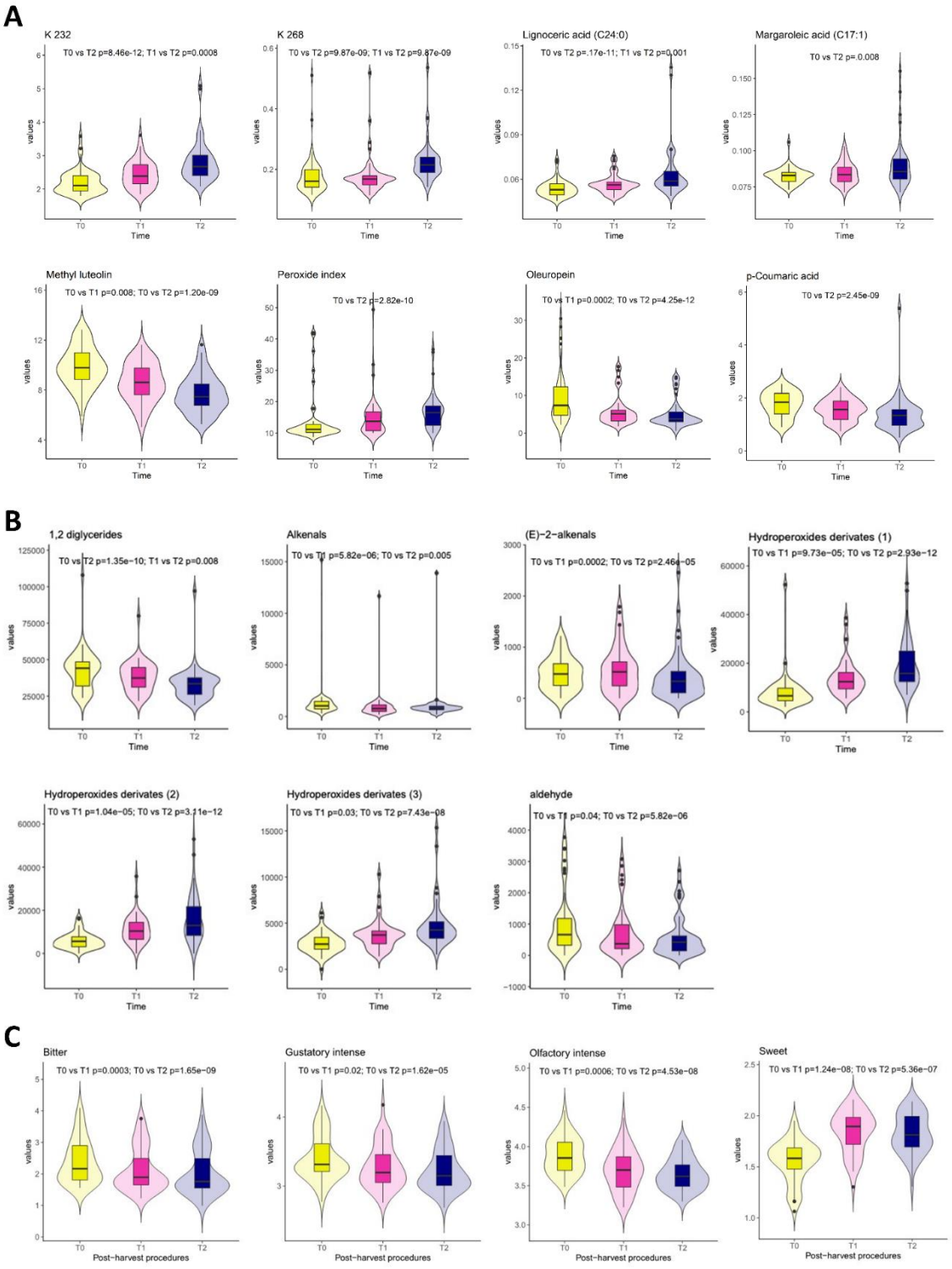


Figure 3: Univariate analysis performed on the three timepoints sampling of olive oils, t0, t1, and t2, respectively. The violin plots are related to A) indirectly quantified molecular features, B) directly quantified molecular features, and C) the sensorial profile. The yellow violins correspond to t0 olive oil samples, the magenta violins to t1 olive oil samples, and the dark blue violin to the t2 olive oil samples.

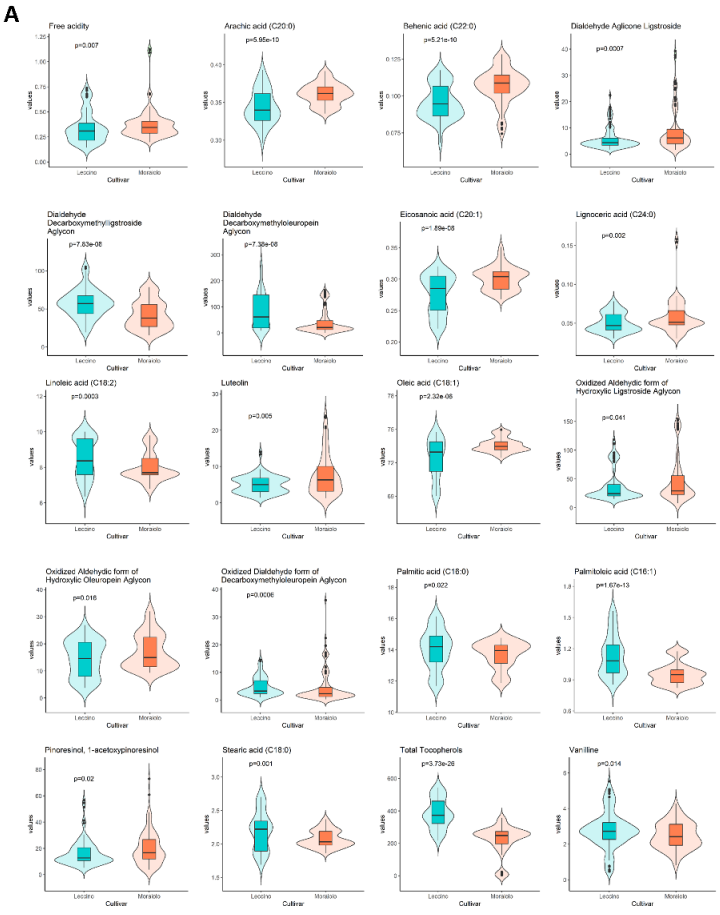

A
Free acidity
p=0.007
Arachic acid (C20:0)
p=5.95e-10
Behenic acid (C22:0)
p=5.21e-10
Dialdehyde Aglicone Ligstroside
p=0.0007
Dialdehyde Decarboxymethyl ligstroside Aglycon
p=7.63e-06
Dialdehyde Decarboxymethyloleuropein Aglycon
p=7.38e-06
Eicosanoic acid (C20:1)
p=1.89e-08
Lignoceric acid (C24:0)
p=0.002
Linoleic acid (C18:2)
p=0.0003
Luteolin
p=0.005
Oleic acid (C18:1)
p=2.32e-06
Oxidized Aldehydic form of Hycroxylic Ligstroside Aglycon
p=0.041
Oxidized Aldehydic form of Hydroxylic Oleuropein Aglycon
p=0.016
Oxidized Dialdehydic form of Decarboxymethyloleuropein Aglycon
p=0.0006
Palmitic acid (C16:0)
p=0.022
Palmitoleic acid (C16:1)
p=1.67e-13
Pinoresinol, 1-acetoxypinoresinol
p=0.02
Stearic acid (C18:0)
p=0.001
Total Tocopherols
p=3.73e-26
Vanilline
p=0.014
B
1,2 diglycerides/triolein
p=8.17e-12
4-hydroxy-trans-alkenals
p=2.67e-06
Aldehydic forms of oleuropein and ligstroside
p=0.003
Beta-sitosterol
p=7.89e-09
Cycloartenol
p=0.001
Cyclopropene
p=4.45e-32
Hydroperoxides derivates (1)
p=0.006
Linoleic acid
p=0.002
Linoleyl/Linolenyl
p=0.002
Squalene
p=9.35e-10
Terpene (1)
p=1.80e-23
Terpene (2)
p=4.69e-16
Terpene (3)
p=4.37e-09
C
Bitter
p=0.003
Olfactory Fruity
p=0.002
Spicy
p=0.303

Figure 4: Univariate analysis performed on the two olive oil cultivars: Leccino and Morellino. The violin plots are related to A) indirectly quantified molecular features, B) directly quantified molecular features, and C) the sensorial profile. The light blue violins correspond to the Leccino olive oil samples, and the light red violins to Morellino olive oil samples.

