Check for updates

Metabolomics

High-Throughput Metabolomics by 1D NMR

Alessia Vignoli[+], Veronica Ghini[+], Gaia Meoni[+], Cristina Licari,
Panteleimon G. Takis, Leonardo Tenori, Paola Turano, and Claudio Luchinat*

Keywords:
fingerprinting · metabolomics ·
NMR · omic sciences ·
profiling
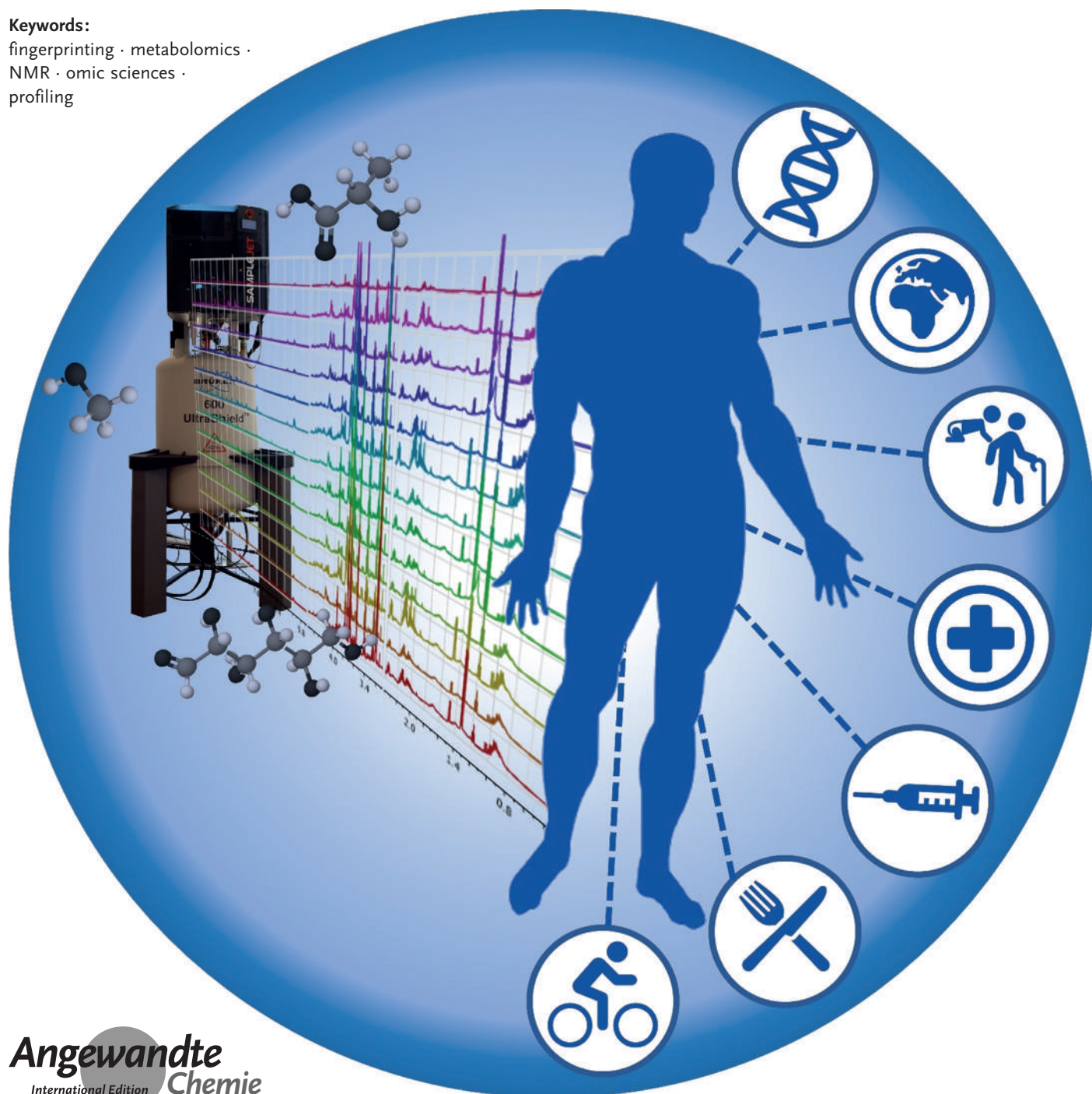
Metabolomics deals with the whole ensemble of metabolites (the metabolome). As one of the -omic sciences, it relates to biology, physiology, pathology and medicine; but metabolites are chemical entities, small organic molecules or inorganic ions. Therefore, their proper identification and quantitation in complex biological matrices requires a solid chemical ground. With respect to for example, DNA, metabolites are much more prone to oxidation or enzymatic degradation: we can reconstruct large parts of a mammoth's genome from a small specimen, but we are unable to do the same with its metabolome, which was probably largely degraded a few hours after the animal's death. Thus, we need standard operating procedures, good chemical skills in sample preparation for storage and subsequent analysis, accurate analytical procedures, a broad knowledge of chemometrics and advanced statistical tools, and a good knowledge of at least one of the two metabolomic techniques, MS or NMR. All these skills are traditionally cultivated by chemists. Here we focus on metabolomics from the chemical standpoint and restrict ourselves to NMR. From the analytical point of view, NMR has pros and cons but does provide a peculiar holistic perspective that may speak for its future adoption as a population-wide health screening technique.

From the Contents

1. What is Metabolomics

Metabolomics is the accepted name for the -omic science that deals with the characterization of the metabolome, in turn defined as the whole set of metabolites in a certain biological system such as a cell, a tissue, an organ, or an entire organism. The term metabonomics, as distinct from metabolomics, is also used when we refer to the science of understanding the network of interactions that metabolites undergo in a living system and is therefore closer to the concept of systems biology. Metabolomics deals with the objects, metabonomics with their interconnections and functions.[1] In practice the two terms are mostly used interchangeably.

Metabolomics is downstream with respect to the other -omic sciences, and this has important implications, as
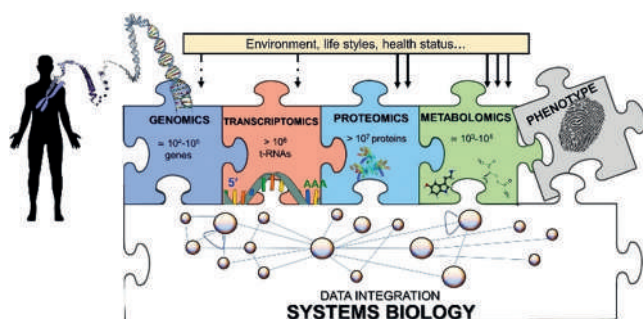
illustrated in Figure 1. Metabolomics is much more influenced by environmental factors, and this makes metabolomic data more difficult to interpret but also richer of information about the health status of an individual. While the genome is (almost) invariant throughout the lifespan of an individual, the metabolome may change as a result of lifestyle, stress and, most importantly, onset of pathologies. On the other hand,

[*] Dr. A. Vignoli[+]
    C.I.R.M.M.P.
    Via Luigi Sacconi 6 50019 Sesto Fiorentino, Florence (Italy)
    Dr. V. Ghini,[+] G. Meoni,[+] C. Licari, Prof. P. Turano, Prof. C. Luchinat
    CERM, University of Florence
    Via Luigi Sacconi 6 50019 Sesto Fiorentino, Florence (Italy)
    E-mail: luchinat@cerm.unifi.it

    Dr. P. G. Takis
    Giotto Biotech
    Via Luigi Sacconi 6 50019 Sesto Fiorentino, Florence (Italy)

    Dr. L. Tenori
    Department of Experimental and Clinical Medicine, University of
    Florence, Largo Brambilla 3, Florence (Italy)

    Prof. P. Turano, Prof. C. Luchinat
    Department of Chemistry "Ugo Schiff", University of Florence
    Via della Lastruccia 3–13 50019 Sesto Fiorentino, Florence (Italy)

[+] These authors contributed equally.

