eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

Doctoral thesis

NMR-based metabolomics in precision medicine. Application to metabolic syndrome and COVID-19.

Presented by:

Chiara Bruzzone

Supervised by:

Óscar Millet Aguilar-Galindo

Faculty of Science and Technology

Department of Biochemistry and Molecular Biology

Doctoral Programme in Molecular Biology and Biomedicine

Leioa, 2023

Contents:

Abstract

NMR based metabolomics in precision medicine. Application to metabolic syndrome and COVID-19.

Precision medicine is considered as an innovative approach to traditional medicine, for a tailored treatment of patient's pathologies according to individual characteristics. Metabolomics, using different technologies like nuclear magnetic resonance (NMR), offers the possibility to observe the metabolic changes, to identify biomarkers and to study the molecular mechanism involved in a specific disease, which are considered important factors for a better diagnosis and treatment. Here we decided to apply NMR-metabolomics to the study of two unrelated disease case studies: metabolic syndrome and COVID-19, the recent worldwide infection caused by SARS-CoV-2 virus.

A large cohort of urine (11,127 individuals) and serum (8470 individuals) samples was used to investigate the molecular signature of the metabolic syndrome. NMR metabolomics showed to be sensitive to this disorder, with all the contributing risk factors involved in the development of MetS represented by at least one of the identified metabolites from the conduced analysis. Disease progression was accompanied by a continuous variation (up- or down-regulation) of the pertinent metabolites, allowing the obtention of a metabolic model that can discriminate between individuals with and without MetS with statistical significance, adding an unprecedented diagnostic molecular dimension to the set of risk factors that currently describe it.

To understand the characteristic aspects of SARS-CoV-2 infection, a cohort of 263 COVID-19 patients in the acute phase of the disease and 280 pre COVID-19 control subjects was analysed. The observed metabolic and lipidomic changes that characterize the infected subjects were consistent with a model in which SARS-CoV-2 infection, in addition to respiratory system impairment, was producing a systemic infection involving different organs dysfunction, dyslipidemia and oxidative stress.

Resumen

La medicina de precisión debe considerarse como un enfoque innovador de la medicina tradicional ya que se centra en un tratamiento más personalizado de las patologías de los pacientes en función de sus características individuales. El estudio del mecanismo molecular asociado a una enfermedad y la identificación de los biomarcadores implicados en el desarrollo y la progresión de dicha patología se consideran factores esenciales para obtener un mejor diagnóstico y una optimización de la terapia. La metabolómica proporciona una imagen completa del organismo y permite observar los cambios metabólicos específicos que caracterizan a un individuo debido a su condición de salud o a factores externos (dieta, medioambiente, ejercicio, etc). En concreto, la metabolómica por resonancia magnética nuclear (RMN) es especialmente adecuada para este tipo de análisis, ya que las mediciones se realizan de forma rápida y sencilla, es una técnica que ofrece la posibilidad de identificar y cuantificar numerosos metabolitos, las muestras no necesitan derivatización y por tanto se acelera la obtención de resultados. En este trabajo decidimos aplicar las ventajas que ofrece esta técnica al estudio de dos enfermedades no relacionadas entre sí: el síndrome metabólico y a la reciente pandemia mundial causada por el virus SARS-CoV-2.

El síndrome metabólico (SM) es una patología compleja que se origina por una combinación de distintos factores de riesgo, como alteraciones del metabolismo de la glucosa, obesidad, niveles elevados de triglicéridos, valores bajos de colesterol HDL así como la presencia de hipertensión. Se ha demostrado que padecer este trastorno puede conllevar un mayor riesgo de desarrollo de enfermedades cardiovasculares y, por lo tanto, un incremento de la mortalidad. Por este motivo y debido al creciente número de personas afectadas por el SM, actualmente esta enfermedad representa un importante problema sanitario a nivel mundial. Sin embargo, y a pesar de todos los estudios realizados en este campo, el SM todavía no se conoce bien. Basta decir que no existe un consenso para su definición médica y el diagnóstico se basa únicamente en la sintomatología compatible.

En este estudio se ha utilizado la metabolómica por RMN para investigar detalladamente las características moleculares del SM. Para ello se ha empleado una

gran cohorte de muestras de orina (11.127 individuos) y suero (8.470 individuos), pertenecientes principalmente a trabajadores de diferentes sectores de las empresas del País Vasco y a otras sub-cohortes de muestras de menor tamaño de otras regiones españolas o de Europa. La cohorte se diseñó para contemplar todas las posibles condiciones intermedias que pueden llevar al desarrollo de SM., desde sujetos sin ningún factor de riesgo (FR) hasta individuos con SM. Por esta razón, para poder describir todas las combinaciones posibles fue necesario clasificar las muestras, creando un código binario que indicara la presencia (1) o la ausencia (0) de cada uno de los cuatro FR (FR_1 = diabetes, FR_2 = obesidad, FR_3 = dislipidemia, FR_4 = hipertensión) generando finalmente16 condiciones diferentes (2^4).

En primer lugar, se analizaron los perfiles medios obtenidos por RMN a partir de la orina, para cada una de las condiciones examinadas y que representan su huella molecular característica. Mediante un análisis PCA no supervisado se obtuvo una separación de los diferentes perfiles según el FR presentado, evidenciando especialmente el gran impacto que la diabetes y la hipertensión tenían en este proceso.

Todas las condiciones generadas se compararon mediante un análisis univariante con los individuos aparentemente sanos (0000), utilizando como variables para las comparaciones la intensidad de los *bins* de los espectros de orina adquiridos. Los *bins* que resultaron estadísticamente significativos en al menos una de las comparaciones de los perfiles con la ausente de factores de riesgo bajo consideración (0000), se representaron en un *heatmap*, y se asignaron los metabolitos asociados a ellos observando los espectros de RMN en el área correspondiente a cada uno de los *bins*. Se pudo observar que, una vez más, los perfiles se agrupaban en cuatro grupos diferentes, de la misma forma que en el PCA. Utilizando como herramienta un análisis de *cluster* no supervisado, se ordenaron los distintos perfiles con una tendencia que se asemejaba a la definición de síndrome metabólico según la OMS, el EGIR y la AACE, con la diabetes como factor más relevante (y obligatorio) para el diagnóstico. De hecho, también fue posible observar una clara separación entre los perfiles diabético (1XXX) y no diabético (0XXX). Además, la variación en la concentración de los metabolitos estaba en consonancia con la progresión hacia estadios más avanzados de SM.

Posteriormente, se describieron las correlaciones encontradas en los metabolitos relacionados con los FR estudiados y el SM, o se investigó una posible explicación de su presencia en la orina. Además, este estudio ha sido capaz de identificar por primera vez la asociación entre el SM y algunos metabolitos. La glucosa fue uno de los metabolitos más destacados en el *heatmap* y su mayor concentración se relacionó, como era de esperar, con la diabetes y la resistencia a la insulina. Entre todos los metabolitos identificados, la mayoría estaban asociados a la diabetes y la obesidad. Por ejemplo, el p-cresol sulfato, el ácido 4-hidroxifenilpirúvico (4-HPPA) y el maltitol se vincularon con problemas en el metabolismo de la glucosa, mientras que el ácido salicilúrico, el ácido nicotinúrico, el N-óxido de la trimetilamina (TMAO), la trigonelina y el triptófano se asociaron con la obesidad. Se buscó una explicación lógica para la asociación de estos metabolitos con los FR asociados al SM o con el propio síndrome, y se observó que determinados metabolitos estaban claramente conectados al menos a uno de los factores de riesgo del SM, aunque en el caso de la dislipidemia y la hipertensión esta asociación se produjo en menor medida. Por tanto, el análisis de orina mediante RMN ha demostrado ser sensible a la detección del SM.

Para profundizar el análisis del deterioro del metabolismo en función del número de FR presentados, se realizó un análisis multivariante considerando otra vez el perfil medio por cada una de las condiciones, pero utilizando solo los *bins* más relevantes según el análisis anterior. Para comparar los perfiles, se calcularon las distancias de correlación de Spearman (escaladas entre 0 y 1) entre cada perfil y el 0000. De esta manera se pudo observar su similitud con esta condición y cuanto de lejos estaba de desarrollar SM. Así, fue posible confirmar la variación del metabolismo de la orina en función del número creciente de factores de riesgo presentados, aunque no todos ellos evidenciaron la misma influencia en la progresión hacia el SM. De hecho, la manifestación concomitante de algunos FR en determinadas condiciones mostró un metabotipo más cambiante y patogénico que otras combinaciones. Éste fue el caso de los perfiles 1111, 1101, 1011 y 1001, que presentaron una firma molecular similar. Esta firma molecular característica nos llevó a sugerir una nueva definición de síndrome metabólico, parcialmente diferente de las existentes, basada no en la mera presencia o ausencia de FR específicos, sino en función de la contribución que cada

uno de ellos posee en los cambios metabólicos producidos. Este análisis refleja el rol principal que desempeña la diabetes en el SM.

Estos resultados permitieron generar unos modelos con el fin de distinguir mejor las condiciones SM y no SM desde un punto de vista metabólico. Los tres modelos construidos, basados meramente en el análisis metabolómico de las muestras de orina medidas, demostraron ser capaces de distinguir a los sujetos con SM obteniendo AUROC con valores entre 0.83 y 0.86. Al observar los modelos, el primero, que incluía los problemas del metabolismo de la glucosa como factor obligatorio, mostró un mejor rendimiento (valor AUROC: 0.863), lo que reflejaba la importancia de la glucosa en orina para este modelo metabólico, también evidenciada previamente. De esta manera este estudio constató que esta novedosa definición de SM añadía una dimensión metabólica a las ya existentes. Algunas de las definiciones utilizadas actualmente (OMS, EGIR y AACE) podrían considerarse más apropiadas para el diagnóstico por la importancia atribuida a los factores de riesgo que son realmente los responsables de la aparición de alteraciones metabólicas. De hecho, la diabetes demostró sin duda ser un factor muy relevante en este síndrome, como evidencian todos los resultados propuestos, pero este estudio en orina desveló que la hipertensión es un factor de riesgo clave en el agravamiento de este trastorno. Además, el proceso de envejecimiento y la enfermedad del hígado graso no alcohólico también se consideraron factores de riesgo que pueden aumentar la probabilidad de presentar SM, pero en nuestro estudio se demostró que no interferían directamente en la discriminación metabólica de este síndrome.

De manera parecida se realizó el análisis en muestras de suero con el objetivo de incrementar el conocimiento de los aspectos metabólicos característicos del SM, y de algunos de los FR implicados en su desarrollo. Al igual que en orina, en primer lugar, se realizó un análisis PCA no supervisado del perfil medio de RMN, obtenido para cada una de las condiciones estudiadas. En este PCA se representó la huella molecular característica que tiene en cuenta la expresión de metabolitos y lipoproteínas, y de esta manera observar la distribución de los diferentes perfiles según el RF presentado. Esto evidenció una clara separación entre los sujetos dislipidémicos y no dislipidémicos

debido a que, en las muestras de suero, los pacientes con dislipidemia mostraban un desequilibrio del perfil de lipoproteínas.

Para investigar más a fondo el perfil metabólico característico de las muestras de suero, se realizó un análisis univariante con el fin de comparar todas las condiciones de SM con respecto a la asintomática (0000). En este caso, se utilizaron como variables las concentraciones de metabolitos, a partir del informe de Bruker obtenido tras cada medición de suero. El análisis univariante del suero, además de la diabetes, que mostró su influencia principalmente a través de la elevada concentración de glucosa detectada en el torrente sanguíneo (y en la orina excretada), evidenció la relevancia de otros dos factores de riesgo, la obesidad y la dislipemia (con una alteración metabólica observada en relación con la presencia de estos dos FR). Así, algunas de las variaciones más relevantes en las concentraciones de metabolitos aparecieron en perfiles caracterizados por la presencia de dislipidemia, obesidad o ambos factores de riesgo presentes al mismo tiempo. Además, como se evidenció anteriormente también en el resultado del análisis univariante de orina, la variación de la concentración de metabolitos se correlacionaba con una progresión hacia la aparición del SM. Muchos de los metabolitos alterados se encontraron asociados con algunos de los factores de riesgo o directamente con el SM, demostrando la sensibilidad de los espectros de protón de RMN tanto en suero como en orina para el estudio de este desorden.

De particular interés fue el estudio de la expresión de las lipoproteínas, en un intento de comprender más profundamente los aspectos de la dislipidemia, ya que resultó ser la mejor discriminante entre los diferentes perfiles de suero. Por eso se realizó un análisis univariante comparando todas las condiciones creadas con respecto a la asintomática (0000) cuantificando las diferentes lipoproteínas. Se observaron cambios importantes a nivel lipoproteico en los pacientes afectados por dislipidemia y SM, con un aumento de los niveles de triglicéridos y especialmente de lipoproteínas VLDL y *small dense* LDL, y una disminución en los niveles de colesterol HDL. Debido a los cambios metabólicos y lipidómicos observados, se estudió, tanto en los metabolitos como en las lipoproteínas, y mediante análisis multivariante, la correlación entre los perfiles intermedios generados para cada una de las condiciones en estudio en comparación con la sana. Se pudo observar una desregulación metabólica más evidente

en el perfil del espectro de suero en función del número creciente de factores de riesgo incluidos, pero en este caso, el FR que causaba el mayor grado de variación fue la dislipidemia. De hecho, la condición 0010, con un solo FR, ya mostraba un cambio significativo con respecto al perfil 0000. A pesar de ello, se pudo observar que la presencia concomitante de determinados FR, junto con la dislipidemia, aumentaba el estadio patogénico de SM. Las condiciones con diabetes y/u obesidad, junto con la dislipidemia, como resulta especialmente evidente para los perfiles 0110 y 1010, fueron más patógenas, mientras que, al contrario de lo observado anteriormente en el análisis de orina, la hipertensión en suero no mostró un impacto tan grave (perfil 0011). El análisis de muestras de suero mediante RMN sigue desarrollándose con la intención de construir un modelo de discriminación de los sujetos afectados por el SM, teniendo en cuenta estas últimas conclusiones. La combinación de los resultados obtenidos del análisis de orina y suero sería óptima para garantizar una mayor precisión en el diseño de un modelo molecular para la discriminación de los subjetos con SM respeto a los no SM, y para saber la probabilidad de que un individuo analizado pueda llegar a desarrollar el SM en función de su perfil metabólico.

La infección por SARS-CoV-2, que comenzó en diciembre de 2019, ha afectado e a la población mundial de una manera nunca antes conocida en la sociedad moderna. Desde el momento del contacto con el virus, los individuos infectados desarrollan sus primeros síntomas después de 2 a 14 días, con un mecanismo de replicación dentro del cuerpo y posterior infección de otros organismos tremendamente eficiente. Muchas personas infectadas no presentan consecuencias graves debido a la enfermedad, pero el 20% de los casos muestran síntomas graves que pueden llevar a la hospitalización, a un tratamiento en cuidados intensivos e incluso causar la muerte.

Para profundizar las causas y características de la enfermedad se estudiaron los cambios metabólicos asociados a ella analizando muestras de suero provenientes de una gran cohorte de muestras de pacientes que incluía 263 pacientes hospitalizados con COVID-19 en la fase aguda y 280 sujetos control cuyas muestras fueron recolectadas previamente a la aparición del COVID-19. Se realizaron diferentes experimentos de RMN para investigar cada aspecto de la huella metabólica característica de esta clase de muestras. A partir de la comparación de los espectros ^{1}H NOESY de las muestras

COVID con las *preCOVID*, se identificaron claras diferencias (incluso mediante un simple análisis visual), que fueron confirmadas posteriormente por el análisis PCA utilizando como variables las cuantificaciones primero de los metabolitos y después de las lipoproteínas.

El posterior análisis univariante realizado también con los metabolitos y las lipoproteínas, evidenció cambios significativos entre las muestras *COVID* y *preCOVID*. Los cuerpos cetónicos, producidos predominantemente en el hígado a partir de acetil-CoA (derivado de la oxidación de ácidos grasos), se encontraban notablemente elevados en los pacientes con COVID-19. Asimismo, el aumento de los ácidos succínico, cítrico, glutámico y pirúvico puede estar relacionado con una desregulación del metabolismo hepático central en los pacientes con COVID-19.

Una de las diferencias más llamativas surge de la distribución de las lipoproteínas: las partículas lipoproteicas muestran un fenotipo alterado en los pacientes COVID-19, con una reducción del tamaño medio de las HDL, un aumento del tamaño medio de las LDL y un mayor nivel de subclases de VLDL de tamaño intermedio. Además, la proporción de Apo-B a Apo-A1, un equilibrio entre partículas aterogénicas y antiaterogénicas, aumentó notablemente en los pacientes con COVID-19. Además. los perfiles de suero de los pacientes examinados mostraron niveles anormalmente elevados de GlycA, que es un biomarcador asociado a los episodios de inflamación de fase aguda.

Nuestros resultados evidencian cambios masivos en los perfiles lipoproteicos y metabolómicos, consistentes con el fenotipo severo encontrado en pacientes COVID-19 que resaltan el carácter sistémico de la enfermedad. Por último, el perfil lipídico sugiere un aumento del riesgo aterogénico y un deterioro del estrés oxidativo.

Chapter 1

Introduction

1.1 Precision Medicine

The US National Research Council in 2011 defined Precision Medicine (PM) as the optimization of the current approach to medical treatment taking into account the susceptibility of certain groups of individuals for a specific pathology and its therapy in order to improve these aspects[1]. Sometimes the term "personalized medicine" has been used as a synonymous of "precision medicine", but this could lead to misinterpretation by referring to the development of a treatment specifically appropriate just for one patient[2]. The first term turns out to be older than the second, but, nowadays, they are often used interchangeably even if the preferred and newest one is precision medicine, which also reflects an evolution in the approach to patient's study[3,4].

Precision medicine is considered a modern medicine strategy with the aim of improving the diagnosis, prognosis, and cure of different pathological condition thanks to the combination of the extra information obtained from different omics techniques like genomics, proteomics, or metabolomics, together with clinical information of the patients, lifestyle habits, diet or social condition[5,6].

Multi-omics data are collected from huge cohort of samples to ensure a better understanding of a given disease and the best therapeutic approach to treat it with the greatest effectiveness[7]. Genomics, transcriptomics, proteomics, and metabolomics are interconnected and each one of these omics technologies allows the study of different aspects of a living system[8–11]. Thanks to genomics it is possible to study the function of genes, their expression and regulation, the whole genome sequence and the identification of the possible risk of insurgence of some type of diseases like cancer or degenerative disorders[5]. In turn, transcriptomics gives information about all the RNA transcripts produced at a specific time in an organism while proteomics allows the investigation of the produced protein in a specific biological context because of the influence of internal or external stimulations[12]. Finally, metabolomics emerges as one of the most powerful techniques to monitor the metabolic changes caused by the

influence of multiple factors allowing the study of the disease progression at a biochemical level[13].

1.2 Metabolomics and precision medicine

Metabolomics consist in the identification and quantification of the small molecules, produced at the cellular level, because of metabolism, and thus present in a biological system and detectable in the blood stream, in urine or other biofluids[14]. The success of metabolomics is related to the advantage offered by this technique complementing the obtained data from the previously mentioned omics approaches[15].

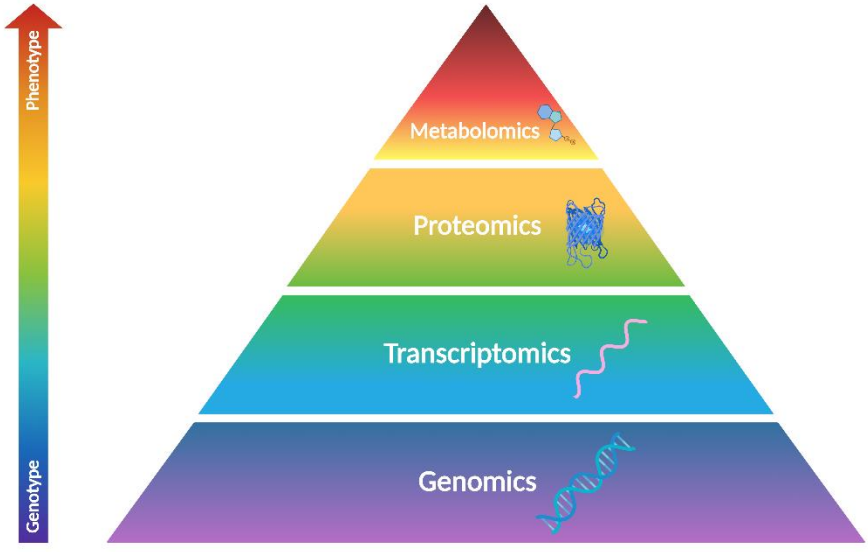

Figure 1.1: Schematic pyramid order existing between the different "omics" approaches. Metabolomics is the technique that best represents the result of the influence of different factors leading to the phenotypic characterization of the individual. Inspired on Carneiro et al. review[176]. (Figure created with Biorender.com).

Metabolomics is the most suitable technique for the detection of all the phenotypical changes that can affect an individual due to the exposure to the environment and can easily detect the metabolic products, called metabolites, produced by the organism (**Figure 1.1**)[16]. This offers an advantage over proteomics, transcriptomics and genomics, since different causes can influence the manifestation of a disorder, including external factors, which are not correctly addressed by these omics

techniques[17,18]. Specifically, the advantage offered by this approach is the monitoring of the changes that can affect an individual due to the influence of external factors, different health conditions, of a drug treatment or simply the impact of diet and lifestyle habits[19]. In this context, metabolomics allows the *in vivo* detection of a huge number of metabolites that can be quantified with only one measurement in a fast and easy way according to the used technique for the analysis[20].

The greatest advantage of metabolomics comes from its ability to provide a complete picture of the organism under study (metabotype) at the time of analysis and to identify biomarkers essentials for a better diagnosis and treatment of the patients[21]. On the other hand, the identification and validation of biomarkers is one of the final goals in precision medicine, key for the diagnosis of a disease, the identification of the most suitable therapy, the understanding of the associated molecular mechanism and the involved pathways[6,15,22]. For all these reasons, metabolomics is considered as a valuable tool in precision medicine, along with other advantages such as providing rapid and reproducible diagnostic tools and presenting predetermined rules for sample handling and management that are some of the main recommendations for precision medicine initiatives[23].

1.2.1 Targeted versus untargeted metabolomics

Metabolomics can be divided into targeted and untargeted analysis. The first one consists in the identification (and quantification whenever possible) of a specific set of compounds (the target) while, in untargeted metabolomics the complete experimental dataset is examined collectively in order to determine the most relevant metabolites and identify them[24].

Targeted metabolomics is usually used to investigate the different expression of a pre-determined set of metabolites given a specific condition. Thanks to statistical analysis, it is possible to compare the concentrations of the compounds of interest for a better understanding of the pathology under study[25]. In turn, untargeted metabolomics is mainly used to compare health versus pathogenic conditions, to identify differences between distinct states trying to find potential biomarkers for the diagnosis of a specific disease[26]. Data analysis enables examining large groups of samples and to identify the

most relevant statical differences[27]. In this way the metabolites highlighted in the comparisons should be assigned so the involved biological pathway could be identified[28]. Metabolites can be assigned either using specific programs like Chenomx, with the help of previous bibliography or by consulting databases like the Human Metabolome Databases (HMDB), which includes information about most of the small molecules present in the human body[29,30]. These databases often contain spectroscopic information of the metabolites, including the mass spectrometry (MS) or nuclear magnetic resonance (NMR) spectra, and they will be better described in Chapter 3.

1.2.2 Analytical methods in metabolomics

The main techniques most often used in the metabolomics field are chromatography coupled to MS and NMR[31]. Indeed, NMR spectroscopy and MS are two powerful techniques in the field of metabolomics due to their ability to detect a huge number of metabolites[32]. Through their application it is possible to identify and quantify (either in an absolute or relative form) a very large number of metabolites in biological samples like urine, serum and plasma, among other biofluids[33]. Compound identification can often be difficult due to the complexity of the spectra generated by the biological matrix[34]. Fortunately, thanks to the resolution of these methods, it is still possible to recognize a great number of molecules. In addition to resolution, these techniques present various advantages and disadvantages but both have excellent properties and can even be complementary to obtain more information and reliable results[35].

As mentioned, MS is generally characterised by the presence of a previous chemical separation techniques like liquid (LC-MS) or gas chromatography (GC-MS), generally chosen according to the matrix to be analysed[36]. Gas chromatography is especially suited for the detection of volatile compounds while liquid chromatography is mainly used for the analysis of polar and non-polar ones[37]. Moreover, to improve the specificity of the analysis it is possible to couple these separation techniques with two consecutives steps of mass spectrometry so called tandem mass spectrometry (GC-MS/MS and LC-MS/MS) which offers advantages for the determination of certain organic compounds and could also provide structural information[38].

NMR spectroscopy doesn't require any previous preparative step before sample analysis, which is directly inserted inside the instrument never coming into contact with it[39]. NMR has many other applications such as in the organic chemistry field for the elucidation of compound structures, or in protein structural elucidation and to understand ligand-binding interactions[40].

Figure 1.2 summarize the main characteristics differences between MS and NMR. NMR has a lower sensitivity than MS: with NMR it is possible to identify hundreds of metabolites in a biological fluid, while with MS it is possible to reach the identification of thousands of compounds. That said, NMR spectroscopy is much more reproducible with respect to MS. Moreover, sample preparation is usually faster and easier in NMR than MS, the latter usually requiring derivatization of the samples and a separation process to ensure a proper detection of the metabolites. In addition, NMR is a non-destructive technique that allows samples recovery (if necessary) of the sample for further experiments, while MS is destructive and doesn't permit to reuse the prepared sample, while requiring much less substance with respect to NMR[35,36,41,42].

Due to the advantages and the continuous improvements of NMR spectroscopy, this technique has been employed more and more in recent years for metabolomics purposes[43]. In this thesis, the mainly used technique for the performed metabolic analyses was NMR. Associated theory and some of the most relevant features will be briefly explained in the following sections.

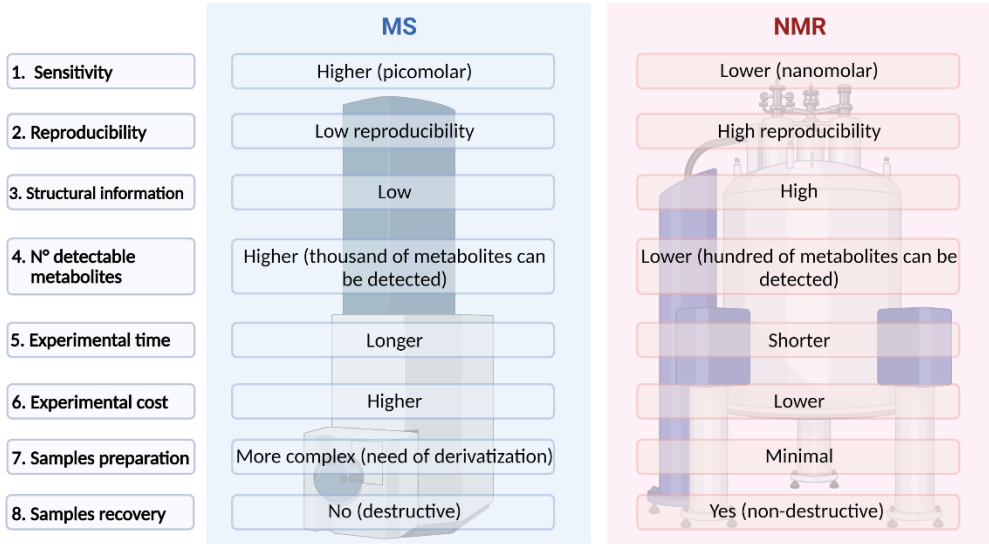

Figure 1.2: Comparison between MS and NMR, when applied to metabolomics. The figure summarizes the most relevant differences between the two techniques highlighting their strengths and weaknesses. (Figure created with BioRender.com).

1.3 NMR-based metabolomics

The use of NMR in the metabolic field dates back to the employment of this technique for phosphoromics (i. e. biological phosphorous determination) applications, including the monitoring of compounds containing this element, characteristic of some class of molecules in cells and tissues[39,44–46]. In addition to the abovementioned application, NMR is also particularly suitable for the study of metabolism at the cellular level thanks to the use of labelled isotopes, not only using phosphorous (^{31}P), but also, with proton (^{1}H), carbon (^{13}C) and nitrogen (^{15}N), which allows a better understanding of the fate of small molecules and of the involved pathways[47–49].

As mentioned, NMR-metabolomics enables the analysis of small molecules existing within different human biofluids such as blood, urine, saliva, faeces, cerebral fluid and many others, particularly useful in clinical metabolomics due the non-invasiveness of this approach[50].

Not only small molecules can be identified but also macromolecules like lipids with the so-called "lipodomics", which can offer several advantages in the investigation of certain diseases[51]. Moreover, it also offers others application like the examination of food components ("foodomics") for quality test analysis, to verify authenticity, safety and aliment origin[52–54]. In addition, NMR-metabolomics can be applied to the study and characterisation of plants extracts for the identification of natural products[55,56]. Moreover, NMR in the analysis of complex biomolecular mixtures, offers the advantage of being a quantitative technique which allows to determine the absolute concentration of the identified metabolites[57,58].

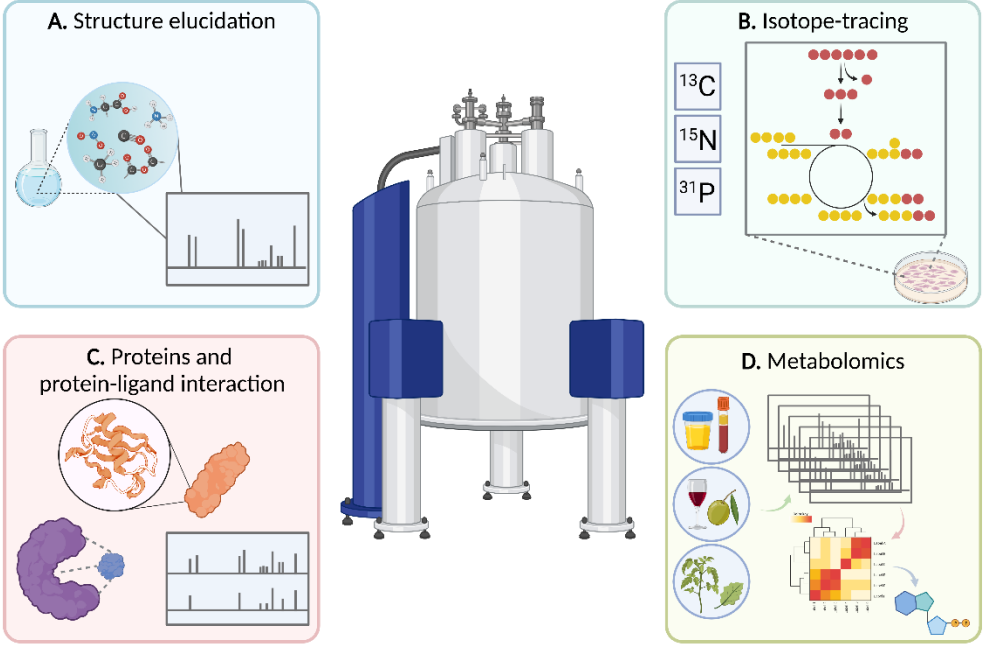

Figure 1.3: NMR spectroscopy application in the study of metabolism. NMR can be used as a useful tool for **A**) Structure elucidation; **B**) Isotope-tracing; **C**) Protein and protein ligand interaction study and finally, **D**) For metabolomics purposes. (Figure created with BioRender.com).

Figure 1.3 highlights some of the possible applications of NMR spectroscopy for a better understanding of the different aspect that characterize the metabolism in complex biological systems. As mentioned, this goes from compounds identification and related structure elucidation, to interaction study of certain metabolites with macromolecules for regulatory purposes and many other aspects[39].

Proton NMR (^{1}H NMR), and especially mono-dimensional experiments (1D NMR), is widely used for metabolomic investigations due to the speed of analysis, the reproducibility of the obtained results that make this technique suitable for inter-laboratory comparisons, thanks to the application of strict standard operating procedures (SOPs), and due to several other advantages that will be explained in more detail in the following sections and chapters[59]. Two-dimensional experiments (2D NMR) can also offer important information for complex matrix investigation and especially for metabolites identification, but they are not frequently used for the subsequent metabolomics analysis. Moreover, depending on the acquired experiment, due to the longer measurement time required, they can generally be acquired only for a subset of samples[60]. All these aspects will be discussed in more detail further on.

1.3.1 Nuclear magnetic resonance principles

NMR is an analytical technique whose principles are based on the presence of intrinsic magnetic properties in certain atomic nuclei. All nuclei, except the ones that have an even number of both protons and neutrons, can give rise to an NMR signal. Nuclei that possess a spin of magnetic quantum number *I* different from zero, like the nuclei with $I = ½$ (like ^{1}H, ^{15}N, ^{13}C, ^{19}F, ^{31}P), are sensitive to the external magnetic fields (B_0) and originate an NMR signal. Indeed, in the presence of B_0 these active NMR nuclei present a precessional motion at a characteristic frequency known as the Larmor Frequency. This frequency arises from the different spin orientation ($I = \pm ½$) associated to two different energy levels (**Figure 1.4**). One level corresponds to the lowest one in term of energy, aligned with the applied magnetic field ($I = +½$), and the other is the highest one ($I = -½$), since the spins are antiparallel to B_0[61]. The energy difference between these two states is quite small and Boltzmann's distribution can explain the difference on the number of nuclei associated to each of these spin states. The transition between the two energy levels is possible thanks to the stimuli of a radiofrequency electromagnetic radiation. After the excitation by the application of a radiofrequency, the return to the basal state is responsible of an energy emission that can be detected as free induction decay (FID) and elaborated with the Fourier transformation (FT) to generate an NMR spectrum.

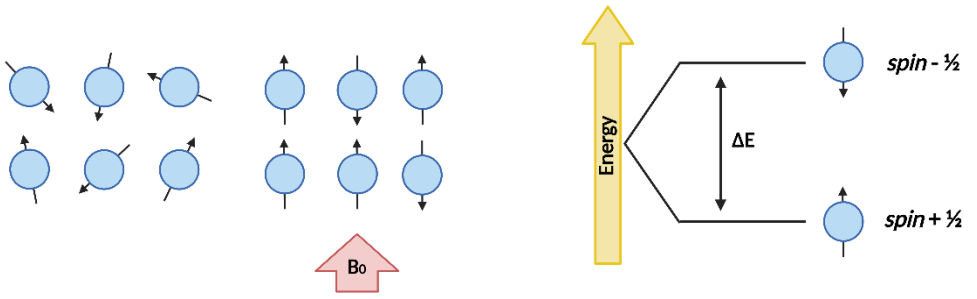

Figure 1.4: Nuclei in the absence of an applied external magnetic field, and in the presence of it (B_0). The nuclei with spin + ½ are aligned with the magnetic field and present lower energy, while the spin - ½ are not aligned and have lower energy. ΔE represents the energy difference between the two spin states. (Figure created with BioRender.com).

The NMR signal produced by nuclei relaxation to equilibrium is influenced by different factors. First of all, it is important to point out that not all of them are surrounded by the same electronic environment[62]. Electrons that surround the nuclei of an atom are able to shield it from the applied external magnetic field, generating a little magnetic field (Bi) usually opposed to B_0. The shielding strength depends on the position of the considered proton in the molecule and especially on the electronegativity of the surrounding atoms. The shift, in terms of frequency, caused by the shielding of the electrons with respect to the Larmor Frequency of the specific nucleus, is the so-called "chemical shift" (CS).

Secondly, the position of a proton in a molecule is important not only for the chemical shift but also for the influence that the surrounding ones can have on it. The presence of two or more adjacent protons in a molecule produces a spin-spin coupling (J-coupling) responsible for the "splitting" of the NMR signal into multiplets that obey the Tartaglia ruling.

These two factors, the CS and the J-coupling, are important concepts that must be taken into account in the analysis of the spectra especially in the case of complex matrices like biofluids, which is typical in NMR-metabolomics. Due to these factors the resulting spectra are complex, and the presence of hundreds of metabolites overlap the many different independent molecular profiles in the NMR spectrum[63].

1.3.2 ^{1}H-NMR spectroscopy for metabolomics purposes

^{1}H-NMR spectroscopy is largely used for metabolomics studies. The great use of this technique is related to pervasive distribution of ^{1}H atoms in the biomolecules, sensitive to the NMR phenomenon, which ultimately allows characterizing a plethora of metabolites[64].

Both 1D and 2D NMR approaches can be used in the field of NMR metabolomics, as previously mentioned, but 1D ^{1}H-NMR spectroscopy is more often used due to the short acquisition time needed to obtain a spectrum from each sample and for the huge amount of information that it contains. In fact, from a 1D ^{1}H spectrum is it possible to identify and quantify hundreds of metabolites[36]. This is particularly beneficial for large scale studies that requires the measurement of a huge number of samples in a fast, easy, and cost-efficient way[65].

Different pulse program can be used to acquire mono dimensional spectra, but one of the most used in the metabolomics field is the 1D Nuclear Overhauser Effect Spectroscopy (^{1}H NOESY) experiment with water suppression that allows molecules identification and quantification in a reproducible way[66]. Another experiment, that is largely used for metabolomics purposes, is the Carr-Purcell-Meiboom-Gill (CPMG) pulse program that is especially useful to eliminate the signal of macromolecules like proteins and lipids to identify the signals from small metabolites[67].

One of the limitations of mono-dimensional NMR spectroscopy is signal overlapping due to the huge number of metabolites present in each biofluid that contribute to the spectrum and to the relatively small difference in chemical shift between them. To resolve this issue, sometimes, it is necessary to relay to signal deconvolution for quantification purposes or to acquire spectra at higher magnetic fields[36].

Moreover, due to this weak point of 1D measurements, 2D NMR methods have also been used to overcome the metabolites identifications problems. In particular, ^{1}H J-resolved (JRES) spectroscopy, spreading the overlapping signals in a second dimension, could represent a solution. This experiment is also characterized by a relatively short acquisition time, compared with other 2D experiments like COSY

(COrrelation SpectroscopY), TOCSY (TOtal Correlation Spectroscopy) or HSQC (Heteronuclear Single Quantum Coherence), which is a key point in high-throughput studies[68]. Numerous approaches to the use of 2D spectroscopy have been recently evaluated to improve metabolites quantification in biological samples for metabolomics purposes. Despite this, NMR research is still constrained in its application to metabolomics investigations. This is probably related to the persistent worries (sometimes unjustified) about data size, to the longer acquisition time required for 2D NMR and to quantification issues associated with the multi-pulse nature of 2D NMR experiments, which results in a peak-dependent analytical response of the nuclei[36,69,70].

Anyway, in order to take advantage of the benefits offered by the use of ^{1}H spectroscopy for metabolomics analysis, both the used instruments in the conduced studies and the analytical and pre-analytical procedures for samples preparation, must be subject of strict controls and protocols, as explained below.

1.3.3 Urine metabolomics

Urine is considered an excellent biofluid for metabolomics studies. This is related to the huge amount of small molecules contained in it like organic acids, sugars, amino acids and some lipids. Different studies have demonstrated that more than 4000 metabolites are present in urine but not all of them can be identified by NMR[71].This allows to study the metabolic changes of an individual often related, for example, to diseases, lifestyle habits, diet or drugs consumption. Moreover, the ease and quantity with which urine can be obtained made it even more suitable for metabolomics purposes.

The huge variability of metabolites concentration that characterise urine is unfortunately also considered a disadvantage in the study of this biofluid. Many confounding factors can have an influence in the metabolite's identification and quantification for metabolomics purposes. For this reason, it is extremely important to follow rigorous SOPs to obtain urine in a specific moment, avoiding diary variation related to different factors, from each subject for the entire duration of the study. Early

morning urine (ideally 12 hours fasting), as previously mentioned, is usually preferred because it is less influenced by food consumption or other factors[72].

Urine metabolic fingerprinting is easily obtained by NMR. This can be used in order to identify changes between different health conditions and for biomarkers identification comparing healthy subjects with an investigated population. Lots of diseases are not detected until advanced stages and generally present invasive medical test for diagnosis. Urine metabolomics offers the possibility to identify early-stage biomarkers for a prompt intervention trying to avoid further complications. Previous studies have been able to identify metabolic changes and biomarkers for different diseases like prostate cancer[73], inborn error of metabolism[74] and many others[75,76].

1.3.4 Serum and plasma metabolomics

Blood contains lot of compounds whose change can be associated to different metabolic states. Serum and plasma analysis in metabolomics studies present great advantages due to the higher stability of these biofluids with respect to urine in terms of metabolites variability in response, for example, to diet or environmental factors. The inter-individual variation are present with minor extent, nevertheless, is still preferable to obtain samples under fasting conditions for a better understanding of the metabolic changes that affect each subject avoiding possible confounding factors. Once again, the use of strict standard operating procedures is of the outmost importance especially because, in blood extraction, as previously said, the coagulation factors (like EDTA or heparin) that may be present in the extraction tubes can produce huge noise signals in the NMR spectra that can overlap with important metabolites peaks[77].

Different studies have been done to understand if serum or plasma are better for metabolomics investigation but both fluids are considered appropriate for this kind of studies[72]. Serum presents a higher concentration of metabolites with respect to plasma, but as regard lipoprotein quantification no differences have been detected[78]. Lipoprotein analysis is an important advantage offered by blood investigation using NMR, that allows the identification and quantification of different lipoproteins subclasses that cannot be analysed with normal biochemistry.

Human lipoproteins contained in blood consists in high-density lipoprotein (HDL), low-density lipoproteins (LDL), intermediate-density lipoprotein (IDL) and very-low density lipoproteins (VLDL). Changes in the levels of these macromolecules are involved in different pathologies like metabolic syndrome, cardiovascular diseases and many others[76].

1.3.5 Standard Operation Procedures (SOPs)

NMR metabolomics studies rely on strict SOPs that must be respected to ensure the reliability of the obtained results[79]. SOPs include specific indications for each step within the metabolomics workflow, from the recollection and storage, sample preparation, instrument calibration, NMR acquisition and, finally, data processing[80]. This is extremely important also for comparisons in collaboration projects with different laboratories and to guarantee reproducibility and reliability of the obtained results[81,82].

1.3.5.1 Sample recollection and storage

Sample recollection is the first step in the metabolomics studies workflow (**Figure 1.5**). Different studies have been done to determine which is the best way to maintain urine or serum samples in order to study the small molecules contained in it avoiding degradation and guaranteeing the reliability of the obtained results[79,83,84]. Conditions include fasting for blood extraction, to give the first morning urine and homogenous material for samples recollection and storage during the whole process to avoid the presence of contaminants[85,86]. Biobanks are in charge of preserving samples according to the associated SOPs up to the measurement time[87,88]. The organization of huge cohort of samples into the biobanks is essential: specific IDs are usually assigned to extraction tubes of blood or urine that has to be linked to the corresponding patient's ID and its associated medical history[89].

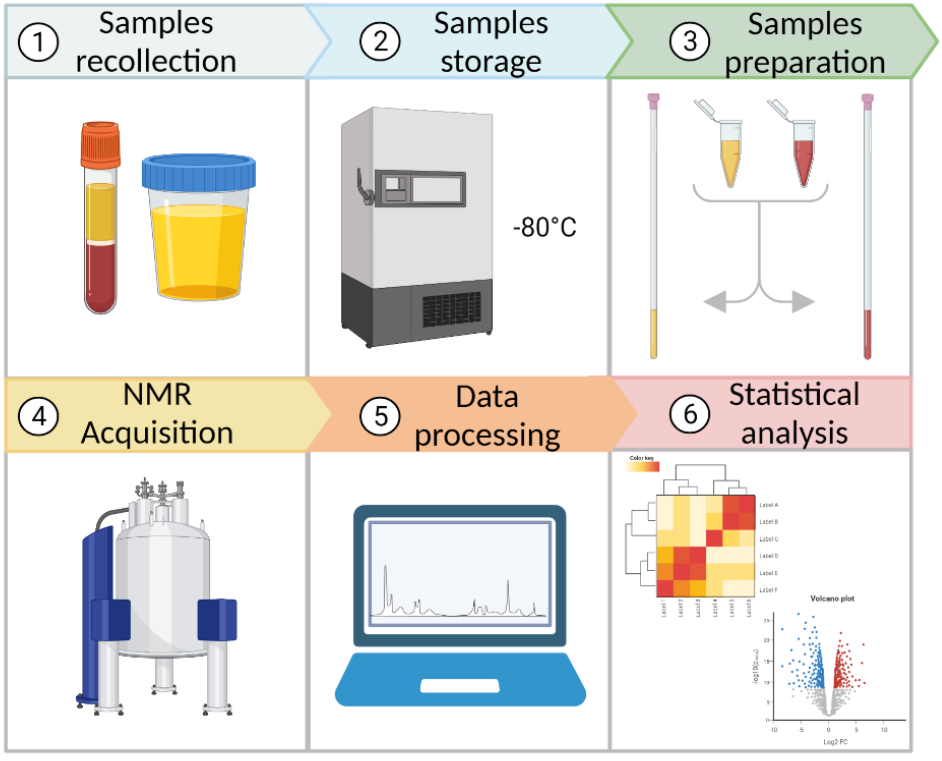

Figure 1.5: Scheme of the main steps that characterize the metabolomic workflow. (Figure created with BioRender.com).

1.3.5.2 SOPs for samples preparation

Blood requires control of contaminants to avoid the presence of strong NMR signals coming, from example, from EDTA, heparin or sodium citrate, which are commonly contained in the extraction tubes and will result in spureous signals in the NMR spectrum[58]. Specific centrifugation protocols are followed to obtain plasma or serum[79]. Moreover, biofluids contain lot of metabolites and lipids that are sensitive to temperature changes and samples are typically maintained at 4°C while managing them until measurement in the NMR spectrometer[90,91]. Actually, sample manipulation is maintained to a minimum to ensure reproducibility: serum and plasma samples are only mixed in a 1:1 ratio with a specific buffer.

Regarding urine, the early morning one is usually used for the majority of metabolomics studies. After recollection, urine is stored at 4 °C to avoid degradation and the aliquots frozen at -80°C until measurement. Urine generally undergoes a

centrifugation process to remove possible contaminants and cellular residues that could alter metabolites quantification[71]. Urine is usually more stable than plasma or serum, therefore this gives the possibility of repeating NMR experiments, if needed, as long as it is kept at 4°C for a limited time period (generally not more than 48 hours).

1.3.5.3 Magnet calibration

NMR spectrometer performance is checked daily to ensure reproducibility. Checks include temperature stability (i.e. by measuring the signal splitting in a methanol sample), quantification (by means of an internal electronic signal, QuantRef) and magnet homogeneity (by shimming a sucrose reference sample)[59,77].

For temperature calibration a deuterated methanol sample is analysed to verify temperature in the range of 175 and 310 K. Hydrogen bonding is affected by temperature that causes a different shift of the protons according to it. The measured value is the OH/CH_3 chemical shift difference, that gives information about the real temperature inside the magnet with the measured methanol sample[61].

Shimming performance of the spectrometers is key in the metabolomic spectra due to the need to compare different spectra in large scale studies but also because of the great amount of metabolites and related peaks that comes out after the measurement of a biological sample[92]. Resolution, sensitivity and water presaturation performances are evaluated by means of different parameters in a sealed glass 2mM sucrose sample in 90% H_2O, 10% D_2O, 2mM NaN_3, and 0.5 mM DSS (2,2-Dimethyl-2-silapentane-5-sulfonate). DSS and water residual signals are necessary to evaluate the suppression performance while sucrose is necessary to evaluate the resolution and sensitivity.

Quantification is calibrated by means of a specific sample called QuantRef (Quantification Reference solution) containing a known amount of specific metabolites. Concentration of the compounds is estimated according to the resulting peaks integrals and using a synthetic signal called ERETIC (Electronic Reference To access In-vivo Concentrations) as calibration compound to avoid possible overlapping with some metabolite peaks[43].

A more in-depth description of these calibration procedures is provided in Chapter 3.