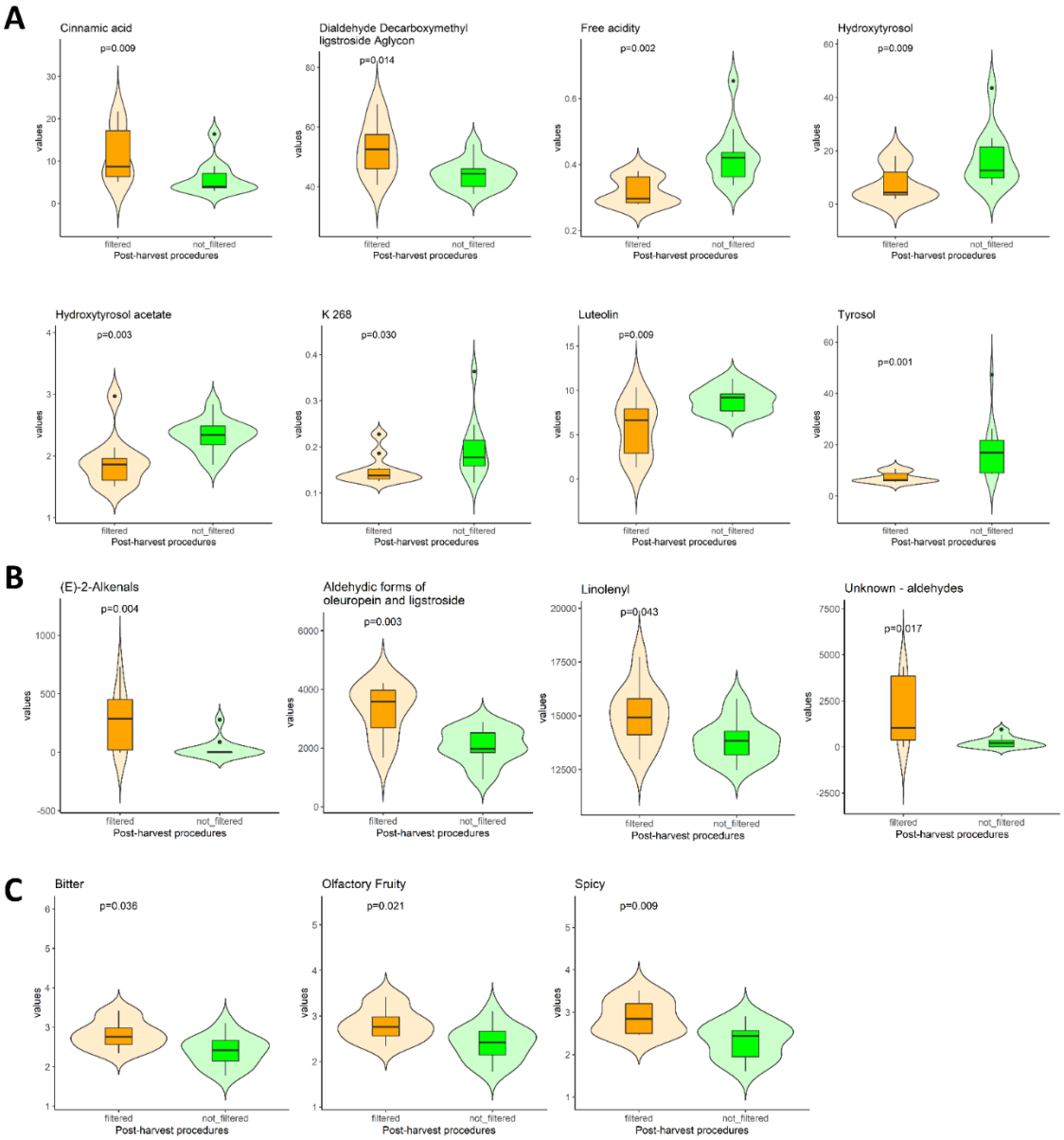


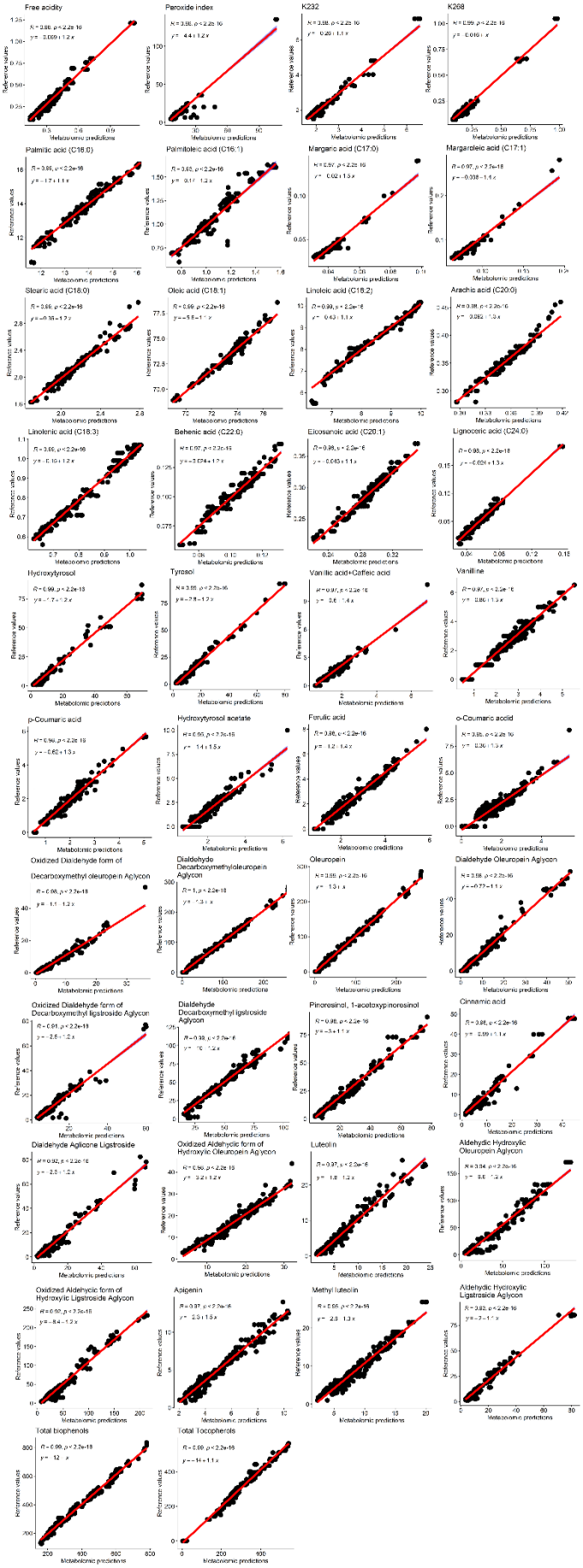
Figure 5: Univariate analysis performed on the two post-harvest procedure: filtration and non-filtration, respectively. The violin plots are related to A) indirectly quantified molecular features, B) directly quantified molecular features, and C) the sensorial profile. The orange violins correspond to the filtered olive oil samples, and the green violins to the non-filtered olive oil samples.

Supplementary Materials

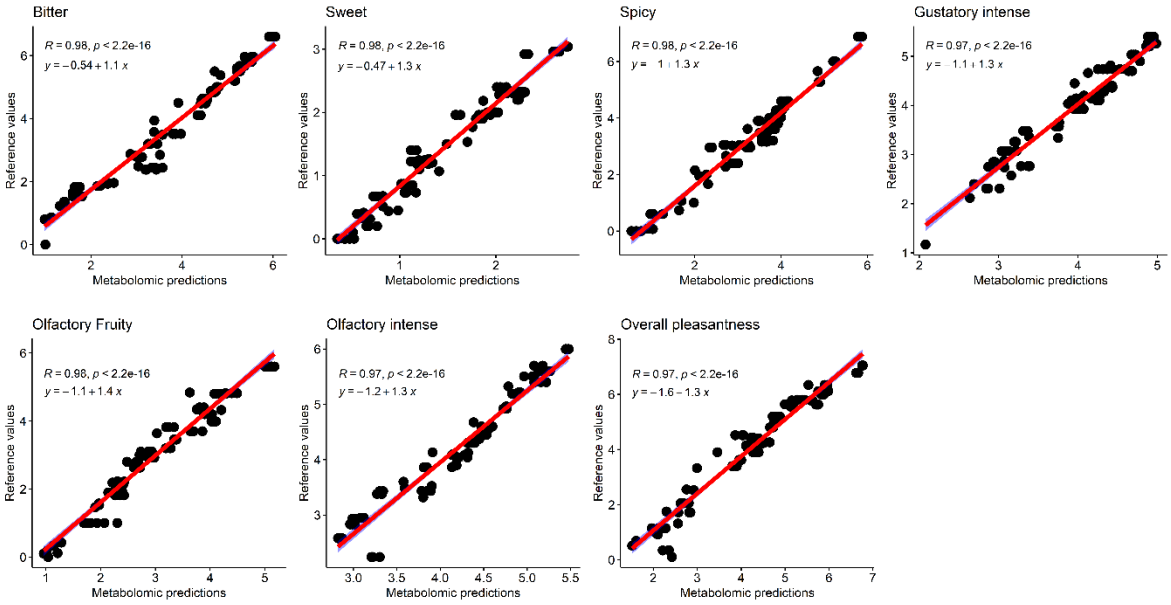
Metabolite names	range ppm	experiment
cycloartenol (1)	0.365-0.3181	zg
cycloartenol (2)	0.5805-0.5575	zg
cyclopropene	0.5557-0.5435	zg
unknown 1	0.6326-0.6107	zg
beta-sitosterol	0.7033-0.684	zg
all acids	0.9279-0.8445	zg
linolenyl/omega3	1.0119-1.0007	zg
all acyl chains (1)	1.4248-1.188	zg
all acyl chains (1)	1.6722-1.5381	zg
squalene	1.7064-1.68	zg
oleic+linoleic	2.0961-1.9435	zg
unkown 2	2.2135-2.1994	zg
all acyl chains (3)	2.3608-2.2512	zg
hydrogenation saturation	2.3857-2.3635	zg
linoleic acid (1)	2.6456-2.6005	zg
linoleyl + linolenyl	2.869-2.7145	zg
linoleic acid (2)	2.958-2.894	zg
unkown 3	3.3204-3.2616	zg
beta-sitosterol + stigmasterol	3.5743-3.4956	zg
1,2 diglycerides	3.7693-3.7171	zg
Acetate	4.004-3.997	zg
Glycerol triglycerides + 1,3 diglycerides	4.2112-4.0678	zg
Glycerol triglycerides + 1,2 diglycerides + linoleic	4.4186-4.2292	zg
Glycerol-13C3	4.5211-4.4584	zg
unkown 4	4.7665-4.7486	zg
1,2 diglycerides/triolein	5.2064-5.0544	zg
Terpene (1)	4.618-4.577	NOESYGPPS
Terpene (2)	4.6873-4.6608	NOESYGPPS
Terpene (3)	4.7386-4.7146	NOESYGPPS
Hydroperoxides derivates (1)	6.0567-5.8958	NOESYGPPS
Hydroperoxides derivates (2)	6.6213-6.5358	NOESYGPPS
Hydroperoxides derivates (3)	6.313-6.2201	NOESYGPPS
aldehyde (1)	7.6692-7.6536	NOESYGPPS
Formaldehyde	8.0268-8.0032	NOESYGPPS
aldehydic forms of oleuropein and ligstroside	9.2601-9.2141	NOESYGPPS
aldehyde (2)	9.296-9.2792	NOESYGPPS
n_alchenale	9.7846-9.7592	NOESYGPPS
4-hydroxy-trans-alkenals	9.6712-9.6348	NOESYGPPS

