Università degli Studi di Genova

Doctorate in:

Sciences and Technologies of Chemistry and Materials

Innovative analytical methods for the study of low and high weight molecules involved in diseases.

PhD student: Chiara Lavarello

Supervisors: Prof. Emanuele Magi

PhD Andrea Petretto

SOMMARIO

1. PRECISION MEDICINE

Personalized or precision medicine is an emerging approach to the treatment and prevention of disease, which takes into account the individual variability of genes, environment and lifestyles of each person[1]. The study of a biological system is one of the major challenges of system biology and seeks to reveal the organization and interactions between the different cellular mechanisms that underlie complex processes. Personalized medicine includes preventive, diagnostic and therapeutic measures optimally tailored to an individual and aimed at applying more effective therapies and reducing side effects. The diffusion and improvement of molecular characterization technologies we have witnessed in recent years allow us to simultaneously study the different levels of biological information. These levels involve the study of DNA (genomics), proteins (proteomics) as well as the analysis of chemical fingerprints left by physiological processes (metabolomics); these areas of study are defined as "omics". Proteins and metabolites are highly dependent on lifestyle, environment, drugs and the physiological state of each individual so they are necessary to fully describe the phenotype of a disease.

The proteomic and metabolomic approaches allow the global identification, not guided by a priori hypothesis, of thousands of proteins and hundreds of metabolites present in a biological fluid and the evaluation of how these can vary in the presence of a specific disease compared to a non-pathological condition, allowing to characterize and discriminate between different study groups[2].

The different functional levels analyzed by omics sciences are integrated with each other in the context of system biology, a new field of science, which offers an innovative approach to the study of biological systems. The reductionist approach that has been used in medicine to date studies cellular components separately and opposes the holistic approach that considers a biological system as a complex network of dynamic interactions between its components, such as genes, mRNA, proteins and metabolites that identify the so-called phenotype, Figure 1.

The understanding of networks of interactions through system biology can lead to a more complete view of the processes involved in the disease. The identification of thousands of proteins and hundreds of metabolites can help in the new disease biomarkers discovery and therefore in development of new methodologies for the diagnosis or treatment of a pathological state.

Both omics and more traditional research identify biomarkers for diagnostic, prognostic, screening and monitoring purposed.

Translational medicine has the task of rapidly and effectively translating scientific discoveries into clinical practice, through biomarkers qualification, verification and validation and through a two-way relationship with scientists, producers, clinicians and epidemiologists[3].

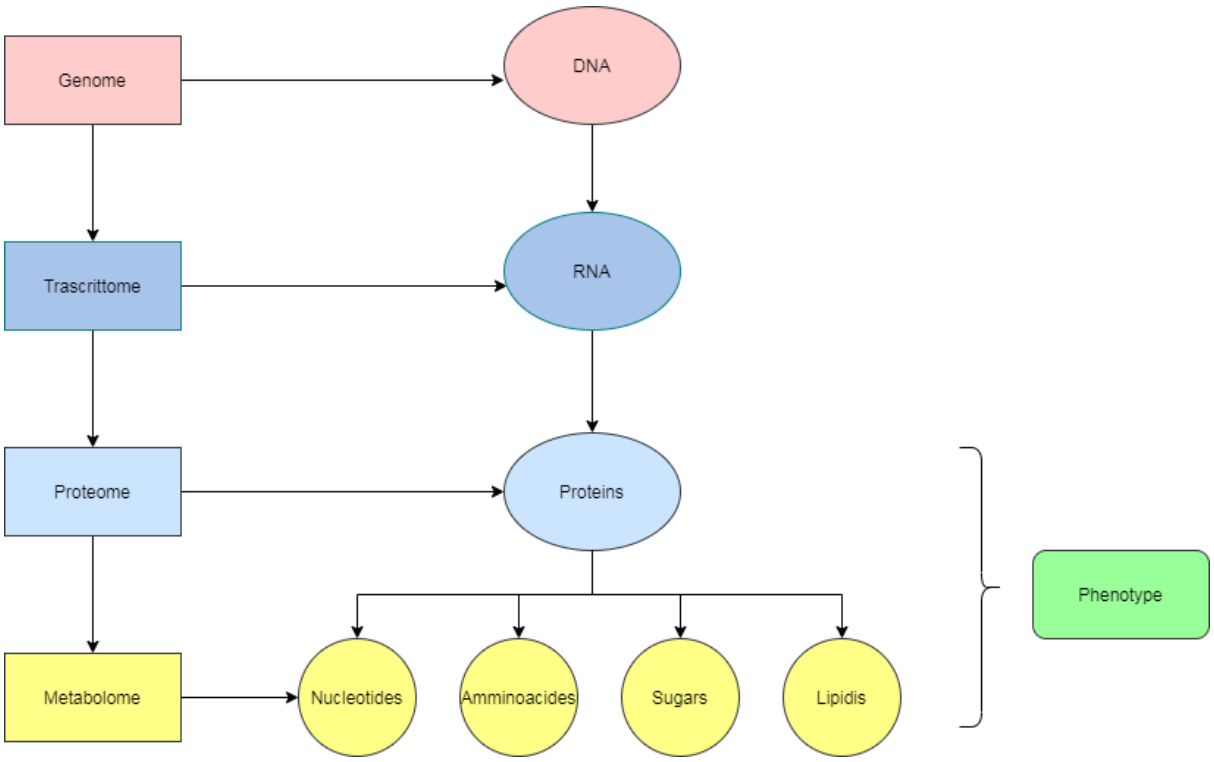

Figure 1: *Omic sciences.*

High Resolution Mass Spectrometry (HRMS) coupled with Ultra High Performance Liquid Chromatography (UHPLC) plays a fundamental role in characterizing and quantifying disease markers as well as in describing the molecular mechanisms of the biological processes in which they are involved[4].

This latest generation technology has become in recent years the investigative tool of excellence, thanks to which it is possible to "untargeted" the study of thousands of analytes in a single run with an accuracy and precision comparable to targeted methods[5].

2. THE PROJECT

My PhD project is the result of a collaboration between the Department of Chemistry and Industrial Chemistry of the University of Genoa and Proteomics and Clinical Metabolomics Laboratory of the Institute for Scientific Research (IRCCS) Giannina Gaslini.

The collaboration is consistent with the new birth of "*Federazione Nazionale degli Ordini dei Chimici e Fisici"*; thanks to 3/18 law the profession of chemist in Italy has been recognized as being of fundamental importance for the health and safety of the community.

My PhD project is divided into two parts that have in common the development of innovative analytical methods in High Resolution Mass Spectrometry with Orbitrap technology, for the study of proteome and metabolome in complex biological matrices.

During the first two years, the proteome of auto inflammatory diseases was studied, while in the last year a spectral MSMS library and Accurate Mass Retention Time (AMRT) Database of small biological molecules were created.

3. INSTRUMENTATIONS

In both projects High Resolution Mass Spectrometer coupled to UHPLC were used to separate and to identify and quantify proteins and metabolites in biological matrix such as cells, cerebrospinal fluids, urine and plasma.

RS Ultimate 3000 HPLC coupled with Orbitrap Fusion Tribrid was used to study the proteome of inflammatory diseases, while Vanquish Horizon UHPLC coupled with the Q-Exactive Hybrid Quadrupole-Orbitrap mass spectrometer was used to build an accurate database of masses and retention times. The choice was dictated by the fact that nano and micro fluxes for proteomics and metabolomics respectively are required.

Liquid chromatography allows the separation of the analytes (whether polar, thermolabile or hardly volatile) contained in the sample on the basis of their distribution between the mobile and stationary phases. The stationary phase is formed by packed solid particles of very fine grain size (between 1.8 and 10 µm), housed inside a column suitable to withstand the high pressures with which the instrument works.

The mobile phase may consist of a pure liquid or a solvent mixture of known composition, which flows through the stationary phase of the chromatographic column. The elution can be isocratic, in which the solvents flow with the same composition throughout the chromatographic run, or polarity gradient, in which the relative quantities of solvents vary during the course of the analysis. The affinity with the stationary phases is given by the different polarity and the different chemical characteristics of the compounds of the mixture to be separated: analytes which have little affinity with the stationary phase are not retained and therefore elute quickly transported by the mobile phase; on the other hand, analytes which have greater affinity will be retained and will come out in a longer time.

Mass spectrometry is an analytical technique that allows the identification and quantification of analytes present in complex mixtures based on the different mass-to-charge ratio (m/z) of ions generated by ionization. The main strengths of these mass spectrometers are their high resolution and mass accuracy. The resolution power of a

spectrometer is the ability to differentiate masses; its expression for Orbitrap[6] is given by the formula[7]:

$$R = \frac{m}{\Delta m}$$

where m is the peak mass and Δm is the peak amplitude at half its height (FWHM: Full With at Half Maximum peak height).

Mass Accuracy is a parameter that directly depends on the instrument's ability to separate adjacent peaks and therefore resolution. The estimate of mass accuracy is expressed by the value of the error on the measured mass, expressed in parts per million (ppm), is determined by the function[8,9]:

$$A = \frac{mt - ms}{mt} = \frac{1}{R}$$

where mt= exact theoretical mass and ms= accurate mass measured instrumentally. The ability to measure the mass of a compound with adequate accuracy directly determines the possibility of obtaining its elementary composition[10]. The instrument must be carefully calibrated: external calibration is performed by direct infusion of a mixture of known compounds (caffeine, MRFA tetrapeptide and Ultramark 1621), in a mass range of 150 to 2000 m/z and allows a mass accuracy of 3 ppm. Improved accuracy of less than 1 ppm can be achieved using an internal calibration.

Orbitrap-based instruments have established themselves firmly in the field of proteomics, biopharmaceuticals, metabolomics and metabolite analysis and made a high resolution mass spectrometer accessible to most life science laboratories. Moreover, it has gained increased popularity in areas of bio-analysis, lipidomics, doping as well as in drug and pesticide residue analysis[11].

3.1 Ultimate 3000 and Vanquish Orizon UHPLC

The two UHPLC are both suitable for processing large quantities of samples thanks to the robustness of the hydraulic system. They differ in the flow they work with. RS-UHPLC Ultimate 3000 is characterized by a wide flow/pressure range allowing the application of UHPLC on nano scale and at the same time allowing to have maximum resolution and analytical speed, even for tryptic digests of maximum complexity. The Vanquish UHPLC for the metabolomics project uses micro fluxes.

The fundamental components are:

- degaser, a system that allows to eliminate any bubbles present in the solvent that could compromise the chromatographic efficiency and possibly the operation of the detectors;
- housing of the analytical and nano pump;
- autosampler;
- the thermostatic compartment in which the analytical column is housed. The column is certainly a fundamental part of the instrument: the columns on the market are many and differ in the type of stationary phase, length, diameter and the different dimensions of the pores of the stationary phase.

  In reverse-phase (RP) chromatography, the stationary phase has a hydrophobic character, while the mobile phase has a polar character. The interactions in RP-HPLC are considered to be the hydrophobic forces and these forces are caused by the energies resulting from the disturbance of the dipolar structure of the solvent. The separation is typically based on the partition of the analyte between the stationary phase and the mobile phase. The solute molecules are in equilibrium between the hydrophobic stationary phase and partially polar mobile phase. The more hydrophobic molecule has a longer retention time while the ionized organic compounds, inorganic ions and polar metal molecules show little or no retention time.

  HILIC or Hydrophilic Interaction Liquid Chromatography is a technique for separation of polar and hydrophilic compounds. HILIC separation is based on passing a mostly

organic mobile phase over a polar stationary phase. This forms a water enriched layer next to the stationary phase. The hydrophilic analytes partition into this layer and as the mobile phase becomes more hydrophilic, they are eluted in order of increasing hydrophilicity.

3.2 Orbitrap Fusion Tribrid e Orbitrap Q-Exactive

The Figure 2 represents the fundamental constituents of the Orbitrap Fusion Tribrid instrument:

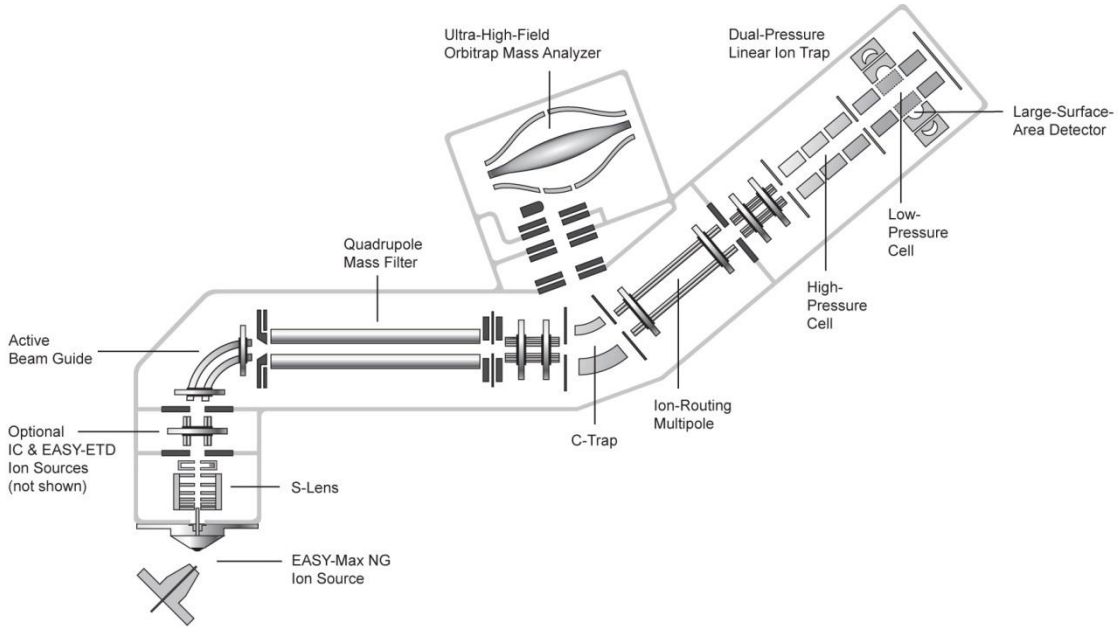

Figure 2: *Orbitrap Fusion Tribrid scheme.*

- Electrospray Interface

ESI (Electrospray Ionization) interface, as well as the Atmospheric Pressure Chemical Ionization (APCI) interface are required to couple liquid chromatography with mass spectrometers and it converts the analytes in solution into gas-phase ions. ESI is a "soft" ionization source: it causes a low sample fragmentation so that the mass spectrum presents protonated [M + H]+ , deprotonated [M- H]- or adduced ions of the analyte with ionic species such as Na+ , K+ etc. ESI is particularly suitable for the ionization of polar compounds and not only small molecules, but also proteins, peptides, industrial polymers thanks to the formation of multicharged ions. Electrospray ionization takes place at atmospheric pressure and the temperature can be increased in order to improve solvent evaporation and declusterization.

The analytes previously separated by chromatography, are injected through a metal capillary into the ionization chamber of the source, the input is facilitated by the application of a high electrical potential that can be negative or positive depending on the analytes. Coaxial to the capillary there is a gas flow (sheath gas) with the function of nebulizing the solution in a thin aerosol at the exit of the capillary. A spray of charged droplets, with a diameter of a few microns (1-60 μm), is thus formed and directed towards the opening on the counter-electrode made up of a capillary and a skimmer. The desolvation is favored by the temperature of the source and the flow of an auxiliary gas which, applied externally has the task of assisting the sheath gas in the nebulization of the sample and the evaporation of the solvent. The process continues until the "Rayleigh limit" is reached, a limit in which the surface tension of the drop can no longer withstand the Coulumbian repulsion between the charges with a consequent "Coulumbian" explosion: the drops disintegrate obtaining smaller dimensions. The mechanism is repeated until all the solvent has evaporated with consequent formation of analyte desolved ions in the gas phase, which can have unit or multiple charges. The ion swipe cone forms a mechanical barrier that protects the transfer capillary from larger droplets and particulate matter. Ions then pass through the ion transfer capillary into the mass detector.

- S – Lens

At the exit of the capillary is the S-Lens which focuses the ions towards the opening of the skimmer. S-lens is a parameter that must be carefully adjusted, since it is of great importance in the intensity with which an ion species will be revealed, representing a discriminat between the degree of declusterization and fragmentation of an ion.

- Active beam guide.

Prevents neutral molecules from entering the quadrupole, improving the robustness of the system.

- Quandrupolo Q.

The quadrupole has a "filter" function and selects ions with an isolation width of less than 0.4 amu to achieve excellent sensitivity and selectivity. In particular, such instruments can select ions virtually instantaneously due to the fast switching times of the quadrupole and it is able to fragment them in HCD mode on a similarly fast time scale. In combination with the

ability to fill the HCD cell or the C-trap with ions while a previous Orbitrap detection cycle is still ongoing, the system achieved a very high duty cycle. By allocating most of the analysis time (up to 70–95%) just for accumulating ions, the sensitivity is maximized and the negative impact that low ion currents could have on acquisition speed and quality of spectra is prevented.

- C – trap.

The C-trap is essentially a "flatpole" ion guide that has both a curved central axis and a slot in the electrode closest to the Orbitrap. The trap is typically operated with ~1 mTorr of nitrogen, allowing very effective collisional cooling of ions in its RF field. Once the ions have accumulated/cooled sufficiently within the C-trap, the RF amplitude is quickly ramped down (in ~100-200 ns). The absence of an RF field means that the ions will no longer be confined radially within the trap's volume. Once the RF amplitude has dropped, the ions must be quickly pulled out of the trap so that they don't strike the trap electrodes. This is accomplished by appropriate DC potentials. For positive ions, the electrode furthest from the Orbitrap is the most positive, and the slotted electrode closest to the Orbitrap is the least positive. As a result, positive ions are pushed toward the Orbitrap, exiting the C-trap through its slotted electrode. They are subsequently accelerated/focused into the Orbitrap. Since the ions had very low kinetic energy within the C-trap, they have a narrow distribution of kinetic energies upon entering the Orbitrap. Additionally, the rapid changes in RF and DC potentials that are used to empty the C-trap mean that the ions enter the Orbitrap in a narrow time window.

- Ion Routing Multipole (Higher Energy Collisional Dissociation, HCD).

The HCD collision cell consists of a multipole mounted inside a connected metal tube, direct line with the C-Trap and filled with nitrogen. Through collisions with the gas molecules, the precursor ions are fragmented, producing high-energy fragmentation spectra that provide information complementary to that obtained from the ion trap. The fragments obtained are sent through C-Trap to Orbitrap and a high-resolution spectrum is acquired.

- Orbitrap.

The Orbitrap[12] is a mass analyzer, consisting of an internal spindle-shaped electrode, placed coaxially to an external bell-shaped electrode. Ions are injected into the volume between the central and outer electrodes essentially along a tangent through a specially machined slot with a compensation electrode (a “deflector”) in one of the outer electrodes. With voltage applied between the central and outer electrodes, a radial electric field bends the ion trajectory toward the central electrode while tangential velocity creates an opposing centrifugal force. With a correct choice of parameters, the ions remain on a nearly circular spiral inside the trap. At the same time, the axial electric field caused by the special conical shape of electrodes pushes ions toward the widest part of the trap initiating harmonic axial oscillations. Outer electrodes are then used as receiver plates for image current detection of these axial oscillations. The digitized image current in the time domain is Fourier-transformed into the frequency domain in the same way as in FTICR and then converted into a mass spectrum.

$$\omega = \sqrt{k\frac{z}{m}}$$

- Dual linear ion trap

The linear ion trap is a low-resolution mass analyzer that can store, isolate and fragment ions and detect them using its detectors. To produce a mass spectrum, direct current (dc) and radio frequency (rf) voltages are varied in the ion trap to sequentially eject ions according to their m/z ratio. By varying only the RF voltage, the analyser can first expel all the ions, except some precursor ions and then exploiting the collisions with the helium molecules it contains, fragmenting them to obtain the ions produced. The ions produced are then ejected from the trap according to their m/z ratio, thus obtaining an MSMS spectrum.

Dual linear ion trap was chosen as a ‘partner’ for the Orbitrap mass analyzer due to its very high sensitivity, superb control of the ion population, short cycle time, and MS^n capability.

Depending on the requirements for the analysis, the two analyzers could be used independently or in concert. An important feature of the instrument is the procedure of

automatic gain control (AGC), wherein a short pre-scan in the linear trap is used to determine the ion current within the mass range of interest, hence enabling storage of a defined number of ions ('AGC target value') in the subsequent analytical scan. The AGC feature, in combination with the precise determination of the ion injection time, allows the instrument to be used for accurate quantitative analyses and ensures the stability and accuracy of m/z measurements by the Orbitrap analyzer, as control of the ion packet decreases the rate of space charging. It is worth pointing out that the MS/MS spectra generated in the linear ion trap and detected either in the linear trap or the Orbitrap analyzer are very similar, with the only major difference being the resolution and mass accuracy of the observed peaks. Meanwhile, the linear trap requires about an order of magnitude fewer ions for an analytically useful spectrum due to its very high transmission and single-ion detection capability[11].

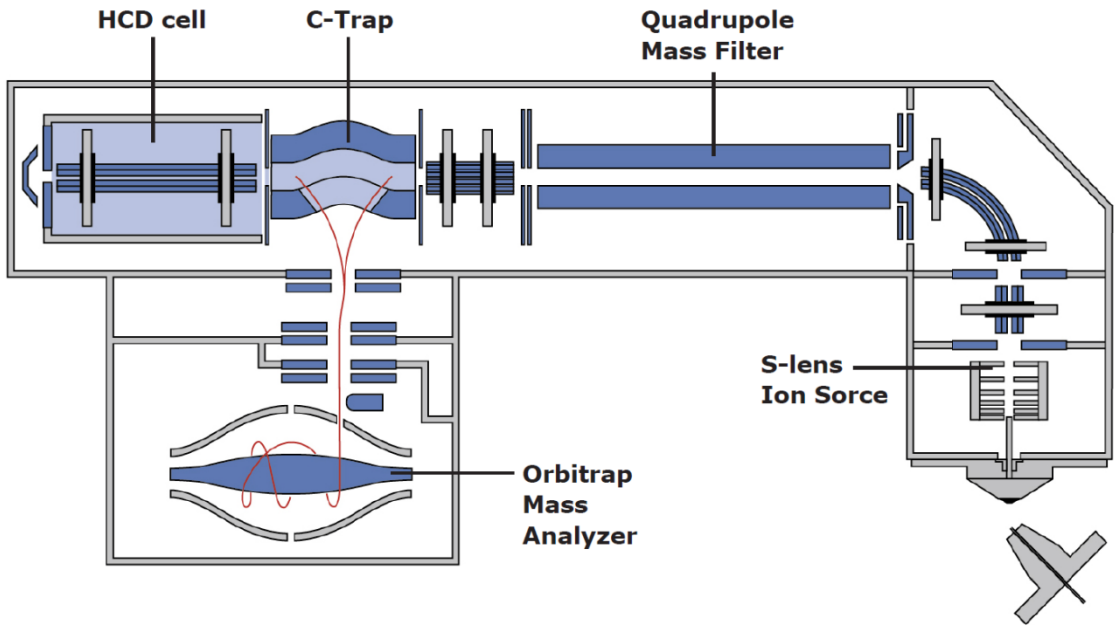

Figure 3: *Orbitrap Q-Exactive scheme.*

Figure 3 shows the scheme of the fundamental constituents of the Orbitrap Q-exactive instrument; as can be easily seen by comparing Figure 2 with Figure 3, the Q-Exactive spectrometer lacks the linear trap. This different configuration allows only certain types of acquisitions, those that would involve the use of the linear trap for low-energy fragmentation and low-resolution but faster fragment ion detection.

3.3 Acquisition mode

In LC-MS/MS methods, three events occur in the mass spectrometer:

1. ionization: analytes elute into the mass spectrometer from the HPLC and are ionized;
2. MS1 scan: the abundance and mass-to-charge ratios (m/z) of all ions eluting at a given time are measured. This acquisition mode can be conducted in both the low resolution ion trap and the high resolution orbital trap, but usually is conducted in Orbitrap. The chromatogram is given by the Total Ion Current (TIC) as a function of the retention time;
3. MS2 scan: some or all detected ions are fragmented, the abundances and m/z's of the fragments in orbital trap or ion trap are measured and recorded.

Different LC-MS/MS methods vary in how ions are selected and measured in the MS2 scan. Figure 4 shows a scheme of how analytes are isolated, fragmented and analyzed by a mass spectrometer working on DDA, MRM, PRM, or DIA modes[13].

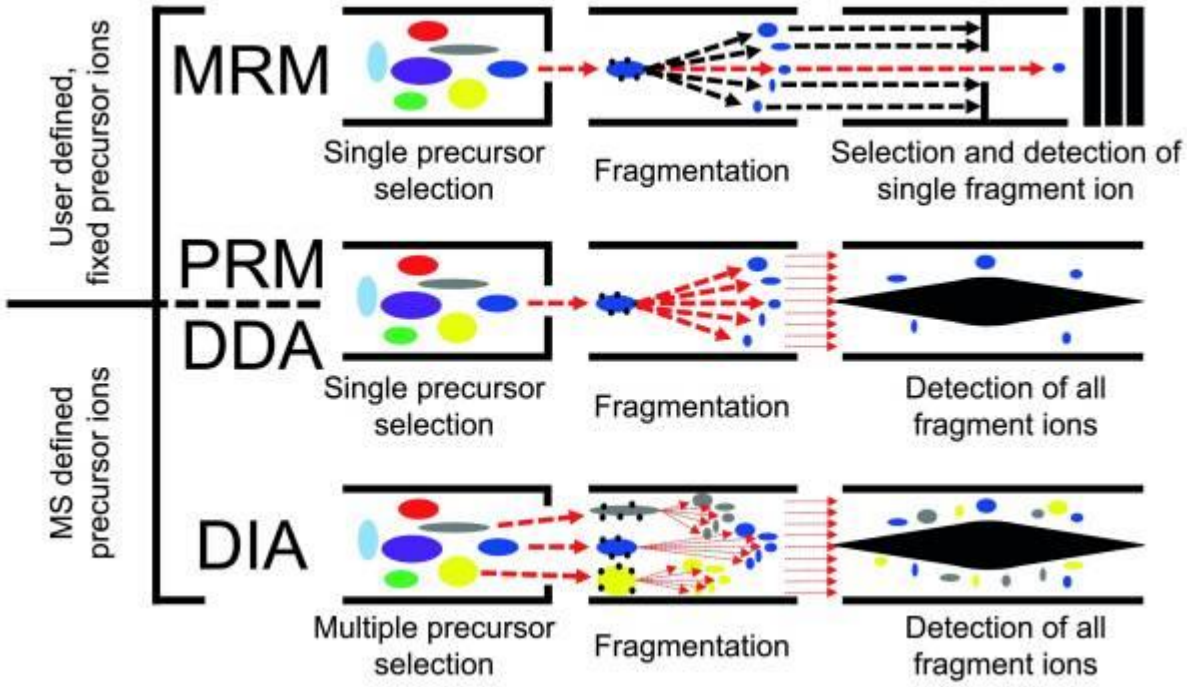

Figure 4: *MRM, PRM, DDA and DIA acquisition mode.*

In DDA, MRM and PRM, single precursor ions are isolated, fragmented and analyzed in an MS2 scan by the mass spectrometer. In DDA mode, the precursor ions are chosen by the instrument on the basis of abundance. In MRM and PRM, the precursor ions to be analyzed are fixed by the user.

DIA is different form the methods above in that all precursor ions within a selected mass range are isolated, fragmented and analyzed in a single MS2 scan. In MS1 scan the analytes ions entering the mass spectrometer at a given time are identified; in MS2 scan the fragments of all (or some) of precursors that are in the mass spectrometer at a given time are identified.

- Data Dependent Analysis (DDA).

In DDA, a subset of the most abundant ions reaching the mass spectrometer detector during an MS1 scan are individually isolated and fragmented in sequential MS2 scans and each MS2 scan can be analyzed with a database search algorithm[14,15]. Currently, most instruments can perform a DDA cycle with one MS1 scan and 10 MS2 scans within 2 seconds. DDA typically yields thousands of analytes identifications. Unfortunately, irreproducibility and imprecision are fundamental to DDA's design; if too many analytes co-elute and appear in a single MS1 scan, then DDA stochastically samples only the most abundant ions and misses the rest[16]. This approach diminishes reproducibility and prevents the measurement of low-abundance peptides[17]. Additionally, to survey as many analytes as possible, DDA deliberately samples each ions only once or twice, preventing precise absolute quantification that requires multiple measurements per analyte.

DDA is an acquisition strategy widely used in the so-called "discovery proteomics" because it allows the identification of analytes with a wide dynamic range[18].