1.3.5.4 NMR measurement

According to the analytical *in vitro* diagnostic research (IVDr) SOPs, after magnet calibration, samples are measured using specific validated NMR pulse sequences for metabolomics studies. These includes specific measurements for urine and serum samples[79]. As regards urine, two experiments are acquired for each sample: a mono dimensional ^{1}H NMR spectrum (*noesypr1d*) for metabolites quantification and a bidimensional one J-resolved ^{1}H spectrum (*jresgpprqf*). For serum samples three experiments are acquired: a one-dimensional NOESY, a JRES (*jresgpprqf*), as for urine, and then an additional ^{1}H NMR Carr-Purcell-Meiboom-Gill experiment (*cpmgpr*) that filters out all the signals arising from high-molecular weight particles, including the lipoproteins. Further information will be given in the materials and methods section (Chapter 3).

1.3.6 Data processing and analysis

Data organization obtained from the analysis of thousands of samples is a complex task but it is essential in order to process the spectra, proceed with an adequate statistical analysis and, finally, for the interpretation of the results. Statistical analysis plays a crucial role, giving the essential tools needed for the understanding of the metabolic changes observables in large population under studies.

Data must be organized in a univocal way so that patient's information can be linked to the corresponding NMR spectra to proceed with the following steps of the analysis. This can be done by assigning a specific code to each one of the patients and associating this code to the corresponding samples ID of urine and serum and naming the acquired NMR spectra in a way that can be linked to the corresponding donor. Under these circumstances, patient information can be related to the produced data so that this can be used for further steps of the analysis (**Figure 1.6**).

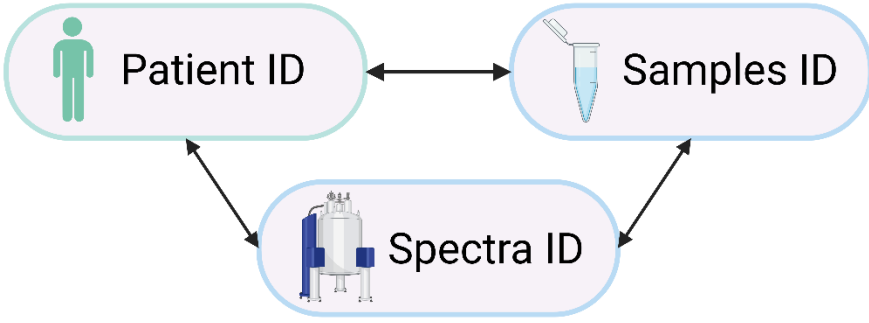

Figure 1.6: Schematic representation of the organization codes (ID) to relate the information coming from each of the steps that compose the samples analysis. (Figure created with BioRender.com).

After NMR data acquisition, spectral processing has to be performed. This includes all the corrections such as baseline optimization, phase correction and chemical shift referencing. Subsequently, spectra must be filtered, aligned, scaled, and, according to the selected analysis procedure, spectra can also be divided into spectral areas of equal size that are called bins. Next, the obtained data are usually scaled and normalized[93]. Additional information about all these procedures will be discussed further in Chapter 3.

Two main statistical methods are used to analyse data according to what is needed to be investigated: multivariate or univariate analysis. In the first case, different variables are analysed at the same time to find differences into, for example, two or more groups in order to identify the putative changes between these populations. Alternatively, univariate analysis takes into account just a variable independently. Another distinction can be made between unsupervised and supervised analysis. These two conditions differ in the way data are organised. An example of a commonly applied unsupervised method is the principal component analysis (PCA), which allows to identify, if present, differences into the groups under study. On the other hand, another example of supervised analysis is the employment of the partial least squares analysis (PLS) or PLS Discriminant Analysis (PLS-DA) and orthogonal PLS-DA (OPLS-DA). Further details of the applied statistical methods for the conduced analysis in this thesis are described in Chapter 3.

The obtained results from statistical analysis are usually interpreted within a biological framework. This contextualization often also includes the metabolites identification. To that aim, different tools can be used: the Chenomx software or specific databases like the Human Metabolome Database (HMDB) that contains information from the majority of small molecules present in the human body (in serum, urine, faeces, cerebral biofluid or saliva) or NMR spectra biobanks like the Biological Magnetic Resonance Bank (BMRB).

1.3.7 Large scale studies

The analysis of large cohorts that can include individuals from the general population or groups of patients with a specific pathology, is essential for the validation of new biomarkers or for a better understanding of the disease under study. The cohort size analysed in metabolomic studies is an important factor under consideration because on it largely depends the reliability of the study and the trustworthy of the obtained results.

To carry out studies that include lot of samples is essential to count on high-throughput techniques enabling a fast and cost efficient analysis. Additionally, all the previously explained SOPs must be strictly applied because the accuracy of the results depends both on NMR performance and samples extraction, organization and storage.

Two main studies have been conducted and included into this thesis project as a case study to show how large scale high-throughput studies can be conducted to obtain models or biomarkers for the discrimination and identification of a disease at an early stage and/or to characterize the metabolic changes that can affect a population under study. Specifically, NMR based metabolomic methodologies are here applied to the investigation of metabolic syndrome and to the viral infection caused by SARS-CoV-2. The characteristics and the mechanisms regulating these two pathologies are discussed in detail of the following sections.

1.4 The Metabolic Syndrome

Metabolic syndrome (MetS) is a multifactorial disorder whose insurgence is related to a variety of risk factors. MetS contributes to a higher probability of development of secondary problems like cardiovascular dysfunction, heart attack, stroke, type 2 diabetes and increased risk of mortality due to all these related pathologies[94].

1.4.1 The definition of metabolic syndrome

Many health organizations have attempted to give a definition of this disorder including the World Health Organization (WHO), the European Group for the Study of Insulin Resistance (EGIR), the National Cholesterol Education Program's Adult Treatment Panel III (NCEP ATP III), the International Diabetes Federation (IDF), and the American Association of Clinical Endocrinology (AACE) (**Table 1.1**)[95–100]. All of them provide slightly different definitions, but they all agree with the presence of four main contributing risk factors: altered glucose metabolism, obesity, dyslipidemia and hypertension (**Figure 1.7**)[101]. According to the WHO, the first one trying to give a definition of this disorder in 1998, MetS should be defined mainly as the presence of problems with the metabolism of the glucose, like impaired glucose tolerance (IGT) or impaired fasting glucose (IFG) as mandatory factors plus other two risk factors including obesity, dyslipidemia, hypertension or microalbuminuria[102]. Later, EGIR agreed with WHO's definition but included insulin resistance as a compulsory requirement for the diagnosis plus two of the previously mentioned additional factors apart from microalbuminuria[103]. NCEP ATP III decided to eliminate the compulsory requirement of a factor for diagnosis and to consider an individual as affected by metabolic syndrome with the presence of three or more factors out of the five usually taken into account (in this case dyslipidemia is not considered as the sum of higher triglycerides and low HDL cholesterol, but they are considered separately)[104]. Finally, IDF used the same criteria for the definition of MetS as the previous organisations like WHO and AACE, but with the difference of not considering problems with glucose metabolism as a compulsory factor but obesity[105].

One of the apparent problems when perusing the described definitions is the lack of a clear notion on what is the best way to define MetS. In addition, these organisations take into account different cut-off values for each of the risk factors that are considered for diagnosis. For example, WHO considers IFG , consisting of a higher fasting glucose value of 100 mg/dL, as problem for the metabolism of the glucose or IGT that is determined thanks to a specific test that consist in the monitoring of glucose levels for 120 minutes after ingesting it (if the levels are higher than 140 mg/dL IGT is diagnosed). On the other hand, EGIR also includes insulin resistance in the mandatory factors for the diagnosis of MetS[102–105].

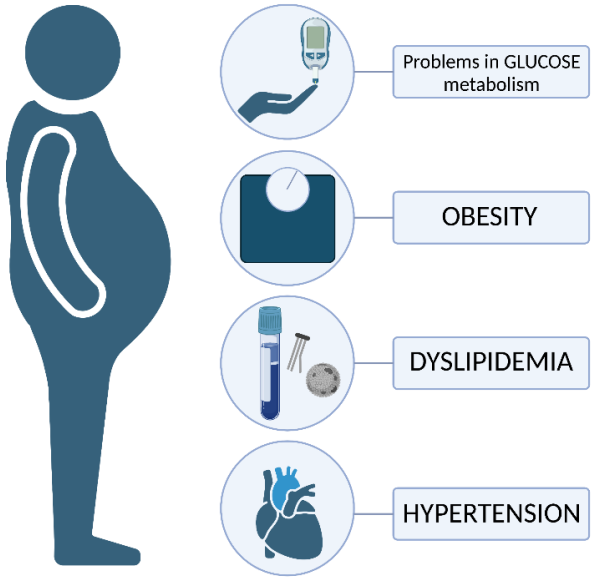

Figure 1.7: Risk factors contributing to the development of MetS. (Figure created with BioRender.com).

Even more complicated is the diagnosis of obesity that, in some cases, refers to the body mass index (BMI) calculated from weight and high of each individual for the diagnosis, while in other the waist circumference or the waist to hip ratio are prefered[106]. Moreover, different values are taken into account between males and females, or ethnicity is considered also to play a crucial role according to some organization like IDF. In addition to it, the dyslipidemia is another factor whose cut-off values are different in each one of the organization's definitions, making difficult to understand which values should be the best one and why[107].

Finally, some organization like WHO or AACE decided to take into consideration additional factors with respect to the usual ones for the diagnosis: the WHO considered microalbuminuria as an additional risk factor in the insurgence of MetS, while the AACE decided to consider as risky the presence of other syndromes like the polycystic ovary, a family history of type 2 diabetes mellitus (T2DM), a sedentary lifestyle or increased age[108].

Table 1.1: Definition criteria for the diagnosis of MetS according to different organizations.

	WHO	EGIR	AACE	NCEP:ATPIII	IDF	HARMONIZED
GLUCOSE METABOLISM (FG mg/dL)	**IGT, IFG, T2DM or lowered insulin sensitivity**†	**IR**‡ FG ≥ 110	**IGT or IFG (but not diabetes)** †	FG ≥ 100	FG ≥ 100 or T2DM	FG ≥ 100 or treatment
OBESITY (BMI kg/m^2, WC cm)	WHR(m) > 0.90 WHR(f) > 0.85 or BMI > 30	WC(m) ≥ 94 WC(f) ≥ 80	BMI > 25	WC(m) ≥ 102 WC(f) ≥ 88	**Elevated WC, ethnicity, and gender specific**†	Elevated WC, population, and country specific
DYSLIPIDEMIA (TG, HDL-C mg/dL)	TG > 150 or HDL-C(m) < 35 HDL-C(f) < 39	TG > 177 or HDL-C < 39	TG ≥ 150 or HDL-C(m) < 40, HDL-C(f) < 50	TG ≥ 150 or HDL-C(m) < 40, HDL-C(f) < 50	TG ≥ 150 or treatment or HDL-C(m) < 40, HDL-C(f) < 50 or treatment	TG ≥ 150 or treatment or HDL-C(m) < 40, HDL-C(f) < 50 or treatment
HYPERTENSION (BP mmHg)	≥ 140/90	≥ 140/90	≥ 130/85	≥ 130/85	≥ 130/85 or treatment	≥ 130/85 or treatment
OTHER FACTORS	Microalbominuria > 30 mg/g	Not relevant	Other risk factors§	Not relevant	Not relevant	Not relevant

Organizations: **WHO**, World Health Organization; **EGIR**, European Group for the Study of Insulin; **AACE**, American Association of Clinical Endocrinology; **NCEP:ATPIII**, National Cholesterol Education Program-Third Adult Treatment Panel; **IDF**, International Diabetes Federation.

†Bold highlighted factors are compulsory for the given definition. Obtained from refs[98,109,110].

‡**IR**, Insulin resistance, defined as hyperinsulinemia: top 25% of fasting insulin values among the nondiabetics.

§Family history of T2DM, sedentary lifestyle, advanced age, ethnic groups susceptible to T2DM, polycystic ovary syndrome.

Other abbreviations: **IFG**, impaired fasting glucose; **IGT**, impaired glucose tolerance; **FG**, fasting plasma glucose; **T2DM**, type 2 diabetes; **WC**, waist circumference; **WHR**, waist-hip ratio; **BMI**, body mass index; **TG**, triglycerides; **HDL-C**, HDL cholesterol; **BP**, blood pressure; **m**, male; **f**, female.

In 2009 different organizations tried to give a harmonized definition of the metabolic syndrome, eliminating the compulsory requirement on the presence of a specific factor for diagnosis and considering only the presence of three conditions out of five. That said, many studies still don't take into proper consideration this solution due to its arbitrary character[98]. For this reason, it is timely to understand the associated metabolic changes produced by each of the risk factors involved in the development of MetS, in order to be able to give a better definition based on real observable differences and not only in pre decided cut-off values.

1.4.2 The spreading of MetS

The study of this syndrome is raising interest due to the exponential increase in the number of individuals suffering from it, starting even in the people of young age[111]. Its prevalence is between the 10 and 40% according to the population under study[112]. This is particularly linked to an increased incidence of obese people that is one of the risk factors related with the insurgence of MetS and with the elevated spreading of T2DM. Regarding obesity, it has been observed a growing number of children and adolescent affected by it, which is reflected in the adult population, leading to a higher number of people presenting this syndrome[96]. In relation to T2DM, it has been evidenced that, for example, due to the speeding of this condition, up to one third of the population in the United States can be considered as affected by MetS[113]. Indeed, many causes contribute to the insurgence of this syndrome, like lifestyle habits, diet, physical activity, genetic susceptibility or an altered circadian rhythm[95]. Unfortunately, it has been noted that MetS increases in lower socio-economic populations, probably due to a misinformation about the importance of good lifestyle habits and healthy diet[114].

1.4.3 The risk factors associated with MetS

As mentioned, the main risk factors that lead to the insurgence of this disease, as previously mentioned, are: problems with the metabolism of the glucose, obesity, an altered concentration of cholesterol and triglycerides and a higher blood pressure (**Figure 1.7**).

1.4.3.1 Glucose metabolism

The higher concentration of fasting glucose (≥ 100 mg/dL) is usually associated with problems in its metabolism. This can be related to insulin resistance or other recognised problems as impaired glucose tolerance or impaired fasting glucose and T2DM. IGT and IFG share common characteristics, being both considered as pre-diabetic states, which start to increase the risk for cardiovascular diseases (CVD) events. The difference between these two conditions is based on the fasting glucose concentration, which is normal in IGT (< 110 mg/dL) and abnormal in IFG, and the postprandial glucose concentration (two hours), which, on the contrary, is abnormal in IGT (from 140 to 199 mg/dL) and normal in IFG[115,116]. Insulin is considered responsible for systemic glucose disposal by controlling whether it is used in muscle, liver or adipose tissue[102]. In case an individual has problem of insulin sensitivity this led to a higher concentration of glucose in the bloodstream, avoiding being conserved into glycogen or any other storage molecules.

1.4.3.2 Obesity

Obesity represents one of the major problems in the insurgence of MetS, in fact some of the definitions include it as one of the mandatory factors for the diagnosis[110]. Subjects characterized by upper body obesity are more prone to develop MetS with respect to those presenting lower body obesity[117,118]. The first one, more frequent in men, is characterized by an excess of subcutaneous and especially visceral fat, while the second one, represented by gluteofemoral fat, is most commonly presented by women but these can be presented by both genders. Abdominal visceral fat is considered as more dangerous because different studies have associated it with a higher risk of incidence for T2D, dyslipidemia and insulin resistance which could contribute to the insurgence of MetS[119]. Moreover, visceral obesity is also considered dangerous due to its association with unfavourable clinical outcomes[120].

Different cut-off values and methods have been suggested for the diagnosis of obesity, sometimes trying to differentiate between sex or ethnic factors. Moreover, it is not clear whether is better to consider the BMI (obtained by dividing the weight in kilograms of the subject by the square of the height in metres), the waist to hip ratio or the waist

circumference to assess if an individual is considered as obese[121]. All of them are obesity descriptors, but, for example, BMI is the most widely used in several studies even if it does not take into account that the abdominal obesity has been more often associated with MetS. On the other hand, this factor is highlighted when considering waist circumference as method.

1.4.3.3 Dyslipidemia

Dyslipidemia is usually associated with an increased blood concentration of triglycerides a lower amount of HDL-cholesterol and higher levels of small density lipoprotein[122]. The latter are considered as particularly dangerous as responsible for the transport of lipids to peripheral cells and because of their low dimension, can reach areas that other lipoproteins cannot[123]. Moreover, they are accumulated into the blood vessels causing the generation of atherosclerotic plaques which significantly increases the possibility of cardiovascular events[124]. In fact, dyslipidaemia is considered to be one of the factors most involved in the onset of secondary problems, such as atherosclerotic cardiovascular disease (ASCVD), related with an increased mortality.

Dyslipidemia is also often associated with insulin resistance and can be a predictor of T2DM. In both these conditions an increased level of VLDL, produced by liver, has been observed. This leads to and increment of LDL which are produced from the metabolism of VLDL: free fatty acids are released to IDL formation and further to LDL[125]. Despite this, alternation of VLDL is not obligatorily related to hyperinsulinemia but it can also be present in individuals that are affected by problems with the metabolism of lipoproteins rich in triglycerides concentration[121].

1.4.3.4 Hypertension

Hypertension is usually diagnosed when the systolic pressure is greater or equal to 130-140 mmHg and the diastolic less than or equal to 85-90 mmHg, according on the considered definition. This risk factor, as the previously described one, play a crucial role in the development of MetS. In fact, it has been observed that up to one third of the individuals that present high blood pressure is affected by MetS[126]. This condition is often associated with the previously described ones, like obesity or insulin resistance

and yet it is not clear whether these may also be some of the causes of increased blood pressure, along with many others. The importance of the diagnosis of this health condition is related to its implication in the development of secondary problems especially at the cardiac level[127].

1.4.4 Associated comorbidities

The importance for the diagnosis of MetS is also related to the comorbidities that are associated with this disorder. Indeed, MetS is responsible for the increased risk of development of CVD and may also herald the onset of diabetes because the correlated risk factors. Despite this, it is difficult to understand up to which extent MetS is also associated to an increase in CVD, the latter also affected by other factors like the alcohol consumption, smoking, the age or more in general, unhealthy lifestyle habits.

MetS is related to an increased risk of mortality. It has been demonstrated that the risk of the manifestation of a cardiac event is much higher in patients affected by MetS rather than in patients that present just one of the risk factors involved in the manifestation of the syndrome[128].

Grouping individuals with similar health conditions is helpful at the clinical level, in order to encourage patients to take measures to improve their lifestyle habits and possibly treating them for their diagnosed disorders avoiding further complications.

MetS has been associated to other diseases, like cancer, chronic kidney diseases or even psychiatric disorders, of particular interest is the Non-Alcoholic Fatty Liver Disease (NAFLD)[129–133]. In fact, this disorder and MetS shares the common risk factors of obesity, insulin resistance, dyslipidemia and hyperglycemia. It is still an open question whether or not NAFLD is the hepatic manifestation of the metabolic syndrome especially due to the shared pathologic conditions. Further studies need to be conducted to resolve this doubt and for a better understanding of this chronic liver disease[134].

1.5 Sars-CoV-2 Infection

In December 2019 some novel cases of pneumonia were detected in Wuhan, China. The isolated virus responsible for it was the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) so called for its genetic similarity to the previous SARS-CoV[135]. The origin of the virus has been debated but it is believed that it originated in the Huanan Seafood Wholesale Market[136]. The World Health Organization in March 2020 declared a state of pandemic due to the rapid spreading of this virus all over the world[137,138]. In fact, SARS-CoV-2 is very easily transmitted from person to person through droplets inhalation together with aerosol emission and consequent contact of the virus with nose, mouth or eyes mucous[139,140]. Currently, the number of people worldwide who have been infected since the beginning of the pandemic, is approximately of 623 million and the number of victims amounts to 6.5 million[141].

1.5.1 COVID-19 associated epidemiology

COVID-19, the disease caused by the infection of SARS-CoV-2 virus, is characterized especially by problems in the respiratory track. The presented symptoms include cough, fever, shortness of breath, cold, headache, muscle aches, general fatigue but also gastrointestinal symptoms as vomit, diarrhoea or many others (**Figure 1.8**)[140]. As the study of this disease and its variants has progressed, many symptoms and health problems related to this infection have been detected[142].

The complexity of this virus has originated many scientific studies, in order to understand its mechanism of infection, the problems it causes and the best clinical approach to treat patients[143]. The main factor for the infection of SARS-CoV-2 consist in the binding of the Spike protein of the virus with the angiotensin-converting-enzyme 2 (ACE2) receptor on the host cells surface. ACE2 is principally expressed into the nasal epithelium cells and in the bronchial epithelia[144]. The SARS-CoV-2 principal target is the respiratory track but it is also present in other organs explaining the onset of secondary issues like renal insufficiency, heart or nervous system problems. Hence, COVID-19 has to be considered as a systemic infection[140,145].

Not all the subjects develop the same kind of symptoms, presenting a mild or severe COVID-19[146]. On the other hand, some patients are completely asymptomatic. This

evidence that the factors related to the onset of specific symptoms depend also on personal characteristics that vary from one individual to another, making some individuals more susceptible to the development of a specific symptomatology, even if affected from the same virus. Moreover, different pre-existing risk factors, like specific health conditions, are considered as involved in the regulation of disease progression [147].

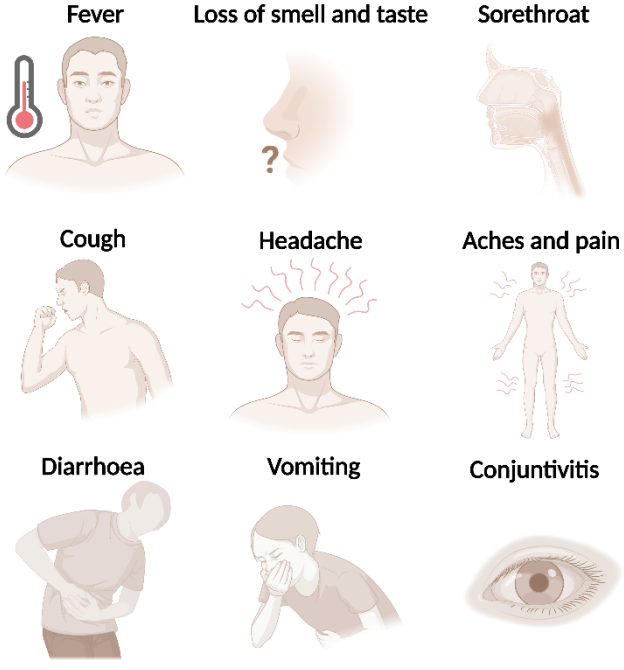

Figure 1.8: Main symptoms that characterize COVID-19 patients. (Figure created with BioRender.com).

One of the causes linked to the severe aggravation of patients is what has been called the "cytokine storm" involving the excessive immune response that can be triggered by this disease[148,149]. When a patient is infected with the SARS-CoV-2 virus, the individual's first response is the one from the innate immune system. Unfortunately, an over-response can cause serious damage to several organs, occurring in some COVID-19 patients. Indeed, the main cytokines involved are the interleukin 2,7 and 8 (IL-2, IL-7, IL-8), the granulocyte macrophage stimulating factor, some chemokines like the

interferon-γ inducible protein-10 (IP-10) and monocyte chemoattractants like CCL2, CCL5 and CCL3 and the tumor necrosis factor-α (TNF-α)[150,151].

The related acute respiratory distress syndrome (ARDS) consisting in an increased release of immune system cells had been associated with organ failure. Therefore, the combination of the "cytokine storm" and ARDS are considered the main causes of death among patients[152].

1.5.2 A precision medicine approach in infectious diseases

As regard infectious diseases, it has been observed that individual's reaction to infection is not always the same but depends on both the pathogen and the specific characteristic of each subject[153]. Metabolism seems to play a crucial role in disease development leading to milder or more severe symptoms depending on the person[154]. For this reason, a metabolomics approach can be used to identify metabolites that could be considered as biomarkers, especially if presented in the early stages of the disease[155].

These general considerations can also be applied in the context of COVID-19. In symptomatic COVID-19 patients, it is possible to identify pre-existing factors in each individual that can correlate with severity. As regard COVID-19, type 2 diabetes (T2DM) for example, has been associated with more severe SARS-CoV-2 infections representing a risk factor for this disease together with others like obesity, hypertension or metabolic syndrome as previously mentioned[154].

1.5.3 Use of NMR-metabolomics in the study of COVID-19

Due to the systemic character of this disease, a better comprehension of COVID-19 is necessary, and the use of NMR-metabolomics could be helpful for this purpose. Thanks to the NMR analysis of samples like plasma or serum from infected individuals it could be possible to understand the causes that leads patients to be affected by COVID-19, the metabolic changes that characterize individuals with this pathology, including the possible differential expression of specific molecules and macromolecules and the identification of the principal causes that leads to severe outcomes. Moreover, NMR metabolomics proved to be useful also for the discrimination of the patients according to the disease severity, the investigation of the

possible effect of drugs treatment or even vaccination and, finally, the characterization of the long-term sequelae that can affect some SARS-CoV-2 infected individuals[156–159]. In the conduced analysis we focused our attention to the description of the characteristic metabolic changes during the acute phase of the disease, further described in Chapter 5[160].

In the context of the use of NMR spectroscopy for COVID-19 investigation, due to some of the previously explained limitation, principally related with signal overlap and low sensitivity, new methodologies have also been developed to overcome these issues[161–164].

Some specific signals in the NMR spectra have been previously associated with the description of inflammation states: the N-acetyl signal of glycoproteins (GlycA and GlycB) and the supramolecular phospholipid composite peak (SPC) of phospholipids. To investigate these aspects more thoroughly, NMR pulse sequences have been specifically designed which, allow a better isolation and examination of these specific NMR signals. Thanks to the use of the optimized technologies, it is possible to monitor the inflammation state of a patient and check their clinical course, thus also being able to have an idea of the patient's recovery time.

1.5.4 MetS and COVID-19

Due to the spread of COVID-19, the infectious disease caused by the SARS-CoV-2 virus that has recently affected the worldwide population at a pandemic level, several studies have been recently conducted trying to understand the possible relationship between this infection and the metabolic syndrome[165–167]. Indeed, subjects with diabetes, hypertension or obesity had more probability to present severe outcomes from SARS-CoV-2 infection[168–171]. Patients with COVID-19 that are affected by MetS, had increased probability to be admitted to the intensive care unit (ICU) due to severe complication and even higher risk of death[172].

Moreover, it has been observed that there is also a correlation between the probability of testing positive for SARS-CoV-2 and subjects with higher BMI. Obesity seems to

be a risk factor for COVID-19 disease, making individuals even more susceptible to be infected[173–175].

For this reason, a more in-depth study of the mechanisms that regulate these two diseases can help to improve the health of patients by avoiding negative outcomes. In the context of this thesis, due to the concomitance of this pathology and due to the availability of samples for investigating this new disease, the metabolic mechanisms related to the onset of COVID-19 and the characteristics of patients in the acute phase of this pathology have been studied using NMR-metabolomics.

Bibliography Chapter 1:

1. Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease. *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease* 1–128 (2011) doi:10.17226/13284.
2. König, I. R., Fuchs, O., Hansen, G., von Mutius, E. & Kopp, M. v. What is precision medicine? *European Respiratory Journal* **50**, 1–12 (2017).
3. Matchett, K. B., Lynam-Lennon, N., William Watson, R. & Brown, J. A. L. Advances in precision medicine: Tailoring individualized therapies. *Cancers (Basel)* **9**, (2017).
4. Engla, N. E. W. & Journal, N. D. A New Initiative on Precision Medicine. *N Engl J Med* 793–795 (2015).
5. McAlister, F. A., Laupacis, A. & Armstrong, P. W. Finding the right balance between precision medicine and personalized care. *Cmaj* **189**, E1065–E1068 (2017).
6. Mussap, M., Noto, A., Piras, C., Atzori, L. & Fanos, V. Slotting metabolomics into routine precision medicine. *Expert Rev Precis Med Drug Dev* **6**, 173–187 (2021).
7. Wishart, D. S. Emerging applications of metabolomics in drug discovery and precision medicine. *Nat Rev Drug Discov* **15**, 473–484 (2016).
8. Lockhart, D. J. & Winzeler, E. A. Genomics, gene expression and DNA arrays. *Nature 2000 405:6788* **405**, 827–836 (2000).
9. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature 2003 422:6928* **422**, 198–207 (2003).
10. Zhang, A., Sun, H., Wang, P., Han, Y. & Wang, X. Modern analytical techniques in metabolomics analysis. *Analyst* **137**, 293–300 (2011).
11. Dong, Z. C. & Chen, Y. Transcriptomics: Advances and approaches. *Science China Life Sciences 2013 56:10* **56**, 960–967 (2013).
12. Hegde, P. S., White, I. R. & Debouck, C. Interplay of transcriptomics and proteomics. *Curr Opin Biotechnol* **14**, 647–651 (2003).
13. Wishart, D. S. Emerging applications of metabolomics in drug discovery and precision medicine. *Nat Rev Drug Discov* **15**, 473–484 (2016).
14. Patti, G. J., Yanes, O. & Siuzdak, G. Metabolomics: the apogee of the omics trilogy. *Nat Rev Mol Cell Biol* **13**, 263–269 (2012).

15. Beger, R. D. *et al.* Metabolomics enables precision medicine: "A White Paper, Community Perspective". *Metabolomics* **12**, (2016).

16. Reo, N. v. NMR-based metabolomics. *Drug Chem Toxicol* **25**, 375–382 (2002).

17. Wild, C. P., Scalbert, A. & Herceg, Z. Measuring the exposome: A powerful basis for evaluating environmental exposures and cancer risk. *Environ Mol Mutagen* **54**, 480–499 (2013).

18. Rappaport, S. M., Barupal, D. K., Wishart, D., Vineis, P. & Scalbert, A. The Blood Exposome and Its Role in Discovering Causes of Disease. *Environ Health Perspect* **122**, 769 (2014).

19. Fiehn, O. Metabolomics-the link between genotypes and phenotypes. *Plant Mol Biol* **48**, 155–171 (2002).

20. Vinayavekhin, N., Homan, E. A. & Saghatelian, A. Exploring disease through metabolomics. *ACS Chem Biol* **5**, 91–103 (2010).

21. Houten, S. M. Metabolomics: Unraveling the chemical individuality of common human diseases. *Ann Med* **41**, 402–407 (2009).

22. Johnson, C. H., Ivanisevic, J. & Siuzdak, G. Metabolomics: Beyond biomarkers and towards mechanisms. *Nat Rev Mol Cell Biol* **17**, 451–459 (2016).

23. Beger, R. D. *et al.* Metabolomics enables precision medicine: "A White Paper, Community Perspective". *Metabolomics* **12**, 149 (2016).

24. Roberts, L. D., Souza, A. L., Gerszten, R. E. & Clish, C. B. Targeted Metabolomics. *Curr Protoc Mol Biol* **98**, 1–34 (2012).

25. Bingol, K. Recent advances in targeted and untargeted metabolomics by NMR and MS/NMR methods. *High Throughput* **7**, (2018).

26. Bingol, K. *et al.* Emerging new strategies for successful metabolite identification in metabolomics. *Bioanalysis* **8**, 557–573 (2016).

27. Smolinska, A., Blanchet, L., Buydens, L. M. C. & Wijmenga, S. S. NMR and pattern recognition methods in metabolomics: From data acquisition to biomarker discovery: A review. *Anal Chim Acta* **750**, 82–97 (2012).

28. Naz, S., Vallejo, M., García, A. & Barbas, C. Method validation strategies involved in non-targeted metabolomics. *J Chromatogr A* **1353**, 99–105 (2014).

29. Wishart, D. S. *et al.* HMDB: the Human Metabolome Database. *Nucleic Acids Res* **35**, D521–D526 (2007).

30. J. Ellinger, J., A. Chylla, R., L. Ulrich, E. & L. Markley, J. Databases and Software for NMR-Based Metabolomics. *Curr Metabolomics* **1**, 28–40 (2013).

31. Markley, J. L. *et al.* The future of NMR-based metabolomics. *Curr Opin Biotechnol* **43**, 34–40 (2017).

32. Lenz, E. M. & Wilson, I. D. Analytical strategies in metabonomics. *J Proteome Res* **6**, 443–458 (2007).

33. Bingol, K. *et al.* Metabolomics beyond Spectroscopic Databases: A Combined MS/NMR Strategy for the Rapid Identification of New Metabolites in Complex Mixtures. *Anal Chem* **87**, 3864–3870 (2015).

34. Wishart, D. S. *et al.* NMR and Metabolomics—A Roadmap for the Future. *Metabolites* **12**, 678 (2022).

35. Raftery, Z. P. D. Comparing and combining NMR spectroscopy and mass spectrometry in metabolomics. *Anal Bioanal Chem* **387**, 525–527 (2007).

36. Emwas, A. *et al.* NMR Spectroscopy for Metabolomics Research. (2019).

37. Jacob, M., Lopata, A. L., Dasouki, M. & Abdel Rahman, A. M. Metabolomics toward personalized medicine. *Mass Spectrom Rev* **38**, 221–238 (2019).

38. McLafferty, F. W. Tandem mass spectrometry. *Science (1979)* **214**, 280–287 (1981).

39. Moco, S. Studying Metabolism by NMR-Based Metabolomics. *Front Mol Biosci* **9**, 1–12 (2022).

40. Cohen, J. S., Jaroszewski, J. W., Kaplan, O., Ruiz-Cabello, J. & Collier, S. W. A history of biological applications of NMR spectroscopy. *Prog Nucl Magn Reson Spectrosc* **28**, 53–85 (1995).

41. Amberg, A. *et al.* NMR and MS methods for metabolomics. *Methods in Molecular Biology* **1641**, 229–258 (2017).

42. Dunn, W. B., Broadhurst, D. I., Atherton, H. J., Goodacre, R. & Griffin, J. L. Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chem Soc Rev* **40**, 387–426 (2011).

43. Crook, A. A. & Powers, R. Quantitative NMR-Based Biomedical Metabolomics: Current Status and Applications. *Molecules* **25**, (2020).

44. Gadian, D. G. & Radda, G. K. NMR Studies of Tissue Metabolism. **50**, 69–83 (2003).

45. Shulman, R. G. *et al.* Cellular Applications of 31P and 13C Nuclear Magnetic Resonance. *Science (1979)* **205**, 160–166 (1979).

46. Hoult, D. I. *et al.* Observation of tissue metabolites using 31P nuclear magnetic resonance. *Nature 1974 252:5481* **252**, 285–287 (1974).

47. Lane, A. N., Higashi, R. M. & Fan, T. W. NMR and MS-based Stable Isotope-Resolved Metabolomics and applications in cancer metabolism. *TrAC Trends in Analytical Chemistry* **120**, 115322 (2019).

48. Buescher, J. M. *et al.* A roadmap for interpreting 13C metabolite labeling patterns from cells. *Curr Opin Biotechnol* **34**, 189–201 (2015).

49. Saborano, R. *et al.* A framework for tracer-based metabolism in mammalian cells by NMR. *Sci Rep* **9**, 1–14 (2019).

50. Letertre, M. P. M., Giraudeau, P. & de Tullio, P. Nuclear Magnetic Resonance Spectroscopy in Clinical Metabolomics and Personalized Medicine: Current Challenges and Perspectives. *Front Mol Biosci* **8**, 880 (2021).

51. Li, J., Vosegaard, T. & Guo, Z. Applications of nuclear magnetic resonance in lipid analyses: An emerging powerful tool for lipidomics studies. *Prog Lipid Res* **68**, 37–56 (2017).

52. Laghi, L., Picone, G. & Capozzi, F. Nuclear magnetic resonance for foodomics beyond food analysis. *TrAC - Trends in Analytical Chemistry* **59**, 93–102 (2014).

53. García-Cañas, V., Simó, C., Herrero, M., Ibáñez, E. & Cifuentes, A. Present and future challenges in food analysis: Foodomics. *Anal Chem* **84**, 10150–10159 (2012).

54. Larive, C. K., Barding, G. A. & Dinges, M. M. NMR spectroscopy for metabolomics and metabolic profiling. *Anal Chem* **87**, 133–146 (2015).

55. Selegato, D. M., Pilon, A. C. & Carnevale Neto, F. Plant Metabolomics Using NMR Spectroscopy. *Methods in Molecular Biology* **2037**, 345–362 (2019).

56. Kim, H. K., Choi, Y. H. & Verpoorte, R. NMR-based plant metabolomics: Where do we stand, where do we go? *Trends Biotechnol* **29**, 267–275 (2011).

57. Giraudeau, P. Challenges and perspectives in quantitative NMR. *Magnetic Resonance in Chemistry* **55**, 61–69 (2017).

58. Wishart, D. S. Quantitative metabolomics using NMR. *TrAC - Trends in Analytical Chemistry* **27**, 228–237 (2008).

59. Kostidis, S., Addie, R. D., Morreau, H., Mayboroda, O. A. & Giera, M. Quantitative NMR analysis of intra- and extracellular metabolism of mammalian cells: A tutorial. *Anal Chim Acta* **980**, 1–24 (2017).

60. Dona, A. C. *et al.* A guide to the identification of metabolites in NMR-based metabonomics/metabolomics experiments. *Comput Struct Biotechnol J* **14**, 135–153 (2016).

61. Claridge, T. D. W. *High-Resolution NMR Techniques in Organic Chemistry*. (2016).

62. McPhee, C., Reed, J. & Zubizarreta, I. Nuclear Magnetic Resonance (NMR). *Developments in Petroleum Science* **64**, 655–669 (2015).

63. Silva, R. A., Pereira, T. C. S., Souza, A. R. & Ribeiro, P. R. 1H NMR-based metabolite profiling for biomarker identification. *Clinica Chimica Acta* **502**, 269–279 (2020).

64. Emwas, A.-H. M., Salek, R. M., Griffin, J. L. & Merzaban, J. NMR-based metabolomics in human disease diagnosis: applications, limitations, and recommendations. doi:10.1007/s11306-013-0524-y.

65. Emwas, A. H. *et al.* Recommendations and Standardization of Biomarker Quantification Using NMR-Based Metabolomics with Particular Focus on Urinary Analysis. *J Proteome Res* **15**, 360 (2016).

66. Mckay, R. T. How the 1D-NOESY suppresses solvent signal in metabonomics NMR spectroscopy: An examination of the pulse sequence components and evolution. *Concepts Magn Reson Part A Bridg Educ Res* **38 A**, 197–220 (2011).

67. Rankin, N. J. *et al.* The emergence of proton nuclear magnetic resonance metabolomics in the cardiovascular arena as viewed from a clinical perspective. *Atherosclerosis* **237**, 287–300 (2014).

68. Parsons, H. M., Ludwig, C. & Viant, M. R. Line-shape analysis of J-resolved NMR spectra: Application tometabolomics and quantification of intensity errors from signal processing and high signal congestion. *Magnetic Resonance in Chemistry* **47**, (2009).

69. Martineau, E. & Giraudeau, P. Fast Quantitative 2D NMR for Untargeted and Targeted Metabolomics. in *Methods in Molecular Biology* vol. 2037 (2019).

70. Martineau, E., Dumez, J. N. & Giraudeau, P. Fast quantitative 2D NMR for metabolomics and lipidomics: A tutorial. *Magnetic Resonance in Chemistry* **58**, (2020).

71. Emwas, A. H. *et al.* Standardizing the experimental conditions for using urine in NMR-based metabolomic studies with a particular focus on diagnostic studies: a review. *Metabolomics* **11**, 872–894 (2015).

72. Stevens, V. L., Hoover, E., Wang, Y. & Zanetti, K. A. Pre-analytical factors that affect metabolite stability in human urine, plasma, and serum: A review. *Metabolites* **9**, (2019).

73. Bruzzone, C. *et al.* 1H NMR-Based Urine Metabolomics Reveals Signs of Enhanced Carbon and Nitrogen Recycling in Prostate Cancer. *J Proteome Res* **19**, 2419–2428 (2020).

74. Embade, N. *et al.* NMR-based newborn urine screening for optimized detection of inherited errors of metabolism. *Sci Rep* **9**, 1–9 (2019).

75. Bruzzone, C. *et al.* SARS-CoV-2 Infection Dysregulates the Metabolomic and Lipidomic Profiles of Serum. *iScience* **23**, (2020).

76. Bruzzone, C. *et al.* A molecular signature for the metabolic syndrome by urine metabolomics. *Cardiovasc Diabetol* **20**, 1–13 (2021).

77. Dona, A. C. *et al.* Precision high-throughput proton NMR spectroscopy of human urine, serum, and plasma for large-scale metabolic phenotyping. *Anal Chem* **86**, 9887–9894 (2014).

78. Jiménez, B. *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* **90**, 11962–11971 (2018).

79. Vignoli, A. *et al.* High-Throughput Metabolomics by 1D NMR. *Angewandte Chemie - International Edition* **58**, 968–994 (2019).

80. Sukumaran, D. K. *et al.* Standard operating procedure for metabonomics studies of blood serum and plasma samples using a 1H-NMR micro-flow probe. *Magnetic Resonance in Chemistry* **47**, S81–S85 (2009).

81. Stavarache, C. *et al.* A Real-Life Reproducibility Assessment for NMR Metabolomics. *Diagnostics 2022, Vol. 12, Page 559* **12**, 559 (2022).

82. Masuda, R. *et al.* Integrative Modeling of Plasma Metabolic and Lipoprotein Biomarkers of SARS-CoV-2 Infection in Spanish and Australian COVID-19 Patient Cohorts. *J Proteome Res* **4**, acs.jproteome.1c00458 (2021).

83. Bernini, P. *et al.* Standard operating procedures for pre-analytical handling of blood and urine for metabolomic studies and biobanks. *J Biomol NMR* **49**, 231–243 (2011).

84. Ghini, V., Quaglio, D., Luchinat, C. & Turano, P. NMR for sample quality assessment in metabolomics. *N Biotechnol* **52**, 25–34 (2019).

85. Barton, R. H., Nicholson, J. K., Elliott, P. & Holmes, E. High-throughput 1 H NMR-based metabolic analysis of human serum and urine for large-scale epidemiological studies: validation study. *Int J Epidemiol* **37**, 31–40 (2008).

86. Maher, A. D., Zirah, S. F. M., Holmes, E. & Nicholson, J. K. Experimental and Analytical Variation in Human Urine in 1H NMR Spectroscopy-Based Metabolic Phenotyping Studies. *Anal Chem* **79**, 5204–5211 (2007).

87. Ghini, V. *et al.* Impact of the pre-examination phase on multicenter metabolomic studies. *N Biotechnol* **68**, (2022).

88. Ghini, V. *et al.* Metabolomic Fingerprints in Large Population Cohorts: Impact of Preanalytical Heterogeneity. *Clinical Chemistry*, **67**, (2021).

89. de Souza, Y. G. & Greenspan, J. S. Biobanking past, present and future: Responsibilities and benefits. *AIDS* **27**, 303–312 (2013).

90. Rotter, M. *et al.* Stability of targeted metabolite profiles of urine samples under different storage conditions. *Metabolomics* **13**, (2017).

91. Beckonert, O. *et al.* Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat Protoc* **2**, 2692–2703 (2007).

92. Bothwell, J. H. F. & Griffin, J. L. An introduction to biological nuclear magnetic resonance spectroscopy. *Biol. Rev* **86**, 493–510 (2011).

93. Emwas, A. H. *et al.* Recommended strategies for spectral processing and post-processing of 1D 1 H-NMR data of biofluids with a particular focus on urine. *Metabolomics* **14**, 1–23 (2018).

94. Day, C. Metabolic syndrome, or What you will: Definitions and epidemiology. *Diab Vasc Dis Res* **4**, 32–38 (2007).

95. Nilsson, P. M., Tuomilehto, J. & Rydén, L. The metabolic syndrome – What is it and how should it be managed? *Eur J Prev Cardiol* **26**, 33–46 (2019).

96. Grundy, S. M. Metabolic syndrome pandemic. *Arterioscler Thromb Vasc Biol* **28**, 629–636 (2008).

97. Alberti, G. Introduction to the metabolic syndrome. *European Heart Journal, Supplement* **7**, 3–5 (2005).

98. Alberti, K. G. M. M. *et al.* Harmonizing the metabolic syndrome: A joint interim statement of the international diabetes federation task force on epidemiology and prevention; National heart, lung, and blood institute; American heart association; World heart federation; International . *Circulation* **120**, 1640–1645 (2009).

99. Alberti, K. G. M. M., Zimmet, P. & Shaw, J. The metabolic syndrome - A new worldwide definition. *Lancet* **366**, 1059–1062 (2005).

100. Strazzullo, P. *et al.* Diagnostic criteria for metabolic syndrome: a comparative analysis in an unselected sample of adult male population. *Metabolism* **57**, 355–361 (2008).

101. Eckel, R. H., Grundy, S. M. & Zimmet, P. Z. The metabolic syndrome. *Lancet* **365**, 1415–1428 (2005).

102. Huang, P. L. A comprehensive definition for metabolic syndrome. *DMM Disease Models and Mechanisms* **2**, 231–237 (2009).

103. Balkau, B. & Charles, M. A. Comment on the provisional report from the WHO consultation. *Diabetic Medicine* **16**, 442–443 (1999).

104. Grundy, S. M. *et al.* Diagnosis and management of the metabolic syndrome: An American Heart Association/National Heart, Lung, and Blood Institute scientific statement. *Circulation* **112**, 2735–2752 (2005).

105. Zimmet, P., Magliano, D., Matsuzawa, Y., Alberti, G. & Shaw, J. The metabolic syndrome: a global public health problem and a new definition. *J Atheroscler Thromb* **12**, 295–300 (2005).

106. Bener, A. *et al.* Obesity index that better predict metabolic syndrome: Body mass index, waist circumference, waist hip ratio, or waist height ratio. *J Obes* **2013**, (2013).

107. Ascaso, J. *et al.* Management of Dyslipidemia in the Metabolic Syndrome Recommendations of the Spanish HDL-Forum. *Am J Cardiovasc Drugs* **7**, 39–58 (2007).

108. Isomaa, B. A major health hazard: The metabolic syndrome. *Life Sci* **73**, 2395–2411 (2003).

109. O'Neill, S. & O'Driscoll, L. Metabolic syndrome: A closer look at the growing epidemic and its associated pathologies. *Obesity Reviews* **16**, 1–12 (2015).

110. Kassi, E., Pervanidou, P., Kaltsas, G. & Chrousos, G. Metabolic syndrome: Definitions and controversies. *BMC Med* **9**, 48 (2011).

111. Noubiap, J. J. *et al.* Global, regional, and country estimates of metabolic syndrome burden in children and adolescents in 2020: a systematic review and modelling analysis. *Lancet Child Adolesc Health* **6**, 158–170 (2022).

112. Bonora, Enzo. & DeFronzo, R. A. *Diabetes Complications, Comorbidities and Related Disorders*. (Springer International Publishing, 2020). doi:10.1007/978-3-030-36694-0.

113. Saklayen, M. G. The Global Epidemic of the Metabolic Syndrome. *Current Hypertension Reports 2018 20:2* **20**, 1–8 (2018).

114. Santos, A. C., Ebrahim, S. & Barros, H. Gender, socio-economic status and metabolic syndrome in middle-aged and old adults. *BMC Public Health* **8**, 62 (2008).

115. Nathan, D. M. *et al.* Impaired Fasting Glucose and Impaired Glucose Tolerance. *Diabetes Care* **30**, 753–759 (2007).

116. Pei, D. *et al.* Impaired Glucose Tolerance and Impaired Fasting Glucose Share Similar Underlying Pathophysiologies. *Tohoku J. Exp. Med* **212**, 349–357 (2007).

117. Tan, G. D., Goossens, G. H., Humphreys, S. M., Vidal, H. & Karpe, F. Upper and Lower Body Adipose Tissue Function: A Direct Comparison of Fat Mobilization in Humans. *Obes Res* **12**, 114–118 (2004).

118. Jensen, M. D. Is Visceral Fat Involved in the Pathogenesis of the Metabolic Syndrome? Human Model. *Obesity* **14**, 20S-24S (2006).

119. Alberti, K. G. M. M., Zimmet, P. & Shaw, J. Metabolic syndrome--a new world-wide definition. A Consensus Statement from the International Diabetes Federation. *Diabet Med* **23**, 469–480 (2006).

120. Koster, A. *et al.* Body fat distribution and inflammation among obese older adults with and without metabolic syndrome. *Obesity* **18**, 2354–2361 (2010).

121. Reaven, G. M. The metabolic syndrome: Is this diagnosis necessary? *American Journal of Clinical Nutrition* **83**, 1237–1247 (2006).