E,E-2,4-alkadienals	9.5609-9.541	NOESYGPPS
(E)-2-alkenals	9.5165-9.506	NOESYGPPS
oxidized dialdehyde form of decarboxymethyloleuropein aglycon	11.82-11.8	NOESYGPPS

Supplementary Table 1: List of compounds directly assign and quantified on olive oil spectrum; the range of ppm and the ^{1}H NMR experiment in which the assignment was performed are also reported


Free acidity
Peroxide index
K232
K268
Palmitic acid (C16:0)
Palmitoleic acid (C16:1)
Margaric acid (C17:0)
Margaroleic acid (C17:1)
Stearic acid (C18:0)
Oleic acid (C18:1)
Linoleic acid (C18:2)
Arachic acid (C20:0)
Linolenic acid (C18:3)
Behenic acid (C22:0)
Eicosanoic acid (C20:1)
Lignoceric acid (C24:0)
Hydroxytyrosol
Tyrosol
Vanillic acid+Caffeic acid
Vanillina
p-Coumaric acid
Hydroxytyrosol acetate
Ferulic acid
o-Coumaric acid
Oxidized Dialdehyde form of Decarboxymethyl oleuropein Aglycon
Dialdehyde Decarboxymethyloleuropein Aglycon
Oleuropein
Dialdehyde Oleuropein Aglycon
Oxidized Dialdehyde form of Decarboxymethyl ligstroside Aglycon
Dialdehyde Decarboxymethyl ligstroside Aglycon
Pinoresinol, 1-acetoxypinoresinol
Cinnamic acid
Dialdehyde Aglicone Ligstroside
Oxidized Aldehydic form of Hydroxylic Oleuropein Aglycon
Luteolin
Aldehydic Hydroxylic Oleuropein Aglycon
Oxidized Aldehydic form of Hydroxylic Ligstroside Aglycon
Apigenin
Methyl luteolin
Aldehydic Hydroxylic Ligstroside Aglycon
Total biophenols
Total Tocopherols

Supplementary Figure S1: Linear regression RF models performed on chemical olive oil parameters. In the multipanel here presented the R2 and the P-value are also reported.



Supplementary Figure S2: Linear regression RF models performed on sensory olive oil parameters. In the multipanel here presented the R^2, the *P*-value, and the regression equation are also reported.

Conclusions

The results presented in this three-year methodological PhD thesis primarily extolled the potential of untargeted metabolomics for various applications in the biomedical field, focusing on the investigation of both human physiological and pathophysiological processes, while also emphasizing the fundamental synergy between chemistry, biochemistry, and different statistical approaches for a better understanding the dynamic, comprehensive, and accurate image of human phenotype, thus increasing knowledge on biological mechanisms of human metabolism in healthy and diseased status.

In this methodological thesis, firstly, untargeted metabolomics has been applied to obtain new insights into different biological and physiological conditions. To investigate the complex physiological aging process and to shed light on the sexual dimorphic mechanisms of aging, in the first studies here proposed **(§4.1.1**, **§4.1.2**, and **§4.1.3**), untargeted metabolomics was applied, integrating both standard univariate and multivariate analyses with the molecular features-association networks. In more detail, applying a holistic metabolomic approach, in the first study presented in this thesis, the untargeted approach proved to be decisive in investigating the difference in terms of architecture and connectivity of sex-specific and clinical and biochemical parameters-specific serum metabolic association networks build considering a cohort of nonagenarians. Secondly, using the same cohort described above, the NMR untargeted metabolomics, combined with classical statistical multivariate and univariate analyses, highlighted the sexual dimorphic metabolic differences in nonagenarians and it was, also, relevant for determining potential biomarkers for predicting the risk of developing geriatric diseases, particularly decreased cognitive function and geriatric depression. Thirdly, the untargeted metabolomic approach reveals to be important to evaluate and identify the potential correlation between different age-increasing groups and: i) the molecular features concentrations; ii) the pairwise molecular features correlations; ii) the pairwise molecular features ratios. This innovative approach in metabolomics was proposed to shed light on the dynamic of aging molecular mechanisms.

Moreover, to better understand the mechanisms underlying human physiology, considering that in the human body several microorganisms exist and coexist, an NMR untargeted metabolomics approach was applied in **§4.1.4** to characterize and understand potential metabolic changes that could determine and affect the human phenotype after assumption of a non-invasive exogenous treatment – probiotics.

As known, the ABO and Rh systems play a fundamental role in transfusion medicine and hematopoietic transplantation and are also related to the pathogenesis and

pathophysiology of various human diseases, such as cardiovascular and oncological diseases. In this scenario, the work **§4.1.5** here presented is the first NMR-based metabolomics work in which the specific associations between circulating levels of plasma metabolites and lipoproteins and the ABO/Rh blood group system were analyzed in a cohort of Italian healthy blood donors.

Metabolomics is also an excellent tool for characterizing and preventing a pathophysiological status. In particular, metabolic perturbations are fundamental events that contribute to ischemic stroke, its progression, and the development of unfavorable outcomes. Using retrospective data from the Italian multicentric observational study, in **§4.1.6** an NMR-based metabolomics approach was applied to identify serum biomarkers of three-month poor outcomes (*i.e.* mortality, development of neurological impairments, hemorrhagic transformation of the cerebral lesion, and non-response to intravenous thrombolysis) in AIS patients, treated with intravenous recombinant tissue plasminogen activator.

As for neurological disorders, cancer research is one of the most important fields investigated by metabolomics, with the aim to discover new biomarkers and refine diagnostic tests and therapies. With these ideas in mind, in **§4.1.7**, an untargeted metabolomics approach was used to investigate the blood metabolite profiles in patients with CRC, polyposis, and healthy controls, using a metabolite–metabolite association networks analysis to investigate and explore the existence of molecular mechanisms underlying these different clinical profiles and with the aim to highlight potential biomarkers of the disease.

To conclude, in **§4.1.8**, is reported a contribution highlighting the various aspects of the NMR-based untargeted metabolomics, focusing on: i) the effects on the individual metabolomic fingerprint of non-invasive treatment; ii) the characterization of metabolic fingerprints of specific diseases; ii) the effects of drugs on the disease fingerprint and on its reversal to a healthy metabolomic status.

In this methodological thesis, **§4.2.1** is dedicated to the development of a robust statistical approach, based on the construction of linear regression Random Forest models, to relate the ^{1}H-NMR bucketed olive oil spectra with the analytical results of several important chemical parameters and sensory profile of olive oil, thus determining an indirect quantification of both chemical and organoleptic parameters.

In conclusion, even if a significant fraction of the presented material is still in preparation and touches on different research areas, this methodological thesis may contribute to the demonstration that untargeted metabolomics, combined with bioinformatic tools and

robust statistical analyses, might be regarded as a complete and powerful analytical technique with fair and realistic possibilities to be structurally implemented in the heterogeneous but integrated field, ranging, in particular, from biomedical to food research.

References

(1) Tyers, M.; Mann, M. From Genomics to Proteomics. *Nature* **2003**, *422* (6928), 193–197. https://doi.org/10.1038/nature01510.

(2) Lindon, J. C.; Nicholson, J. K. Systems Biology Metabonomics. *Nature (London)* **2008**, *455* (7216), 1054–1056. https://doi.org/10.1038/4551054a.

(3) Vignoli, A.; Ghini, V.; Meoni, G.; Licari, C.; Takis, P. G.; Tenori, L.; Turano, P.; Luchinat, C. High-Throughput Metabolomics by 1D NMR. *Angewandte Chemie International Edition* **2019**, *58* (4), 968–994. https://doi.org/10.1002/anie.201804736.

(4) Roberts, L. D.; Souza, A. L.; Gerszten, R. E.; Clish, C. B. Targeted Metabolomics. *Current Protocols in Molecular Biology* **2012**, *98* (1), 30.2.1-30.2.24. https://doi.org/10.1002/0471142727.mb3002s98.