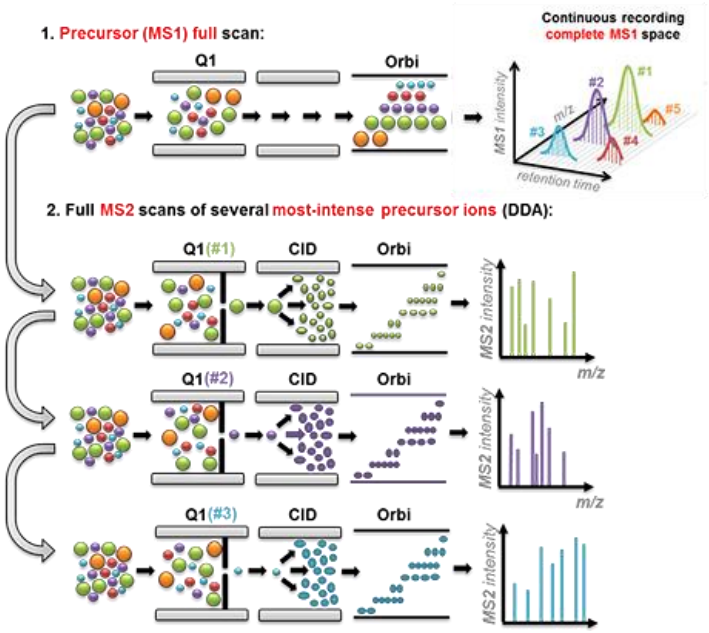

Figure 5: *DDA acquisition mode.*

- Selected Reaction Monitoring (SRM) and PRM (Parallel Reaction Monitoring).

SRM and PRM are targeted mass spectrometry techniques that are emerging in the field of proteomics as a complement to untargeted shotgun methods. SRM and PRM are particularly useful when predetermined sets of proteins, such as those constituting cellular networks or sets of candidate biomarkers, need to be measured across multiple samples in a consistent, reproducible and quantitatively precise manner[19]. Instead of selecting the top *n* precursors in an MS1 scan for further fragmentation in MS2 scans, these methods select only precursors and fragments with the m/z and elution time that match a pre-specified analyte of interest. In MRM, for each MS1 scan, a subset of fragment ions is measured in the subsequent MS2 scan[20], whereas in PRM, all of the fragment ions are measured[21]. In both, the same precursors are selected and fragmented multiple times to acquire more precise quantification of fewer peptides, compared with DDA[22].

- Data Indipendent Analysis (DIA).

In DIA mode, for each cycle, the instrument focuses on a narrow mass window of precursors and acquires MS/MS data from all precursors detected within that window. This mass

window is then stepped across the entire mass range, systematically collecting MS/MS data from every mass and from all detected precursors. The most common method to generate DIA data is called SWATH (SequentialWindowed Acquisition of All Theoretical Fragment ions), in which the mass spectrometer divides the mass range into small mass windows.

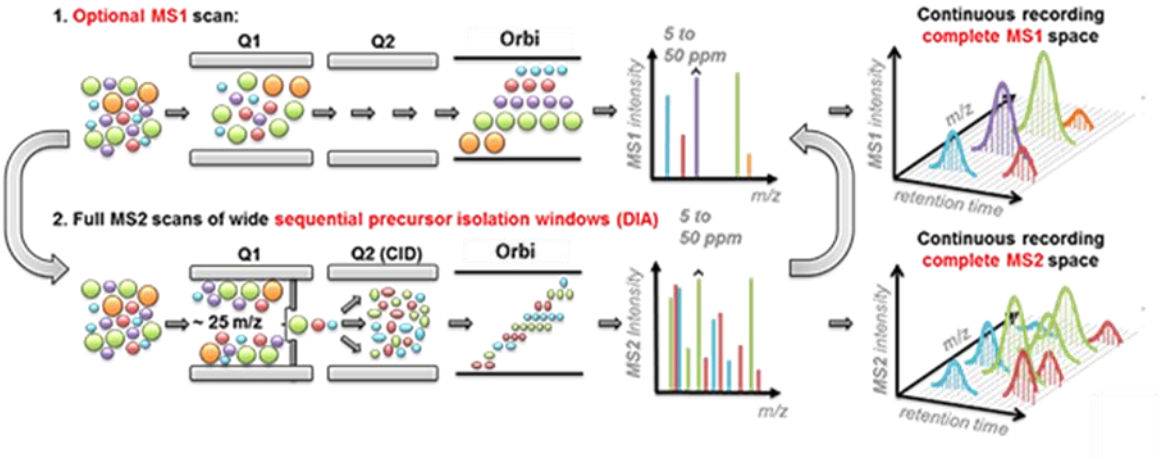

Figure 6: *DIA acquisition mode.*

In SWATH, the mass analyzer cover 400-1200 m/z as an overall mass range and the mass spectrometer steps with 2-4seconds cycle time by precursor acquisition windows. During each cycle, the mass spectrometer thus fragments all precursors from the quadrupole isolation window (such as 475 – 500 m/z for 25 wide windows) and a complete, high accuracy fragment ion spectrum of all precursors selected in that isolation window is recorded. The same precursor isolation window is fragmented over and over at each cycle during the entire chromatographic separation, thus providing a time-resolved recording of the fragment ions of all the precursors that elute on the chromatography. The SWATH MS data consists therefore of highly multiplexed fragment ion maps that are deterministically recorded over the user-defined mass precursor mass range and chromatographic separation[13].

4 PROTEOMICS OF AUTOINFLAMMATORY DISEASES

4.1 Introduction

Proteomics is defined as the study of the globality of proteins, in their functional complexity. The term proteome was introduced by Wilkins[23] to identify proteins (prote-) expressed by a genome (-oma). Knowledge of the sequence of the entire human genome is only the first step in understanding the functioning of the cellular machine at the molecular level; the proteome can therefore be seen as the end product of the process of DNA translation/transcription to proteins and is complementary to genomic information in understanding cellular mechanisms.

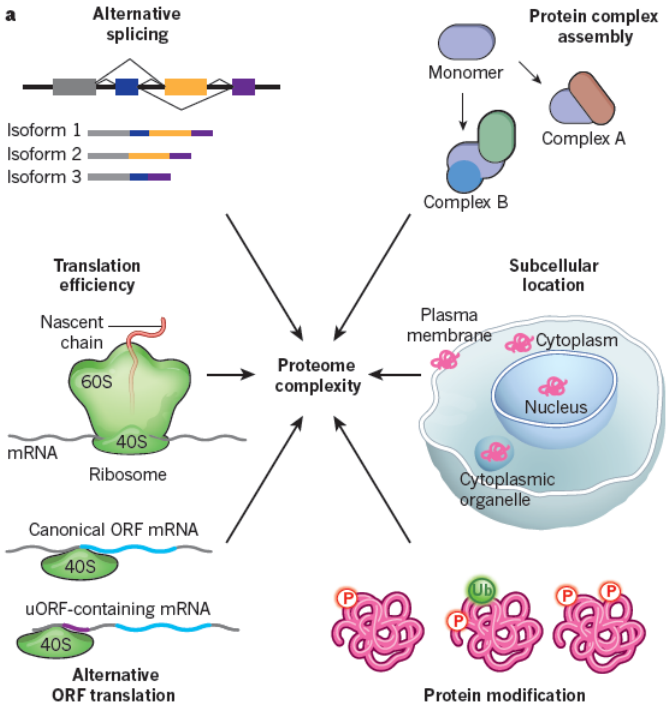

Figure 7: *Factors involved in determining the proteome complexity.*

The proteome is extremely complex, in fact there are many potential variability factors between a gene and its corresponding active product. These include splicing, co- and post-

translational modifications, sub-cellular localization, mRNA translation efficiency, assembly of proteins with different compositions, Figure 7.

The cellular proteome is a complex microcosm of structural and regulatory networks that requires continuous surveillance and modulation to meet cellular needs. It is therefore essential that the protein flow of the cell remains in equilibrium to ensure the correct functioning of the cells. Genetic alterations ranging from chromosome imbalance to oncogene activation can affect the speed and capacity of biogenesis and protein degradation systems, which often result in proteome imbalance. In a balanced proteome (Figure 8a), the level of protein production does not exceed the capacity of protein depletion systems. To achieve this balance, many factors contribute to protein generation or degradation. Events and cell states such as chromosomal imbalance, oncogenic activation and translation errors alter the proteome in ways that promote proteotoxic stress and lead to an imbalance in the proteome (Figure 8b).

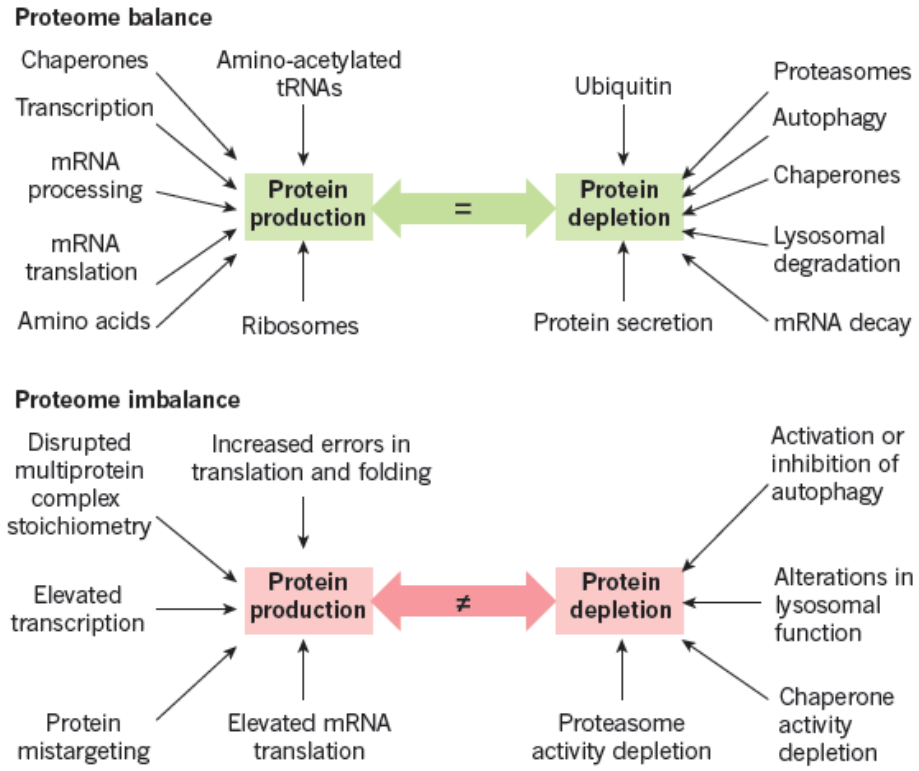

Figure 8: *Possible causes of proteome balance dysregulation*

Numerous biological processes are simultaneously active in every living cell. Each of them includes synthetic, catalytic and regulatory functions that are almost always performed by proteins further organized into higher order structures and networks. For decades, the structures and functions of selected proteins have been studied using biochemical and biophysical methods. However, the properties and behavior of the proteome as an integrated system have remained largely unknown. Powerful mass spectrometry-based technologies provide informations on the composition, structure, function and control of the proteome, explaining the complex biological processes and phenotypes of a physiological and/or pathological state. Proteomics based on high-resolution mass spectrometry is suitable for the study of post-transdutional changes because leads to characteristic changes in mass and can be localized with the resolution of a single amino acid.

To be successful, clinical proteomic projects must link observed phenotypes to modern molecular medicine through the analysis of complex proteoforms. Clinical phenotypes are affected by both familial inheritance (genotype) and environmental effects (that is, there can be differing molecular causes for the same underlying disease).

The two principal approaches to identifying and characterizing proteins using mass spectrometry are the “bottom-up”, which analyze peptides by proteolytic digestion and “top-down”, which analyze intact proteins.

Bottom-up proteomics is the most mature and most widely used approach for protein identification and characterization. The bottom-up approach provides an indirect measurement of proteins through peptides derived from proteolytic digestion of intact proteins. Intact proteins are digested into peptides prior to introduction into the mass spectrometer where they are then detected and fragmented[24].

Shotgun proteomics refers to a use of bottom-up proteomics techniques to study the whole proteins in a complex mixture, such as serum, urine, and cell lysates, etc. It utilizes HPLC in combination with mass spectrometry technology. The most distinctive feature of shotgun proteomics is that it enables identification and comparative quantification of a wide range of proteins at the same time with minimal protein separation needed. Protein mixtures are first digested by protease, and the resulting peptides are separated in HPLC, followed by tandem

MS/MS analysis to identify the sequence of each peptide. The identified peptide sequences are compared with database, in searching for the corresponding protein identity[25].

Top-down proteomics can characterize intact proteins from complex biological systems. This approach routinely allows for nearly 100% sequence coverage and full characterization of proteoforms, the specific molecular form of the protein resulting from combinations of genetic variation, alternative splicing, and post-translational modifications. Fragmentation for tandem mass spectrometry is accomplished by electron-capture dissociation (ECD) or electron-transfer dissociation (ETD). The proteins are extracted from cell or tissue lysates, separated by either gel or LC, and directly analyzed by MS for a complete view of all proteoforms including those with PTMs and sequence variations[26].

Essentially two approaches are used for protein quantification : label free quantification and chemical or metabolic label.

Label-free quantification is a method of determining the relative amount of proteins in two or more biological samples, but unlike other quantitative methods, it does not use a stable isotope for chemical binding and labeling of proteins. Label-free quantitative proteomics approach provides a powerful tool to resolve and identify thousands of proteins from a complex biological sample. In this approach, proteins are first digested with a protease into a peptide mixture, which is subsequently analyzed by tandem MS (MS/MS) and identified by database searching. Relative protein abundance is determined by either spectral counting or chromatographic peak intensity measurements.

The label free approach is opposed by the label methods such as chemical labeling or metabolic labeling. Heavy stable isotope labels can be introduced in vivo by growing cells or even whole organisms in the presence of amino acids or nutrients carrying such stable isotopes. Metabolic labeling is often the preferred labeling technique, since incorporation occurs at the earliest possible moment in the sample preparation process, thereby minimizing the error in quantification.

The advantage of chemical labeling over metabolic labeling is the possibility to label a wide range of different sample types, since incorporation of the label is performed only after

harvesting cells and subsequent purification of proteins. Chemical labeling is essentially based on similar mechanisms as metabolic labeling, except that the label is introduced into proteins or peptides by a chemical reaction. Alternatively, the heavy stable isotope label can be introduced into the peptide during an enzymatic reaction with heavy water ($H_2{}^{18}O$). In all PhD experiments label free approach was carried out.

Shotgun and label free proteomic experiments were performed during the PhD project.

4.2 State of the art

Autoinflammatory diseases are a group monogenic inflammatory conditions characterized by an early onset during childhood. Under the term “periodic fevers” are gathered some monogenic diseases characterized by periodic or recurrent episodes of systemic inflammation causing fever often associated with rash, serositis (peritonitis, pleuritis), lymphadenopathy, arthritis, and other clinical manifestations[27].

Most of these conditions are caused by mutations of genes regulating the activation of innate immunity leading to an unprovoked activation of inflammatory response[28]. A genetic defect is a modification of a gene by an event known as mutation. This mutation alters the function of the gene, which gives incorrect information to the organism giving rise to the disease. The most common autoinflammatory diseases are the so called monogenic recurrent fevers, namely: Familial Mediterranean Fever (FMF), TNF receptor-associated periodic syndrome (TRAPS) and mevalonate kinase deficiency/hyperimmunoglobulin D syndrome (MKD/HIDS).

Figure 9 shows a simplified scheme of the pathogenesis of the main autoinflammatory syndromes. The mutated proteins are marked with an asterisk and green lettering denotes the diseases associated with them.

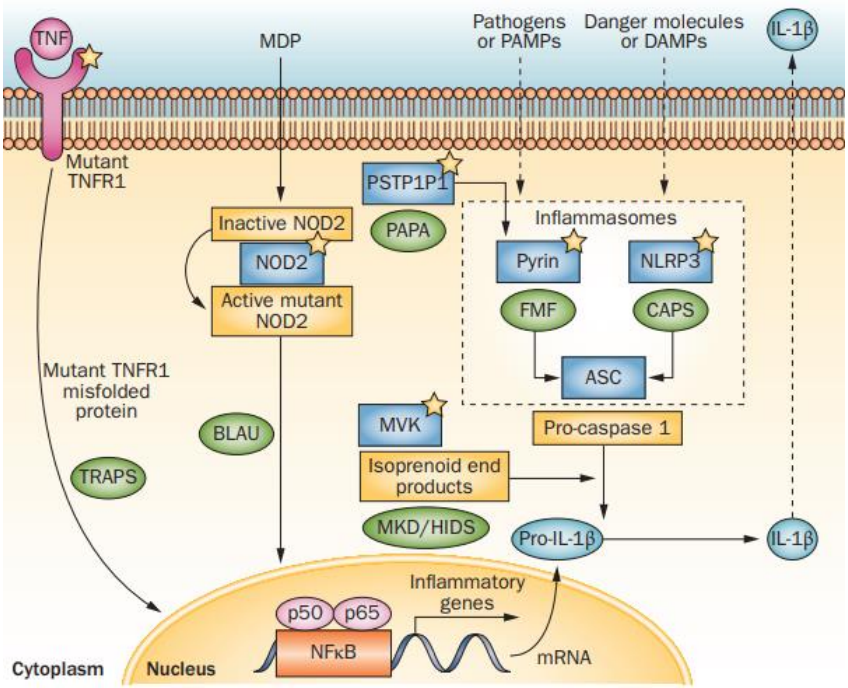

Figure 9: *Scheme of the pathogenesis of the main autoinflammatory syndromes.*

In CAPS genetic diseases the mutation on the CIAS1 gene alters the production of a specific protein called Cryopyrin and affects its function. The mutated ASC protein stimulates the proinflammatory cytokinic response of IL-1β promoting the inflammatory process causing fever and inflammation in sick patients.

Periodic cryopyrin-associated syndromes (CAPS) include a group of rare autoinflammatory diseases including Familial Autoinflammatory Cold Syndrome (FCAS), Muckle-Wells Syndrome (MWS) and Chronic Infantile Neurological Cutaneous Articular Skin Syndrome (CINCA), also known as neonatal onset multisystemic inflammatory disease (NOMID). These syndromes were initially described as distinct clinical entities despite some clinical similarities: patients often present with overlapping symptoms including fever, hives-like rashes (pseudo-orticaria) and joint involvement of varying severity associated with systemic inflammation. These three diseases exist on a continuum of severity: FCAS is the mildest condition, CINCA (NOMID) the most severe, and patients with MWS have an intermediate phenotype. Characterizations of these conditions at the molecular level have demonstrated mutations of the same gene in all three disorders.

MEFV is the first gene mutated discovered in FMF patients, the mutation has a autosomal recessive inheritance and leads to a loss of function of the protein[29,30]. The gene encodes for Pyrin protein; Pyrin after activation forms a Pyrin-inflammasome that is able to process IL-1β independently from NLRP3. Recent studies showed that Pyrin can sense signal changes of cytoskeletal organization by bacterial toxins via RhoA. Toxins can inactivate RhoA and lead to actin depolymerization, in this manner pyrin senses bacterial infection and responds by nucleating an inflammasome. Mutated Pyrin forms substantially reduce the binding affinity to PKNs (serine/threonine protein kinases that phosphorylates pyrin) and thus escape from the inhibitory phosphorylation due to the link with the inhibitory 14.3.3 protein[31,32].

14-3-3 proteins are a family of conserved regulatory molecules that are expressed in all eukaryotic cells. 14-3-3 proteins have the ability to bind a multitude of functionally diverse signaling proteins, including kinases, phosphatases, and transmembrane receptors. More than 200 signaling proteins have been reported as 14-3-3 ligands.

In patients with TRAPS, the mutation is a heterozygous mutation in TNFRSF1A which encode for the TNFR1 receptor. The result is a misfolded molecule that is retained in the endoplasmic reticulum and the consequent oxidative stress and defective autophagy represents a trigger for the production and the activation of pro-inflammatory cytokines.

MKD is caused by biallelic loss-of-function mutations in the MVK gene, which encodes mevalonate kinase and is also linked to pyrin-inflammasome activation. The identification of the molecular defects associated to recurrent fevers sheds new lights in the pathogenesis of these conditions. However, the understanding of all the intracellular consequences secondary to the mutations of the genes involved is far to be completely achieved. A defect in MVK gene leads to RhoA inactivation and diminished PI3K (phosphoinositide 3-kinase) activity; this potentiates pyrin-inflammasome activation[33,34].

In the last few years, there has been a massive expansion of large-scale proteomic data that has facilitated the comprehension of the complexity of the intracellular communication and the identification of new markers of disease. Proteomics based on high-resolution mass spectrometry has maturated into attractive technology that allows the quantification of thousands of proteins and offers a hypothesis-free and systems-wide analysis[35,36,37].

Studying the monocytes of these patients using a quantitative proteomics we will try to identify new pathways and mechanisms that could be involved in the pathogenesis of these disorders.

4.3 The aim of the project

The diagnosis of autoinflammatory diseases is complex and often based on the exclusion of other diseases. The autoinflammatory diseases involve genetic mutations and malfunctions of the proteins encoded by genes. It is easy to understand how a proteomic analysis is of fundamental interest to understand the molecular mechanisms underlying genetic diseases but not only, the study of proteome also represents the hope to find biomarkers of disease that allow to predict the disease before the onset of symptoms. There are also some diseases that have symptoms very similar to the above mentioned diseases but whose genetic mutation is not yet known. The aim of this project is also to identify a protein pattern characteristic of each known disease and compare it with the proteomic analyses that will be carried out on samples of patients suffering from the still unknown diseases to help in the choice of treatments to be administered to patients.

4.4 Sample Preparation

Fifteen patients with FMF (5), TRAPS (5), and MKD (5) were evaluated and compared with Fifteen healthy donors (HDs). Disease activity of patients was evaluated at the moment of the sampling by considering clinical manifestations and laboratory tests including C-reactive protein (CPR) and serum amyloid A (SAA). Almost all patients were analyzed in a symptom-free period and during an ongoing treatment. Patients with TRAPS and MKD were in treatment with Canakinumab (except Case No 14 and 15), patients with FMF were under colchicine treatment (except Case No 3),Table 1. The study was approved by the ethics board of G. Gaslini institute.

Table 1: Patients

ID	Disease	Mutation	CPR (mg/dL)	SAA (mg/L)	Drug
1	FMF	M694V/M694V	1,82	145	Colchicine
2	FMF	M694V/V726A	1,16	26	Colchicine
3	FMF	M694V/V726A	<0,46	15,3	None
4	FMF	M680I/V726A	<0,46	14,2	Colchicine
5	FMF	M694V/M694V	<0,46	4,2	Colchicine
6	TRAPS	C55Y	<0,46	2,07	Canakinumab
7	TRAPS	C88Y	<0,46	1,8	Canakinumab
8	TRAPS	C52Y	<0,46	5,39	Canakinumab
9	TRAPS	T50M	<0,46	9,94	Canakinumab
10	TRAPS	C55Y	<0,46	3,86	Canakinumab
11	MKD	L265R/V377I	<0,46	4,1	Canakinumab
12	MKD	C605insG/V377I	1,21	135	Canakinumab
13	MKD	C785_790delC/V377I	<0,46	7,2	Canakinumab

14	MKD	V310M/V377I	3,08	130	None
15	MKD	V377I omo	<0,46	16,8	None

Peripheral blood mononuclear cells (PBMCs) of patients and controls were isolated by gradient (Ficoll Hystpaque) and then purified with CD14 microbeads (Miltenyi Biotec) positive selection as indicated in the manual. Monocytes were analyzed in two different conditions: i) unstimulated after the sampling, ii) after LPS stimulation. This latter condition was used to test the responsiveness of monocytes to a TLR ligand in order to minimize the impact of disease activity and/or ongoing treatment. For this reason, monocytes were cultured for 4h in RPMI 1640 with 5% fetal calf serum in presence or absence of lipopolysaccharide (LPS, 100 ng/ml)[38,39].

Tryptophan fluorescence emission assay was used for protein quantification.

Samples with protein amounts below 15 µg were discarded. The samples were processed following the method proposed in the literature, called in-Stage Tip[40]. The solubilization solvent, however, has been modified to obtain a better lysis of the sample.

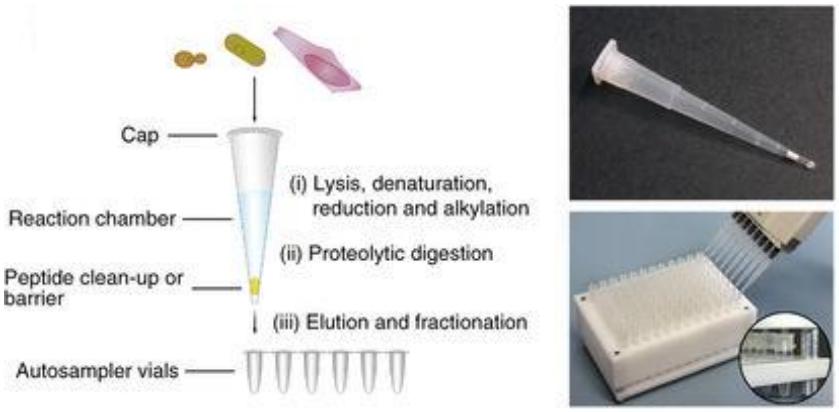

Figure 10: *In Stage Tips sample preparation method.*

The protein pellet is solubilized in 25 µl of 2% (W/V) SDC, 10 mM TCEP, 40mM CAA, 100 mM Tris pH 8.5. In a single step the cells are then lysed, reduced and alkylated through three

cycles of freezing the sample for 10 minutes at -20 °C and subsequent heating to 99 °C for 10 minutes and sonication in an ultrasonic bath for 30 minutes. The sample was transferred into the tips containing 2 disks of a styrenedivinylbenzene resin modified with sulfonic acid groups to make it hydrophilic. The SDB-RPS disks are intended to purify the peptides and normalized their quantity in the different samples because the disk area is proportional to the amount of peptides retained, 16 gauge diameter is able to retain 7 µg peptides (Figure 10). The lysate was diluted 1:2 with 10% (v/v) ACN, 25 mM Tris pH 8.5 and the peptide digested protein material with 0.5 µg lysine C and 0.5 µg trypsin at 37 °C overnight. The sample was diluted 1:2 because trypsin works well with up to 1% SCD and then acidified with 100 µl trifluoroacetic acid (TFA) to block the enzyme action. The peptides are eluted with 60 µl of 5% ammonium hydroxide and 80% acetonitrile (ACN) in water. The sample was lyophilized and taken back with 10 µl of 98% water (H2O), 2% ACN, 0.1% formic acid (AF). Considering that the containment capacity of each disc is 7 µg peptides, the final concentration of the sample injected by mass is approximately 1.4 µg/µl.

Tryptophan fluorescence emission assay.

Protein concentrations were determined by tryptophan fluorescence emission at 350 nm using an excitation wavelength of 295 nm.

Briefly, 1 µl of sample was solubilized in 200 µl of 8 M urea and 0.1 µg/µl tryptophan was used to build a standard calibration curve (0.25–1.5 µl).

Protein concentration in samples was estimated considering the emission of 0.1 µg/µl tryptophan equivalent to the emission of 7 µg/µl of human protein extract, assuming that tryptophan accounts for 1.3% on the human protein amino acid composition, on average. In bottom-up proteomics, protein determination is the prerequisite for optimal protein digestion and the quantification of the generated peptides is important for mass-spectrometry-based analysis.

The direct quantitation of tryptophan by measuring its fluorescence is an easy, fast and reliable way to quantify total protein and total peptide.

4.6 Mass spectrometer set up

Peptides were analyzed by nano-UHPLC-MS/MS using an Ultimate 3000 RSLC with easy-spray column (75 µm x 500 mm, 2 µm particle size, Thermo Scientific) and with a 180 minute non-linear gradient of 5-45 % solution B (80% acetonitrile, 20% H_2O with 5% dimethyl sulfoxide and 0.1% formic acid) at a flow rate of 250 nl/min, Table 2. Solution A (98% H_2O, 2% acetonitrile with 0.1% formic acid) varies as a function of B solution.

To reduce the counter pressure in the column the separation of peptides takes place at a temperature of 55 °C.

Eluting peptides were analyzed using an Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific Instruments, Bremen, Germany). Orbitrap detection was used for both MS1 and MS2 measurements at resolving powers of 120000 and 30000 (at m/z 200), respectively (Table 3).

Data-dependent MS/MS analysis was performed in top speed mode with a 3 sec. cycle time, during which precursors detected within the range of m/z 375−1500 were selected for activation in order of abundance. Quadrupole isolation with a 1.4 m/z isolation window was used, and dynamic exclusion was enabled for 45 s. Automatic gain control targets were 2,5e5 for MS1 and 5e4 for MS2, with 50 and 45 ms maximum injection times, respectively.

The signal intensity threshold for MS2 was 1e4. HCD was performed using 30% normalized collision energy. One microscan was used for both MS1 and MS2 events.

Table 2: HPLC gradient.

RT	% B
0	2
3	2
7	5
140	30
165	45
165.1	80

170	80
170.1	2
180	2

The addition of DMSO in solvents has allowed the identification and quantification of 10% more proteins; increasing the nebulization of the analytes in the source.

Table 3: MS parameters.

Ion Source	
Ion Source Type	Nano ESI
Positive Ion (V)	2200
Ion Transfer Tube(°C)	300
Master Scan MS-OT	
Detector Type	Orbitrap
Resolution	120000
Scan Range (m/z)	375-1500
AGC Target	2.5e5
Maximum Injection Time (ms)	50
Dynamic Exclusion	
Exclude after n times	1
Exclusion Duration (s)	45
Intensity Threshold	1.0e4
DDA Mode	Cycle Time

ddMS^2 MS OT	
Isolation Mode	1.4
Activation Type	HCD
HCD Energy (%)	30
Detector	IT
Speed	Rapid
AGC Target	5.0e4
Maximum Injection Time (ms)	45

4.7 Data analysis

MaxQuant software was used to process the raw data, setting a false discovery rate (FDR) of 0.01 for the identification of proteins, peptides and PSM (peptide-spectrum match); moreover, a minimum length of six amino acids for peptide identification was required.