122. Ginsberg, H. N., Zhang, Y. L. & Hernandez-Ono, A. Metabolic syndrome: focus on dyslipidemia. *Obesity (Silver Spring)* **14 Suppl 1**, 2006 (2006).

123. Rizzo, M., Pernice, V., Frasheri, A. & Berneis, K. Atherogenic lipoprotein phenotype and LDL size and subclasses in patients with peripheral arterial disease. *Atherosclerosis* **197**, 237–241 (2008).

124. Hanefeld, M. & Schaper, F. Dyslipidemia in the Metabolic Syndrome. *The Metabolic Syndrome at the Beginning of the XXI Century: A Genetic and Molecular Approach* 347–358 (2005)

125. Adiels, M., Olofsson, S. O., Taskinen, M. R. & Borén, J. Overproduction of very low-density lipoproteins is the hallmark of the dyslipidemia in the metabolic syndrome. *Arterioscler Thromb Vasc Biol* **28**, 1225–1236 (2008).

126. Yanai, H. *et al.* The underlying mechanisms for development of hypertension in the metabolic syndrome. *Nutr J* **7**, 10 (2008).

127. Morse, S. A., Zhang, R., Thakur, V. & Reisin, E. Hypertension and the metabolic syndrome. *American Journal of the Medical Sciences* **330**, 303–310 (2005).

128. Grundy, S. M. Pre-Diabetes, Metabolic Syndrome, and Cardiovascular Risk. *J Am Coll Cardiol* **59**, 635–643 (2012).

129. Penninx, B. W. J. H. & Lange, S. M. M. Metabolic syndrome in psychiatric patients: overview, mechanisms, and implications. *Dialogues Clin Neurosci* **20**, 63–73 (2018).

130. John, A. P., Koloth, R., Dragovic, M. & Lim, S. C. B. Prevalence of metabolic syndrome among Australians with severe mental illness. *Medical Journal of Australia* **190**, 176–179 (2009).

131. Zhang, X. & Lerman, L. O. The Metabolic Syndrome and Chronic Kidney Disease. *Transl Res* **183**, 14 (2017).

132. Pothiwala, P., Jain, S. K. & Yaturu, S. Metabolic Syndrome and Cancer. *Metab Syndr Relat Disord* **7**, 279 (2009).

133. Dietrich, P. & Hellerbrand, C. Non-alcoholic fatty liver disease, obesity and the metabolic syndrome. *Best Pract Res Clin Gastroenterol* **28**, 637–653 (2014).

134. Yki-Järvinen, H. Non-alcoholic fatty liver disease as a cause and a consequence of metabolic syndrome. *Lancet Diabetes Endocrinol* **2**, 901–910 (2014).

135. Velavan, T. P. & Meyer, C. G. The COVID-19 epidemic. *Tropical Medicine & International Health* **25**, 278 (2020).

136. Worobey, M. *et al.* The Huanan Seafood Wholesale Market in Wuhan was the early epicenter of the COVID-19 pandemic. *Science* **377**, 951–959 (2022).

137. Cucinotta, D. & Vanelli, M. WHO Declares COVID-19 a Pandemic. *Acta Biomed* **91**, 157–160 (2020).

138. da Silva Torres, M. K. *et al.* The Complexity of SARS-CoV-2 Infection and the COVID-19 Pandemic. *Front Microbiol* **13**, 30 (2022).

139. Ciotti, M. *et al.* The COVID-19 pandemic. **57**, 365–388 (2020).

140. Esakandari, H. *et al.* A comprehensive review of COVID-19 characteristics. *Biol Proced Online* **22**, 1–10 (2020).

141. WHO Coronavirus (COVID-19) Dashboard | WHO Coronavirus (COVID-19) Dashboard With Vaccination Data. https://covid19.who.int/.

142. Zhang, J. *et al.* Clinical Characteristics of COVID-19 Patients Infected by the Omicron Variant of SARS-CoV-2. *Front Med (Lausanne)* **9**, (2022).

143. Li, W. *et al.* Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature 2003 426:6965* **426**, 450–454 (2003).

144. Sungnak, W., Huang, N., Bécavin, C., Berg, M. & HCA Lung Biological Network. SARS-CoV-2 Entry Genes Are Most Highly Expressed in Nasal Goblet and Ciliated Cells within Human Airways. *ArXiv* **26**, 681–687 (2020).

145. Ashraf, U. M. *et al.* Sars-cov-2, ace2 expression, and systemic organ invasion. *Physiol Genomics* **53**, 51–60 (2021).

146. Gallo Marin, B. *et al.* Predictors of COVID-19 severity: A literature review. *Rev Med Virol* **31**, 1–10 (2021).

147. Jackson, C. B., Farzan, M., Chen, B. & Choe, H. Mechanisms of SARS-CoV-2 entry into cells. *Nature Reviews Molecular Cell Biology 2021 23:1* **23**, 3–20 (2021).

148. Coperchini, F., Chiovato, L., Croce, L., Magri, F. & Rotondi, M. The cytokine storm in COVID-19: An overview of the involvement of the chemokine/chemokine-receptor system. *Cytokine Growth Factor Rev* **53**, 25 (2020).

149. Rothan, H. A. & Byrareddy, S. N. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *J Autoimmun* **109**, 102433 (2020).

150. Lamers, M. M. & Haagmans, B. L. SARS-CoV-2 pathogenesis. *Nature Reviews Microbiology 2022 20:5* **20**, 270–284 (2022).

151. Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**, 497 (2020).

152. Montazersaheb, S. *et al.* COVID-19 infection: an overview on cytokine storm and related interventions. *Virology Journal 2022 19:1* **19**, 1–15 (2022).

153. Al-Mozaini, M. A. & Mansour, M. K. Personalized medicine Is it time for infectious diseases? *Saudi Med J* **37**, (2016).

154. Ayres, J. S. A metabolic handbook for the COVID-19 pandemic. *Nature Metabolism 2020 2:7* **2**, 572–585 (2020).

155. Ward, R. A. *et al.* Harnessing the Potential of Multiomics Studies for Precision Medicine in Infectious Disease. *Open Forum Infect Dis* **8**, (2021).

156. Bizkarguenaga, M. *et al.* Uneven metabolic and lipidomic profiles in recovered COVID-19 patients as investigated by plasma NMR metabolomics. *NMR Biomed* **35**, 1–10 (2022).

157. Ghini, V. *et al.* Serum NMR Profiling Reveals Differential Alterations in the Lipoproteome Induced by Pfizer-BioNTech Vaccine in COVID-19 Recovered Subjects and Naïve Subjects. *Front Mol Biosci* **9**, 1–8 (2022).

158. Rendeiro, A. F. *et al.* Metabolic and Immune Markers for Precise Monitoring of COVID-19 Severity and Treatment. *Front Immunol* **12**, (2022).

159. Correia, B. S. B. *et al.* 1H qNMR-Based Metabolomics Discrimination of Covid-19 Severity. *J Proteome Res* **21**, 1640–1653 (2022).

160. Bruzzone, C. *et al.* SARS-CoV-2 Infection Dysregulates the Metabolomic and Lipidomic Profiles of Serum. *iScience* **23**, (2020).

161. Nitschke, P. *et al.* J-Edited DIffusional Proton Nuclear Magnetic Resonance Spectroscopic Measurement of Glycoprotein and Supramolecular Phospholipid Biomarkers of Inflammation in Human Serum. *Anal Chem* **94**, 1333–1341 (2022).

162. Lodge, S. *et al.* Diffusion and Relaxation Edited Proton NMR Spectroscopy of Plasma Reveals a High-Fidelity Supramolecular Biomarker Signature of SARS-CoV-2 Infection. *Anal Chem* **93**, 3976–3986 (2021).

163. Lodge, S. *et al.* Low Volume in Vitro Diagnostic Proton NMR Spectroscopy of Human Blood Plasma for Lipoprotein and Metabolite Analysis: Application to SARS-CoV-2 Biomarkers. *J Proteome Res* (2021).

164. Nitschke, P. *et al.* Direct low field J-edited diffusional proton NMR spectroscopic measurement of COVID-19 inflammatory biomarkers in human serum. *Analyst* **147**, 4213–4221 (2022).

165. Ghoneim, S., Butt, M. U., Hamid, O., Shah, A. & Asaad, I. The incidence of COVID-19 in patients with metabolic syndrome and non-alcoholic steatohepatitis: A population-based study. *Metabol Open* **8**, 100057 (2020).

166. Marhl, M., Grubelnik, V., Magdič, M. & Markovič, R. Diabetes and metabolic syndrome as risk factors for COVID-19. *Diabetes Metab Syndr* **14**, 671 (2020).

167. Costa, F. F. *et al.* Metabolic syndrome and COVID-19: An update on the associated comorbidities and proposed therapies. *Diabetes Metab Syndr* **14**, 809–814 (2020).

168. Yanai, H. Metabolic Syndrome and COVID-19. *Cardiol Res* **11**, 360 (2020).

169. Swamy, S. *et al.* Hypertension and COVID-19: Updates from the era of vaccines and variants. *J Clin Transl Endocrinol* **27**, 100285 (2022).

170. Yang, J. K. *et al.* Plasma glucose levels and diabetes are independent predictors for mortality and morbidity in patients with SARS. *Diabetic Medicine* **23**, 623–628 (2006).

171. Huttunen, R. & Syrjänen, J. Obesity and the risk and outcome of infection. *Int J Obes (Lond)* **37**, 333–340 (2013).

172. Wu, S., Zhou, K., Misra-Hebert, A., Bena, J. & Kashyap, S. R. Impact of Metabolic Syndrome on Severity of COVID-19 Illness. **20**, 191–198 (2022).

173. Yanai, H. Significant Correlations of SARS-CoV-2 Infection With Prevalence of Overweight/Obesity and Mean Body Mass Index in the SARS-CoV-2 Endemic Countries. *Cardiol Res* **11**, 412 (2020).

174. Leong, A. *et al.* Cardiometabolic Risk Factors for COVID-19 Susceptibility and Severity: A Mendelian Randomization Analysis. *medRxiv* 2020.08.26.20182709 (2020).

175. Scalsky, R. J. *et al.* Baseline Cardiometabolic Profiles and SARS-CoV-2 Risk in the UK Biobank. *medRxiv* 2020.07.25.20161091 (2020).

176. Carneiro, G., Radcenco, A. L., Evaristo, J. & Monnerat, G. Novel strategies for clinical investigation and biomarker discovery: A guide to applied metabolomics. *Horm Mol Biol Clin Investig* **38**, (2020).

Chapter 2

Hypothesis and Objectives

2.1 Hypothesis

Metabolic syndrome and SARS-CoV-2 infection can be considered as two complex disorders. Our hypothesis was that, both these pathologies produce important metabolic alterations which can be analysed by NMR-based metabolomics. The study of the latter is fundamental for a better understanding of the involved mechanism in these disorders which could lead to a personalized treatment of the affected subjects.

2.2 Objectives

2.2.1 A molecular discrimination of the metabolic syndrome by urine and serum metabolomics

The main goals of this project were:

- The identification of new biomarkers and the validation of the existing ones through the study of the relative contributing risk factors involved in the development of MetS.
- To determine, at the molecular level, differences between peoples with and without MetS.
- To design a tool for the determination of a "MetS score" able to define how likely a subject is to develop MetS based on the metabolomic analysis of serum and/or urine samples.
- To evaluate the putative metabolic relationship with MetS of some other risk factors, like aging and NAFLD, that may enhance MetS probability.

2.2.2 Metabolomic and Lipidomic dysregulation caused by SARS-CoV-2 infection

The main goals of this project were:

- To investigate the characteristic metabolic and lipidomic serum profile in hospitalized patients diagnosed with COVID-19 and showing clear manifestations of the disease.
- The identification of new biomarkers specifically associated with this disease.

- To evaluate the possibility of the development of secondary problems, such as an increased atherosclerotic risk or liver damage, due to the produced metabolic and lipidomic changes by SARS-CoV-2 infection.

Chapter 3

Materials and Methods

3.1 Samples cohorts

Different cohorts of samples were included into the conducted studies. All samples were obtained following the principles of the Declaration of Helsinki, which protects patients by establishing a set of ethical standards for medical research involving the use of human samples. All the patients that participated into the studies gave their informed consent and their data were anonymized to protect their privacy. Moreover, the projects were approved by the corresponding ethics committee.

For each of the samples included into the studies a number of metadata were obtained, according to the cohort, to help with patients' classification and to direct the analyses. Biochemical data were obtained from aliquots of serum and urine samples collected in the same sample extraction day.

3.1.1 Metabolic syndrome study cohorts

The cohorts of samples included in this study are listed below:

- OSARTEN: including urine and serum samples from the working population of the Basque Country, man and women from 19 to 66 years old. These samples were recollected by the Osarten Kooperatiba Elkartea (Mondragon Cooperative) as additional aliquots during the routine medical check-ups of their employees. No particular exclusion criteria were applied with the exception of having suffered from a serious illness like cancer or ictus in the 3 months preceding the sample collection. Detailed information about the urine and serum samples included into the study is listed in **Table A1** and **Table A2** respectively in the Appendix.
- OBENUTIC: the acronym of this cohort comes from "Obesidad, Nutrición y Tecnologías de la Información y Comunicación", a study on obesity conducted by the Preventive and Public Health department of the Faculty of Medicine of Valencia. This cohort include urine samples from men and women with a range of age between 18 and 60 of the Community of Valencia and with a BMI between 20-35 kg/m^2. Exclusion criteria were applied, like suffering from any

infectious-contagious disease, physical or mental incapacitating illness, being in pregnancy or lactation, having a diagnosis of cancer, thyroid disorders and/or type I diabetes and Cushing's disease or using medication altering blood concentration of lipids. More detailed information is listed in **Table A3** of the Appendix.

- PREDIMED: consists in urine samples from a bigger prevention study based on the effect of the consumption of Mediterranean diet. For our study the included subjects presented an age between 55 and 80 years old for men and from 60 to 80 years old for women. Moreover, individuals must present type 2 diabetes or a diagnosed higher cardiovascular risk. Detailed information is listed in **Table A4** in the Appendix.
- KIROLGETXO: represented by urine samples from senior individuals, older than 65 years old, practicing sport routinely at the Sport Centre of Getxo. Information of these subjects is listed in **Table A5** in the Appendix.
- NAFLD: includes urine samples from men and women between 18-75 years old from different part of Europe who underwent a liver biopsy. More information of these subjects is included in **Table A6** in the Appendix.
- MetS long: consists in serum samples from the biobanks of the University of Navarra and from the University Hospital Ramón y Cajal. Individuals presenting one or more risk factors for the development of the metabolic syndrome like diabetes, obesity, dyslipidemia or hypertension were recruited to form this cohort. The included subjects did not suffered from a serious illness like cancer or ictus in the 3 months preceding the sample collection. More information of these individuals is included in **Table A7** in the Appendix.
- PORTUGAL: including serum samples form senior subjects recovered in retirement homes from a specific region of Portugal (Beira Interior region). More information of these subjects is included in **Table A8** in the Appendix.

For the majority of these cohorts, the biochemical data were available in order to classify patients according to their risk factors involved in the development of the metabolic syndrome.

3.1.2 COVID-19 study cohorts

The cohorts of samples included in this study are listed above:

- *preCOVID*: includes serum samples recruited in 2018/2019, before the onset of the pandemic, from some patients of the OSARTEN cohort, previously described in the metabolic syndrome project cohorts. These samples were used as controls in order to compare the unhealthy patients with healthy ones, recollected long before COVID-19 spread. Information regarding the selected samples is listed in **Table A9** in the Appendix.
- *COVID*: includes serum samples from hospitalized COVID positive subjects in the acute phase of the infection recruited in the Basurto and Cruces University Hospitals during the first wave of the pandemic. Patients were presenting symptoms compatible with COVID-19 and were testing positive to RT-PCR (Real-Time reverse-transcription Polymerase Chain Reaction) targeting the viral RNA of SARS-CoV-2 virus on nasal swab. Information regarding the selected samples is listed in **Table A9** in the Appendix.

3.2 NMR measurements

NMR was the selected technology to conduce the performed samples measurements in the metabolic studies carried out in this thesis. Some of the basic theoretical principles on which this technique was based were explained in Chapter 1. Here we described the used spectrometers and the experimental procedures that characterize this measurement process.

3.2.1 The Equipment

For the analysis of the serum and urine samples two Bruker Biospin NMR spectrometers were used:

- A 600 MHz AVANCE III HD (IVDr), called 601, coupled with an automatic sample changer (SampleJet). This spectrometer is equipped with a Double

Resonance Broadband Probe (BBI) with a z gradient coil and with 3 channels: ^{1}H, ^{2}H, X (BB-F).

- A 600 MHz AVANCE NEO (IVDr), called 602, coupled with an automatic sample changer (SampleJet). This spectrometer is equipped with a Double Resonance Broadband Probe (BBI) with a z gradient coil and with 4 channels: ^{1}H, ^{2}H, X (BB-F), ^{13}C or ^{31}P.

Both magnets were used for urine and serum acquisition according to need. Furthermore, it was proven that, by following strict SOPs, it was possible to use both NMR magnets without any variability in the measurements[1]. The 600 MHz spectrometers are considered as a good compromise for the obtained spectra sensitivity and resolution in relation to the related cost. For this reason, 600 MHz spectrometers are among the most used ones in the metabolomic field.

The SampleJet consist in a robotic unit coupled to the spectrometers with five refrigerated rack positions (each with a capacity of 96 NMR tubes, for a total of 480 tubes) that allows to maintain samples at the desired temperature before and after measurement. Additionally, the SampleJet could host up to 96 tubes without spinner in a not refrigerated carousel. Non-temperature-sensitive samples that must be measured daily (the calibration tube described further) were stored in this area.

Bruker TopSpin software is used in order to control the NMR spectrometers and to proceed with data analysis, acquisition, processing and a wide range of additional functions. Moreover, IconNMR software allows the fully automated NMR spectra acquisition thanks also to its connection with the SampleJet robotic unit. The use of these programs is essential especially for high throughput studies with a large number of samples.

3.2.2 Magnet calibration

As briefly explained in the introduction (Chapter 1), before samples NMR measurements it was necessary to check the performance of the spectrometer that was going to be used. In order to do that, three different samples were measured daily: methanol, QuantRef and sucrose.

3.2.2.1 Temperature calibration

Temperature calibration was realized measuring a methanol sample. According to which biofluid was measured a specific temperature was set and the NMR spectrometers was left to stabilize at least all night long before starting with the temperature calibration, to avoid fluctuations and ensure that the magnet reached the correct temperature. For urine measurements the magnet was set at 300K while for serum/plasma at 310K. The exact calibration is important because small molecules and lipoproteins are sensible to temperature changes that can provoke chemical shifts. It's important to ensure that all the samples from the same project were measured always at the optimal temperature in order to avoid these shifts.

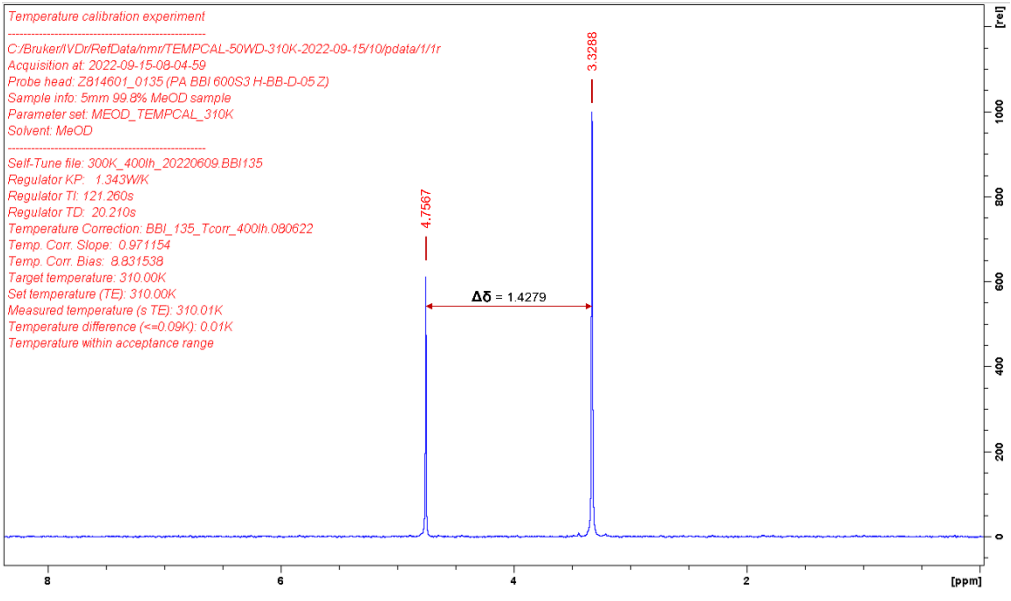

Figure 3.1: Representative NMR spectra of deuterated methanol automatically measured to determine the real temperature of the magnet (300K for urine measurements; 310K for serum measurements). If the calculated distance is in line with the selected temperature, the message "temperature within acceptable range" appears in the display.

A 5 mm sealed NMR tube containing 99.8% deuterated methanol (MeOD) was preheated for 300 seconds before proceeding with measurement. To verify if the temperature inside the magnet was exactly the set one, deuterated methanol was used due to the known temperature dependence of its NMR signals. In fact, the OH-CH_2 chemical shift difference is inversely proportional to the increase in temperature,

(decrease as the temperature rise) due to its association with the reduced hydrogen bounding[2]. A quick pre-set experiment from Bruker was acquired in the IconNMR automatic routine. After that, the spectrum was automatically processed (line broadening of 0.3 Hz) and, to check the real temperature inside the magnet, the distance in ppm between the CH_3 and OH methanol signals was automatically calculated ($\Delta\delta$) in TopSpin. Thanks to that, it was possible to know if temperature was within the acceptable range and to continue with the following step of the magnet calibration (**Figure 3.1**)[2,3].

3.2.2.2 QuantRef calibration

To guarantee the optimal absolute quantification of the metabolites present in the samples, QuantRef (quantification reference solution) was calibrated daily. Moreover, this is also important for comparisons between NMR platforms and to assess the reliability of the results over time to compare samples concentrations.

Quantification is possible thanks to the presence of an external synthetic radio frequency signal called ERETIC (Electronic Reference To access In-vivo Concentrations). In this way an additional peak appears in the NMR spectra, far away from the NMR peaks of the metabolites in the solution, avoiding overlapping problems. QuantRef sample consists in a mixture of metabolites commonly present in biofluids at a known concentration and stable over time inside a 5 mm sealed NMR tube. To allow quantification in complex matrix, like the ones of biofluids, was necessary to calibrate the QuantRef before each samples set. A NOESY experiment was acquired in IconNMR automation mode after sample preheating for 300 seconds at the selected temperature, according to the biofluid that was going to be measured. NMR acquisition was done with the same parameters further used for the acquisition of the urine or serum samples. The resulting spectra was checked in order to determine if the quality of the measurement was within the acceptable range (**Figure 3.2**). In order to determine it, the half height of the DSS signal has to be lower than 1 Hz and the NMR signals of the known metabolites present in solution are integrated in order to assess the *in vivo* concentration of the ERETIC factor applying the following formula:

$$ERETIC = \frac{I_{Ref} \cdot SW_{Ref} \cdot M_{Ref}}{SI_{Ref} \cdot C_{Ref} \cdot NH_{Ref}}$$

Where *Ref* are the reference metabolites of known concentration, *I* represents the absolute integral, *SW* the spectral with, *M* the reference substance's molecular weight, *SI* the size of the real spectrum, which shows the number of data points after Fourier transformation, *C* the known concentration of the reference substances and *NH* the number of protons giving rise to the considered NMR signal[4,5].

From the present metabolites and the considered NMR signals, an average ERETIC factor is obtained. To determine if the measured QuantRef results is within the acceptable range, the maximum internal deviation of the calculated ERETIC factor after the daily measurement must be smaller than the 4%, otherwise the experiment must be repeated and all the parameters checked. If the calibration results as acceptable, the ERETIC factor can be used for all the following quantifications of urine or serum samples. In our case, the described procedure is performed automatically by the TopSpin software.

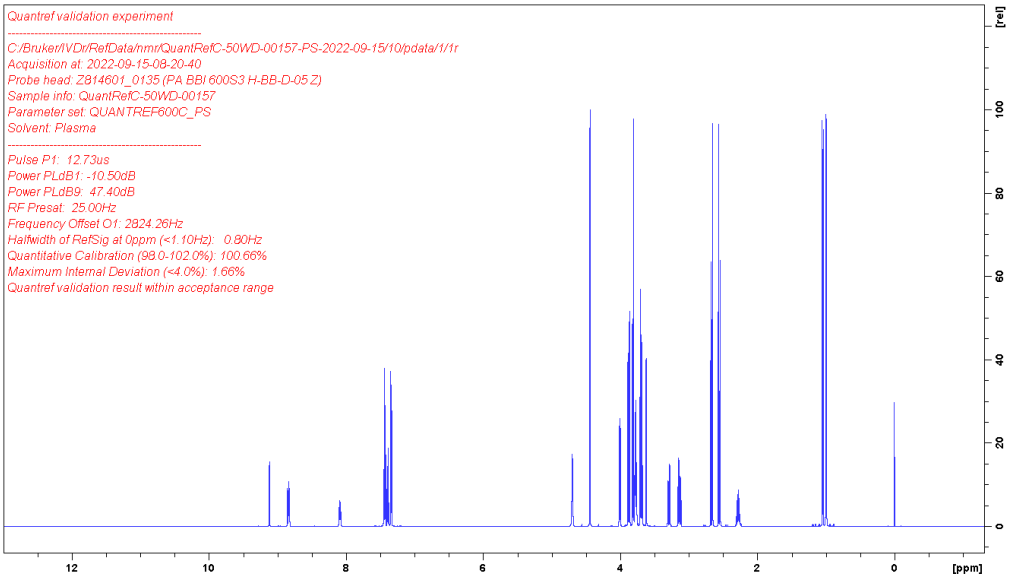

Figure 3.2: Representative NMR spectra of QuantRef sample. If the half height of the DSS signal is lower than 1 Hz and the calculated internal deviation of the coefficient of the ERETIC factor is smaller than the 4%, the message "QuantRef within acceptable range" appears in the display.

3.2.2.3 Shim Performance and Water Suppression Test

To check the integrity, the stability and the performance of the NMR system different NMR measurements were performed on a sucrose sample (0.5 mM DSS, 2 mM NaN_3 in 90% H_2O and 10% D_2O) in the automation mode of IconNMR.

The first measured experiment was necessary for the O1 offset optimization measuring a 1D NMR experiment with water presaturation, by a long relaxation delay and only 1 scan. The magnetic field homogeneity is checked looking at the half height of the DSS signal that has to be lower than 1 Hz. After that, it was necessary to evaluate the water suppression performance without which water signal would be the only one visible in the spectrum. An experiment processed with 8 scans cycles and the previously set saturation frequency, power and acquisition parameters was automatically performed. From this experiment different important parameters were checked (**Figure 3.3**): the signal-to-noise looking at the anomeric peak of sucrose, that has to present a value higher of 300. The resolution, measuring the splitting of the signal, that has to be better of the 15% and finally the water hump, measured at 50% and 10% of the DSS signal intensity, that must not be bigger than 30 and 50 Hz respectively. Finally, a last gradient experiment was acquired in order to check the probe integrity.

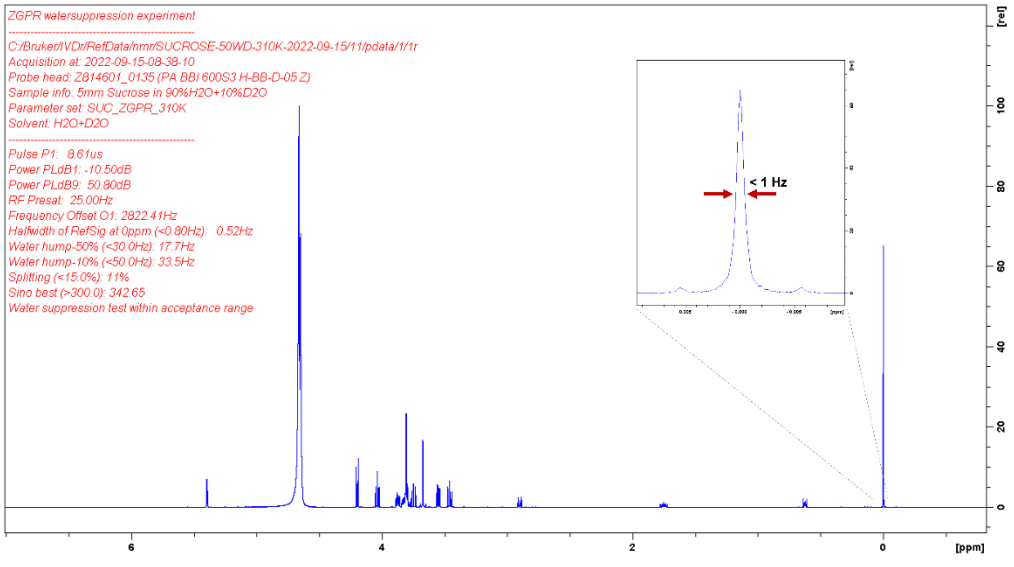

Figure 3.3: Representative ^1H NMR spectra of sucrose measurement during daily calibration. All the checked parameters are listed on the left and an amplified section of DSS signal is showed. If all the parameters are in the correct range, the message "water suppression test within acceptable range" appears in the display.

3.2.3 NMR samples preparation

Samples were prepared following strict standard operating procedure to obtain reliable results as previously mentioned in Chapter 1.

3.2.3.1 Urine samples preparation

Urine was preserved in the freezer at -80°C. After thawing at room temperature for 30 minutes, samples were centrifuged at 6000 RPM for 5 minutes at 4°C. 630 µL of the urine supernatant was transferred to a 1.5 ml Eppendorf with 70 µL of urine buffer (1.5 M KH_2PO_4/K_2HPO_4, 2 mM NaN_3, 1% TSP (trimethylsilylpropionic acid-d_4 sodium salt) in 70% D_2O, pH 7.4 ± 0.1). After mixing the buffered urine with the vortex for a couple of seconds, 600 µL of well mixed sample was transferred into a 5 mm NMR-tube. NMR tubes containing the prepared samples were further manually shaken for several seconds and stored inside the rack at 5°C in the SampleJet until acquisition (**Figure 3.4**).

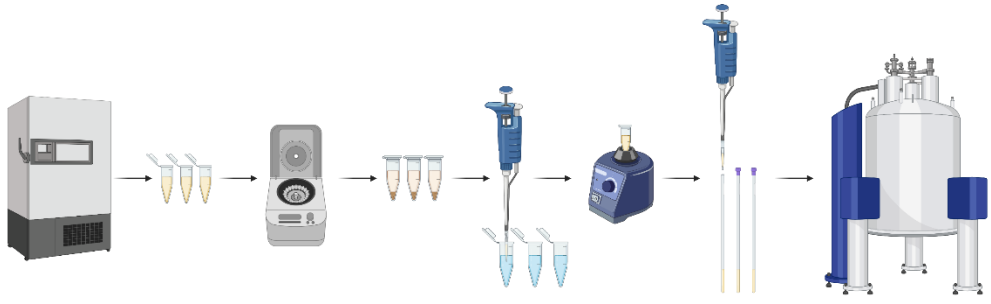

Figure 3.4: Schematic representation of the urine samples preparation. (Figure created with BioRender.com).

3.2.3.2 Serum samples preparation

Blood samples were collected into EDTA free extraction tubes to avoid contamination for the following NMR measurements. Serum aliquots obtained after blood centrifugation were stored in the -80°C freezer and then used to prepare NMR samples both manually and using the SamplePro Tube robotic system (Bruker Biospin).

For manual preparation, samples were defrosted at room temperature for 30 minutes. NMR samples were prepared mixing the serum buffer (75 mM Na_2HPO_4, 2mM NaN_3,

4.6 mM TSP in H_2O and 10% D_2O, pH 7.4 ±0.1) with the serum sample at a 1:1 (v/v) ratio in a 5 mm NMR tube for a final volume of 600 µL. NMR samples were gently shaken for several seconds and kept at 5°C in the SampleJet until measurement (**Figure 3.5**).

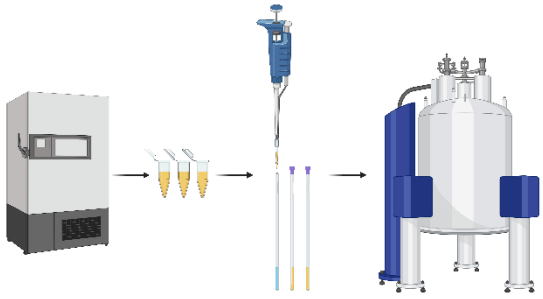

Figure 3.5: Schematic representation of serum sample preparation. (Figure created with BioRender.com).

Serum samples preparation with the SamplePro Tube (SPT) was done automatically mixing the serum buffer with the samples in a 1:1 ratio (v/v), as for manual samples, for a final volume of 600 µL into a 5 mm NMR tube. Prepared samples were removed from the cooling area of the SPT, tube's cups were sealed with POM balls and samples were manually gently shaken for several seconds. After that, samples were kept at 5°C in the SampleJet until measurement.

SamplePro Tube operation required procedures for the setting of the instrument functionality for the preparation of samples at the correct ratio, and a daily calibration. "SPT Calibration Day", the procedure for the correct setting of instrument functionality, must be done every three months or each time changes were made to the NMR magnet or the QuantRef was recalibrated. This consisted in the preparation of some control samples form a serum pool both manually and with the SPT. Samples prepared manually were done mixing the serum buffer with the serum pool at the perfect 1:1 ratio calculated using a precision balance and taking into account the serum density. A specific amount of serum buffer was added into a glass vial and weighted using the precision balance. The obtained value of was divided by 1,025 to know the

exact serum pool amount in weight that must be added to the glass vial already containing the serum buffer. The obtained mix of serum buffer and serum pool was gently mixed to homogenise the pool sample, dived into 5 mm NMR tubes for a final volume of 600 µL in each one of them and a NOESY experiment was measured for the obtained samples.

Meanwhile, other control samples with the same serum pool were prepared using the SPT with three different set up for the additive amount of 290, 300 and 310 µL (usually 3 samples with each set up) to build up a calibration curve. The prepared samples were gently manually mixed and a NOESY experiment was acquired for each one of them. After the acquisition of all the samples, the one prepared with the balance and the one prepared with the SPT, a MatLab script was run to build up the mentioned calibration curve and to extrapolate the value of additive amount that must be set in the SamplePro Tube to prepare samples at the exact 1:1 ratio. To verify that the established additive volume was the correct one, three samples were prepared with this amount of additive with the SPT and measured by NMR. Their validity was checked with the "SPT daily calibration" test, explained below.

Regarding the SamplePro Tube daily calibration, this consisted in two steps: the preparation of four NMR samples with a blank solution (1.1% NaCl water solution with 0,02% w/v [ca. 3mM] NaN_3) with the same method used for real samples preparation, mixing the blank solution with buffer in a 1:1 ratio for a final volume of 600 µL inside the 5 mm NMR tube, needed just to check visually if the amount of sample was equal in all the four NMR tube prepared. Secondly, a set of three control samples from a serum pool (the same one previously used for the "SPT Calibration Day") were prepared with the method later used for real samples preparation to check the SPT performance. A NOESY experiment was acquired for each serum control sample and, after measurement, a MatLab scrip was run in order to compare the prepared samples with the previously prepared one using the precision balance on the "SPT Calibration Day". If the variability of the acquired samples was within an acceptable range, real samples could be prepared using the SPT.

3.2.4 NMR acquisition

After preparation, samples were acquired according to the defined standard operating procedures to ensure the comparison between different laboratories and the reliability of results in long-term studies.

3.2.4.1 Urine samples NMR measurements

All the urine samples were acquired at 300K. For each sample two experiment were measured: a 1D ^{1}H spectrum with water presaturation using the NOESY pulse sequence (*noesygppr1d*) and a 2D JRES (*jresgpprqf*). For a subset of samples, to help with metabolites identification, a 2D ^{1}H TOCSY experiment was acquired.

1D-NOESY experiment is the most used one for metabolomics measurements of biofluid, as previously mentioned in Chapter 1[6]. This is related especially to the characteristics of this pulse sequence for the optimal water suppression, an important factor for this kind of samples, with the implication of little optimization processes[7]. In fact, water signal must be well suppressed to guarantee the optimal metabolites identification and quantification due to the impact that the presence of water signal could have on the baseline of the spectra[8]. Moreover, water peak could overlap with metabolite's ones making impossible their identification.

3.2.4.2 Serum samples NMR measurements

All the serum samples were acquired at 310 K. Three different experiments were recorded for each sample: a 1D ^{1}H spectrum with water presaturation using the NOESY pulse sequence (*noesygppr1d*), a 2D JRES spectrum (*jresgpprqf*) and a 1D ^{1}H Carr-Pulcell-Maiboom_Gill (CPMG) spin-echo experiment (*cpmgpr1d*). The latter allowed the selective removing of the macromolecules NMR signals in the resulting spectra enabling a better observation of the small molecules. Indeed, this experiment was designed according to the characteristic very short relaxation time of macromolecules. This permits to observe only the small molecules NMR signals, which indeed have larger relaxation time, by properly choosing the T2 filter.

3.2.5 Bruker Reports

After the measurement of each sample, Bruker reports were obtained. Both for urine and serum measurement a Quality Control (QC) report was given to check that the experiments performed meet the agreed standards and could be included in the study. Moreover, specific reports were given, according to the measured biofluid, for the quantification of the metabolites in solution and, in the case of serum, also for lipoproteins quantification.

3.2.5.1 Urine report

As regard Urine, the Bruker IVDr BioBank QC (B.I.BioBankQC^{TM}) report verified the quality of the measured experiment with a number of specific tests: NMR experiment parameter test, NMR experiment quality test, NMR preparation quality test, matrix identity, integrity and contamination test, medication test, protein background test and finally the further indicative parameter test. The NMR experiment parameter test was used to check that the correct values were used for the acquisition. The NMR spectral quality parameters test checked the shimming performance looking at TSP linewidth and symmetry together with the absolute residual water signal intensity value. Moreover, sample preparation was verified by means of TSP concentration. The matrix identity and integrity were examined to discard bacterial growth or degradation. In addition, the presence of contaminants or medication related metabolites was controlled because, during the pre-analytic procedures, some undesired compounds could be introduced impairing spectra fingerprinting. The presence of disturbing background proteins in urine causing baseline alteration was also checked. Finally, further indicative parameters were examined including other common contaminants like acetone, acetoacetic acid, 3-hydroxybutiric acid or glucose that should be not present in fasting conditions.

In addition to the B.I.BioBankQC^{TM} report, after urine measurement two kind of reports (B.I.Quant-UR b^{TM} or B.I.Quant-UR e^{TM}) with different number (50 or 150) of quantified metabolites could be obtained. Both reports, although useful in verifying the presence of certain metabolites, were not, however, used in our urine analysis. For the

purpose of our work, we preferred instead to use the binning strategy, explained further.

3.2.5.2 Serum reports

As for urine, the B.I.BioBankQC™ report for serum samples was obtained after each acquisition to verify the quality of the measured experiments. The obtained report, in this case, was achieved performing a slightly different number of tests with respect to the ones previously mentioned for urine: the NMR experiment parameter and quality test, the NMR preparation quality test and the matrix identity, integrity and contamination test. The NMR experiment parameters test was used to check again that the optimal values were used for the acquisition. The NMR spectral quality parameters test was used to check the shimming performance but, in this case, instead of the TSP, alanine doublet was controlled. Moreover, the absolute value of the intensity of water residual signal was used to check water suppression. The NMR sample preparation quality parameters test was performed to check that samples were prepared with the correct buffer-sample ratio (1:1 v/v). To that aim, TSP concentration was calculated together with the intensity of the background protein in the spectral range from 6 to 12 ppm and the alanine shift. The matrix identity test was then used to recognize different matrices commonly implied in blood recollection tubes: EDTA-plasma, citrate plasma and heparin plasma and serum. Moreover, the matrix integrity parameters test was used to check if samples were not degraded or prone to bacterial growth. Finally, the matrix contamination parameter test ensured the absence of metabolites like isopropanol coming from pre-analytical procedures or other contaminants (0.8-1.25 ppm) related to low quality cryovials used for storage.

In addition to the previous one, another report was also obtained in order to quantify 41 metabolites that could be present in the serum samples: the Bruker IVDr Quantification in Plasma/Serum B.I.Quant-PS™. This included small molecules grouped according to their chemical class: alcohol, amines, amino acids and related derivatives, carboxylic acids, essential nutrient, keto acids and sugars and related derivatives, sulfones and technical additives (listed in **Table A11** in the Appendix). The concentration of each metabolite, if present and detectable according to their limit of detection (LOD), was reported in mmol/L.

Finally, in the serum analysis an additional report was obtained: the Bruker IVDr Lipoprotein Subclass Analysis B.I.LISA™. Thanks to this additional report the quantification of 112 lipoproteins parameters was possible (listed in **Table A12** in the Appendix). Specifically, it was possible to assess information about the main VLDL, IDL, LDL and HDL classes, six VLDL subclasses (VLDL-1 to VLDL-6), six LDL sub-classes (LDL-1 to LDL-6) and four HDL-subclasses (HDL-1 to HDL-4). Lipoproteins were sorted according to the increasing density and consequent decreasing size (look at **Table 3.1**).

Table 3.1: Lipoproteins density is expressed in kg/L and they are sorted according to the increasing density and decreasing size. VLDL (Very Low Density Lipoprotein); IDL (Intermediate Desnity Lipoprotein); LDL (Low Density Lipoprotein); HDL (High Density Lipoprotein).

Lipoprotein Main Fractions					
VLDL		**IDL**	**IDL**		**HDL**
0.950 - 1.006		1.006 - 1.019	1.019 - 1.063		1.063 - 1.210
Very Low Density Lipoprotein subfractions					
LDL-1	**LDL-2**	**LDL-3**	**LDL-4**	**LDL-5**	**LDL-6**
1.019 - 1.031	1.031 - 1.034	1.034 - 1.037	1.037 - 1.040	1.040 - 1.044	1.044 - 1.063
High Density Lipoprotein subfractions					
HDL-1		**HDL-2**	**HDL-3**		**HDL-4**
1.063 - 1.100		1.100 - 1.112	1.112 - 1.125		1.125 - 1.210

3.2.6 Metabolites and lipoprotein quantification

Quantification by NMR considered the number of protons which could give rise to an NMR signal, and the intensity of the peak, which depends not only on the number of protons, but also on the concentration of the metabolite under investigation[9]. Moreover, metabolite peak intensity had to be related to a reference signal with known concentration. This signal could be internal or external. In our study, metabolites were quantified by means of the ERETIC calibration signal, previously explained.

Quantification is not always an easy process due to the overlapping signals typical of biofluids. For this reason, the use of automated systems that consider the possible overlap of certain signals, and use signal deconvolution tools to optimise quantification, could be helpful. In our study, the quantification was automatically

carried out using Bruker's software, which took this information into account, and obtained within the previously described reports.

As regard the lipoprotein quantification, this was realized following the same principles explained so far. However, the quantification of lipoproteins, proved to be more complex. It was necessary to study the signal profile coming from methyl and methylene groups in greater detail to proceed with the correct deconvolution of the signal. Indeed, the B.I.LISA™ report quantification used a regression model established on the combination of a training data set based on NMR measurements and ultracentrifugation data.

3.2.7 Metabolites identification

Metabolites identification was realised consulting databases like the Biological Magnetic Resonance Databank and the Human Metabolome Database, both free and available online[10]. The BMRB includes NMR spectra of a great number of small molecules and allows to search for metabolites starting from specific chemical shift and giving as result a list of compounds in the searched area. This option is also available in HMDB. Moreover, the HMDB contains information about most of the small molecules that could be find in human body. For each metabolite it is possible to know the biofluid in which was detected, a description, the biological properties, its normal concentration range, the references of the studies that detected the metabolite and, if available, the NMR spectra with the related chemical shift and multiplicity of the signals.

Metabolites identification was also possible thanks to the use of the Chenomx software for the NMR mixture analysis. The latter allows the metabolites identification by the fitting of libraries of compounds with the desired spectra.

3.3 Samples processing and analysis

NMR metabolomics must often dial with important challenges managing huge amount of data. For this reason, to obtain reliable results, specific steps must be followed according with pre-established standard operating procedures. As previously explained

in the introduction, metabolomics workflow after samples acquisition, consisted in spectral processing and data analysis.

3.3.1 Spectral processing

The spectral processing included different steps for the optimization of the acquired NMR spectra eliminating small errors coming from the acquisition. NMR spectra processing must be done exactly in the same way in all the samples included into a study. This consisted in chemical shift referencing and phase and baseline correction[11].

Chemical shift referencing must be performed looking at the internal reference compound signal, like TSP, which must be referenced to zero. Alternatively, the NMR peak of a compound, known to be in solution and not sensitive to pH changes, can be referenced to the correct chemical shift. This process was particularly important for the correct alignment of the spectra and for the identification of metabolites[12]. As regard phase correction, this was necessary in order to correct possible distortion, ensuring the perfect symmetry of all the NMR peaks in the spectra for the following correct integration for quantification purposes[13]. Finally, baseline correction must be executed to eliminate possible alterations that may have an effect, once again, on metabolite quantification due to incorrect signal integration[11].

In the routine acquisition performed, spectra processing was automatically done using Topspin, thus ensuring the reduction of errors due to erroneous sample-dependent processing, which would lead to further mistakes in the subsequent analysis process.

3.3.1.1 Spectral binning

One of the following post-processing steps to proceed with the statistical analysis was the spectral binning. This step consisted in the segmentation of the NMR spectra into spectral areas of the desired width (generally from 0.02 to 0.05 ppm) called bins. For urine samples, segmentation of the ^{1}H NMR spectra (NOESY), already processed by TopSpin, was done between 0.5 and 9.5 ppm, apart from the area of water signal from 4.7 to 5.0 ppm to avoid strong signals. Each spectrum was divided in 290 buckets of equal size of 0.03 ppm. Each bin was summarized as the average of its points and divided by the total spectrum intensity (sum of all points removing water region). This

normalization minimized the effect of possible concentration differences. The process of binning was applied to proceed with some of the following statistical analysis using the generated bins as variables.

3.3.2 Statistical analysis

Statistical analysis was performed to obtain information regarding the examined cohort of samples. Univariate and multivariate analysis were applied, and different methods were used to conduct these studies including supervised and unsupervised techniques[14]. Here follows an explanation of the main statistical tools used in the context of this thesis.

3.3.2.1 Elimination of outlier samples

Outlier elimination was executed to avoid distortion of statistical analysis. To that aim, a multivariate clustering algorithm was applied after Pareto Scaling, which must be used to normalise the distribution of NMR signals, especially in complex spectra such as those of biofluids where some metabolites with high concentrations may prevail those of less intense small molecules, affecting subsequent analyses.

The used multivariate clustering algorithm was the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) with scaled bins as input variables for metabolic syndrome cohorts or metabolites and lipoproteins for COVID-19 study cohorts[15]. This algorithm tried to identify groups with high density, characterized by a large number of neighbours, taking into account the multivariate space composed by the variables. The groups with low density or isolated samples were marked as extreme and considered as outlier.

This process was executed on the samples of the metabolic syndrome cohorts (Chapter 4) and on the COVID-19 project cohorts (Chapter 5). A total of 9,367 (94%) samples in the OSARTEN cohort, 960 (98%) for PREDIMED, 465 (96%) for OBENUTIC, 234 (100%) in NAFLD and 101 (100%) for KIROLGETXO in each of the examined cohorts, were considered as valid for the following statistical analysis after outliers exclusion. As regard the COVID project only 6 outliers were detected in the *COVID* cohort and none in the *preCOVID* one.