(5) Schrimpe-Rutledge, A. C.; Codreanu, S. G.; Sherrod, S. D.; McLean, J. A. Untargeted Metabolomics Strategies – Challenges and Emerging Directions. *J Am Soc Mass Spectrom* **2016**, *27* (12), 1897–1905. https://doi.org/10.1007/s13361-016-1469-y.

(6) Bingol, K. Recent Advances in Targeted and Untargeted Metabolomics by NMR and MS/NMR Methods. *High Throughput* **2018**, *7* (2), 9. https://doi.org/10.3390/ht7020009.

(7) Kosmides, A. K.; Kamisoglu, K.; Calvano, S. E.; Corbett, S. A.; Androulakis, I. P. Metabolomic Fingerprinting: Challenges and Opportunities. *Crit Rev Biomed Eng* **2013**, *41* (3), 205–221.

(8) Zacharias, H. U.; Altenbuchinger, M.; Gronwald, W. Statistical Analysis of NMR Metabolic Fingerprints: Established Methods and Recent Advances. *Metabolites* **2018**, *8* (3), 47. https://doi.org/10.3390/metabo8030047.

(9) Larive, C. K.; Barding, G. A.; Dinges, M. M. NMR Spectroscopy for Metabolomics and Metabolic Profiling. *Anal. Chem.* **2015**, *87* (1), 133–146. https://doi.org/10.1021/ac504075g.

(10) Sengupta, A.; Weljie, A. M. NMR Spectroscopy Based Metabolic Profiling of Biospecimens. *Curr Protoc Protein Sci* **2019**, *98* (1), e98. https://doi.org/10.1002/cpps.98.

(11) Wishart, D. S. Metabolomics for Investigating Physiological and Pathophysiological Processes. *Physiological Reviews* **2019**, *99* (4), 1819–1875. https://doi.org/10.1152/physrev.00035.2018.

(12) Zhang, X.; Li, Q.; Xu, Z.; Dou, J. Mass Spectrometry-Based Metabolomics in Health and Medical Science: A Systematic Review. *RSC Adv.* **2020**, *10* (6), 3092–3104. https://doi.org/10.1039/C9RA08985C.

(13) Rampler, E.; Abiead, Y. E.; Schoeny, H.; Rusz, M.; Hildebrand, F.; Fitz, V.; Koellensperger, G. Recurrent Topics in Mass Spectrometry-Based Metabolomics and Lipidomics—Standardization, Coverage, and Throughput. *Anal. Chem.* **2021**, *93* (1), 519–545. https://doi.org/10.1021/acs.analchem.0c04698.

(14) Li, J.; Vosegaard, T.; Guo, Z. Applications of Nuclear Magnetic Resonance in Lipid Analyses: An Emerging Powerful Tool for Lipidomics Studies. *Progress in Lipid Research* **2017**, *68*, 37–56. https://doi.org/10.1016/j.plipres.2017.09.003.

(15) Johnson, C. H.; Ivanisevic, J.; Siuzdak, G. Metabolomics: Beyond Biomarkers and towards Mechanisms. *Nat Rev Mol Cell Biol* **2016**, *17* (7), 451–459. https://doi.org/10.1038/nrm.2016.25.

(16) 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. *Analyst* **2016**, *141* (15), 4558–4561. https://doi.org/10.1039/C6AN00393A.

(17) Stringer, K. A.; Younger, J. G.; McHugh, C.; Yeomans, L.; Finkel, M. A.; Puskarich, M. A.; Jones, A. E.; Trexel, J.; Karnovsky, A. Whole Blood Reveals More Metabolic Detail of the Human Metabolome than Serum as Measured by 1H-NMR Spectroscopy: Implications for Sepsis Metabolomics. *Shock* **2015**, *44* (3), 200–208. https://doi.org/10.1097/SHK.0000000000000406.

(18) Li, S.; Todor, A.; Luo, R. Blood Transcriptomics and Metabolomics for Personalized Medicine. *Computational and Structural Biotechnology Journal* **2016**, *14*, 1–7. https://doi.org/10.1016/j.csbj.2015.10.005.

(19) *Food Analysis Laboratory Manual | SpringerLink*. https://link.springer.com/book/10.1007/978-1-4419-1463-7 (accessed 2022-08-27).

(20) 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.

(21) Ghini, V.; Quaglio, D.; Luchinat, C.; Turano, P. NMR for Sample Quality Assessment in Metabolomics. *N Biotechnol* **2019**, *52*, 25–34. https://doi.org/10.1016/j.nbt.2019.04.004.

(22) Ghini, V.; Abuja, P. M.; Polasek, O.; Kozera, L.; Laiho, P.; Anton, G.; Zins, M.; Klovins, J.; Metspalu, A.; Wichmann, H. E.; Gieger, C.; Luchinat, C.; Zatloukal, K.; Turano, P. Impact of the Pre-Examination Phase on Multicenter Metabolomic Studies. *New biotechnology* **2022**, *68*, 37–47. https://doi.org/10.1016/J.NBT.2022.01.006.

(23) Hotelling, H. Analysis of a Complex of Statistical Variables into Principal Components. *Journal of Educational Psychology* **1933**, *24* (6), 417–441. https://doi.org/10.1037/h0071325.

(24) Trygg, J.; Wold, S. Orthogonal projections to latent structures (O-PLS). *Journal of Chemometrics* **2002**, *16* (3), 119–128. https://doi.org/10.1002/cem.695.

(25) Westerhuis, J. A.; van Velzen, E. J. J.; Hoefsloot, H. C. J.; Smilde, A. K. Multivariate Paired Data Analysis: Multilevel PLSDA versus OPLSDA. *Metabolomics* **2010**, *6* (1), 119–128. https://doi.org/10.1007/s11306-009-0185-z.

(26) Evgeniou, T.; Pontil, M. Support Vector Machines: Theory and Applications. In *Machine Learning and Its Applications: Advanced Lectures*; Paliouras, G., Karkaletsis, V., Spyropoulos, C. D., Eds.; Lecture Notes in Computer Science; Springer: Berlin, Heidelberg, 2001; pp 249–257. https://doi.org/10.1007/3-540-44673-7_12.

(27) Breiman, L. Random Forests. *Machine Learning 2001 45:1* **2001**, *45* (1), 5–32. https://doi.org/10.1023/A:1010933404324.

(28) Ali, J.; Khan, R.; Ahmad, N.; Maqsood, I. Random Forests and Decision Trees. *International Journal of Computer Science Issues(IJCSI)* **2012**, *9*.

(29) 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. *Evidence-Based Complementary and Alternative Medicine* **2013**, *2013*, e298183. https://doi.org/10.1155/2013/298183.

(30) Kruskal, W. H.; Wallis, W. A. Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association* **1952**, *47* (260), 583–621. https://doi.org/10.1080/01621459.1952.10483441.

(31) Wilcoxon, F. Individual Comparisons by Ranking Methods. *Biometrics Bulletin* **1945**, *1* (6), 80–83. https://doi.org/10.2307/3001968.

(32) Neuhäuser, M. Wilcoxon–Mann–Whitney Test; 2011; pp 1656–1658. https://doi.org/10.1007/978-3-642-04898-2_615.

(33) Oberg, A. L.; Mahoney, D. W. Linear Mixed Effects Models. In *Topics in Biostatistics*; Ambrosius, W. T., Ed.; Methods in Molecular Biology™; Humana Press: Totowa, NJ, 2007; pp 213–234. https://doi.org/10.1007/978-1-59745-530-5_11.

(34) Wanichthanarak, K.; Jeamsripong, S.; Pornputtapong, N.; Khoomrung, S. Accounting for Biological Variation with Linear Mixed-Effects Modelling Improves the Quality of Clinical Metabolomics Data. *Comput Struct Biotechnol J* **2019**, *17*, 611–618. https://doi.org/10.1016/j.csbj.2019.04.009.

(35) 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 Bioinformatics* **2016**, *53*, 14.13.1-14.13.18. https://doi.org/10.1002/0471250953.bi1413s53.

(36) Pang, Z.; Chong, J.; Zhou, G.; de Lima Morais, D. A.; Chang, L.; Barrette, M.; Gauthier, C.; Jacques, P.-É.; Li, S.; Xia, J. MetaboAnalyst 5.0: Narrowing the Gap between Raw Spectra and Functional Insights. *Nucleic Acids Research* **2021**, *49* (W1), W388–W396. https://doi.org/10.1093/nar/gkab382.

(37) Cuperlovic-Culf, M. *NMR Metabolomics in Cancer Reasearch*; 2012.

(38) Oakman, C.; Tenori, L.; Claudino, W. M.; Cappadona, S.; Nepi, S.; Battaglia, A.; Bernini, P.; Zafarana, E.; Saccenti, E.; Fornier, M.; Morris, P. G.; Biganzoli, L.; Luchinat, C.; Bertini, I.; Leo, A. D. 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. *Annals of Oncology* **2011**, *22* (6), 1295–1301. https://doi.org/10.1093/annonc/mdq606.

(39) Larkin, J. R.; Anthony, S.; Johanssen, V. A.; Yeo, T.; Sealey, M.; Yates, A. G.; Smith, C. F.; Claridge, T. D. W.; Nicholson, B. D.; Moreland, J.-A.; Gleeson, F.; Sibson, N. R.; Anthony, D. C.; Probert, F. Metabolomic Biomarkers in Blood Samples Identify Cancers in a Mixed Population of Patients with Nonspecific Symptoms. *Clinical Cancer Research* **2022**, *28* (8), 1651–1661. https://doi.org/10.1158/1078-0432.CCR-21-2855.