Figure 1. The flow of information in systems biology proceeds from the genome to the transcriptome, the proteome and finally to the metabolome. From left to right they are increasingly variable during an individual lifespan, and all concur to the phenotype.
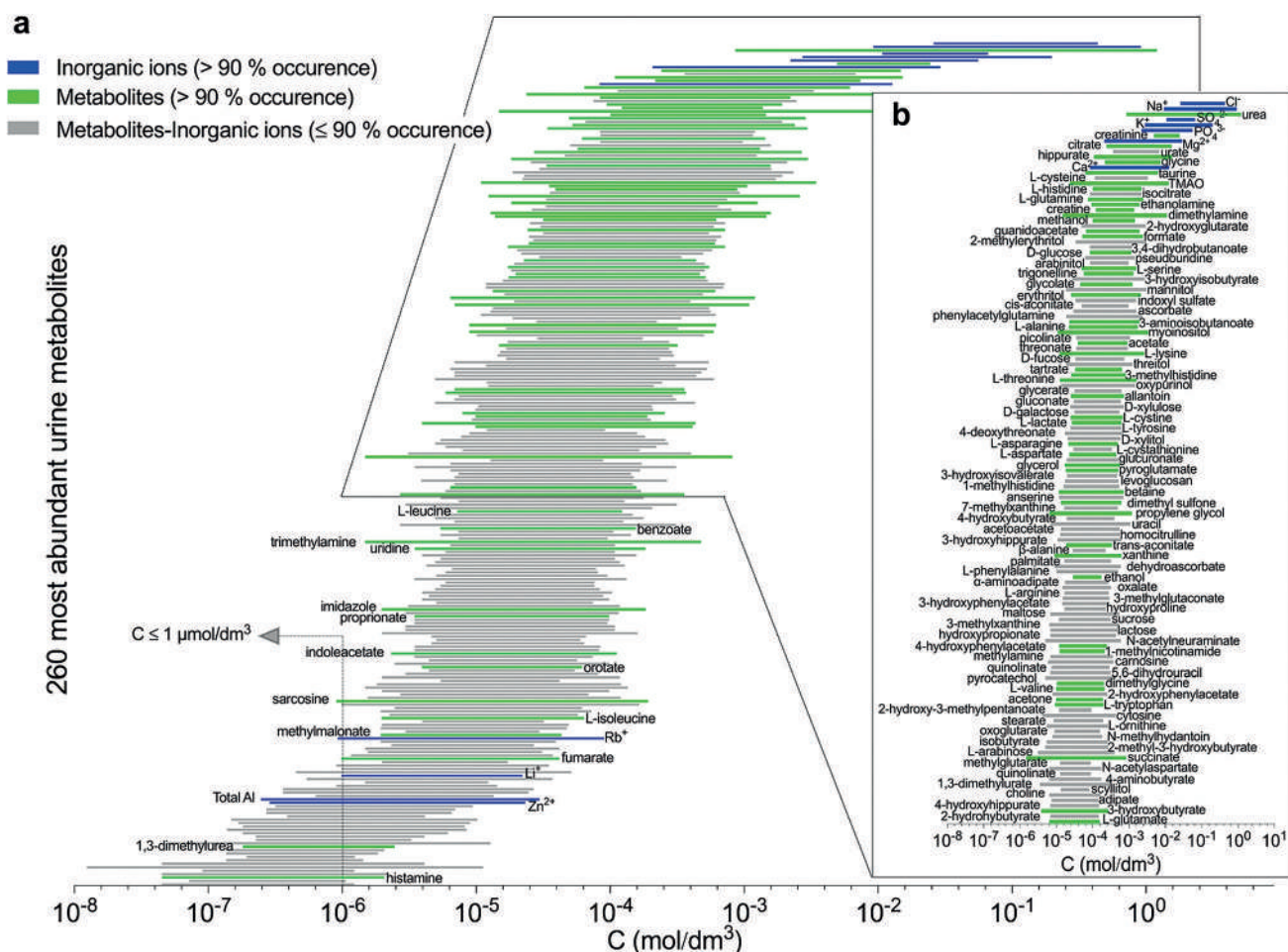
Figure 2. a) Concentration ranges of the 260 most abundant metabolites in urine as determined by LC-MS, NMR, GC-MS, ICP-MS and HPLC.[18] Metabolites are sorted according to their mean absolute concentration values in urine. Green and blue bars highlight the organic metabolites and the inorganic ions that appear at high occurrence in urine, respectively; gray bars those at lower occurrence.[18] b) Enlargement of the first 136 urine metabolites, with mean concentration value > 30 μM. Adapted from Ref. [14].

Claudio Luchinat is Professor of Chemistry at the University of Florence, co-founder and Director of the Center of Magnetic Resonance (CERM), and President of the Interuniversity Consortium on Magnetic Resonance of Metallo Proteins (CIRMMP). His research interests include bioinorganic chemistry, NMR-based structural biology, and NMR of paramagnetic species. In the last decade his research has also focused on NMR-based metabolomics, demonstrating the existence of individual metabolic fingerprints and of specific metabolic fingerprints of several diseases. The NMR metabolomics team from left to right: Leonardo Tenori, Gaia Meoni, Veronica Ghini, Claudio Luchinat, Paola Turano, Alessia Vignoli, Panteleimon G. Takis, Cristina Licari.

while the number of biological objects increases from 10^4–10^5 genes to ca. 10^7 proteins, it decreases down to 10^3–10^4 metabolites[1,2] (Figure 1), thus condensing the information. The distribution of metabolites in a biological fluid (urine) is shown as an example in Figure 2.

Although the majority of studies of metabolomics deal with humans, and therefore with human health and diseases, metabolomics also applies to animals,[3] plants[4] and microorganisms,[5,6] and therefore also impacts on other important fields such as agriculture, food production and environment.[7–9] There are very many review articles that, as it can be imagined, describe metabolomics from many different viewpoints.[10] The present one attempts at looking at metabolomics mainly from a chemical viewpoint. We are well aware that the whole is greater than the sum of all its parts, that is, understanding metabolomics implies a systemic view of the biological entity and cannot be limited to correctly identifying and quantifying as many chemical objects as possible in a system. On the other hand, expertise in analytical chemistry, medicinal chemistry, chemometrics, molecular reactivity, good laboratory practice, and in at least one of the two major instrumental metabolomic techniques (MS and NMR)

is essential to properly collect, handle and store the metabolomic samples, perform the experiments, analyze the results, and even detect flaws that might escape an eye that is not chemically trained. Metabolomics can bring to great discoveries, but we need to be sure that the discovery is based on solid ground.

Of the two main analytical techniques, MS and NMR,[11] we will only deal with NMR which is our main field of research, but we think that the two techniques are very complementary, and the weaknesses of one are compensated by the strengths of the other,[10,12,13] as illustrated in Table 1.

Table 1: Main features of NMR and MS applied to metabolomics.[12,13]

Technology	NMR	MS
Reproducibility	Very high	Fair
Detection limit	Micromolar range	Picomolar range
Sample preparation	Minimal	Several steps: often requires chromatographic separation and sample derivatization
Volume of the original sample used	0.1–0.5 mL	0.01–0.2 mL
Types of molecules detected	Any molecules containing NMR active nuclei	Most organic and some inorganic
Types of experiments	All metabolites above detection limit are observed simultaneously	Several, tailored for specific chemical species
Ambiguous/false identification	Origin: compounds with degenerate chemical shifts, chemical shift variability due to experimental conditions (pH, temperature, ionic strength), presence of only one singlet signal. Experimental approaches: 2D experiments[14]	Origin: compounds (e.g. isomers) that can match a given atomic composition or a parent ion mass.[15] Experimental approaches: LC-MS/MS

The main reason to stick to one of them is that, because the merits of the two techniques are intrinsically different, doing metabolomics by NMR results in developing an "NMR-vision" of metabolomics that is different from the "MS-vision". By doing this we hope to contribute to increase the "biodiversity" of metabolomic approaches, to the advantage of the scientific communities (the medical community above all) that are progressively attracted by it but still need to appreciate and exploit the fact that metabolomics comes in very different flavors.

2. Metabolomics by NMR

A major application of NMR in metabolomics is the detection of all small molecules in a sample that are above

a given concentration threshold (Table 1) using one-dimensional ^{1}H pulse sequences. Two main approaches are employed: i) solution NMR, for the detection of soluble metabolites in biofluids, cell lysates or polar/apolar tissue extracts and ii) HR-MAS (High Resolution Magic Angle Spinning), for the measurement of metabolites in semi-solid samples, like intact tissues.