Andromeda engine[41], incorporated into MaxQuant software[42], was used to search MS/MS spectra against the Uniprot human database. For protein digestion, allowing for cleavage N-terminal to proline, trypsin was chosen as enzyme. Cysteine carbamidomethylation was selected as fixed modification, whereas acetylation protein N-terminal, oxidation (M) and deamidation (N, Q) have been selected as variable modifications. A tolerance of 7 ppm was set for the mass deviation of the precursor ions, while the maximum mass deviation for MS2 events was 0.5 Da.

Algorithm MaxLFQ[43] was chosen for protein quantification, with the activated option 'match between runs' to reduce the number of the missing proteins. All bioinformatics analyses were done with the Perseus software of the MaxQuant computational platform.

The Perseus software platform supports biological and biomedical researchers in interpreting protein quantification, interaction and post-translational modification data. Perseus contains a comprehensive portfolio of statistical tools for high-dimensional omics data analysis covering normalization, pattern recognition, time-series analysis, cross-omics comparisons and multiple-hypothesis testing. A machine learning module supports the classification and validation of patient groups for diagnosis and prognosis, and it also detects predictive protein signatures[44,45].

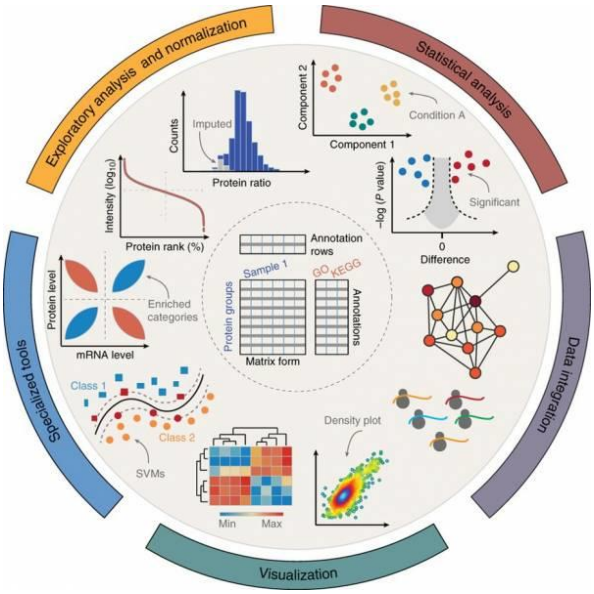

Figure 11: *Data analysis scheme.*

Protein groups were filtered to require 70% valid values in at least one experimental group. The label-free intensities were expressed as base log2, and empty values were imputed with random numbers from a normal distribution for each column, to best simulate low abundance values close to noise level. For each group, a t-test with permutation-based FDR of 0.05 and a s0 of 0.1 was used.

The Venn diagram of identified proteins was calculated using an online tool. A Student's t-test analysis (FDR of 0.05 and S0 of 0.1) was performed to characterize differences and reciprocal relationships between the two groups in each disease. The significant proteins were plotted using a Volcano plot.

We also performed a Cytoscape analysis to identify molecular interaction pathways and biological functions of proteins founded in monocytes.

Cytoscape is an open source software platform for visualizing molecular interaction networks and biological pathways and integrating these networks with annotations, gene expression profiles and other state data. Although Cytoscape was originally designed for biological research, now it is a general platform for complex network analysis and visualization. Cytoscape core distribution provides a basic set of features for data integration, analysis, and visualization. Additional features are available as Apps (formerly

called Plugins). Apps are available for network and molecular profiling analyses, new layouts, additional file format support, scripting, and connection with databases[46].

In order to study modulated protein interaction networks STRING database was used in Cytoscape; the analysis was integrated with a DISEASES database query in order to highlight proteins already known to be related to each diseases and proteins not yet linked to the diseases. STRING provides a network view on functional protein associations, based on direct (physical) and indirect (functional) protein-protein interactions[47]. This tool uses both known and predicted interactions (derived from indirect evidences of gene co-occurrence, fusion events, co-expression and conserved neighborhood), to whom a confidence score is attributed by comparing to a reference set of trusted true associations (KEGG database).

ClueGo and CluePedia plugins of Cytoscape were used to build a functionally grouped networks based on GO terms, Reactome, Kyoto Encyclopedia of Genes and Genomes and WikiPathways with terms as nodes linked based on their κ score level (≥0.3) . Only the most significant term in each group is presented using ClueGo. Associated proteins were visualized using CluePedia.

4.8 Results

The protein content of monocytes of patients and healthy donors, once isolated from peripheral blood, is analyzed and quantified by tandem mass spectrometry.

For each disease (FMF, TRAPS, MKD), monocytes from patients were compared to healthy donors (HD) in two different conditions: i) un-stimulated monocytes, ii) LPS-stimulated monocytes. Each group consists of 5 biological replicates belonging to 5 different patients. Globally, at a peptide and protein false discovery rate (FDR) of 1%, we identified about 5000 proteins, with an average of more than 4000 proteins per group. The numerical distribution of each experiment is reported through the Venn diagram, Figure 12. Through the Venn diagram we do not see the modulated proteins but the presence / absence of the proteins in the different experimental conditions.

In unstimulated and LPS-stimulated FMF samples 4507 proteins are common among all conditions; 25 and 22 proteins are exclusive of LPS-stimulated and untreated FMF samples respectively.

In unstimulated and LPS-stimulated TRAPS samples 4092 proteins are common among all conditions; 26 and 36 proteins are exclusive of LPS-stimulated and untreated TRAPS samples respectively.

In unstimulated and LPS-stimulated MKD samples 4223 proteins are common among all conditions; 79 and 41 proteins are exclusive of LPS-stimulated and untreated MKD samples respectively.

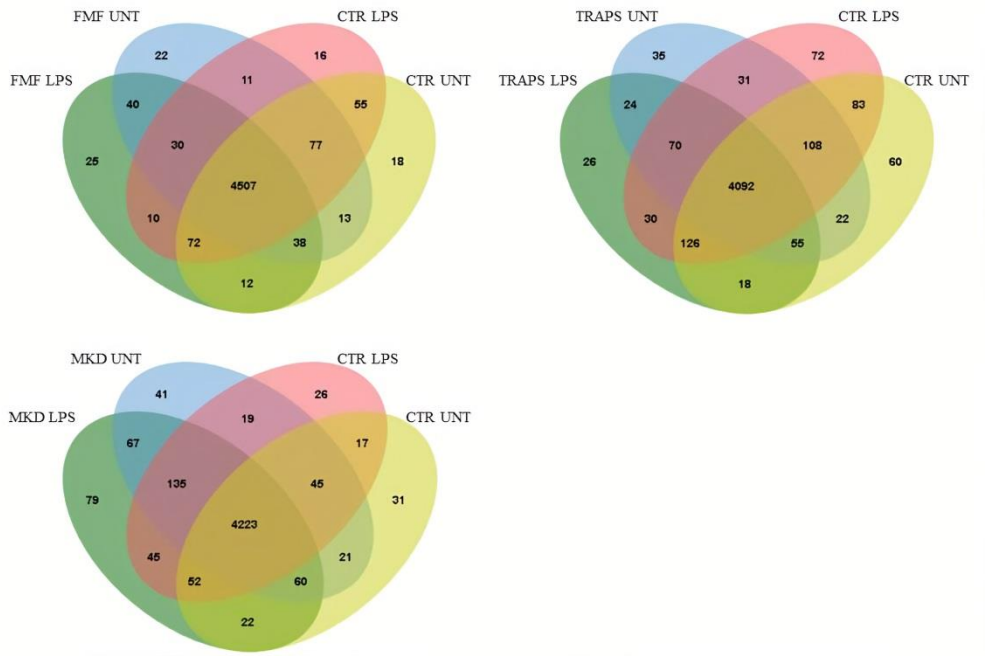

Figure 12: *Venn diagrams of unstimulated and LPS-stimulated FMF, TRAPS and MKD samples.*

As will be described later, the differences are associated for Gene Ontology to the disease and to specific processes of immune response.

In Figure 13 for each disease the t-test was used to identify the significantly modulated proteins, comparing patients and healthy controls, for unstimulated (left panels) and LPS stimulated (right panels) cells. In the case of FMF the unstimulated cells modulated 115 proteins while in the LPS-stimulated cells produced 99 proteins. Similarly for the TRAPS the proteins are 16 and 90 as well as for MKD are 36 and 415 in unstimulated and LPS-stimulated cells, respectively.

The T-test is a parametric statistical test with the aim of verifying whether the average value of a distribution deviates significantly from a certain reference value.

The statistically significant proteins from T-test are highlighted in the Volcano Plot as black pellets. For more clarity please refer to the supplementary tables where for each statistically significant protein the value of the p-value and the T-test difference is reported. The Volcano plot is constructed by plotting on the y-axis the negative of the logarithm on base 10 of the

p-value and on the abscissae the logarithm of the fold-change between the two conditions. In this type of graph the lower values of p-value (corresponding to high significance) are higher. A protein is more modulated when it has a high ordinate value and a value very different from zero on the abscissae.

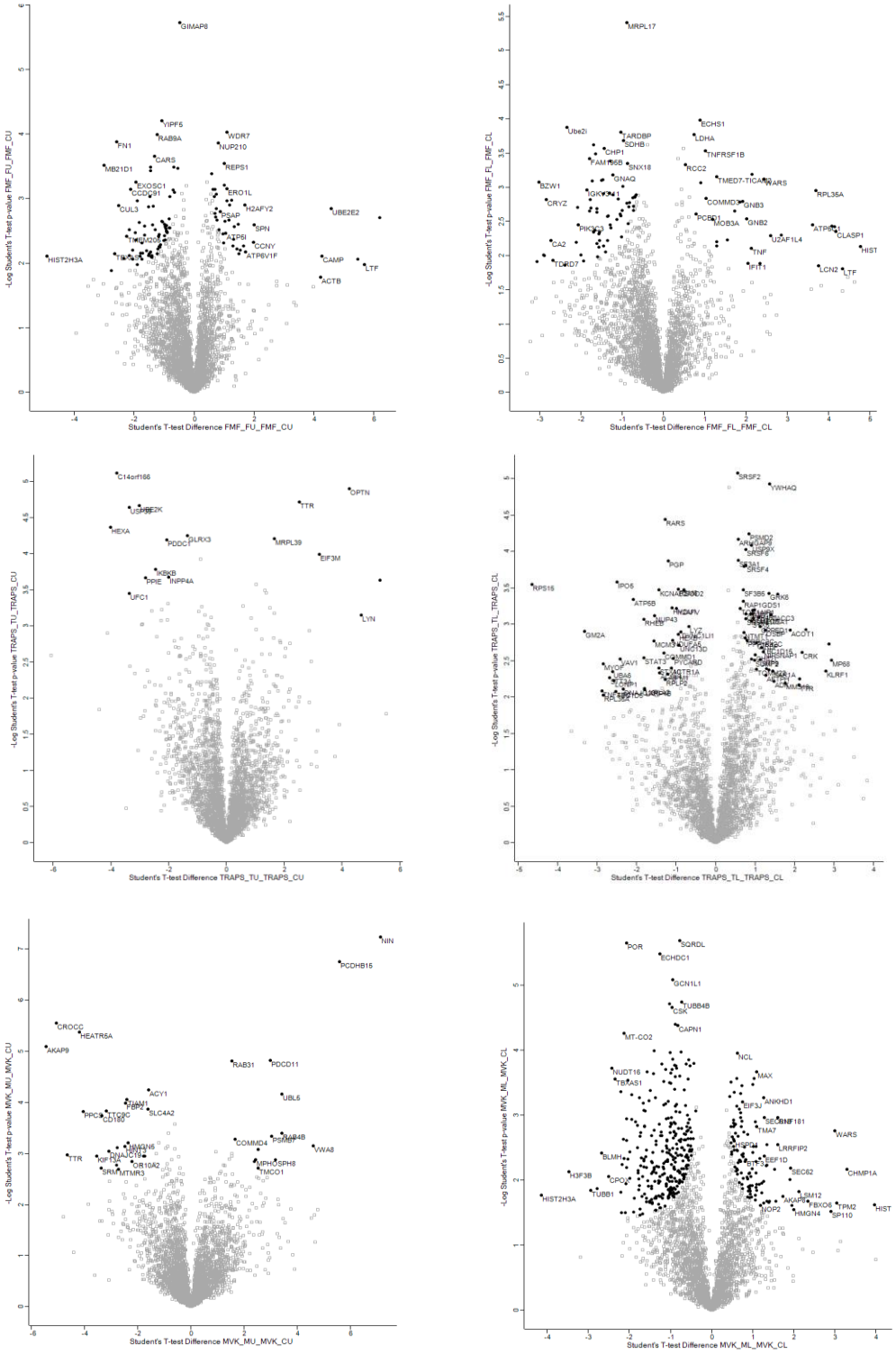

Figure 13: *Volcano plots of untreated and LPS-treated patients versus untreated and LPS-treated healthy donors in FMF, TRAPS and MKD samples.*

In supplement tables 1-6 at the end of the thesis are reported the significantly modulated proteins.

However, LPS treatment maximizes differences between up-regulated proteins in monocytes of HRF patients and HD, independently from the disease activity and/or ongoing treatments.

The Tables below summarizing interesting proteins significantly modulated in FMF, TRAPS and MKD patients compared with healthy controls in monocytes untreated or treated with LPS. In red and blue are the proteins up and down regulated respectively.

Table 4: Interesting modulated proteins in untreated and LPS-treated FMF patients. In red and blue are the proteins up and down regulated respectively.

Genes related to FMF	UNT	LPS	Interesting genes	UNT	LPS	Various	UNT	LPS
MEFV			IFIT1			MRPL17		
PYCARD			IFIT2			RPL9		
IL1RN			IFIT3			RPL35A		
IL1B			ISG15			RPL36		
CASP1			MB21D1			RPS15		
IL18			PRDX6			PPIE		
TNF			GNAQ			YY1		
PSTPIP1			SHC1			DAXX		
RHOA			PTPN11			HAT1		
IKBKB			MIF			HIST2H3A		
MAPK3			CDK9			HIST1H2BK		
STAT3			FN1			RB1		
STAT5B			ELANE			SHDB		
PIK3R1			PPP1CA			GUSB		
PKN1			SHARPIN			LYZ		
PRKCA			RHOT1			Ca2		
TLR2			TMEM173			TUBB2A		
S100A12			TAX1BP3			SPN		
C5AR1			CUL3			CCNY		
DNM1L						ERAP1		
TNFRSF1B						TTC37		
TTR						GRHPR		
TPT1						ELL		
TRIM21								
B2M								
SLC25A1								

Table 5: Interesting modulated proteins in untreated and LPS-treated TRAPS patients. In red and blue are the proteins up and down regulated respectively.

Genes related to TRAPS	UNT	LPS	Interesting genes	UNT	LPS	Various	UNT	LPS
MEFV			TTR			SF3A1		
PYCARD			CRK			SF3B5		
TNFRSF1B			YWHAQ			SRSF4		
ADAM17			OPTN			SRSF11		
STAT3			VAV1			SRSF6		
TPT1			RAP1A			NUP43		
LYZ			USP9X			RPS15		
IKBKB			ACTR1A			LYPLA2		
			SRP9			RPL35A		
			EIF3M			RPLP2		
			LYN			UFC1		
						C14orf166		

Table 6: Interesting modulated proteins in untreated and LPS-treated MKD patients. In red and blue are the proteins up and down regulated respectively

Genes related to MVD	UNT	LPS	Interesting genes	UNT	LPS	Various	UNT	LPS
IL1RN			LYN			RPL8		
IL1B			SYK			CLNS1A		
PSTPIP1			GNA13			MRPL44		
MAPK1			GNA12			HIST2H3A		
MAP2K3			PTPN18			PIGS		
CD14			PTPN11			SEPHS1		
NFKB1			VAV1			QARS		
RHOA			PSMB4			TARS		
LRPAP1			ACTR1A			YARS		
TMEM173			ACTR3A			HSD17B10		
TLR2			VASP			HSPD1		
TLR8			SEC61B			LMNB1		
RAB14			XRCC5			FLOT1		
RAB7			SAMHD1			PGRMC1		
RAB1A			NEDD8			AP1G2		
ATG7			USP4			DCTN6		
Mefv			RBCK1			MFN2		
Pycard			TCEB2			MTCH2		
TNFRSF1B			ZNRF2			HNRNPF		
Casp1			UBE3C			CTSD		
IL18			TRIM4			SRSF3		
TNF			LMAN2			CPNE2		
			RAP2B			MTDH		
			NCF4			CA2		
			SND1			MYO1F		
			FBXO6			MYO1G		
			CPT1A			MYO5A		
			CPT2			LRPPRC		
			ECI1			TCP1		
			CSK			TOP2B		
			DNM2			NCL		
			VPS16			NME2		
			VPS33A			ELAVL1		
			VPS29			UQCRC2		
			VPS18			NDUFA9		
			STX7			NDUFS2		
			ATP6V1A			CBX5		
			CTSZ			SLC25A3		
			XPO1			HMGA1		
			ARL8B			LSM12		
			CAPZA2			MMAB		
			VAMP7			MSN		
			AP3M1			IDH3G		
			COPA			ACADVL		

In order to highlight the differences in protein expression the graphs in Figure 14 have been constructed. The chart on the ordinates shows the ratio between the intensities of the proteins present in the untreated monocytes of the patients and those present in the control and on the abscissae the ratio between the intensities of the proteins present in the LPS-stimulated monocytes of the patients compared to healthy donors.

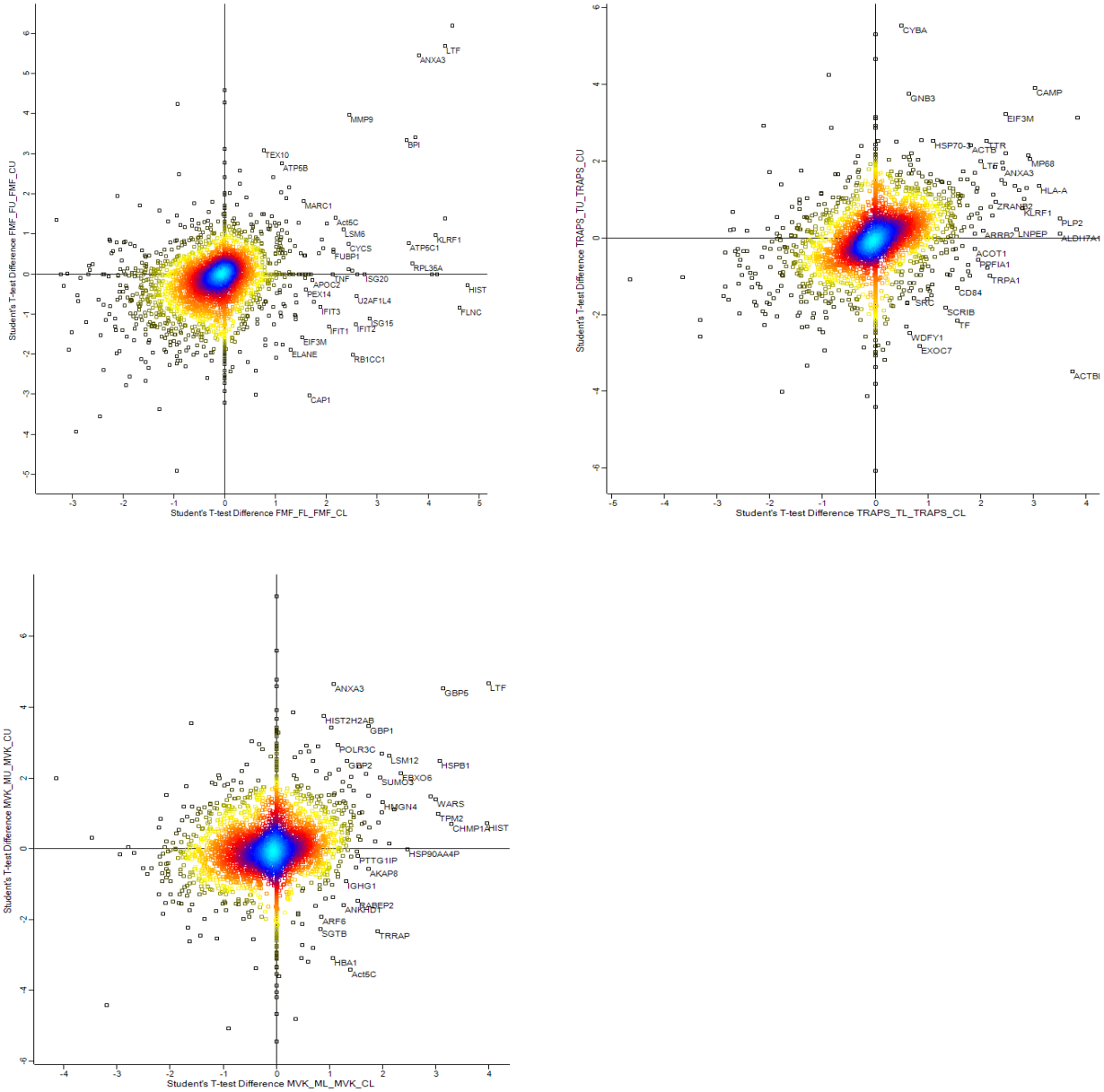

Figure 14: *Stimulus-independent and modulated proteins in patients and donors.*

Dividing the graph into quadrants we can say that:

- in the upper right quadrant, we find the up regulated proteins characteristic of the disease, independently from the stimulus,
- in the lower left quadrant are the up regulated proteins in healthy controls, regardless of the stimulus.

In this way it was possible to compare the LPS-treated and untreated samples by disease in a single graph because they were both normalized with healthy controls.

In all pathologies diseases, several proteins known to be related to the disease have been found significantly modulated, confirming the method used.

A Cytoscape analysis with ClueGo app using STRING and DISEASE database was performed to identify proteins interactions and to highlight proteins already related to autoinflammatory diseases.

The figures below show the modulated proteins and their interactions for untreated and LPS-treated FMF, TRAPS and MKD diseases.

In particular the Figure 15 shows the modulated proteins between LPS-treated FMF patients and LPS-treated healthy donors; the Figure 16 shows the modulated proteins between untreated FMF patients and untreated healthy donors; the Figure 18 shows the modulated proteins between LPS-treated TRAPS patients and LPS-treated healthy donors; the Figure 19 shows the modulated proteins between untreated TRAPS patients and untreated healthy donors; the Figure 21 shows the modulated proteins between LPS-treated MKD patients and LPS-treated healthy donors; the Figure 22 shows the modulated proteins between untreated MKD patients and untreated healthy donors.

The figures 15, 16, 18, 19, 21, 22 should be read as follows.

The proteins represented with a circle are those known from literature, while those represented with a square are those not yet related to the diseases under examination. In red the proteins that are up regulated in the disease samples and in blue those up regulated in the samples of healthy donors. The size of the nodes is a function of the FDR.

Figure 17, Figure 20, and Figure 23 show the most modulated pathways in FMF, TRAPS, MVK diseases respectively. The three figures have the following reading. The large circles represent the modulated pathways, while the small circles represent the significant proteins related to pathways. In red are the significant up-regulated proteins in patients and in blue are the significant up-regulated proteins in healthy donors. Pathways are also colored in proportion to the expression of the proteins that belong to them. The size of the circles depends on how many proteins belonging to the pathway have been quantified.

To give an explanatory example in Figure 17 FasL pathway/HSP regulation correlates directly with four proteins FAF1, DAXX, RB1, TNF, of which only one is up-regulated FMF1 in healthy donors because it is blue in color, the other 3 are up-regulated in diseased samples. The pathway color is for ¾ red (3 out of 4 proteins are colored red) and ¼ is colored blue. The RAC1/PAK1/p38/MMP2 pathway correlates 8 proteins of which two are up-regulated blue in healthy control and six are up-regulated red in diseased samples; the percentage color of the pathway is the same as above but the size is larger because it correlates with a double number of proteins. The pathway Regulation of Microtubule Cytoskeleton correlates with three proteins, all of which are up-regulated in diseased samples and therefore its color is exclusively red.

In FMF monocytes treated with LPS , Figure 14, pro-inflammatory cytokines such as IL1β and TNFα were found up-regulated; also MEFV and RhoA are differently modulated according to the literature[31,48,49], in this case MEFV is up-regulated and RhoA down-regulated independently by the stimulus. Some proteins related to RhoA pathway are differently modulated such as Fibronectin (FN1), GNAQ, PTPN11 that is able to activate RhoA but after stimulation with LPS is down-regulated. PTPN11 is also able to regulate MAPKinase pathway and PI3K-Akt pathway. A very interesting result is the high expression of several proteins related to the type 1 interferon (IFN-1) pathway such as IFIT1, IFIT2, IFIT3, ISg15 all up-regulted after LPS stimulation. Also cGAS (MB21D1) that regulate the IFN-1 production in up-regulated after LPS stimulation.

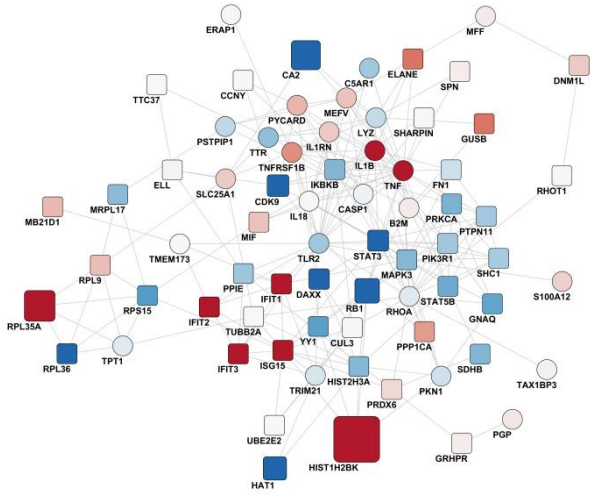

Figure 15: *Protein network in LPS-treated FMF patients.*

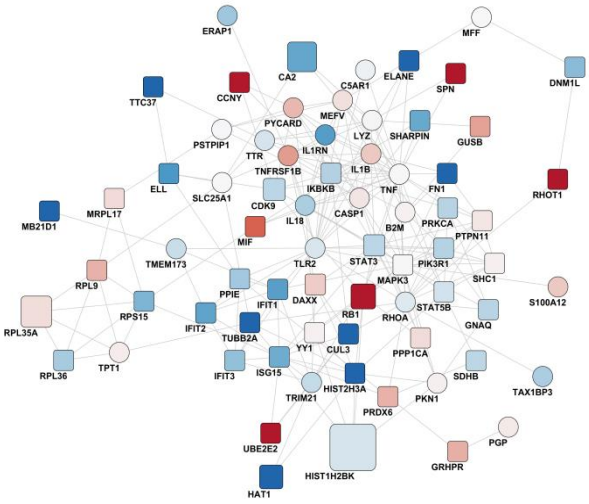

Figure 16: *Protein network in untreated FMF patients.*

The pathway analysis, Figure 17 highlight the regulation of microtubule cytoskeleton differently regulated in FMF patients and different pathway correlated with cytokines.

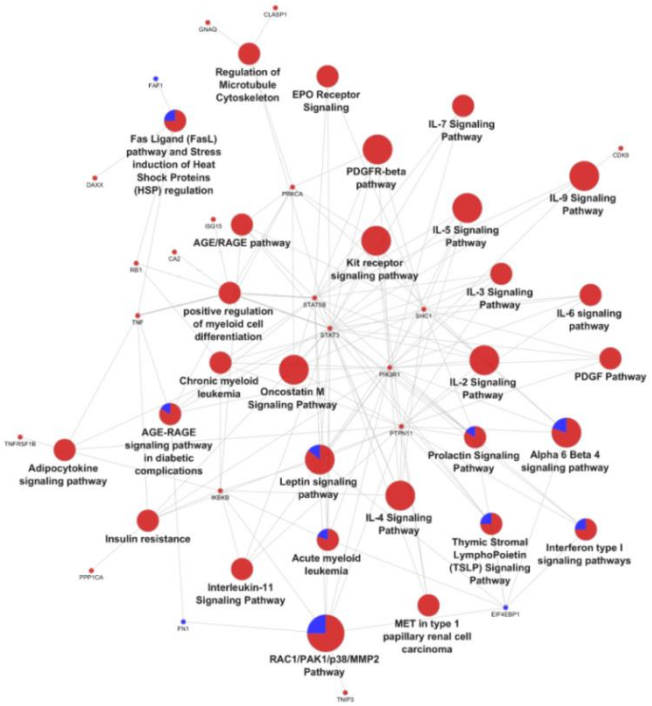

Figure 17: *Modulated pathways in FMF disease.*

From Figure 18 and Figure 19 it is interesting to see that in samples of FMF patients the proteins that most modulate between the unstimulated and PLS stimulated samples are OPTN related to autophagy and maintenance of the Golgi apparatus, ADAM17 related to leukocyte degranulation, LYN related to chemokine signaling pathway and leukocyte degranulation and UFC1 associated to an autosomal recessive disorder apparent soon after birth or in early infancy.