(40) Salmerón, A. M.; Tristán, A. I.; Abreu, A. C.; Fernández, I. Serum Colorectal Cancer Biomarkers Unraveled by NMR Metabolomics: Past, Present, and Future. *Anal. Chem.* **2022**, *94* (1), 417–430. https://doi.org/10.1021/acs.analchem.1c04360.

(41) Banoei, M. M.; Mahé, E.; Mansoor, A.; Stewart, D.; Winston, B. W.; Habibi, H. R.; Shabani-Rad, M.-T. NMR-Based Metabolomic Profiling Can Differentiate Follicular Lymphoma from Benign Lymph Node Tissues and May Be Predictive of Outcome. *Sci Rep* **2022**, *12* (1), 8294. https://doi.org/10.1038/s41598-022-12445-5.

(42) 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. *International Journal of Cardiology* **2013**, *168* (4), e113–e115. https://doi.org/10.1016/j.ijcard.2013.08.042.

(43) 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.

(44) Lema, C.; Andrés, M.; Aguadé-Bruix, S.; Consegal, M.; Rodriguez-Sinovas, A.; Benito, B.; Ferreira-Gonzalez, I.; Barba, I. 1H NMR Serum Metabolomic Profiling of Patients at Risk of Cardiovascular Diseases Performing Stress Test. *Sci Rep* **2020**, *10* (1), 17838. https://doi.org/10.1038/s41598-020-74880-6.

(45) 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.

(46) Maltesen, R. G.; Wimmer, R.; Rasmussen, B. S. A Longitudinal Serum NMR-Based Metabolomics Dataset of Ischemia-Reperfusion Injury in Adult Cardiac Surgery. *Sci Data* **2020**, *7*, 198. https://doi.org/10.1038/s41597-020-0545-0.

(47) 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.

(48) Yang, Z.; Wang, J.; Chen, J.; Luo, M.; Xie, Q.; Rong, Y.; Wu, Y.; Cao, Z.; Liu, Y. High-Resolution NMR Metabolomics of Patients with Subjective Cognitive Decline plus: Perturbations in the Metabolism of Glucose and Branched-Chain Amino Acids. *Neurobiology of Disease* **2022**, *171*, 105782. https://doi.org/10.1016/j.nbd.2022.105782.

(49) Abreu, A. C.; Navas, M. M.; Fernández, C. P.; Sánchez-Santed, F.; Fernández, I. NMR-Based Metabolomics Approach to Explore Brain Metabolic Changes Induced by Prenatal Exposure to Autism-Inducing Chemicals. *ACS Chem. Biol.* **2021**, *16* (4), 753–765. https://doi.org/10.1021/acschembio.1c00053.

(50) 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.

(51) Gómez-Cebrián, N.; Vázquez Ferreiro, P.; Carrera Hueso, F. J.; Poveda Andrés, J. L.; Puchades-Carrasco, L.; Pineda-Lucena, A. Pharmacometabolomics by NMR in Oncology: A Systematic Review. *Pharmaceuticals* **2021**, *14* (10), 1015. https://doi.org/10.3390/ph14101015.

(52) Alarcon-Barrera, J. C.; Kostidis, S.; Ondo-Mendez, A.; Giera, M. Recent Advances in Metabolomics Analysis for Early Drug Development. *Drug Discovery Today* **2022**, *27* (6), 1763–1773. https://doi.org/10.1016/j.drudis.2022.02.018.

(53) Casadei-Gardini, A.; Del Coco, L.; Marisi, G.; Conti, F.; Rovesti, G.; Ulivi, P.; Canale, M.; Frassineti, G. L.; Foschi, F. G.; Longo, S.; Fanizzi, F. P.; Giudetti, A. M. 1H-NMR Based Serum Metabolomics Highlights Different Specific Biomarkers between Early and Advanced Hepatocellular Carcinoma Stages. *Cancers (Basel)* **2020**, *12* (1), 241. https://doi.org/10.3390/cancers12010241.

(54) Song, Z.; Wang, H.; Yin, X.; Deng, P.; Jiang, W. Application of NMR Metabolomics to Search for Human Disease Biomarkers in Blood. *Clinical Chemistry and Laboratory Medicine (CCLM)* **2019**, *57* (4), 417–441. https://doi.org/10.1515/cclm-2018-0380.

(55) Zhang, J.; Du, Y.; Zhang, Y.; Xu, Y.; Fan, Y.; Li, Y. 1H-NMR Based Metabolomics Technology Identifies Potential Serum Biomarkers of Colorectal Cancer Lung Metastasis in a Mouse Model. *CMAR* **2022**, *14*, 1457–1469. https://doi.org/10.2147/CMAR.S348981.

(56) Darst, B. F.; Koscik, R. L.; Hogan, K. J.; Johnson, S. C.; Engelman, C. D. Longitudinal Plasma Metabolomics of Aging and Sex. *Aging (Albany NY)* **2019**, *11* (4), 1262–1282. https://doi.org/10.18632/aging.101837.

(57) Lin, H.-T.; Cheng, M.-L.; Lo, C.-J.; Hsu, W.-C.; Lin, G.; Liu, F.-C. 1H NMR Metabolomic Profiling of Human Cerebrospinal Fluid in Aging Process. *Am J Transl Res* **2021**, *13* (11), 12495–12508.

(58) Ghini, V.; Tenori, L.; Pane, M.; Amoruso, A.; Marroncini, G.; Squarzanti, D. F.; Azzimonti, B.; Rolla, R.; Savoia, P.; Tarocchi, M.; Galli, A.; Luchinat, C. Effects of Probiotics Administration on Human Metabolic Phenotype. *Metabolites* **2020**, *10* (10). https://doi.org/10.3390/metabo10100396.

(59) Rådjursöga, M.; Lindqvist, H. M.; Pedersen, A.; Karlsson, G. B.; Malmodin, D.; Brunius, C.; Ellegård, L.; Winkvist, A. The 1H NMR Serum Metabolomics Response to a Two Meal Challenge: A Cross-over Dietary Intervention Study in Healthy Human Volunteers. *Nutr J* **2019**, *18*, 25. https://doi.org/10.1186/s12937-019-0446-2.

(60) Tomassini, A.; Capuani, G.; Delfini, M.; Miccheli, A. Chapter 11 - NMR-Based Metabolomics in Food Quality Control. In *Data Handling in Science and Technology*; Marini, F., Ed.; Chemometrics in Food Chemistry; Elsevier, 2013; Vol. 28, pp 411–447. https://doi.org/10.1016/B978-0-444-59528-7.00011-9.

(61) Zhou, C.-Y.; Bai, Y.; Wang, C.; Li, C.-B.; Xu, X.-L.; Pan, D.-D.; Cao, J.-X.; Zhou, G.-H. 1H NMR-Based Metabolomics and Sensory Evaluation Characterize Taste Substances of Jinhua Ham with Traditional and Modern Processing Procedures. *Food Control* **2021**, *126*, 107873. https://doi.org/10.1016/j.foodcont.2021.107873.

(62) Ferrari, E.; Foca, G.; Vignali, M.; Tassi, L.; Ulrici, A. Adulteration of the Anthocyanin Content of Red Wines: Perspectives for Authentication by Fourier Transform-Near InfraRed and 1H NMR Spectroscopies. *Analytica Chimica Acta* **2011**, *701* (2), 139–151. https://doi.org/10.1016/j.aca.2011.05.053.

(63) Longobardi, F.; Ventrella, A.; Napoli, C.; Humpfer, E.; Schütz, B.; Schäfer, H.; Kontominas, M. G.; Sacco, A. Classification of Olive Oils According to Geographical Origin by Using 1H NMR Fingerprinting Combined with Multivariate Analysis. *Food Chemistry* **2012**, *130* (1), 177–183. https://doi.org/10.1016/j.foodchem.2011.06.045.

(64) Hu, Y.; Wang, S.; Wang, S.; Lu, X. Application of Nuclear Magnetic Resonance Spectroscopy in Food Adulteration Determination: The Example of Sudan Dye I in Paprika Powder. *Sci Rep* **2017**, *7* (1), 2637. https://doi.org/10.1038/s41598-017-02921-8.

(65) Picone, G.; Mengucci, C.; Capozzi, F. The NMR Added Value to the Green Foodomics Perspective: Advances by Machine Learning to the Holistic View on Food and Nutrition. *Magnetic Resonance in Chemistry* **2022**, *60* (7), 590–596. https://doi.org/10.1002/mrc.5257.

(66) Macias, S.; Kirma, J.; Yilmaz, A.; Moore, S. E.; McKinley, M. C.; McKeown, P. P.; Woodside, J. V.; Graham, S. F.; Green, B. D. Application of 1H-NMR Metabolomics for the Discovery of Blood Plasma Biomarkers of a Mediterranean Diet. *Metabolites* **2019**, *9* (10), 201. https://doi.org/10.3390/metabo9100201.

(67) 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.

(68) 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 40* (1), 1–15. https://doi.org/10.1080/01652176.2019.1707907.

(69) Tran, H.; McConville, M.; Loukopoulos, P. Metabolomics in the Study of Spontaneous Animal Diseases. *J Vet Diagn Invest* **2020**, *32* (5), 635–647. https://doi.org/10.1177/1040638720948505.