2.1. Types of Samples

The types of samples are many; the most common and significant are listed in Table 2, Figures 3 and 4. The samples can be classified in terms of molecular/biochemical complexity, from cells to tissues and biofluids. The NMR detectable
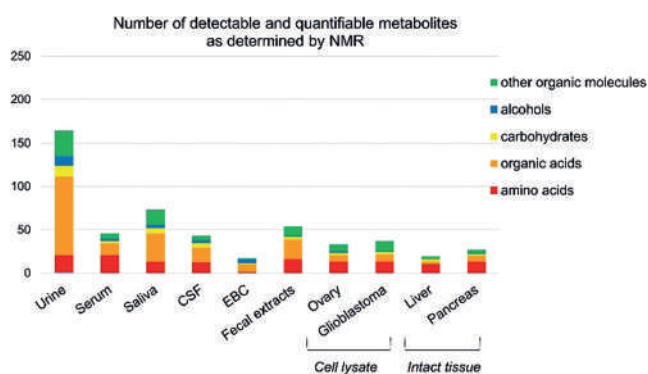

Figure 3. Number of detectable and quantifiable metabolites with $\geq 50\%$ occurrence in the ^{1}H NMR spectra of different biological fluids (urine,[18] serum,[19] saliva,[20] CSF,[21] EBC,[22] fecal extracts, cell lysates (e.g. ovary and glioblastoma cells)[23,24] and intact tissues (e.g. liver and pancreas).[25,26]

part of the metabolome corresponds to tens-hundreds of molecules, mainly belonging to the class of amino acids, carbohydrates, alcohols and organic acids (Figure 3). NMR can also contribute to the definition of the overall lipid composition of a biosystem, the so-called lipidomics.[16,17] Sample-specific considerations can be applied. Different types of cells are characterized by different metabolomes, but the measurable molecules are end-products or intermediates of the main metabolic pathways. Intensity variations in intracellular metabolites (endo-metabolome) induced by treatment with drugs, overexpression of a protein, genetic manipulation, etc. can be directly related to the up- or down-regulation of specific pathways. For a comprehensive picture of the biochemistry of cells, the complementary information provided by analyzing culture media (exo-metabolome) is extremely important.[27] The NMR spectra of tissues reflect organ-specific biochemistry (aerobic respiration, lipid metabolism, etc.) but are also affected by the heterogeneity of the tissue composition, the contribution from the extracellular matrix and the tissue microenvironment, so that their biochemical interpretation is more difficult than in cultured cells. Nevertheless, tissue samples are extremely valuable as direct reporters of the diseased organ, where variations in the metabolome with respect to a healthy status are expected to

Table 2: Type of samples, pre-analytical procedures, NMR sample preparation, NMR spectral acquisition.

Type of sample	Pre-analytical procedures[a]	Analytical procedures NMR sample preparation[b]	NMR spectral acquisition[c]
URINE	Collect the first urine of the morning after a minimum of 8 h fasting. Specify if collected in different daytimes or not fasting. Centrifuge the urine within 120 min after the collection at 1000–3000 RCF for 5 min at +4°C and/or filtrate the samples by 0.20 µm cut-off filter. Recover the urine in sterile condition making 1 mL aliquots. Store at −80°C.[29]	Thaw at room temperature and shake. Centrifuge at 14000 RCF for 5 min at 4°C.[30] Add 630 µl of sample to 70 µL of potassium phosphate buffer (1.5 M K_2HPO_4, 100% (v/v) ^2H_2O, 2 mM NaN_3, 5.8 mM TMSP; pH 7.4).[d] Transfer 600 µL of each mixture into a 5 mm NMR tube.[30]	[31]Recommended magnetic field: 600 MHz Recommended acquisition temperature: 300 K **1D NOESY-presat (Bruker: noesygppr1d)** Scans: 32 Data points: 65 536 Spectral width: 12 019.230 Hz Acquisition time: 2.73 s Relaxation delay: 4 s Mixing time: 0.01 s **2D J-RES (Bruker: jresgpprqf)** Scans: 2 Data points direct dimension (F2): 8192 Data points indirect dimension (F1): 40 Spectral width direct dimension (F2): 10026.738 Hz Spectral width indirect dimension (F2): 78.000 Hz Increment for delay: 12820.51 µs Acquisition time direct dimension (F2): 0.41 s Acquisition time indirect dimension (F1): 0.26 s Relaxation delay: 2 s
BLOOD	**SERUM** Collect blood into anticoagulant free tubes. Allow the blood to clot in an upright position for 30–60 min at RT. Spin centrifuge within 30 min from collection at 1500 RCF for 10 min at RT.[29] **PLASMA** Collect blood into tubes contain an anti-coagulant (preferred EDTA; avoid heparin). Centrifuge within 30 min from collection at 820 RCF for 10 min at 4°C.[29] For both serum and plasma: recover supernatant in sterile condition making 0.5 mL aliquots. Store at −80°C.[29]	Thaw at room temperature and shake. Add 350 µl of sample to 350 µL of sodium phosphate buffer (70 mM Na_2HPO_4; 20% (v/v) ^2H_2O, 6.1 mM NaN_3; 4.6 mM TMSP; pH 7.4).[d] Transfer 600 µL of each mixture into a 5 mm NMR tube.[30]	[31]Recommended magnetic field: 600 MHz Recommended acquisition temperature: 310 K **1D NOESY-presat (Bruker: noesygppr1d)** Scans: 32 Data points: 98 304 Spectral width: 18 028.846 Hz Acquisition time: 2.73 s Relaxation delay: 4 s Mixing time: 0.01 s **1D CPMG (Bruker: cpmgpr1d)** Scans: 32 Data points: 73 728 Spectral width: 12 019.230 Hz Acquisition time: 3.07 s Relaxation delay: 4 s Total spin-echo delay: 80 ms **1D Diffusion-edited (Bruker: ledbpgppr2s1d)** Scans: 32 Data points: 98 304 Spectral width: 18 028.846 Hz Acquisition time: 2.73 s Relaxation delay: 4 s Square gradient strength: 80% of the maximum gradient strength (53.5 G cm^{-1}) Square gradient length: 1.5 ms Diffusion time: 0.119 s **2D J-RES (Bruker: jresgpprqf)** For all the parameters see urine.

Table 2: (Continued)

Type of sample	Pre-analytical procedures[a]	Analytical procedures NMR sample preparation[b]	NMR spectral acquisition[c]
SALIVA	Collect saliva after refraining from eating, drinking, smoking, or using oral hygiene products for at least 1 h. Rinse the mouth twice with water before spitting the saliva in sterile tubes making 1 mL aliquots. Freeze asap at −20°C and then store in liquid nitrogen.[32]	Thaw at room temperature and shake. Centrifuge at 14000 RCF for 5 min at +4°C. Add 630 µl of sample to 70 µL of potassium phosphate buffer (1.5 M K$_2$HPO$_4$, 100% (v/v) ^{2}H$_2$O, 2 mM NaN$_3$, 5.8 mM TMSP; pH 7.4). Transfer 600 µL of each mixture into a 5 mm NMR tube.[33]	Recommended magnetic field: 600 MHz Recommended acquisition temperature: 300 K **1D NOESY-presat (Bruker: noesygppr1d)** Scans: 128 For all the other parameters see urine. **2D J-RES (Bruker: jresgpprqf)** For all the parameters see urine.
CSF	Collect CSF via lumbar puncture in sterile polypropylene tubes. Centrifuge asap at 2000 RFC for 10 min at RT. Collect the supernatant in sterile condition making 0.5 mL aliquots. Store at −80°C.[34]	Thaw at room temperature and shake. Add 750 µL of sample to 150 µL of potassium phosphate buffer (0.9 M K$_2$HPO$_4$, 60% (v/v) ^{2}H$_2$O, 1.2 mM NaN$_3$, 3.5 mM TMSP; pH 7.4).[d] Transfer 600 µL of each mixture into a 5 mm NMR tube.	Recommended magnetic field: 600 MHz Recommended acquisition temperature: 300 K **1D NOESY-presat (Bruker: noesygppr1d)** Scans: 256 For all the other parameters see blood. **2D J-RES (Bruker: jresgpprqf)** For all the parameters see urine.
EBC	Collect EBC after refraining from eating for at least 3 h using a condenser equipped with a saliva trap. Rinse the mouth twice with water before breathing through a mouthpiece for 15 min making a 1.5 mL aliquot. Transfer the EBC into glass vials closed with 20 mm butyl rubber lined with poly-tetrafluoroethylene septa and crimped with perforated aluminum seals. Before sealing, remove volatile substances by a gentle nitrogen gas flow for 3 min. Freeze asap in liquid nitrogen and then store at −80°C.[35]	Thaw at room temperature and shake. Centrifuge at 14000 RCF for 5 min at +4°C. Add 540 µl of sample to 60 µL of ^{2}H$_2$O containing 5.8 mM TMSP. Transfer 600 µL of each mixture into a 5 mm NMR tube.	Recommended magnetic field: 600 MHz Recommended acquisition temperature: 300 K **1D NOESY-presat (Bruker: noesygppr1d)** Scans: 512 For all the other parameters see urine. **2D J-RES (Bruker: jresgpprqf)** For all the parameters see urine.
FECES	Collect feces in sterile containers. Add 2 mL of PBS/^{2}H$_2$O buffer (0.1 M, pH 7.4) to 1 g of each feces sample and homogenize the mixture by vortexing for 60 s.[36] Sonicate for 30 min and centrifuge at 10000 RCF for 5 min at RT. Collect the supernatant in sterile condition making 1 mL aliquots. Store at −80°C storage.[37]	Thaw at room temperature and shake. Centrifuge at 14000 RCF for 5 min at +4°C. Add 630 µl of sample to 70 µL of potassium phosphate buffer (1.5 M K$_2$HPO$_4$, 100% (v/v) ^{2}H$_2$O, 2 mM NaN$_3$, 5.8 mM TMSP; pH 7.4). Transfer 600 µL of each mixture into a 5 mm NMR tube.	Recommended magnetic field: 600 MHz Recommended acquisition temperature: 300 K **1D NOESY-presat (Bruker: noesygppr1 d)** Scans: 128–256 For all the other parameters see urine. **2D J-RES (Bruker: jresgpprqf)** For all the parameters see urine.

be most evident. The latter feature is shared with some compartmentalized biofluids; for example, cerebrospinal fluid (CSF) reflects the biochemistry of the central nervous systems, and exhaled breath condensate (EBC) that of respiratory pathways. The saliva metabolome not only changes in correlation with oral disorders but also in the presence of distant pathologies.[28] An important contribution to the metabolome of fecal extracts originates from metabolites resulting from the gut microbiota.