3.3.2.2 Univariate analysis

Various methods could be used to analyse the obtained data, the following were the most used in this thesis:

- Box plots: were used for the representation of the distribution of an examined variable, as an alternative to histograms. These were particularly useful because of their graphical representation which was giving information about the median and interquartile range[16].
- Violin plots: were used as an alternative to box plot for a better visualization of the distribution because they allowed to also evidence the density of the examined data. The included characteristics curves (kernel density plot), surrounding the distribution of the values, gave an estimation of the probability that another variable in the examined population exhibits a value in this area. Wider areas indicated a higher probability while skinner sections a lower one[17].
- Scatter plots: were used to indicate the correlation between two variables. They could include a regression line indicating their tendency.
- Heatmaps: were used to visualize the result of the conduced univariate analysis when several variables were compared. They included several information and the application of different test, in the statistical analysis process, was necessary. These were generally performed to confirm a hypothesis under investigation. The Wilcoxon test (nonparametric) and the t-test (parametric) were frequently applied in order to identify bins (or other variables) with statistically significant differences between two groups. Both reached very similar values with the studied metabolic data and were therefore used interchangeably. A p-value below 0.05 (less than 5% of probability that the null hypothesis [no difference] was true) was used as threshold to determine if the observed differences were statistically significat. In the context of our analysis, asterisks within the cells in the described heatmap, represented the calculated p-value (*: p-value < 0.05; **: p-value < 0.01; ***: p-value < 0.001; ****: p-value < 0.0001). This process was executed after the application of another method: the false discovery rate (FDR) correction used to minimize

type I errors due to multiple comparisons[18]. Moreover, the binary logarithm of the fold-changes (log_2FC) was used to detect in which direction was going the magnitude of the observed differences allowing to the detect the up or down regulated variables, represented with a colour code in the heatmap. Finally, variables were grouped according to hierarchical clustering algorithm based on distances between different profiles (Euclidean distances or Pearson correlation distances) with a multivariate approach[14]. Hierarchical clustering was applied in order to group similar profiles following the complete (calculating distances between the two most distant points in each cluster, finding similar ones) or Ward's method (minimum variance method, aimed to generate compact clusters). The result of the clustering process generated dendrograms which were coupled to heatmaps in order to detect patterns.

- Forest plot: were used to display the results of univariate analysis. They included a summarized overview of the result of the study of different independent variables, all addressing the same question (the difference against a control condition). Each of the obtained result from an individual variable analysis was displayed in the horizontal orientation reporting the name and the result of the studied data. The vertical reference line represented no change against the reference group. The result of each individual difference was represented by a circular point incorporating also a horizontal line indicating the associated standard error. If the represented point was black filled, the observed difference was considered as statistically significant.

3.3.3.3 Multivariate analysis

The main multivariate analysis used in this thesis are described below:

- Principal Component Analysis (PCA): was generally used in order to detect possible intrinsic patterns within the dataset. Thanks to this unsupervised approach it was possible to observe in an easy way the distribution of the data in the Principal Component (PC) spaces. PC1 and PC2 were generated to represent the direction of maximum variance of the data being PC2 orthogonal with respect to the first PC. Thanks to this statistical tool it was possible to

visualize groups of samples or spectra sharing similar characteristics that led them to cluster together in the PC space[14,19].

- Orthogonal partial least-squares discriminate analysis (OPLS-DA): was applied to observe differences between groups of samples or spectra in a supervised way. In this case, with respect to the previously explained PCA, the observed separation is forced, giving previous information of the samples under analysis[20]. However, this approach could overfits the data, necessitating careful validation[21]. To select the number of orthogonal components and validate the predictive capacity a repeated double cross validation process (rdCV) was performed.
- Spearman correlation graph: was used as graphical representation from the calculated multivariate Spearman distances between average profiles of conditions. The difference from the apparently healthy condition was calculated for each profile. The resulting distances were scaled (range 0-1) and represented by a colour code in the graph (created with *igraph* of R package version 1.2.6) connecting adjacent conditions by the increasing number of presented RF.
- Random forest algorithm: was used to build up a binary classification model. This kind of supervised machine learning algorithm are commonly used in order to generate predictive models. In the context of the conduced analysis, first the data were randomly divided into training (75%) and testing (25%) sets. The final model was built from the selected parameters using the entire training set and evaluated on the testing set to estimate its performance on unseen data. This final performance was summarized in receiver operating characteristic (ROC) curves, including their AUCs (Area Under the Curve) with pertaining 95% confidence intervals and cut-off points to maximize the associated specificity and sensitivity parameters[22]. This was necessary in order to understand how much the built model was able to discriminate between the conditions under examination, or in our case, the ability to discriminate the target condition (metabolic syndrome). The higher the AUC value, the more correctly the model was generated for the desired purpose.

3.4 Microalbuminuria analysis

A semi-quantitative analysis using a test strip was performed by the collaborating laboratory of OSARTEN Kooperativa Elkartea to each urine sample for the detection of proteinuria. The obtained results were considered as negative/positive if the value of proteinuria, identified as microalbuminuria as further explained in Chapter 4, was lower/higher than 10 mg/dL.

Moreover, to estimate the glomerular filtration rate (E-GFR) from the available serum creatinine concentrations, the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation was applied, here reported[23]:

$$GFR = 141 \times \min(Ser/\kappa, 1)^{\alpha} \times \max(Ser/\kappa, 1)^{-1.209} \times 0.993^{Age} \times 1.018\ [if\ female]\_1.159\ [if\ black]$$

Where: *Scr* is serum creatinine, κ is 0.7 for females and 0.9 for males, α is -0.329 for females and -0.411 for males, min indicates the minimum of Scr/κ or 1, and max indicates the maximum of Scr/κ or 1.

Bibliography Chapter 3:

1. Masuda, R. *et al.* Integrative Modeling of Plasma Metabolic and Lipoprotein Biomarkers of SARS-CoV-2 Infection in Spanish and Australian COVID-19 Patient Cohorts. *J Proteome Res* **4**, acs.jproteome.1c00458 (2021).
2. Claridge, T. D. W. *High-Resolution NMR Techniques in Organic Chemistry: Third Edition*. (2016).
3. Bruzzone, C. *et al.* SARS-CoV-2 Infection Dysregulates the Metabolomic and Lipidomic Profiles of Serum. *iScience* **23**, (2020).
4. Teipel, J. C. *et al.* Application of 1h nuclear magnetic resonance spectroscopy as spirit drinks screener for quality and authenticity control. *Foods* **9**, (2020).
5. Lachenmeier, D. W. *et al.* Fully automated identification of coffee species and simultaneous quantification of furfuryl alcohol using NMR spectroscopy. *J AOAC Int* **103**, (2020).
6. Mckay, R. T. How the 1D-NOESY suppresses solvent signal in metabonomics NMR spectroscopy: An examination of the pulse sequence components and evolution. *Concepts Magn Reson Part A Bridg Educ Res* **38 A**, 197–220 (2011).
7. Beckonert, O. *et al.* Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat Protoc* **2**, 2692–2703 (2007).
8. Giraudeau, P., Silvestre, V. & Akoka, S. Optimizing water suppression for quantitative NMR-based metabolomics: a tutorial review. *Molecules*, **25** (2020).
9. Zheng, C., Zhang, S., Ragg, S., Raftery, D. & Vitek, O. Identification and quantification of metabolites in 1H NMR spectra by Bayesian model selection. *Bioinformatics* **27**, 1637 (2011).
10. Wishart, D. S. *et al.* HMDB 3.0—The Human Metabolome Database in 2013. *Nucleic Acids Res* **41**, D801–D807 (2013).
11. Emwas, A.-H. *et al.* Recommended strategies for spectral processing and post-processing of 1D 1H-NMR data of biofluids with a particular focus on urine. *Metabolomics 2018 14:3* **14**, 1–23 (2018).
12. Emwas, A. H. *et al.* Recommendations and Standardization of Biomarker Quantification Using NMR-Based Metabolomics with Particular Focus on Urinary Analysis. *J Proteome Res* **15**, 360–373 (2016).
13. Wishart, D. S. Quantitative metabolomics using NMR. *TrAC Trends in Analytical Chemistry* **27**, 228–237 (2008).

14. Zacharias, H. *et al.* Current Experimental, Bioinformatic and Statistical Methods used in NMR Based Metabolomics. *Curr Metabolomics* **1**, 253–268 (2013).

15. Ester M, Kriegel H P, S. J. A density-based algorithm for discovering clusters in large spatial databases with noise, in: Proceedings of Second International Conference on Knowledge Discovery and Data Mining. *Kdd* **96**, (1996).

16. Boddy, R. & Smith, G. *Statistical Methods in Practice: for Scientists and Technologists*. *Statistical Methods in Practice* (2009).

17. Hintze, J. L. & Nelson, R. D. Violin Plots: A Box Plot-Density Trace Synergism. *Am Stat* **52**, 181 (1998).

18. 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)* **57**, 289–300 (1995).

19. Lindon, J. C., Nicholson, J. K., Holmes, E. & Everett, J. R. Metabonomics: Metabolic Processes Studied by NMR Spectroscopy of Biofluids. *Inc. Concepts Magn Reson* **12**, 289 (2000).

20. Worley, B. & Powers, R. PCA as a practical indicator of OPLS-DA model reliability. *Curr Metabolomics* **4**, 97 (2016).

21. Westerhuis, J. A. *et al.* Assessment of PLSDA cross validation. *Metabolomics* **4**, (2008).

22. Florkowski, C. M. Sensitivity, Specificity, Receiver-Operating Characteristic (ROC) Curves and Likelihood Ratios: Communicating the Performance of Diagnostic Tests. *Clin Biochem Rev* **29**, S83 (2008).

23. Levey, A. S. *et al.* A new equation to estimate glomerular filtration rate. *Ann Intern Med* **150**, (2009).

Chapter 4

A molecular discrimination of the metabolic syndrome by urine and serum metabolomics

Results and Discussion

4.1 Metabolic syndrome investigation by urine metabolomics

As explained in Chapter 1, metabolic syndrome, is considered as a multimorbid disease with different risk factors (RF) involved in its onset. Different organizations have tried to give a definition for its diagnostic (**Table 1.1**) but, to date, no consensus has yet been reached on a specific one. Within this context, NMR-based metabolomics of urine samples was used to investigate the characteristic metabolic fingerprint of the metabolic syndrome.

The presented results are partially based on our publication "A molecular signature for the metabolic syndrome by urine metabolomics"[1].

4.1.1 Urine samples cohort description and classification

A large cohort of more than 11,000 urine samples was examined to understand the characteristics aspects of MetS. Individuals that took part in this study were mainly volunteers of the general population and subjects with some of the relevant RFs involved in the development of MetS.

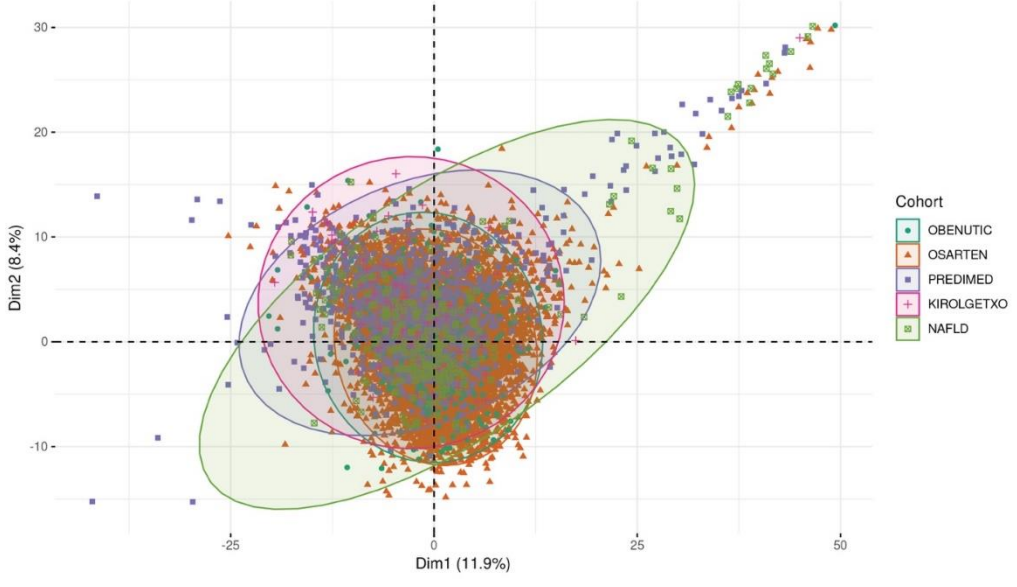

Figure 4.1: PCA of the different cohorts involved in the study. Urine samples from each subject were measured by NMR and spectra were compared.

As a result, urine samples came from various subcohorts (OSARTEN, OBENUTIC, PREDIMED KIROLGETXO and NAFLD) previously described in Chapter 3, from different regions of Spain and from other areas of Europe (NAFLD cohort). Samples were recollected in different areas, always adhering to the strict standard extraction procedures previously described (Chapter 1 and 3), and an unsupervised PCA analysis was performed on the entire urine NMR dataset to confirm that they were comparable (**Figure 4.1**). **Figure 4.1** evidences that the included subcohorts do not present significant differences, making them all comparable into the present study.

In order to investigate the factors contributing to the development of MetS, it was first necessary to classify the obtained samples according to the general characteristics associated with each of the donors. In fact, the various organisations that attempted to give a definition of the metabolic syndrome (Chapter 1, **Table 1.1**) agreed on the importance of four main RFs: alterations of the metabolism of the glucose (which was often summarised in the text only as "diabetes"), obesity, dyslipidemia and hypertension. For this reason, to study the metabolic changes produced by each of each of the RF separately, or the effect caused by the presence of more than one of them in an individual, it was necessary to classify the samples into groups that describe all the possible combinations.

Table 4.1: Risk factors and created conditions for the study of MetS. Abbrevations: IFG, impaired fasting glucose; IGT, impaired glucose tolerance; T2DM, type 2 diabetes; IR, insulin resistance.

Conditions (RF_1, RF_2, RF_3, RF_4)*			
RF_1 Diabetes	**RF_2 Obesity**	**RF_3 Dyslipidemia**	**RF_4 Hypertension**
• fasting plasma glucose > 110 mg/dL • previously diagnosed T2DM, IFG, IGT, IR, or taking medication for hyperglycemia	• BMI > 30 kg/m^2	• triglycerides > 150 mg/dL • HDL cholesterol < 34.75 mg/dL in men or < 38.61 in women • previously diagnosed hypercholesterolemia, hyperlipidemia, hypertriglyceridemia or taking medication for dyslipidemia	• blood pressure ≥ 140/90 mmHg • previously diagnosed hypertension or taking medication for hypertension

0/1 = absence/presence of the risk factor			
0000	apparently healthy	**1001**	diabetes + hypertension
0001	hypertension	**1010**	diabetes + dyslipidemia
0010	dyslipidemia	**1100**	diabetes + obesity
0100	obesity	**0111**	obesity + dyslipidemia + hypertension
1000	diabetes	**1011**	diabetes + dyslipidemia + hypertension
0011	dyslipidemia + hypertension	**1101**	diabetes + obesity + hypertension
0101	obesity + hypertension	**1110**	diabetes + obesity + dyslipidemia
0110	obesity + dyslipidemia	**1111**	diabetes + obesity + dyslipidemia + hypertension

A binary code was used to indicate the presence (1) or absence (0) of each of the four RFs (RF_1, RF_2, RF_3, RF_4) generating 16 different conditions (2^4), as shown in **Table 4.1**.

Following the explained criteria, the 0000 condition represented the apparently healthy subjects, without any associated RF while, for example, the 0100 represented the condition in which subject were presenting only obesity, the 1001 the diabetic and hypertensive individuals with two RFs and the 1101 the diabetic, obese and hypertensive condition with three RFs, and so on. This simple notation clearly describes how each individual participating into the study was classified according to the considered RFs.

Table 4.2 shows the number of samples for each of the 16 conditions from the different subcohorts of urine samples included into the study (OSARTEN, PREDIMED and OBENUTIC) also stratifying them by gender. A statical significant number of samples was reached for each of the condition under consideration, with the lowest number of 62 subjects for the 1110 condition.

Table 4.2: Number of urine samples for each of the conditions under study.

	[ALL valid]	**female**	**male**	**N**
	N=10792	***N=4351***	***N=6441***	
MetS condition:				10792
0000	6925 (64.17%)	2935 (67.46%)	3990 (61.95%)	
0001	692 (6.41%)	276 (6.34%)	416 (6.46%)	
0010	733 (6.79%)	120 (2.76%)	613 (9.52%)	
0011	170 (1.58%)	53 (1.22%)	117 (1.82%)	
0100	504 (4.67%)	232 (5.33%)	272 (4.22%)	
0101	310 (2.87%)	169 (3.88%)	141 (2.19%)	
0110	170 (1.58%)	37 (0.85%)	133 (2.06%)	
0111	148 (1.37%)	62 (1.42%)	86 (1.34%)	
1000	282 (2.61%)	89 (2.05%)	193 (3.00%)	
1001	188 (1.74%)	83 (1.91%)	105 (1.63%)	
1010	84 (0.78%)	18 (0.41%)	66 (1.02%)	
1011	90 (0.83%)	32 (0.74%)	58 (0.90%)	
1100	92 (0.85%)	44 (1.01%)	48 (0.75%)	
1101	202 (1.87%)	111 (2.55%)	91 (1.41%)	
1110	62 (0.57%)	17 (0.39%)	45 (0.70%)	
1111	140 (1.30%)	73 (1.68%)	67 (1.04%)	

4.1.2 Urine ^{1}H NMR spectrum for the study of MetS

Urine metabolites, excreted by the kidney, include the monitorization of waste substances from bloodstream as products of metabolism like amino acids, sugars,

xenobiotics, foods processing derivates and many others, with all of them identifiable and quantifiable by NMR metabolomics[2]. This makes this technique particularly suitable for the metabolic investigation of MetS.

An unsupervised PCA analysis of the mean NMR profiles, obtained for each of the examined conditions and representing their characteristic fingerprinting, shows a separation of the different profiles according to the presented RFs (**Figure 4.2**).

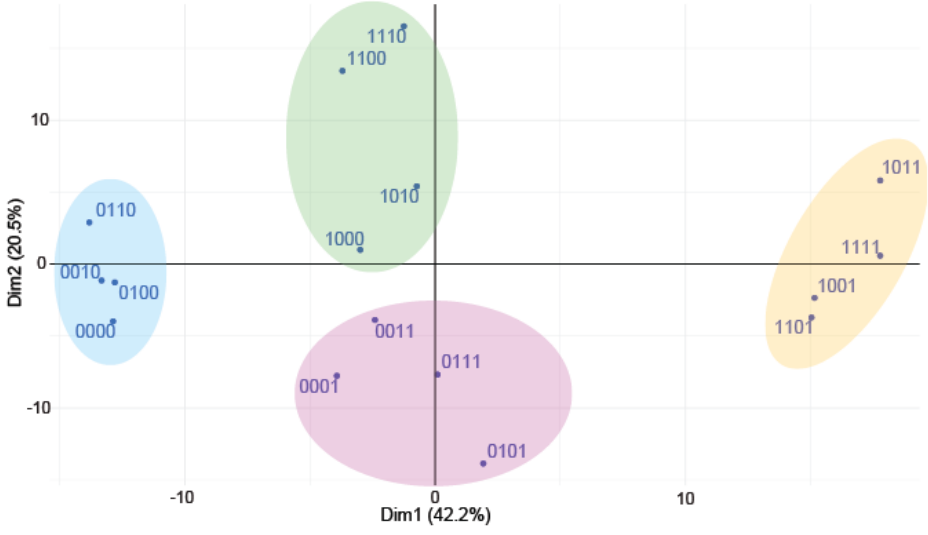

Figure 4.2: Univariate PCA analysis of the mean profile of each of the 16 conditions under study from urine. Coloured ellipses evidenced clustering of subjects with diabetes (green), diabetes and hypertension (yellow), hypertension (pink) and none of the previous conditions (blue).

Figure 4.2 illustrates four different groups in the PCA plot highlighting especially the impact of diabetes and hypertension in this clustering of conditions. Indeed, the green ellipse groups four profiles characterised by the presence of diabetes, the pink one surrounded four conditions affected by hypertension and the yellow highlighted other four profiles presenting both these RFs, while the blue one is characterized by the presence of four conditions without any of the mentioned RFs. This separation highlights the impact of diabetes and hypertension and the sensitivity of the NMR spectra to their presence, in line with previous observations, while, on the other hand,

obesity and dyslipidemia showed a lower degree of modification of the urine spectra[3–5].

4.1.3 Univariate analysis of the urine metabolome

For a better understanding of the observed separations, univariate analysis was conducted to study the differences between each of the profiles with respect to the apparently healthy condition (0000). The heatmap in **Figure 4.3** represents the result of this analysis, using as variables for the comparisons the spectral intensity of the bins from the acquired proton urine spectra. The abscise axis shows the examined conditions while the ordinate one indicates the relevant bins, both sorted according to unsupervised cluster analysis. The represented bins in the heatmap are the ones that came out as statistically significant in at least one of the comparisons of the profiles with the apparently healthy one, and the metabolites associated to them were assigned, looking at the NMR spectra in the corresponding area for each one of the bins.

Table 4.3 shows the contribution of the identified metabolites in the discrimination of the studied condition and its association with some of the examined RFs. As some of them are represented by different bins in the heatmap, their relative abundance was calculated taking into account only the most significant one. Moreover, asterisks inside the squares are used to indicate the attributed p-value obtained from the comparison of each of the variables from the different profiles with respect to the ones of the apparently healthy condition, reflecting the statistical significance of the variation. On the other hand, the colour of the squares is used to indicate the fold change, referring to the bar legend: red squares represent the upregulated bins, while the blue ones are downregulated. Generally, all the bins belonging to the same metabolite show similar fold changes. That said, small magnitude differences in the fold change that can be observed, depending on the heterogeneity of metabolites profile that can be attributed to a given bin. Finally, regarding citric acid, it was observed a contrasting result with its upregulation in the bin at 2.66 ppm and downregulation at 2.57 ppm. This could be explained by the sensitive nature of this metabolite to pH and osmolarity, which were generally associated with changes in the chemical shift of the NMR signals of the metabolites, leading to the involvement of different bins and their relative intensity[6].

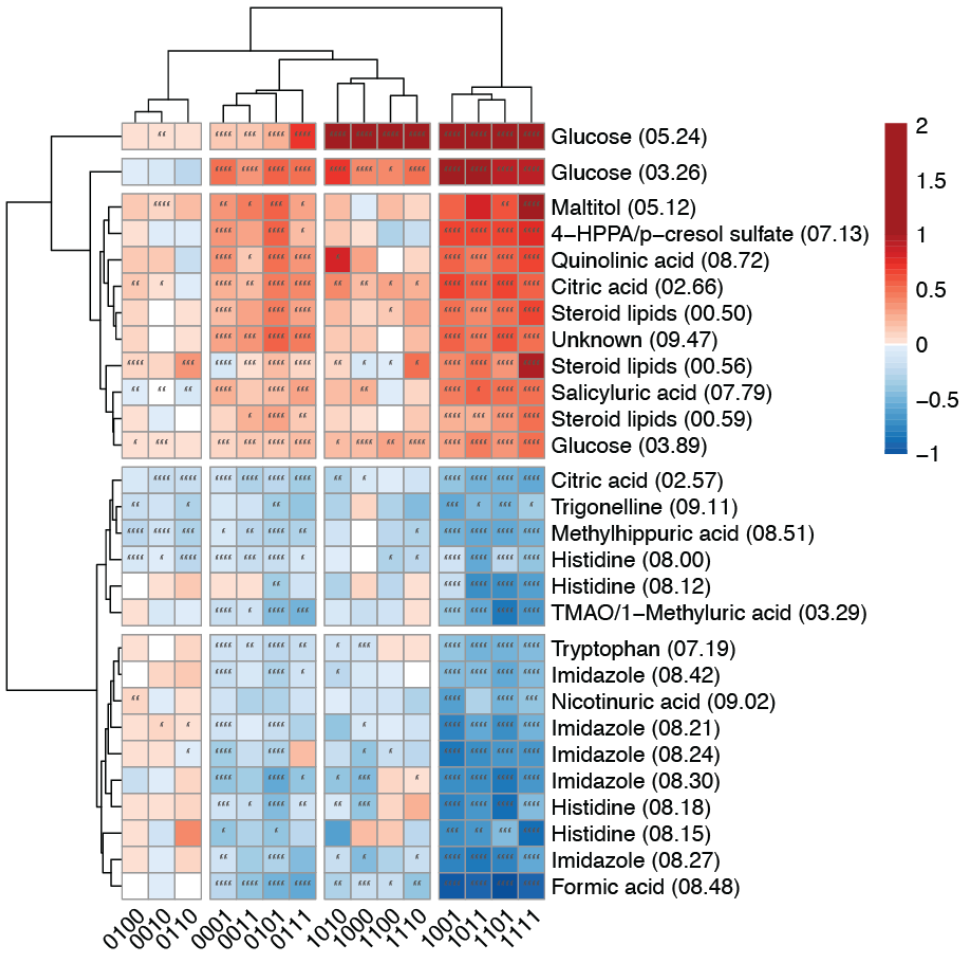

Figure 4.3: Heatmap of the comparison of the different urine conditions with the apparently healthy one (0000). Conditions are represented in the abscise axis while in the ordinate are listed the relevant bins (and the associated assigned metabolites), both sorted according to cluster analysis. Asterisks in the cells indicates the significance of the variation by the p-value (*: p-value < 0.05; **: p-value < 0.01; ***: p-value < 0.001; ****: p-value < 0.0001) while the fold-change is shown by the used colour-code indicated in the bar legend. Dendodrams on the top and on the left evidenced the clustering of the different profiles and bins (metabolites) respectively.

The heatmap of **Figure 4.3** enables extracting several important results. First, it is possible to observe that, once again, as seen in the PCA, the profiles cluster in four different groups. Going from right to left, the first four profiles are characterized by the presence of diabetes and hypertension (1111,1101, 1011, 1001), followed by other four profiles with diabetes (1110,1100, 1000, 1010), and by other four all presenting hypertension (0111, 0101, 0011, 0001), concluding with three profiles characterized by the absence of the so far mentioned RFs (0110, 0010, 0100). The described order is

a result of unsupervised cluster analysis, that sort the different profiles with a tendency that resembles the MetS definition according to WHO, EGIR and AACE, with diabetes as the most relevant (and mandatory) factor for the diagnosis. In fact, it is also possible to observe a clear separation between the diabetic (1XXX) and non-diabetic (0XXX) profiles. Moreover, the variation in the metabolite's concentration is in line with the progression through MetS, with an increased impaired imbalance according to the order of the conditions with greater upregulation/downregulation trending to right, for the profiles with more risk factors.

Table 4.3: Summary of the assigned metabolites and their found association with some of the RFs or MetS.

Metabolite*	Variable importance in the model	log_2FC†	P-value†	Associated RF‡, MetS
Glucose	1056.86	1.66 (1.37, 1.94)	5.88e-97	MetS, RF_1
Formic acid	436.74	-0.79 (-0.87, -0.71)	3.53e-77	n. a.
Steroid lipids	364.47	0.57 (0.3, 0.86)	3.68e-31	RF_3
TMAO/1-Methyluric acid	218.32	-0.54 (-0.7, -0.38)	1.58e-30	RF_2 [7]
Trigonelline	201.66	-0.4 (-0.5, -0.3)	1.38e-06	RF_2 [8]
Tryptophan	198.95	-0.38 (-0.44, -0.31)	1.38e-38	MetS, RF_1 [9]
Quinolinic acid	192.41	0.41 (0.24, 0.59)	1.99e-17	RF_2 [10]
Imidazole	184.05	-0.57 (-0.7, -0.43)	1.20e-26	RF_4 [11]
Histidine	181.71	-0.56 (-0.75, -0.37)	8.42e-16	MetS, $RF_{1,4}$ [11,12]
4-HPPA/p-cresol sulfate	171.38	0.53 (0.4, 0.67)	1.56e-19	RF_1 [13,14]
Salicyluric acid	164.22	0.42 (0.29, 0.56)	8.77e-14	RF_2 [15]
Maltitol	155.43	0.65 (0.45, 0.85)	2.22e-05	RF_1 [13]
Methylhippuric acid	153.23	-0.45 (-0.54, -0.36)	3.79e-21	n. a.
Nicotinuric acid	146.41	-0.38 (-0.5, -0.27)	1.18e-09	MetS, RF_2 [16]

n.a: not applicable.
*For metabolites with more than one associated bin, results with the higher abs(log_2FC) were showed.
†Binary logarithms of fold-changes (log_2FC), their 95% confidence intervals and p-values were calculated between MetS and non-MetS conditions.
‡ The indicated RFs correspond to most frequently associated with the metabolite, and the reported numbers represent at least one of the bibliographic references reporting the metabolite's association with the pertaining RFs or to MetS. More information is provided further.

In **Table 4.3** the found correlations of the evidenced metabolites with the studied RFs and MetS are shown. Most of the identified metabolites are related with the molecular pathophysiology of the studied RFs (**Table 4.3**), while some of them result associated with MetS for the first time in the present study. **Figure 4.4** represents a graphical abstract of the main identified relationships between the RFs and the metabolites.

Glucose was one of the most remarkable metabolites in the heatmap and its increased concentration was unsurprisingly associated with diabetes and insulin resistance[17].

As regard the identified amino acids, previous results revealed the association of branched-chain amino acids (BCAA) and aromatic amino acids (AAA) with MetS[18–20]. These have been related also with obesity, T2DM and insulin resistance (especially obesity-related), which could be responsible of an increase in protein degradation at the muscular level[18,21]. In fact, tryptophan was also inversely associated with plasma adiponectin levels, which in turn is related with an activation of the proteolysis pathways[21–23]. Moreover, several studies have associated tryptophan with MetS and its associated features like problems in glucose metabolism and alteration of the blood pressure[9,10,24–26].

Also, histidine was associated with MetS, insulin resistance, T2DM, and an adverse cardiometabolic risk, so that it was even considered for treatment of T2DM[18,19,27,28]. This metabolite was also related with hypertension, together with altered levels of other amino acids (BCAA and AAA) but unfortunately it was still difficult to understand the underlying mechanism[11,12,29].

Moreover, different studies have been conducted on the trimethylamine N-oxide (TMAO) and its correlation with MetS or the RFs associated with it[7,30,31]. This metabolite, together with trimethylamine (TMA), from which it is synthesised in the liver by gut microbial metabolism starting from choline and L-carnitine, is related with the insurgence of metabolic diseases[32–34]. These results in a dysregulation of glucose metabolism, most often leading to obesity[35,36]. This was further confirmed by its association with BMI and liver fat content[37]. Moreover, the presence of TMAO at urinary and plasma levels, was evidenced as a biomarker for NAFLD, insulin resistance and cardiovascular diseases[38,39].

Trigonelline has also been considered as an obesity-regulator probably by the lipid catabolism control[8]. This metabolite also showed an effect on gut microbiota, especially in the regulation of choline metabolism to produce TMA and TMAO[40].

Among the identified metabolites, salicyluric acid was also already linked with obesity, diet, or other metabolic processes[15,41]. Salicylic and salicyluric acid were frequently found in human urine from salicylates excretion[41,42]. Glycine conjugation pathway, the detoxification mechanism used to eliminate waste substances by facilitating their

solubilisation, was considered as one of the main metabolic routes of salicylic acid, leading to the formation of salicyluric acid[30,43]. In fact, aspirin, hydrolysed to salicylic acid, can be considered the major source for the presence of salicyluric acid in urine, giving rise to this metabolite and others to a lesser extent[44]. Moreover, salicylic acid was found in lower concentration in the serum of obese children, influencing glucose metabolism, in a 1-year nutritional intervention study to improve the cardiometabolic risk of these young patients[15].

p-cresol sulfate, a uremic toxin produced by the intestinal microbes from tyrosine metabolism, has been related with insulin resistance and increased cardiovascular mortality, due to its possible involvement with lipotoxicity phenomena[13,45–47].

The 4-hydroxyphenylpyruvic acid (4-HPPA) was associated with insulin resistance, together with p-cresol sulphate, due to its relationship with tyrosine metabolism from which is synthetized[14].

As regard nicotinuric acid, a previous study proposed this metabolite as potential biomarker for MetS due to its altered levels in the urine of the investigated subjects, correlating this metabolite also with BMI, blood pressure and some lipoproteins (triglycerides, HDL and LDL)[16].

The increase in steroid lipids observed in urine was instead associated with dyslipidaemia and obesity, even if these RFs could be more thoroughly analysed especially by the following serum analysis, which better reflected lipoproteins expression[30,48–50].

When considering quinolinic acid, an increased level of this metabolite was found as positively correlated with higher BMI[10]. Additionally, quinolinic acid was related with tryptophan metabolism, especially with kynurenic acid from which it is produced. Kynurenic acid was in turn associated with high BMI, but also with inflammatory processes, principally in the lungs, liver and kidneys, being even considered a marker for renal insufficiency[51–53]. Due to the known proinflammatory state associated with MetS, and to obesity, this could justify the presence of quinolinic acid in urine, as metabolism product of kynurenic acid[54].

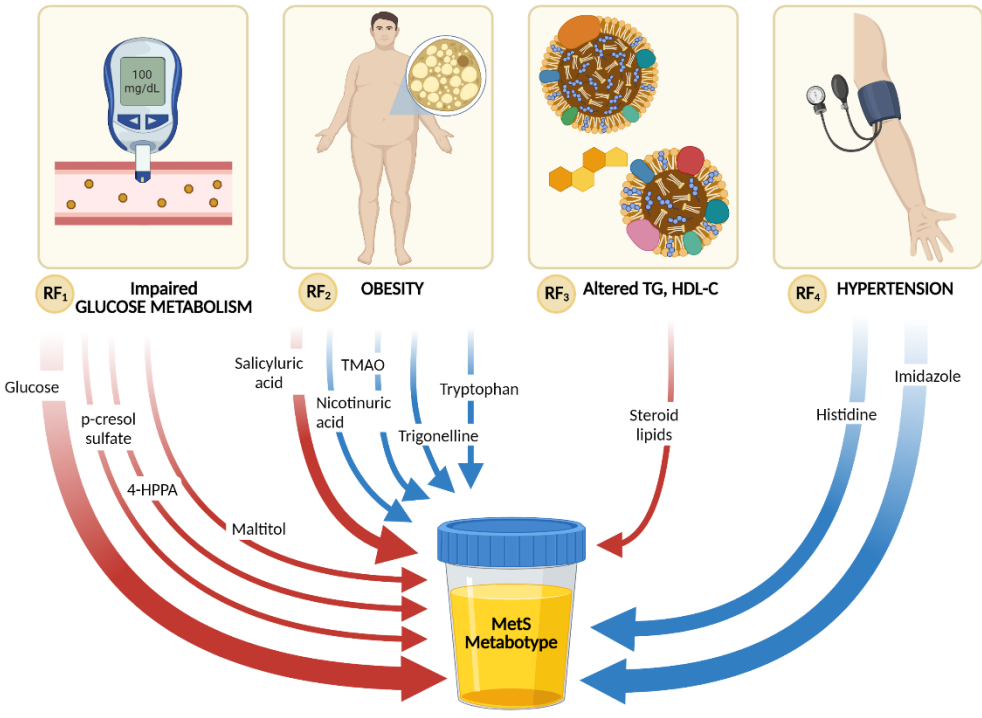

Figure 4.4: Graphical abstract of the association of the investigated risk factors with the identified metabolites. Each RF has at least one metabolite in urine that is altered contributing to MetS metabotype. Metabolites that are associated with more than one RF are represented in the graphic as linked with the one that present the most evident association. Red and blue arrows correspond to up- and down-regulated metabolites in urine respectively. (Created with BioRender.com)

Imidazole has been associated with hypertension for different implied mechanisms. The reduction in the levels of this metabolite can be related with a dysregulation in the concentration of endogenous ligands of the imidazoline and α_2-adrenogenic receptor, which were connected with hypertension episodes[11,55]. In addition, the imidazole moiety was identified as one of the possible competitors for angiotensin binding to its receptor, which was known to be produced by the renin-angiotensin system for the regulation of blood pressure, causing its increase due to the provoked vasoconstriction[56,57]. Finally, the imidazole moiety was evidenced as one of the components of a metabolite, the imidazole propionate, which was associated with overweight problems and especially with T2DM[58]. Indeed, this metabolite, produced from histidine by the gut microbiota, was involved in the impairment of glucose metabolism[59,60].

Maltitol is commonly used as a low-caloric sweetener, often used as substitute in food industry in order to try to avoid the increasing obesity leading to a higher number of people affected by MetS. For this reason, its consumption may be preferable for persons at risk of T2DM and this could justify its presence in the examined urine[61,62]. Despite this, sweeteners interaction with metabolism must be considered by evaluating the entity of its benefits and considering also the role of microbiota in their processing[62,63].

Finally, in the case of methyl hippuric and formic acids, no previous association with MetS had been found. Both of them were described as products of gut microbial metabolism, but their role still needs to be investigated[64]. However, a possible association between formic acid and obesity was found, evidencing a downregulation of this metabolite in obese subjects[65,66].

The proposed explanations for the presence of the altered metabolites are merely hypothetical, but the goal was to attempt to justify the possible reason for their alteration. Nevertheless, urine analysis by NMR proved to be a suitable criterion for the investigation of the metabolic syndrome.

4.1.4 A molecular signature of MetS by urine analysis

To deeply analyse the observed metabolism deterioration in relation with the studied RFs, multivariate analysis was conducted scrutinizing the median spectra profile for each condition calculated by considering the values of the most important bins obtained in the described univariate analysis. A correlation map was built by progressively sorting the 16 studied profiles according to the increasing number of risk factors presented in the various existing combinations, subdividing all conditions in 5 columns, differing by the presence of just one RF (**Figure 4.5**). Conditions were connected by lines and were coloured considering their Superman's correlation distance from the apparently healthy condition (0000). The associated colour legend described the more distant (red shapes) or similar (green shapes) each metabolic condition was to that of the 0000 profile.

Looking at the graph it is possible to confirm the variation of urine metabolome according to the increasing number of presented RFs, but not all of them exert the same influence on MetS progression. In fact, the concomitant manifestation of some RFs in certain conditions reveal a more different and pathogenic metabotype than other combinations. This is the case of the 1111, 1101, 1011 and 1001 profiles, evidenced by the orange ellipse in **Figure 4.5**, that presented a similar molecular signature.

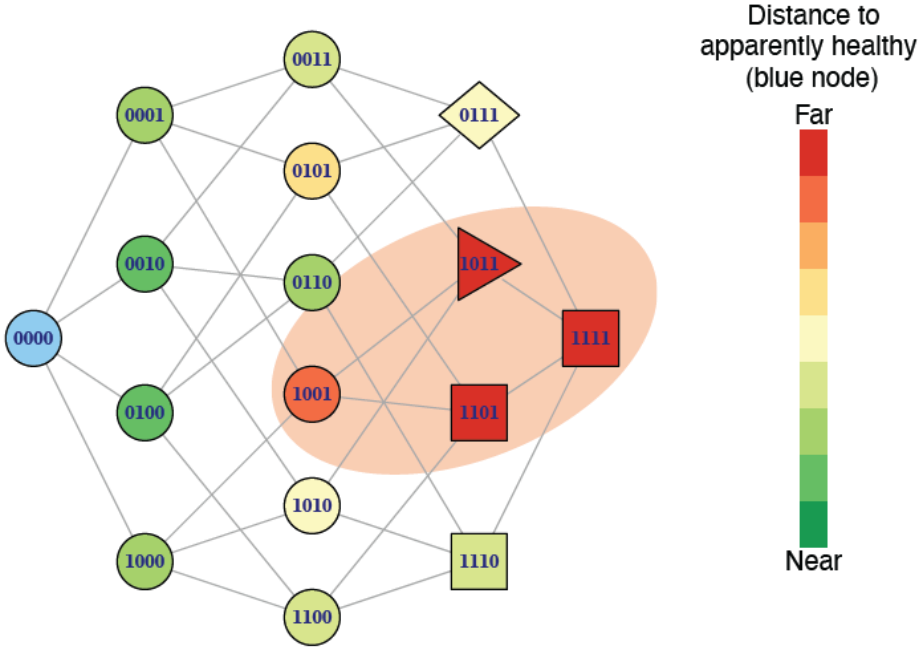

Figure 4.5: Spearman correlation map for urine profiles evidencing the distance of each condition from the apparently healthy one (0000), colour coded according to it, as indicated by the bar legend. MetS conditions according to the WHO, EGIR and AACE are represented by squares and triangles; these last two with the addition of the rhombus represent MetS according to the NCEP:ATPIII and harmonized definitions, while squares and rhombus are the profiles presenting MetS according to the IDF. The orange ellipse grouped the conditions that most resembled the metabolic syndrome according to our metabolic definition.

The triangular, rhombus or squares profiles are the ones considered as affected by MetS according to the described existing definitions (Chapter 1, **Table 1.1**). For instance, conditions like 0111 or 1110 do not present a great difference to the apparently healthy profiles, looking more like intermediate profiles even with fewer RFs. On the other hand, the 1001 condition, presenting just diabetes and hypertension show a significant MetS status, more closely resembling MetS manifestation in its worst scenario.

This characteristic molecular signature led us to suggest a new definition of metabolic syndrome, partially differing from the existing ones, no longer based on the mere presence or absence of given RFs but calculated according to the contribution of each of them in the produced metabolic changes.

4.1.5 A metabolic model for the determination of MetS

Using the large spectra database obtained from the measurement of the urine samples from the subjects included into the study (OSARTEN, PREDIMED and OBENUTIC subcohorts), we tried to build a metabolic model for the determination of MetS, based on the statistically significant bins/metabolites that came out as relevant from the univariate analysis.

First, we determined the number of subjects affected by MetS according to the existing definitions considering a pool of 10,792 individuals. In our analysis, due to the accessible metadata, we were just able to classify patients for MetS according to three groups of organization's definitions with respect to the 5 different ones explained in Chapter 1, **Table 1.1**.

The first criteria for MetS definition grouped the WHO, EGIR and AACE definitions of MetS and included the 1111, 1011, 1101 and 1110 profiles, which were represented as squares and triangles in **Figure 4.5**. Using these definitions, 494 subjects were determined as affected by MetS.

The second criteria included the NCEP:ATPIII and harmonized definitions, represented by the 1111, 1011, 1101, 1110 profiles as before, plus the 0111 condition. This criterion corresponds to the inclusion of the squares, triangles and rhombus profiles in **Figure 4.5** and was populated by 642 cases.

Finally, the third and last criteria was represented by the IDF definition with the 1111, 1101, 1110 and 0111 conditions, with squares and rhombus in **Figure 4.5** and 552 identified subjects with MetS according to this diagnostic.

Three metabolic models were created, considering the existing criteria for the definition of MetS, in order to better distinguish the MetS and non-MetS conditions on a metabolic point of view. The 75% of the cohort, including 8,094 individuals, was

used to train each of the generated models, while the 25% of the cohort, with 2,698 subjects, was used to test it. **Figure 4.6 A-C** showed the generated ROC curves from of each of the created models.

To calculate the ability of the three models to distinguish MetS, the three models were used to screen the cohort of individuals under study. (**Figure 4.6 D-F**). A "MetS probability" value from 0 to 1 was given for each of the subjects, and the obtained scores were represented in a smoothed histogram.

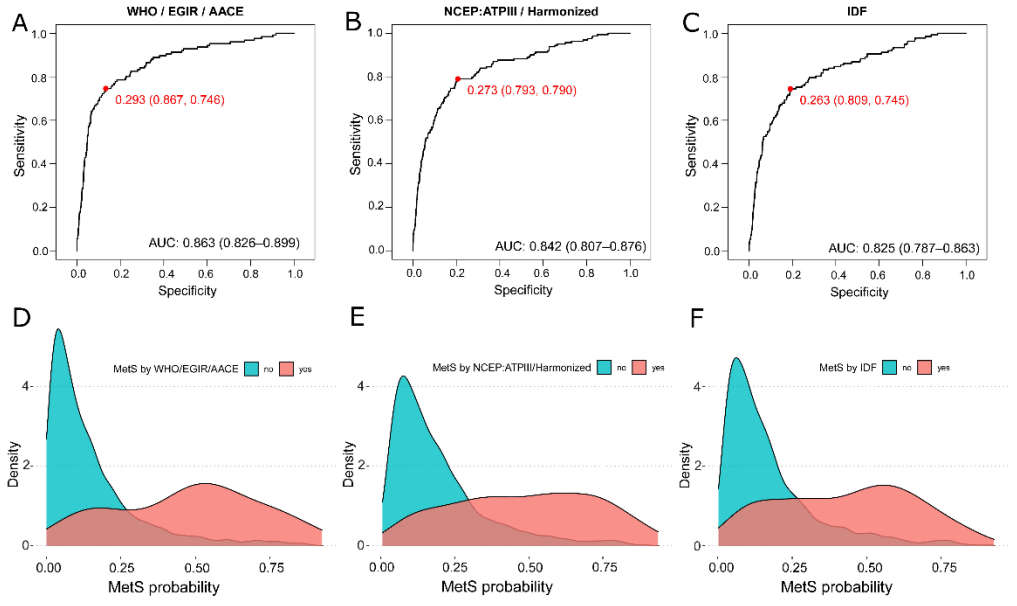

Figure 4.6: Metabolic syndrome models generated thanks to the measured spectra database. **A-C**) ROC curves for the three groups of definitions under consideration: **A**) WHO, EGIR and AACE; **B**) NCEP:ATPIII and Harmonized; **C**) IDF. **D-F**) Smoothed histograms (Kernel density based) evidencing the probability distribution for the application of the built MetS models to the full cohort, for the three groups of definitions: **D**) WHO, EGIR and AACE; **E**) NCEP:ATPIII and Harmonized; **F**) IDF. Subjects affected by MetS are represented in red, while the ones without MetS are blue.

Looking at **Figure 4.6 D-F**, it is possible to observe that subjects without MetS, with low scores values, cluster together for each of the tested models, whereas the ones with MetS have a wider distribution, probably justified by the heterogeneous nature of this syndrome, but with a clear tendency towards higher values of the score.

Remarkably, the three built-in models, merely based on the metabolomic analysis of the measured urine samples, proved to be able to distinguish subjects with metabolic syndrome with AUROC values between 0.83 and 0.86. The observed divergency may be related to the existing differences between the created definition for MetS, which considers the molecular dimension of this disorder, with respect to the existing standard ones. In fact, by looking at the models, the first one, including the problems with glucose metabolism as a mandatory factor, shows a better performance (AUROC value: 0.863), reflecting the importance of glucose in urine for this metabolic model, also previously evidenced. Indeed, rising the values of hyperglycemia to 110 mg/dL, preferred by some existing definition, instead of the 100 mg/dL, considered for the conduced analysis, the AUROC values increased to 0.86-0.92 (**Figure A1** in the Appendix), with better performance results with respect to previous study[67].

Our definition of metabolic syndrome adds a metabolic dimension to the existing ones. Some of the currently used definitions (WHO, EGIR and AACE) could be considered as more appropriate for diagnosis because of the relevance attributed to the RFs that are actually responsible of the most important metabolic changes. In fact, diabetes undoubtedly shows to be a very relevant factor in this syndrome, as evidenced by all the results proposed, but a risk factor that in urine also proved to be key in the aggravation of this disorder was hypertension.

4.1.6 Potential caveats and limitations

4.1.6.1 Effect of gender in MetS determination

Gender differences emerge as a possible limitation in our study. In fact, previous studies observed differences in the manifestation of MetS according to gender, mainly due to the different probability of presenting some of the RF involved in the onset of this disorder, such as visceral obesity[68,69].

To investigate this aspect, we repeated the univariate analysis of the comparison of all the different conditions against the apparently healthy one, but this time segregating samples by gender. **Figure 4.7** shows the resulting heatmaps obtained from this analysis. The heatmaps dived for gender render the same results as the one obtained

for the analysis of the full cohort, proving that gender differences do not present any effect on the results for the metabolomic discrimination of MetS.