(70) Scarsella, E.; Segato, J.; Zuccaccia, D.; Swanson, K. S.; Stefanon, B. An Application of Nuclear Magnetic Resonance Spectroscopy to Study Faecal Canine Metabolome. *Italian Journal of Animal Science* **2021**, *20* (1), 887–895. https://doi.org/10.1080/1828051X.2021.1925602.

(71) Ratcliffe, R. G.; Shachar-Hill, Y. Probing Plant Metabolism with Nmr. *Annual Review of Plant Physiology and Plant Molecular Biology* **2001**, *52* (1), 499–526. https://doi.org/10.1146/annurev.arplant.52.1.499.

(72) Selegato, D. M.; Pilon, A. C.; Carnevale Neto, F. Plant Metabolomics Using NMR Spectroscopy. *NMR-Based Metabolomics* **2019**, 345–362. https://doi.org/10.1007/978-1-4939-9690-2_19.

(73) Wilkop, T. E.; Wang, M.; Heringer, A.; Singh, J.; Zakharov, F.; Krishnan, V. V.; Drakakaki, G. NMR Spectroscopy Analysis Reveals Differential Metabolic Responses in Arabidopsis Roots and Leaves Treated with a Cytokinesis Inhibitor. *PLOS ONE* **2020**, *15* (11), e0241627. https://doi.org/10.1371/journal.pone.0241627.

(74) *J. C. Lindon, J. K. Nicholson and E. Holmes, "The Handbook of Metabonomics and Metabolomics," Access Online via Elsevier, Amsterdam, 2011.* - *References* - *Scientific Research*

Publishing. https://www.scirp.org/%28S%28351jmbntvnsjt1aadkposzje%29%29/reference/referencespapers.aspx?referenceid=1060152 (accessed 2022-08-23).

(75) Krishnan, P.; Kruger, N. J.; Ratcliffe, R. G. Metabolite Fingerprinting and Profiling in Plants Using NMR. *Journal of Experimental Botany* **2005**, *56* (410), 255–265. https://doi.org/10.1093/jxb/eri010.

(76) Emwas, A.-H.; Roy, R.; McKay, R. T.; Tenori, L.; Saccenti, E.; Gowda, G. A. N.; Raftery, D.; Alahmari, F.; Jaremko, L.; Jaremko, M.; Wishart, D. S. NMR Spectroscopy for Metabolomics Research. *Metabolites* **2019**, *9* (7), 123. https://doi.org/10.3390/metabo9070123.

(77) Capozzi, F.; Laghi, L.; Belton, P. S. Magnetic Resonance in Food Science: Defining Food by Magnetic Resonance. *Magnetic resonance in food science: defining food by magnetic resonance*. **2015**.

(78) *NMR-Based Metabolomics*; 2018. https://doi.org/10.1039/9781782627937.

(79) 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.

(80) 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.

(81) 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.

(82) Crook, A. A.; Powers, R. Quantitative NMR-Based Biomedical Metabolomics: Current Status and Applications. *Molecules* **2020**, *25* (21), 5128. https://doi.org/10.3390/molecules25215128.

(83) Mckay, R. T. How the 1D-NOESY Suppresses Solvent Signal in Metabonomics NMR Spectroscopy: An Examination of the Pulse Sequence Components and Evolution. *Concepts in Magnetic Resonance Part A* **2011**, *38A* (5), 197–220. https://doi.org/10.1002/cmr.a.20223.

(84) 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.

(85) Wu, D. H.; Chen, A. D.; Johnson, C. S. An Improved Diffusion-Ordered Spectroscopy Experiment Incorporating Bipolar-Gradient Pulses. *Journal of Magnetic Resonance, Series A* **1995**, *115* (2), 260–264. https://doi.org/10.1006/jmra.1995.1176.

(86) Ludwig, C.; Viant, M. R. Two-Dimensional J-Resolved NMR Spectroscopy: Review of a Key Methodology in the Metabolomics Toolbox. *Phytochem Anal* **2010**, *21* (1), 22–32. https://doi.org/10.1002/pca.1186.

(87) 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.

(88) 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.

(89) Bruker Announces AVANCE™-IVDr as a Standardized NMR Platform for Clinical Screening and In Vitro Diagnostics - 2013 - Wiley Analytical Science.

(90) 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. *Analytica Chimica Acta* **2012**, *750*, 82–97. https://doi.org/10.1016/j.aca.2012.05.049.

(91) Sousa, S. A. A.; Magalhães, A.; Ferreira, M. M. C. Optimized Bucketing for NMR Spectra: Three Case Studies. *Chemometrics and Intelligent Laboratory Systems* **2013**, *122*, 93–102. https://doi.org/10.1016/j.chemolab.2013.01.006.

(92) Dieterle, F.; Ross, A.; Schlotterbeck, G.; Senn, H. Probabilistic Quotient Normalization as Robust Method to Account for Dilution of Complex Biological Mixtures. Application in 1H NMR Metabonomics. *Anal. Chem.* **2006**, *78* (13), 4281–4290. https://doi.org/10.1021/ac051632c.

(93) Wishart, D. S.; Feunang, Y. D.; Marcu, A.; Guo, A. C.; Liang, K.; Vázquez-Fresno, R.; Sajed, T.; Johnson, D.; Li, C.; Karu, N.; Sayeeda, Z.; Lo, E.; Assempour, N.; Berjanskii, M.; Singhal, S.; Arndt, D.; Liang, Y.; Badran, H.; Grant, J.; Serra-Cayuela, A.; Liu, Y.; Mandal, R.; Neveu, V.; Pon, A.; Knox, C.; Wilson, M.; Manach, C.; Scalbert, A. HMDB 4.0: The Human Metabolome Database for 2018. *Nucleic Acids Res.* **2018**, *46* (D1), D608–D617. https://doi.org/10.1093/nar/gkx1089.

(94) 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.

(95) 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.

(96) *BBIOREFCODE*. https://www.bruker.com/es/products-and-solutions/mr/nmr-clinical-research-solutions/bbiorefcode.html (accessed 2022-08-21).

(97) Takis, P. G.; Vuckovic, I.; Tan, T.; Denic, A.; Lieske, J. C.; Lewis, M. R.; Macura, S. NMRpQuant: An Automated Software for Large Scale Urinary Total Protein Quantification by One-Dimensional 1H NMR Profiles. *Bioinformatics* **2022**, btac502. https://doi.org/10.1093/bioinformatics/btac502.

(98) 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.

(99) Ravanbakhsh, S.; Liu, P.; Bjordahl, T. C.; Mandal, R.; Grant, J. R.; Wilson, M.; Eisner, R.; Sinelnikov, 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.

(100) Tardivel, P. J. C.; Canlet, C.; Lefort, G.; Tremblay-Franco, M. A.; 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), Non Paginé. https://doi.org/10.1007/s11306-017-1244-5.

(101) Weljie, A. M.; Newton, J.; Mercier, P.; Carlson, E.; Slupsky, C. M. Targeted Profiling: Quantitative Analysis of 1H NMR Metabolomics Data. *Anal. Chem.* **2006**, *78* (13), 4430–4442. https://doi.org/10.1021/ac060209g.

(102) Geladi, P. Herman Wold, the Father of PLS. *Chemometrics and Intelligent Laboratory Systems* **1992**, *15* (1), vii–viii. https://doi.org/10.1016/0169-7439(92)80021-U.

(103) Wold, S.; Sjöström, M.; Eriksson, L. PLS-Regression: A Basic Tool of Chemometrics. *Chemometrics and Intelligent Laboratory Systems* **2001**, *58* (2), 109–130. https://doi.org/10.1016/S0169-7439(01)00155-1.

(104) Abdi, H. Partial Least Squares Regression and Projection on Latent Structure Regression (PLS Regression). *WIREs Computational Statistics* **2010**, *2* (1), 97–106. https://doi.org/10.1002/wics.51.

(105) Cunningham, P.; Delany, S. J. K-Nearest Neighbour Classifiers: 2nd Edition (with Python Examples). *ACM Comput. Surv.* **2022**, *54* (6), 1–25. https://doi.org/10.1145/3459665.

(106) Kalpić, D.; Hlupić, N.; Lovrić, M. Student's t-Tests. In *International Encyclopedia of Statistical Science*; Lovric, M., Ed.; Springer: Berlin, Heidelberg, 2011; pp 1559–1563. https://doi.org/10.1007/978-3-642-04898-2_641.

(107) St»hle, L.; Wold, S. Analysis of Variance (ANOVA). *Chemometrics and Intelligent Laboratory Systems* **1989**, *6* (4), 259–272. https://doi.org/10.1016/0169-7439(89)80095-4.

(108) Friedman, M. A Comparison of Alternative Tests of Significance for the Problem of m Rankings. *The Annals of Mathematical Statistics* **1940**, *11* (1), 86–92. https://doi.org/10.1214/aoms/1177731944.

(109) Straube, J.; Gorse, A.-D.; Team, P. C. of E.; Huang, B. E.; Cao, K.-A. L. A Linear Mixed Model Spline Framework for Analysing Time Course 'Omics' Data. *PLOS ONE* **2015**, *10* (8), e0134540. https://doi.org/10.1371/journal.pone.0134540.

(110) Bonferroni, C. E. *Il calcolo delle assicurazioni su gruppi di teste*; Tipografia del Senato, 1935.

(111) Bland, J. M.; Altman, D. G. Multiple Significance Tests: The Bonferroni Method. *BMJ* **1995**, *310* (6973), 170. https://doi.org/10.1136/bmj.310.6973.170.