Other significantly modulated proteins are involved in vesicle transport (ACTR1A, TPT1, NUP43), and MAP Kinase pathway (RAP1A, VAV1, CRK).

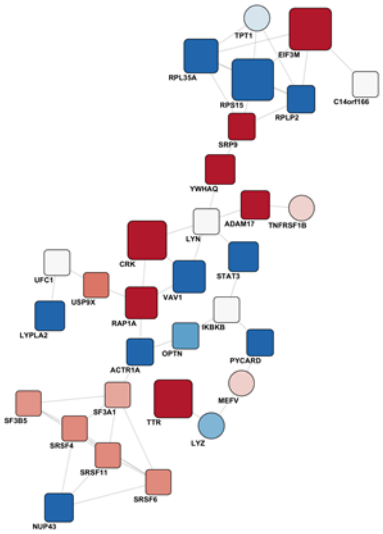

Figure 18: *Protein network in LPS-treated TRAPS patients.*

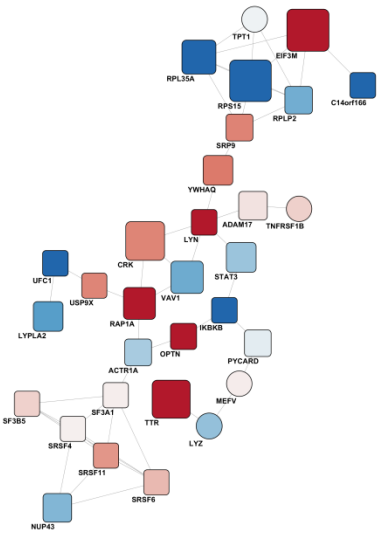

Figure 19: *Protein network in untreated TRAPS patients.*

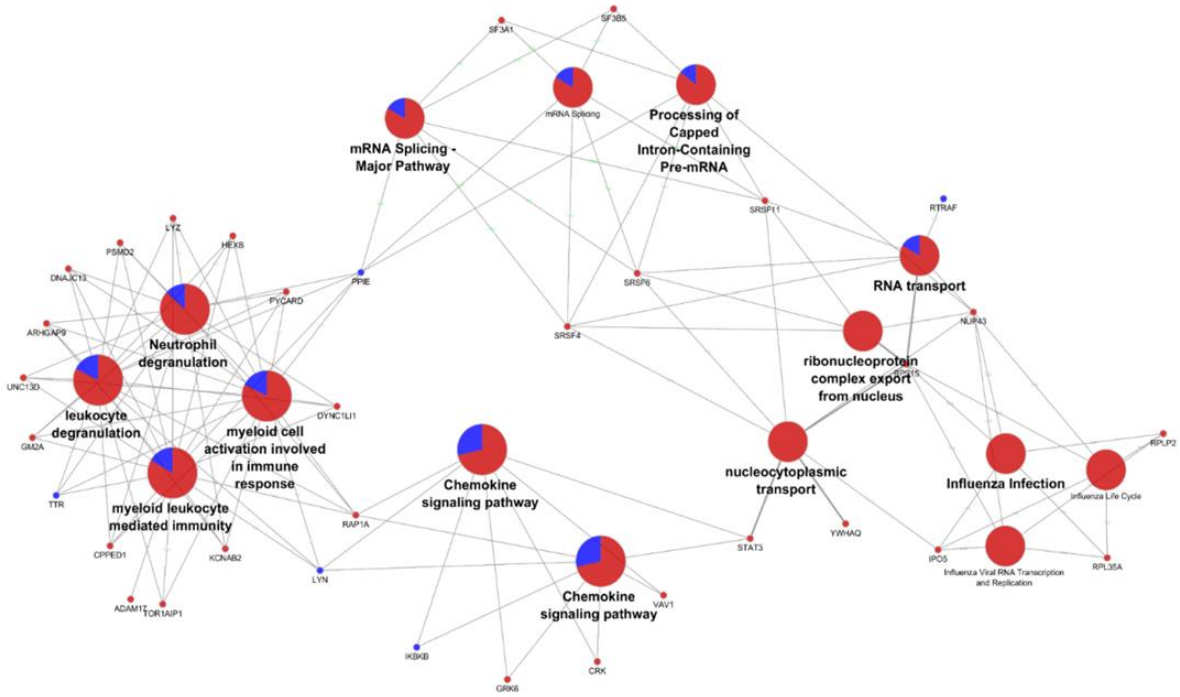

Figure 20: *Modulated pathways in TRAPS disease.*

The MKD complexity is evident from the number of proteins and pathways involved, Figure 21, Figure 22, and Figure 23. In fact both metabolic and inflammatory pathway results regulated in patients. Prenylation is found to be related to this disease and in fact several are the proteins linked to prenylation founded regulated in patients monocytes mevalonate kinase deficiency. Different regulated proteins are related to the fatty acid metabolism (CPT1A, CPT2, ECI1, HADHA, SLC25A1, HSD17B4, ACAA2), many are those related to the transport of vesicles (ACTR1A, LMAN2, VPS16, VPS33, VPS29, VPS18, CTSZ, VAMP7, AP3M1, COPA, COPB1, SEC23B, AP-A1, DYNC1H1), or at the mitochondrial function (QARS, TARS, YARS, HSD17B10, HSPD1, LRPPRC, UQCRC2, NDUFA9, NDUFS2, SLC25A3).

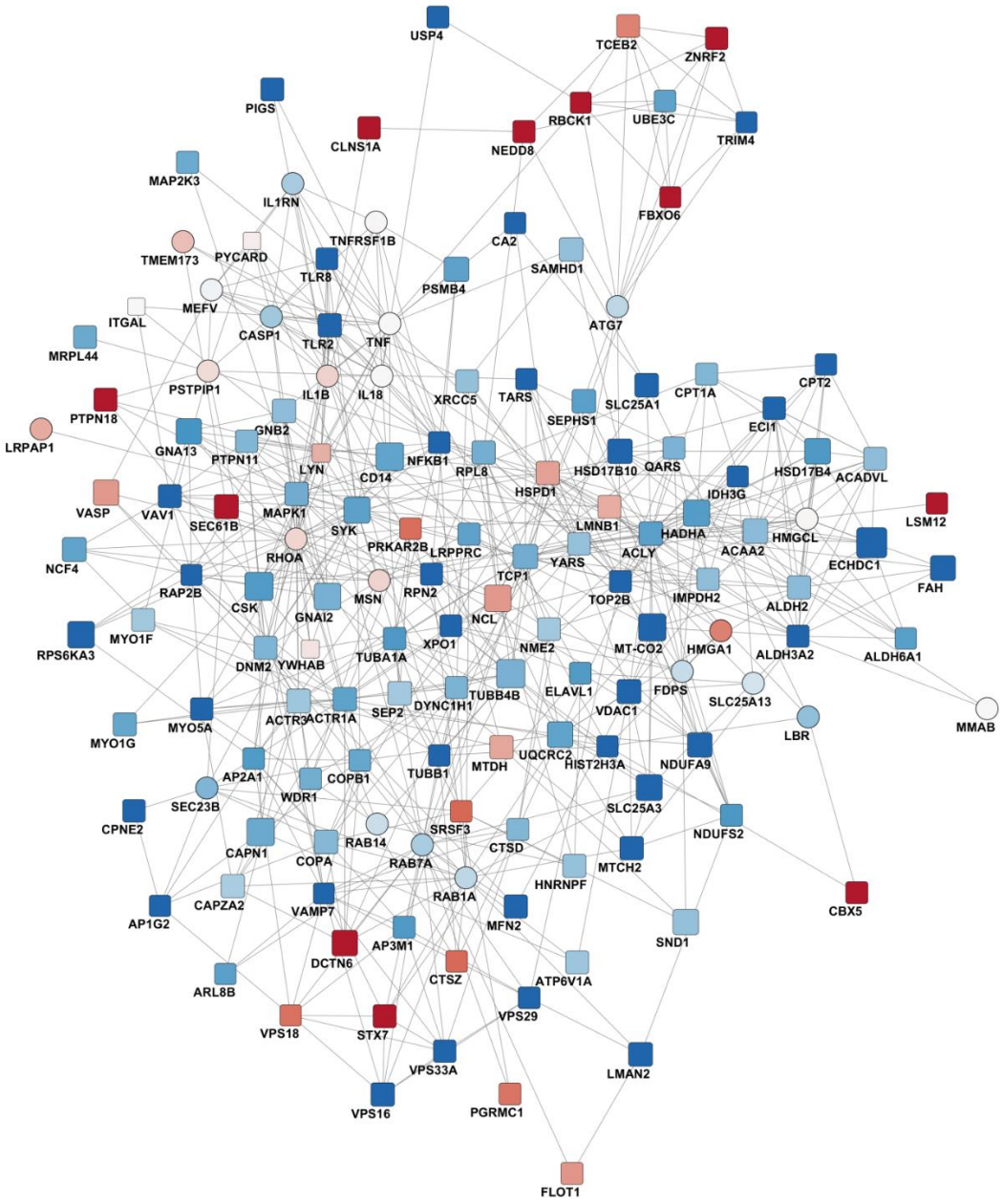

Figure 21: *Protein network in LPS-treated MKD patients.*

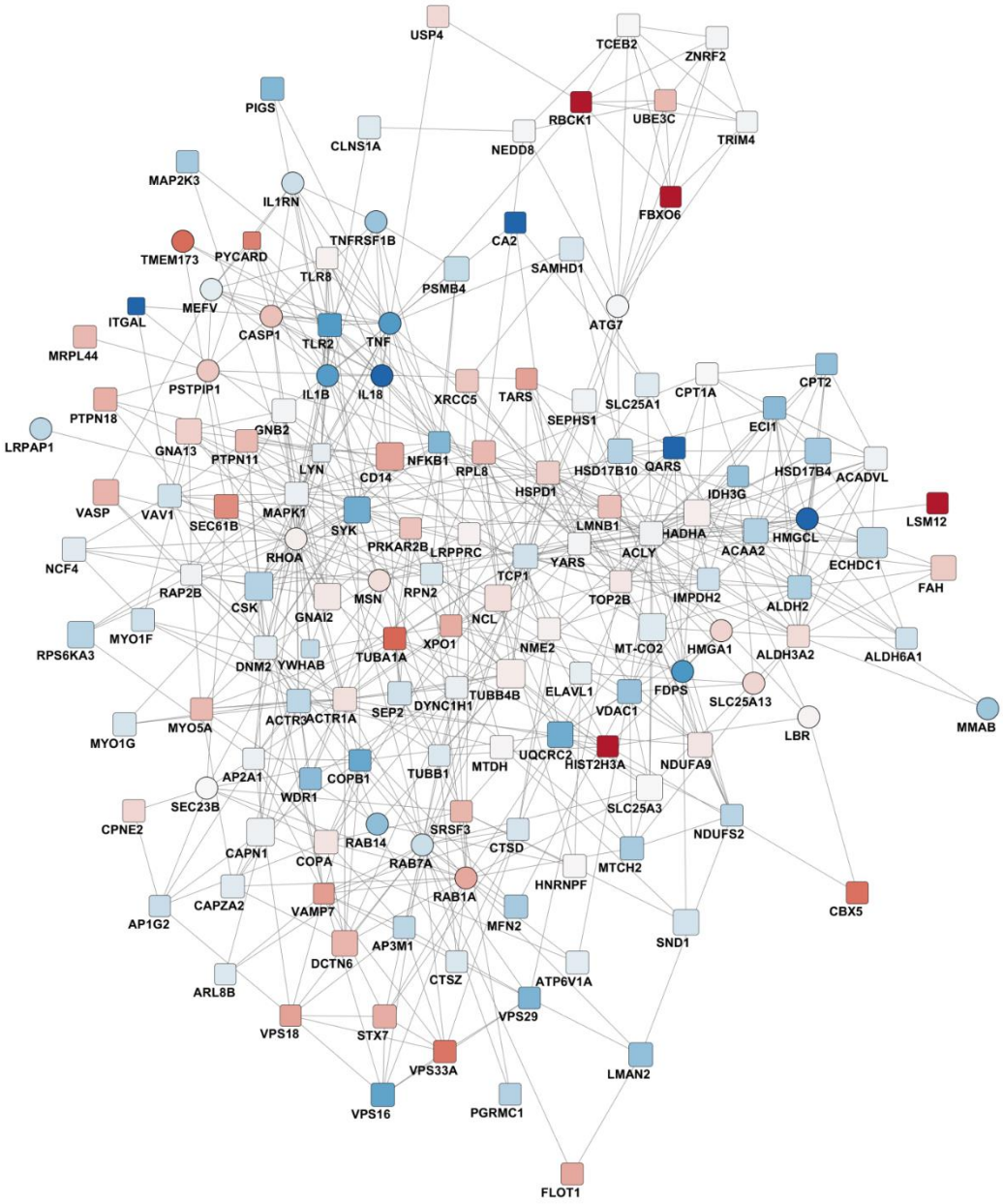

Figure 22: *Protein network in untreated MKD patients.*

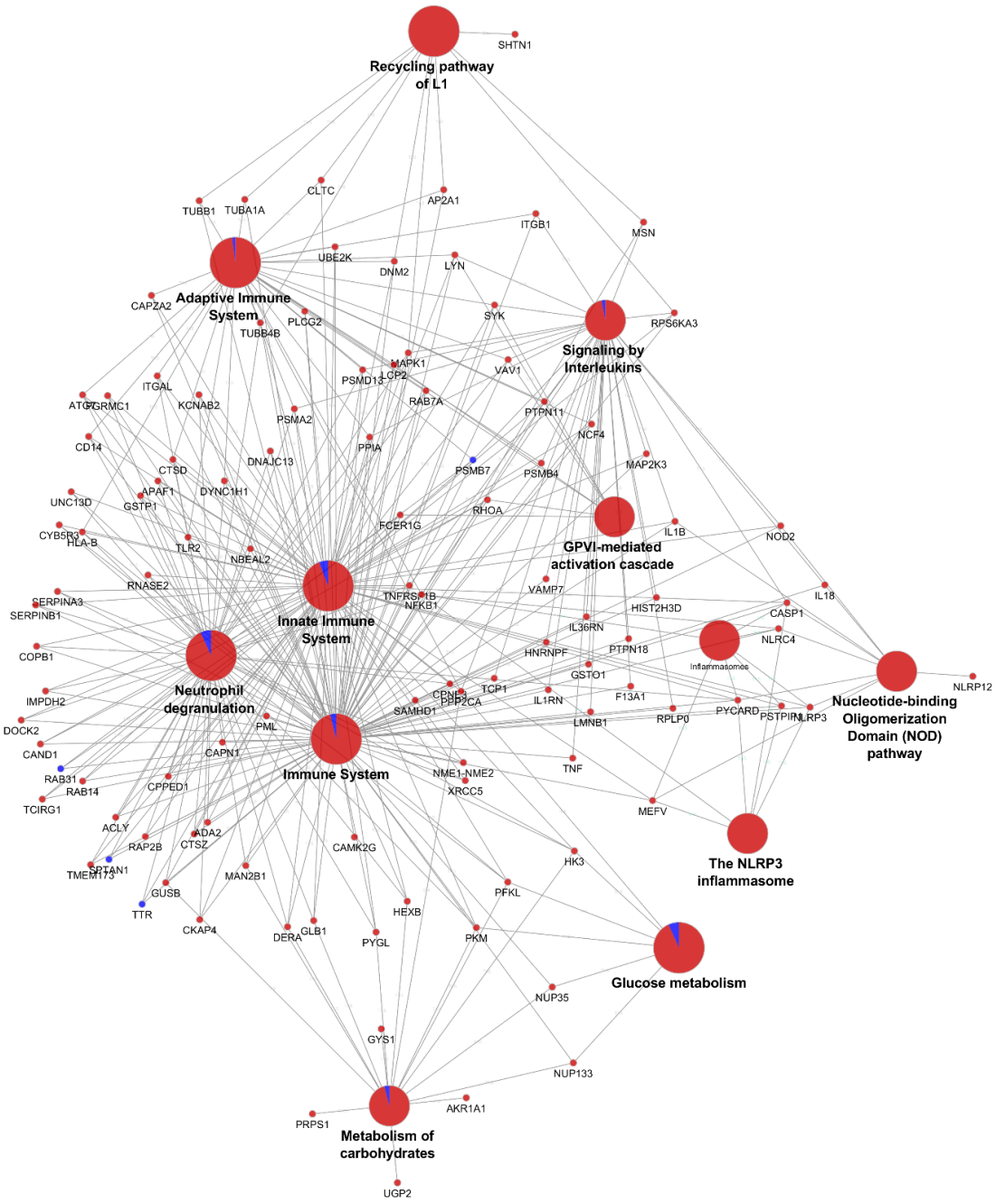

Figure 23: *Modulated pathways in MKD disease.*

4.9 Discussion

The study of the proteome of autoinflammatory diseases was carried out in collaboration with the pediatric and rheumatology clinical operating unit.

My research work consisted in the development of a method ranging from sample preparation to data analysis for high-resolution mass spectrometry-based proteomics. We have identified modulated proteins in patients with FMF, TRAPS and MVK and the pathways most affected by the disease. The effort made was not only to find a method for sample preparation and mass spectrometry analysis but also to provide a sufficiently clear interpretation of the data to doctors and biologists.

It was not possible to compare the different diseases with each other because the samples and mass spectrometer analysis were performed in different times and their comparison was deemed unreliable.

From this consideration we then decided to introduce isotopic peptides that allow the comparison between analysis carried out at different times, as is already the case for metabolomics experiments.

The point of arrival of my work is the starting point of the pediatric and rheumatology clinical operating unit; on the results obtained from my PhD project, colleagues will direct at least part of their future research work. The data will then be validated by colleagues through other types of analysis such as cytofluorimetric methods.

It cannot be excluded that mass spectrometry can also be used for the validation of possible biomarkers with targeted methods such as MRM.

For a completeness of writing below I report part of the conclusions that biologists and doctors have drawn from the data obtained during the project. I want to emphasize that this part was not done by me.

In LPS-treated FMF monocytes, high expression of IL1β, TNFα and proteins related to the type 1 interferon pathway, such as IFIT1/2/3 and ISG15, recalls treatment targets of FMF[50,51,52]. The higher concentration of the IL-6-induced STAT3/CDK9 complex in LPS-treated HD monocytes suggests a lower activation of the IL-6-mediated inflammatory response in

FMF monocytes, thus acting against the use of flavonoids as CDK9 inhibitors in FMF patients. However, an increased oxidative stress in FMF patients seems confirmed by the high concentration of the macrophage migration inhibitory factor (MIF) in untreated FMF monocytes[53].

In LPS-treated TRAPS monocytes, up-regulation of the shedding protein ADAM17 suggests an ineffective attempt to correct the well-known TNF receptor 1 shedding defect[54]. Moreover, the up-regulated amyloidogenic TTR recalls the highest prevalence of systemic amyloidosis in TRAPS patients[55]. Differently from FMF data, pyrin inflammasome is not up-regulated but even inhibited by the up-regulated 14-3-3 protein theta, a well-known pyrin inhibitor.

MKD is a metabolic autoinflammatory disease and showed a higher complexity of proteins regulation. Essentially, the significantly modulated pathways are involved in the intermediate cell metabolism, ubiquitination, intracellular trafficking, and inflammation. In particular, NEDD8, ZNRF2, FBXO6, LSM12 and RBCK1, also known as HOIL-1, mediate ubiquitination, and SEC61B, STX7, DCTN6, AP1G2, TUBB1, MYO5A, LMAN2, VPS16/29/33A and CPNE2 are linked to the intracellular trafficking regulation, both emerging pathways in autoinflammation[56,57]. On the other hand, up-regulated PTPN18 can bind PSTPIP1, even modestly un-regulated in untreated and LPS-treated MKD monocytes, whose mutations have been related to the well-known IL-1-mediated autoinflammatory pyogenic sterile arthritis, pyoderma gangrenosum, and acne (PAPA) syndrome. Finally, the up-regulated CBX5, also known as haptoglobin 1 alpha, have been associated with several inflammatory disorders due to its ability to modulate the helper T-cell type 1 and 2 balance and antioxidant properties. Down-regulated proteins in the LPS-treated MKD monocytes are mainly related to the NFkB-mediated inflammatory response (NFKB1, TLR2/8, RAP2B and TRIM4), as well as DNA transcription and RNA nuclear export (CLNS1A, TOP2B, TARS and XPO1), maybe explain the reduced microRNA-mediated silencing complex activity of the mevalonate pathway[58]. These evidences recall the increased susceptibility to infections of MKD patients. Furthermore, several down-regulated proteins participate to the fatty acid oxidation and glycolipid biosynthesis (PIG-S, ALDH3A2, RPN2, CPT2, ECI1) supporting a

prenylation defects of small GTPases in MKD[59,34]. Finally, the modulated MT-CO2, MFN2, MTCH2, IDH3G, NDUFA9, VDAC1, ECHDC1 and SLC25A1/3 proteins confirmed previous evidences of an instable mitochondrial membrane and altered autophagy regulation in MKD monocytes.

5 Establishing a Shareable Spectral MSMS Library and Accurate Mass Retention Time (AMRT) Database for Pediatric Metabolomics Analysis.

5.1 Introduction

Omics technologies aim to identify and quantify huge numbers of compounds extracted from complex matrices, most commonly being biological, plant or food-based materials. MS methods can be tailored for compound classes by manipulating the source parameters, setting conditions on MS/MS triggers or manipulating the pre-MS separation conditions. Alternatively, a completely untargeted approach may be used to provide unbiased profiling. Universally, quick MS/MS scan speeds benefit these analyses.

Metabolomics is an emerging approach that allows the comprehensive study of the low molecular weight molecules within an organism[60,61] despite that the first metabolomic approach dates back to 1506 when physician recognizes pathology on the basis of colors, smells and tastes of urine, Figure 24.

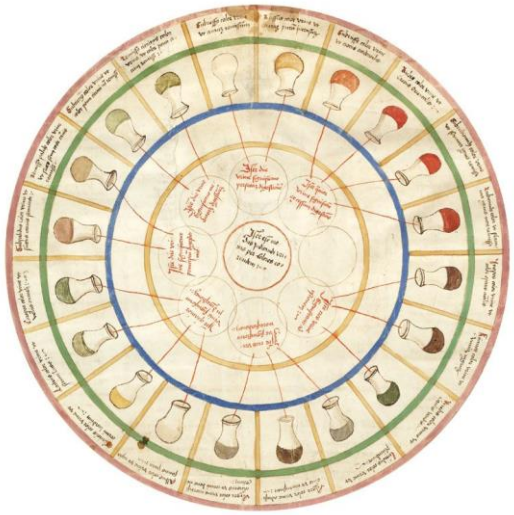

Figure 24: *The Urine wheel - Ullrich Pinder Epiphanie Medicorum 1506.*

Metabolomics, the newest addition to the “omics” domain and the closest to the observed phenotype, reflects changes occurring at all molecular levels, as well as influences resulting from other internal and external factors.

The metabolomic approaches are divided in *untargeted*, which includes metabolomic screening and fingerprinting and *targeted*.

The targeted approach is normally applied on information obtained from the previous approaches, where they are some metabolites recognized as biologically relevant (e.g. potential biomarkers).

It is applied to measure a set of metabolites defined or to answer a question specific biochemistry.

Metabolic fingerprinting studies the metabolome according to a holistic approach and from a limited point of biological knowledge first.

The experimental design is aimed at acquiring a large subset of metabolites (100–1,000s) followed by reading the data to define biological differences. For interpretation, univariate statistical analyzes or multivariate. This approach aims to discover new ways metabolic or biomarker (fishing expedition).

The targeted approach is normally hypothesis-driven while the fingerprinting studies are question-driven; the metabolic screening is a compromise in fact it aims to determine a large set of compounds to the same metabolic class.

The chosen approach will determine how to draw the experiments, prepare the samples and which analytical technique use. It is essential to plan the project and the experiments with statisticians, analytical chemists, biochemists and doctors to collect the correct number of samples and treat them in the most appropriate way to get as much information as possible.

The metabolome is considered as a final output of genetic, epigenetic and non-genetic factors[62]. A positive diagnosis of a disease is followed by treatment using drugs produced on a large scale and administered at a standardized and universally accepted dosage. These conventional drugs are developed to treat general symptoms or the disease as determined by the mean results obtained over large population groups. However, it is crucial to

understand that due to, for example, genetics and a variety of other factors such as individual diet, habits, gender etc., not all diseases affect all individuals in the same manner[63] and neither do all individuals respond to treatment in the same way[64].

By comparing the metabolite profiles of two or more disease phenotypes, metabolomics can be applied to identify biomarkers related to the perturbation being investigated. These biomarkers can, in turn, be used to develop personalized prognostic, diagnostic, and treatment approaches, and can also be applied to the monitoring of disease progression, treatment efficacy, predisposition to drug-related side effects, and potential relapse[3,65,66] .

From a clinical point of view, the metabolomic approach has three important applications potential:

- The first application, identifying the set of metabolites that discriminate between different groups of subjects, makes possible the metabolic description of a particular disease phenotype. The discriminating metabolites can therefore be identified as potential pathology biomarkers. Considering the close correlation between the metabolome and the phenotype, many expectations of metabolomics studies are aimed at defining early metabolic disturbances that occur even before their appearance symptoms of the disease, potentially offering a diagnostic tool early.
- The second major application is the determination of predictive profiles of pathology which, applied to the general population, could allow the development of what is called "molecular epidemiology".
- The third important clinical potential is related to the identification of individual metabolic characteristics they may be able to predict the efficacy and toxicity of a drug treatment according to what it is defined as “pharmaco-metabolomics” [67,68]. The purpose is to identify measurable markers before therapeutic intervention that are able to predict the effect and / or toxicity that the drug will have on the individual, allowing to optimize the therapeutic choices for each individual patient.

To summarize the ultimate goal of personalized medicine is to enable clinicians to prescribe the right medicine to the right patient at the right time with maximum efficacy and minimal toxicity, or to predict the susceptibility to disease onset among populations in advance.

5.2 State of the art

High resolution mass spectrometry is increasingly used instrument for metabolomics thanks to its high performances allowing the analysis of complex matrices and the detection of hundred metabolites[69].

Untargeted metabolomics usually involves comparing the metabolome of control and test groups to identify differences between their metabolite profiles. These metabolomic differences may be relevant to specific biological conditions.

In an untargeted metabolomics workflow, analytical reproducibility and stringent data analysis are the keys to a successful experiment. High analytical reproducibility means that the data are a direct expression of biological variance, both the chromatographic separations and mass spectrometers must have high reproducibility to minimize analytical variation.

Because untargeted metabolomics aims to be comprehensive, it is imperative that the analytical technologies perform the following functions:

- cover a breadth of metabolites with diverse physico-chemical properties,
- possess a large dynamic range to analyze the varying abundances of metabolites,
- possess high resolution accurate mass capability to separate isobaric species,
- perform HRAM MSn for compound identification and structural elucidation.

There are typically four steps in an untargeted metabolomics experiment:

1. Experimental design consists in separating the biological study into multiple analytical experiments and analyzing quality controls for accuracy and precision of results to ensure successful experimental design[70].
2. Sample preparation is a must when using mass spectrometry-based techniques. There are multiple techniques based on sample type, and there are entire methods for extracting specific classes of metabolites[70,71].
3. Analytical procedures. Small molecule compounds are chemically and structurally diverse, so the mass spectrometer must be combined with a separation technique to reduce sample complexity prior to MS analysis. Depending on the separation technique,

specific metabolite classes may be discriminated against and not appear at their normal abundance.

4. Data analysis. Huge amounts of complex raw data must be analyzed in a stringent manner to remove experimental artifacts. To accomplish this task, stringent statistical tools and software are required for accurate compound identification and quantitation.

As endogenous metabolites are diverse in their physico-chemical properties, as well as varying in abundance, a true comprehensive metabolomics study will require orthogonal sample preparation and separation techniques. In reality, most experiments are not comprehensive and there is a bias towards certain classes of compounds.

The mass spectrometer can also introduce bias into the study. Utilizing positive or negative ion mode can affect which compounds can be ionized and detected. It is always recommended to analyzed positive and negative ions to avoid this bias.

Untargeted metabolomics calls for utilization of exact mass measurements of precursor and fragment ions to enable the true identification of metabolites, versus features[72,73]. High-resolution mass spectrometry and high scan speeds are critical to increasing the depth of coverage. The ability to search data with a narrow, 1–3 ppm mass tolerance and include spectral accuracy based on the mass and intensity of detected natural isotopes can be critical to properly identifying each species[74]. In addition, both fragmentation techniques, HCD and CID, have been shown to provide complementary structural information that may be used to better annotate metabolites. It is most common to perform metabolomics using DDA, although recent work has published results using data-independent acquisition. While the dynamic range for quantitation of the DDA and DIA data was similar for the metabolites targeted, DIA may offer the possibility to identify more species with lower S/N, at the cost of more difficult deconvolution.

The increased presence of Orbitrap-based instruments in the metabolomics field, combined with the rise of advanced software, has led to universal databases[75,76,77,78] akin to those developed by NIST for chemical ionization. mzCloud is a freely available MS/MS database that enables wide-spread searching of HRAM data, with >8000 entries

available[79,80]. ChemSpider offers an alternate search engine of freely sourced data with over 26 million entries ranging from high to low mass error.