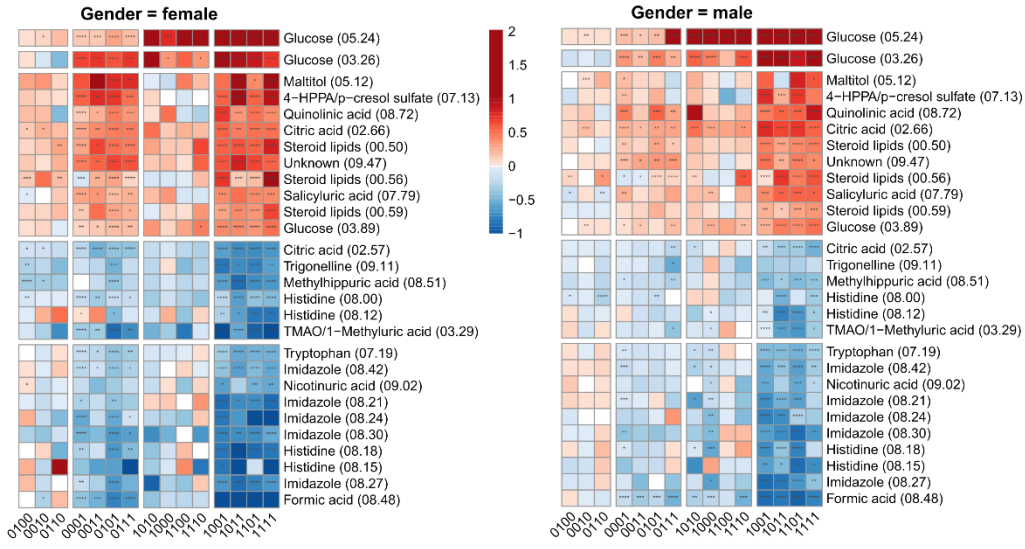

Figure 4.7: Heatmaps divided by gender: as indicated, women heatmap on the left and men heatmap on the right. Conditions (in the abscise axis) and bins/metabolites (in the ordinate axis) are sorted according to cluster analysis. Fold change is colour-coded according to the bar legend. The statistical significance of the variation from the apparently healthy profile was calculated for each condition and determined by the p-value, shown inside the squares.

4.1.6.2 Effect of age in MetS determination

Aging could also emerge as another possible limitation for the conduced study, due to its known influence in many diseases, including MetS[70]. In fact, while the OSARTEN and OBENUTIC cohorts were balanced in terms of age, the PREDIMED cohort included older subjects on average. This could cause errors in the carried-out study, leading to a bias in the results, partially reporting on the ageing process. In order to investigate this aspect, an independent cohort called KIRLOGETXO, not previously used for the design of the created models, was analysed. This cohort, previously described in Chapter 3, includes older individuals between 60 and 85 years of age with a healthy lifestyle and regular physical activity (more information in **Table A5** in the Appendix). Precisely because of this, these subjects had no risk factors (n = 34) or only one (n = 40).

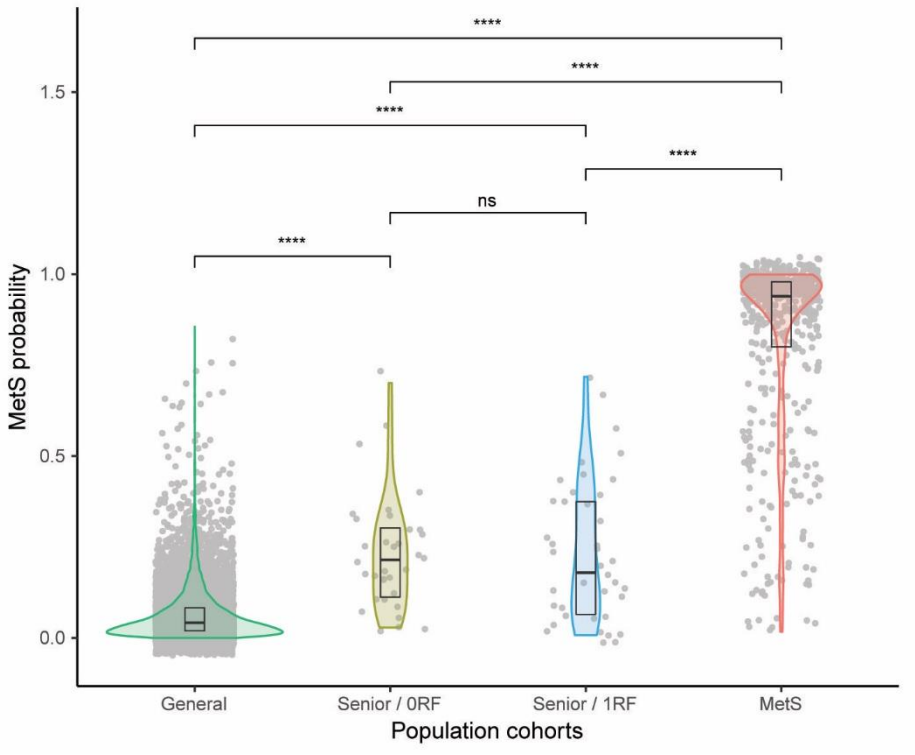

Figure 4.8: KIROLGETXO cohort screening with the MetS model. The probability distributions for suffering for MetS were calculated for: the general population (0000 individuals represented in green), senior population with no risk factors (yellow), senior population with 1 RF (blue) and population with MetS (red).

Figure 4.8 shows the result of the application of the created MetS model with the best performance to the screening of the KIROLGETXO cohort. It was possible to observe that only a very limited number of individuals could be considered as affected by MetS, debunking the possibility of the hypothesised influence of age on the generated metabolic model.

4.1.7 Investigation of the microalbuminuria and impaired renal function association with MetS

According to the WHO definition of metabolic syndrome (described in Chapter 1, **Table 1.1**), microalbuminuria has to be considered as an additional RF, that must be evaluated for diagnosis. This is related to the association of microalbuminuria with kidney dysfunction, chronic kidney disease (CKD), cardiovascular disorders and

especially due to its frequent association with T2DM or, more in general, with metabolic disorders[71].

To investigate the role of this factor, proteinuria values (> 10 mg/dL) were obtained for the OSARTEN cohort, by the collaborating laboratory of OSARTEN Kooperativa Elkartea, by a semi-quantitative analysis using test strips. This cohort represented one of the largest in terms of sample size, although particularly characterised by healthy individuals, but still with enough subjects to populate all the conditions under investigation.

Urine proteins composition is quite heterogenous, but albumin was considered as the predominant one and dip-stick tests evidenced to be sensitive for it[72,73]. Because of this, microalbuminuria and proteinuria were assimilated as equal.

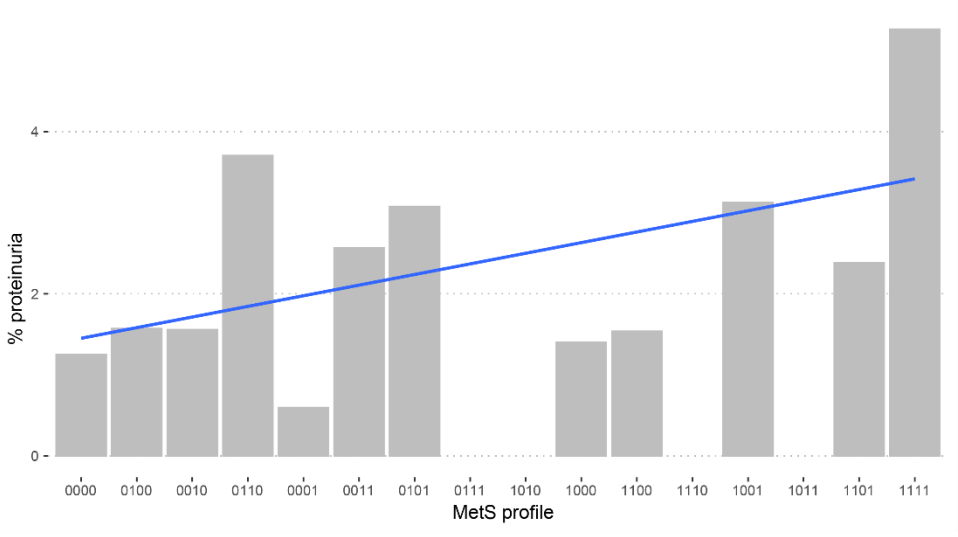

Figure 4.9: Percentage of samples with albuminuria are represented by bar plot for each of the conditions in the OSARTEN II cohort. Conditions are sorted according to heatmap order of Figure 4.3. Blue line evidence the linear regression of the values assuming equidistance in the conditions towards MetS.

Figure 4.9 shows an increase in the microalbuminuria percentage for the condition with the most altered number of risk factors (1111), representing the full MetS profile. Microalbuminuria may therefore be associated with metabolic syndrome, as suggested by the WHO, but its presence shall be mainly considered as a side effect for its association with other involved RFs in the onset of MetS, like, for example, with

hypertension[74]. That said, the highest percentage of samples presenting microalbuminuria was equivalent to 10% for the 1111 condition, highlighting the secondary importance of this factor. As the excessive elimination of albumin in urine reflected kidney damage with reduced glomerular filtration, to estimate the glomerular filtration rate, the Chronic Kidney Disease Epidemiology Collaboration equation was used (reported in Chapter 3)[75]. This was calculated for the OSARTEN and OBENUTIC cohorts, for which serum creatinine concentration value was accessible.

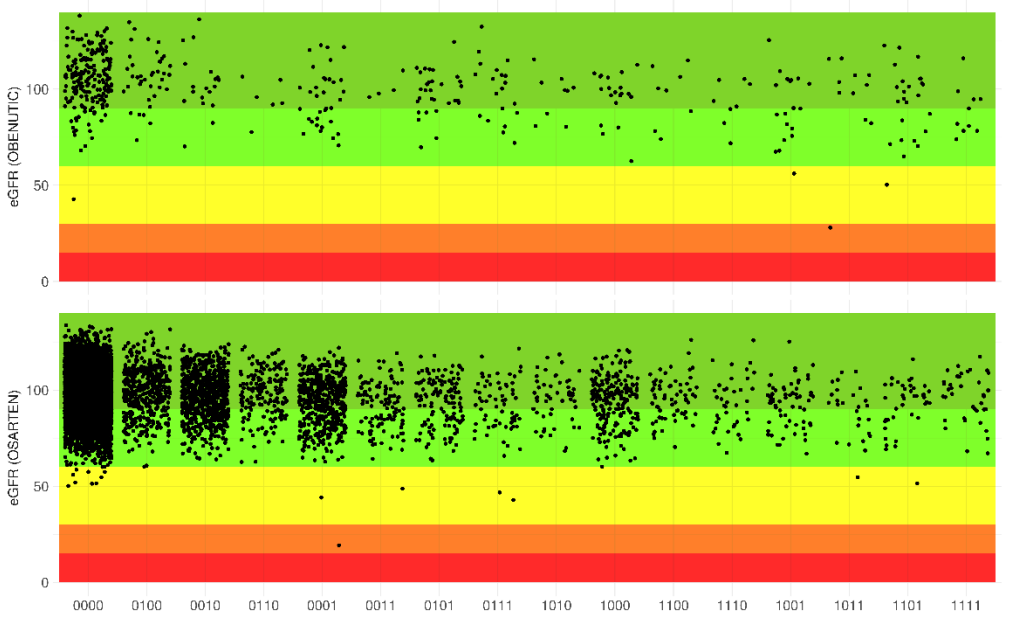

Figure 4.10: E-GFR distribution values for the OSARTEN II and OBENUTIC cohorts, as a function of the MetS conditions. E-GFR values were colour coded according to the categories (G1 to G5): G1, in dark green, represents normal or high E-GFR; G2, in light green, shows the mildly decreased E-GFR; G3, in yellow, evidences moderate to severely decreased E-GFR; G4, in orange, the severely decreased E-GFR and finally G5, in red, the kidney failure.

Figure 4.10 shows the distribution of the E-GFR values for the two analysed cohorts. A scale of five range values was created, from G1 to G5 to indicate the rate of glomerular filtration of each sample according to its MetS condition, where G1 represented the normal or high GFR and G5 the worst scenario with kidney failure. Most of the subjects analysed (75%) were included in G1 area, the 24.8% was in G2 category with a mildly decreased GFR and, the remaining small percentage of subjects, fell in G3 area. Almost no subjects presented a severe decreased GFR in G4 scale and no one evidenced kidney failure (G5).

The presented results thus show that the reported metabolic changes are uncoupled to a putative renal function impairment from the investigated subjects.

4.1.8 The relationship between NAFLD and MetS

Previous studies already wondered whether NAFDL is the hepatic manifestation of metabolic syndrome or not[76,77]. This could be explained by the co-involvement of MetS risk factors in the contribution to NAFLD progression.

In order to better understand this aspect, a cohort of patients with NAFLD, including 234 urine samples, was analysed (more information in **Table A5** in the Appendix). The studied subjects were diagnosed by means of a previous biopsy and classified according to reference methods for their charachterization[78].

Thanks to the MetS criteria including the WHO, EGIR and AACE definitions, the examined samples were divided in two main subcohorts: NAFLD subjects with MetS (*NAFLD with MetS*) or without MetS (*NAFLD without MetS*). These subcohorts were then screened with the created metabolic model with the best performance for de detection of MetS, to determine the percentage of subjects affected by it according to our criteria.

Figure 4.11 shows the MetS probability distribution for the investigated subcohorts, for the general population, mainly characterized by healthy subjects (0000), and MetS population, whose NAFLD status is unknown. By looking at the probability distribution of subjects with NAFLD but without MetS, it is possible to see that, on average, these have a low probability for MetS according to our model, resembling more the general population distribution profile and suggesting that the metabotype associated with NAFLD is different from that for MetS. On the other hand, *NAFLD with MetS* subjects have a more disperse and complex probability distribution, reinforcing our previous hypothesis that RFs involved in metabolic syndrome and those for the progression of NAFLD lead to a similar metabotype.

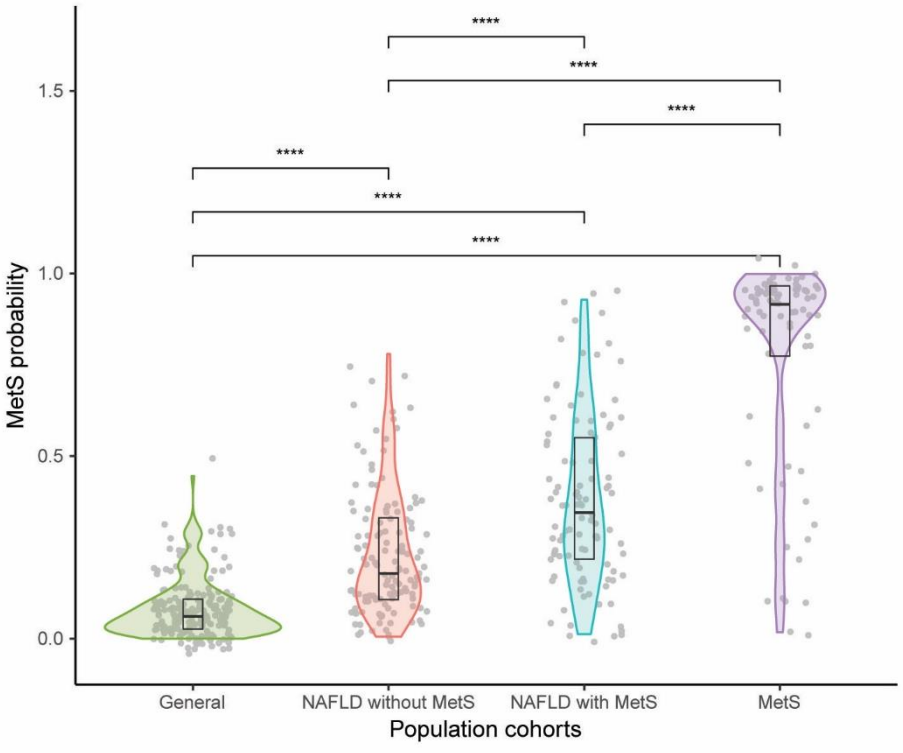

Figure 4.11: NAFLD cohort screening with the built MetS model. Probability distributions for suffering for MetS were calculated for: the general population (0000 individuals represented in green); *NAFLD without MetS*, according to WHO definition, in red; *NAFLD with MetS*, according to WHO definition, in blue and MetS population in purple.

Based on the obtained results, it could be suggested that MetS and NAFLD share some similar features and symptomatology but are characterised by a different metabolic profile. Despite this, due to the overlap between these two conditions, further investigations should be performed for a better understanding of this relationship.

4.2 Metabolic syndrome investigation by serum metabolomics

Urine analysis may be biased by its sensitivity to certain RFs (i. e. diabetes) as compared to other ones (i. e. dyslipidaemia). To that end, a better understanding of the characteristic metabolic aspects of MetS may benefit from serum metabolomics as well.

4.2.1 Serum samples cohort description and classification

A cohort of almost 8,500 serum samples was examined to further investigate the molecular aspects of MetS. The subjects included into this analysis came from three different subchorts: OSARTEN, MetS-long and PORTUGAL (described in Chapter 3). OSARTEN cohort included the serum samples from the same individuals involved in urine analysis. These were volunteers of the general working population, including mainly healthy subjects and individuals with some of the relevant RFs associated to the development of MetS. For the OBENUTIC and PREDIMED subcohorts, serum samples were not available and it was necessary to expand the dataset with more serum samples from individuals that were presenting some of the relevant RFs involved in the onset and progression of MetS. The MetS-long subcohort fulfilled these criteria, including serum samples from individuals selected precisely for this purpose. On the other hand, the PORTUGAL cohort includes older subjects from Portugal, and adds a more homogeneous age distribution in the sample's cohort for the performed study.

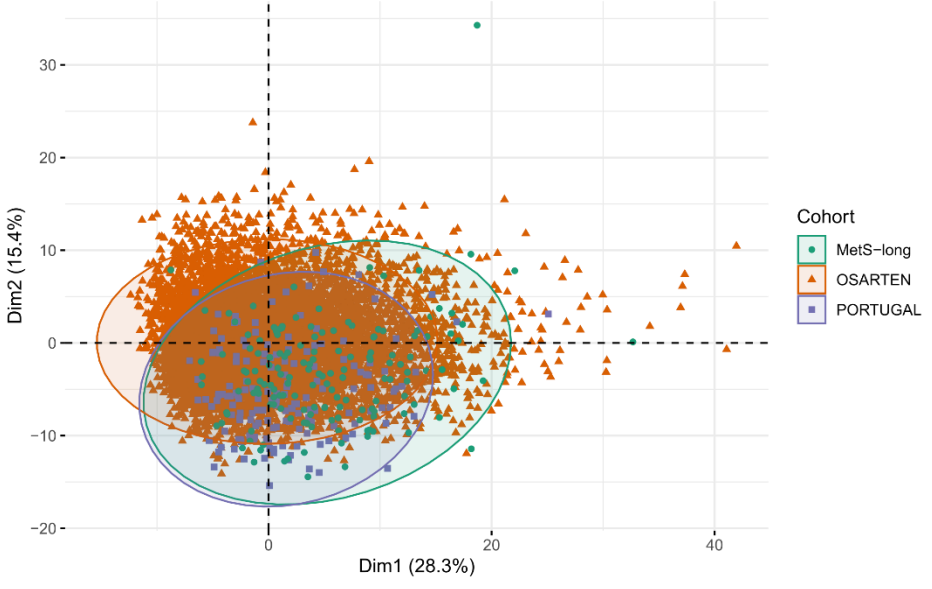

Figure 4.12: PCA of the different cohorts involved in the study. Serum samples from each subject were measured by NMR and spectra were compared.

Considering that the mentioned samples subcohorts are from different areas of Europe, an unsupervised PCA analysis was conducted to the acquired serum NMR spectra dataset, to ascertain that the cohorts are comparable (**Figure 4.12**).

Figure 4.12 shows that the measured serum samples from the subcohorts included into the analysis do not show significant differences, but only a slight shift possibly explained by the very nature of the included cohorts, as the MetS long one was enriched by subjects with the highest number of risk factors and the Portuguese cohort includes older subjects, but still making them all comparable for the present study.

In order to proceed with serum analysis, first it was necessary to classify the samples into 16 different profiles according to the criterion previously used for urine examination. **Table 4.4** shows the allocation of samples into the created conditions, also separated by gender, according to the established criteria previously explained in **Table 4.1**. Similar to urine, the condition with the lowest number of samples is the 1110 with 48 subjects, yet still adequate for the analysis.

Table 4.4: Number of serum samples for each of the conditions under study.

	[ALL]	**female**	**male**	**N**
	N=8470	***N=3116***	***N=5354***	
MetS conditions:				8470
0000	5564 (65.69%)	2313 (74.23%)	3251 (60.72%)	
0001	443 (5.23%)	135 (4.33%)	308 (5.75%)	
0010	663 (7.83%)	111 (3.56%)	552 (10.31%)	
0011	156 (1.84%)	40 (1.28%)	116 (2.17%)	
0100	396 (4.68%)	163 (5.23%)	233 (4.35%)	
0101	130 (1.53%)	42 (1.35%)	88 (1.64%)	
0110	168 (1.98%)	27 (0.87%)	141 (2.63%)	
0111	74 (0.87%)	21 (0.67%)	53 (0.99%)	
1000	272 (3.21%)	79 (2.54%)	193 (3.60%)	
1001	128 (1.51%)	37 (1.19%)	91 (1.70%)	
1010	83 (0.98%)	9 (0.29%)	74 (1.38%)	
1011	81 (0.96%)	38 (1.22%)	43 (0.80%)	
1100	85 (1.00%)	32 (1.03%)	53 (0.99%)	
1101	75 (0.89%)	24 (0.77%)	51 (0.95%)	
1110	48 (0.57%)	10 (0.32%)	38 (0.71%)	
1111	104 (1.23%)	35 (1.12%)	69 (1.29%)	

4.2.2 Serum ^{1}H NMR spectrum for the study of MetS

Serum samples analysis was conducted in a similar way to the urine study, to investigate the metabolites and lipoproteins that characterize the subjects affected by MetS or that presented some of the RFs involved in its onset. In this case, two different

types of one-dimensional experiments were measured for each sample, the NOESY, as for urine samples, and a CPMG, filtering out macromolecules that characterize serum samples, generally used for a better analysis of the metabolites present in this biofluid (for more experimental details see Chapter 3).

Unsupervised PCA analysis of the mean NMR profile, obtained for each of the studied conditions and representing the characteristic fingerprinting of metabolites and lipoproteins expression, was conducted to observe the distribution of the different profiles according to the presented RFs. **Figure 4.13** shows a clear separation in the PCA space between the dyslipidemic (pink and blue ellipse) and not dyslipidemic subjects (green and yellow ellipse): patients with dyslipidaemia, present a lipoprotein imbalance in blood as compared to those without lipoprotein alterations, that results in a strong impact on the NMR profile of the measured samples.

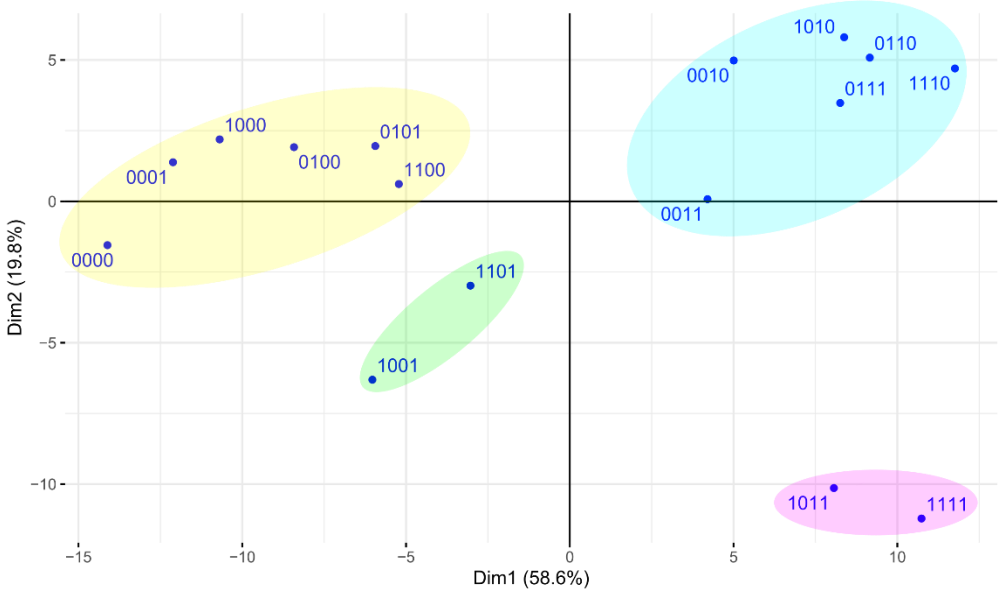

Figure 4.13: Univariate PCA analysis of the mean profile of each of the 16 conditions under study from serum. Dyslipidemia caused a clear separation of the profiles in the PCA space: pink and blue ellipse are grouping dyslipidemic subjects while green and yellow ellipse are evidencing individuals without this RF.

In addition to dyslipidaemia, some RFs, or the combination of some of them, still seem to have some effect on the separation of the profiles. In fact, pink ellipse groups two profiles (1011 and 1111), both considered as affected by MetS and characterized by

the presence of diabetes and hypertension, previously evidenced as relevant aspects in the manifestation of this disorder according to our urine analysis. This is also observed in green ellipse, that evidences two profiles again characterized by the presence of these two RFs (1001, 1101) but not exhibiting dyslipidaemia and, therefore, they show up far apart in the PCA space. On the other hand, the blue ellipse shows a group of profiles all characterized by the presence of dyslipidemia. While, on the other side the yellow ellipse highlights subjects without this RF, as suggested before.

4.2.2 Serum metabolome univariate analysis

To investigate further the characteristic metabolic profile of the serum samples, univariate analysis was conducted in order to compare all the MetS conditions with respect to the asymptomatic one (0000). In this case, metabolites concentrations obtained after every serum measurement, were used as variables.

Figure 4.14 shows the heatmap obtained as result of this analysis. Abscise axis shows the examined serum conditions while in the ordinate axis were listed all the quantified metabolites, both sorted according to unsupervised cluster analysis.

Coloured squares indicate the fold change calculated form the comparison of each of the variables with the apparently healthy condition, referencing to the associated bar legend: red squares represent the upregulated metabolites while the blue ones correspond to the downregulated.

As for urine univariate analysis (**Figure 4.3**), it is possible to observe a clear separation between the diabetic conditions (1XXX, on the right) and the non-diabetic ones (0XXX), evidencing the impact of this RF in the profile's discrimination and the importance of this health condition in the produced metabolic changes.

Univariate serum analysis (**Figure 4.14**) showed its influence mainly through the elevated detected glucose concentration in the bloodstream (diabetes, also in the excreted urine), highlighting also the relevance of other two RFs, obesity and dyslipidaemia, due to the important metabolic alteration observed in relation to the presence of these two RFs in the analysed MetS conditions.

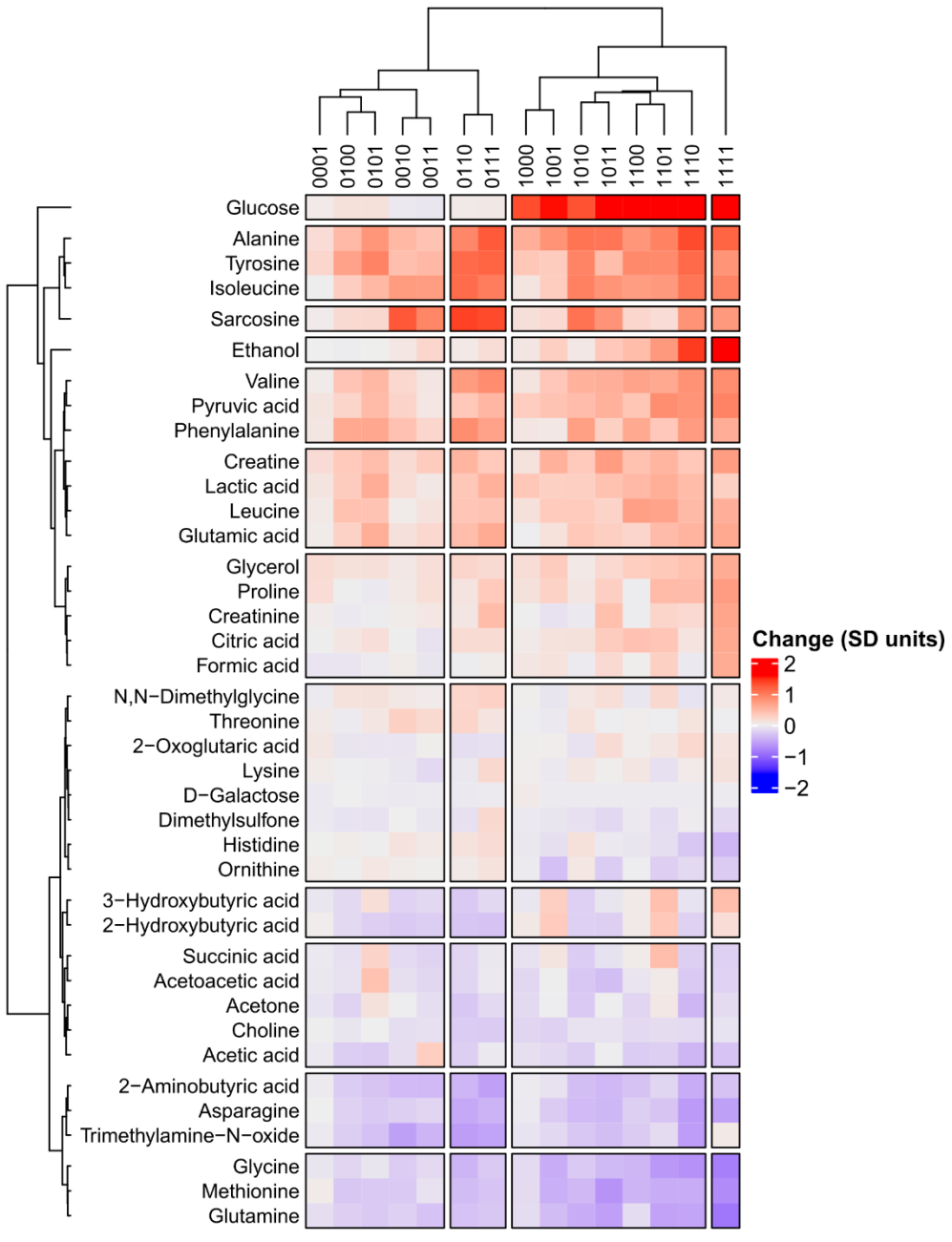

Figure 4.14: Heatmap of the comparison of the different serum conditions with the apparently healthy one (0000). Conditions are represented in the abscise axis while in the ordinate are listed the metabolites, both sorted according to cluster analysis. Fold-change for each of the comparison is evidenced by the colour-code indicated in the bar legend. Dendodrams on the top and on the left evidence the clustering of the different profiles and metabolites respectively.

Moreover, as previously shown in the result of urine univariate analysis, the variation in metabolite's concentration was in line with the progression through MetS. In fact,

the 1111, 1110, 1101 and 1011 conditions, the only ones considered as affected by MetS according to most of the already existing definitions and by our urine analysis, were all clustered together and in an ordered sequence according to the unsupervised cluster analysis and all presented important variations at the metabolic level. Condition 0111, still considered as MetS by some of the existing definitions (NCEP:ATPIII, IDF and harmonized), felt in a slightly more distant area of the heatmap, with condition 0110, not MetS, but still presented significant alterations in some metabolites concentration, probably due to the concomitant presence of obesity and especially dyslipidaemia.

Many of the altered metabolites were previously found as associated with MetS. BCAA, including leucine, isoleucine and valine, and AAA, consisting in tyrosine and phenylalanine, are upregulated, especially in the serum of the subjects affected by MetS, like the 1111,1110, 1101 and 1011 profiles. Moreover, important changes in these metabolite's concentrations are also observed in the 0110 and 0111 profiles or, in general, more prominently in profiles characterised by the presence of diabetes and especially in those with obesity and dyslipidaemia. In fact, as mentioned in the discussion of urine analysis, BCAA and AAA were reported to be upregulated in the serum of the subjects affected by MetS and in the ones presenting diabetes and obesity[79,80]. Their higher circulating levels are associated with a dysfunction in carbohydrates metabolism. Moreover, they are related with an increased stimulation of gluconeogenesis[28,30].

Another metabolite significantly upregulated in serum analysis is alanine, also considered as highly gluconeogenic and lending to glucose intolerance in obesity[81]. In fact, alanine, together with pyruvic acid, also partially raised, was previously found as increased in obese individuals[28]. It has been hypothesised that this metabolite could be produced by pyruvate transamination into alanine through glutamate stimulation and this process could be influenced by the described increase in BCAA catabolic flux[28,30,81]. Moreover, pyruvic acid, traditionally produced by glycolysis, in addition to alanine production, could follow different paths, including the entry in the Krebs cycle, whose intermediate metabolites were also dysregulated in the conduced serum

univariate analysis, like citric and succinic acid, evidencing a possible disfunction in this pathway.

Glutamate (or glutamic acid) and glutamine concentration were altered according to serum analysis. Both these metabolites and the glutamine/glutamate ratio have been previously suggested as potential biomarker for T2DM. Moreover, increased glutamate concentration is related with visceral obesity and in the development of MetS, being obesity particularly involved in glutamate metabolism[82,83]. In general, both these amino acids are considered as crucial in human metabolism and its regulation due to their involvement in multiple functions. In relation to the observed metabolic changes in the context of our analysis, it is important to also consider the role of glutamine, which appears to be downregulated in the synthesis of alanine through the pathway involving glutamic and pyruvic acid production, potentially explaining the increase of these metabolites[84]. Moreover, the elevated levels of glutamate could be a product of BCAA catabolism, which is involved in glucagon release by the pancreatic α cells and promoting the described transamination of pyruvate to alanine[81].

Altered concentration of several other metabolites suggest an alteration of the one-carbon metabolism pathway, already observed in other metabolic diseases and in obesity[85]. Glycine and methionine downregulation, especially in the 1111 MetS profile, underline a dysregulation in the methyl-donors' metabolites. Glycine deficiency could lead to important health problems and it has been associated with IR[86]. Moreover, this metabolite is involved in different pathways for the biosynthesis of other small molecules like glutathione or creatine. Alternatively, methionine concentration is involved in the conversion of SAM (S-adenosylmethionine) to SAH (S-adenosylhomocysteine), which is in turn related with GNMT (glycine-N-methyltransferase) metabolic mechanism[86,87]. The latter has been shown as responsible of the sarcosine production, highly increased especially in MetS profiles, particularly the ones characterized by the presence of obesity and/or dyslipidemia[88]. Finally, other metabolites like formic acid, histidine or ornithine have been previously associated with the regulation of folate metabolism pathway or with glycine catabolism[86,89].

An alteration in the ketone bodies like, 3-hydroxybutiric acid, acetone and acetoacetic acid has also been observed, especially in the case of 3-hydroxyburic acid, whose highest levels were observed in the 1111 condition (and to a lesser extent in the 1101, 1001 and 0101 profiles). The increase of this metabolite was previously found as associated with hyperinsulinemia and T2DM[90,91]. Moreover, ketone bodies, produced from fatty acids in the liver, were generally considered as important regulators of the metabolism as associated with β-oxidation of fatty acids, tricarboxylic acid cycle (TCA), lipogenesis and glucose metabolism[92].

Finally, the dysregulation in some metabolite's concentration like ethanol, lactic acid, acetate and TMAO could be associated with an alteration in the gut microbiota. Ethanol increased levels, especially in the 1111 and 1110 MetS profiles could derive from microbiome activity, as evidenced in some cases of liver diseases (NAFLD), that previously reported an increase in ethanol concentration because of *Lactobacillus* or *Klebsiella pneumoniae*, leading to a degenerative progression of the disease[93,94].

As regard lactic acid, this metabolite could be produced from glucose metabolism from pyruvic acid, or it could originate form bacteria like *Lactobacillus*[95]. Its altered concentration was previously associated with carbohydrate metabolism dysfunction, MetS, IR, obesity and diabetes[96,97]. Moreover, acetate production was also associated with microbial fermentation and is considered as involved in human energy regulation through cholesterol synthesis, lipogenesis and adipocyte's accumulation[98]. Altered levels of this metabolites were also found as associated with increased in ethanol levels[99]. Lastly, as further evidence of microbial influence, TMAO highest levels are seen in the 1111 condition. In fact, its production, as mentioned before, was based on the conversion of choline into TMA, which was in turn converted into TMAO in the liver by gut microbiome.

The described mechanisms could be considered as responsible of the observed metabolic changes in the studied serum conditions. As in the case of urine analysis, many of the significantly altered metabolites were previously associated at least to one of the investigated risk factors or with MetS, highlighting the sensitivity of ^{1}H NMR spectra for the study of this disorder.

4.2.3 Serum lipidome univariate analysis

Of particular interest was the study of lipoproteins, particularly for a better characterization of dyslipidaemia, which emerged as a discriminating RF in serum profiles distribution in the PCA analysis (**Figure 4.13**). So far, it was not possible to investigate thoroughly this RF because of the lower concentration of lipoproteins in urine, generally found in the bloodstream.

Univariate analysis was conducted to compare the created MetS conditions (according to the established criteria explained in **Table 4.1**) with the apparently healthy one (0000) but, in this case, the employed variables were the quantified lipoproteins obtained from the Bruker IVDr B.I.LISA™ report (more information in Chapter 3).

Figure 4.15 shows the result of this analysis, including an extensive number of lipoproteins, lipoproteins subfractions and parameters (up to 112) determined thanks to the regression model applied by Bruker as method for the quantification of the mentioned variables. In addition to the quantification of the main lipoprotein classes, including the VLDL, IDL, LDL and the HDL, and their related subfractions VLDL from 1 to 5, LDL form 1 to 6 and HDL from 1 to 4, whose numbers correspond to the increased density and decreased size (**Figure 4.16**). Interestingly, the obtained reports, allow the quantification of the amount of each of the lipoprotein components, triglycerides, cholesterol, free cholesterol, phospholipids, apolipoprotein B (Apo-B), apolipoprotein A1 (Apo-A) and apolipoprotein A2 (Apo-A2), whose percentage differ depending on the considered lipoprotein but also on the presence of metabolic dysregulations caused by different health conditions, providing important information largely inaccessible through the use of other techniques (i.e. enzymatic and colorimetric assays, ultracentrifugation, electrophoresis).

As the previous heatmaps, coloured squares, according to the bar legend, represent the obtained fold changes from the comparisons of the variables (in the abscise axis) for each condition (in the ordinate axis) with the apparently healthy one (0000). In this case, unsupervised cluster analysis separates according to dyslipidaemia (XX1X or XX0X) instead of diabetes, as expected also from the previous result of the PCA analysis (**Figure 4.13**).

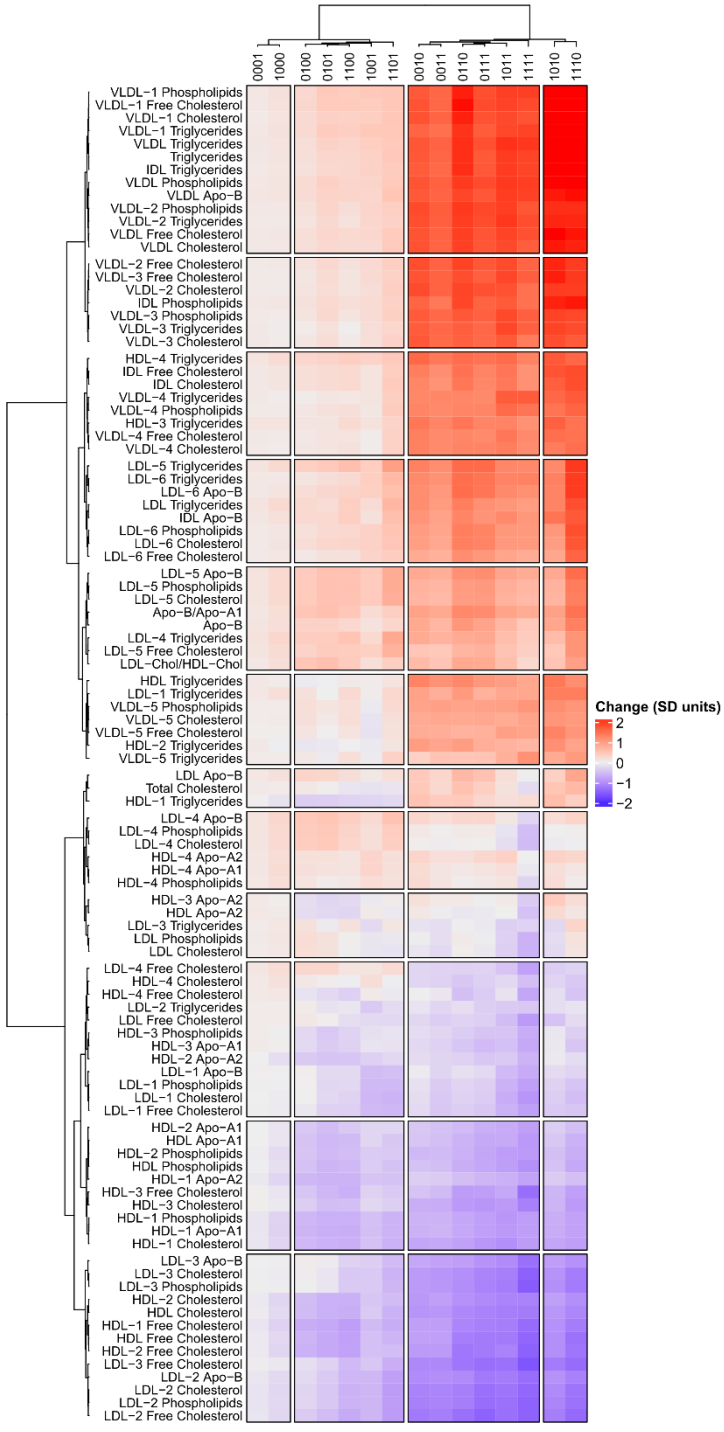

Figure 4.15: Heatmap of the comparison of the different serum conditions with the apparently healthy one (0000). Conditions are represented in the abscise axis while in the ordinate are listed the lipoproteins and related parameters, both sorted according to cluster analysis. Fold-change for each of the comparison is evidenced by the colour-code indicated in the bar legend. Dendodrams on the top and on the left evidence the clustering of the different profiles and lipoproteins respectively.

The obtained results report on a significant increase in VLDL particles (especially VLDL-1 and VLDL-2, followed by VLDL-3), IDL, and to a lesser extent of LDL, predominantly LDL-5 and LDL-6 which could be considered as the smaller dense LDL. On the other hand, it is possible to observe a lower level of LDL-2 and LDL-3 particles and of the HDL lipoproteins (HDL-1 and HDL-2).

The described dysregulation is particularly evident in the profiles of the individuals affected by MetS, according to our criteria, including the 1111, 1110 and 1011 profiles, but also in the 0111 condition and in all those affected by dyslipidaemia.

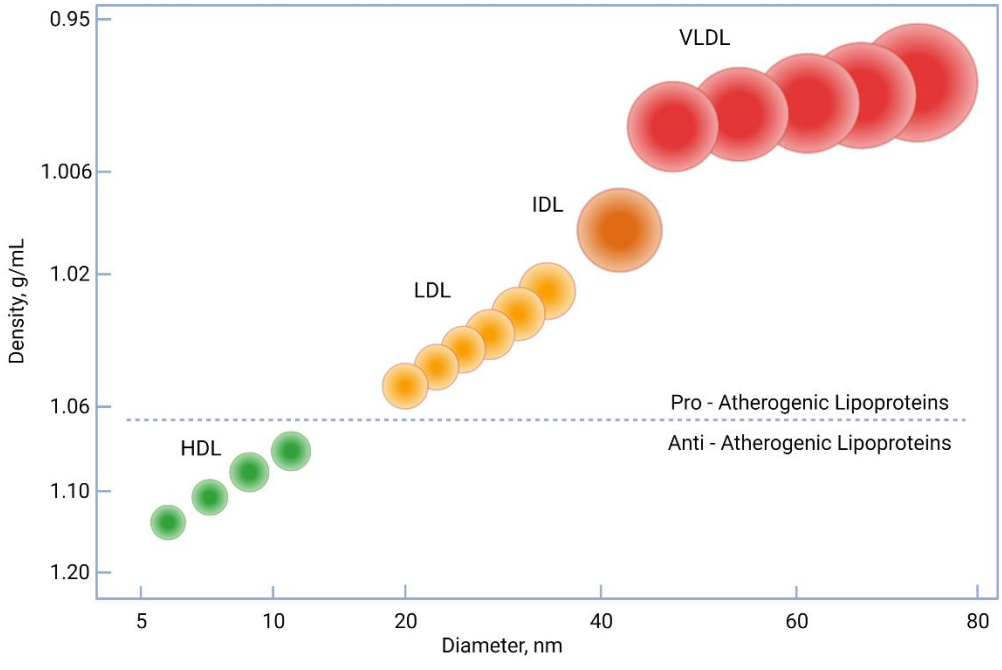

Figure 4.16: Size and density of the quantified lipoproteins according to the Bruker B.I.LISA™ report (figure modified from Feingold K. R., 2021[117]). LDL, IDL and VLDL are considered as pro-atherogenic lipoproteins due to their transport of lipid to the peripheral tissues, while HDL are the only anti-atherogenic lipoproteins thanks to their reverse cholesterol transport from peripheral tissues to the liver. VLDL, Very low density lipoproteins; IDL intermediate density lipoproteins; LDL, low density lipoproteins; HDL, high density lipoproteins.

The observed changes are in line with previous studies reporting a characteristic fingerprint in the lipoproteins expression for the subjects affected by MetS that was generally associated to the increased levels of triglycerides, typical of this syndrome,

and that is also particularly evident in our results, as shown in heatmap from the conduced univariate analysis[100–102].

The observed lipoproteins variations may result from insulin resistance, possibly generated by the co-presence of other factors such as obesity or diabetes, which could cause an increased synthesis of triglycerides in the liver from free fatty acids, but the causes that could lead to this increase in MetS subjects could be several and still need to be thoroughly investigated[101].

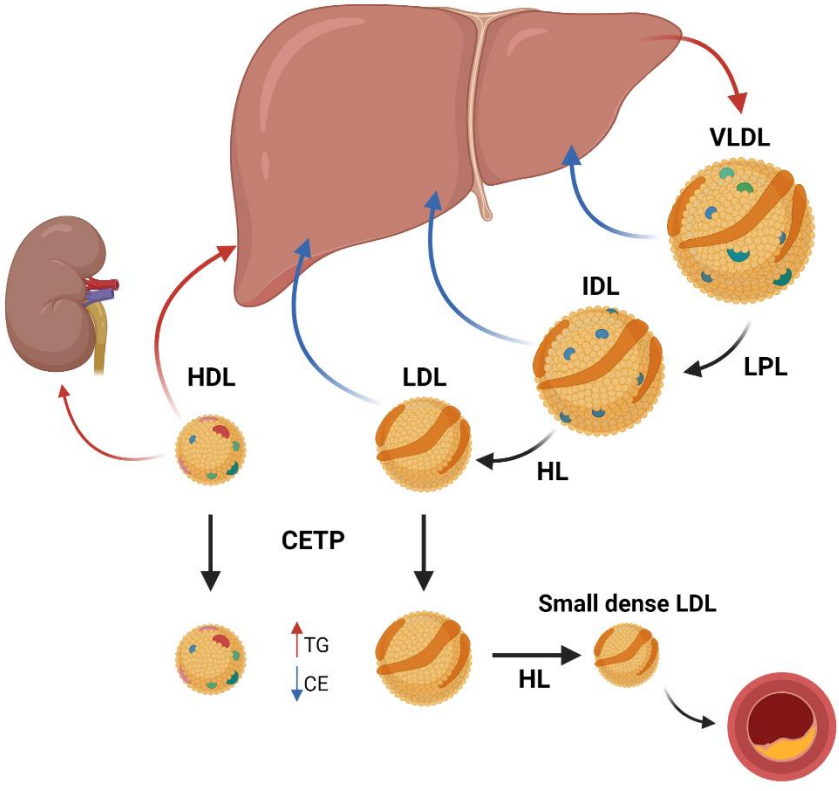

Figure 4.17: Lipoprotein metabolism in MetS subjects. Red arrow represents the increase in VLDL lipoproteins due the increased triglycerides. IDL, generated from VLDL by the action of LPL, can be further hydrolysed by HL creating LDL. Blue arrows evidence a decreased hepatic uptake of the produced lipoproteins in MetS subjects (VLDL, IDL, LDL). Moreover, by the action of CETP, cholesteryl esters can be replaced by TG in LDL and HDL. The created TG-rich-LDL are considered as good substrate for the HL that generates small dense LDL, which are associated with an increased cardiovascular risk. On the other hand, the TG-rich-HDL are more prone to be catabolised either in the liver or in the kidney. VLDL, Very low density lipoproteins; IDL, intermediate density lipoproteins; LDL, low density lipoproteins; HDL, high density lipoproteins; LPL, lipoprotein lipase; HL, hepatic lipase; CETP, cholesteryl ester transfer protein; TG, triglycerides; CE, cholesteryl esters.