(112) 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**, *57* (1), 289–300.

(113) Hauke, J.; Kossowski, T. Comparison of Values of Pearson's and Spearman's Correlation Coefficients on the Same Sets of Data. *Quaestiones Geographicae* **2011**, *30* (2), 87–93. https://doi.org/10.2478/v10117-011-0021-1.

(114) Mair, P.; Wilcox, R. Robust Statistical Methods in R Using the WRS2 Package. *Behav Res Methods* **2020**, *52* (2), 464–488. https://doi.org/10.3758/s13428-019-01246-w.

(115) Suarez-Diez, M.; Saccenti, E. Effects of Sample Size and Dimensionality on the Performance of Four Algorithms for Inference of Association Networks in Metabonomics. *J. Proteome Res.* **2015**, *14* (12), 5119–5130. https://doi.org/10.1021/acs.jproteome.5b00344.

(116) Akhand, M. A. H.; Nandi, R. N.; Amran, S. M.; Murase, K. Context Likelihood of Relatedness with Maximal Information Coefficient for Gene Regulatory Network Inference. In *2015 18th International Conference on Computer and Information Technology (ICCIT)*; 2015; pp 312–316. https://doi.org/10.1109/ICCITechn.2015.7488088.

(117) Hoffman, J. M.; Lyu, Y.; Pletcher, S. D.; Promislow, D. E. L. Proteomics and Metabolomics in Ageing Research: From Biomarkers to Systems Biology. *Essays Biochem.* **2017**, *61* (3), 379–388. https://doi.org/10.1042/EBC20160083.

(118) Jové, M.; Maté, I.; Naudí, A.; Mota-Martorell, N.; Portero-Otín, M.; De la Fuente, M.; Pamplona, R. Human Aging Is a Metabolome-Related Matter of Gender. *J. Gerontol. A Biol. Sci. Med. Sci.* **2016**, *71* (5), 578–585. https://doi.org/10.1093/gerona/glv074.

(119) Molino Lova, R.; Sofi, F.; Pasquini, G.; Gori, A. M.; Vannetti, F.; Abbate, R.; Gensini, G.; Macchi, C. The Mugello Study, a Survey of Nonagenarians Living in Tuscany: Design, Methods and Participants' General Characteristics. *European journal of internal medicine* **2013**, *24*. https://doi.org/10.1016/j.ejim.2013.09.008.

(120) Canfield, C.-A.; Bradshaw, P. C. Amino Acids in the Regulation of Aging and Aging-Related Diseases. *Translational Medicine of Aging* **2019**, *3*, 70–89. https://doi.org/10.1016/j.tma.2019.09.001.

(121) Fukagawa, N. K. Protein and Amino Acid Supplementation in Older Humans. *Amino Acids* **2013**, *44* (6), 1493–1509. https://doi.org/10.1007/s00726-013-1480-6.

(122) Aboyans Victor; Criqui Michael H.; Abraham Pierre; Allison Matthew A.; Creager Mark A.; Diehm Curt; Fowkes F. Gerry R.; Hiatt William R.; Jönsson Björn; Lacroix Philippe; Marin Benôıt; McDermott Mary M.; Norgren Lars; Pande Reena L.; Preux Pierre-Marie; Stoffers H.E. (Jelle); Treat-Jacobson Diane. Measurement and Interpretation of the Ankle-Brachial Index. *Circulation* **2012**, *126* (24), 2890–2909. https://doi.org/10.1161/CIR.0b013e318276fbcb.

(123) Brinton, R. D. Estrogen Regulation of Glucose Metabolism and Mitochondrial Function: Therapeutic Implications for Prevention of Alzheimer's Disease. *Adv. Drug Deliv. Rev.* **2008**, *60* (13–14), 1504–1511. https://doi.org/10.1016/j.addr.2008.06.003.

(124) Rettberg, J. R.; Yao, J.; Brinton, R. D. Estrogen: A Master Regulator of Bioenergetic Systems in the Brain and Body. *Front Neuroendocrinol* **2014**, *35* (1), 8–30. https://doi.org/10.1016/j.yfrne.2013.08.001.

(125) Chiurchiù, V.; Leuti, A.; Maccarrone, M. Bioactive Lipids and Chronic Inflammation: Managing the Fire Within. *Front Immunol* **2018**, *9*. https://doi.org/10.3389/fimmu.2018.00038.

(126) Garcia, E.; Shalaurova, I.; Matyus, S. P.; Oskardmay, D. N.; Otvos, J. D.; Dullaart, R. P. F.; Connelly, M. A. Ketone Bodies Are Mildly Elevated in Subjects with Type 2 Diabetes Mellitus and Are Inversely Associated with Insulin Resistance as Measured by the Lipoprotein Insulin Resistance Index. *J Clin Med* **2020**, *9* (2). https://doi.org/10.3390/jcm9020321.

(127) Agostini, D.; Zeppa Donati, S.; Lucertini, F.; Annibalini, G.; Gervasi, M.; Ferri Marini, C.; Piccoli, G.; Stocchi, V.; Barbieri, E.; Sestili, P. Muscle and Bone Health in Postmenopausal Women: Role of Protein and Vitamin D Supplementation Combined with Exercise Training. *Nutrients* **2018**, *10* (8). https://doi.org/10.3390/nu10081103.

(128) Auro, K.; Joensuu, A.; Fischer, K.; Kettunen, J.; Salo, P.; Mattsson, H.; Niironen, M.; Kaprio, J.; Eriksson, J. G.; Lehtimäki, T.; Raitakari, O.; Jula, A.; Tiitinen, A.; Jauhiainen, M.; Soininen, P.; Kangas, A. J.; Kähönen, M.; Havulinna, A. S.; Ala-Korpela, M.; Salomaa, V.; Metspalu, A.; Perola, M. A Metabolic View on Menopause and Ageing. *Nat Commun* **2014**, *5* (1), 4708. https://doi.org/10.1038/ncomms5708.

(129) Socha, E.; Koba, M.; Kośliński, P. Amino Acid Profiling as a Method of Discovering Biomarkers for Diagnosis of Neurodegenerative Diseases. *Amino Acids* **2019**, *51* (3), 367–371. https://doi.org/10.1007/s00726-019-02705-6.

(130) Park, S.; Sadanala, K. C.; Kim, E.-K. A Metabolomic Approach to Understanding the Metabolic Link between Obesity and Diabetes. *Mol Cells* **2015**, *38* (7), 587–596. https://doi.org/10.14348/molcells.2015.0126.

(131) Cui, X.; Yu, X.; Sun, G.; Hu, T.; Likhodii, S.; Zhang, J.; Randell, E.; Gao, X.; Fan, Z.; Zhang, W. Differential Metabolomics Networks Analysis of Menopausal Status. *PLoS One* **2019**, *14* (9). https://doi.org/10.1371/journal.pone.0222353.

(132) 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.

(133) Hutson, S. M.; Sweatt, A. J.; Lanoue, K. F. Branched-Chain [Corrected] Amino Acid Metabolism: Implications for Establishing Safe Intakes. *J. Nutr.* **2005**, *135* (6 Suppl), 1557S-64S. https://doi.org/10.1093/jn/135.6.1557S.

(134) Mittelstrass, K.; Ried, J. S.; Yu, Z.; Krumsiek, J.; Gieger, C.; Prehn, C.; Roemisch-Margl, W.; Polonikov, A.; Peters, A.; Theis, F. J.; Meitinger, T.; Kronenberg, F.; Weidinger, S.; Wichmann, H. E.; Suhre, K.; Wang-Sattler, R.; Adamski, J.; Illig, T. Discovery of Sexual Dimorphisms in Metabolic and Genetic Biomarkers. *PLoS Genet* **2011**, *7* (8), e1002215. https://doi.org/10.1371/journal.pgen.1002215.

(135) Ganna, A.; Fall, T.; Salihovic, S.; Lee, W.; Broeckling, C. D.; Kumar, J.; Hägg, S.; Stenemo, M.; Magnusson, P. K. E.; Prenni, J. E.; Lind, L.; Pawitan, Y.; Ingelsson, E. Large-Scale Non-Targeted Metabolomic Profiling in Three Human Population-Based Studies. *Metabolomics* **2015**, *12* (1), 4. https://doi.org/10.1007/s11306-015-0893-5.

(136) Janssen, C. I. F.; Kiliaan, A. J. Long-Chain Polyunsaturated Fatty Acids (LCPUFA) from Genesis to Senescence: The Influence of LCPUFA on Neural Development, Aging, and Neurodegeneration. *Prog Lipid Res* **2014**, *53*, 1–17. https://doi.org/10.1016/j.plipres.2013.10.002.

(137) Cybulska, A. M.; Skonieczna-Żydecka, K.; Drozd, A.; Rachubińska, K.; Pawlik, J.; Stachowska, E.; Jurczak, A.; Grochans, E. Fatty Acid Profile of Postmenopausal Women Receiving, and Not Receiving, Hormone Replacement Therapy. *Int J Environ Res Public Health* **2019**, *16* (21). https://doi.org/10.3390/ijerph16214273.

(138) Johnson, A. A.; Stolzing, A. The Role of Lipid Metabolism in Aging, Lifespan Regulation, and Age-Related Disease. *Aging Cell* **2019**, *18* (6), e13048. https://doi.org/10.1111/acel.13048.