Systemic biofluids, such as urine or blood serum and plasma, in general have a fainter biochemical correlation with a diseased organ or apparatus, but present two main advantages: the simple, noninvasive or minimally invasive collection, and the ability to reflect the overall response of the individual to a disease status. On the other hand, urine and blood are largely different in terms of chemical composition, with urine metabolites being heavily influenced by environmental factors such as food and liquid intake, while blood has a better defined and stable metabolome.

From the above considerations it is clear that preserving the chemical composition of the in vivo metabolome, and ensuring spectral reproducibility, are key factors for the significance of metabolomic analysis. In general, the entire workflow can be subdivided into two main phases, the pre-

Table 2: (Continued)

Type of sample	Pre-analytical procedures[a]	Analytical procedures NMR sample preparation[b]	NMR spectral acquisition[c]
CELLS	**ENDO-METABOLOME** Place cell plates into ice and rinse with PBS. Scrape and collect cells with PBS supplemented with 1% Protease Inhibitor Cocktail and 1% Phosphatase Inhibitor Cocktail. Sonicate and ultra-centrifugate for 1 h at 4°C. Collect the supernatant in sterile condition making 0.6 mL aliquots. Store at −80°C freezer.[23]	Thaw in ice and shake. Add 540 µl of sample to 60 µL of ^2H_2O containing 5.8 mM TMSP. Transfer 600 µL of each mixture into a 5 mm NMR tube.[23]	Recommended magnetic field: 900 MHz Recommended acquisition temperature: 300 K **1D NOESY-presat** **(Bruker: noesygppr1d)** Scans: 128–256 Data points: 110060 Spectral width: 17942.584 Hz Acquisition time: 3.07 s Relaxation delay: 4 s Mixing time: 0.01 s **1D CPMG** **(Bruker: cpmgpr1d)** Scans: 128–256 Data points: 110060 Spectral width: 17942.584 Hz Acquisition time: 3.07 s Relaxation delay: 4 s Total spin-echo delay: 80 ms
	EXO-METABOLOME Removal of the growing medium. Collect the medium in sterile condition making 1 mL aliquots. Store at −80°C.	Thaw at room temperature and shake. Add 350 µl of sample to 350 µL of sodium phosphate buffer (70 mM Na_2HPO_4; 20% (v/v) ^2H_2O, 6.1 mM NaN_3; 4.6 mM TMSP; pH 7.4). Transfer 600 µL of each mixture into a 5 mm NMR tube.	Recommended magnetic field: 900 MHz Recommended acquisition temperature: 310 K **1D NOESY-presat** **(Bruker: noesygppr1d)** Scans: 32–64 Data points: 110060 Spectral width: 17942.584 Hz Acquisition time: 3.07 s Relaxation delay: 4 s Mixing time: 0.01 s
TISSUES	Collect tissues in the most aseptic conditions possible. Cut tissue samples ≤0.5 cm in any single dimension. Snap freeze in liquid nitrogen within 30 min from collection. Stored in liquid nitrogen vapor.	Trim frozen tissue samples (10–15 mg) to fit HR-MAS ZrO_2 rotor insert capacity. Fill the insert with ^2H_2O containing 5.8 mM TMSP. Cover the rotor inserts with plug and plug-restraining screw and insert it into the 4 mm rotor for HR-MAS.[38]	[31,38]Recommended magnetic field: HR-MAS 600 MHz Recommended acquisition temperature: 277 K **1D ZGPR** Scans: 128–256 Data points: 32768 Spectral width: 12019.230 Hz Acquisition time: 1.36 s Relaxation delay: 2 s Rotor speed: 4 MHz **1D CPMG** **(Bruker: cpmgpr1d)** Scans: 128–256 Data points: 32768 Spectral width: 12019.230 Hz Acquisition time: 3.07 s Relaxation delay: 1.36 s Total spin-echo delay: 94 ms Rotor speed: 4 MHz

[a] It is extremely important to: 1) minimize as much as possible the time between sample collection, processing and storage keeping the samples at 4°C, if not differently specified; 2) use additive-free tubes and laboratory materials to avoid sample contamination. [b] Minimize as much as possible the time between sample preparation and NMR spectral acquisition. [c] 1) Acquire NMR spectra in automatic mode using a refrigerated (4–6°C) sample changer. 2) Acquire NMR spectra of samples in a totally random way, interleaving the samples of different groups to avoid batch effects. [d] According to Bruker guidelines.

analytical and the analytical phases. Both require specific procedures for accurate downstream analyses but are characterized by different degrees of development.

2.1.1. SOPs for Pre-Analytical Treatment

Molecular analyses of biological samples are necessarily preceded by the so-called pre-analytical phase, which includes
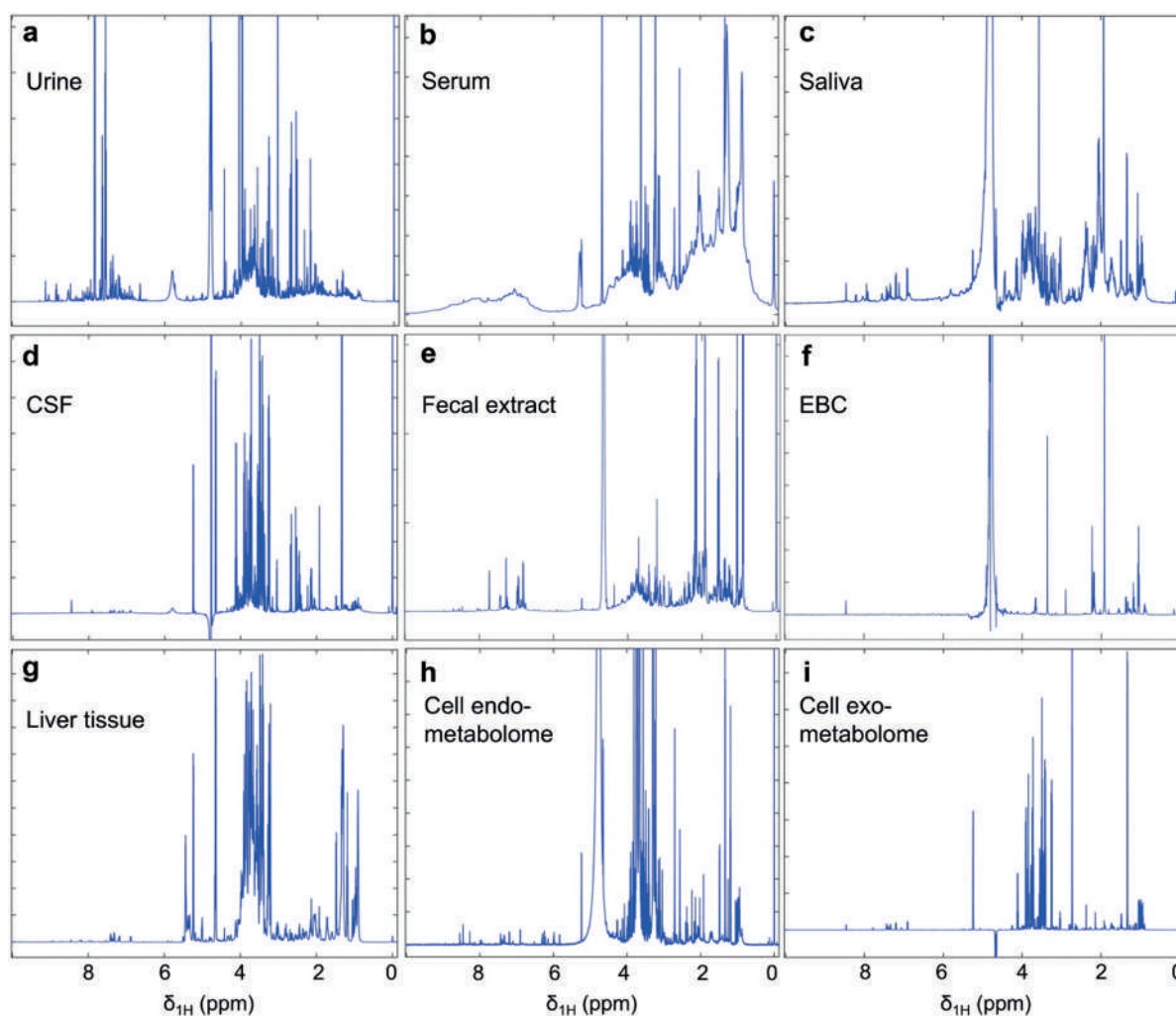
Figure 4. ^{1}H NMR spectra of: a) urine, b) serum, c) saliva, d) CSF, e) fecal extract, f) EBC, g) liver tissue, h) cell lysate (endo-metabolome) and i) cell media (exo-metabolome). The NMR spectra are recorded at 600 MHz, except for cell lysates and media (900 MHz).

several steps such as primary sample collection, processing, transport, and storage. The impact of non-appropriate pre-analytical procedures on downstream metabolomic analyses can be heavy and is one of the main reasons that makes it difficult to compare metabolomic data collected in multi-center studies. Some of the molecules forming the metabolome are indeed very sensitive to sample conditions, and their levels can change drastically from collection to analyses. It is therefore important to develop simple and validated standard operating procedures (SOPs) for each type of sample to be strictly followed to ensure that the subsequent assay will determine the metabolome in the original sample and not an artificial profile generated during the pre-examination process.