However, one of the major bottlenecks of metabolomics is the identification process of compounds, that is necessary to draw any biological conclusion from untargeted metabolomics data[81] . Data processing workflows incorporate several defined steps such as noise filtering, peak detection, peak deconvolution, retention time alignment and finally feature annotation. Feature annotation is performed by comparing an experimental mass measurement to a database of known metabolites in order to generate potential candidates. HRMS measurement alone is not sufficient to provide metabolite information, indeed, additional information such as fragmentation pattern data and their matching with reference fragmentation spectra are essential. Moreover, given the wide variety of chromatographic and mass spectrometric conditions that can be employed to generate metabolomic experiments, it is essential that established metabolomics laboratories implement fixed robust and reliable protocols that yield stable retention times[82].

Standardization in metabolomics is still an unmet need but many efforts have been made towards realizing this aim. In 2007, the Metabolomics Standards Initiative (MSI) defined the rules to achieve a consistent communication regarding metabolite identification confidence, a level system has been proposed and discussed[83], starting with unique features (level 5) followed by molecular formula prediction (level 4). Tentative structure identification (level 3) on the MS1 level can be performed by *m*/*z* database searching (PubChem and ChemSpider). For putative identification (level 2) on the MS2 level several databases of experimental fragmentation data are available online such as METLIN, the Human Metabolome Database (HMDB) and mzCloud. However, these databases have several drawbacks. Full usability often requires commercial software and, in addition, available MS2 spectra have been generated with specific instruments or settings, precluding the transfer to other equipment. Importantly, consistent mass accuracy, resolution, collision energy (CE) and fragmentation are needed to ensure comparability. For ideal matching and as part of a validated identification (level 1), the MS2 spectra of reference standards and samples should be generated under identical experimental conditions. Level 1 being the highest confident

identification when a structure is confirmed with a minimum of two independent and orthogonal data from a pure reference standard under identical analytical conditions[84].

Moreover, not all known metabolites are covered by available spectral libraries. Therefore, an in-house library is of great value for enhanced metabolite identification, even though the generation can be time consuming.

5.3 The aim of the project

Metabolomics is a relatively new approach compared to other "-omics". With the rapid development of analytical technology and increasing passion of precision medicine, metabolomics has been extensively trialed in both the clinical and research studies. The individual metabotype at the baseline is informative for predicting responses to drug therapy and personal susceptibility to diseases during the follow-up. Meanwhile, the different metabolic profiles between baseline and post-treatment could reveal novel mechanisms of drugs by integrating the altered metabolites with involved metabolic pathways.

For understanding processes and mechanisms in biological systems, High Resolution Mass Spectrometry gains more and more importance, but it is necessary to improve the construction of a simple, robust and standardized method from sample preparation to an accurate metabolites identification.

The main challenge in untargeted approaches is a reliable metabolite identification.

The purpose of this work is to build, optimize and share a spectral MSMS library with Accurate Mass Retention Time (AMRT) databases, using metabolic standards (MSMLS, IROA Technologies, Bolton, MA, USA), selected for their involvement in key biological processes, in order to simplify a robust metabolite identification and the development of quantitative methods.

The PhD thesis reports the analytical steps needed to build the database, which are easily usable also on new and different molecules, so that research groups can customize their libraries. The effectiveness of the process has been tested on real pediatric plasma and urine samples.

5.4 Chemicals

Ammonium formate, acetonitrile, methanol and formic acid (LC-MS grade) were purchased from Sigma Aldrich Srl (Milan, Italy). Water was purified by reverse osmosis and filtrated through a Milli-Q purification system (Millipore, Milford, MA, USA).

The library of metabolite standards (Mass of Standards Mass Spectrometry Metabolite, MSMLS) was purchased from IROA Technologies (Bolton, MA, USA).

4 reverse phase and 4 HILIC columns were tested, Table 7.

Table 7: Chromatographic columns.

Name	Stationary phase	Diameter x lenght (mm)	Porosity (µm)	Company
ACQUITY BEH	C18	2.1 X 100	1.7	Waters S.p.A.
Hypersil GOLD aQ	C18	2.1 X 100	2.6	ThermoFisher Scientific
Hypersil GOLD CN	C18	2.1 X 100	1.9	ThermoFisher Scientific
Accucore Polar Premium	C18	2.1 X 150	2.6	ThermoFisher Scientific
ACQUITY BEH	Amide	2.1 X 150	1.7	Waters S.p.A.
Accucore Amide	Amide	2.1 X 150	2.6	ThermoFisher Scientific
SeQuant	ZIC-pHILIC	2.1 X 150	5	Merck S.p.a.
Shodex Asahipak	NH2P-50 2D	2.0 X 150	5	Phenomenex

5.5 Workflow

This paragraph shows the steps to build the spectral library and the AMRT database. Subsequently each step will be explained in detail.

- MSMS spectral library construction using mzVault software,
- choice of chromatographic conditions (columns and mobile phases) by analyzing IROA spiked plasma samples,
- AMRT database construction,
- database validation on real neuroblastoma plasma and epileptic urine samples.

For each molecule we obtain the accurate molecular mass, 24 fragmentation spectra at different energies (12 positive and 12 negative), the positive and negative best fragmentation energy and the retention time on reverse phase and HILIC columns.

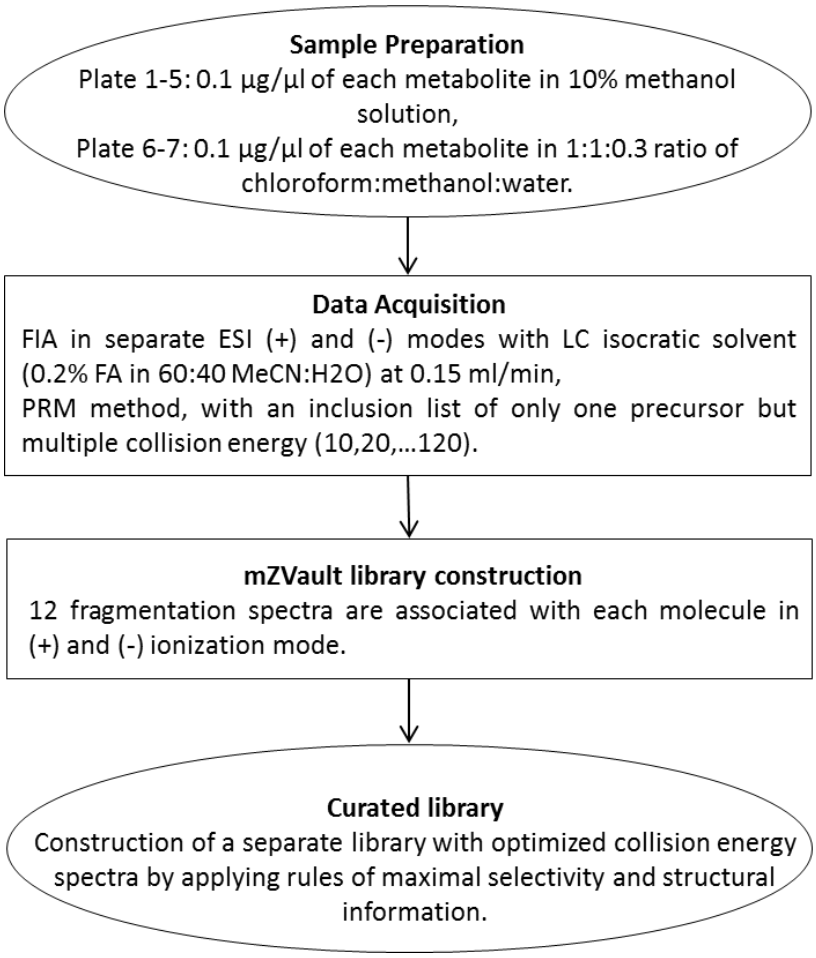

Figure 25: *Steps to build the MSMS spectral library and the AMRT database.*

5.5.1 MSMS spectral library construction

MSMLS contains over 634 unique small molecule compounds, provided at 5 µg per well, arranged in 7 polypropylene plates with alphanumeric assigned positions. Forty-two compounds were excluded from analyses: 5 of them were out of our defined mass range (70-1000 m/z) and 37 were duplicates. Compounds were dissolved using two different solutions (10% methanol for plates 1-5 and chloroform:methanol:water 1:1:0.3 for plates 6 and 7) in order to obtain a 0.1 µg/µL concentration.

LC-HRMS analysis was carried out using a Vanquish Horizon UHPLC system coupled to a Q-Exactive Plus, Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermofisher Scientific,

Milan, Italy). Ionization was obtained using an Electrospray (ESI) source both in positive and negative mode.

In order to build the MSMLS spectral library, the compounds, contained in the corresponding wells, were analyzed by flow injection analysis (FIA) using the ESI source coupled with isocratic mobile phase 60/40 ACN/H20, 0.2% FA, at a flow rate of 0.15 ml/min, for a run time of 0.5 min. These analytical conditions have been optimized on a tryptophan standard, prepared in the same way used for the library.

The source parameters were chosen to gain the maximum signal with acceptable stability (CV<5%) Table 8

Table 8: Easy Spray source parameters.

Parameter	Value
sheath gas (u.a.)	30
auxiliary gas (u.a.)	10 u.a.
temperatura di auxiliary gas temperature (°C)	300
capillary temperature (°C)	300
spray voltage + (V)	4000
Spray voltage – (V)	3000

The molecular ion was selected and fragmented using 12 different collision energies from 10 to 120 eV, with steps of 10 eV. Few compounds do not generate the molecular ion and therefore it has been considered, if present, its adduct.

Data were acquired in MS1 Full Scan mode with a resolution of 70000, AGC 1e6, mass range 70-1000 m/z and a maximum injection time of 200 ms followed by a PRM event with an inclusion list of only one precursor but multiple collision energies (NCE from 10 to 120 eV)

with a resolution 17500, AGC 5e5, maximum injection time of 60 ms and an isolation windows of 1.2 m/z.

The fragmentation library was built using mzVault 2.1 (ThermoFisher Scientific, Milan, Italy).

The best collision energy for each compound in positive and negative mode was also chosen applying the rule of maximal selectivity and structural information.

5.5.2 Chromatographic conditions

The challenge of metabolomics as already mentioned is to identify and quantify hundreds of small molecules with very different physical characteristics. The choice of chromatographic conditions such as chromatographic columns, mobile phases and duration of chromatographic runs is fundamental to the success of the analysis.

50 µl of plasma from healthy donors are extracted with 150 µl of cold methanol (-20 °C), centrifuged at 14,000 rpm for 10 min and the supernatant is spiked with the 634 molecules of our interest.

Several chromatographic conditions were tested using the pooled plasma samples the spiked with the metabolite library. In particular, 8 different chromatographic columns and two different mobile phases (as specified in Table 9) were tested. The linear gradient for reversed phase columns started with 1% B and in 15 minutes increased up to 100% with a flow rate of 250 µl/min, then the columns were normalized for 5 minutes with 1% phase B. The linear gradient for HILIC columns started with 90% B and decrease to 30% B in 15 minutes with a flow rate of 200 µl/min, the columns was then normalized with 90 % phase B for 9 min. Separately for each polarity the experiments were done in data-dependent acquisition mode alternating MS and MS/MS experiments. A maximum of 5 MS/MS experiments were triggered per MS scan. The intensity threshold was set at 1.6e5 using an isolation window of 1.4 Da. The m/z values of signals already selected for MS/MS were put on an exclusion list for 20 s. 70000 and 17500 resolution, 1e6 and 1e5 AgC, 100 ms and 50 maximum injection time were used for MS1 and MS2 scan respectively, Table 10.

To choose the best conditions, chromatographic suitability criteria were applied: symmetry parameters, width at peak height 50%, symmetry threshold above 90%, min peak width 1.8

s, max peak width 3.6 s. The conditions that yielded the best results were then used to annotate the RT by analyzing 9 mixtures composed by non-isobaric molecules in solvent and in plasma.

Table 9: Chromatographic conditions.

	SOLVENT A1	**SOLVENT A2**	**SOLVENT B1**	**SOLVENT B2**
C18 BEH	0.1% FA in H2O	0.1% FA in H2O	0.1% FA in MeOH	0.1% FA in MeCN
C18 aQ	0.1% FA in H2O	0.1% FA in H2O	0.1% FA in MeOH	0.1% FA in MeCN
C18 Accucore Polar	0.1% FA in H2O	0.1% FA in H2O	0.1% FA in MeOH	0.1% FA in MeCN
Hypersil GOLD CN	0.1% FA in H2O	0.1% FA in H2O	0.1% FA in MeOH	0.1% FA in MeCN
BEH Amide 1.7µm	5 mM ammonium formate pH 2	5 mM ammonium formate pH 8	MeCN	MeCN
ZIC - pHILIC 5µm	5 mM ammonium formate pH 2	5 mM ammonium formate pH 8	MeCN	MeCN
Shodex Asahipak	5 mM ammonium formate pH 2	5 mM ammonium formate pH 8	MeCN	MeCN
Accucore Amide	5 mM ammonium formate pH 2	5 mM ammonium formate pH 8	MeCN	MeCN

Table 10: Mass spectrometer set up.

Master Scan MS-OT	
Detector Type	Orbitrap
Resolution	70000

Scan Range (m/z)	70-1000
AGC Target	1.0e6
Maximum Injection Time (ms)	100

Dynamic Exclusion	
Exclude after n times	1
Exclusion Duration (s)	20
Intensity Threshold	1.6e5
DDA Mode	Cycle Time

$ddMS^2$ MS OT	
Isolation Mode	1.4
Activation Type	HCD
HCD Energy (%)	Stepped 30, 40, 50
Detector	OT
Resolution	15000
AGC Target	1.0e5
Maximum Injection Time (ms)	50

The chosen columns are summarized in Table 11. HILIC overcome the challenge of retaining and separating extremely polar compounds. HILIC are designed to retain metabolites that are too polar to be retained by RP chromatography. Amide columns facilitate the retention of polar analytes spanning a wide range in polarity, structural moiety and pK_a.

Table 11: Chosen columns.

	Particle Type	Ligand Type	pH range	Pressure tollerance
BEH Amide 1.7µm 2.1 X 150 mm	Ethylene Bridged Hybrid (BEH)	Trifunctional Amide	2-11	1000 bar
C18 BEH 1.7µm 2.1 X 100 mm	Ethylene Bridged Hybrid (BEH)	C18	1-12	1240 bar

5.5.3 AMRT database

The 634 molecules belonging to the MSMLS library have been divided into 9 mixtures so that no isobar molecules are present in the same mixture. In the supplementary table 7 the column D indicates the mixture to which each molecule belongs. Each mixture has been analyzed under the previously selected chromatographic conditions. The mixtures have been acquired in data dependent using an inclusion list that allows to give a fragmentation priority to the molecules belonging to the mixture under examination. The retention times of each molecule on the RP and HILIC columns have been recorded and are shown supplementary table 7, column G and H respectively.

5.5.4 Untargeted metabolomic analysis of pediatric plasma and urine samples

To validate the MSMS spectral library and the AMRT database, experiments on the most two used biological fluids plasma and urine were made.

Urine metabolomics has emerged as a prominent field for non-invasive biomarkers discovery that can detect subtle metabolic changes brought by a specific disease or therapeutic intervention.

Urine is a sterile, transparent amber-colored fluid generated by the kidneys. The kidneys extract excess water, sugars and a variety of other soluble wastes from the bloodstream. Urine generally is composed of metabolic breakdown products from a large number of foods,

drugs, drinks, endogenous waste metabolites, environmental contaminants and bacterial by-products.

Blood is composed of a cellular component including red and white blood cells and platelets, and a liquid carrier, named plasma. Plasma accounts for about 50–55% of blood volume. Comparing to plasma, serum is less viscous don't have fibrinogen, prothrombin and other clotting proteins. Serum contains a large number of substances including nutrients (such as lipids, carbohydrates and amino acids), proteins and peptides (such as globulins, albumins, lipoproteins, hormones and enzymes), electrolytes, organic wastes and many other small organic molecules dissolved in them. As a primary carrier of small molecules in the body, serum plays a crucial role in the transportation of dissolved gases, hormones, nutrients and metabolic wastes, it also assists in the regulation of the ion composition and pH of interstitial fluids, the stabilization of body temperature, the defense against toxins and pathogens and the restriction of fluid losses at injury sites. Since blood enter every tissue and every organ in the body, it acts as an important liquid highway for all the molecules secreted, excreted or discarded by different tissues in response to different physiological or pathological stresses.

In urine and plasma samples we have performed all stages of a real experiment but finally we will focalize our attention on the leve1 metabolites identification. To test the database 43 urine samples belong to epileptic pediatric patients treated with two different drugs plasma and 37 plasma samples belong to neuroblastoma patients at different stage disease have been sampled.

50 µl of plasma/urine were extracted with 150 µl of cold methanol (-20 °C), centrifuged (4 °C) and, after protein precipitation, the supernatant was recovered.

The analytical study design is summarized in Figure 25.

Different strategies—at best in combination—are generally accepted: addition of internal standards (IS) into each analyzed sample; continuous measurement of system suitability tests (SST) over the whole analytical batch containing a defined number of metabolites ideally evenly distributed over the chromatographic run; and inclusion of QC pool samples prepared from all experimental samples, hence representing the average of the data set. The

importance of experimental design and QC samples in large-scale and MS-driven untargeted metabolomic studies of humans. Further, randomization of the samples at least for analysis is an important step in untargeted metabolomics to prevent technical/instrumental biases.

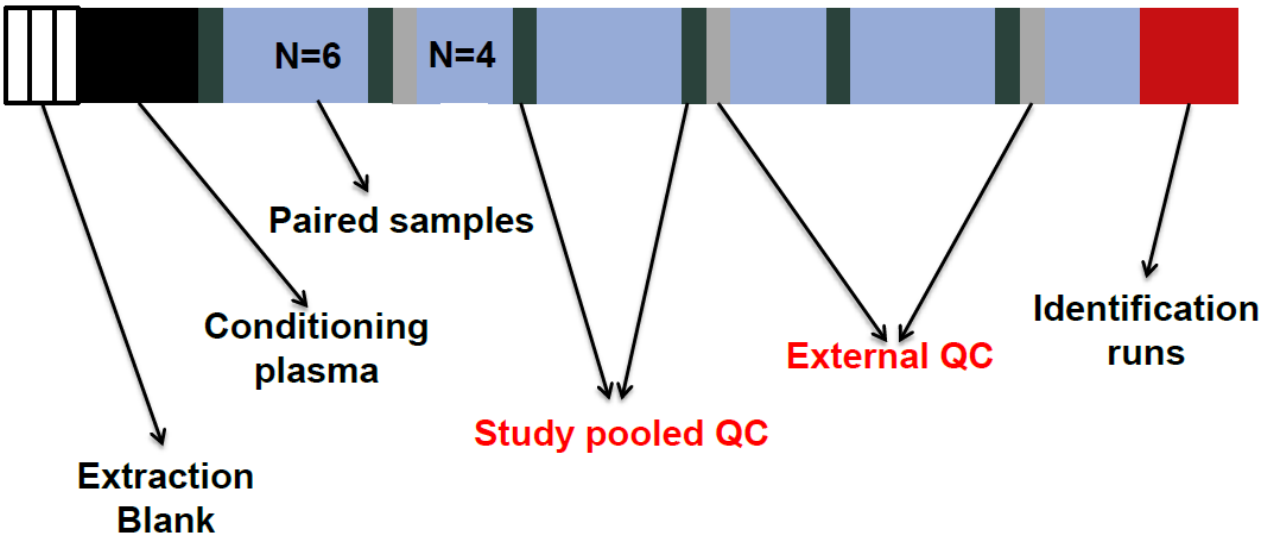

Figure 26: *Analytical study design.*

- Procedure blanks.

In the first 3-5 runs the procedure blank was analyzed in MS1 acquisition mode because the most 20 intense ions from blank samples will be put in an exclusion mass list of identification method. So the most intense ions of blank samples will not be fragmented in the identification runs.

- Plasma samples.

8-10 plasma samples runs were performed to condition the columns.

- Study pooled QC

About 10 µl of each samples (depending on the experiment length) was mixed together in a pool sample which is analyzed every 4 samples. The usefulness of the QC study pooled is to check that the instrumental performance remains constant during the duration of all

analyses. Additionally, these samples will be used to filter out features that will have a CV>25%.

- External pooled QC.

External pooled QC is a mixtures of plasma or urine or serum of healthy donors which we have made in large quantities. This external pool is used in all our experiment and acquired every 8-10 samples.

The usefulness of this pool is to be able to compare different analysis batch by normalizing them to the pool.

- Samples.

The samples were acquired in full scan mode in Orbitrap using 70000 resolution, 1e6 AcG and 100 as maximum injection time.

- Identification runs on study pooled QC.

To identify the metabolites, the plasma/urine were analyzed in fullScan-ddMS2 mode with fragmentation priority. 70000 and 17500 resolution, 1e6 and 2e5 AgC, 100 and 65 ms were used for MS1 and MS2 scan respectively. The loop count was set at 5, the m/z values of signals already selected for MS/MS were put on an exclusion list for 5 s. A maximum of 5 MS/MS experiments were triggered per MS scan. The intensity threshold for triggering MS/MS was set at 1.6e5 using an isolation window of 1.4. If no further inclusion list entries are identified in a scan event will be selected other masses. A normalized stepped collision energy of 20, 40, 80 was used.

With Skyline software[22], inclusion lists have been created with a retention time range of +/- 0.5 minutes around the retention time in scheduled mode; to each metabolite was assigned the best fragmentation energy obtained by manual review of the spectral library.

Skyline is a freely-available and open source Windows client application for building Selected Reaction Monitoring (SRM) / Multiple Reaction Monitoring (MRM), Parallel Reaction Monitoring (PRM), Data Independent Acquisition (DIA/SWATH) and DDA with MS1 quantitative methods and analyzing the resulting mass spectrometer data. It aims to employ

cutting-edge technologies for creating and iteratively refining targeted methods for large-scale quantitative mass spectrometry studies in life sciences[22].

Compound Discoverer software ver. 3.1 (Thermo Fisher Scientific) was used for data processing including retention time alignment, unknown compound detection, compound grouping across all samples, prediction of elemental compositions for all compounds, filling gaps across all samples, and background subtraction (using blank samples). Compound Discoverer was used, Mass List, Spectral Library (ddMS2), mzCloud (ddMS2) and ChemSpider (KEGG[85], BioCyc[86], HMD[75] as database) were used to identify level 2 and 3 compounds. Level 1 metabolites were identify using the AMRT database constructed.

5.6 Results

Using mzVault, the spectral MSMS library of 530 molecules was build, the molecular ion or its adduct was ionized in positive and negative mode and was fragmented with 12 different energies. The library containing the list of compounds, the molecular ions and the fragmentation spectra will be share in a public format. The best energy for each compound was also recorded (G and H columns in the supplementary table 7). Some molecules belonging to the database, have significantly different best fragmentation energies from those commonly used in discovery experiments (I and J columns in the supplementary table 7).

16 chromatographic condition were tested, the two columns which allowed us to obtain the best results were: ACQUITY C18 BEH as mobile phases A 0.1%FA in H2O and B 0.1% FA in ACN and ACQUITY BEH Amide using in A 5mM ammonium formate in H2O pH3 and in B ACN.

Using the established MSMS and chromatographic conditions (as describe in paragraph 5.5.3) a total of 359 and 191 compounds were detected by reversed phase and HILIC columns respectively analyzed in both positive and negative polarities.

Moreover, using our defined protocol, plasma and urine samples derived from patients were analyzed.

A total of 132 and 124 level 1 compounds belonging to AMRT database were identified in plasma and urine samples and listed in the supplementary tables 7. Figure 26 and Figure 27 (Venn diagram) shows the number of metabolites identified in plasma and urine respectively and highlights the importance of combining the two different columns and MS polarities in order to increase the number of identified compounds[87].

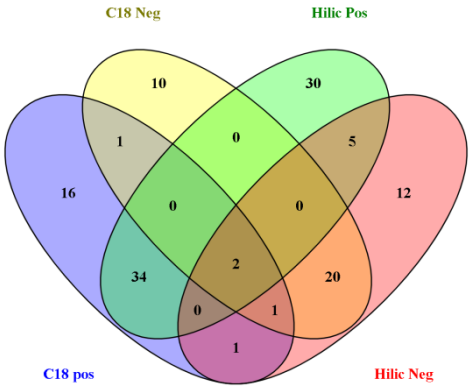

Figure 27: *Plasma level 1 metabolites.*

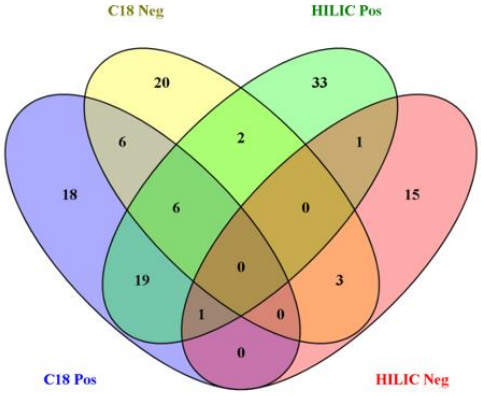

Figure 28: *Urine level 1 metabolites.*

5.7 Discussion

Untargeted metabolomic analysis is a powerful technique in that it allows the collection of data without pre-existing knowledge and it is thus very useful for several applications. One of the bottlenecks of the untargeted strategy is represented by chemical structure identification. Chromatographic conditions and analytical methods such as sample preparation steps, chromatographic conditions and instrument platforms have an impact on the results directly influencing the subset of metabolites selected. However, identification of metabolites is crucial for drawing the correct biological conclusions so one of the major efforts for laboratories that want to approach untargeted metabolomic is the standardization of operative protocols for analysis and workflows used to obtain a transparent and confident metabolite assignment. In this PhD project we have developed a standardized protocol to build a MSMS spectral library and retention time database.

The in-house spectral MSMS library and the Accurate Mass Retention Time database can be used in small molecule identification software such as Compound Discoverer, MSDials[88], XCMS and improve the identification phase allowing the level 1 identification of about 400 metabolites.

Moreover, for each molecule the choice of the best fragmentation energy for both positive and negative polarities allows the optimization of the fragmentation energy in clinical matabolomics experiments. Usually in discovery experiments a stepped energy of 30, 40, 50 eV is used. However, our data suggests that for many molecules these energies are not optimal and could affect the identification of metabolites, especially in complex matrices such as plasma and urine.

The validation analysis was performed on the most two common biological fluids, the level 1 metabolites identified with CV<20% are about 1 hundred; the number is not so high, but it allows a solid core of identified metabolites.

The number of level 2 metabolites in plasma and urine samples is about 400 but these are not subject of my PhD project.

This protocol might be also useful for those who want to approach metabolomic untargeted analysis in fact it constitutes a reliable basis for the identification of level 1 of about 400 metabolites.

On the other hand spectral MSMS library and the retention time database can also be used by software developers to training and testing retention time prediction software[89],[90].

The spectral library is proving to be very useful in ongoing projects and will certainly need to be increased with other metabolites of biological interest.

The first metabolites of which we will acquire fragmentation spectra and retention times will be drugs, of which we have already bought a library of 1280 drugs (LOPAC library of pharmacologically active compounds, Merck KGaA). LOPAC is biologically annotated collection of inhibitors, receptor ligands, pharma-developed tools, and covers all major drug target classes.

The MSMS spectral library and the AMRT database represent the first brick to an accurate level 1 metabolites identification.

6. Conclusions

Improvements in technology have fueled the proliferation of omics applications. Omics is a wide domain involving specialized and high-throughput biotechnological methods, instruments and algorithms. These techniques are often used to measure and study complex biological systems and their interactions. The omics platforms (genomics, proteomics and metabolomics) offers a systematic assessment of the interactions between genetic variations, central metabolism and environmental exposures to explain the initiation and progression of a disease. It is understood that the causation in systems biology should determine how and why a process occurs in a certain way rather than identify separate molecular components.

During the PhD project we have studied the proteome of autoinflammatory disease, most of these conditions are caused by mutations of genes regulating the activation of innate immunity leading to an unprovoked activation of inflammatory response. It was possible to identify and quantify thousands of proteins thanks to the development of a miniaturized sample preparation and the development of a mass spectrometry method suitable for the small quantities of starting material we had available.

Some proteins already known in literature have been confirmed as significantly modulated and some proteins have been identified could be key points in the mechanism of autoinflammatory diseases.

Mass spectrometry has once again proved to be a useful tool in the complex path of understanding the molecular mechanisms underlying diseases.

Metabolomics is the omic science closest to explaining the phenotype of a disease and therefore most used in precision medicine. Metabolomics allows the identification of prognostic and diagnostic biomarkers of the diseases. The biggest challenge is identifying features with high accuracy degree. During the PhD project, about 600 molecules that play a fundamental role in the metabolic processes were chosen to be analyzed with level 1 identification according to the MSI guidelines. The database thus created can be used in metabolomics experiments on real samples. The database was tested on plasma and urine

samples but it could be useful also on other matrix such as liquor, amniotic fluids, saliva and tears.