This increase in triglycerides is widely considered as responsible for the overproduction of large VLDL (VLDL-1) in the liver, with a high triglycerides content, as observed also by our analysis. Their synthesis has been elucidated by several studies and involves a multistep process, starting with the lipidation of the Apo-B (more specifically Apo-B100), in the endoplasmic reticulum (ER), which is generally considered as the key structural protein for VLDL lipoprotein assembly. The generated pre-VLDL particles, through their transition to the Golgi complex, turns into triglycerides-poor VLDL (VLDL-2), which can be released from the cells or transformed into triglycerides-rich VLDL (VLDL-1) by further lipidation[100,101]. In the presence of MetS or, in general of dyslipidaemia, a reduced clearance of the VLDL-1 and VLDL-2 is observed, due to the increased level of triglycerides and an altered function of the lipoprotein lipase (LPL), leading to an accumulation of this lipoproteins in the bloodstream[101]. The described process can be responsible of the observed increase in VLDL-1, VLDL-2, VLDL Apo-B and especially of the triglycerides content.

Cholesterol percentage also showed to be increased in the upregulated VLDLs in the MetS and dyslipidemic subjects. This can be related to the intrinsic composition of these lipoproteins, as well as phospholipids, which also seem to be upregulated. Moreover, previous studies reported an increase in the production of VLDL lipoproteins, not only in relation to the presence of triglycerides, but also due to higher levels of cholesterol[101,103].

The observed IDL increase, in addition to the so far described alterations, can be related to the VLDL overproduction. In fact, IDL could be generated from triglyceride rich VLDL hydrolyzation through the action of LPL and this can be followed by the generation of cholesterol rich LDL by the hepatic lipase (HL), always preserving the structural protein Apo-B in both IDL and LDL[100,104]. The hepatic uptake of the produced lipoproteins (VLDL, IDL and LDL) was decreased in the MetS patients, explaining their upregulation in the conducted analysis[105–107]. Moreover, previous studies suggested that the production of a population of LDL particles with an altered conformation of the Apo-B results in a decreased affinity for the binding to LDL receptors, also justifying their higher levels in the bloodstream. It was observed that

their permanence in the circulation, could make them more easily accessible to the cholesteryl ester transfer protein, substituting cholesteryl esters with triglycerides. The generated triglycerides rich LDL have a greater affinity for the HL, generating small dense LDL (like LDL-5 and LDL-6), which also showed to be upregulated in our serum analysis and associated with the highest atherosclerotic risk[101,108,109]. The described process could be responsible of the observed downregulation in the cholesterol rich LDL-3 and LDL-2 lipoproteins, which presumably are converted into the small dense triglyceride rich LDL-5 and LDL-6 lipoproteins by the HL (**Figure 4.17**).

Finally, a decreased level of HDL was observed and could be justified by an increased clearance of these lipoproteins. As explained, for the generation of the triglycerides rich LDL lipoproteins through the action of cholesteryl ester transfer protein (CETP), it was observed the same mechanism for the HDL particles, where cholesteryl esters were replaced by triglycerides. The generated lipoproteins particles were evidenced as more prone to catabolisation, producing their decreased level in the bloadstream[101].

We hypothesize that the increase in the risk of cardiovascular problems that characterise the patients affected by MetS is related to the observed lipoprotein dysregulation, especially to the increased amount of small dense atherogenic LDL (LDL-5 and LDL-6). Previous studies related their adverse effect to different properties of these particles like their decreased size, responsible of an increased accumulation and retention in the walls of the blood vessels, a reduction in recognition by the LDL-receptor responsible for their clearance and their increased predisposition to oxidation[110].

4.2.4 A molecular signature of MetS by serum analysis

Due to the observed metabolic and lipidomic changes, it was interesting to observe, by means of multivariant analysis, the correlation between the average profiles generated by each of the conditions under study compared to the healthy one, considering both metabolites and lipoproteins as variables.

A correlation map was built, following the previous explained criteria (in **Figure 4.5**), by sorting the profiles according to the increased number of RFs that characterize each condition, generating 5 columns, each differing by the presence of one RF (**Figure 4.18**). In the map, conditions are connected by lines and coloured coded according to their Superman's correlation distance from the apparently healthy condition (0000) following the bar legend.

The represented graph shows a more evident metabolic dysregulation in the serum spectra profile in relation with the increasing number of RFs, but in this case, dyslipidaemia is the main discriminating RF, causing the greatest degree of variation. Indeed, 0010 condition, with only one RF, already shows a significant change with respect to the 0000 profile. Despite this, it is possible to observe that the concomitant presence of certain risk factors, together with dyslipidaemia, is more pathogenic than others.

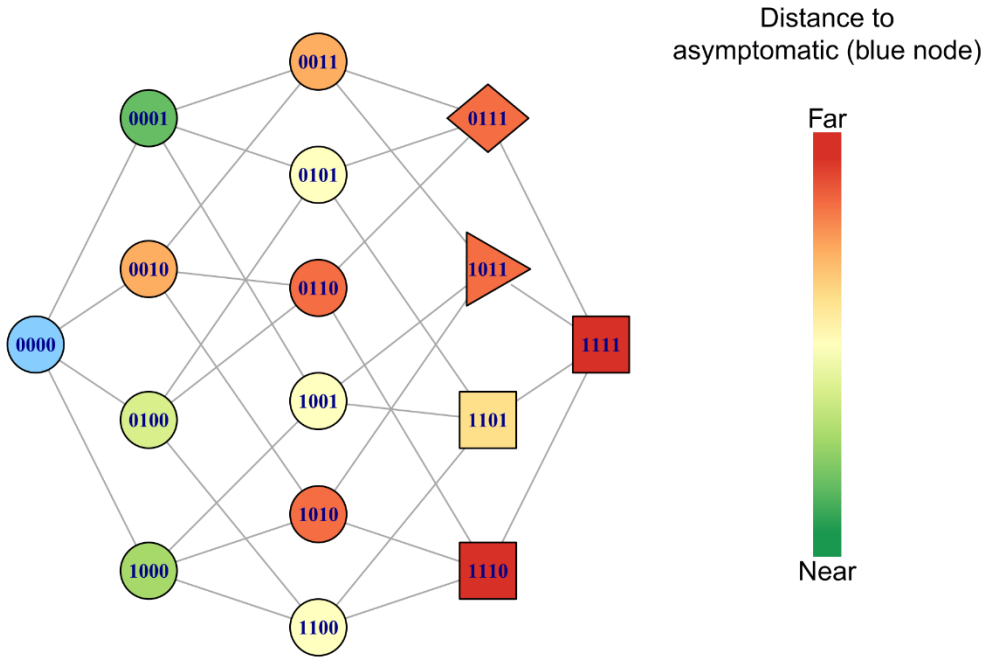

Figure 4.18: Spearman correlation map for serum profiles evidencing the distance of each condition from the apparently healthy one (0000), colour coded according to it, as indicated by the bar legend. MetS conditions according to the WHO, EGIR and AACE are represented by squares and triangles; these last two with the addition of the rhombus represent MetS according to the NCEP:ATPIII and harmonized definitions, while squares and rhombus are the profiles presenting MetS according to the IDF.

Conditions with diabetes and/or obesity, together with dyslipidaemia, as especially evident for the 0110 and 1010 profiles were more pathogenic whereas, unlike previous urine analysis, hypertension in serum did not exhibit such a severe impact (0011 profile).

Since in this case dyslipidaemia (and no longer diabetes) was the main cause of the differentiation of the profiles from the apparently healthy cohort, squares, triangle and rhombus (representing the profiles affected by metabolic syndrome according to previously described existing definitions, **Table 1.1**) do not all show the same degree of severity, and 1110 and 1111 conditions show the most different metabolic and lipidomic fingerprinting with respect to the 0000 condition. Despite this, both these profiles are considered as affected by MetS according to all the described criteria.

The previous urine analysis showed how diabetes, whose presence was considered as mandatory for diagnosis according to many of the existing definitions (**Table 1.1**), confirmed its relevance through the produced metabolic variation, and hypertension also seemed to play an important role in this aspect. In serum analysis, diabetes still results in important metabolic changes, but the importance of dyslipidaemia is also emphasised, with obesity as a secondary factor.

The obtained results reflect how the observed metabolic changes, in relation to the presence of certain RFs, are important to understand which factors should be considered as most relevant for the diagnosis of MetS and should therefore be taken more seriously, especially for prevention and treatment of this syndrome.

4.3 Final considerations

Up to date, there is a lack of (molecular) instruments for the efficient evaluation of the stage of metabolic syndrome that may characterise a subject[111]. Indeed, nowadays, there is a growing interest for the identification of molecular biomarkers and diagnostic tools for an early detection of the factors that could lead to the development of MetS and try to avoid this to happen. This is even more complex when dealing with multifactorial diseases that may depend on environmental, physical, metabolic but also genetic or cellular factors as metabolic syndrome[30,112,113]. Yet, the existing definitions

often only consider a number of factors that are "generically" applied for diagnosis, but for such complex syndromes more integrative approaches may be necessary.

The conducted study aimed to describe all the MetS-related phenotypes from a metabolic point of view in order to attempt to design a tool that would be able to define in a more personalised way how the metabolic changes that characterise an individual, based on the presence of certain RFs, may made a person more likely to develop this disorder. This could be achieved by integrating the results obtained from NMR metabolomics analysis of urine and serum samples.

The metabolic phenotyping of MetS also allowed the identification of new biomarkers for the characterization of this disorder and the corroboration of others previously found. Different studies have been carried out to study MetS, including by means of NMR spectroscopy, but our study included an unprecedented number of analysed samples that gave greater statistical power and reliability to the obtained results, thanks also to the constant application of established SOPs (extensively explained in Chapter 1 and 3) in each of the steps of our analysis process.

Moreover, the conducted analysis may be useful for understanding more deeply the molecular mechanisms involved in this syndrome and the different factors that contribute to the development of this disorder. Although further experiments are still needed to fully elucidate certain molecular processes and pathways dysregulations.

According to the obtained results, diabetes and dyslipidaemia shall be considered as key components for the development of MetS, and the presence of these RFs can lead to a rapid progression of this disorder, especially causing a significant increase in the risk of insurgence of secondary problems associated with this syndrome, such as T2DM or CVD.

MetS insurgence have often been related to erroneous lifestyle habits that, if persisted, may influence a genetic and phenotypic susceptibility[30,112,114,115]. Diet, according to nutrigenomics, could even produce specific genetically programmed responses in relation to certain food[30]. Indeed, prevention by means of healthier nutritional habits, or even by personalized nutrition to target specific metabolism deviations, together

with a regular physical activity appeared to be one of the most effective[30,116]. On the contrary, when the disease is developed, the intervention by medication is also needed.

Bibliography Chapter 4:

1. Bruzzone, C. *et al.* A molecular signature for the metabolic syndrome by urine metabolomics. *Cardiovasc Diabetol* **20**, 1–13 (2021).
2. Bouatra, S. *et al.* The Human Urine Metabolome. *PLoS One* **8**, (2013).
3. Singh, S., Khera, R., Allen, A. M., Murad, M. H. & Loomba, R. Comparative effectiveness of pharmacological interventions for nonalcoholic steatohepatitis: A systematic review and network meta-analysis. *Hepatology* **62**, 1417–1432 (2015).
4. McPherson, S. *et al.* Evidence of NAFLD progression from steatosis to fibrosing-steatohepatitis using paired biopsies: Implications for prognosis and clinical management. *J Hepatol* **62**, 1148–1155 (2015).
5. Lent-Schochet, D., McLaughlin, M., Ramakrishnan, N. & Jialal, I. Exploratory metabolomics of metabolic syndrome: A status report. *World J Diabetes* **10**, 23–36 (2019).
6. Lindon, J. C., Nicholson, J. K. & Holmes, E. The Handbook of Metabonomics and Metabolomics. *The Handbook of Metabonomics and Metabolomics* (2007) doi:10.1016/B978-0-444-52841-4.X5000-0.
7. Barrea, L. *et al.* Trimethylamine-N-oxide (TMAO) as novel potential biomarker of early predictors of metabolic syndrome. *Nutrients* **10**, 1–19 (2018).
8. Choi, M., Mukherjee, S. & Yun, J. W. Trigonelline induces browning in 3T3-L1 white adipocytes. *Phytotherapy Research* **35**, 1113–1124 (2021).
9. Mallmann, N. H., Lima, E. S. & Lalwani, P. Dysregulation of Tryptophan Catabolism in Metabolic Syndrome. *Metab Syndr Relat Disord* **16**, (2018).
10. Favennec, M. *et al.* The kynurenine pathway is activated in human obesity and shifted toward kynurenine monooxygenase activation. *Obesity (Silver Spring)* **23**, 2066–2074 (2015).
11. SCHÄFER, S. G. *et al.* Why Imidazoline Receptor Modulator in the Treatment of Hypertension? *Ann N Y Acad Sci* **763**, 659–672 (1995).
12. Sun, S. *et al.* Metabolic Syndrome and Its Components Are Associated With Altered Amino Acid Profile in Chinese Han Population. *Front Endocrinol (Lausanne)* **12**, 795044 (2021).
13. Koppe, L. *et al.* p-Cresyl sulfate promotes insulin resistance associated with CKD. *Journal of the American Society of Nephrology* **24**, 88–99 (2013).
14. Yousri, N. A. *et al.* A systems view of type 2 diabetes-associated metabolic perturbations in saliva, blood and urine at different timescales of glycaemic control. *Diabetologia* **58**, 1855 (2015).

15. Vizzari, G. *et al.* Circulating salicylic acid and metabolic profile after 1-year nutritional–behavioral intervention in children with obesity. *Nutrients* **11**, 1–11 (2019).

16. Huang, C. F., Cheng, M. L., Fan, C. M., Hong, C. Y. & Shiao, M. S. Nicotinuric acid: A potential marker of metabolic syndrome through a metabolomics-based approach. *Diabetes Care* **36**, 1729–1731 (2013).

17. O'Neill, S. & O'Driscoll, L. Metabolic syndrome: A closer look at the growing epidemic and its associated pathologies. *Obesity Reviews* **16**, 1–12 (2015).

18. Ntzouvani, A. *et al.* Amino acid profile and metabolic syndrome in a male Mediterranean population: A cross-sectional study. *Nutrition, Metabolism and Cardiovascular Diseases* **27**, 1021–1030 (2017).

19. Peddinti, G. *et al.* Early metabolic markers identify potential targets for the prevention of type 2 diabetes. *Diabetologia* **60**, 1740–1750 (2017).

20. Reddy, P., Leong, J. & Jialal, I. Amino acid levels in nascent metabolic syndrome: A contributor to the pro-inflammatory burden. *J Diabetes Complications* **32**, 465–469 (2018).

21. Wang, X., Hu, Z., Hu, J., Du, J. & Mitch, W. E. Insulin Resistance Accelerates Muscle Protein Degradation: Activation of the Ubiquitin-Proteasome Pathway by Defects in Muscle Cell Signaling. *Endocrinology* **147**, 4160–4168 (2006).

22. Combs, T. P., Marliss, E. B., Lamarche, M. & Chevalier, S. Adiponectin is an Independent Predictor of Circulating Branched-Chain Amino Acids in Men. *Can J Diabetes* **40**, S55–S56 (2016).

23. Nakamura, H. *et al.* Plasma amino acid profiles are associated with insulin, C-peptide and adiponectin levels in type 2 diabetic patients. *Nutr Diabetes* **4**, 133 (2014).

24. Wang, W. *et al.* Dietary Tryptophan and the Risk of Metabolic Syndrome: Total Effect and Mediation Effect of Sleep Duration. *Nat Sci Sleep* **13**, 2141 (2021).

25. Fears, R. & Murrellt, E. A. Tryptophan and the control of triglyceride and carbohydrate metabolism in the rat. *British Journal of Nutrition* **43**, 349–356 (1980).

26. Ardiansyah, Shirakawa, H., Inagawa, Y., Koseki, T. & Komai, M. Regulation of blood pressure and glucose metabolism induced by L-tryptophan in stroke-prone spontaneously hypertensive rats. *Nutr Metab (Lond)* **8**, 1–7 (2011).

27. Yamada, C. *et al.* Association between insulin resistance and plasma amino acid profile in non-diabetic Japanese subjects. *J Diabetes Investig* **6**, 408–415 (2015).

28. Newgard, C. B. *et al.* A Branched-Chain Amino Acid-Related Metabolic Signature that Differentiates Obese and Lean Humans and Contributes to Insulin Resistance. *Cell Metab* **9**, 311–326 (2009).

29. Lin, C. *et al.* The causal associations of circulating amino acids with blood pressure: a Mendelian randomization study. *BMC Med* **20**, 1–11 (2022).

30. Hernandez-Baixauli, J. *et al.* Detection of early disease risk factors associated with metabolic syndrome: A new era with the NMR metabolomics assessment. *Nutrients* **12**, (2020).

31. Gao, X., Tian, Y., Randell, E., Zhou, H. & Sun, G. Unfavorable associations between serum trimethylamine N-oxide and L-carnitine levels with components of metabolic syndrome in the Newfoundland population. *Front Endocrinol (Lausanne)* **10**, 1–12 (2019).

32. Sonnenburg, J. L. *et al.* Glycan foraging in vivo by an intestine-adapted bacterial symbiont. *Science (1979)* **307**, 1955–1959 (2005).

33. da Silva, H. E. *et al.* Nonalcoholic fatty liver disease is associated with dysbiosis independent of body mass index and insulin resistance. *Sci Rep* **8**, (2018).

34. Chen, S. *et al.* Trimethylamine N-Oxide Binds and Activates PERK to Promote Metabolic Dysfunction. *Cell Metab* **30**, 1141-1151.e5 (2019).

35. Chiang, J. Y. L. Regulation of bile acid synthesis: pathways, nuclear receptors, and mechanisms. *J Hepatol* **40**, 539–551 (2004).

36. Wang, Z. *et al.* Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472**, 57–63 (2011).

37. Randrianarisoa, E. *et al.* Relationship of serum trimethylamine N-oxide (TMAO) levels with early atherosclerosis in humans. *Sci Rep* **6**, 1–9 (2016).

38. Lin, H., An, Y., Hao, F., Wang, Y. & Tang, H. Correlations of Fecal Metabonomic and Microbiomic Changes Induced by High-fat Diet in the Pre-Obesity State. *Sci Rep* **6**, (2016).

39. Schiattarella, G. G. *et al.* Gut microbe-generated metabolite trimethylamine-N-oxide as cardiovascular risk biomarker: a systematic review and dose-response meta-analysis. *Eur Heart J* **38**, 2948–2956 (2017).

40. Anwar, S. *et al.* Trigonelline inhibits intestinal microbial metabolism of choline and its associated cardiovascular risk. *J Pharm Biomed Anal* **159**, 100–112 (2018).

41. Chaturvedi, A. K. & Liu, R. H. Substance Misuse: Urine Analysis. *Encyclopedia of Forensic and Legal Medicine: Second Edition* 442–452 (2015).

42. Heavner, J. E. & Cooper, D. M. Pharmacology of Analgesics. *Anesthesia and Analgesia in Laboratory Animals* 97–123 (2008).

43. Badenhorst, C. P. S., Erasmus, E., van der Sluis, R., Nortje, C. & van Dijk, A. A. A new perspective on the importance of glycine conjugation in the metabolism of aromatic acids. *Drug metabolism reviews*, **46**, 343–361 (2014).

44. Alabi, Q. K. & Adeyemi, W. J. Vernonia amygdalina (Del) as an antioxidant, aspirin toxicity, and oxidative stress. *Toxicology* 491–504 (2021).

45. Wu, I. W. *et al.* Serum free p-cresyl sulfate levels predict cardiovascular and all-cause mortality in elderly hemodialysis patients--a prospective cohort study. *Nephrol Dial Transplant* **27**, 1169–1175 (2012).

46. Liabeuf, S. *et al.* Free p-cresylsulphate is a predictor of mortality in patients at different stages of chronic kidney disease. *Nephrol Dial Transplant* **25**, 1183–1191 (2010).

47. Bammens, B., Evenepoel, P., Keuleers, H., Verbeke, K. & Vanrenterghem, Y. Free serum concentrations of the protein-bound retention solute p-cresol predict mortality in hemodialysis patients. *Kidney Int* **69**, 1081–1087 (2006).

48. Olszanecka, A., Kawecka-Jaszcz, K. & Czarnecka, D. Association of free testosterone and sex hormone binding globulin with metabolic syndrome and subclinical atherosclerosis but not blood pressure in hypertensive perimenopausal women. *Arch Med Sci* **12**, 521 (2016).

49. Blouin, K. *et al.* Contribution of age and declining androgen levels to features of the metabolic syndrome in men. *Metabolism* **54**, 1034–1040 (2005).

50. Marchand, G. B. *et al.* Increased body fat mass explains the positive association between circulating estradiol and insulin resistance in postmenopausal women. *Am J Physiol Endocrinol Metab* **314**, E448–E456 (2018).

51. Darcy, C. J. *et al.* An observational cohort study of the kynurenine to tryptophan ratio in sepsis: association with impaired immune and microvascular function. *PLoS One* **6**, (2011).

52. Meyer, K. C. *et al.* Tryptophan metabolism in chronic inflammatory lung disease. *J Lab Clin Med* **126**, 530–540 (1995).

53. Gelpi, M. *et al.* Tryptophan catabolism and immune activation in primary and chronic HIV infection. *BMC Infect Dis* **17**, 1–8 (2017).

54. Esposito, K. & Giugliano, D. The metabolic syndrome and inflammation: Association or causation? *Nutrition, Metabolism and Cardiovascular Diseases* **14**, 228–232 (2004).

55. Bousquet, P., Hudson, A., García-Sevilla, J. A. & Li, J. X. Imidazoline receptor system: The past, the present, and the future. *Pharmacol Rev* **72**, 50–79 (2020).

56. Timmermans, P. B. M. W. M. *et al.* Angiotensin II receptor antagonists: From discovery to antihypertensive drugs. *Hypertension* **18**, III-136-III–142 (1991).

57. Okazaki, T. *et al.* Studies on nonpeptide angiotensin II receptor antagonists. IV. Synthesis and biological evaluation of 4-acrylamide-1H-imidazole derivatives. *Chem Pharm Bull (Tokyo)* **46**, 973–981 (1998).

58. van Son, J. *et al.* Plasma Imidazole Propionate Is Positively Correlated with Blood Pressure in Overweight and Obese Humans. *Nutrients* **13**, (2021).

59. Molinaro, A. *et al.* Imidazole propionate is increased in diabetes and associated with dietary patterns and altered microbial ecology. *Nat Commun* **11**, (2020).

60. Koh, A. *et al.* Microbially Produced Imidazole Propionate Impairs Insulin Signaling through mTORC1. *Cell* **175**, 947-961.e17 (2018).

61. Franz, M. J. *et al.* Evidence-based nutrition principles and recommendations for the treatment and prevention of diabetes and related complications. *Diabetes Care* **26 Suppl 1**, (2003).

62. Gómez-Fernández, A. R., Santacruz, A., Jacobo-Velázquez, D. A., Correspondence, D. A. & Jacobo-Velázquez, T. The complex relationship between metabolic syndrome and sweeteners. *J Food Sci* **86**, 1511–1531 (2021).

63. Chukwuma, C. I., Ibrahim, M. A. & Islam, M. S. Maltitol inhibits small intestinal glucose absorption and increases insulin mediated muscle glucose uptake ex vivo but not in normal and type 2 diabetic rats. *Int J Food Sci Nutr* **68**, 73–81 (2017).

64. Martin, F.-P. J., Collino, S., Rezzi, S., Kochhar, S. & Pandol, S. J. Metabolomic applications to decipher gut microbial metabolic influence in health and disease. **3**, (2012).

65. Pietzke, M. *et al.* Stratification of cancer and diabetes based on circulating levels of formate and glucose. *Cancer Metab* **7**, (2019).

66. Pietzke, M., Meiser, J. & Vazquez, A. Formate metabolism in health and disease. *Mol Metab* **33**, 23 (2020).

67. Pujos-Guillot, E. *et al.* Systems Metabolomics for Prediction of Metabolic Syndrome. *J Proteome Res* **16**, 2262–2272 (2017).

68. Strack, C. *et al.* Gender differences in cardiometabolic health and disease in a cross-sectional observational obesity study. *Biol Sex Differ* **13**, 1–11 (2022).

69. Lee, S., Ko, Y., Kwak, C. & Yim, E. S. Gender differences in metabolic syndrome components among the Korean 66-year-old population with metabolic syndrome Public health, nutrition and epidemiology. *BMC Geriatr* **16**, 1–8 (2016).

70. Hildrum, B., Mykletun, A., Hole, T., Midthjell, K. & Dahl, A. A. Age-specific prevalence of the metabolic syndrome defined by the International Diabetes Federation and the National Cholesterol Education Program: The Norwegian HUNT 2 study. *BMC Public Health* **7**, 1–9 (2007).

71. Sheng, C. S. *et al.* Microalbuminuria in relation to the metabolic syndrome and its components in a Chinese population. *Diabetol Metab Syndr* **3**, 3–8 (2011).

72. Guh, J. Y. Proteinuria versus albuminuria in chronic kidney disease. *Nephrology* **15**, 53–56 (2010).

73. Miller, W. G. *et al.* Current Issues in Measurement and Reporting of Urinary Albumin Excretion. *Clin Chem* **55**, 24–38 (2009).

74. Saadi, M. M. *et al.* Association of microalbuminuria with metabolic syndrome: A cross-sectional study in Bangladesh. *BMC Endocr Disord* **20**, 1–7 (2020).

75. Levey, A. S. *et al.* A new equation to estimate glomerular filtration rate. *Ann Intern Med* **150**, (2009).

76. Chen, S. H. *et al.* Relationship between nonalcoholic fatty liver disease and metabolic syndrome. *J Dig Dis* **12**, 125–130 (2011).

77. Yki-Järvinen, H. Non-alcoholic fatty liver disease as a cause and a consequence of metabolic syndrome. *Lancet Diabetes Endocrinol* **2**, 901–910 (2014).

78. Dietrich, P. & Hellerbrand, C. Non-alcoholic fatty liver disease, obesity and the metabolic syndrome. *Best Pract Res Clin Gastroenterol* **28**, 637–653 (2014).

79. Zheng, Y. *et al.* Cumulative consumption of branched-chain amino acids and incidence of type 2 diabetes. *Int J Epidemiol* **45**, 1482 (2016).

80. Cheng, S. *et al.* Serum metabolic profiles in overweight and obese women with and without metabolic syndrome. *Diabetol Metab Syndr* **6**, (2014).

81. Sookoian, S. & Pirola, C. J. Alanine and aspartate aminotransferase and glutamine-cycling pathway: Their roles in pathogenesis of metabolic syndrome. *World J Gastroenterol* **18**, (2012).

82. Maltais-Payette, I., Boulet, M. M., Prehn, C., Adamski, J. & Tchernof, A. Circulating glutamate concentration as a biomarker of visceral obesity and associated metabolic alterations. *Nutr Metab (Lond)* **15**, (2018).

83. Tulipani, S. *et al.* Biomarkers of Morbid Obesity and Prediabetes by Metabolomic Profiling of Human Discordant Phenotypes. *Clinica Chimica Acta* **463**, 53–61 (2016).

84. Newsholme, P., Procopio, J., Ramos Lima, M. M., Pithon-Curi, T. C. & Curi, R. Glutamine and glutamate—their central role in cell metabolism and function. *Cell Biochem Funct* **21**, 1–9 (2003).

85. Miccheli, A. *et al.* Urinary 1H-NMR-based metabolic profiling of children with NAFLD undergoing VSL#3 treatment. *International Journal of Obesity 2015 39:7* **39**, 1118–1125 (2015).

86. Ducker, G. S. & Rabinowitz, J. D. One-Carbon Metabolism in Health and Disease. *Cell Metab* **25**, 27–42 (2017).

87. Mentch, S. J. *et al.* Histone Methylation Dynamics and Gene Regulation Occur through the Sensing of One-Carbon Metabolism. *Cell Metab* **22**, 861–873 (2015).

88. Yeo, E. J. & Wagner, C. Purification and properties of pancreatic glycine N-methyltransferase. *Journal of Biological Chemistry* **267**, 24669–24674 (1992).

89. Alves, A., Bassot, A., Bulteau, A. L., Pirola, L. & Morio, B. Glycine Metabolism and Its Alterations in Obesity and Metabolic Diseases. *Nutrients* **11**, (2019).

90. Shearer, J. *et al.* Metabolomic profiling of dietary-induced insulin resistance in the high fat–fed C57BL/6J mouse. *Diabetes Obes Metab* **10**, 950–958 (2008).

91. Kotronen, A. *et al.* Liver fat and lipid oxidation in humans. *Liver International* **29**, 1439–1446 (2009).

92. Cotter, D. G., Schugar, R. C. & Crawford, P. A. Ketone body metabolism and cardiovascular disease. *Am J Physiol Heart Circ Physiol* **304**, 1060–1076 (2013).

93. Yuan, J. *et al.* Fatty Liver Disease Caused by High-Alcohol-Producing Klebsiella pneumoniae. *Cell Metab* **30**, 675-688.e7 (2019).

94. Meijnikman, A. S. *et al.* Microbiome-derived ethanol in nonalcoholic fatty liver disease. *Nature Medicine 2022 28:10* **28**, 2100–2106 (2022).

95. Flint, H. J., Duncan, S. H., Scott, K. P. & Louis, P. Links between diet, gut microbiota composition and gut metabolism. *Proceedings of the Nutrition Society* **74**, 13–22 (2015).

96. Crawford, S. O. *et al.* Association of blood lactate with type 2 diabetes: the Atherosclerosis Risk in Communities Carotid MRI Study. *Int J Epidemiol* **39**, 1647–1655 (2010).

97. Adeva-Andany, M. *et al.* Comprehensive review on lactate metabolism in human health. *Mitochondrion* **17**, 76–100 (2014).

98. Plovier, H. & Cani, P. D. Microbial Impact on Host Metabolism: Opportunities for Novel Treatments of Nutritional Disorders? *Microbiol Spectr* **5**, (2017).

99. Boets, E. *et al.* Systemic availability and metabolism of colonic-derived short-chain fatty acids in healthy subjects: a stable isotope study. *J Physiol* **595**, 541–555 (2017).

100. Adiels, M., Olofsson, S. O., Taskinen, M. R. & Borén, J. Overproduction of very low-density lipoproteins is the hallmark of the dyslipidemia in the metabolic syndrome. *Arterioscler Thromb Vasc Biol* **28**, 1225–1236 (2008).

101. Kolovou, G. D., Anagnostopoulou, K. K. & Cokkinos, D. v. Pathophysiology of dyslipidaemia in the metabolic syndrome. *Postgrad Med J* **81**, 358 (2005).

102. Ginsberg, H. N., Zhang, Y. L. & Hernandez-Ono, A. Metabolic syndrome: focus on dyslipidemia. *Obesity (Silver Spring)* **14 Suppl 1**, (2006).

103. Prinsen, B. H. C. M. T. *et al.* Endogenous cholesterol synthesis is associated with VLDL-2 apoB-100 production in healthy humans. *J Lipid Res* **44**, 1341–1348 (2003).

104. Brown, M. S. & Goldstein, J. L. A receptor-mediated pathway for cholesterol homeostasis. *Science* **232**, 34–47 (1986).

105. Watts, G. F. *et al.* Differential regulation of lipoprotein kinetics by atorvastatin and fenofibrate in subjects with the metabolic syndrome. *Diabetes* **52**, 803–811 (2003).

106. Chan, D. C., Watts, G. F., Redgrave, T. G., Mori, T. A. & Barrett, P. H. R. Apolipoprotein B-100 kinetics in visceral obesity: Associations with plasma apolipoprotein C-III concentration. *Metabolism* **51**, 1041–1046 (2002).

107. Riches, F. M. *et al.* Hepatic secretion of very-low-density lipoprotein apolipoprotein B-100 studied with a stable isotope technique in men with visceral obesity. *Int J Obes Relat Metab Disord* **22**, 414–423 (1998).

108. Lamarche, B. *et al.* Fasting Insulin and Apolipoprotein B Levels and Low-Density Lipoprotein Particle Size as Risk Factors for Ischemic Heart Disease. *JAMA* **279**, 1955–1961 (1998).

109. Packard, C. J. Triacylglycerol-rich lipoproteins and the generation of small, dense low-density lipoprotein. *Biochem Soc Trans* **31**, 1066–1069 (2003).

110. Chait, A., Brazg, R. L., Tribble, D. L. & Krauss, R. M. Susceptibility of small, dense, low-density lipoproteins to oxidative modification in subjects with the atherogenic lipoprotein phenotype, pattern B. *Am J Med* **94**, 350–356 (1993).

111. Quesada-Vázquez, S., Hernandez-Baixauli, J., Navarro-Masip, E. & Escoté, X. NMR Metabolomics for Marker Discovery of Metabolic Syndrome. 1–29 (2022) doi:10.1007/978-3-030-81304-8_39-1.

112. Marti, A., Martinez-González, M. A. & Martinez, J. A. Interaction between genes and lifestyle factors on obesity. *Proc Nutr Soc* **67**, 1–8 (2008).

113. Vassallo, P., Driver, S. L. & Stone, N. J. Metabolic Syndrome: An Evolving Clinical Construct. *Prog Cardiovasc Dis* **59**, 172–177 (2016).

114. Oladejo, A. O. Overview of the metabolic syndrome; an emerging pandemic of public health significance. *Ann Ib Postgrad Med* **9**, 78 (2011).

115. Hossain, P., Kawar, B. & el Nahas, M. Obesity and diabetes in the developing world--a growing challenge. *N Engl J Med* **356**, 213–215 (2007).

116. Farhud, D. & Zarif Yeganeh, M. Nutrigenomics and Nutrigenetics. *Iranian J Publ Health* **39**, 1–14 (2010).

117. Feingold, K. R. & Grunfeld, C. Introduction to Lipids and Lipoproteins. *Endotext* (2021).

Chapter 5

Metabolomic and Lipidomic dysregulation caused by SARS-CoV-2 infection

Results and Discussion

5.1 SARS-CoV-2 infection

As previously explained in the introduction, SARS-CoV-2 infection has recently affected the worldwide population, with an official disease start in December 2019[1]. Individuals develop their first symptoms after 2 to 14 days from infection and, from this very moment they constitute transmission vector of the disease thanks to a very efficient virus propagation mechanism, ultimately leading to the infection of several other subjects[2,3]. Most infected people do not present serious consequences induced by the disease, but about 20% of the cases will develop severe symptoms that may lead to hospitalisation, intensive care treatment and may even cause death[4].

In order to investigate deeper the disease causes and characteristics, the metabolic associated changes were studied analysing serum samples coming from hospitalized patients in the acute phase of the disease. The presented results are partially based on our publication "SARS-CoV-2 Infection Dysregulates the Metabolomic and Lipidomic Profiles of Serum"[5].

5.1.1 *COVID* and *preCOVID* cohorts description

A cohort of 263 COVID-19 (named here *COVID* cohort) patients, previously described in Chapter 3, was studied, using NMR spectroscopy, in order to identify the possible metabolic and lipidomic changes in the serum induced by this disease. Serum samples were recollected at their hospital arrival and all the subjects included into the study were presenting COVID-19 related symptoms (listed in **Table A9** in the Appendix) and tested positive after RT-PCR test for the detection of the RNA particles of SARS-CoV-2 virus on nasal swab samples.

As a collaboration with the laboratory of Cancer Immunology and Immunotherapy at CIC bioGUNE, a subset of the COVID-19 patient cohort (n = 43) was also tested for the presence of specific antibodies against the virus. Serum antibodies like the immunoglobulin M (IgM) or IgG against the spike and nucleocapsid (N) structural proteins of the virus, start to be produced from 1 to 3 weeks after the infection as a part of the natural immune response [6]. IgM are the first ones to be synthesized, with a shorter average life, while IgG, generally produced just 2-3 days after, evidenced a longer

duration, ensuring long-term immunity[7]. From the 43 individuals analysed for the presence of antibodies from the *COVID* cohort, 21 samples (48%) evidenced the presence of IgG and just 11 samples (26%) were showing both IgG and IgM. In total, almost the 75% of the tested individuals were already presenting antibodies. This is not totally surprising since it is well-known that during the first week many individuals do not yet produce sufficient antibodies, whereas these gradually increase from the second week, raising the sensitivity of their detection in the conduced tests[8]. Indeed, all patients included into this study well represented the most acute phase of the disease, between approximately 14 and 21 days from the infection.

280 healthy subjects were included into the study in order to compare the *COVID* patients with a control cohort. These serum samples (named here as *preCOVID* cohort) were recollected during 2018-2019, before the start of the pandemic, from the working population of the Basque Country, during the yearly medical check-up. The *preCOVID* cohort was built selecting subjects of the OSARTEN cohort, balancing the number of serum samples with the *COVID* cohort. As previously described in Chapter 3, no particular exclusion criteria were used to include samples into the OSARTEN cohort, a part from being affected by a serious disease (like cancer) or having suffered a stroke in the 3 months preceding the sample donation.

5.1.2 SARS-CoV-2 infection alters the metabolic profile of patients

Different NMR experiment were recorded in order to investigate the serum metabolism of the donors. First, a ^{1}H NOESY was collected under quantitative conditions for all the samples under consideration. Visual inspection of the ^{1}H NOESY spectra already evidence the existence of differences between *COVID* and *preCOVID* samples, as shown in **Figure 5.1**.

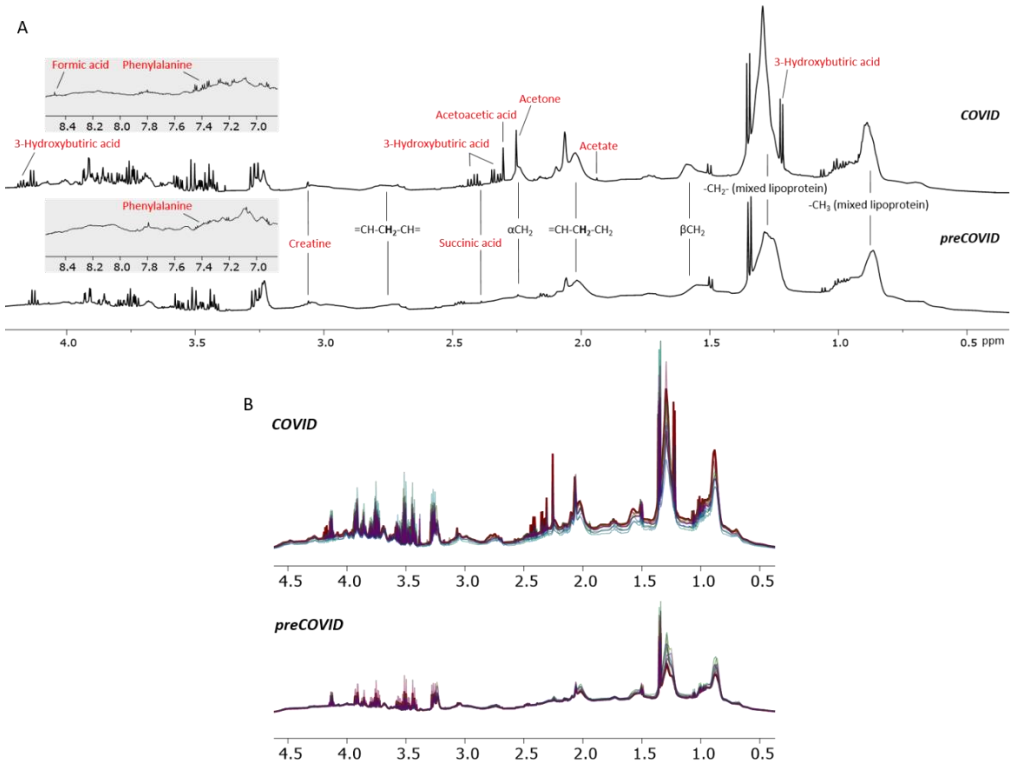

Figure 5.1: A) Metabolites and lipoproteins identification of two representing ^{1}H NOESY serum spectra from a *COVID* and *preCOVID* individual. **B**) NOESY experiment overlap of a selected subset of serum samples for the comparison between *COVID* and *preCOVID* cohorts.

Several metabolites and lipoproteins could be identified in the examined spectra, as evidenced in **Figure 5.1A**. Thanks to the Bruker B.I.Quant-PS^{TM} report, the quantification of up to 41 metabolites was done on each measured sample and, the PCA of these quantified metabolites (**Figure 5.2A**), showed a substantial separation between the *COVID* and *preCOVID* subjects confirming the existing differential metabolites expression between the two cohorts. **Figure 5.2B** evidenced the 10 most relevant metabolites which had the greatest impact to the first two components separation in the PCA. The direction of the arrows represented the distribution of their weight in the two components, while the colour highlighted the percentage of their contribution.

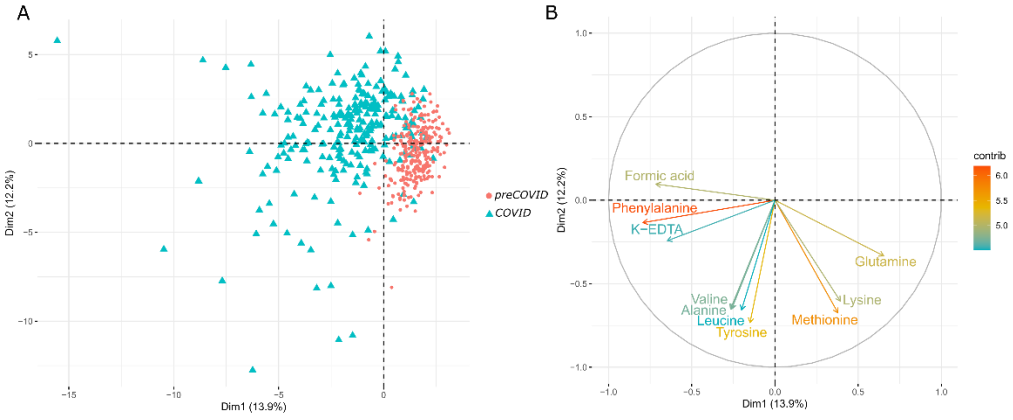

Figure 5.2 A) Score plot of the principal components from the PCA analysis of serum samples. Blue triangles represents the *COVID* cohort while the red dots the *preCOVID* one. **B**) Loading plots of the top 10 variables with the highest contribution in the PCA of serum metabolites. Arrows direction indicate the weight of the metabolite in the two components while the colour evidence the percentage of their contribution.

The 41 metabolites quantified by the Bruker report were examined by univariate analysis to confirm the different concentration of small molecules within the two cohorts of samples (**Figure 5.3**). In fact, several metabolites showed an altered concentration with respect to the control cohort, evidencing the dysregulated metabolism caused by SARS-CoV-2 in the serum after infection, possibly induced by the progression in viral generation and replication[9]. In particular, an increase in the production of ketone bodies was observed in the *COVID* cohort as compared to the *preCOVID* one. Acetoacetic acid (from 1.14 x 10^{-2} to 5.54 x 10^{-2} mmol/L; $p < 0.0001$; 385%), acetone (from 2.75 x 10^{-2} to 6.45 x 10^{-2} mmol/L; $p < 0.0001$; 134%) and 3-hydroxybutiric acid (from 6.6 x 10^{-2} to 2.7 x10^{-1} mmol/L; $p < 0.0001$; 302%) were considerably elevated. Fasting conditions increase the production of ketone bodies, which are principally sensitized by mitochondria in liver cells from Acetyl-CoA, derived from fatty acids oxidation[10]. The examined samples were recollected under uncontrolled fasting condition, as the blood extraction was carried out on patients upon their arrival at the hospital due to their serious health state. That said, *preCOVID* samples were taken at fasting. Taking this into account, the difference in the expression of the ketone bodies shall be strictly related to the presence of this disease. In fact, it was observed that, the condition of anorexia and fasting are normal responses of the

individual due to the presence of acute infections, which lead to the production of ketone bodies that have also evidenced an impact on the immune response, inflammation and oxidative stress[11–13]. The accumulation of these metabolites was evidenced in other studies that associate their production to COVID patients' severity, showing their increase in mild and severe cases, with a negative correlation with the hospitalization time and the mortality rate[11,14].

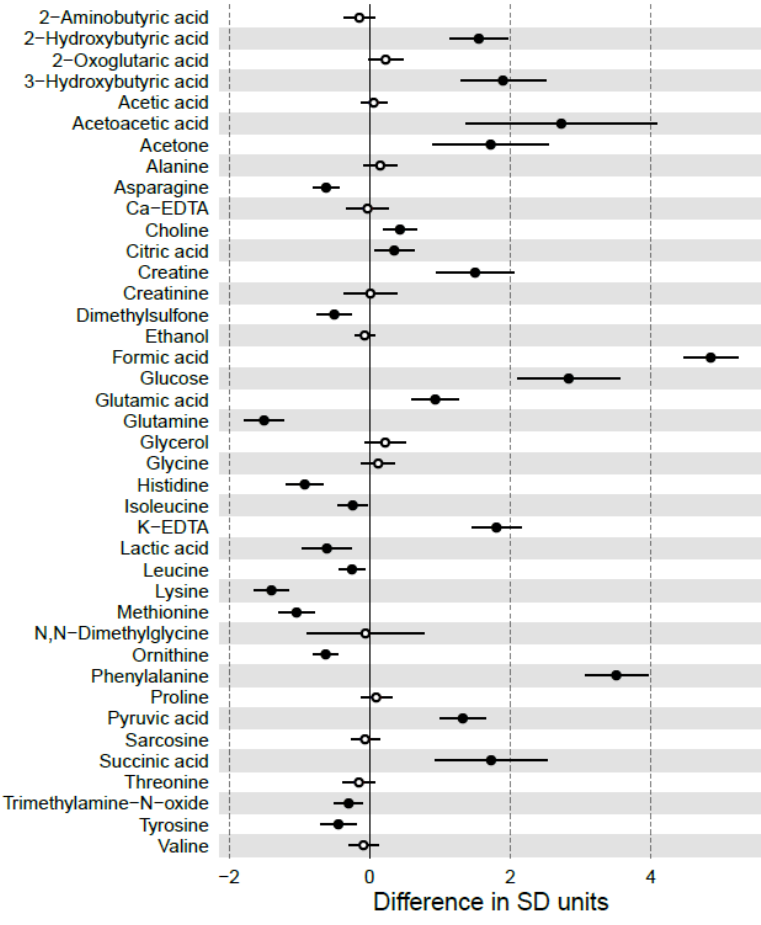

Figure 5.3: Effect of COVID-19 on metabolites expression. The horizontal axis indicates the number of standard deviations increase or decrease on average for each variable in *COVID* individuals. Circles indicate the specific mean increase or decrease value, and the horizontal lines represented the 95% confidence interval. The filled circles (totally black) represent the variables with a statistically significant difference (p-value < 0.05).

A significant increase in glucose levels was also present in COVID-19 patients (of 8.19 versus 4.89 mmol/L, $p < 0.0001$, 68%). This was consistent with what was observed also by other studies [15–17] and suggest that this increase in glucose concentration in the serum of the patients can be associated with a diabetic or pre-diabetic state and could also contribute to several other comorbidities. Moreover, glucose evidenced to be a predictor of severity, increasing the risk of fatal outcomes[18]. Finally, different studies suggested its association with the acute respiratory distress syndrome (ARDS) consisting in an increased release of immune system cells that leads to organ failure[19–22]. In fact, increased glucose levels are related to a higher viral replication rate and cytokine release[23]. Despite this, the mechanisms related to the higher serum concentration of this metabolite in COVID-19 patients is still unclear, but it was hypothesised that the presence of the ACE2 receptor in the pancreatic islets cells could lead to their infection giving rise to hyperglycaemia[17,24].