The development of validated procedures requires the systematic simulation of different pre-analytical situations. In this way one can identify the critical steps and parameters that may influence the levels of the metabolites that are prone to degradation (with consequent accumulation of their degradation products). Systematic studies exist for some of the most common samples;[39,40] they have revealed enzymatic activities in the samples as main sources of variations. For

blood derivatives, the presence of cells during sample processing heavily affects the concentration of glucose and lactate; therefore, serum and plasma harvesting should be initiated within 30 min from blood collection. Residual cellular activity is a problem also in urine; mild centrifugation and/or filtration upon sample collection are therefore required. While maintaining samples at low temperature throughout the pre-analytical process is obviously an efficient strategy to slow down degradation reactions, one should avoid sample freezing before removal of cellular components, to prevent cell breaking and consequent release of enzymes into the biofluid.[41] In tissues, it is obviously impossible to preclude cellular/enzymatic activities, and the best strategy consists in flash freezing samples in liquid nitrogen upon collection. This treatment reduces to a minimum the detrimental effects of the so-called post-resection or cold ischemia. More difficult to avoid is the impact of the intraoperative warm ischemia, that also alters the levels of metabolites associated to oxidative stress and apoptosis.[42] If warm ischemia times can hardly be standardized during surgical procedures, their accurate annotation is recommended. Finally, particular attention should be devoted in the selection

of sample collection vials that should not contain stabilizers or other contaminants that can interfere with the NMR analysis, giving rise to detectable NMR signals.[43]

A noteworthy example of activities focused on pre-analytics was conducted within the FP7 project SPIDIA (standardization and improvement of generic pre-analytical tools and procedures for in vitro diagnostics), which has led to the first example of technical specifications for pre-examination processes in metabolomics.[29]

2.1.2. SOPs for Analytical Treatment and NMR Acquisition

The analytical step initiates with the NMR sample preparation, usually starting from cryopreserved samples. To ensure intra- and inter-laboratory comparability, several efforts have been made to develop standardized procedures for i) NMR sample preparation and ii) spectral acquisition. The former is designed to obtain NMR-quality samples; buffering is required to easily reference chemical shift values to existing databases; dilution has to be defined for subsequent absolute quantification. The latter relies on instrumental optimization, NMR pulse sequence selection, and choice of acquisition parameters. Instrument producers have been very active in the field. The selection of a "recommended" magnetic field facilitates comparison of spectra acquired in different laboratories and with available spectral databases (see Section 3.3.2). The use of 600 MHz spectrometers represents the best compromise between a good spectral sensitivity and resolution and affordable instrumental cost, and it is therefore considered the standard field for biofluid and tissue analyses. The economical aspect is essential for a technique that aims at being translated into clinical practices. Higher fields are often used to achieve better performances, especially with non-clinical samples, such as cell cultures. Instrumental optimization includes an ample range of aspects: probe design, sample changer and software tools for automatic acquisition (temperature control, optimization of shim, water suppression, etc.) of about hundred samples. In most biofluids, low mass metabolites coexist with high mass biomolecules, such as lipids, proteins, and lipoproteins. Three NMR pulse sequences are used in these cases to selectively observe the different components: i) the nuclear Overhauser effect spectroscopy (NOESY) pulse sequence yields a spectrum in which both signals of metabolites and high molecular weight molecules are visible; the Carr-Purcell–Meiboom–Gill (CPMG) pulse sequence enables the selective observation of small molecule components in solutions containing macromolecules (via T_2 filtering); the DIFFUSION-EDITED sequence permits the selective observation of macromolecular components in solutions containing small molecules. As summarized in Table 2, for several sample types, it is a common practice to acquire all three spectra. In defining the total acquisition time, one shall consider the stability of the sample under the selected experimental conditions. It should be pointed out that some authors prefer to physically remove macromolecular components via centrifugation with 3000 MWCO Amicon Ultra-0.5 filters followed by NOESY acquisition rather than relying of CMPG filtering.[44,45] The recording temperature is clearly an

issue that, in the case of HR-MAS, sums up to destructive friction/centrifugation effects induced by spinning the sample in the rotor. Thus, the selection of recycling times between different scans is again a compromise between magnetization recovery (ideally $5 \times T_1$ of the slowest relaxing signals) and acquisition of a number of scans sufficient for good S/N before sample degradation. Typical spectra in Table 2 require 4–32 min acquisition. Refrigerated automatic sample changers, where NMR tubes are kept at 6°C until acquisition, have been developed to facilitate the automatic acquisition of large sets of samples. Along the lines of a general standardization of the entire process involving metabolomic studies, there are ongoing efforts in the community to achieve open data standards and accessible repositories that allow researchers to store, exchange, and compare metabolomic data with pertinent metadata information.[39,46] For example, MetaboLights (https://www.ebi.ac.uk/metabolights/) or Metabolomics Workbench (http://www.metabolomicsworkbench.org) are repositories of metabolomic experiments and derived information.

3. Analysis of the Spectra

3.1. Processing

In the metabolomic work-flow, the data processing step follows the acquisition of the raw spectra and is used to transform the data in a form suitable for subsequent statistical analyses (Figure 5). Each NMR spectrum must be properly adjusted for phase and baseline. Both operations are usually performed automatically; manual adjustment is discouraged for metabolomic data, because it can introduce operator-biased artefacts. The spectra need also to be properly aligned to a reference signal, using a chemical shift standard, ideally not interacting with any sample component (for instance deuterated trimethylsilylpropanoic acid (TMSP) may bind macromolecules), thus in samples like serum/plasma, tissues and cell extracts an alternative internal reference is preferred (e.g. the anomeric signal of glucose).[47] For the same reason, use of TMSP as a standard for quantification is discouraged. For the absolute quantification of metabolites, alternative approaches have been proposed, such as the production of an artificial NMR signal based upon PULCON method[48] or the ERETIC method.[49]