Bibliography

1. Cappelletti P. Medicina di Precisione e Medicina di Laboratorio. *Rivista Italiana della Medicina di Laboratorio*. 2016. doi:10.1007/s13631-016-0131-9

2. Fiehn O. Metabolomics - The link between genotypes and phenotypes. *Plant Molecular Biology*. 2002. doi:10.1023/A:1013713905833

3. Davis KD, Aghaeepour N, Ahn AH, et al. Discovery and validation of biomarkers to aid the development of safe and effective pain therapeutics: challenges and opportunities. *Nature Reviews Neurology*. 2020. doi:10.1038/s41582-020-0362-2

4. Guo L, Milburn M V., Ryals JA, et al. Plasma metabolomic profiles enhance precision medicine for volunteers of normal health. *Proceedings of the National Academy of Sciences*. 2015. doi:10.1073/pnas.1508425112

5. Junot C, Fenaille F, Colsch B, Bécher F. High resolution mass spectrometry based techniques at the crossroads of metabolic pathways. *Mass Spectrometry Reviews*. 2014. doi:10.1002/mas.21401

6. Perry RH, Cooks RG, Noll RJ. Orbitrap mass spectrometry: Instrumentation, ion motion and applications. *Mass Spectrometry Reviews*. 2008. doi:10.1002/mas.20186

7. Skoog DA, Holler FJ, Crouch SR. *principles of instrumental analysis sixth edition*.; 2007. doi:10.1017/CBO9781107415324.004

8. Erve JCL, Gu M, Wang Y, DeMaio W, Talaat RE. Spectral Accuracy of Molecular Ions in an LTQ/Orbitrap Mass Spectrometer and Implications for Elemental Composition Determination. *Journal of the American Society for Mass Spectrometry*. 2009. doi:10.1016/j.jasms.2009.07.014

9. Su X, Lu W, Rabinowitz JD. Metabolite Spectral Accuracy on Orbitraps. *Analytical Chemistry*. 2017. doi:10.1021/acs.analchem.7b00396

10. Makarov A, Scigelova M. Coupling liquid chromatography to Orbitrap mass spectrometry. *Journal of Chromatography A*. 2010. doi:10.1016/j.chroma.2010.02.022

11. Hecht ES, Scigelova M, Eliuk S, Makarov A. Fundamentals and Advances of Orbitrap Mass Spectrometry. In: *Encyclopedia of Analytical Chemistry*. ; 2019. doi:10.1002/9780470027318.a9309.pub2

12. Makarov A. Electrostatic axially harmonic orbital trapping: A high-performance technique of mass analysis. *Analytical Chemistry*. 2000. doi:10.1021/ac991131p

13. Wolf-Yadlin A, Hu A, Noble WS. Technical advances in proteomics: New developments in data-independent acquisition. *F1000Research*. 2016. doi:10.12688/f1000research.7042.1

14. Bateman NW, Goulding SP, Shulman NJ, et al. Maximizing peptide identification

events in proteomic workflows using data-dependent acquisition (DDA). *Molecular and Cellular Proteomics*. 2014. doi:10.1074/mcp.M112.026500

15. Mann M, Hendrickson RC, Pandey A. Analysis of proteins and proteomes by mass spectrometry. *Annual Review of Biochemistry*. 2001. doi:10.1146/annurev.biochem.70.1.437

16. Venable JD, Dong MQ, Wohlschlegel J, Dillin A, Yates JR. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nature Methods*. 2004. doi:10.1038/nmeth705

17. Johnson D, Boyes B, Fields T, Kopkin R, Orlando R. Optimization of data-dependent acquisition parameters for coupling high- speed separations with LC-MS/MS for protein identifications. *Journal of Biomolecular Techniques*. 2013. doi:10.7171/jbt.13-2402-003

18. Cristea IM, Graham D. Virology meets Proteomics. *Proteomics*. 2015. doi:10.1002/pmic.201570103

19. Picotti P, Aebersold R. Selected reaction monitoring-based proteomics: Workflows, potential, pitfalls and future directions. *Nature Methods*. 2012. doi:10.1038/nmeth.2015

20. Lange V, Picotti P, Domon B, Aebersold R. Selected reaction monitoring for quantitative proteomics: A tutorial. *Molecular Systems Biology*. 2008. doi:10.1038/msb.2008.61

21. Peterson AC, Russell JD, Bailey DJ, Westphall MS, Coon JJ. Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Molecular and Cellular Proteomics*. 2012. doi:10.1074/mcp.O112.020131

22. Adams KJ, Pratt B, Bose N, et al. Skyline for Small Molecules: A Unifying Software Package for Quantitative Metabolomics. *Journal of Proteome Research*. 2020. doi:10.1021/acs.jproteome.9b00640

23. Godovac-Zimmermann J. 8th Siena meeting. From Genome to Proteome: Integration and Proteome Completion. In: *Expert Review of Proteomics*. ; 2008. doi:10.1586/14789450.5.6.769

24. Fournier ML, Gilmore JM, Martin-Brown SA, Washburn MP. Multidimensional separations-based shotgun proteomics. *Chemical Reviews*. 2007. doi:10.1021/cr068279a

25. Compton PD, Kelleher NL. Spinning up mass spectrometry for whole protein complexes. *Nature Methods*. 2012. doi:10.1038/nmeth.2216

26. Catherman AD, Skinner OS, Kelleher NL. Top Down proteomics: Facts and perspectives. *Biochemical and Biophysical Research Communications*. 2014.

doi:10.1016/j.bbrc.2014.02.041

27. Gattorno M, Federici S, Pelagatti MA, et al. Diagnosis and management of autoinflammatory diseases in childhood. In: *Journal of Clinical Immunology*. ; 2008. doi:10.1007/s10875-008-9178-3

28. Gül A. Dynamics of inflammatory response in autoinflammatory disorders: Autonomous and hyperinflammatory states. *Frontiers in Immunology*. 2018. doi:10.3389/fimmu.2018.02422

29. Peckham D, Scambler T, Savic S, McDermott MF. The burgeoning field of innate immune-mediated disease and autoinflammation. *Journal of Pathology*. 2017. doi:10.1002/path.4812

30. Aksentijevich I, Kastner DL. Genetics of monogenic autoinflammatory diseases: Past successes, future challenges. *Nature Reviews Rheumatology*. 2011. doi:10.1038/nrrheum.2011.94

31. Park YH, Wood G, Kastner DL, Chae JJ. Pyrin inflammasome activation and RhoA signaling in the autoinflammatory diseases FMF and HIDS. *Nature Immunology*. 2016. doi:10.1038/ni.3457

32. Sag E, Bilginer Y, Ozen S. Autoinflammatory Diseases with Periodic Fevers. *Current Rheumatology Reports*. 2017. doi:10.1007/s11926-017-0670-8

33. Mandey SHL, Schneiders M, Koster J, Waterham HR. Mutational spectrum and genotype-phenotype correlations in mevalonate kinase deficiency. *Human Mutation*. 2006. doi:10.1002/humu.20361

34. Akula MK, Shi M, Jiang Z, et al. Control of the innate immune response by the mevalonate pathway. *Nature Immunology*. 2016. doi:10.1038/ni.3487

35. Rieckmann JC, Geiger R, Hornburg D, et al. Social network architecture of human immune cells unveiled by quantitative proteomics. *Nature Immunology*. 2017. doi:10.1038/ni.3693

36. Aebersold R, Mann M. Mass-spectrometric exploration of proteome structure and function. *Nature*. 2016. doi:10.1038/nature19949

37. Meissner F, Mann M. Quantitative shotgun proteomics: Considerations for a high-quality workflow in immunology. *Nature Immunology*. 2014. doi:10.1038/ni.2781

38. Tassi S, Carta S, Vené R, Delfino L, Ciriolo MR, Rubartelli A. Pathogen-Induced Interleukin-1β Processing and Secretion Is Regulated by a Biphasic Redox Response. *The Journal of Immunology*. 2009. doi:10.4049/jimmunol.0900578

39. Tassi S, Carta S, Delfino L, et al. Altered redox state of monocytes from cryopyrin-associated periodic syndromes causes accelerated IL-1β secretion. *Proceedings of the National Academy of Sciences of the United States of America*. 2010.

doi:10.1073/pnas.1000779107

40. Kulak NA, Pichler G, Paron I, Nagaraj N, Mann M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nature Methods*. 2014. doi:10.1038/nmeth.2834

41. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen J V., Mann M. Andromeda: A peptide search engine integrated into the MaxQuant environment. *Journal of Proteome Research*. 2011. doi:10.1021/pr101065j

42. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*. 2008. doi:10.1038/nbt.1511

43. Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Molecular and Cellular Proteomics*. 2014. doi:10.1074/mcp.M113.031591

44. Tyanova S, Temu T, Sinitcyn P, et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods*. 2016. doi:10.1038/nmeth.3901

45. Rudolph JD, Cox J. A Network Module for the Perseus Software for Computational Proteomics Facilitates Proteome Interaction Graph Analysis. *Journal of Proteome Research*. 2019. doi:10.1021/acs.jproteome.8b00927

46. Otasek D, Morris JH, Bouças J, Pico AR, Demchak B. Cytoscape Automation: Empowering workflow-based network analysis. *Genome Biology*. 2019. doi:10.1186/s13059-019-1758-4

47. Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research*. 2015. doi:10.1093/nar/gku1003

48. Jamilloux Y, Lefeuvre L, Magnotti F, et al. Familial Mediterranean fever mutations are hypermorphic mutations that specifically decrease the activation threshold of the Pyrin inflammasome. *Rheumatology (United Kingdom)*. 2018. doi:10.1093/rheumatology/kex373

49. Chae JJ, Cho YH, Lee GS, et al. Gain-of-Function Pyrin Mutations Induce NLRP3 Protein-Independent Interleukin-1β Activation and Severe Autoinflammation in Mice. *Immunity*. 2011. doi:10.1016/j.immuni.2011.02.020

50. Zhang X, Bogunovic D, Payelle-Brogard B, et al. Human intracellular ISG15 prevents interferon-α/β over-amplification and auto-inflammation. *Nature*. 2015. doi:10.1038/nature13801

51. John SP, Sun J, Carlson RJ, et al. IFIT1 Exerts Opposing Regulatory Effects on the Inflammatory and Interferon Gene Programs in LPS-Activated Human Macrophages. *Cell Reports*. 2018. doi:10.1016/j.celrep.2018.09.002

52. Karadeniz H, Güler AA, Atas N, et al. Tofacitinib for the treatment for colchicine-resistant familial Mediterranean fever: case-based review. *Rheumatology International*. 2020. doi:10.1007/s00296-019-04490-7

53. Savran Y, Sari I, Leyla Kozaci D, Gunay N, Onen F, Akar S. Increased levels of macrophage migration inhibitory factor in patients with familial mediterranean fever. *International Journal of Medical Sciences*. 2013. doi:10.7150/ijms.6116

54. McDermott MF, Aksentijevich I, Galon J, et al. Germline mutations in the extracellular domains of the 55 kDa TNF receptor, TNFR1, define a family of dominantly inherited autoinflammatory syndromes. *Cell*. 1999. doi:10.1016/S0092-8674(00)80721-7

55. Papa R, Lachmann HJ. Secondary, AA, Amyloidosis. *Rheumatic Disease Clinics of North America*. 2018. doi:10.1016/j.rdc.2018.06.004

56. Aksentijevich I, Zhou Q. NF-κB pathway in autoinflammatory diseases: Dysregulation of protein modifications by ubiquitin defines a new category of autoinflammatory diseases. *Frontiers in Immunology*. 2017. doi:10.3389/fimmu.2017.00399

57. di Masi A, De Simone G, Ciaccio C, D'Orso S, Coletta M, Ascenzi P. Haptoglobin: From hemoglobin scavenging to human health. *Molecular Aspects of Medicine*. 2020. doi:10.1016/j.mam.2020.100851

58. Shi Z, Ruvkun G. The mevalonate pathway regulates microRNA activity in Caenorhabditis elegans. *Proceedings of the National Academy of Sciences of the United States of America*. 2012. doi:10.1073/pnas.1202421109

59. Carvalho BC, Oliveira LC, Rocha CD, et al. Both knock-down and overexpression of Rap2a small GTPase in macrophages result in impairment of NF-κB activity and inflammatory gene expression. *Molecular Immunology*. 2019. doi:10.1016/j.molimm.2019.02.015

60. Schrimpe-Rutledge AC, Codreanu SG, Sherrod SD, McLean JA. Untargeted Metabolomics Strategies—Challenges and Emerging Directions. *Journal of the American Society for Mass Spectrometry*. 2016. doi:10.1007/s13361-016-1469-y

61. Kaddurah-Daouk R, Kristal BS, Weinshilboum RM. Metabolomics: A global biochemical approach to drug response and disease. *Annual Review of Pharmacology and Toxicology*. 2008. doi:10.1146/annurev.pharmtox.48.113006.094715

62. Matsuda F. Technical Challenges in Mass Spectrometry-Based Metabolomics. *Mass Spectrometry*. 2016. doi:10.5702/massspectrometry.s0052

63. Jirtle RL, Skinner MK. Environmental epigenomics and disease susceptibility. *Nature*

Reviews Genetics. 2007. doi:10.1038/nrg2045

64. Dworkin RH, McDermott MP, Farrar JT, O'Connor AB, Senn S. Interpreting patient treatment response in analgesic clinical trials: Implications for genotyping, phenotyping, and personalized pain treatment. *Pain*. 2014. doi:10.1016/j.pain.2013.09.019

65. Jacob M, Lopata AL, Dasouki M, Abdel Rahman AM. Metabolomics toward personalized medicine. *Mass Spectrometry Reviews*. 2017.

66. Tolstikov V, James Moser A, Sarangarajan R, Narain NR, Kiebish MA. Current status of metabolomic biomarker discovery: Impact of study design and demographic characteristics. *Metabolites*. 2020. doi:10.3390/metabo10060224

67. Everett JR. Pharmacometabonomics in humans: A new tool for personalized medicine. *Pharmacogenomics*. 2015. doi:10.2217/pgs.15.20

68. Carraro S, Piacentini G, Lusiani M, et al. Exhaled air temperature in children with bronchopulmonary dysplasia. *Pediatric Pulmonology*. 2010. doi:10.1002/ppul.21317

69. Dettmer K, Aronov PA, Hammock BD. Mass spectrometry-based metabolomics. *Mass Spectrometry Reviews*. 2007. doi:10.1002/mas.20108

70. Dunn WB, Wilson ID, Nicholls AW, Broadhurst D. The importance of experimental design and QC samples in large-scale and MS-driven untargeted metabolomic studies of humans. *Bioanalysis*. 2012. doi:10.4155/bio.12.204

71. Yi L, Dong N, Yun Y, et al. Chemometric methods in data processing of mass spectrometry-based metabolomics: A review. *Analytica Chimica Acta*. 2016. doi:10.1016/j.aca.2016.02.001

72. Cajka T, Fiehn O. Toward Merging Untargeted and Targeted Methods in Mass Spectrometry-Based Metabolomics and Lipidomics. *Analytical Chemistry*. 2016. doi:10.1021/acs.analchem.5b04491

73. Hamzeiy H, Cox J. What computational non-targeted mass spectrometry-based metabolomics can gain from shotgun proteomics. *Current Opinion in Biotechnology*. 2017. doi:10.1016/j.copbio.2016.11.014

74. Xu Y, Heilier JF, Madalinski G, et al. Evaluation of accurate mass and relative isotopic abundance measurements in the LTQ-Orbitrap mass spectrometer for further metabolomics database building. *Analytical Chemistry*. 2010. doi:10.1021/ac100271j

75. Wishart DS, Feunang YD, Marcu A, et al. HMDB 4.0: The human metabolome database for 2018. *Nucleic Acids Research*. 2018. doi:10.1093/nar/gkx1089

76. Horai H, Arita M, Kanaya S, et al. MassBank: A public repository for sharing mass spectral data for life sciences. *Journal of Mass Spectrometry*. 2010. doi:10.1002/jms.1777

77. Guijas C, Montenegro-Burke JR, Domingo-Almenara X, et al. METLIN: A Technology Platform for Identifying Knowns and Unknowns. *Analytical Chemistry*. 2018. doi:10.1021/acs.analchem.7b04424

78. Blaženović I, Kind T, Ji J, Fiehn O. Software tools and approaches for compound identification of LC-MS/MS data in metabolomics. *Metabolites*. 2018. doi:10.3390/metabo8020031

79. de Souza LP, Naake T, Tohge T, Fernie AR. From chromatogram to analyte to metabolite. How to pick horses for courses from the massive web resources for mass spectral plant metabolomics. *GigaScience*. 2017. doi:10.1093/gigascience/gix037

80. Little JL, Williams AJ, Pshenichnov A, Tkachenko V. Identification of «known unknowns» utilizing accurate mass data and chemspider. *Journal of the American Society for Mass Spectrometry*. 2012. doi:10.1007/s13361-011-0265-y

81. Dunn WB, Erban A, Weber RJM, et al. Mass appeal: Metabolite identification in mass spectrometry-focused untargeted metabolomics. *Metabolomics*. 2013. doi:10.1007/s11306-012-0434-4

82. Sumner LW, Amberg A, Barrett D, et al. Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics*. 2007. doi:10.1007/s11306-007-0082-2

83. Folberth J, Begemann K, Jöhren O, Schwaninger M, Othman A. MS2 and LC libraries for untargeted metabolomics: Enhancing method development and identification confidence. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*. 2020. doi:10.1016/j.jchromb.2020.122105

84. Fiehn O, Robertson D, Griffin J, et al. The metabolomics standards initiative (MSI). *Metabolomics*. 2007. doi:10.1007/s11306-007-0070-6

85. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Research*. 2012. doi:10.1093/nar/gkr988

86. Caspi R, Altman T, Billington R, et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Research*. 2014. doi:10.1093/nar/gkt1103

87. Bardou P, Mariette J, Escudié F, Djemiel C, Klopp C. Jvenn: An interactive Venn diagram viewer. *BMC Bioinformatics*. 2014. doi:10.1186/1471-2105-15-293

88. Tsugawa H, Cajka T, Kind T, et al. MS-DIAL: Data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nature Methods*. 2015. doi:10.1038/nmeth.3393

89. Bonini P, Kind T, Tsugawa H, Barupal DK, Fiehn O. Retip: Retention Time Prediction for

Compound Annotation in Untargeted Metabolomics. *Analytical Chemistry*. 2020. doi:10.1021/acs.analchem.9b05765

90. Narayanaswamy P, Teo G, Ow JR, et al. MetaboKit: a comprehensive data extraction tool for untargeted metabolomics. *Molecular Omics*. 2020. doi:10.1039/d0mo00030b

Supplementary tables

Supplementary Table 1			
Protein names	**Gene names**	**Student's T-test p-value**	**Student's T-test Differenc**
Translation initiation factor eIF-2B subunit delta	EIF2B4	2,6491	-1,08
BCL2/adenovirus E1B 19 kDa protein-interacting protein 3-like	BNIP3L	2,64253	-1,04628
Proteasomal ubiquitin receptor ADRM1	ADRM1	3,42823	-1,45877
Thromboxane-A synthase	TBXAS1	2,14246	-2,65853
Ubiquitin-protein ligase E3A	UBE3A	2,34873	-1,01414
Baculoviral IAP repeat-containing protein 1	NAIP	2,35951	-0,962915
Exosome complex component CSL4	EXOSC1	3,25102	-1,94119
von Willebrand factor A domain-containing protein 9	VWA9	3,21249	1,0017
Myeloid-associated differentiation marker	MYADM	2,46876	-0,901073
COP9 signalosome complex subunit 1	GPS1	2,57068	-1,6512
Actin beta	ACTB	1,78203	4,23435
Lactotransferrin	LTF	1,97549	5,68489
Serine/threonine-protein kinase MARK2	MARK2	2,1941	-1,57728
3-phosphoinositide-dependent protein kinase 1	PDPK1	2,29002	-1,1278
DNA (cytosine-5)-methyltransferase 3A	DNMT3A	2,88017	-1,36095
Translocation protein SEC62	SEC62	2,69098	-0,83733
Protein transport protein Sec24C	SEC24C	3,09638	-0,6466
Sjoegren syndrome/scleroderma autoantigen 1	SSSCA1	2,66599	1,16753
Cathelicidin antimicrobial peptide	CAMP	2,11231	4,27064
28S ribosomal protein S7, mitochondrial	MRPS7	2,15092	-1,50467
ORM1-like protein 3	ORMDL3	2,12002	-1,40518
Actinin alpha 4	ACTN4	2,06711	-2,05933
Transmembrane protein 205	TMEM205	2,42134	-2,26277
Dynamin-1-like protein	DNM1L	2,31057	-1,73744
Histone acetyltransferase type B catalytic subunit	HAT1	1,88924	-2,76738
2-deoxynucleoside 5-phosphate N-hydrolase 1	DNPH1	2,31399	0,988897
Syntaxin-6	STX6	2,06827	-2,33611
Uncharacterized protein KIAA0513	KIAA0513	2,1456	-1,23255
Galactosylgalactosylxylosylprotein	B3GAT3	2,22009	-1,21531

3-beta-glucuronosyltransferase 3			
Thioredoxin domain-containing protein 12	TXNDC12	2,84474	0,747147
Fibronectin	FN1	3,87831	-2,59544
Non-histone chromosomal protein HMG-14	HMGN1	2,56373	-0,967116
Prosaposin	PSAP	2,79447	0,875078
Annexin A3	ANXA3	2,0623	5,45663
Macrophage migration inhibitory factor	MIF	2,60382	1,46934
Leukosialin	SPN	2,59416	2,01043
High mobility group protein HMG-I/HMG-Y	HMGA1	2,41716	-1,1831
Diacylglycerol kinase alpha	DGKA	2,87246	-1,53102
Peroxiredoxin-6	PRDX6	2,77405	0,686326
60S ribosomal protein L9	RPL9	3,02732	0,67707
Prohibitin	PHB	3,38393	0,576388
Replication factor C subunit 5	RFC5	2,15914	-1,77835
Cysteine--tRNA ligase, cytoplasmic	CARS	3,65283	-1,31223
RNA-binding protein 25	RBM25	3,0307	-0,811191
Ras-related protein Rab-9°	RAB9A	3,98751	-1,23249
RNA polymerase II elongation factor ELL	ELL	2,09464	-1,41141
Methionine--tRNA ligase, cytoplasmic	MARS	2,12011	-1,51632
ATP synthase subunit e, mitochondrial	ATP5I	2,46174	1,055
40S ribosomal protein S15	RPS15	2,60154	-0,989971
Histone H3.2	HIST2H3A	2,11215	-4,91382
Mitochondrial 2-oxoglutarate/malate carrier protein	SLC25A11	2,25039	-1,14861
Sialic acid-binding Ig-like lectin 14	SIGLEC14	2,19903	-2,05611
Heat shock protein 75 kDa, mitochondrial	TRAP1	2,53508	-0,936696
TAR DNA-binding protein 43	TARDBP	2,54596	-0,825465
Eukaryotic translation initiation factor 4E-binding protein 1	EIF4EBP1	3,06487	0,743391
Inositol-tetrakisphosphate 1-kinase	ITPK1	2,65068	1,03663
Cullin-3	CUL3	2,89467	-2,51483
Tubulin beta-2A chain	TUBB2A	2,10634	-1,80102
Nuclear transcription factor Y subunit gamma	NFYC	2,71195	0,712255
V-type proton ATPase subunit F	ATP6V1F	2,17341	1,71373
Putative coiled-coil-helix-coiled-coil-helix domain-containing protein CHCHD2P9, mitochondrial	CHCHD2P9	2,63145	-1,8274

Ubiquitin-associated protein 2	UBAP2	2,17427	-1,44638
Rootletin	CROCC	2,96526	1,10605
Vacuolar protein sorting-associated protein 53 homolog	VPS53	2,51984	-2,16291
Rap1 GTPase-activating protein 2	RAP1GAP2	2,36566	1,31747
Tetratricopeptide repeat protein 37	TTC37	2,12155	-2,16444
Vacuolar protein sorting-associated protein 13C	VPS13C	3,49152	-0,673022
UPF0317 protein C14orf159, mitochondrial	C14orf159	1,97671	-1,88973
Coiled-coil domain-containing protein 91	CCDC91	3,14619	-2,12756
Nitrilase homolog 1	NIT1	2,48105	-0,895555
Copine-8	CPNE8	2,15429	-1,72138
Cytokine receptor-like factor 3	CRLF3	2,42804	-0,981761
LysM and putative peptidoglycan-binding domain-containing protein 2	LYSMD2	2,56517	1,33936
Mitochondrial Rho GTPase 1	RHOT1	2,26671	1,64559
Misshapen-like kinase 1	MINK1	2,36364	-1,95205
Oxidation resistance protein 1	OXR1	2,4458	-1,16105
Eukaryotic translation initiation factor 4E type 3	EIF4E3	2,06687	-1,74517
Cyclic GMP-AMP synthase	MB21D1	3,51982	-3,01569
GTPase IMAP family member 8	GIMAP8	5,72817	-0,482056
Cyclin-Y	CCNY	2,32371	1,97999
ATP-binding cassette sub-family F member 1	ABCF1	2,48835	-0,929731
Retinol dehydrogenase 11	RDH11	2,10449	-1,42718
Nuclear pore membrane glycoprotein 210	NUP210	3,86408	0,818427
AP-3 complex subunit sigma-1	AP3S1	2,27	1,30312
Protein YIPF5	YIPF5	4,20425	-1,07156
Coiled-coil domain-containing protein 47	CCDC47	2,65329	-1,0721
RalBP1-associated Eps domain-containing protein 1	REPS1	3,54157	1,00333
Copine-2	CPNE2	2,45587	0,974257
ERO1-like protein alpha	ERO1L	3,15015	1,10495
Zinc finger RNA-binding protein	ZFR	3,03535	-1,48686
Ubiquitin-conjugating enzyme E2 E2	UBE2E2	2,84301	4,58013
Calcium and integrin-binding protein 1	CIB1	2,43848	-1,66321
N-terminal Xaa-Pro-Lys N-methyltransferase 1	NTMT1	2,52139	0,823927

RNA-binding protein 4	RBM4	3,14836	0,634053
Sharpin	SHARPIN	2,26113	-1,16509
Phosducin-like protein 3	PDCL3	2,90454	1,1818
Translation initiation factor IF-3, mitochondrial	MTIF3	2,2022	1,54116
MIP18 family protein FAM96A	FAM96A	2,97977	1,25787
Calcyclin-binding protein	CACYBP	3,47277	-0,546935
Integrator complex subunit 7	INTS7	2,47123	-1,96085
Aurora kinase A-interacting protein	AURKAIP1	2,54637	-0,819901
Cell growth-regulating nucleolar protein	LYAR	2,96601	-1,88782
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex assembly factor 4	NDUFAF4	3,49299	-1,45721
Core histone macro-H2A.2	H2AFY2	2,90307	1,70402
Glyoxylate reductase/hydroxypyruvate reductase	GRHPR	3,14723	0,703817
Protein NipSnap homolog 3A	NIPSNAP3A	2,5962	-1,38459
U6 snRNA-associated Sm-like protein LSm7	LSM7	2,14617	1,48891
Glucocorticoid modulatory element-binding protein 2	GMEB2	2,49792	-0,921182
DNA helicase INO80	INO80	2,58241	-0,919009
FAS-associated factor 1	FAF1	2,66964	0,754553
Peptidyl-prolyl cis-trans isomerase E	PPIE	3,13143	-0,697405
Sialic acid-binding Ig-like lectin 9	SIGLEC9	2,52277	-1,3022
WD repeat-containing protein 7	WDR7	4,03392	1,10474
Myeloblastin	PRTN3	2,21776	1,43862

Supplementary Table 2			
Protein names	**Gene names**	**Student's T-test p-value**	**Student's T-test Difference**
Apolipoprotein C-II	APOC2	2,65529	1,73212
AP-2 complex subunit mu	AP2M1	2,70471	-2,07333
Ig kappa chain C region	IGKV3-11	2,96108	-1,84914
Transmembrane emp24 domain-containing protein 7	TMED7-TICAM2	3,15646	1,2845
Killer cell lectin-like receptor subfamily F member 1	KLRF1	2,42047	4,14973
Thromboxane-A synthase	TBXAS1	2,34402	-1,54743
Phosphatidylinositol 3-kinase	PIK3C3	2,45059	-2,0696

Basic leucine zipper and W2 domain-containing protein 1	BZW1	3,07285	-3,00868
Myeloid-associated differentiation marker	MYADM	2,3412	-1,68624
COP9 signalosome complex subunit 1	GPS1	2,65123	-1,78145
G protein subunit beta 2	GNB2	2,54111	2,01711
Protein FAM195B	FAM195B	3,41629	-1,7937
Lactotransferrin	LTF	1,8043	4,34023
Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-3	GNB3	2,79654	1,91601
2,5-phosphodiesterase 12	PDE12	2,60798	-1,04542
Aminoacyl-tRNA synthetases	WARS	3,12091	2,4317
Spermatogenesis-defective protein 39 homolog	VIPAS39	3,48995	-1,63996
Translocating chain-associated membrane protein 1	TRAM1	2,66561	-0,725739
REST corepressor 1	RCOR1	2,00124	-2,87949
Cytochrome B5 type B	CYB5B	2,12077	-1,56355
Myelin basic protein	MBP	1,91818	-1,93226
ATP-dependent RNA helicase DDX39A	DDX39A	2,85283	-0,690964
Insulin-like growth factor 2 mRNA-binding protein 3	IGF2BP3	2,01345	-2,89959
Inhibitor of nuclear factor kappa-B kinase subunit beta	IKBKB	3,013	-0,974699
Cytochrome b-c1 complex subunit 8	UQCRQ	2,71608	-0,836841
Histone H2B type 1-K	HIST1H2BK	2,13071	4,77234
Mitochondrial import receptor subunit TOM70	TOMM70A	2,84649	-0,732833
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 3	NDUFA3	2,43581	4,07581
L-lactate dehydrogenase A chain	LDHA	3,77584	0,746957
Carbonic anhydrase 2	CA2	2,22256	-2,72336
Tumor necrosis factor	TNF	2,10769	2,12428
Ubiquitin-like protein ISG15	ISG15	2,30345	2,85696
Retinoblastoma-associated protein	RB1	2,19513	-2,1079
Epoxide hydrolase 1	EPHX1	2,82814	-1,09211
Beta-glucuronidase	GUSB	2,20811	1,2849
Neutrophil elastase	ELANE	2,1428	1,29188
Interferon-induced protein with tetratricopeptide repeats 1	IFIT1	1,88922	2,04699
Protein kinase C alpha type	PRKCA	2,27297	-1,02004