Other metabolites like succinic acid (156%), citric acid (12%), glutamic (33%) and pyruvic acid (67%) showed to be increased in COVID-19 patients. The high concentration of these metabolites in infected individuals evidenced a dysregulation of the central carbon metabolism. In fact, this is involved in purine biosynthesis which is necessary to fulfil the viral demand for high RNA replication[25]. On the other hand, glutamine evidenced to be decreased (by 4%). This metabolite plays an equally important role from a metabolic point of view and, in fact, the glutamine/glutamate ratio was associated with several comorbidities like hypertension, obesity and cardiovascular diseases and it has also been considered as a biomarker for liver damage due to the possible influence of alpha-glutathione S-transferase, generally present in liver failure, on this ratio[16,26]. In fact, glutamine is involved in different metabolic processes like purine synthesis for DNA and RNA components, glutathione production and energy generation via TCA cycle through the production of glutamate, further converted into alpha-chetoglutarate[27–29]. Finally, glutamine availability is also controlled by skeletal muscles in catabolic processes[30].

Essential amino acids like leucine, isoleucine, lysine, histidine and methionine were decreased in the *COVID* cohort. The reason related to this is still unclear, but this could be associated to the weight loss typical of these hospitalized patients, that made

essential amino acids one of the main energy sources for muscles, reducing their serum levels[31].

On the other hand, other metabolites like phenylalanine and 2-hydroxyturic acid were significantly increased. In the first case, it has been hypothesized that the higher concentration of phenylalanine (by the 81%) can be related to the activation of the immune system and, more generally, to inflammation, as evidenced in previous studies[32,33]. As regard the increase in 2-hydroxybutiric acid, this was associated to ketoacidosis conditions, previously described also as characteristic of this disease, to glutathione synthesis or, more in general to oxidative stress conditions[32]. A previous work also used the so called "Fisher's ratio", calculated from the sum of valine, leucine and isoleucine divided by the sum of phenylalanine and tyrosine, as a tool for the determination of liver dysfunction[16,34]. In fact, the increase on aromatic amino acids, like phenylalanine, was associated to catabolic processes observed in hepatic fibrosis[35]. Altogether, the last described changes evidence a general metabolic stress induced by the SARS-CoV-2 infection.

5.1.3 Alterations in the lipoprotein composition found in SARS-CoV-2 infected patients

The lipoprotein composition of the serum samples of COVID-19 patients was also analysed in order to understand if SARS-CoV-2 infection was miss-regulating their expression. Employing the Bruker IVDr Lipoprotein Subclass Analysis B.I.LISA™ report, as previously explained in Chapter 3, the analysis of up to 112 lipoprotein parameters was possible by means of spectra deconvolution. This included the quantification of different lipoproteins and the related lipoprotein subfraction like the phospholipids, cholesterol, triglycerides and apolipoproteins like Apo-A1, Apo-A2 and Apo-B.

A PCA comparing the *COVID* and *preCOVID* cohorts for the investigated lipoprotein parameters showed a reasonable separation between the two groups of samples evidencing the differential expression of these macromolecules in the serum of the

infected individuals and, the lipoproteins that mainly contribute to this separation are highlighted in **Figure 5.4**.

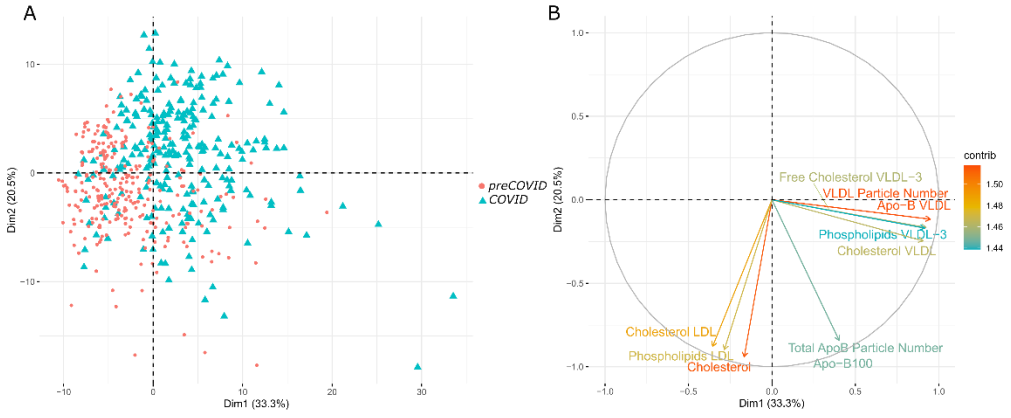

Figure 5.4: A) Score plot of the principal components from the PCA analysis of serum samples. Blue triangles represent the *COVID* cohort while the red dots the *preCOVID* one. **B)** Loading plots of the top 10 variables with the highest contribution in the PCA of serum lipoproteins. Arrows direction indicate the weight of the lipoproteins in the two components while the colour evidence the percentage of their contribution.

To investigate lipoprotein composition in more detail, univariate analysis was conducted for this dataset, as previously done for the metabolites. As shown in **Figure 5.5**, COVID-19 patients evidenced a marked increase in triglycerides, especially in the mean concentration of TG-HDL, TG-IDL, TG-LDL (the latter increasing by a factor of 2) and TG-VLDL. On the other hand, a decrease in the total cholesterol (TC, including the cholesterol and cholesteryl esters) was also observed, mainly in its major carriers TC-LDL and TC-HDL, for the latter especially in its subfractions 3 and 4, while TC-VLDL and TC-IDL were slightly increased.

The phospholipid profile, when examined in the main lipoprotein classes, showed a behaviour similar to the total cholesterol one with a decrease in the HDL and LDL components and an increase in IDL and VLDL ones.

In summary, the described results evidence an altered lipoprotein expression and a severe dyslipidemic profile induced upon SARS-CoV-2 infection, with an increase in triglycerides and especially in VLDL particles, particularly in the VLDL mean size subclasses, and a reduction in the HDL mean size subfractions.

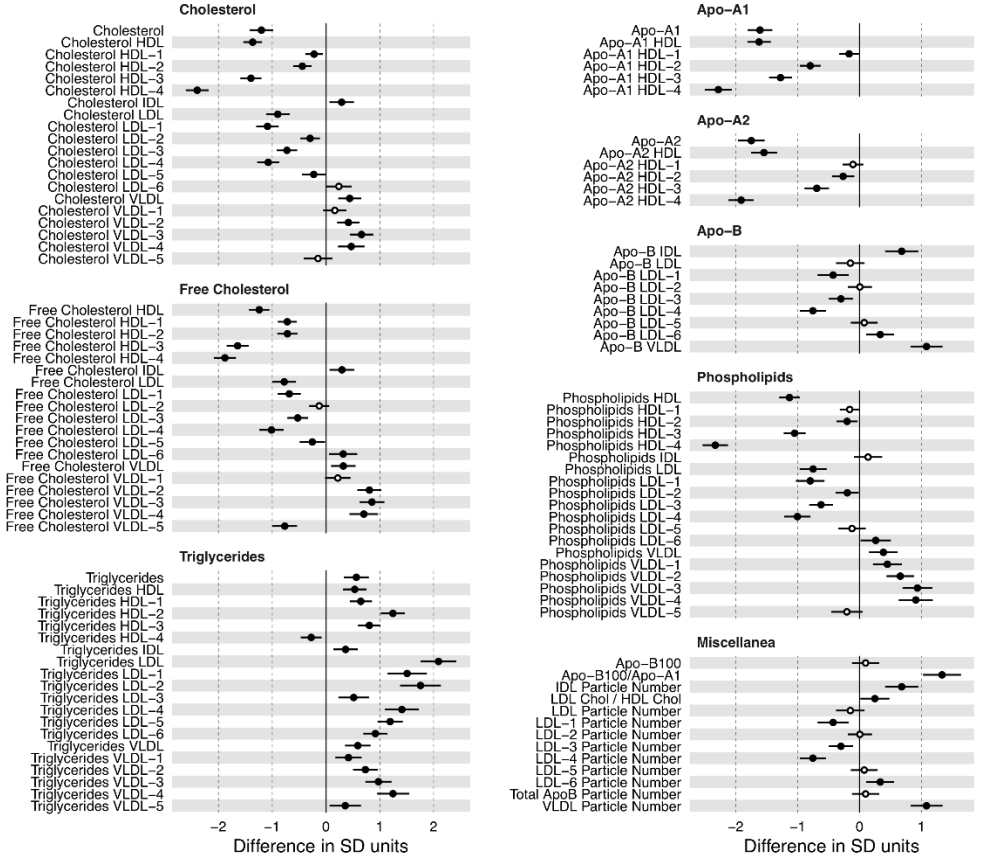

Figure 5.5: Effect of SARS-CoV-2 infection on lipoproteins and lipoprotein subclasses expression. The horizontal axis indicates the number of standard deviations increase or decrease on average for each variable in *COVID* individuals. Circles indicate the specific mean increase or decrease value, and the horizontal lines represented the 95% confidence interval. The filled circles (totally black) represented the variables with a statistically significant difference (p-value < 0.05).

These characteristic changes, observed only in COVID-19 patients, again point towards the presence of liver damage, as previously suggested from the differences in the expression of some metabolites like the essential amino acid or the glutamine to glutamate ratio. In fact, higher levels of VLDL lipoproteins are consistent with an increase in their production and reduced clearance, probably related to insulin resistance problems which could cause the increase in glucose concentrations previously mentioned and observed also in other studies[16]. It has been reported that all these alterations may be associated with the ongoing infection that caused a significant increase in the immune response which, in turn, influenced lipid expression reducing

cholesterol production and the clearance in triglycerides-rich lipoproteins[36,37]. Moreover, the observed triglycerides accumulation can be associated with a reduced hepatic ability in the mitochondria for the oxidation of acetyl-CoA, which was in turn used for the ketone bodies synthesis, previously described. The altered levels of succinic and pyruvic acid, observed before, are also consistent with the proposed mitochondrial dysfunction and/or the impaired central metabolism observed.

Additionally, our group observed an increase expression of porphyrins levels in COVID-19 patients, specifically in Uroporphyrin I (URO I), Coproporphyrin I and III (COP I and III). This is evidenced also in thrombocytopenia and in porphyria disease, which was associated with mitochondrial impairment, liver damage and oxidative stress[38–40].

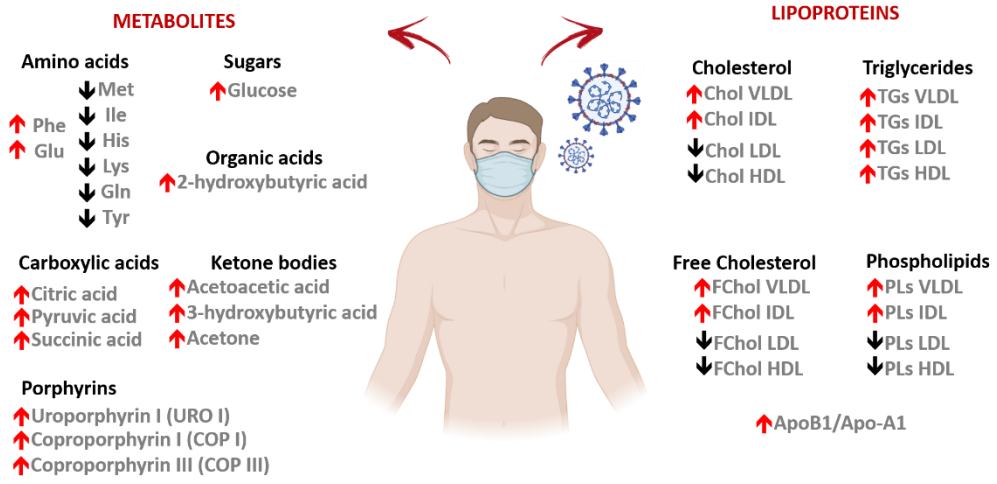

Figure 5.6: Graphical summary of the observed metabolites and lipoproteins differential expression in COVID-19 patients. Phe, phenylalanine; Glu, glutamate; Met, methionine; Ile, isoleucine; His, histidine; Lys, lysine; Gln, glutamine; Tyr, tyrosine; Chol, cholesterol; FChol, free cholesterol; TGs triglycerides; PLs, phospholipids.

Finally, apolipoproteins were also analysed and differences in their expression were observed. Apo-A1 and Apo-A2 are decreased in the *COVID* cohort, showing a similar profile. Both these lipoproteins can be found in high density lipoproteins, especially Apo-A1, which oversees the elimination of the excess of peripheral cholesterol leading it to the liver with the so-called "reverse cholesterol transport"[41]. On the contrary, Apo-B does the opposite work, and its generally associated to VLDL particles[42]. The Apo-

B100 to Apo-A1 ratio, which indicated the ratio between the atherogenic and anti-atherogenic particles respectively, was highly increased (by 2-fold) in COVID-19 patients, evidencing a significant risk for cardiovascular events.

Altogether, the observed changes (summarised in **Figure 5.6**) evidenced a severe pathogenic state with an increased atherogenic risk which could lead patients to severe outcomes[37].

5.1.4 A metabolic discrimination model for COVID-19 patients.

The observed changes between the *COVID* and *preCOVID* cohorts described until now based on the unsupervised and univariate analysis, were further corroborated by OPLS-DA analysis with one predictive and one orthogonal component, performed using the complete set of quantified metabolites and lipoproteins subclasses (**Figure 5.7**). As expected, the comparison between the *COVID* and *preCOVID* cohorts presented a high separation degree, statistically significant (p-value < 0.01) and with good predictability ($AUROC_{validation}$ = 0.9777, look **Table A13** in the Appendix). **Figure 5.7B** showed the loading plot obtained from the OPLS-DA analysis in **Figure 5.7A**, evidencing the set of variables clusters (lipoproteins and metabolites) that most contributed to the separation of the two cohorts of analysed samples.

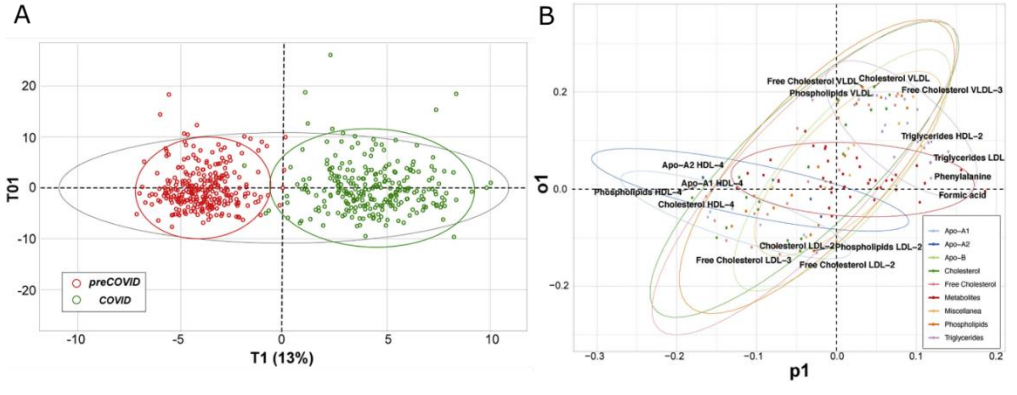

Figure 5.7: (**A**) OPLS-DA score plot of the comparison between *COVID* (green) and *preCOVID* (red) cohorts for the full set of metabolites and lipoprotein subclasses. The main component versus the first orthogonal component are shown. (**B**) Loading plot of the OPLS-DA. Each variable (metabolites or the lipoprotein subclasses) is represented with different colours. 95% of the members of each type are surrounded by ellipses and the four variables that most contribute to the component were labelled for each direction.

Once again, the variables previously identified with the univariate analysis were the ones used in the discrimination model to analyse the two cohort of samples. As regard metabolites, formic acid and phenylalanine were the most increased ones in *COVID* cohort. As previously mentioned, formic acid is one of the most important metabolites in the one carbon metabolism which appeared to be miss regulated, while phenylalanine was associated with inflammation and COVID-19 severity[43,44].

As regard lipoproteins, some characteristics component among triglycerides, cholesterol, phospholipids and VLDL particles were the main responsible for the differences between the analysed cohorts, as shown before by univariate analysis, confirming once again a characteristic dyslipidemia associated to COVID-19 patients.

5.1.5 Inflammation markers

We then attempted to obtain information about inflammation markers from the ^{1}H-NMR spectra. Indeed, looking at the NMR spectra of the analyzed samples, it was possible to observe a great intensity difference between the *COVID* and *preCOVID* cohorts in the α-1-acid glycoprotein A (Glyc A) signal which was significantly increased in the SARS-CoV-2 positive patients, as shown in **Figure 5.8**.

This characteristic signal, firstly identified in 1987 by *Sadler at al.*[45], originates from the acetylation of different glycoproteins, specifically from the acetyl group of *N*-acetylglucosamine, *N*-acetylgalactosamine and *N*-acetylneuraminic acid[46]. Its higher intensity in the *COVID* cohort postulates it as a biomarker for the acute phase of COVID-19, in line with what was observed also by other studies[16,45]. In fact, altered glycosylation patterns have been observed during inflammation processes associated not only with COVID-19, but also with other severe diseases like cancer, cardiovascular or metabolic disorders[47–49].

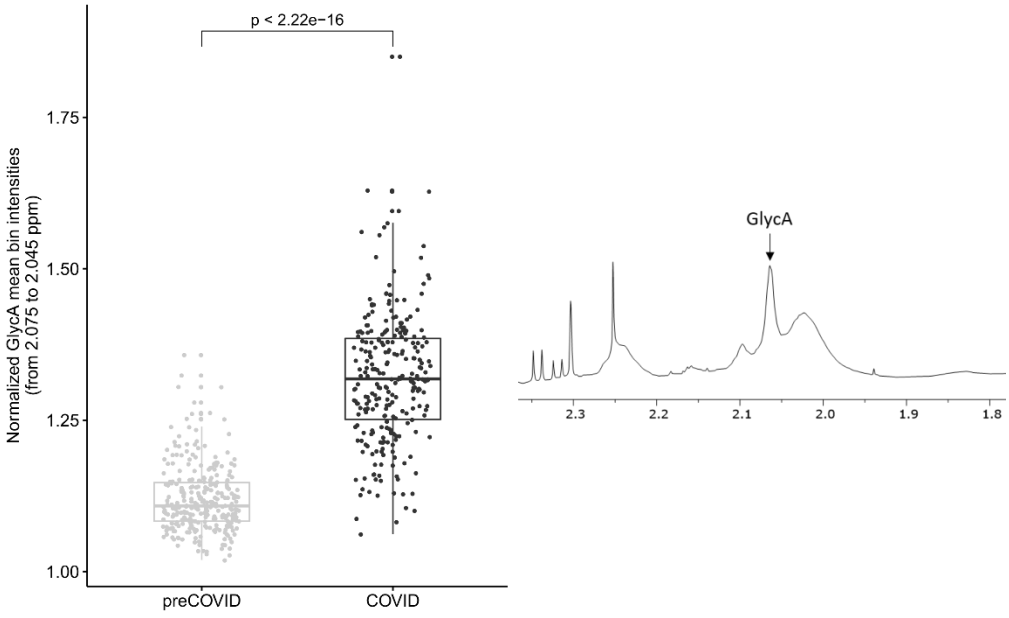

Figure 5.8: Glyc-A expression in *COVID* and *preCOVID* samples. Signal was integrated for the measured serum samples in order to compare its level in the two cohorts. On the right, representative ^{1}H NMR spectral region evidencing the studied signal in a serum sample.

5.1.6 Study limitations and potential caveats analysis

5.1.6.1 Age limitation

Different studies evidenced that COVID-19 severity is directly correlated with age[50–52]. This is mainly due to the onset of comorbidities in elderly subjects, that aggravate the prognosis of infected patients [53–55]. To investigate if the observed changes in the expression of metabolites and lipoproteins were corelated with age, two sub-cohort of 112 samples were analysed (additional information in **Table A10** in the Appendix). These two "new" sub-cohorts were created to be perfectly balanced for age and gender distribution. **Figure 5.9** shows an equivalent qualitative separation between the two cohorts for both metabolites (**Figure 5.9A**) and lipoproteins (**Figure 5.9B**) in the PCA and in the OPLS-DA (**Figure 5.9C** and **D**) as compared to the full cohorts, confirming that age is not a compromising factor for the obtained results.

5.1.6.2 Sample recollection conditions

Another potential concern about the conduced study was the heterogeneous conditions in which samples were recollected during the drastic situation that hospitals had to face during the first wave of the COVID-19 pandemic.

The importance of a proper adherence to specific standard operating procedures throughout the sample handling process was explained extensively in the introduction. Unfortunately, we could not ensure that samples taken in hospitals followed all the necessary standards to guarantee secondary contamination-free results. For this reason, we wanted to confirm that the reported observations were not partially related to sample mishandling. To that end, a second cohort of samples of almost 400 patients, coming from a different hospital, was analysed and the results for the expression of lipoproteins and metabolites were compared with the ones of the comparison between *COVID* and *preCOVID*.

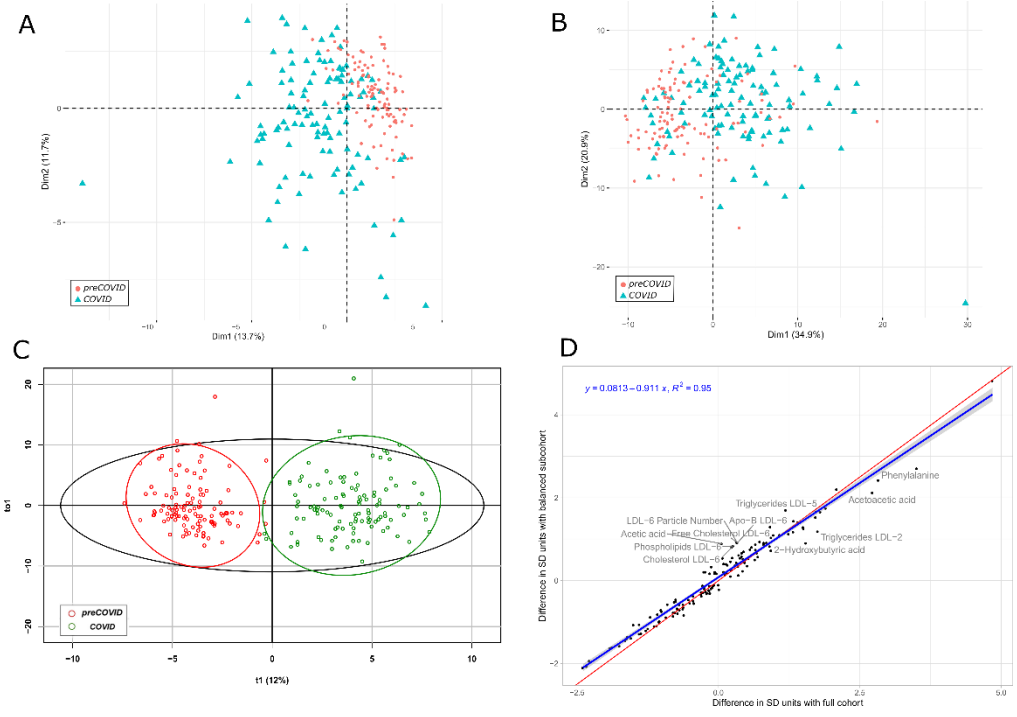

Figure 5.9: Sub-cohorts analysis for age caveat. **A**) PCA analysis for the serum metabolites and **B**) lipoproteins, representing the first two components. **C**) Score plot of the OPLS-DA for the full set of metabolites and lipoproteins. **D**) Effect of COVID-19 over the metabolites and lipoprotein subclasses in the new sub-cohorts and the full cohorts. Blu line is the linear model of both analyses and the red one is the identity. Samples with absolute difference higher than 0.5 between both analyses are labelled.

The comparison between the validation cohort with the *preCOVID* one reported equivalent results, as shown in **Figure 5.10**. This result suggests that the differences in the samples metabolomic and lipidomic profile of the infected patients arise from the SARS-CoV-2 infection and not from handling issues.

Moreover, thanks to the analysis of this large quantity of samples, the integration of all cohorts (*COVID*, *preCOVID* and validation) converted our study into one of the largest metabolic analyses carried out for the study of this disease.

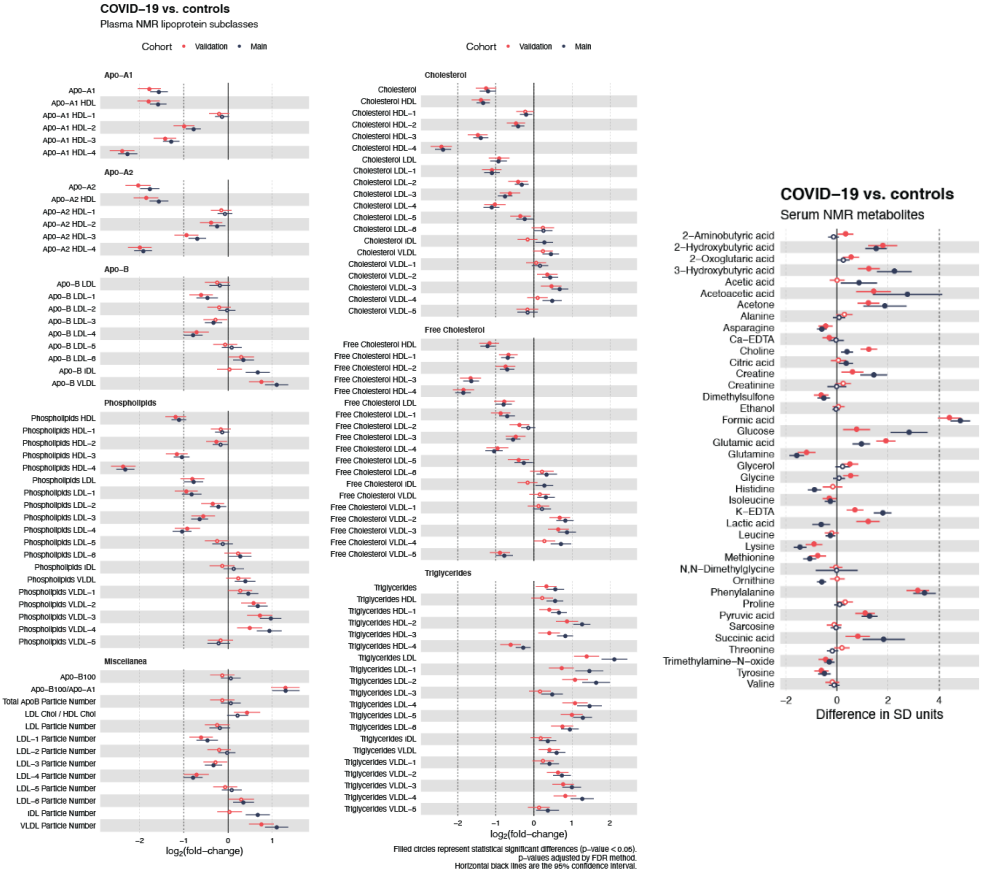

Figure 5.10: Validation cohort analysis coming from a different hospital. The validation cohort was compared against the *preCOVID* one and the average effect of COVID-19 for the lipoprotein and metabolites expression was evidenced comparing it to the results of the *COVID* cohort. The horizonal axis shows the standard deviation of the increase or decrease of each variable in COVID-19 patients and the circles are positioned in the mean increase/decrease. Filled circles represent the statistically significant variables (p-value < 0.05). Black bars represent the 95% confidence interval.

5.1.6.3 Storage stability over time

Samples included in the *preCOVID* cohort were recollected long before the onset of COVID-19 pandemic (2 years before). As already mentioned for the recollection conditions of the samples, storage time could also have an effect on the serum metabolome[56]. To investigate such effect on the sample's stability, a PCA was performed to compare the first samples recollected in 2016 with the last ones obtained in 2018 (**Figure 5.11**).

As shown in **Figure 5.11**, it was not possible to separate the first recollected samples (2016) from the last ones (2018), indicating that the freezing time at -80°C does not significantly alter the nature of the samples. The same results were observed also in previous works conducted analysing quality control samples (QC) over time, also present in our routine analysis in order to carefully monitor the conducted studies[57,58].

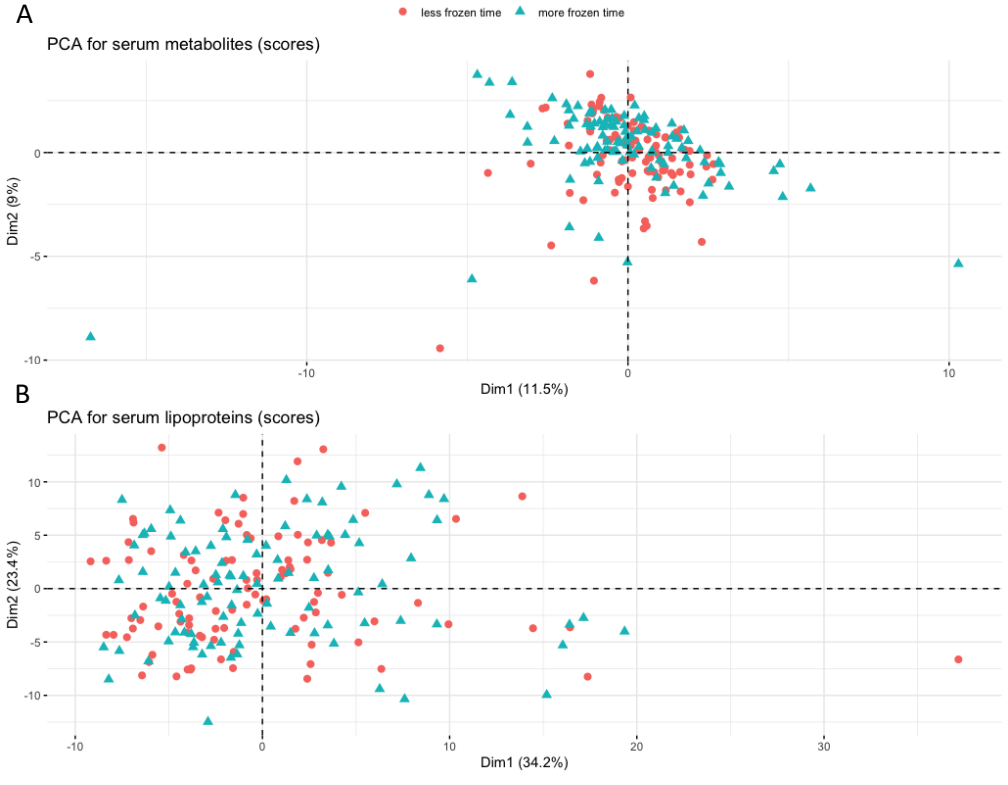

Figure 5.11: PCA of the full set of **A**) metabolites **B**) lipoproteins of the two subset of control samples recollected in 2016 and 2018, at 400 days of difference for the freezing period.

5.1.7 Final considerations

The spread out of SARS-CoV-2 infection and the evidenced differential phenotypic response induced by this disease, highlighted the need to investigate it more profoundly using the "omics" techniques, particularly used in the field of precision medicine, as previously explained also in Chapter 1.

Metabolomics emerged as a powerful tool for the investigation o the characteristics metabolic changes that affect subjects infected by SARS-CoV-2, leading to the identification of biomarkers for the diagnosis and the prognosis of the disease.

Our results demonstrated to be in line with the ones obtained by others research groups, employing NMR and/or MS, highlighting the damage caused by this disease not only in the respiratory tract, but also involving several other organs such as kidneys, liver and the cardiovascular system[16,59–62]. These studies evidenced a similar metabolic and lipidomic signature characterized by an increase in glucose, glutamate, phenylalanine, formic acid and ketone bodies, as regard metabolites, and in VLDL and IDL when considering lipoproteins. On the other hand, they agreed in the observation of a reduced level of glutamine, lactic acid, histidine and in LDL and HDL components[59]. Moreover, some of these changes, were also observed in previous studies on SARS-CoV-1 and MERS-CoV, which evidenced liver damage and changes in the lipid's levels [63,64]. Finally, the extreme inflammatory response that characterize COVID-19 patients was investigated in order to find the association between inflammatory markers like interleukin-18 (IL-18), IL-6, interferon gamma (IFN-γ), interferon gamma inducible protein-10 (IP-10) and RANTES (Regulated on Activation, Normal T Expressed and Secreted, also known as CCL5) with the observed metabolic changes. A positive and negative correlation was observed with some LDL and HDL subfractions respectively[65].

Further investigations have been trying to find an association between the identified biomarkers and COVID-19 severity. A differential expression in some metabolites levels was observed between mild and severe patients, especially in kynurenine pathway and lipids such as carnitines and phosphatidylcholine[66–71]. These important changes, associated with the worsening of the disease symptoms, were also correlated

with specific markers of the immune response (i.e. chemokines, IL-6)[65,72]. This could have a predictive power for the prognosis of the infected subjects, improving medical interventions[72].

Metabolomics has also been employed to investigate patient's recovery. Infected subjects were followed for several months after hospital discharge (in some cases, even more than 1 year), with multiples check-ups to study their metabolic normalization, the so-called "phenoreversion"[73,74]. It was observed that this may take also several months, depending above all on the severity of the disease suffered and the associated respiratory problems[39,73,75,76].

Finally, there is a growing interest in those cases where patients have not been able to fully recover a normal metabolism after the disease or have even developed a post-acute COVID-19 syndrome[77]. The long COVID effects are being investigated and, among the most important prognostic biomarkers, emerges lipoprotein's role[78].

Bibliography Chapter 5:

1. Zhu, N. *et al.* A Novel Coronavirus from Patients with Pneumonia in China, 2019. *New England Journal of Medicine* **382**, 727–733 (2020).
2. Munster, V. J., Koopmans, M., van Doremalen, N., van Riel, D. & de Wit, E. A Novel Coronavirus Emerging in China — Key Questions for Impact Assessment. *New England Journal of Medicine* **382**, 692–694 (2020).
3. Lauer, S. A. *et al.* The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Ann Intern Med* **172**, 577–582 (2020).
4. Richardson, S. *et al.* Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. *JAMA* **323**, 2052–2059 (2020).
5. Bruzzone, C. *et al.* SARS-CoV-2 Infection Dysregulates the Metabolomic and Lipidomic Profiles of Serum. *iScience* **23**, (2020).
6. Röltgen, K. *et al.* Defining the features and duration of antibody responses to SARS-CoV-2 infection associated with disease severity and outcome. *Sci Immunol* **5**, (2020).
7. Ma, H. *et al.* COVID-19 diagnosis and study of serum SARS-CoV-2 specific IgA, IgM and IgG by chemiluminescence immunoanalysis. *medRxiv*, 20064907 (2020).
8. Ejazi, S. A., Ghosh, S. & Ali, N. Antibody detection assays for COVID-19 diagnosis: an early overview. *Immunol Cell Biol* **99**, 21–33 (2021).
9. Wang, T. *et al.* COVID-19 metabolism: Mechanisms and therapeutic targets. *MedComm (Beijing)* **3**, (2022).
10. Kumari, A. Beta Oxidation of Fatty Acids. *Sweet Biochemistry* 17–19 (2018).
11. Valdés, A. *et al.* Metabolomics study of COVID-19 patients in four different clinical stages. *Scientific Reports 2022 12:1* **12**, 1–11 (2022).
12. Puchalska, P. & Crawford, P. A. Multi-dimensional Roles of Ketone Bodies in Fuel Metabolism, Signaling, and Therapeutics. *Cell Metab* **25**, 262–284 (2017).
13. Karagiannis, F. *et al.* Impaired ketogenesis ties metabolism to T cell dysfunction in COVID-19. *Nature* **609**, 801–807 (2022).
14. Li, J. *et al.* COVID-19 infection may cause ketosis and ketoacidosis. *Diabetes Obes Metab* **22**, 1935–1941 (2020).

15. Danaei, G. *et al.* National, regional, and global trends in fasting plasma glucose and diabetes prevalence since 1980: Systematic analysis of health examination surveys and epidemiological studies with 370 country-years and 2·7 million participants. *The Lancet* **378**, 31–40 (2011).

16. Kimhofer, T. *et al.* Integrative Modeling of Quantitative Plasma Lipoprotein, Metabolic, and Amino Acid Data Reveals a Multiorgan Pathological Signature of SARS-CoV-2 Infection. *J Proteome Res* **19**, 4442–4454 (2020).

17. Bode, B. *et al.* Glycemic Characteristics and Clinical Outcomes of COVID-19 Patients Hospitalized in the United States. *J Diabetes Sci Technol* **14**, 813 (2020).

18. Wang, W. *et al.* Elevated glucose level leads to rapid COVID-19 progression and high fatality. *BMC Pulm Med* **21**, (2021).

19. Grasselli, G. & Zanella, A. Critically ill patients with COVID-19 in New York City. *Lancet* **395**, 1740 (2020).

20. Cummings, M. J. *et al.* Epidemiology, clinical course, and outcomes of critically ill adults with COVID-19 in New York City: a prospective cohort study. *Lancet* **395**, 1763 (2020).

21. Zhou, F. *et al.* Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* **395**, 1054 (2020).

22. Wu, C. *et al.* Risk Factors Associated With Acute Respiratory Distress Syndrome and Death in Patients With Coronavirus Disease 2019 Pneumonia in Wuhan, China. *JAMA Intern Med* **180**, 1 (2020).

23. Codo, A. C. *et al.* Elevated Glucose Levels Favor SARS-CoV-2 Infection and Monocyte Response through a HIF-1α/Glycolysis-Dependent Axis. *Cell Metab* **32**, 437 (2020).

24. Yang, J. K., Lin, S. S., Ji, X. J. & Guo, L. M. Binding of SARS coronavirus to its receptor damages islets and causes acute diabetes. *Acta Diabetol* **47**, 193 (2010).

25. Perła-Kaján, J. & Jakubowski, H. COVID-19 and One-Carbon Metabolism. *Int J Mol Sci* **23**, 23 (2022).

26. Matsuyama, T., Yoshinaga, S. K., Shibue, K. & Mak, T. W. Comorbidity-associated glutamine deficiency is a predisposition to severe COVID-19. *Cell Death Differ* **28**, 3199 (2021).

27. Curi, R. *et al.* Molecular mechanisms of glutamine action. *J Cell Physiol* **204**, 392–401 (2005).

28. Li, X. *et al.* Dysregulation of glutamine/glutamate metabolism in COVID-19 patients: A metabolism study in African population and mini meta-analysis. *J Med Virol* (2022) doi:10.1002/jmv.28150.

29. Xiao, D. *et al.* The glutamine-alpha-ketoglutarate (AKG) metabolism and its nutritional implications. *Amino Acids 2016 48:9* **48**, 2067–2080 (2016).

30. Cruzat, V., Rogero, M. M., Keane, K. N., Curi, R. & Newsholme, P. Glutamine: Metabolism and Immune Function, Supplementation and Clinical Translation. *Nutrients* **10**, (2018).

31. Atila, A. *et al.* The serum amino acid profile in COVID-19. *Amino Acids* **53**, 1569–1588 (2021).

32. Shi, D. *et al.* The serum metabolome of COVID-19 patients is distinctive and predictive. *Metabolism* **118**, 154739 (2021).

33. Huang, S. S. *et al.* Phenylalanine- and leucine-defined metabolic types identify high mortality risk in patients with severe infection. *International Journal of Infectious Diseases* **85**, 143–149 (2019).

34. Fischer, J. *et al.* The effect of normalization of plasma amino acids on hepatic encephalopathy in man. *Surgery* **80**, 77–91 (1976).

35. Nakamura, Y. *et al.* Cluster analysis of indicators of liver functional and preoperative low branched-chain amino acid tyrosine ration indicate a high risk of early recurrence in analysis of 165 hepatocellular carcinoma patients after initial hepatectomy. *Surgery* **150**, 250–262 (2011).

36. Mead, J. R., Irvine, S. A. & Ramji, D. P. Lipoprotein lipase: structure, function, regulation, and role in disease. *J Mol Med (Berl)* **80**, 753–769 (2002).

37. Masana, L. *et al.* Low HDL and high triglycerides predict COVID-19 severity. *Scientific Reports* | **11**, 7217 (123AD).

38. San Juan, I. *et al.* Abnormal concentration of porphyrins in serum from COVID-19 patients. *Br J Haematol* (2020).

39. Bizkarguenaga, M. *et al.* Uneven metabolic and lipidomic profiles in recovered COVID-19 patients as investigated by plasma NMR metabolomics. *NMR Biomed* **35**, 1–10 (2022).

40. Connors, J. M. & Levy, J. H. COVID-19 and its implications for thrombosis and anticoagulation. *Blood* **135**, 2033–2040 (2020).

41. Marcovina, S. & Packard, C. J. Measurement and meaning of apolipoprotein AI and apolipoprotein B plasma levels. *J Intern Med* **259**, 437–446 (2006).

42. Behbodikhah, J. *et al.* Apolipoprotein B and Cardiovascular Disease: Biomarker and Potential Therapeutic Target. *Metabolites* **11**, 690 (2021).

43. Pietzke, M., Meiser, J. & Vazquez, A. Formate metabolism in health and disease. *Mol Metab* **33**, 23–37 (2020).

44. Luporini, R. L. *et al.* Phenylalanine and COVID-19: Tracking disease severity markers. *Int Immunopharmacol* **101**, 108313 (2021).

45. Bell, J. D., Brown, J. C. C., Nicholson, J. K. & Sadler, P. J. Assignment of resonances for 'acute-phase' glycoproteins in high resolution proton NMR spectra of human blood plasma. *FEBS Lett* **215**, 311–315 (1987).

46. Unione, L., Ardá, A., Jiménez-Barbero, J. & Millet, O. NMR of glycoproteins: profiling, structure, conformation and interactions. *Curr Opin Struct Biol* **68**, 9–17 (2021).

47. Fuertes-Martín, Correig, Vallvé & Amigó. Human Serum/Plasma Glycoprotein Analysis by 1H-NMR, an Emerging Method of Inflammatory Assessment. *J Clin Med* **9**, 354 (2020).

48. Gruppen, E. G., Connelly, M. A., Sluiter, W. J., Bakker, S. J. L. & Dullaart, R. P. F. Higher plasma GlycA, a novel pro-inflammatory glycoprotein biomarker, is associated with reduced life expectancy: The PREVEND study. *Clinica Chimica Acta* **488**, 7–12 (2019).

49. Suman, S., Sharma, R. K., Kumar, V., Sinha, N. & Shukla, Y. Metabolic fingerprinting in breast cancer stages through 1H NMR spectroscopy-based metabolomic analysis of plasma. *J Pharm Biomed Anal* **160**, 38–45 (2018).

50. Statsenko, Y. *et al.* Impact of Age and Sex on COVID-19 Severity Assessed From Radiologic and Clinical Findings. *Front Cell Infect Microbiol* **11**, 777070 (2022).

51. Liu, K., Chen, Y., Lin, R. & Han, K. Clinical features of COVID-19 in elderly patients: A comparison with young and middle-aged patients. *Journal of Infection* **80**, 14–18 (2020).

52. Davies, N. G. *et al.* Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nat Med* **26**, 1205–1211 (2020).

53. Zhou, F. *et al.* Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* **395**, 1054 (2020).

54. Yang, X. *et al.* Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. *Lancet Respir Med* **8**, 475–481 (2020).

55. Wang, J. *et al.* Clinical and CT findings of COVID-19: differences among three age groups. *BMC Infect Dis* **20**, (2020).

56. Wang, X. *et al.* Influence of Storage Conditions and Preservatives on Metabolite Fingerprints in Urine. *Metabolites* **9**, (2019).

57. Loo, R. L. *et al.* Quantitative In-Vitro Diagnostic NMR Spectroscopy for Lipoprotein and Metabolite Measurements in Plasma and Serum:

Recommendations for Analytical Artifact Minimization with Special Reference to COVID-19/SARS-CoV-2 Samples. *J Proteome Res* **19**, 4428–4441 (2020).

58. Pinto, J. *et al.* Human plasma stability during handling and storage: impact on NMR metabolomics. *Analyst* **139**, 1168–1177 (2014).

59. Schmelter, F. *et al.* Metabolic and Lipidomic Markers Differentiate COVID-19 From Non-Hospitalized and Other Intensive Care Patients. *Front Mol Biosci* **8**, (2021).

60. Shen, B. *et al.* Proteomic and Metabolomic Characterization of COVID-19 Patient Sera. *Cell* **182**, 59-72.e15 (2020).

61. Wu, D. *et al.* Plasma metabolomic and lipidomic alterations associated with COVID-19. *Natl Sci Rev* **7**, 1157–1168 (2020).

62. Schmelter, F. *et al.* Metabolic markers distinguish COVID-19 from other intensive care patients and show potential to stratify for disease risk. *medRxiv*, 21249645 (2021).

63. Kukla, M. *et al.* COVID-19, MERS and SARS with Concomitant Liver Injury—Systematic Review of the Existing Literature. *J Clin Med* **9**, 1420 (2020).

64. Wu, Q. *et al.* Altered Lipid Metabolism in Recovered SARS Patients Twelve Years after Infection. *Sci Rep* **7**, 1–12 (2017).

65. Lodge, S. *et al.* 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* (2021).

66. Caterino, M. *et al.* Dysregulation of lipid metabolism and pathological inflammation in patients with COVID-19. *Sci Rep* **11**, (2021).

67. Wu, J. *et al.* Lipidomic signatures align with inflammatory patterns and outcomes in critical illness. *Nature Communications 2022 13:1* **13**, 1–18 (2022).

68. D'alessandro, A. *et al.* Biological and clinical factors contributing to the metabolic heterogeneity of hospitalized patients with and without covid-19. *Cells* **10**, 2293 (2021).

69. Danlos, F. X. *et al.* Metabolomic analyses of COVID-19 patients unravel stage-dependent and prognostic biomarkers. *Cell Death Dis* **12**, (2021).

70. Albóniga, O. E. *et al.* Metabolic Snapshot of Plasma Samples Reveals New Pathways Implicated in SARS-CoV-2 Pathogenesis. *J Proteome Res* **21**, 623–634 (2022).

71. Dillard, L. R. *et al.* Leveraging metabolic modeling to identify functional metabolic alterations associated with COVID-19 disease severity. *Metabolomics* **18**, (2022).

72. Sindelar, M. *et al.* Longitudinal metabolomics of human plasma reveals prognostic markers of COVID-19 disease severity. *Cell Rep Med* **2**, (2021).

73. Lodge, S. *et al.* Diffusion and Relaxation Edited Proton NMR Spectroscopy of Plasma Reveals a High-Fidelity Supramolecular Biomarker Signature of SARS-CoV-2 Infection. *Anal Chem* **93**, 3976–3986 (2021).

74. Bruzzone, C., Conde, R., Embade, N., Mato, J. M. & Millet, O. Metabolomics as a powerful tool for diagnostic, pronostic and drug intervention analysis in COVID-19. *Front Mol Biosci* **10**, (2023).

75. Xu, J. *et al.* Plasma Metabolomic Profiling of Patients Recovered From Coronavirus Disease 2019 (COVID-19) With Pulmonary Sequelae 3 Months After Discharge. *Clinical Infectious Diseases* **2019**, 1–12 (2021).

76. Holmes, E. *et al.* Incomplete Systemic Recovery and Metabolic Phenoreversion in Post-Acute-Phase Nonhospitalized COVID-19 Patients: Implications for Assessment of Post-Acute COVID-19 Syndrome. *J Proteome Res* (2021).

77. Masuda, R. *et al.* Exploration of Human Serum Lipoprotein Supramolecular Phospholipids Using Statistical Heterospectroscopy in n-Dimensions (SHY- n): Identification of Potential Cardiovascular Risk Biomarkers Related to SARS-CoV-2 Infection. *Anal Chem* **94**, 4426–4436 (2022).

78. Bai, Y. *et al.* Lipidomic alteration of plasma in cured COVID-19 patients using ultra high-performance liquid chromatography with high-resolution mass spectrometry. *Biosci Rep* **41**, (2021).

Chapter 6
Conclusions

6.1 A molecular discrimination of the metabolic syndrome by urine and serum metabolomics

- Urine and serum NMR metabolomics are sensitive to MetS, due to the observed differential manifestation of the risk factors (RFs) in the urine and serum derived metabotype.
- New biomarkers associated with MetS were found and some of the previously related with MetS or with the associated risk factors have been validated.
- All the contributing RFs were represented by at least one metabolite that varied significantly, according to the metabolomic analysis of urine and serum samples.
- Disease progression was accompanied by a continuous variation (up- or down-regulation) of the identified and quantified metabolites as a function of the conditions, reflecting a more pronounced alteration in the subjects affected by MetS.
- Not all the RFs equally contribute to the progression to MetS. Urine analysis underlines the impact of the glucose metabolism alterations and of hypertension, while serum analysis highlights the impact of dyslipidaemia.
- Based on the analysis of the urine samples, a metabolic model was created for the discrimination between individuals with and without MetS and the assignment of a "MetS score" for the determination of how likely an individual is to develop MetS. The integration of the results obtained from serum analysis shall contribute to the creation to a more reliable model in order to improve the discrimination of the subjects affected by MetS and the assigned of the "MetS score".
- Aging and non-alcoholic fatty liver disease (NAFLD) were also considered as RFs that enhance MetS but they do not directly interfere with the metabolic discrimination of the syndrome. However, in the case of NAFLD, the simultaneous presence of this disorder and MetS could confound the metabolic definition of MetS due to a partial overlap between the risk factors that contribute to the development of MetS with NAFLD associated comorbidities.