3.2. Normalization for Different Types of Spectra/Samples

In order to compare signal—or bucket—intensities in different samples, one should refer to the same amount of total sample. In biofluids this seems an obvious task, as one prepares the NMR sample starting from the same volume of the original fluid (Table 2). In practice, a preliminary step of correction for dilution effects is needed due to the presence of large physiological variations in concentration. For examples, urines exhibit significant metabolite concentration fluctuations depending on the hydration state of the individual (in turn related to water/food intake, physical exercise and
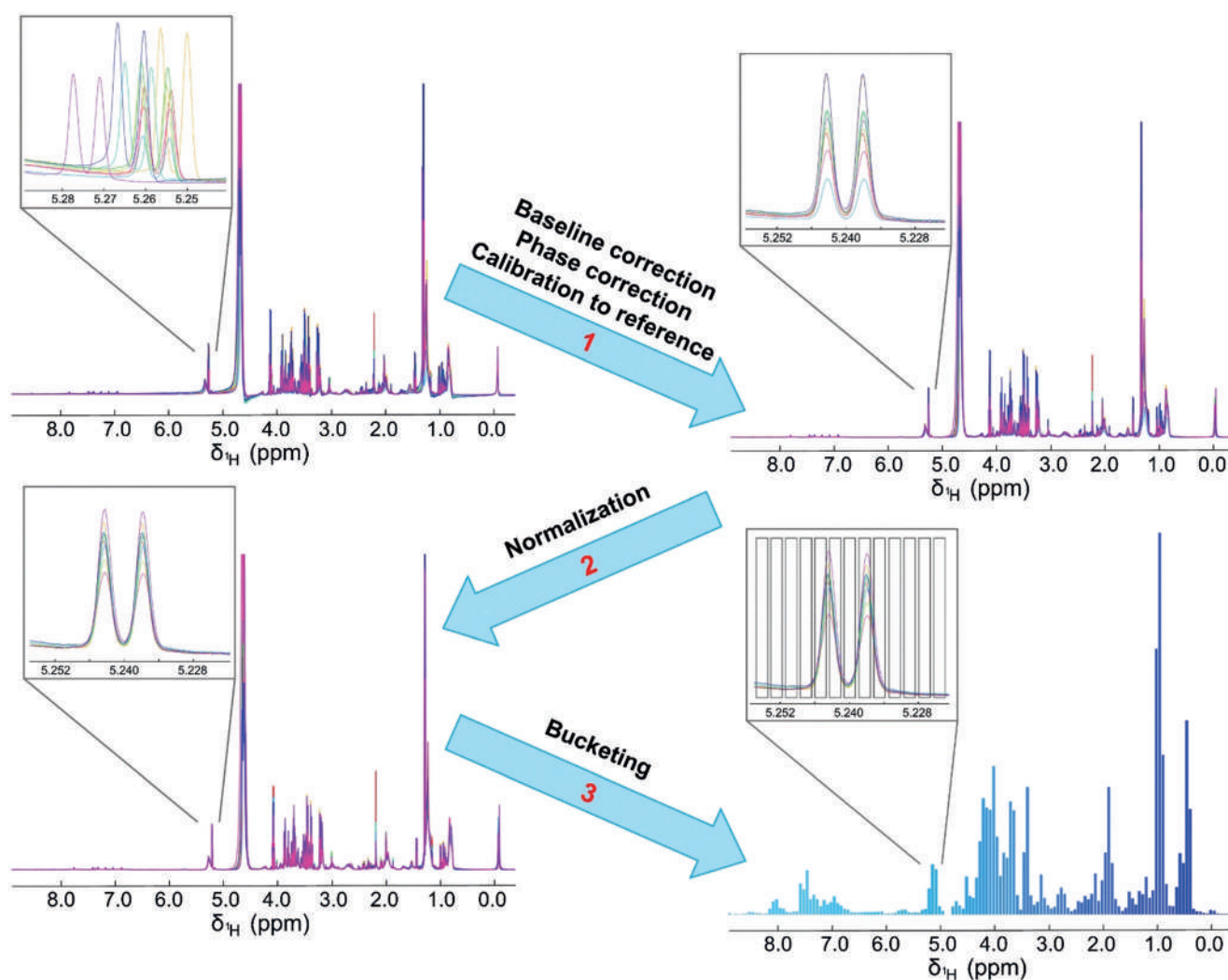
Figure 5. Key stages of NMR spectral processing: 1) baseline correction, phase correction, calibration to an internal reference peak (e.g. glucose, TMSP, etc.); 2) normalization; 3) bucketing.

sweating, etc.),[50] therefore, a global intensity correction by means of normalization is a critical initial step. For biofluids under stricter physiological control (like serum/plasma) normalization is less crucial, but could be beneficial to compensate for non-physiological sources of variations emerging from (small) experimental inaccuracies and/or technical artifacts.[51] Total area normalization is a common practice but there are more sophisticated and reliable methods. In the quest for the "optimal" normalization method, a plethora of strategies/algorithms have been developed. Table S1 (Supporting information) shows and briefly describes 23 state-of-the-art methods belonging to five main categories, including those not explicitly developed for metabolomics but borrowed from other fields. Although several comparisons are reported in the literature, a definitive consensus on which normalization method to use in which case is still lacking. For biofluids, Probabilistic Quotient Normalization (PQN),[52] Quantile Normalization,[53] Cubic Splines Normalization (CSN),[54] and Pairwise Log-Ratios (PLR),[55] are generally described as good choices[50,51,55,56] that outperform total area normalization. For instance,

PQN has been successfully used with EBC,[57,58] saliva,[33] and urine[59] samples. For tissues, cell lysates or cell/tissue extracts, exact quantification of the starting material is not straightforward. For cells, it can be assumed that an increase of cell number produces a linear increase of metabolite signal intensities, thus data normalization is usually performed to the number of cells.[60,61] Alternatively, one could refer to the total DNA or total protein content.[61] On the same line, for tissues, normalization according to the sample weight is recommended.[62] If this information is not available, total area normalization represents a good method, as the total spectral area can be considered proportional to the number of cells or the weight of tissues.

3.3. Different Experimental Approaches in Metabolomics

Depending on the biological question at issue, metabolomic analysis can be planned with two different methodological approaches: targeted and untargeted analyses (Figure 6). The targeted approach involves the monitoring of a panel of
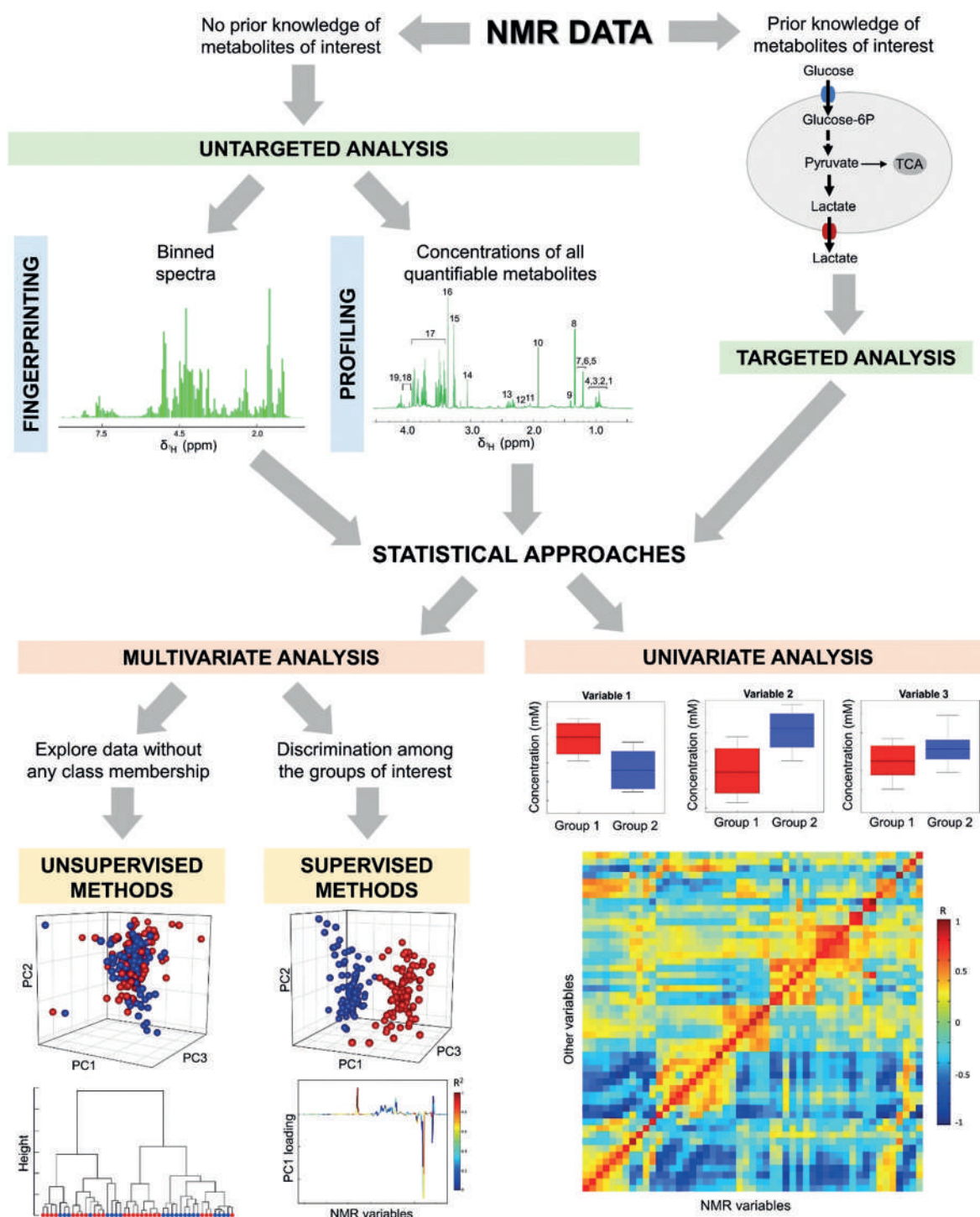
Figure 6. Different metabolomic strategies. If knowledge of the metabolites or metabolic pathways of interest is available, a targeted approach, which involves the analysis of only these specific metabolites, is the preferred choice. In the absence of prior knowledge, the problem can be addressed by analyzing the spectrum, with a so-called untargeted approach. Untargeted analysis can be achieved via metabolic fingerprinting or profiling. The former is a global evaluation of all of the features of a binned spectrum without identification of single metabolites; the latter deals with the analysis of all quantifiable metabolites. Each of these three different sets of data can be addressed by either multivariate or univariate analyses. Multivariate methods are routinely used to visualize biological data, to identify possible clusters, and to build predictive models; these can be divided into two main categories: unsupervised analyses to explore data without any class membership and supervised analyses to discriminate among known groups of interest. Conversely, univariate methods are used to identify metabolites and thus metabolic pathways that are altered or correlated with specific biological conditions.