60S ribosomal protein L35a	RPL35A	2,95199	3,69147
Tumor necrosis factor receptor superfamily member 1B	TNFRSF1B	3,53611	1,02287
Succinate dehydrogenase [ubiquinone] iron-sulfur subunit, mitochondrial	SDHB	3,68361	-0,970757
Oxysterol-binding protein 1	OSBP	2,47088	-0,857971
Immunoglobulin alpha Fc receptor	FCAR	1,86006	-2,38474
Transcriptional repressor protein YY1	YY1	2,45033	-1,29518
Phosphatidylinositol 3-kinase regulatory subunit alpha	PIK3R1	2,72729	-0,716925
SHC-transforming protein 1	SHC1	2,88778	-0,660619
Enoyl-CoA hydratase, mitochondrial	ECHS1	3,98197	0,886471
ATP synthase subunit gamma, mitochondrial	ATP5C1	2,45458	3,61616
Signal transducer and activator of transcription 3	STAT3	2,69252	-1,61647
Guanine nucleotide-binding protein G(q) subunit alpha	GNAQ	3,18343	-1,22688
Cyclin-dependent kinase 9	CDK9	3,08943	-1,72175
Signal transducer and activator of transcription 5B	STAT5B	2,53191	-1,12713
Pterin-4-alpha-carbinolamine dehydratase	PCBD1	2,61119	0,791427
Serine/threonine-protein phosphatase PP1-alpha catalytic subunit	PPP1CA	3,06993	0,90153
U6 snRNA-associated Sm-like protein LSm6	LSM6	1,88535	2,34564
40S ribosomal protein S15	RPS15	2,58079	-1,33849
SUMO-conjugating enzyme UBC9	Ube2i	3,88061	-2,33522
Signal peptidase complex catalytic subunit SEC11A	SEC11A	2,88991	-1,23157
Neutrophil gelatinase-associated lipocalin	LCN2	1,85332	3,75308
Tyrosine-protein phosphatase non-receptor type 11	PTPN11	2,86577	-0,697657
Quinone oxidoreductase	CRYZ	2,82319	-2,84019
Mitochondrial-processing peptidase subunit alpha	PMPCA	2,40649	-1,27734
TAR DNA-binding protein 43	TARDBP	3,80335	-1,02279
Cytoskeleton-associated protein 5	CKAP5	2,77314	-0,818856
Putative heat shock protein HSP 90-alpha A4	HSP90AA4P	3,18752	2,13964
BRO1 domain-containing protein BROX	BROX	2,05345	-1,40511
Microtubule-associated protein 1S	MAP1S	2,9133	-1,2984
Type-1 angiotensin II receptor-associated protein	AGTRAP	1,98753	-1,59994
UPF0317 protein C14orf159, mitochondrial	C14orf159	1,91081	-3,06433
CLIP-associating protein 1	CLASP1	2,35413	4,17119

Putative phospholipase B-like 2	PLBD2	2,57368	-1,352
Tudor domain-containing protein 7	TDRD7	1,9284	-2,67805
Splicing factor U2AF 26 kDa subunit	U2AF1L4	2,29238	2,59262
Nuclear pore complex protein Nup133	NUP133	2,82081	-1,76479
Ubiquitin carboxyl-terminal hydrolase 7	USP7	3,38129	-1,2865
MOB kinase activator 3A	MOB3A	2,52959	1,19402
TNFAIP3-interacting protein 3	TNIP3	2,22684	1,54706
Retinoid-binding protein 7	RBP7	2,00695	-1,99031
Sorting nexin-18	SNX18	3,35005	-0,862465
Calcineurin B homologous protein 1	CHP1	3,56584	-1,433
Gamma-soluble NSF attachment protein	NAPG	2,17734	-1,61833
Death-inducer obliterator 1	DIDO1	2,5816	-1,03425
Protein PBDC1	PBDC1	2,66488	-1,00433
UBX domain-containing protein 6	UBXN6	2,36709	-1,66088
Rab3 GTPase-activating protein non-catalytic subunit	RAB3GAP2	2,63818	-1,61327
Rab GTPase-binding effector protein 2	RABEP2	2,22598	-1,54912
Complement component C1q receptor	CD93	2,6965	-1,76187
39S ribosomal protein L17, mitochondrial	MRPL17	5,41162	-0,887653
Structural maintenance of chromosomes protein 4	SMC4	2,22263	-1,35577
CKLF-like MARVEL transmembrane domain-containing protein 6	CMTM6	3,1136	-1,45964
ER membrane protein complex subunit 3	EMC3	2,91134	-1,45108
Protein RCC2	RCC2	3,3307	0,539761
COMM domain-containing protein 3	COMMD3	2,83432	1,0291
Death domain-associated protein 6	DAXX	2,32283	-1,55871
Ragulator complex protein LAMTOR2	LAMTOR2	2,78285	1,85818
WW domain-binding protein 11	WBP11	2,1449	-1,45081
ARF GTPase-activating protein GIT1	GIT1	2,76892	-0,881892
60S ribosomal protein L36	RPL36	3,10071	-1,50514
Mitochondrial carrier homolog 2	MTCH2	3,62523	-1,69682

Supplementary Table 3			
Protein names	**Gene names**	**Student's T-test p-value (-log)**	**Student's T-test Difference**

U4/U6.U5 tri-snRNP-associated protein 2	USP39	4,64476	-3,37603
Beta-hexosaminidase	HEXA	4,36552	-4,00728
Inhibitor of nuclear factor kappa-B kinase subunit beta	IKBKB	3,78714	-2,44653
Glutaredoxin-3	GLRX3	4,24581	-1,35854
Transthyretin	TTR	4,71492	2,5282
Tyrosine-protein kinase Lyn	LYN	3,1506	4,66983
Ubiquitin-conjugating enzyme E2 K	UBE2K	4,66944	-3,01738
Eukaryotic translation initiation factor 3 subunit M	EIF3M	3,98755	3,2244
Parkinson disease 7 domain-containing protein 1	PDDC1	4,19438	-2,06145
Optineurin	OPTN	4,89887	4,25239
Type I inositol 3,4-bisphosphate 4-phosphatase	INPP4A	3,67887	-2,00061
39S ribosomal protein L39, mitochondrial	MRPL39	4,20496	1,66681
Peptidyl-prolyl cis-trans isomerase E	PPIE	3,66788	-2,80125
UPF0568 protein C14orf166	C14orf166	5,11778	-3,80364
Ubiquitin-fold modifier-conjugating enzyme 1	UFC1	3,4539	-3,37135

Supplementary Table 4			
Protein names	**Gene names**	**Student's T-test p-value**	**Student's T-test Difference**
Acyl-coenzyme A thioesterase 1	ACOT1	2,91761	1,8944
Killer cell lectin-like receptor subfamily F member 1	KLRF1	2,35408	2,78495
Zinc finger Ran-binding domain-containing protein 2	ZRANB2	2,92689	2,29108
Ubiquitin-like modifier-activating enzyme 6	UBA6	2,34883	-2,60249
Phosphoglycolate phosphatase	PGP	3,87107	-1,19608
GTP-binding protein Rheb	RHEB	3,06775	-1,81305
Protein PRRC2C	PRRC2C	2,8109	0,757051
ATP synthase F1 subunit beta	ATP5B	3,34007	-2,09055
Serine/threonine-protein phosphatase	PPP4C	2,37124	1,35351
Serine/arginine-rich splicing factor 2	SRSF2	5,07821	0,553397
Sulfatase-modifying factor 2	SUMF2	2,50839	0,993273
Lon protease homolog, mitochondrial	LONP1	2,22189	-2,56644
Transcription initiation factor TFIID subunit 4	TAF4	2,52059	0,902915
Importin-5	IPO5	3,58239	-2,49514
Long-chain-fatty-acid--CoA ligase 4	ACSL4	2,33436	-1,4109

Syntaxin-10	STX10	2,39669	-1,44128
DnaJ homolog subfamily C member 13	DNAJC13	2,1122	-2,3428
Acyl-protein thioesterase 2	LYPLA2	2,1156	-1,79507
Apoptosis-inducing factor 1, mitochondrial	AIFM1	2,31319	-1,21197
Transthyretin	TTR	2,17034	2,11562
Superoxide dismutase [Mn], mitochondrial	SOD2	3,47564	-0,795899
60S acidic ribosomal protein P2	RPLP2	2,24316	-1,28098
Uroporphyrinogen decarboxylase	UROD	3,79867	0,718335
Beta-hexosaminidase subunit beta	HEXB	2,85845	-0,943622
Histone H2A.V	H2AFV	3,2165	-0,997572
Prolyl 4-hydroxylase subunit alpha-1	P4HA1	3,20409	0,991196
Proto-oncogene vav	VAV1	2,52521	-2,42122
Ganglioside GM2 activator	GM2A	2,90487	-3,32058
60S ribosomal protein L35a	RPL35A	2,02909	-2,83904
Oxysterol-binding protein 1	OSBP	2,92113	1,26053
DNA replication licensing factor MCM3	MCM3	2,76898	-1,55605
14-3-3 protein theta	YWHAQ	4,92329	1,3736
Proteasome subunit beta type	PSM8	3,48298	-0,942297
Signal transducer and activator of transcription 3	STAT3	2,54145	-1,81738
G protein-coupled receptor kinase 6	GRK6	3,42739	1,3553
Adapter molecule crk	CRK	2,61655	2,1844
Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3A	STT3A	2,26965	-2,68238
Signal recognition particle 9 kDa protein	SRP9	2,51648	1,1249
Rap1 GTPase-GDP dissociation stimulator 1	RAP1GDS1	3,31793	0,695963
Arginine--tRNA ligase, cytoplasmic	RARS	4,44209	-1,27305
Adenosine kinase	ADK	2,21462	1,47635
6.8 kDa mitochondrial proteolipid	MP68	2,5083	2,93227
Alpha-centractin	ACTR1A	2,40034	-1,12017
Lysozyme C	LYZ	2,97255	-0,677104
Ras-related protein Rap-1A	RAP1A	3,41596	1,57156
40S ribosomal protein S15	RPS15	3,55048	-4,64983
Disintegrin and metalloproteinase domain-containing protein 17	ADAM17	2,39435	1,27999

Tumor necrosis factor alpha-induced protein 2	TNFAIP2	2,08198	-2,86997
Serine/arginine-rich splicing factor 11	SRSF11	3,07982	0,768003
Serine/arginine-rich splicing factor 4	SRSF4	3,80635	0,766953
Syntaxin-5	STX5	3,04753	0,905947
26S proteasome non-ATPase regulatory subunit 2	PSMD2	4,24552	0,847585
Serine/arginine-rich splicing factor 6	SRSF6	4,02532	0,769634
Voltage-gated potassium channel subunit beta-2	KCNAB2	3,46974	-1,43573
Cold-inducible RNA-binding protein	CIRBP	2,74965	1,03243
Splicing factor 3A subunit 1	SF3A1	3,87414	0,572037
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 5	NDUFA5	2,78074	-1,08153
Putative oxidoreductase GLYR1	GLYR1	3,1385	0,765809
Torsin-1A-interacting protein 1	TOR1AIP1	3,22079	0,621812
Ubiquitin-associated protein 2	UBAP2	3,14002	0,946587
Aftiphilin	AFTPH	2,29981	1,27546
Protein unc-13 homolog D	UNC13D	2,71693	-0,935266
COMM domain-containing protein 1	COMMD1	2,60889	-1,30781
Nucleoporin Nup43	NUP43	3,12088	-1,54425
TBC1 domain family member 15	TBC1D15	2,67981	1,16709
Serine/threonine-protein phosphatase 4 regulatory subunit 1	PPP4R1	3,14244	0,889703
TBC1 domain family member 5	TBC1D5	2,07304	-2,53172
La-related protein 4B	LARP4B	2,1023	-1,80506
Probable ubiquitin carboxyl-terminal hydrolase FAF-X	USP9X	4,08315	0,908494
Protein YIPF5	YIPF5	3,19344	0,935504
Mediator of RNA polymerase II transcription subunit 8	MED8	3,09076	1,09395
MMS19 nucleotide excision repair protein homolog	MMS19	2,19116	1,76638
Protein NipSnap homolog 1	NIPSNAP1	2,62543	1,20919
Serine/threonine-protein phosphatase CPPED1	CPPED1	2,96519	1,12858
Rho GTPase-activating protein 9	ARHGAP9	4,16595	0,576696
RNMT-activating mini protein	FAM103A1	3,08062	0,949156
N-terminal Xaa-Pro-Lys N-methyltransferase 1	NTMT1	2,88891	0,713439
Splicing factor 3B subunit 5	SF3B5	3,47688	0,699482
Chitinase domain-containing protein 1	CHID1	2,57685	1,00737

Protein phosphatase 1 regulatory subunit 12C	PPP1R12C	2,7828	0,802731
GTP-binding protein SAR1a	SAR1A	2,35959	1,43886
Mitochondrial import receptor subunit TOM22 homolog	TOMM22	2,38449	1,0312
Myoferlin	MYOF	2,45846	-2,84924
Apoptosis-associated speck-like protein containing a CARD	PYCARD	2,52971	-1,07121
Zinc finger protein 330	ZNF330	2,52607	0,902349
Hypoxia up-regulated protein 1	HYOU1	3,22441	-1,1079
Transforming acidic coiled-coil-containing protein 3	TACC3	3,13558	1,40283
Cytoplasmic dynein 1 light intermediate chain 1	DYNC1LI1	2,89763	-0,883852

Supplementary Table 5			
Protein names	**Gene names**	**Student's T-test p-value (-log)**	**Student's T-test Difference**
A-kinase anchor protein 9	AKAP9	5,09991	-5,45275
COMM domain-containing protein 4	COMMD4	3,28256	1,655
von Willebrand factor A domain-containing protein 8	VWA8	3,1464	4,5966
Heat shock protein family A member 8	HSPA8	3,11882	-2,77979
HEAT repeat-containing protein 5A	HEATR5A	5,3754	-4,19927
Transmembrane and coiled-coil domain-containing protein 1	TMCO1	2,71126	2,52264
Ras-related protein Rab-4B	RAB4B	3,39498	3,42834
Fructose-1,6-bisphosphatase isozyme 2	FBP2	3,98947	-2,4602
Transthyretin	TTR	2,97015	-4,66502
Anion exchange protein 2	SLC4A2	3,86877	-1,61614
Spermidine synthase	SRM	2,71081	-3,37023
High mobility group nucleosome-binding domain-containing protein 5	HMGN5	3,20722	-2,36694
Aminoacylase-1	ACY1	4,24728	-1,58658
T-lymphoma invasion and metastasis-inducing protein 1	TIAM1	4,06037	-2,40614
Myotubularin-related protein 3	MTMR3	2,69437	-2,73138
Ras-related protein Rab-31	RAB31	4,81537	1,53276

Spectrin alpha chain, non-erythrocytic 1	SPTAN1	2,76857	-2,80785
Protein RRP5 homolog	PDCD11	4,82902	2,97891
Rootletin	CROCC	5,55965	-5,07579
Ninein	NIN	7,24404	7,13879
Tetratricopeptide repeat protein 9C	TTC9C	3,83862	-3,18524
Mitochondrial import inner membrane translocase subunit TIM14	DNAJC19	3,04649	-3,09963
Proteasome subunit beta type-7	PSMB7	3,33627	3,04137
CD180 antigen	CD180	3,74148	-3,34947
M-phase phosphoprotein 8	MPHOSPH8	2,88289	2,43691
Ubiquitin-like protein 5	UBL5	4,15996	3,42452
Mitochondrial fission factor	MFF	2,83858	2,37847
Kinesin-like protein KIF13A	KIF13A	2,95261	-3,54405
Olfactory receptor 10A2	OR10A2	2,84901	-2,22888
Conserved oligomeric Golgi complex subunit 4	COG4	2,87489	3,17967
Phosphopantothenate--cysteine ligase	PPCS	3,81922	-4,05984
Histidine triad nucleotide-binding protein 3	HINT3	3,14403	-2,48642
3-oxoacyl-[acyl-carrier-protein]synthase, mitochondrial	OXSM	2,95555	-1,79409
Protocadherin beta-15	PCDHB15	6,76271	5,59694
General transcription factor 3C polypeptide 5	GTF3C5	2,95287	-1,7429

Supplementary Table 6			
Protein names	**Gene names**	**Student's T-test p-value (-log)**	**Student's T-test Difference**
40S ribosomal protein S9	RPS9	3,6641	-1,56307
Proteasome subunit alpha type	PSMA2	1,98229	-0,758658
Zinc finger protein 787	ZNF787	2,89637	1,0644
40S ribosomal protein S24	RPS24	2,43325	-1,99903
Clathrin heavy chain	CLTC	3,06215	-0,871793
GMP reductase	GMPR2	2,85874	-0,977707
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 10, mitochondrial	NDUFA10	2,2165	-1,33103
ATPase ASNA1	ASNA1	1,95426	-0,795187

Selenoprotein H	C11orf31	3,29106	0,445317
Proteasome activator complex subunit 2	PSME2	1,73527	0,931124
Calcium/calmodulin-dependent protein kinase type II subunit gamma	CAMK2G	2,31833	-1,06501
Isovaleryl-CoA dehydrogenase, mitochondrial	IVD	1,72859	-1,19169
Integrin-linked protein kinase	ILK	2,62889	-0,740769
Sulfotransferase	SULT1A4	2,26674	-1,4465
UPF0693 protein C10orf32	C10orf32-ASMT	1,98054	1,27572
Serine/arginine-rich splicing factor 7	SRSF7	2,86164	0,583411
Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDa subunit	DDOST	2,2896	-1,07467
DNA-directed RNA polymerase	POLR2A	3,52349	-1,16738
Thromboxane-A synthase	TBXAS1	3,55787	-2,35376
Fatty acid synthase	FASN	1,9973	-1,56089
Shootin-1	KIAA1598	2,02041	0,883891
Aconitate hydratase, mitochondrial	ACO2	2,64385	-0,568282
Small ubiquitin-related modifier 3	SUMO3	1,61329	1,95791
Pituitary tumor-transforming gene 1 protein-interacting protein	PTTG1IP	2,17257	1,53228
Unconventional myosin-Ig	MYO1G	2,56182	-0,850527
Polyadenylate-binding protein	PABPC4	3,38119	-1,486
Branched-chain-amino-acid aminotransferase	BCAT2	1,91725	-1,76094
Septin-11	11-set	1,72841	-1,04883
Annexin	ANXA6	2,49616	-0,943773
Protein PRRC2C	PRRC2C	2,91316	0,74127
Casein kinase II subunit alpha	CSNK2A1	2,93641	-0,547729
Complement receptor type 1	CR1	2,32872	-2,04296
Transducin beta-like protein 2	TBL2	2,31121	-0,550451
Transformation/transcription domain-associated protein	TRRAP	2,01142	1,9173
Protein transport protein Sec23A	SEC23A	2,95576	-1,05487
2,5-phosphodiesterase 12	PDE12	2,046	-0,666538
CAD protein	CAD	2,90084	-1,85107
Phosphatidylinositol 4,5-bisphosphate 3-kinase	PIK3CD	2,62517	-1,09182

catalytic subunit delta isoform			
Eukaryotic translation initiation factor 2A	EIF2A	2,45773	0,625928
Translocation protein SEC62	SEC62	2,18751	1,92992
Aminoacyl-tRNA synthetases	WARS	2,76214	3,00623
Sp110 nuclear body protein	SP110	1,51842	2,91582
AMP deaminase 2	AMPD2	2,13904	0,600659
Methyltransferase-like protein 7A	METTL7A	2,63543	-2,1536
RNA-binding protein with serine-rich domain 1	RNPS1	2,85446	0,561779
Beta-hexosaminidase	HEXA	1,95113	-2,08827
Proteasome-associated protein ECM29 homolog	KIAA0368	1,70311	-1,46049
REST corepressor 1	RCOR1	2,03264	0,763764
Ubiquitin thioesterase OTUB1	OTUB1	2,1114	-0,781033
Tubulin-specific chaperone D	TBCD	1,91019	-1,46345
Protein SSXT	SS18	1,92048	0,988246
Histone H3	H3F3B	2,12753	-3,47977
Lon protease homolog, mitochondrial	LONP1	1,77697	-1,1821
Transmembrane protein 205	TMEM205	1,7569	-1,23679
Mitochondrial import inner membrane translocase subunit TIM44	TIMM44	2,9368	-1,94859
Unconventional myosin-If	MYO1F	2,44848	-0,495675
Membrane-associated progesterone receptor component 1	PGRMC1	2,05226	0,858472
Dynactin subunit 6	DCTN6	3,56454	1,02879
High mobility group nucleosome-binding domain-containing protein 4	HMGN4	1,5482	1,99868
Lysosomal alpha-mannosidase	MAN2B1	3,21644	-1,05135
Fructose-1,6-bisphosphatase isozyme 2	FBP2	2,12745	-1,42789
Apoptotic protease-activating factor 1	APAF1	3,31953	-1,94591
Exportin-1	XPO1	1,95858	-1,28982
FYN-binding protein	FYB	1,93861	0,74564
Actin-related protein 2/3 complex subunit 3	ARPC3	1,78974	-0,900927
Serine palmitoyltransferase 1	SPTLC1	1,81264	-0,802845
Syntaxin-7	STX7	2,52298	1,08607
U4/U6 small nuclear ribonucleoprotein Prp3	PRPF3	2,22094	1,32621
Sjoegren syndrome nuclear autoantigen 1	SSNA1	1,78858	0,937857

A-kinase anchor protein 8	AKAP8	1,7539	1,7313
Isocitrate dehydrogenase [NAD] subunit beta, mitochondrial	IDH3B	3,15707	-1,1633
Intron-binding protein aquarius	AQR	1,87165	-2,78409
Procollagen-lysine,2-oxoglutarate 5-dioxygenase 3	PLOD3	1,50536	-2,09434
Toll-like receptor 2	TLR2	2,55182	-1,35848
H/ACA ribonucleoprotein complex subunit 4	DKC1	2,01105	-1,06335
WD repeat-containing protein 1	WDR1	1,97443	-0,770498
Copine-3	CPNE3	1,55489	-1,74597
DnaJ homolog subfamily C member 13	DNAJC13	2,36391	-1,6024
NADH dehydrogenase [ubiquinone] iron-sulfur protein 2, mitochondrial	NDUFS2	2,06543	-0,99331
Core histone macro-H2A.1	H2AFY	3,75721	-0,528851
NADH dehydrogenase [ubiquinone] iron-sulfur protein 3, mitochondrial	NDUFS3	3,06566	-1,62435
Interferon-inducible double-stranded RNA-dependent protein kinase activator A	PRKRA	1,67361	-1,5785
Acyl-protein thioesterase 1	LYPLA1	2,42778	-1,43566
U5 small nuclear ribonucleoprotein 200 kDa helicase	SNRNP200	3,70966	-0,9646
Eukaryotic translation initiation factor 3 subunit J	EIF3J	3,2084	0,74896
AP-1 complex subunit gamma-like 2	AP1G2	1,75285	-1,69973
CAAX prenyl protease 1 homolog	ZMPSTE24	1,85534	-1,77427
Flotillin-1	FLOT1	2,09824	0,637147
Glutaredoxin-3	GLRX3	1,75863	-1,28176
TOX high mobility group box family member 4	TOX4	1,76785	1,06289
Mitofusin-2	MFN2	2,60447	-1,35445
Reticulon-3	RTN3	1,89289	1,07761
Vesicle-associated membrane protein-associated protein B/C	VAPB	1,65561	1,39322
NAD kinase	NADK	1,51436	-1,90481
AP-2 complex subunit alpha-1	AP2A1	1,86978	-0,968019
Peroxisomal membrane protein 11B	PEX11B	2,01916	-1,77389
L-lactate dehydrogenase A chain	LDHA	2,39312	-0,764338
NADH-cytochrome b5 reductase 3	CYB5R3	2,49853	-1,81116
Cytochrome c oxidase subunit 2	MT-CO2	4,25766	-2,13599

Coagulation factor XIII A chain	F13A1	2,20862	-1,28347
Aspartate aminotransferase, mitochondrial	GOT2	3,79135	-1,20688
Carbonic anhydrase 2	CA2	1,78639	-1,62715
Alpha-1-antichymotrypsin	SERPINA3	2,21278	-0,619604
Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2	RPN2	2,10215	-1,14633
Guanine nucleotide-binding protein G(i) subunit alpha-2	GNAI2	3,97126	-0,748285
Aldehyde dehydrogenase, mitochondrial	ALDH2	2,62104	-0,633041
60S acidic ribosomal protein P0	RPLP0	2,41834	-1,03796
Integrin beta-1	ITGB1	3,85059	-0,559403
Glycogen phosphorylase, liver form	PYGL	3,48884	-0,993727
Nucleophosmin	NPM1	2,82436	0,529038
Cathepsin D	CTSD	2,07031	-0,680798
Calpain-1 catalytic subunit	CAPN1	4,38378	-0,817368
Tubulin beta chain	TUBB	2,36384	-0,63584
Beta-hexosaminidase subunit beta	HEXB	2,28612	-1,46273
Tyrosine-protein kinase Lyn	LYN	1,78399	-1,02416
Tropomyosin beta chain	TPM2	1,64658	3,05169
Fumarate hydratase, mitochondrial	FH	2,46166	-0,662382
Annexin A6	ANXA6	3,76216	-0,862628
Beta-glucuronidase	GUSB	1,79523	-0,847849
Monocyte differentiation antigen CD14	CD14	4,40179	-0,876353
Annexin A5	ANXA5	4,71894	-1,03138
Glutathione S-transferase P	GSTP1	3,9671	-1,00345
Chromosome transmission fidelity protein 8 homolog isoform 2	CHTF8	2,36213	0,763079
Non-secretory ribonuclease	RNASE2	3,36359	-2,21522
Cytochrome c oxidase subunit 5B, mitochondrial	COX5B	1,87498	0,748083
60 kDa heat shock protein, mitochondrial	HSPD1	2,59983	0,559345
Histidine--tRNA ligase, cytoplasmic	HARS	1,85024	-1,05112
Inosine-5-monophosphate dehydrogenase 2	IMPDH2	2,17477	-0,604184
Xaa-Pro dipeptidase	PEPD	1,87983	-0,917921
X-ray repair cross-complementing protein 5	XRCC5	2,34754	-0,580843
Glycogen [starch] synthase, muscle	GYS1	1,88678	-0,991416

Hematopoietic lineage cell-specific protein	HCLS1	2,11451	0,722314
Alcohol dehydrogenase [NADP(+)]	AKR1A1	2,12813	-0,77733
Pyruvate kinase PKM	PKM	2,99865	-0,588267
Proto-oncogene vav	VAV1	2,49495	-1,09344
Carbonyl reductase [NADPH] 1	CBR1	2,88512	-0,574039
Beta-galactosidase	GLB1	2,20238	-0,907407
NADPH--cytochrome P450 reductase	POR	5,65244	-2,06288
1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase gamma-2	PLCG2	2,09533	-0,627927
Fumarylacetoacetase	FAH	3,27055	-1,37436
Calpain-2 catalytic subunit	CAPN2	2,65472	-0,924127
ATP-dependent 6-phosphofructokinase, liver type	PFKL	2,37931	-0,72436
T-complex protein 1 subunit alpha	TCP1	3,04807	-0,777268
Nucleolin	NCL	3,95183	0,622529
Hexokinase-1	HK1	2,94551	-1,09912
Casein kinase II subunit alpha	CSNK2A2	2,14554	-0,821901
Nuclear factor NF-kappa-B p105 subunit	NFKB1	1,90645	-1,10802
Transcription factor BTF3	BTF3	2,318	0,827761
Lamin-B1	LMNB1	2,43947	0,495247
Integrin alpha-L	ITGAL	2,48374	-0,843106
Voltage-dependent anion-selective channel protein 1	VDAC1	2,86398	-1,27154
Nucleoside diphosphate kinase B	NME2	2,3839	-0,497314
Cytochrome b-c1 complex subunit 2, mitochondrial	UQCRC2	3,59401	-0,881662
Splicing factor, proline- and glutamine-rich	SFPQ	2,58002	0,475691
NAD-dependent malic enzyme, mitochondrial	ME2	2,23534	-0,596182
Carnitine O-palmitoyltransferase 2, mitochondrial	CPT2	1,79399	-1,19811
ATP synthase F(0) complex subunit B1, mitochondrial	ATP5F1	2,16714	-1,39646
ATP synthase subunit alpha, mitochondrial	ATP5A1	2,64284	-0,513721
Threonine--tRNA ligase, cytoplasmic	TARS	1,75598	-1,26384
Proteasome subunit beta type-4	PSMB4	3,1436	-0,890987
Mitogen-activated protein kinase 1	MAPK1	2,61753	-0,784326
Protein PML	PML	1,8994	0,901271
Elongation factor 1-delta	EEF1D	2,36834	1,27918
Endoplasmic reticulum resident protein 29	ERP29	2,92356	0,541735