- The obtained results generate an unprecedent molecular dimension to the definition of MetS.
- The present investigation may improve medical decision for an early intervention and treatment of MetS.

6.2 Metabolomic and Lipidomic dysregulation caused by SARS-CoV-2 infection

- Patients in the acute phase of the disease show marked changes in the metabolomic and lipidomic profiles of the analysed serum samples.
- GlycA signal, a biomarker for acute systemic inflammation and previously associated with cardiovascular diseases, is significantly increased in the COVID-19 patients, at least during acute phase.
- The observed metabolic changes, in addition to the known impairment of the respiratory system, are related with multiple-organ specific dysfunctions, highlighting the systemic character of the disease.
- Lipoprotein profile alteration and redistribution, characterized by a higher level of TG leading to an increase in VLDL subclasses with intermediate size, suggest a possible increase in the atherogenic risk and an impaired oxidative stress.
- The observed metabolic changes may improve medical treatment of the infected subjects, in a more personalized way.
- Further work is needed for the identification of specific severity-related biomarkers. This could improve the treatment of the patients affected by COVID-19 avoiding their aggravation and consequent long term health problems or pathologies such as “long COVID”.

Appendix

Table A1: General characteristics of the OSARTEN cohort for urine samples included into the study. Biochemical data were extracted from a blood test performed at the same time of urine samples collection. Additional information was achieved from questionnaires complementation.

	[ALL valid] ***N=9367***	**female** ***N=3432***	**male** ***N=5935***	**N**
Age (years)	43.04±9.16	44.06±8.73	42.45±9.35	9367
Weight (cm)	74.85±14.04	64.39±11.29	80.89±11.74	9367
Height (cm)	171.91±8.99	163.59±6.25	176.72±6.48	9367
BMI (kg/m²)	25.22±3.76	24.06±4.04	25.89±3.42	9367
Smoker	1877 (20.06%)	652 (19.03%)	1225 (20.66%)	9355
Alcohol consumption:				9345
Never	1325 (14.18%)	730 (21.33%)	595 (10.05%)	
Social drinker	6966 (74.54%)	2500 (73.04%)	4466 (75.41%)	
Only during meals	829 (8.87%)	171 (5.00%)	658 (11.11%)	
Daily intake of alcohol	225 (2.41%)	22 (0.64%)	203 (3.43%)	
Hypertension	985 (10.54%)	257 (7.50%)	728 (12.30%)	9343
Medicated for hypertension	86 (3.30%)	32 (3.04%)	54 (3.47%)	2610
Medicated for diabetes	70 (2.68%)	7 (0.66%)	63 (4.05%)	2609
Proteins detected in urine	126 (1.35%)	44 (1.29%)	82 (1.38%)	9359
ALT (U/L)	22.56±13.65	16.58±9.33	26.02±14.53	9367
Basophils (10⁹/L)	0.04±0.02	0.04±0.02	0.04±0.02	9367
Cholesterol (mg/dL)	193.49±34.48	191.87±34.34	194.42±34.52	9367
HDL Cholesterol (mg/dL)	61.19±15.98	68.97±15.41	56.68±14.51	9367
LDL Cholesterol (mg/dL)	112.48±30.82	106.85±30.32	115.94±30.63	8878
Non-HDL Cholesterol (mg/dL)	132.30±35.97	122.90±34.05	137.74±35.93	9367
Mean corpuscular hemoglobin concentration (g/dL)	33.91±0.87	33.50±0.82	34.14±0.81	9367
Mean corpuscular volume (fL)	90.15±4.25	90.82±4.42	89.76±4.10	9367
Creatinine (mg/dL)	0.87±0.15	0.73±0.10	0.94±0.12	9367
Eosinophils (10⁹/L)	0.23±0.16	0.21±0.16	0.24±0.17	9367
Erythrocytes (10⁹/L)	4.83±0.41	4.50±0.31	5.03±0.34	9367
ESR (mm/h)	7.98±5.46	10.94±6.63	6.27±3.69	9367
GGT (U/L)	22.20±20.33	15.85±14.70	25.88±22.14	9366
Glucose (mg/dL)	85.68±12.00	84.19±10.08	86.54±12.90	9367
Hematocrit (%)	43.48±3.14	40.76±2.42	45.06±2.31	9367
Hemoglobin (g/dL)	14.75±1.21	13.65±0.91	15.38±0.86	9367
Leukocytes (10⁹/L)	6.69±1.70	6.59±1.66	6.75±1.72	9367
Lymphocytes (10⁹/L)	2.33±0.66	2.24±0.63	2.38±0.67	9367
Monocytes (10⁹/L)	0.62±0.19	0.57±0.17	0.65±0.19	9367
Neutrophils (10⁹/L)	3.48±1.26	3.54±1.29	3.45±1.25	9367
Platelets (10⁹/L)	237.40±51.90	248.79±54.72	230.81±49.00	9367
Mean platelet volume (fL)	8.41±0.69	8.47±0.70	8.37±0.68	9366
Red cell distribution (U)	13.32±0.78	13.38±0.88	13.29±0.71	9367
Triglycerides (mg/dL)	96.72±63.06	78.63±37.34	107.18±71.92	9367
Urate (mg/dL)	5.10±1.26	4.13±0.92	5.66±1.07	9367

Table A2: General characteristics of the OSARTEN cohort for serum samples included into the study. Biochemical data were extracted from a blood test performed at the same time of serum samples collection. Additional information was achieved from questionnaires complementation.

	[ALL] ***N=8157***	**female** ***N=2964***	**male** ***N=5193***	**N**
Age (years)	43.22±9.19	44.22±8.79	42.65±9.36	8157
Weight (kg)	75.60±14.22	65.10±11.37	81.59±12.04	8157
Height (kg)	172.01±9.02	163.68±6.34	176.77±6.52	8157
BMI (kg/m²)	25.44±3.81	24.30±4.05	26.10±3.49	8157
Smoker	1645 (20.19%)	556 (18.79%)	1089 (20.99%)	8148
Alcohol consumption:				8139
Never	1139 (13.99%)	628 (21.23%)	511 (9.86%)	
Social drinker	6074 (74.63%)	2166 (73.23%)	3908 (75.43%)	
Only during meals	729 (8.96%)	144 (4.87%)	585 (11.29%)	
Daily intake of alcohol	197 (2.42%)	20 (0.68%)	177 (3.42%)	
Hypertension	931 (11.41%)	243 (8.20%)	688 (13.25%)	8157
Medicated for hypertension	83 (3.62%)	34 (3.67%)	49 (3.59%)	2291
Medicated for diabetes	71 (3.10%)	9 (0.97%)	62 (4.55%)	2290
Medicated for hypercholesterolemia	68 (2.97%)	26 (2.81%)	42 (3.07%)	2292
Proteins detected in urine	115 (1.42%)	42 (1.46%)	73 (1.41%)	8078
ALT (U/L)	22.79±13.88	16.62±9.46	26.32±14.74	8156
Basophils (10⁹/L)	0.04±0.02	0.04±0.02	0.04±0.02	8157
Cholesterol (mg/dL)	193.60±34.35	191.98±34.00	194.53±34.51	8157
HDL Cholesterol (mg/dL)	60.60±15.83	68.40±15.29	56.14±14.35	8157
LDL Cholesterol (mg/dL)	112.74±30.67	107.37±30.07	116.01±30.56	7689
Non-HDL Cholesterol (mg/dL)	133.01±35.95	123.58±33.93	138.38±35.96	8157
Mean corpuscular hemoglobin concentration (g/dL)	33.90±0.87	33.49±0.82	34.14±0.81	8157
Mean corpuscular volume (fL)	90.15±4.27	90.83±4.42	89.76±4.13	8157
Creatinine (mg/dL)	0.87±0.16	0.73±0.10	0.94±0.12	8157
Eosinophils (10⁹/L)	0.23±0.17	0.21±0.16	0.24±0.17	8157
Erythrocytes (10⁹/L)	4.84±0.42	4.50±0.31	5.03±0.34	8157
ESR (mm/h)	7.98±5.49	10.95±6.75	6.29±3.68	8156
GGT (U/L)	22.60±20.91	15.92±14.94	26.41±22.79	8155
Glucose (mg/dL)	86.11±12.55	84.60±10.74	86.97±13.40	8157
Hematocrit (%)	43.54±3.13	40.81±2.44	45.09±2.31	8157
Hemoglobin (g/dL)	14.77±1.21	13.67±0.92	15.39±0.86	8157
Leukocytes (10⁹/L)	6.73±1.70	6.60±1.67	6.80±1.71	8157
Lymphocytes (10⁹/L)	2.34±0.67	2.25±0.64	2.40±0.68	8157
Monocytes (10⁹/L)	0.62±0.19	0.57±0.17	0.65±0.20	8157
Neutrophils (10⁹/L)	3.50±1.26	3.53±1.28	3.48±1.24	8157
Platelets (10⁹/L)	238.41±51.92	249.95±55.08	231.83±48.83	8157
Mean platelet volume (fL)	8.40±0.68	8.46±0.70	8.36±0.67	8156
Red cell distribution (U)	13.33±0.79	13.40±0.89	13.29±0.71	8157
Triglycerides (mg/dL)	98.80±65.07	79.24±38.06	109.97±74.04	8157
Urate (mg/dL)	5.14±1.27	4.15±0.94	5.70±1.08	8157

Table A3: General characteristics of the OBENUTIC cohort. Biochemical data were extracted from a blood test performed at the same time of urine samples collection. Additional information was achieved from questionnaires complementation.

	[ALL valid] ***N=465***	**female** ***N=307***	**male** ***N=158***	**N**
Age (years)	46.14±13.67	46.42±12.94	45.61±15.01	465
BMI (kg/m²)	27.87±5.40	27.32±5.58	28.94±4.89	465
Waist (cm)	92.31±15.18	88.16±13.48	100.49±15.06	458
Smoker	93 (20.62%)	72 (23.92%)	21 (14.00%)	451
Systolic blood pressure (mmHg)	124.91±17.37	121.00±16.83	132.46±15.91	461
Diastolic blood pressure (mmHg)	78.62±10.69	76.94±9.71	81.89±11.74	461
Hypertension	76 (17.31%)	37 (12.76%)	39 (26.17%)	439
Medicated for hypertension	78 (17.33%)	35 (11.71%)	43 (28.48%)	450
Diabetes	21 (4.79%)	13 (4.48%)	8 (5.41%)	438
Medicated for diabetes	14 (3.12%)	8 (2.69%)	6 (3.97%)	448
Diagnosed cholesterol	132 (30.28%)	84 (29.17%)	48 (32.43%)	436
Medicated for cholesterol	67 (14.99%)	40 (13.47%)	27 (18.00%)	447
Any cardiovascular disease	15 (3.42%)	7 (2.41%)	8 (5.41%)	438
ALT (U/L)	25.29±18.61	20.85±11.96	33.86±25.11	463
AST (U/L)	25.85±11.04	23.33±7.15	30.73±14.95	461
Cholesterol (mg/dL)	212.93±40.11	216.62±40.51	205.77±38.45	465
HDL Cholesterol (mg/dL)	59.85±14.13	64.21±13.53	51.37±11.12	465
LDL Cholesterol (mg/dL)	138.50±32.52	138.86±33.00	137.79±31.67	464
Creatinine (mg/dL)	0.76±0.18	0.67±0.10	0.93±0.19	465
GGT (U/L)	30.74±32.46	27.65±33.59	36.73±29.32	461
Glucose (mg/dL)	94.66±19.07	92.68±16.35	98.50±23.05	465
Leukocytes (10^9/L)	6.43±2.51	6.47±2.89	6.33±1.49	443
Triglycerides (mg/dL)	109.42±58.70	102.63±52.88	122.64±66.87	463
Uric acid (mg/dL)	5.33±1.41	4.79±1.16	6.39±1.25	465

Table A4: General characteristics of the PREDIMED cohort. Biochemical data were extracted from a blood test performed at the same time of urine samples collection.

	[ALL valid] ***N=960***	**female** ***N=612***	**male** ***N=348***	**N**
Age group (years):				960
55-63	330 (34.38%)	213 (34.80%)	117 (33.62%)	
64-69	306 (31.87%)	191 (31.21%)	115 (33.05%)	
70-80	324 (33.75%)	208 (33.99%)	116 (33.33%)	
Diabetes	490 (51.04%)	279 (45.59%)	211 (60.63%)	960
Obesity	492 (51.25%)	338 (55.23%)	154 (44.25%)	960
Dyslipidemia	303 (31.56%)	177 (28.92%)	126 (36.21%)	960
Hypertension	798 (83.12%)	524 (85.62%)	274 (78.74%)	960

Table A5. General characteristics of the KIROLGETXO cohort. Data have been collected from questionaries.

	[ALL valid] ***N=101***	**female** ***N=83***	**male** ***N=18***	**N**
Age (years)	71.14±5.46	70.87±5.49	72.39±5.29	101
Weight (cm)	66.81±12.22	65.08±11.73	74.78±11.51	101
Height (cm)	161.12±7.35	159.65±6.66	167.89±6.68	101
BMI (kg/m²)	25.67±3.95	25.50±4.04	26.48±3.47	101
Ethnic group:				101
caucasian	100 (99.01%)	82 (98.80%)	18 (100.00%)	
hispanic	1 (0.99%)	1 (1.20%)	0 (0.00%)	
Smoker	4 (3.96%)	3 (3.61%)	1 (5.56%)	101
Alcohol consumption:				101
Never	26 (25.74%)	23 (27.71%)	3 (16.67%)	
Social drinker	50 (49.50%)	46 (55.42%)	4 (22.22%)	
Only during meals	16 (15.84%)	10 (12.05%)	6 (33.33%)	
Daily intake of alcohol	9 (8.91%)	4 (4.82%)	5 (27.78%)	
Physical exercise several days per week	101 (100.00%)	83 (100.00%)	18 (100.00%)	101
Diabetes	5 (4.95%)	2 (2.41%)	3 (16.67%)	101
Medicated for diabetes	7 (7.00%)	4 (4.82%)	3 (17.65%)	100
Hypercholesterolemia	38 (37.62%)	29 (34.94%)	9 (50.00%)	101
Medicated for hypercholesterolemia	25 (25.00%)	19 (22.89%)	6 (35.29%)	100
Hypertension	32 (32.00%)	25 (30.12%)	7 (41.18%)	100
Medicated for hypertension	29 (29.29%)	24 (28.92%)	5 (31.25%)	99
Any cardiovascular disease	13 (13.00%)	9 (10.98%)	4 (22.22%)	100
Medicated for cardiovascular disease	9 (9.57%)	5 (6.49%)	4 (23.53%)	94

Table A6: General characteristics of the NAFLD cohort. Biochemical data were extracted from a blood test performed at the same time of urine samples collection. Information on liver state were obtained from the conduced liver biopsies. Additional information was achieved from questionnaires complementation.

	[ALL valid] ***N=234***	**female** ***N=89***	**male** ***N=145***	**N**
Age (years)	54.49±11.92	58.04±11.38	52.30±11.75	234
BMI (kg/m²)	32.26±7.63	32.17±7.45	32.32±7.76	233
Waist (cm)	110.36±14.03	108.65±14.76	111.30±13.58	175
Ethnic group:				231
African	1 (0.43%)	1 (1.14%)	0 (0.00%)	
Arab	3 (1.30%)	1 (1.14%)	2 (1.40%)	
Bangladeshi	1 (0.43%)	0 (0.00%)	1 (0.70%)	
Caribbean	1 (0.43%)	1 (1.14%)	0 (0.00%)	
Indian	4 (1.73%)	1 (1.14%)	3 (2.10%)	
Other Asian background	2 (0.87%)	1 (1.14%)	1 (0.70%)	
Other Black/African/Caribbean background	3 (1.30%)	0 (0.00%)	3 (2.10%)	
Pakistani	2 (0.87%)	0 (0.00%)	2 (1.40%)	
White	214 (92.64%)	83 (94.32%)	131 (91.61%)	
Diabetes	124 (52.99%)	59 (66.29%)	65 (44.83%)	234
Medicated for diabetes	96 (41.20%)	41 (46.07%)	55 (38.19%)	233
Medicated for hypercholesterolemia	108 (46.15%)	47 (52.81%)	61 (42.07%)	234
Medicated for hypertension	123 (52.56%)	54 (60.67%)	69 (47.59%)	234

	[ALL valid] ***N=234***	**female** ***N=89***	**male** ***N=145***	**N**
NAFLD Activity Score (NAS)	4.18±1.53	4.38±1.58	4.06±1.49	233
NAS - Steatosis:				234
1	89 (38.03%)	31 (34.83%)	58 (40.00%)	
2	83 (35.47%)	31 (34.83%)	52 (35.86%)	
3	62 (26.50%)	27 (30.34%)	35 (24.14%)	
NAS - Lobular inflammation:				234
0	28 (11.97%)	12 (13.48%)	16 (11.03%)	
1	139 (59.40%)	51 (57.30%)	88 (60.69%)	
2	64 (27.35%)	24 (26.97%)	40 (27.59%)	
3	3 (1.28%)	2 (2.25%)	1 (0.69%)	
NAS - Hepatocellular ballooning:				233
0	42 (18.03%)	13 (14.77%)	29 (20.00%)	
1	125 (53.65%)	43 (48.86%)	82 (56.55%)	
2	66 (28.33%)	32 (36.36%)	34 (23.45%)	
Fibrosis stage:				234
0	40 (17.09%)	14 (15.73%)	26 (17.93%)	
1	23 (9.83%)	6 (6.74%)	17 (11.72%)	
1a	18 (7.69%)	10 (11.24%)	8 (5.52%)	
1b	17 (7.26%)	5 (5.62%)	12 (8.28%)	
1c	16 (6.84%)	4 (4.49%)	12 (8.28%)	
2	31 (13.25%)	7 (7.87%)	24 (16.55%)	
3	60 (25.64%)	26 (29.21%)	34 (23.45%)	
4	29 (12.39%)	17 (19.10%)	12 (8.28%)	
NAFLD diagnosis:				234
Steatosis	86 (36.75%)	31 (34.83%)	55 (37.93%)	
NASH	148 (63.25%)	58 (65.17%)	90 (62.07%)	
Albumin (g/L)	44.01±3.65	42.78±3.44	44.77±3.58	204
ALT (U/L)	57.02±37.22	51.69±35.03	60.28±38.26	229
AST (U/L)	40.62±24.72	41.86±28.31	39.87±22.32	229
Glucose (mg/dL)	120.66±44.00	132.41±50.87	114.04±38.29	172
Triglycerides (mg/dL)	190.88±293.61	243.42±455.02	158.37±98.11	225

Table A7: General characteristics of the MetS long cohort. Biochemical data were extracted from a blood test performed at the same time of serum samples collection.

	[ALL] ***N=154***	**female** ***N=51***	**male** ***N=103***	**N**
Age (years)	67.21±12.90	71.53±10.72	65.08±13.39	154
Weight (kg)	89.61±23.00	81.52±17.04	93.62±24.53	154
Height (kg)	165.76±9.87	156.27±7.23	170.45±7.29	154
BMI (kg/m²)	32.55±7.28	33.40±6.44	32.13±7.66	154
Diabetes: yes	154 (100.00%)	51 (100.00%)	103 (100.00%)	154
Dyslipidemia	41 (69.49%)	13 (65.00%)	28 (71.79%)	59
Hypertension	115 (74.68%)	38 (74.51%)	77 (74.76%)	154
Waist (cm)	109.76±14.14	107.88±14.04	110.74±14.17	135
Glucose (mg/dL)	141.24±53.05	145.59±61.06	139.07±48.74	153
Cholesterol (mg/dL)	133.99±67.35	128.37±74.33	136.77±63.82	154
HDL Cholesterol (mg/dL)	73.74±36.38	83.84±38.80	68.85±34.40	95
LDL Cholesterol (mg/dL)	120.78±60.67	140.06±62.45	111.44±57.99	95
Triglycerides (mg/dL)	142.43±72.31	160.26±81.77	133.60±65.79	154
Systolic blood pressure (mmHg)	142.62±16.14	140.92±17.05	143.30±16.02	42
Diastolic blood pressure (mmHg)	79.60±10.27	74.92±11.64	81.47±9.22	42

Table A8: General characteristics of the PORTUGAL cohort. Biochemical data were extracted from a blood test performed at the same time of serum samples collection. Additional information was achieved from questionnaires complementation.

	[ALL] ***N=159***	**female** ***N=101***	**male** ***N=58***	**N**
Age (years)	82.33±7.73	83.17±7.23	80.88±8.39	159
Weight (kg)	65.31±12.98	63.42±11.33	68.62±14.97	159
Height (kg)	155.79±8.59	151.69±6.74	162.91±6.59	159
BMI (kg/m²)	26.87±4.65	27.53±4.53	25.72±4.66	159
Diabetes	47 (32.87%)	36 (39.13%)	11 (21.57%)	143
Dyslipidemia	72 (48.98%)	53 (56.38%)	19 (35.85%)	147
Hypertension	116 (76.32%)	73 (76.04%)	43 (76.79%)	152
Medicated for hypertension	125 (81.70%)	82 (83.67%)	43 (78.18%)	153
Medicated for diabetes	37 (24.18%)	30 (30.61%)	7 (12.73%)	153
Medicated for hypercholesterolemia	65 (42.48%)	48 (48.98%)	17 (30.91%)	153
Glucose (mg/dL)	101.10±28.19	102.35±29.96	98.97±25.01	154
Cholesterol (mg/dL)	164.09±47.91	170.84±45.54	152.60±50.03	154
HDL Cholesterol (mg/dL)	56.19±13.20	57.40±14.56	54.14±10.27	154
LDL Cholesterol (mg/dL)	85.23±41.00	88.57±39.14	79.54±43.76	154
Triglycerides (mg/dL)	113.32±53.04	124.34±56.54	94.58±40.50	154
Smoker	7 (4.43%)	1 (1.00%)	6 (10.34%)	158
Systolic blood pressure (mmHg)	126.34±20.95	125.70±21.34	127.35±20.46	148
Diastolic blood pressure (mmHg)	70.18±11.13	69.40±11.20	71.42±10.99	148

Table A9: Metadata from the analyzed individuals of the *preCOVID* and *COVID* cohorts. Biochemical data were extracted from a blood test performed at the same time of serum samples collection. Symptoms were registered at hospital admission.

	preCOVID N=280	***COVID N=263***	**p-value**	**N**
Main info				
Gender (female)	146 (52.14%)	116 (45.14%)	0.124	537
Age (years)	48.89±11.00	64.81±16.64	<0.001	537
Total hospitalization days	-	13.72±19.60	-	257
Days in ICU	-	5.12±16.52	-	257
Smoker	45 (16.07%)	17 (6.61%)	0.001	537
Pneumonia:			-	248
unilateral	-	33 (13.31%)	-	
bilateral	-	173 (69.76%)	-	
Death	-	24 (9.34%)	-	257
Signs at admission				
Temperature (ºC)	-	36.47±0.91	-	249
Breathing freq. (n x min.)	-	22.29±7.98	-	65
Heart rate (n x min.)	-	90.82±17.36	-	250
Systolic blood pressure (mm Hg)	-	135.72±23.08	-	250
Diastolic blood pressure (mm Hg)	-	78.24±12.73	-	251
Comorbidities				
Cardiovascular	-	68 (26.46%)	-	257
Cerebrovascular	-	18 (7.00%)	-	257
Diabetes	18 (6.43%)	64 (24.90%)	<0.001	537
EPOC	-	29 (11.28%)	-	257
Hypertension	50 (18.12%)	116 (45.14%)	<0.001	533
Immunodeficiency	-	11 (4.28%)	-	257
Liver failure	-	4 (1.56%)	-	257
Neoplasm	-	31 (12.06%)	-	257
Renal insufficiency	-	21 (8.17%)	-	257
Symptoms				
Clouding of consciousness	-	20 (7.78%)	-	257
Conjunctival congestion	-	2 (0.78%)	-	257
Diarrhea	-	75 (29.18%)	-	257
Disorientation	-	11 (4.28%)	-	257
Dry cough	-	135 (52.53%)	-	257
Fatigue	-	148 (57.59%)	-	257
Fever	-	177 (68.87%)	-	257
Headache	-	49 (19.07%)	-	257
Hemoptysis	-	1 (0.39%)	-	257
Lymphadenopathy	-	3 (1.17%)	-	257
Myalgia	-	75 (29.18%)	-	257
Nasal congestion	-	11 (4.28%)	-	257
Nausea Vomiting	-	38 (14.84%)	-	256
Odynophagia	-	31 (12.06%)	-	257
Oropharyngeal congestion	-	1 (0.39%)	-	257
Productive cough	-	58 (22.57%)	-	257
Shaking chills	-	55 (21.40%)	-	257
Skin rash	-	3 (1.17%)	-	257
Blood test				
Albumin (g/dL)	-	3.69±0.42	.	136

	preCOVID N=280	*COVID N=263*	p-value	N
ALT (U/L)	22.26±11.90	34.09±26.08	<0.001	528
APTT (s)	-	24.27±4.83	.	252
Bilirubin (mg/dL)	0.50±0.23	0.74±0.46	0.002	164
C-reactive protein (mg/L)	-	77.32±71.33	.	255
Creatinine (mg/dL)	0.82±0.16	1.06±0.81	<0.001	535
Creatine phosphokinase (U/L)	-	158.26±331.91	.	233
D-dimer (ng/mL)	-	2508.72±9149.53	.	248
Ferritin (ng/mL)	47.29±49.13	670.80±772.35	<0.001	254
Glucose (mg/dL)	88.26±16.07	136.09±87.94	<0.001	535
Interleukin 6 (pg/mL)	-	19.50±24.78	-	12
Lactate dehydrogenase (U/L)	-	318.70±205.66	-	247
Leukocytes (10^9/L)	6.91±1.81	7.58±5.15	0.050	534
Lymphocytes (10^9/L)	2.41±0.77	1.20±1.30	<0.001	534
Monocytes (10^9/L)	0.63±0.19	0.42±0.24	<0.001	535
Neutrophils (10^9/L)	3.59±1.37	6.02±6.02	<0.001	534
Platelets (10^9/L)	237.43±45.86	218.03±109.86	0.009	534
Procalcitonin (ng/mL)	-	0.37±0.78	-	130
Protein (g/dL)	-	6.35±0.57	-	145
Prothrombin activity (%)	-	87.85±21.82	-	234
Urea (mg/dL)	-	44.20±29.86	-	255

Table A10: Metadata from the analyzed individuals of the *preCOVID* and *COVID* balanced subcohorts. Biochemical data were extracted from a blood test performed at the same time of serum samples collection. Symptoms were registered at hospital admission.

	preCOVID N=112	*COVID* N=112	p-value	N
Main info				
Gender (female)	59 (52.68%)	59 (52.68%)	1.000	224
Age (years)	49.86±10.66	49.86±10.66	1.000	224
Total hospitalization days	-	8.77±13.41	.	112
Days in ICU	-	2.63±9.13	.	112
Smoker	20 (17.86%)	13 (11.61%)	0.258	224
Pneumonia:			.	109
unilateral	-	19 (17.43%)		
bilateral	-	69 (63.30%)		
Death	-	2 (1.79%)	.	112
Signs at admission				
Temperature (ºC)	-	36.42±0.74	.	109
Breathing freq. (n x min.)	-	21.26±6.14	.	23
Heart rate (n x min.)	-	94.53±15.43	.	108
Systolic blood pressure (mm Hg)	-	136.31±22.68	.	109
Diastolic blood pressure (mm Hg)	-	82.37±11.87	.	109
Comorbidities				
Cardiovascular	-	10 (8.93%)	.	112

	preCOVID N=112	*COVID* N=112	p-value	N
Cerebrovascular	-	3 (2.68%)	.	112
Diabetes	7 (6.25%)	12 (10.71%)	0.337	224
EPOC	-	8 (7.14%)	.	112
Hypertension	20 (18.02%)	25 (22.32%)	0.526	223
Immunodeficiency	-	6 (5.36%)	.	112
Liver failure: No	-	112 (100.00%)	.	112
Neoplasm	-	6 (5.36%)	.	112
Renal insufficiency	-	5 (4.46%)	.	112
Symptoms				
Clouding of conciousness	-	2 (1.79%)	.	112
Conjunctival congestion	-	1 (0.89%)	.	112
Diarrhea	-	44 (39.29%)	.	112
Disorientation: No	-	112 (100.00%)	.	112
Dry cough	-	74 (66.07%)	.	112
Fatigue	-	63 (56.25%)	.	112
Fever	-	85 (75.89%)	.	112
Headache	-	36 (32.14%)	.	112
Hemoptysis: No	-	112 (100.00%)	.	112
Lymphadenopathy	-	1 (0.89%)	.	112
Myalgia	-	45 (40.18%)	.	112
Nasal congestion	-	7 (6.25%)	.	112
Nausea Vomiting	-	20 (18.02%)	.	111
Odynophagia	-	19 (16.96%)	.	112
Oropharyngeal congestion	-	1 (0.89%)	.	112
Productive cough	-	17 (15.18%)	.	112
Shaking chills	-	25 (22.32%)	.	112
Skin rash	-	3 (2.68%)	.	112
Blood test				
Albumin (g/dL)	-	3.89±0.36	.	64
ALT (U/L)	21.67±11.32	37.82±26.75	<0.001	220
APTT (s)	-	23.23±2.68	.	111
Bilirubin (mg/dL)	0.55±0.27	0.74±0.41	0.160	72
C-reactive protein (mg/L)	-	52.48±58.37	.	112
Creatinine (mg/dL)	0.82±0.17	0.89±0.47	0.144	224
Creatine phosphokinase (U/L)	-	162.12±407.84	.	108
D-dimer (ng/mL)	-	1319.18±6822.22	.	110
Ferritin (ng/mL)	47.29±49.13	579.11±710.03	<0.001	117
Glucose (mg/dL)	89.25±15.33	126.21±105.66	<0.001	224
Interleukin 6 (pg/mL)	-	21.50±26.85	.	10
Lactate dehydrogenase (U/L)	-	298.60±253.94	.	110
Leukocytes (10^9/L)	6.87±1.63	6.53±2.70	0.264	223
Lymphocytes (10^9/L)	2.45±0.75	1.25±0.70	<0.001	224
Monocytes (10^9/L)	0.61±0.17	0.39±0.18	<0.001	224
Neutrophils (10^9/L)	3.56±1.20	5.38±7.80	0.016	223

Appendix

	preCOVID N=112	*COVID* N=112	p-value	N
Platelets (10^9/L)	239.20±51.42	220.15±114.36	0.110	224
Procalcitonin (ng/mL)	-	0.14±0.15	.	45
Protein (g/dL)	-	6.48±0.48	.	65
Prothrombin activity (%)	-	96.79±11.65	.	103
Urea (mg/dL)	-	31.79±19.04	.	112

Table A11: List of quantified metabolites by the Bruker IVDr Quantification in Plasma/Serum B.I.Quant-PS™ report.

Metabolite	Concentration unit
2-Aminobutyric acid	mmol/L
2-Hydroxybutyric acid	mmol/L
2-Oxoglutaric acid	mmol/L
3-Hydroxybutyric acid	mmol/L
Acetic acid	mmol/L
Acetoacetic acid	mmol/L
Acetone	mmol/L
Alanine	mmol/L
Asparagine	mmol/L
Ca-EDTA	mmol/L
Choline	mmol/L
Citric acid	mmol/L
Creatine	mmol/L
Creatinine	mmol/L
D-Galactose	mmol/L
Dimethylsulfone	mmol/L
Ethanol	mmol/L
Formic acid	mmol/L
Glucose	mmol/L
Glutamic acid	mmol/L
Glutamine	mmol/L
Glycerol	mmol/L
Glycine	mmol/L
Histidine	mmol/L
Isoleucine	mmol/L
K-EDTA	mmol/L
Lactic acid	mmol/L
Leucine	mmol/L
Lysine	mmol/L
Methionine	mmol/L
N,N-Dimethylglycine	mmol/L
Ornithine	mmol/L
Phenylalanine	mmol/L
Proline	mmol/L
Pyruvic acid	mmol/L
Sarcosine	mmol/L
Succinic acid	mmol/L
Threonine	mmol/L
Trimethylamine-N-oxide	mmol/L
Tyrosine	mmol/L
Valine	mmol/L

Table A12: List of quantified lipoproteins and lipoproteins subfractions in the Bruker IVDr Lipoprotein Subclass Analysis (B.I.-LISA™) report. Abbreviations: LDL: low-density lipoprotein; HDL: high-density lipoprotein; VLDL: very low-density lipoprotein; IDL: intermediate-density lipoprotein.

Key	Class/subclass	Compound	Concentration unit
TPTG	Total Plasma	Triglycerides	mg/dL
TPCH	Total Plasma	Cholesterol	mg/dL
LDCH	LDL	Cholesterol	mg/dL
HDCH	HDL	Cholesterol	mg/dL
TPA1	Total Plasma	Apolipoprotein-A1	mg/dL
TPA2	Total Plasma	Apolipoprotein-A2	mg/dL
TPAB	Total Plasma	Apolipoprotein-B100	mg/dL
LDHD	Ratio LDL and HDL Cholesterol	LDL Cholesterol / HDL Cholesterol	-/-
ABA1	Ratio of Apolipoproteins A1 and B100	Apolipoprotein-A1 / Apolipoprotein-B100	-/-
TBPN	Apolipoprotein-B100 carrying particles	Particle Number	nmol/L
VLPN	VLDL	Particle Number	nmol/L
IDPN	IDL	Particle Number	nmol/L
LDPN	LDL	Particle Number	nmol/L
L1PN	LDL-1	Particle Number	nmol/L
L2PN	LDL-2	Particle Number	nmol/L
L3PN	LDL-3	Particle Number	nmol/L
L4PN	LDL-4	Particle Number	nmol/L
L5PN	LDL-5	Particle Number	nmol/L
L6PN	LDL-6	Particle Number	nmol/L
VLTG	VLDL Class	Triglycerides	mg/dL
IDTG	IDL Class	Triglycerides	mg/dL
LDTG	LDL Class	Triglycerides	mg/dL
HDTG	HDL Class	Triglycerides	mg/dL
VLCH	VLDL Class	Cholesterol	mg/dL
IDCH	IDL Class	Cholesterol	mg/dL
LDCH	LDL Class	Cholesterol	mg/dL
HDCH	HDL Class	Cholesterol	mg/dL
VLFC	VLDL Class	Free Cholesterol	mg/dL
IDFC	IDL Class	Free Cholesterol	mg/dL
LDFC	LDL Class	Free Cholesterol	mg/dL
HDFC	HDL Class	Free Cholesterol	mg/dL
VLPL	VLDL Class	Phospholipids	mg/dL
IDPL	IDL Class	Phospholipids	mg/dL
LDPL	LDL Class	Phospholipids	mg/dL
HDPL	HDL Class	Phospholipids	mg/dL
HDA1	HDL Class	Apolipoprotein-A1	mg/dL
HDA2	HDL Class	Apolipoprotein-A2	mg/dL
VLAB	VLDL Class	Apolipoprotein-B100	mg/dL
IDAB	IDL Class	Apolipoprotein-B100	mg/dL
LDAB	LDL Class	Apolipoprotein-B100	mg/dL
V1TG	VLDL-1 Subclass	Triglycerides	mg/dL
V2TG	VLDL-2 Subclass	Triglycerides	mg/dL
V3TG	VLDL-3 Subclass	Triglycerides	mg/dL
V4TG	VLDL-4 Subclass	Triglycerides	mg/dL
V5TG	VLDL-5 Subclass	Triglycerides	mg/dL
V1CH	VLDL-1 Subclass	Cholesterol	mg/dL
V2CH	VLDL-2 Subclass	Cholesterol	mg/dL

Key	Class/subclass	Compound	Concentration unit
V3CH	VLDL-3 Subclass	Cholesterol	mg/dL
V4CH	VLDL-4 Subclass	Cholesterol	mg/dL
V5CH	VLDL-5 Subclass	Cholesterol	mg/dL
V1FC	VLDL-1 Subclass	Free Cholesterol	mg/dL
V2FC	VLDL-2 Subclass	Free Cholesterol	mg/dL
V3FC	VLDL-3 Subclass	Free Cholesterol	mg/dL
V4FC	VLDL-4 Subclass	Free Cholesterol	mg/dL
V5FC	VLDL-5 Subclass	Free Cholesterol	mg/dL
V1PL	VLDL-1 Subclass	Phospholipids	mg/dL
V2PL	VLDL-2 Subclass	Phospholipids	mg/dL
V3PL	VLDL-3 Subclass	Phospholipids	mg/dL
V4PL	VLDL-4 Subclass	Phospholipids	mg/dL
V5PL	VLDL-5 Subclass	Phospholipids	mg/dL
L1TG	LDL-1 Subclass	Triglycerides	mg/dL
L2TG	LDL-2 Subclass	Triglycerides	mg/dL
L3TG	LDL-3 Subclass	Triglycerides	mg/dL
L4TG	LDL-4 Subclass	Triglycerides	mg/dL
L5TG	LDL-5 Subclass	Triglycerides	mg/dL
L6TG	LDL-6 Subclass	Triglycerides	mg/dL
L1CH	LDL-1 Subclass	Cholesterol	mg/dL
L2CH	LDL-2 Subclass	Cholesterol	mg/dL
L3CH	LDL-3 Subclass	Cholesterol	mg/dL
L4CH	LDL-4 Subclass	Cholesterol	mg/dL
L5CH	LDL-5 Subclass	Cholesterol	mg/dL
L6CH	LDL-6 Subclass	Cholesterol	mg/dL
L1FC	LDL-1 Subclass	Free Cholesterol	mg/dL
L2FC	LDL-2 Subclass	Free Cholesterol	mg/dL
L3FC	LDL-3 Subclass	Free Cholesterol	mg/dL
L4FC	LDL-4 Subclass	Free Cholesterol	mg/dL
L5FC	LDL-5 Subclass	Free Cholesterol	mg/dL
L6FC	LDL-6 Subclass	Free Cholesterol	mg/dL
L1PL	LDL-1 Subclass	Phospholipids	mg/dL
L2PL	LDL-2 Subclass	Phospholipids	mg/dL
L3PL	LDL-3 Subclass	Phospholipids	mg/dL
L4PL	LDL-4 Subclass	Phospholipids	mg/dL
L5PL	LDL-5 Subclass	Phospholipids	mg/dL
L6PL	LDL-6 Subclass	Phospholipids	mg/dL
L1AB	LDL-1 Subclass	Apolipoprotein-B100	mg/dL
L2AB	LDL-2 Subclass	Apolipoprotein-B100	mg/dL
L3AB	LDL-3 Subclass	Apolipoprotein-B100	mg/dL
L4AB	LDL-4 Subclass	Apolipoprotein-B100	mg/dL
L5AB	LDL-5 Subclass	Apolipoprotein-B100	mg/dL
L6AB	LDL-6 Subclass	Apolipoprotein-B100	mg/dL
H1TG	HDL-1 Subclass	Triglycerides	mg/dL
H2TG	HDL-2 Subclass	Triglycerides	mg/dL
H3TG	HDL-3 Subclass	Triglycerides	mg/dL
H4TG	HDL-4 Subclass	Triglycerides	mg/dL
H1CH	HDL-1 Subclass	Cholesterol	mg/dL
H2CH	HDL-2 Subclass	Cholesterol	mg/dL
H3CH	HDL-3 Subclass	Cholesterol	mg/dL
H4CH	HDL-4 Subclass	Cholesterol	mg/dL
H1FC	HDL-1 Subclass	Free Cholesterol	mg/dL
H2FC	HDL-2 Subclass	Free Cholesterol	mg/dL
H3FC	HDL-3 Subclass	Free Cholesterol	mg/dL
H4FC	HDL-4 Subclass	Free Cholesterol	mg/dL
H1PL	HDL-1 Subclass	Phospholipids	mg/dL

Key	Class/subclass	Compound	Concentration unit
H2PL	HDL-2 Subclass	Phospholipids	mg/dL
H3PL	HDL-3 Subclass	Phospholipids	mg/dL
H4PL	HDL-4 Subclass	Phospholipids	mg/dL
H1A1	HDL-1 Subclass	Apolipoprotein-A1	mg/dL
H2A1	HDL-2 Subclass	Apolipoprotein-A1	mg/dL
H3A1	HDL-3 Subclass	Apolipoprotein-A1	mg/dL
H4A1	HDL-4 Subclass	Apolipoprotein-A1	mg/dL
H1A2	HDL-1 Subclass	Apolipoprotein-A2	mg/dL
H2A2	HDL-2 Subclass	Apolipoprotein-A2	mg/dL
H3A2	HDL-3 Subclass	Apolipoprotein-A2	mg/dL
H4A2	HDL-4 Subclass	Apolipoprotein-A2	mg/dL

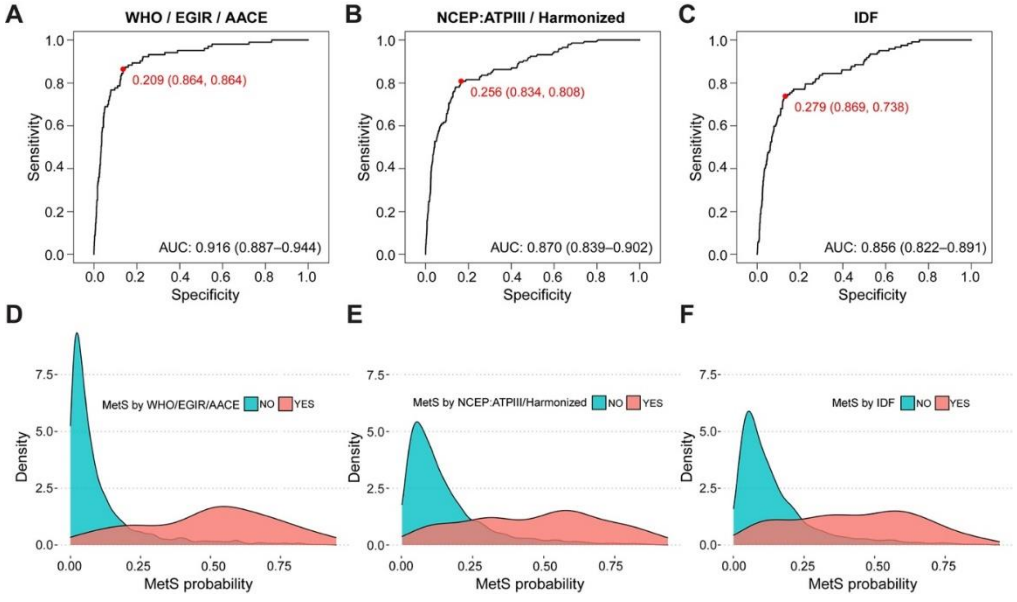

Figure A1: Probability distribution test with a glucose cut-off value of 110 mg/dL for the three built MetS models. **A-C**) (ROC) curves for the three definitions under consideration: **A**) WHO, EGIR and AACE; **B**) NCEP:ATPIII and Harmonized; **C**) IDF. **D-F**) Smoothed histograms evidencing the probability distributions of the MetS model applied to the full cohort for the three definitions under consideration: **D**) WHO, EGIR, and AACE; **E**) NCEP:ATPIII and Harmonized; **F**) IDF. Red and blue colours indicated samples with/without MetS according to the given definition.

Table A13: Performance Metrics from OPLS-DA through a Repeated Double Cross-Validation Process. Value columns are the mean value; p value columns were obtained from permutation tests.

Metric	Training		Validation	
	Value	**p-value**	**Value**	**p-value**
AUROC	0.98	<0.01	0.977	<0.01
Accuracy	0.928	<0.01	0.923	<0.01
Sensitivity	0.874	<0.01	0.867	<0.01
Specificity	0.978	<0.01	0.975	<0.01

List of abbreviations:

1D NMR	Mono-dimensional NMR experiment
^{1}H NMR	Proton NMR
2D NMR	Two-dimensional NMR experiment
4-HPPA	4-hydroxyphenylpyruvic acid
AAA	Aromatic Amino Acids
AACE	American Association of Clinical Endocrinology
ACE2	Angiotensin-Converting-Enzyme 2
Apo-A1/A2	Apolipoprotein A1/A2
Apo-B (Apo-B100)	Apolipoprotein B
ARDS	Acute Respiratory Distress Syndrome
ASCVD	Atherosclerotic Cardiovascular Disease
AUC	Area Under the Curve
B_0	External magnetic field
BCAA	branched-chain amino acids
Bi	Little internal magnetic field
BMI	Body Mass Index
BMRB	Biological Magnetic Resonance Bank
BP	Blood Pressure
CCL 2, 5 or 3	Chemokine (C-C motif) ligand 2, 5 or 3
CE	Cholesteryl Esters
CETP	Cholesteryl Ester Transfer Protein
CKD	Chronic Kidney Disease
CKD-EPI	Chronic Kidney Disease Epidemiology Collaboration
COSY	Correlation SpectroscopY
CPMG	Carr-Purcell-Meiboom-Gill
CS	Chemical Shift
CVD	Cardiovascular Diseases
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DSS	Sodium 2,2-dimethyl-2-silapentane-5-sulphonate
E-GFR	Estimated Glomerular Filtration Rate
EGIR	European Group for the Study of Insulin Resistance
ERETIC	Electronic Reference To access In-vivo Concentrations
f	female
FG	Fasting plasma Glucose
FID	Free induction decay
FT	Fourier transformation

GC-MS	Gas Chromatography copled to Mass Spectrometry
GC-MS/MS	Gas Chromatography copled to tandem Mass Spectrometry
Glyc-A	α-1-acid Glycoprotein A
HDL	High-Density Lipoprotein
HDL-C	HDL cholesterol
HL	Hepatic Lipase
HMDB	Human Metabolome Databases
HSQC	Heteronuclear Single Quantum Coherence
I	Magnetic quantum number
ICU	Intensive Care Unit
IDF	International Diabetes Federation
IDL	Intermediate-Density Lipoprotein
IFG	Impaired Fasting Glucose
IFN-γ	Interferon gamma
IgM/G	Immunoglobulin M/G
IGT	Impaired Glucose Tolerance
IL	Interleukin
IP-10	Interferon gamma Inducible Protein-10
J-coupling	Spin-spin coupling
JRES	J-resolved spectroscopy
LC-MS	Liquid Chromatography copled to Mass Spectrometry
LC-MS/MS	Liquid Chromatography copled to tandem Mass Spectrometry
LDL	Low-Density Lipoprotein
LOD	Limit Of Detection
LPL	Lipoprotein Lipase
m	male
MetS	Metabolic syndrome
MS	Mass Spectrometry
NAFLD	Non-Alcoholic Fatty Liver Disease
NCEP ATP III	National Cholesterol Education Program's Adult Treatment Panel III
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser Effect Spectroscopy
OPLS-DA	Orthogonal PLS-DA
PCA	Principal Component Analysis
PLS	Partial Least Squares analysis
PM	Precision Medicine
PLS-DA	PLS Discriminant Analysis
QuantRef	Quantification Reference solution
RANTES	Regulated on Activation, Normal T Expressed and Secreted
RF	Risk factor

ROC	Receiver Operating Characteristic Curves
RT-PCR	Real-Time reverse-transcription Polymerase Chain Reaction
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SOPs	Standard Operating Procedures
SPT	SamplePro Tube
T2DM	Type 2 Diabetes Mellitus
TC	Total cholesterol
TG	Triglycerides
TMA	Trimethylamine
TMAO	Trimethylamine N-oxide
TNF-α	Tumor Necrosis Factor-α
TOCSY	TOtal Correlation Spectroscopy
TSP	Trimethylsilylpropionic acid-d_4 sodium salt
TSP	Trimethylsilylpropanoic acid
VLDL	Very-Low Density Lipoproteins
WC	Waist Circumference
WHO	World Health Organization
WHR	Waist-Hip Ratio
ΔE	Energy difference