metabolites selected a priori on the basis of known metabolic pathways or pre-identified biomarkers that are undoubtedly associated with the disease or condition of interest.[11] These selected metabolites must be unambiguously assigned and

quantified in the samples. Conversely, untargeted metabolomics provides a global view of a sample by analyzing all (or as many as possible) measurable analytes present, including unknown chemicals.[63] The latter purpose can be achieved via two strategies: fingerprinting or profiling. Fingerprinting is a global, rapid evaluation of an NMR spectrum as a whole that can be considered a "fingerprint" of all (assigned or unassigned) detectable metabolites present in that biological sample.[64] This can be achieved only by transforming NMR spectra in matrices of data for example, through bucketing, that is, a procedure used to reduce the total number of variables and to compensate for small misalignments in the spectra. One bucket (or bin) is a little portion of the spectrum, usually with a width of 0.02 or 0.04 ppm. The integrated spectral intensity within each bin is then calculated to obtain the variables used for feeding the statistical algorithms. Although many more sophisticated algorithms for bucketing exist, this equidistant binning is the most commonly used method[65] and often works quite well despite its simplicity.[66] Alternatively, some practitioners prefer to work with full resolution spectra, totally avoiding binning. In such cases, specific algorithms[67] for peak alignment are needed (e.g. the very efficient *i*coshift[68]), to assure that peaks can be compared across multiple spectra. The fingerprinting approach is essentially utilized to provide sample classification. Indeed, metabolic fingerprinting aims at discriminating specimens in relation to different biological conditions (e.g. presence/absence of disease, before/after treatment), in turn characterizing a specific health state with a unique metabolic pattern.[11] Metabolic profiling, in contrast, deals with the determination of the concentrations of all quantifiable metabolites in a biological sample. Profiling provides considerably more meaningful data from a biochemical perspective since it also enables the identification of metabolites and metabolic pathways associated with a specific physiological or pathological condition. However, it is important to mention that the spectral processing required to deconvolute 1D NMR spectra, in order to obtain concentrations, may not be straightforward and is not yet completely automated. The molecules quantifiable via profiling are significantly less numerous than those contributing to the fingerprint (in the case of urine $< 50\%$).[64] The latter, therefore, is the best tool for sample classification and to build statistical models.

3.3.1. Multivariate Analysis for Classification and Modeling

Starting from the available data (full or binned NMR spectra, or a list of concentrations) arranged in rows (samples) and columns (variables) in a data matrix, the principal aims of metabolomics can be summarized in four goals: i) visualize the overall differences, trends, relationships, and correlations among different samples; ii) detect whether there is a significant difference between investigated groups (e.g., healthy vs. diseased subjects); iii) highlight the spectral regions mostly contributing to these differences and, iv) construct a predictive model for the correct classification of new samples.[69] Multivariate statistics is the key to achieve these goals, by either i) unsupervised or ii) supervised methods. Unsupervised methods are utilized to summarize, explore, and

discover clusters or trends in the data unlabeled with any class membership; therefore, no prior assumptions or knowledge of the data are needed.[70] Unsupervised methods usually represent the first step in data analysis, helping to visualize the data and to discover possible outliers. One effective class of exploratory approaches involve the projection of the data in a new space using just few dimensions (data reduction). This includes classical methods, such as Principal Component Analysis (PCA),[71–74] and Independent Component Analysis (ICA),[75,76] as well as promising new ones developed for metabolomics, such as Group-Wise Principal Component Analysis (GPCA).[77] Another interesting class of approaches is data clustering, which consists in the assignment of a set of samples into subsets (so-called clusters) so that samples belonging to the same cluster are similar in some sense (i.e. they share similar metabolic features). Classical clustering methods, such as K-Means (KM)[78–80] and Partition Around Medoids (PAM),[81] have played a historical role in the data mining field. More recent methods include Spectral Clustering (SC)[82,83] and KODAMA.[84,85] Another option to obtain visually interpretable maps that capture inherent relationships among observables are the Self-Organizing Maps (SOM),[86,87] developed in the neural networks field.

In contrast, supervised approaches use a priori knowledge to generate models that are tightly focused on the effects of interest. The goal of supervised data analysis is to find a rule to extend the knowledge already available from pre-existing samples to new samples, so that predictions can be made (i.e. to classify normal vs. abnormal samples for diagnosis and/or prognosis). In doing so, the characteristics of the training set (data already available) are defined, "learned" by the statistical method, and applied to predict the test set (the unlabeled samples to be classified). Supervised analysis includes methods based on projection and data reduction, such as Partial Least Squares (PLS)[88–91] and its variant Orthogonal Partial Least Squares (OPLS);[92] methods based on machine learning, such as K-Nearest Neighbors (K-NN);[93,94] and methods based on neural networks architectures, such as the potentially ground-breaking paradigm of deep learning.[95,96] An overview of the main multivariate statistical techniques employed in metabolomics is reported in Table S2. It is worth noting that all the multivariate algorithms can be used on full resolution spectra, on the bin data matrix, or a panel of quantified metabolites, depending on the chosen approach.

To demonstrate the performance of a predictive model, proper validation of the results is performed using cross-validation strategies,[97] or an external and completely independent dataset. Moreover, a permutation test is often advocated to calculate the statistical significance of the results. Several measures of the model quality exist, such as those derived from the confusion matrix (i.e. sensitivity, specificity, and accuracy). The use of independent training and validation sets is the preferred approach in the medical community, but requires collecting samples from different hospitals, even from different countries. In this way, weaknesses due to for example, not perfectly identical SOPs (see Section 2) can become apparent. Conversely, artificially dividing in two groups samples collected under identical

conditions is not affected by this type of problems. Thus, the selection of the best approach strongly depends on the aim of the research: cross-validation on a unique collection of samples should be recommended for exploratory studies, where the aim is to see whether a fingerprint for a specific condition exists. Independent validation is then recommended to further demonstrate that the fingerprint is still strong also in the presence of "environmental noise". Immediately jumping to this second strategy may lead to conclude that a fingerprint does not exists while it may be only obscured by sub-optimal procedures that could be easily fixed.

3.3.2. Identification, Quantification, Univariate Analysis

The process of metabolite identification in NMR spectra, especially for less common biospecimens, is not straightforward due to the high complexity of their ^{1}H-NMR spectra. Usually many resonances can be directly assigned in one-dimensional spectra based on chemical shifts and multiplicity. This task is facilitated by databases, often freely available: in particular, for human metabolites, the Human Metabolome Database (HMDB)[98–100] is becoming the de-facto standard reference database. For doubtful cases, the addition of standard molecules (spiking) to the NMR sample could be of help. However, two-dimensional spectra are often required to assign new metabolites, as briefly summarized below (and reviewed in more details in Ref. [101]): i) [1]H-[1]H J-resolved (J-RES), to provide information about multiplicity and coupling patterns, ii) [1]H-[1]H Correlation Spectroscopy (COSY) and [1]H-[1]H Total Correlation Spectroscopy (TOCSY), to provide, respectively, short and long range scalar connectivities, iii) natural-abundance heteronuclear experiments that use ^{1}H in combination with other nuclei, for example, [13]C-[1]H Heteronuclear Single Quantum Coherence Spectroscopy (HSQC), Heteronuclear Multiple-Quantum Correlation (HMQC) and Heteronuclear Multiple-Bond Correlation Spectroscopy (HMBC), to obtain information on the direct scalar coupling between ^{13}C and ^{1}H nuclei.[101] The quest for a completely automatic assignment tool is an active research field, with promising achievements (Figure 7, see also Section 7).[14] Of course, as shown in Table 1, NMR and MS are two complementary approaches and combined MS/NMR can be used to correctly identify a wide range of metabolites.[102,103]

Further, while NMR spectroscopy is an intrinsically quantitative technique (signals are proportional to the concentration of nuclei), the ability to accurately and reproducibly quantify signals from metabolites in complex mixtures is complicated by spectral crowding and signal overlapping, as well as from raising of the baseline caused by the presence of large molecular mass components. In order to solve the latter problem, a sample treatment has been proposed for serum/plasma, that includes a centrifugation step at the level of sample preparation to remove large molecules.[44] In our view, any sample manipulation represents a potential source of experimental errors; therefore, we prefer broad signal suppression by CMPG. Nevertheless, a careful estimation of the effect of spectral acquisition parameters on signal

Figure 7. The automated assignment of several ^{1}H NMR signals of > 60 urine metabolites, performed by the urine shift predictor.[14] The urine spectrum is divided into 3 regions: a) 9.5–5.5 ppm, b) 4.5–3.0 ppm and c) 3.0–0.9 ppm.

intensity for the different metabolites shall be conducted. Spectral crowding and signal overlap are unavoidable in the absence of chemical separation.