High affinity immunoglobulin epsilon receptor subunit gamma	FCER1G	2,38926	0,746802
HLA class I histocompatibility antigen, B-53 alpha chain	HLA-B	2,7472	0,790903
Leukocyte elastase inhibitor	SERPINB1	3,1684	-0,859692
Coronin-1A	CORO1A	2,26719	0,63936
cAMP-dependent protein kinase type II-beta regulatory subunit	PRKAR2B	1,83878	0,900121
Cytochrome b-c1 complex subunit 1, mitochondrial	UQCRC1	3,39441	-0,660538
3-hydroxyisobutyrate dehydrogenase, mitochondrial	HIBADH	2,68811	-1,24953
14-3-3 protein beta/alpha	YWHAB	1,67599	1,40275
Cleavage stimulation factor subunit 2	CSTF2	1,74765	0,979084
Oxygen-dependent coproporphyrinogen-III oxidase, mitochondrial	CPOX	2,05769	-2,51356
DNA-directed RNA polymerase II subunit RPB9	POLR2I	2,39122	-0,987918
V-type proton ATPase catalytic subunit A	ATP6V1A	2,47061	-0,527253
T-complex protein 1 subunit zeta	CCT6A	2,77191	-0,578092
Malate dehydrogenase, mitochondrial	MDH2	3,06233	-0,384426
Trifunctional enzyme subunit alpha, mitochondrial	HADHA	3,86287	-0,949674
Tyrosine-protein kinase CSK	CSK	4,6583	-0,96348
Isoleucine--tRNA ligase, cytoplasmic	IARS	2,15091	-1,79023
Enoyl-CoA delta isomerase 1, mitochondrial	ECI1	2,26472	-1,16629
Leucine-rich PPR motif-containing protein, mitochondrial	LRPPRC	2,16936	-0,89247
3-ketoacyl-CoA thiolase, mitochondrial	ACAA2	3,14964	-0,626313
Glycerol-3-phosphate dehydrogenase, mitochondrial	GPD2	2,4452	-1,19893
Tyrosine-protein kinase SYK	SYK	3,76973	-0,901762
Chromobox protein homolog 5	CBX5	2,10899	1,08514
Probable 28S rRNA (cytosine(4447)-C(5))-methyltransferase	NOP2	1,6149	1,19726
Dual specificity mitogen-activated protein kinase kinase 3	MAP2K3	2,15019	-0,813216
Glucosamine-6-phosphate isomerase 1	GNPDA1	2,94327	-2,06763
F-actin-capping protein subunit alpha-2	CAPZA2	2,63881	-0,444223
Eukaryotic translation initiation factor 1A, X-chromosomal	EIF1AX	1,67984	1,34877

Glutamine--tRNA ligase	QARS	2,08611	-0,715545
Protein PRRC2A	PRRC2A	2,70659	0,880562
Signal recognition particle 9 kDa protein	SRP9	1,84039	0,937887
Alanine--tRNA ligase, cytoplasmic	AARS	2,04659	-0,991529
Serine--tRNA ligase, cytoplasmic	SARS	3,79332	-0,612337
Very long-chain specific acyl-CoA dehydrogenase, mitochondrial	ACADVL	2,39053	-0,604834
Selenide, water dikinase 1	SEPHS1	2,74704	-0,910801
Carnitine O-palmitoyltransferase 1, liver isoform	CPT1A	2,06977	-0,688274
Vasodilator-stimulated phosphoprotein	VASP	3,56988	0,618461
Dynamin-2	DNM2	2,5002	-0,707044
T-complex protein 1 subunit delta	CCT4	3,40388	-0,896176
Isocitrate dehydrogenase [NAD] subunit gamma, mitochondrial	IDH3G	1,7228	-1,38255
Host cell factor 1	HCFC1	2,33379	1,18256
Fatty aldehyde dehydrogenase	ALDH3A2	2,11063	-1,79095
Peroxisomal multifunctional enzyme type 2	HSD17B4	3,64602	-0,989813
Vesicle-associated membrane protein 7	VAMP7	1,57467	-1,24284
Ribosomal protein S6 kinase alpha-3	RPS6KA3	3,99285	-1,39619
Heterogeneous nuclear ribonucleoprotein M	HNRNPM	3,1392	0,613544
Heterogeneous nuclear ribonucleoprotein F	HNRNPF	2,54012	-0,572794
Hexokinase-3	HK3	2,82374	-0,598927
Tricarboxylate transport protein, mitochondrial	SLC25A1	2,76965	-1,40092
ATP-citrate synthase	ACLY	2,75442	-0,910218
Coatomer subunit beta	COPB1	1,98502	-0,845007
Coatomer subunit alpha	COPA	2,7146	-0,661044
Methylosome subunit pICln	CLNS1A	2,28733	1,07814
Tyrosine--tRNA ligase, cytoplasmic	YARS	2,22958	-0,564491
Adenosine kinase	ADK	2,61725	-2,0628
Puromycin-sensitive aminopeptidase	NPEPPS	2,30021	-0,695954
IgG receptor FcRn large subunit p51	FCGRT	1,48641	-1,68566
Methionine--tRNA ligase, cytoplasmic	MARS	1,7956	-1,01287
AP-1 complex subunit sigma-2	AP1S2	3,79375	-0,432019
Protein transport protein Sec61 subunit beta	SEC61B	2,96839	1,28473

Ribose-phosphate pyrophosphokinase 1	PRPS1	1,57901	-1,25981
Ubiquitin-conjugating enzyme E2 K	UBE2K	1,99856	-0,768679
Actin-related protein 3	ACTR3	2,80776	-0,519711
Alpha-centractin	ACTR1A	2,68743	-0,886678
COP9 signalosome complex subunit 2	COPS2	2,74075	-1,22798
Ras-related protein Rap-2b	RAP2B	1,65553	-1,45289
Protein max	MAX	3,66876	1,10512
60S ribosomal protein L27	RPL27	2,11708	-1,35084
40S ribosomal protein S7	RPS7	1,73713	-1,04498
Serine/threonine-protein phosphatase PP1-alpha catalytic subunit	PPP1CA	2,44289	-0,545941
40S ribosomal protein S13	RPS13	1,94349	-1,10148
Histone H4	HIST1H4A	2,47475	-0,484609
60S ribosomal protein L23	RPL23	3,64318	-1,41488
Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-2	GNB2	2,74817	-0,609241
60S ribosomal protein L10a	RPL10A	2,34054	-0,890566
60S ribosomal protein L8	RPL8	2,59349	-0,765245
Peptidyl-prolyl cis-trans isomerase A	PPIA	2,46366	0,548413
Guanine nucleotide-binding protein G(s) subunit alpha isoforms short	GNAS	2,68883	-1,20628
Serine/threonine-protein phosphatase 2A 55 kDa regulatory subunit B alpha isoform	PPP2R2A	2,70699	-0,640564
40S ribosomal protein S21	RPS21	2,96434	0,700949
Guanine nucleotide-binding protein subunit beta-2-like 1	GNB2L1	2,4237	-1,0795
SUMO-conjugating enzyme UBC9	Ube2i	1,81859	-1,01179
Serine/threonine-protein phosphatase 2A catalytic subunit alpha isoform	PPP2CA	2,60837	-1,08207
Nuclease-sensitive element-binding protein 1	YBX1	3,09281	0,670355
Signal peptidase complex catalytic subunit SEC11A	SEC11A	2,20306	-1,17365
Tubulin beta-4B chain	TUBB4B	4,74287	-0,732298
Immunoglobulin-binding protein 1	IGBP1	1,80974	1,22574
Glutathione S-transferase omega-1	GSTO1	1,97652	-0,715778
DNA-dependent protein kinase catalytic subunit	PRKDC	3,60594	-0,776316

28S ribosomal protein S36, mitochondrial	MRPS36	3,33779	0,712056
Serine/arginine-rich splicing factor 3	SRSF3	1,921	0,936574
Histone H3.2	HIST2H3A	1,77461	-4,14398
Phosphate carrier protein, mitochondrial	SLC25A3	3,70505	-1,12527
Transcriptional activator protein Pur-alpha	PURA	2,49857	-0,721097
Methylmalonate-semialdehyde dehydrogenase [acylating], mitochondrial	ALDH6A1	1,81749	-0,926159
DNA topoisomerase 2-beta	TOP2B	2,03779	-1,06745
Single-stranded DNA-binding protein, mitochondrial	SSBP1	1,91343	-0,834792
Tyrosine-protein phosphatase non-receptor type 11	PTPN11	2,18192	-0,681671
Cytoskeleton-associated protein 4	CKAP4	2,28852	0,70689
Mitochondrial-processing peptidase subunit alpha	PMPCA	1,98611	-1,26331
Vesicular integral-membrane protein VIP36	LMAN2	2,84059	-1,80768
Heat shock protein 75 kDa, mitochondrial	TRAP1	2,20098	-1,40645
Protein flightless-1 homolog	FLII	2,88011	-0,990279
Lymphocyte cytosolic protein 2	LCP2	2,46652	0,61867
Ubiquitin carboxyl-terminal hydrolase 4	USP4	2,06133	-1,2616
Serine/arginine-rich splicing factor 9	SRSF9	2,43055	0,998055
Serine/arginine-rich splicing factor 6	SRSF6	2,05764	0,961327
Voltage-gated potassium channel subunit beta-2	KCNAB2	1,86417	-1,03915
Cytoplasmic dynein 1 intermediate chain 2	DYNC1I2	2,2055	0,565186
V-type proton ATPase 116 kDa subunit a isoform 3	TCIRG1	2,73568	-1,00609
Diacylglycerol kinase zeta	DGKZ	2,18066	-1,16102
Stromal interaction molecule 1	STIM1	2,51708	0,529726
Bleomycin hydrolase	BLMH	2,4213	-2,6743
Ubiquitin-associated protein 2-like	UBAP2L	2,51005	0,794757
Cytoplasmic dynein 1 heavy chain 1	DYNC1H1	2,08893	-0,718312
Guanine nucleotide-binding protein subunit alpha-13	GNA13	3,499	-1,02924
RNA-binding protein 39	RBM39	2,35072	-0,576593
Protein disulfide-isomerase A5	PDIA5	2,29842	-1,65076
Pre-mRNA-splicing regulator WTAP	WTAP	3,47656	1,00766
Septin-2	02-set	2,73964	-0,493138
116 kDa U5 small nuclear ribonucleoprotein component	EFTUD2	2,0319	-0,905758

Neutrophil cytosol factor 4	NCF4	2,8733	-0,884125
Transcription elongation factor B polypeptide 2	TCEB2	2,40538	0,768398
Ubiquitin-protein ligase E3C	UBE3C	1,86271	-0,88677
Scaffold attachment factor B1	SAFB	2,39901	0,49421
Splicing factor 1	SF1	2,1738	0,615689
ELAV-like protein 1	ELAVL1	1,98018	-0,963947
NEDD8	NEDD8	2,17462	0,993753
39S ribosomal protein L23, mitochondrial	MRPL23	2,08323	-1,19942
Hsp90 co-chaperone Cdc37	CDC37	2,66815	0,4987
Kynureninase	KYNU	2,52648	-1,09881
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 9, mitochondrial	NDUFA9	3,07419	-1,64304
UTP--glucose-1-phosphate uridylyltransferase	UGP2	3,22477	-1,51934
Macrophage-expressed gene 1 protein	MPEG1	1,79612	-0,976171
Protein LSM12 homolog	LSM12	1,8273	2,12645
Putative oxidoreductase GLYR1	GLYR1	2,53032	-0,515744
Presequence protease, mitochondrial	PITRM1	2,03165	-1,9872
RRP12-like protein	RRP12	1,46907	-1,74184
Engulfment and cell motility protein 2	ELMO2	2,17375	-0,731397
FAD synthase	FLAD1	1,5441	-1,37372
Microtubule-associated protein 1S	MAP1S	1,50047	-2,20191
Atlastin-3	ATL3	2,73621	-1,72114
Twinfilin-2	TWF2	3,72044	-0,856113
Pre-mRNA-processing-splicing factor 8	PRPF8	2,36247	-0,940774
La-related protein 1	LARP1	2,51979	1,01923
Biogenesis of lysosome-related organelles complex 1 subunit 2	BLOC1S2	1,70029	1,06084
Nesprin-3	SYNE3	2,01681	-0,676152
Neurobeachin-like protein 2	NBEAL2	1,9098	-1,54626
Vacuolar protein sorting-associated protein 13C	VPS13C	2,82177	-1,00394
Protein unc-13 homolog D	UNC13D	2,62309	-0,666353
Staphylococcal nuclease domain-containing protein 1	SND1	3,56017	-0,578772
Eukaryotic translation initiation factor 3 subunit M	EIF3M	1,65273	1,27447
Zinc finger CCCH-type antiviral protein 1	ZC3HAV1	2,8254	0,726645

Elongation factor Tu GTP-binding domain-containing protein 1	EFTUD1	2,20326	-1,65264
Myosin-14	MYH14	1,99775	-1,29583
KDEL motif-containing protein 2	KDELC2	1,85801	-0,96589
Protein LYRIC	MTDH	2,63261	0,528735
Cullin-associated NEDD8-dissociated protein 1	CAND1	1,98656	-1,02126
Ankyrin repeat and KH domain-containing protein 1	ANKHD1	3,26598	1,26275
Chromatin complexes subunit BAP18	BAP18	3,35271	0,550715
EH domain-binding protein 1-like protein 1	EHBP1L1	3,50919	0,47355
Cohesin subunit SA-2	STAG2	2,061	-1,67936
ADP-ribosylation factor GTPase-activating protein 1	ARFGAP1	2,31082	0,69476
Serine/threonine-protein phosphatase 6 regulatory ankyrin repeat subunit B	ANKRD44	1,72405	-1,3506
Parkinson disease 7 domain-containing protein 1	PDDC1	1,74804	-1,05617
NHL repeat-containing protein 2	NHLRC2	1,87036	-0,990679
Saccharopine dehydrogenase-like oxidoreductase	SCCPDH	3,26675	-1,76188
Dedicator of cytokinesis protein 8	DOCK8	2,96461	-0,979904
Nucleoporin NUP53	NUP35	2,30424	0,672071
E3 ubiquitin-protein ligase ZNRF2	ZNRF2	2,23478	1,35541
Splicing factor U2AF 26 kDa subunit	U2AF1L4	1,9353	1,16037
Nuclear pore complex protein Nup133	NUP133	1,67874	-1,50911
Sec1 family domain-containing protein 1	SCFD1	2,1278	-1,41071
PEST proteolytic signal-containing nuclear protein	PCNP	2,10385	0,909601
GTPase IMAP family member 1	GIMAP1	1,6353	-1,29585
UBX domain-containing protein 4	UBXN4	2,28551	0,822291
Dedicator of cytokinesis protein 2	DOCK2	2,89095	-0,907926
Translational activator GCN1	GCN1L1	5,08812	-0,941544
Pre-mRNA-splicing factor ATP-dependent RNA helicase PRP16	DHX38	2,34844	-1,71038
Ubiquitin carboxyl-terminal hydrolase 7	USP7	2,19789	-2,06639
Transcription elongation factor A protein-like 3	TCEAL3	1,95649	1,21938
Synapse-associated protein 1	SYAP1	1,93329	0,731824
Vacuolar protein sorting-associated protein 33A	VPS33A	2,24503	-2,17975
U8 snoRNA-decapping enzyme	NUDT16	3,72204	-2,43102
Copine-2	CPNE2	2,28787	-1,11777

RNA-binding protein 14	RBM14	2,82584	0,479664
Vacuolar protein sorting-associated protein 35	VPS35	1,98205	-0,870391
GPI transamidase component PIG-S	PIGS	2,46451	-1,59008
Chloride channel CLIC-like protein 1	CLCC1	1,8732	0,925029
3-hydroxyacyl-CoA dehydrogenase type-2	HSD17B10	3,18232	-1,11657
Histone H2B type 1-M	HIST1H2BM	1,6272	3,96336
Tyrosine-protein phosphatase non-receptor type 18	PTPN18	2,54493	1,35423
Protein NipSnap homolog 1	NIPSNAP1	2,32923	-1,65041
Serine/threonine-protein phosphatase CPPED1	CPPED1	2,09883	-0,678389
Partner of Y14 and mago	WIBG	2,88337	0,534585
ADP-dependent glucokinase	ADPGK	3,18701	-1,66906
COP9 signalosome complex subunit 4	COPS4	2,42094	-0,875684
Mini-chromosome maintenance complex-binding protein	MCMBP	1,96243	0,679536
Transmembrane protein 43	TMEM43	1,66103	-1,50206
Charged multivesicular body protein 4a	CHMP4A	2,67597	0,6476
RanBP-type and C3HC4-type zinc finger-containing protein 1	RBCK1	1,6776	1,56482
E3 ubiquitin-protein ligase TRIM4	TRIM4	1,64549	-1,45091
Phospholysine phosphohistidine inorganic pyrophosphate phosphatase	LHPP	1,72819	-1,87908
Thioredoxin-related transmembrane protein 4	TMX4	2,10657	-1,46578
Vacuolar protein sorting-associated protein 16 homolog	VPS16	2,68981	-1,46442
Tubulin beta-1 chain	TUBB1	1,84806	-2,94456
39S ribosomal protein L44, mitochondrial	MRPL44	2,38241	-0,816819
Nicotinamide/nicotinic acid mononucleotide adenylyltransferase 1	NMNAT1	2,1957	-1,5928
Calcyclin-binding protein	CACYBP	1,84313	1,10445
Manganese-transporting ATPase 13A1	ATP13A1	2,7988	-0,725937
Charged multivesicular body protein 1a	CHMP1A	2,1657	3,30591
Toll-like receptor 8	TLR8	1,93982	-1,10091
F-box only protein 6	FBXO6	1,68245	2,35402
Epimerase family protein SDR39U1	SDR39U1	1,86874	-1,4544
Protein FAM114A2	FAM114A2	2,28072	-0,888541

Phenylalanine--tRNA ligase beta subunit	FARSB	2,17824	-0,909596
Ethylmalonyl-CoA decarboxylase	ECHDC1	5,48376	-1,26085
Mycophenolic acid acyl-glucuronide esterase, mitochondrial	ABHD10	1,72156	-1,58812
ADP-ribosylation factor-like protein 8B	ARL8B	1,74521	-0,919146
ATP-dependent RNA helicase DDX18	DDX18	1,75179	-1,34623
THUMP domain-containing protein 1	THUMPD1	2,116	0,930059
Nuclear distribution protein nudE homolog 1	NDE1	2,31384	-0,925975
UDP-glucose:glycoprotein glucosyltransferase 1	UGGT1	2,71331	-0,868593
Adenosine deaminase CECR1	CECR1	2,8293	-1,63256
E3 ubiquitin-protein ligase RNF181	RNF181	2,96526	1,61198
Vacuolar protein sorting-associated protein 18 homolog	VPS18	1,74559	0,875457
Leucine--tRNA ligase, cytoplasmic	LARS	3,54143	-2,0452
Vacuolar protein sorting-associated protein 29	VPS29	1,98758	-1,37989
Cathepsin Z	CTSZ	1,90528	0,921716
Protein NipSnap homolog 3A	NIPSNAP3A	2,33258	-2,12625
Signal recognition particle subunit SRP68	SRP68	1,75838	-0,944346
Vacuolar protein sorting-associated protein 51 homolog	VPS51	1,81778	-2,21265
Tyrosine-protein kinase BAZ1B	BAZ1B	1,60595	-1,14092
Hematological and neurological expressed 1 protein	HN1	3,45913	0,744209
Apoptosis-associated speck-like protein containing a CARD	PYCARD	1,53768	-1,35594
26S proteasome non-ATPase regulatory subunit 13	PSMD13	2,83225	-1,62535
Zinc finger CCCH domain-containing protein 4	ZC3H4	3,04223	0,711782
Translation machinery-associated protein 7	TMA7	2,82494	1,09774
AP-3 complex subunit mu-1	AP3M1	2,04928	-0,973914
Thyroid hormone receptor-associated protein 3	THRAP3	2,15996	0,623545
Acyl-coenzyme A thioesterase 9, mitochondrial	ACOT9	2,79518	-0,808127
Deoxyribose-phosphate aldolase	DERA	1,83758	-1,00717
Mitochondrial fission 1 protein	FIS1	2,46892	-1,03182
Deoxynucleoside triphosphate triphosphohydrolase SAMHD1	SAMHD1	2,76034	-0,5844
Unconventional myosin-Va	MYO5A	2,22747	-1,72203

Leucine-rich repeat flightless-interacting protein 2	LRRFIP2	2,55254	1,61824
Mitochondrial carrier homolog 2	MTCH2	2,68093	-1,74277
Sulfide:quinone oxidoreductase, mitochondrial	SQRDL	5,68902	-0,769156
Apoptotic chromatin condensation inducer in the nucleus	ACIN1	2,72152	0,584889

Supplementary Table 7	
Urine	**Plasma**
102 - D-Tartaric acid	107 - L-Leucine
106 - Theophylline	110 - D-Tryptophan
107 - L-Leucine	12 - Uric acid
11 - D-Aspartic acid	126 - Pyrrolidonecarboxylic acid
110 - D-Tryptophan	130 - Caffeine
113 - Allantoin	137 - L-Kynurenine
119 - Asymmetric dimethylarginine	138 - Pyroglutamic acid
12 - Uric acid	14 - L-Serine
120 - Glucosamine	143 - Putrescine
124 - 1-Methylhistidine	144 - L-Lysine
127 - 4-Pyridoxic acid	151 - Norvaline
128 - Quinolinic acid	156 - L-Arginine
129 - Methylguanidine	16 - Citrulline
130 - Caffeine	161 - D-Ornithine
136 - Urocanic acid	168 - L-Histidine
137 - L-Kynurenine	171 - L-Pipecolic acid
139 - 4-Acetamidobutanoic acid	193 - Putrescine
14 - L-Serine	20 - Inosine
144 - L-Lysine	208 - N6,N6,N6-Trimethyl-L-lysine
149 - Glutarylcarnitine	217 - L-Carnitine
150 - L-Cystathionine	23 - L-Isoleucine
158 - L-Fucose	25 - L-Glutamic acid
160 - N-Methyl-L-glutamic acid	252 - Ornithine
162 - Xanthosine	277 - L-Acetylcarnitine
167 - Galactaric acid	298 - 2-Aminoisobutyric acid
168 - L-Histidine	306 - 5-Aminopentanoic acid
17 - Taurine	307 - L-Norleucine
171 - L-Pipecolic acid	315 - 3-Methoxytyrosine
192 - Gluconic acid	338 - L-Histidine
194 - 4-Trimethylammoniobutanoic acid	339 - L-Allothreonine
2 - L-Glutamine	35 - Pipecolic acid
208 - N6,N6,N6-Trimethyl-L-lysine	38 - L-Methionine

217 - L-Carnitine	382 - Cortisol
221 - Cyclic AMP	393 - Bilirubin
227 - N-Alpha-acetyllysine	406 - Indoleacetic acid
23 - L-Isoleucine	420 - Glycocholic acid
240 - 1-Methylnicotinamide	424 - Chenodeoxycholic acid glycine conjugate
259 - Asymmetric dimethylarginine	46 - Xanthine
26 - Ascorbic acid	470 - Indoleacetic acid
263 - Indoxyl sulfate	494 - Adipic acid
267 - Thiamine monophosphate	499 - mono-Methyl glutarate
27 - p-Hydroxyphenylacetic acid	50 - L-Tryptophan
272 - N-Acetyl-L-alanine	513 - Indoleacetaldehyde
274 - Hydroxypyruvic acid	52 - L-Proline
275 - D-Mannopyranose, 2-amino-2-deoxy-	561 - 4-Hydroxycinnamic acid
277 - L-Acetylcarnitine	59 - Niacinamide
278 - Riboflavin	6 - L-Threonine
290 - N-Acetylputrescine	62 - L-Phenylalanine
291 - N-Acetylgalactosamine	67 - Hypoxanthine
292 - N-Acetylglutamic acid	68 - Creatine
295 - Acetylcysteine	73 - L-Lysine
297 - Pantothenic acid	74 - L-Tyrosine
298 - 2-Aminoisobutyric acid	77 - L-Valine
305 - Ureidopropionic acid	79 - L-Homoserine
306 - 5-Aminopentanoic acid	9 - L-Kynurenine
309 - Adenosine	125 - Galactitol
316 - Alpha-Lactose	132 - D-Glucuronic acid
317 - 3-Hydroxybutyric acid	18 - Gluconolactone
318 - Imidazoleacetic acid	192 - Gluconic acid
337 - Guanidinosuccinic acid	2 - L-Glutamine
338 - L-Histidine	206 - Glyceric acid
34 - Malonic acid	26 - Ascorbic acid
344 - Citramalic acid	274 - Hydroxypyruvic acid
345 - Anserine	31 - Gluconic acid
347 - 5-Hydroxylysine	317 - 3-Hydroxybutyric acid
35 - Pipecolic acid	319 - beta-D-Galactopyranuronic acid
361 - Stachyose	333 - Oxalic acid
373 - Beta-Glycerophosphoric acid	34 - Malonic acid
379 - Gluconolactone	344 - Citramalic acid
38 - L-Methionine	379 - Gluconolactone
380 - 4-Hydroxybenzoic acid	390 - Maleic acid
381 - Tyramine	43 - Glyceric acid
385 - Xanthurenic acid	444 - Methylmalonic acid
390 - Maleic acid	462 - Pimelic acid

396 - Kynurenic acid	479 - Allyl isothiocyanate
40 - Adenine	480 - Itaconic acid
403 - Tryptamine	481 - Azelaic acid
406 - Indoleacetic acid	487 - Diacetyl
41 - 5'-Methylthioadenosine	5 - Citric acid
416 - 2-Pyrocatechuic acid	504 - Citraconic acid
420 - Glycocholic acid	519 - Sebacic acid
444 - Methylmalonic acid	538 - myo-Inositol
447 - 1H-Indole-3-acetamide	550 - D-Tagatose
448 - Hippuric acid	72 - D-Malic acid
454 - 3-(2-Hydroxyphenyl)propanoic acid	94 - trans-Aconitic acid
46 - Xanthine	109 - Betaine
468 - Dihydrobiopterin	11 - D-Aspartic acid
470 - Indoleacetic acid	119 - Asymmetric dimethylarginine
474 - Hydroquinone	120 - Glucosamine
484 - Phenylacetaldehyde	124 - 1-Methylhistidine
493 - Suberic acid	136 - Urocanic acid
496 - N-Acetylleucine	150 - L-Cystathionine
498 - Benzyl alcohol	166 - N-Methyl-D-aspartic acid
5 - Citric acid	17 - Taurine
50 - L-Tryptophan	184 - Palmitoyl-DL-carnitine
500 - Indole-3-methyl acetate	194 - 4-Trimethylammoniobutanoic acid
501 - Mevalonic acid	200 - Phosphorylcholine
502 - Vanillylmandelic acid	21 - Gamma-Aminobutyric acid
515 - 3-Hydroxyphenylacetic acid	22 - Cytosine
518 - Salicylic acid	227 - N-Alpha-acetyllysine
52 - L-Proline	240 - 1-Methylnicotinamide
527 - 5-Hydroxyindoleacetic acid	259 - Asymmetric dimethylarginine
555 - L-Arabitol	272 - N-Acetyl-L-alanine
58 - Carnosine	275 - D-Mannopyranose, 2-amino-2-deoxy-
6 - L-Threonine	297 - Pantothenic acid
607 - Indole	351 - Trigonelline
608 - Cholic acid	366 - D-Galactosamine
62 - L-Phenylalanine	386 - 2-Methylpropanal
632 - Cortisone	471 - δ-Valerolactone
65 - L-Aspartic acid	48 - L-Cystine
67 - Hypoxanthine	58 - Carnosine
68 - Creatine	601 - 12-Hydroxydodecanoic acid
70 - Guanosine	607 - Indole
72 - D-Malic acid	65 - L-Aspartic acid
74 - L-Tyrosine	70 - Guanosine
77 - L-Valine	76 - L-Asparagine

78 - Guanine	8 - N-Acetylneuraminic acid
8 - N-Acetylneuraminic acid	87 - Diethanolamine
80 - Pyridoxine	88 - 3-Aminoisobutanoic acid
87 - Diethanolamine	92 - Creatinine
88 - 3-Aminoisobutanoic acid	127 - 4-Pyridoxic acid
91 - Guanidoacetic acid	160 - N-Methyl-L-glutamic acid
92 - Creatinine	167 - Galactaric acid
94 - trans-Aconitic acid	216 - 3-Phosphoglyceric acid
	238 - Orotic acid
	263 - Indoxyl sulfate
	395 - Pregnenolone sulfate
	44 - Orotic acid
	515 - 3-Hydroxyphenylacetic acid
	555 - L-Arabitol
	557 - D-Ribose
	71 - Dihydrouracil