(139) Grabner, G. F.; Zimmermann, R.; Schicho, R.; Taschler, U. Monoglyceride Lipase as a Drug Target: At the Crossroads of Arachidonic Acid Metabolism and Endocannabinoid Signaling. *Pharmacol Ther* **2017**, *175*, 35–46. https://doi.org/10.1016/j.pharmthera.2017.02.033.

(140) Li, Z.; Agellon, L. B.; Allen, T. M.; Umeda, M.; Jewell, L.; Mason, A.; Vance, D. E. The Ratio of Phosphatidylcholine to Phosphatidylethanolamine Influences Membrane Integrity and Steatohepatitis. *Cell Metabolism* **2006**, *3* (5), 321–331. https://doi.org/10.1016/j.cmet.2006.03.007.

(141) Lum, H.; Sloane, R.; Huffman, K. M.; Kraus, V. B.; Thompson, D. K.; Kraus, W. E.; Bain, J. R.; Stevens, R.; Pieper, C. F.; Taylor, G. A.; Newgard, C. B.; Cohen, H. J.; Morey, M. C. Plasma Acylcarnitines Are Associated with Physical Performance in Elderly Men. *J Gerontol A Biol Sci Med Sci* **2011**, *66* (5), 548–553. https://doi.org/10.1093/gerona/glr006.

(142) Sommer, F.; Bäckhed, F. The Gut Microbiota--Masters of Host Development and Physiology. *Nat Rev Microbiol* **2013**, *11* (4), 227–238. https://doi.org/10.1038/nrmicro2974.

(143) Collins, M. D.; Gibson, G. R. Probiotics, Prebiotics, and Synbiotics: Approaches for Modulating the Microbial Ecology of the Gut. *Am J Clin Nutr* **1999**, *69* (5), 1052S-1057S. https://doi.org/10.1093/ajcn/69.5.1052s.

(144) Franchini, M.; Favaloro, E. J.; Targher, G.; Lippi, G. ABO Blood Group, Hypercoagulability, and Cardiovascular and Cancer Risk. *Critical Reviews in Clinical Laboratory Sciences* **2012**, *49* (4), 137–149. https://doi.org/10.3109/10408363.2012.708647.

(145) Lindström, S.; Wang, L.; Smith, E. N.; Gordon, W.; van Hylckama Vlieg, A.; de Andrade, M.; Brody, J. A.; Pattee, J. W.; Haessler, J.; Brumpton, B. M.; Chasman, D. I.; Suchon, P.; Chen, M.-H.; Turman, C.; Germain, M.; Wiggins, K. L.; MacDonald, J.; Braekkan, S. K.; Armasu, S. M.; Pankratz, N.; Jackson, R. D.; Nielsen, J. B.; Giulianini, F.; Puurunen, M. K.; Ibrahim, M.; Heckbert, S. R.; Damrauer, S. M.; Natarajan, P.; Klarin, D.; The Million Veteran Program; de Vries, P. S.; Sabater-Lleal, M.; Huffman, J. E.; The CHARGE Hemostasis Working Group; Bammler, T. K.; Frazer, K. A.; McCauley, B. M.; Taylor, K.; Pankow, J. S.; Reiner, A. P.; Gabrielsen, M. E.; Deleuze, J.-F.; O'Donnell, C. J.; Kim, J.; McKnight, B.; Kraft, P.; Hansen, J.-B.; Rosendaal, F. R.; Heit, J. A.; Psaty, B. M.; Tang, W.; Kooperberg, C.; Hveem, K.; Ridker, P. M.; Morange, P.-E.; Johnson, A. D.; Kabrhel, C.; Trégouët, D.-A.; Smith, N. L.; on behalf of the INVENT Consortium. Genomic and Transcriptomic Association Studies Identify 16 Novel Susceptibility Loci for Venous Thromboembolism. *Blood* **2019**, *134* (19), 1645–1657. https://doi.org/10.1182/blood.2019000435.

(146) Franchini, M.; Lippi, G. The Intriguing Relationship between the ABO Blood Group, Cardiovascular Disease, and Cancer. *BMC Medicine* **2015**, *13* (1), 7. https://doi.org/10.1186/s12916-014-0250-y.

(147) Franchini, M.; Mengoli, C.; Capuzzo, E.; Terenziani, I.; Bonfanti, C.; Lippi, G. Correlation between ABO Blood Group, and Conventional Hematological and Metabolic Parameters in Blood Donors. *Semin Thromb Hemost* **2016**, *42* (1), 75–86. https://doi.org/10.1055/s-0035-1564843.

(148) Chen, Y.; Chen, C.; Ke, X.; Xiong, L.; Shi, Y.; Li, J.; Tan, X.; Ye, S. Analysis of Circulating Cholesterol Levels as a Mediator of an Association Between ABO Blood Group and Coronary Heart Disease. *Circulation: Cardiovascular Genetics* **2014**, *7* (1), 43–48. https://doi.org/10.1161/CIRCGENETICS.113.000299.

(149) Bartimaeus, E. S.; Waribo, H. A. Relationship between ABO Blood Groups and Lipid Profile Level in Healthy Adult Residents in Port Harcourt Metropolis, Nigeria. *Journal of Applied Sciences and Environmental Management* **2017**, *21* (6), 1003–1011. https://doi.org/10.4314/jasem.v21i6.1.

(150) Gillum, R. F. Blood Groups, Serum Cholesterol, Serum Uric Acid, Blood Pressure, and Obesity in Adolescents. *J Natl Med Assoc* **1991**, *83* (8), 682–688.

(151) Liu, N.; Zhang, T.; Ma, L.; Wang, H.; Li, H. Association between ABO Blood Groups and Risk of Coronavirus Disease 2019. *Medicine (Baltimore)* **2020**, *99* (33), e21709. https://doi.org/10.1097/MD.0000000000021709.

(152) Kanbay, M.; Yildirir, A.; Ulus, T.; Bilgi, M.; Kucuk, A.; Muderrisoglu, H. Rhesus Positivity and Low High-Density Lipoprotein Cholesterol: A New Link? *Asian Cardiovasc Thorac Ann* **2006**, *14* (2), 119–122. https://doi.org/10.1177/021849230601400208.

(153) Mashahit, M.; Khattab, D.; Zaki, O.; Owis, A. Relation between ABO/Rh Blood Groups and Metabolic Syndrome. *Fayoum University Medical Journal* **2020**, *5* (1), 15–31. https://doi.org/10.21608/fumj.2020.77466.

(154) 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.

(155) 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.

(156) 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.

(157) 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.

(158) 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.

(159) *Association between change in plasma triglyceride levels and risk of stroke and carotid atherosclerosis: systematic review and meta-regression anal... - PubMed - NCBI*. https://www.ncbi.nlm.nih.gov/pubmed/20457452 (accessed 2020-05-12).

(160) Berthet, Q.; Rigollet, P. Optimal Detection of Sparse Principal Components in High Dimension. *Ann. Statist.* **2013**, *41* (4), 1780–1815. https://doi.org/10.1214/13-AOS1127.

(161) 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.

(162) Sung, H.; Ferlay, J.; Siegel, R. L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians n/a* (n/a). https://doi.org/10.3322/caac.21660.

(163) Siegel, R.; Desantis, C.; Jemal, A. Colorectal Cancer Statistics, 2014. *CA Cancer J Clin* **2014**, *64* (2), 104–117. https://doi.org/10.3322/caac.21220.

(164) Zhu, J.; Djukovic, D.; Deng, L.; Gu, H.; Himmati, F.; Chiorean, E. G.; Raftery, D. Colorectal Cancer Detection Using Targeted Serum Metabolic Profiling. *J. Proteome Res.* **2014**, *13* (9), 4120–4130. https://doi.org/10.1021/pr500494u.

(165) Tang, Z.; Xu, Z.; Zhu, X.; Zhang, J. New Insights into Molecules and Pathways of Cancer Metabolism and Therapeutic Implications. *Cancer Communications* **2021**, *41* (1), 16–36. https://doi.org/10.1002/cac2.12112.

(166) La Vecchia, S.; Sebastián, C. Metabolic Pathways Regulating Colorectal Cancer Initiation and Progression. *Seminars in Cell & Developmental Biology* **2020**, *98*, 63–70. https://doi.org/10.1016/j.semcdb.2019.05.018.

(167) Neurauter, G.; Grahmann, A. V.; Klieber, M.; Zeimet, A.; Ledochowski, M.; Sperner-Unterweger, B.; Fuchs, D. Serum Phenylalanine Concentrations in Patients with Ovarian Carcinoma Correlate with Concentrations of Immune Activation Markers and of Isoprostane-8. *Cancer Lett* **2008**, *272* (1), 141–147. https://doi.org/10.1016/j.canlet.2008.07.002.

(168) Ploder, M.; Neurauter, G.; Spittler, A.; Schroecksnadel, K.; Roth, E.; Fuchs, D. Serum Phenylalanine in Patients Post Trauma and with Sepsis Correlate to Neopterin Concentrations. *Amino Acids* **2008**, *35* (2), 303–307. https://doi.org/10.1007/s00726-007-0625-x.

(169) Long, Y.; Sanchez-Espiridion, B.; Lin, M.; White, L.; Mishra, L.; Raju, G. S.; Kopetz, S.; Eng, C.; Hildebrandt, M. A. T.; Chang, D. W.; Ye, Y.; Liang, D.; Wu, X. Global and Targeted Serum Metabolic Profiling of Colorectal Cancer Progression. *Cancer* **2017**, *123* (20), 4066–4074. https://doi.org/10.1002/cncr.30829.