Tools for (semi) automatic quantitation have been developed. Among them there are BATMAN,[104] BAYESIL,[44] ASICS[105] and the NMR Suite Software Package (Chenomx Inc., Edmonton, Canada). The first three are mostly automated computational tools based on Bayesian inference (BATMAN and BAYESIL) or linear models (ASICS). ASICS is freely available. BATMAN is also freely available, it performs quite well, but it is computationally very demanding. BAYESIL is faster, but it is commercial and requires a dedicated kit for the preparation of the samples.

The B.I. (Bruker IVDr) platform for analysis and quantification of metabolites in biofluids (urine, CSF, serum/plasma) has been recently released by Bruker BioSpin.

NMR Suite is another commercial software. It is a complete computer-assisted tool for analysis and deconvolution of the spectra, permitting the user-guided fitting, integration, and quantitation of the selected peaks. However, being mostly manual, it is time consuming and it needs a skilled operator. As witnessed by the total number of citations, BATMAN, BAYESIL and NMR Suite (61, 50, 90 Web of Science citations, respectively) are the most commonly used packages by practitioners.

In order to find a metabolite (or a panel of metabolites) that can be considered as a possible biomarker for the specific pathology or condition under investigation, each metabolite needs to be analyzed independently of all the others without considering any possible interactions. Univariate statistical analysis is the key to achieve this goal: statistical tests, correlation analysis and Receiver Operating Characteristic (ROC) curves are the most used approaches.

On the biological assumption that metabolite concentrations are not normally distributed, non-parametric tests are often utilized. The Wilcoxon–Mann–Whitney[106] test is the non-parametric analogue of the classical t-test to compare the distribution of a metabolite in two groups. The null hypothesis (H0) claims that two randomly selected samples from two populations actually belong to the same population; the alternative hypothesis (Ha) states that H0 is false and the two population are distinct. A variant of this test also exists, called the Wilcoxon signed ranks test,[107] that is suited to compare two matched samples from a repeated measurement in the presence of a paired study design. When the groups are more than two, Kruskal–Wallis test[108] (the analogue of the parametric Analysis of Variance) or Friedman test[109] (for paired samples) are employed. A summary of the main statistical tests for univariate analysis of metabolites is reported in Table S3.

The output of all these tests is a *P*-value, that expresses the probability of obtaining a result equal to or more extreme than what was actually observed, assuming the null hypothesis true. Conventionally, when the *P*-value is less than the significance level of 0.05 or 0.01 the null hypothesis is rejected, and the metabolite is deemed statistically different in the groups of interest. When several metabolites are tested together, to avoid random false positives, multiple testing corrections need to be adopted: Bonferroni[110] and Benjamini–Hochberg[111] are the most widespread methods. However, *P*-values need to be taken with care: they are often misused, misunderstood and misinterpreted;[112] for this reason statisticians have proposed replacing (or accompanying) *P*-values with effect size[113] (a standardized measure of the magnitude of the observed phenomenon) as an alternative (complementary) measure of evidence.

ROC curves are graphical representations that illustrate the diagnostic ability of a binary classifier (e.g. the concentration of a biomarker to pinpoint a disease or the efficacy of a composite prognostic score to catch the risk of developing a disease). The curve is created by plotting sensitivity versus one minus specificity for all possible thresholds of the test.

Accuracy is measured by the area under the ROC curve (AUC). An area of 1 represents a perfect test; an area of 0.5 represents a worthless test. ROC analysis could also be employed to check the performance of a multivariate supervised classifier. In this case, the cross-validated output of the classifier (a continuous response vector) is used as the input for the calculation of the ROC curve. Again, an AUC near 1 indicates a good classifier. ROC analysis is of particular importance because it provides a simple tool to select optimal models and to discard suboptimal ones, thus permitting a direct cost/benefit analysis of a diagnostic test.

Pearson correlations can be calculated to test whether there is an association (linear dependence), between metabolites and clinical data or biological features. Correlations are expressed by a coefficient (R) which ranges between $+1$ (totally correlated), 0 (no correlation) and -1 (totally anticorrelated). From metabolite to metabolite correlation maps, metabolic networks with a biological meaning can be inferred.[114–117]

4. Biomedical and Biochemical Significance

Metabolomics is focused on the analysis of intermediates and end products of metabolism in the form of endogenous (gene-derived metabolites), exogenous (environmentally derived metabolites) and gut microbiota-derived metabolites. For this reason, metabolites, unlike genes and proteins, are easier to correlate with the phenotype and act as direct signatures of biochemical activity, since they play a central role in disease development, cellular signaling and physiological control.[118] Metabolomics monitors the global outcome of all exogenous and endogenous factors, without making assumptions about the effect of any single contribution to that outcome. This makes metabolomics a perfect instrument to investigate and understand the molecular mechanisms of human health and disease.[1] Metabolomics can contribute to precision medicine for the comprehension of individual susceptibility to drug administration, nutrition and life style interventions, influence of environmental factors, as well as to the characterization of the metabolic signature of diseases for diagnostic and prognostic purposes; and to the understanding of the biochemical causes at the basis of different pathophysiological conditions.

4.1. Individual Fingerprint

One of the main findings of metabolomics is that each individual is "chemically different" from any other in terms of small molecules composition of its systemic biofluids like urine, blood and saliva.[33, 119, 120] Our research group has significantly contributed to define and characterize this strong chemical signature, called individual metabolic phenotype or metabotype.[119, 121–125] In urine, a biofluid whose composition is strongly affected by daily variations, the presence of an invariant part of the human metabolic phenotype, characteristic of each individual has been established through multivariate statistical analysis of several
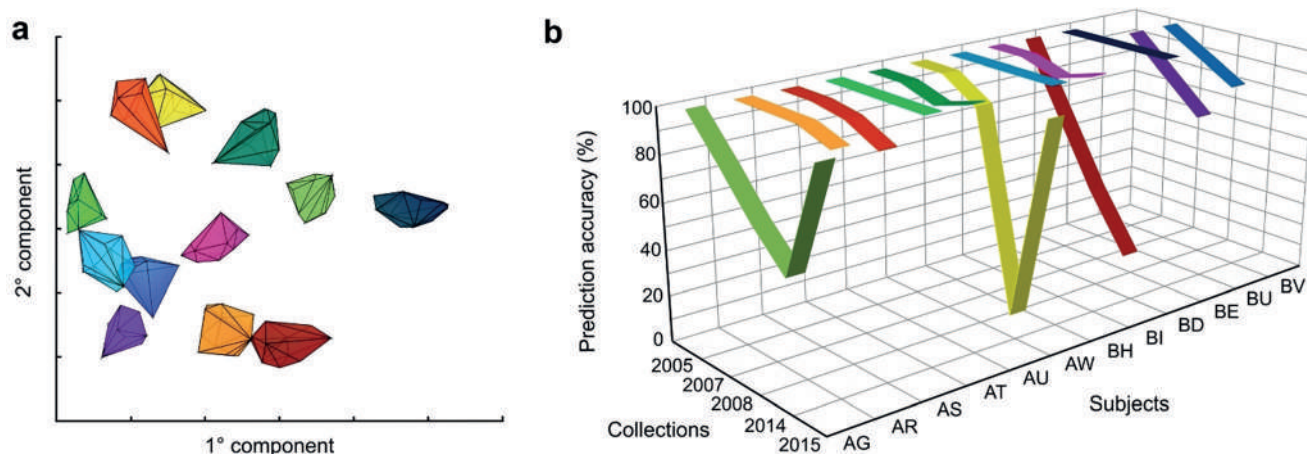
Figure 8. The NMR-derived urine individual metabolic phenotype and its stability over time. a) Multiple urine samples collected from 12 healthy donors (each identified by a given color) over a period of 20 days occupy a well-defined portion of the metabolic space (PCA-CA score plot), thus indicating that intraindividual variations are much smaller than interindividual differences. This is due to an invariant part of the metabolome characteristic of each individual, which identifies the individual phenotype. b) The individual phenotype over the time scale of 10 years is very stable in the absence of physiopathological conditions that can cause abrupt deviations (subjects AG, AW, BD). If this condition is over, the individual phenotype reverts back (AG, AW); adapted from Ref. [122].

samples collected on different days. This has led to the definition of the "metabolic space", where samples from each subject occupy a well-defined and circumscribed sub-space and each individual can be discriminated from all others with near 100% accuracy (Figure 8a). The individual metabolic phenotype is the result of a combination of many factors such as genotype, host metabolism, gut microflora composition, dietary habits, physical activity etc. The collection of multiple urine samples is fundamental to identify the invariant part from the day-to-day variability of the subjects. We have shown that around 20 urine samples per individual, collected in different days, are enough to achieve an individual discrimination of 97–98%. By using 40 urine samples per person values above 99% can be obtained.[119, 121, 122] Using multiple urine collections in the same day for a few days (e.g., 4–5 samples per day, for 10 days) is substantially equivalent to 1 sample per day for 40 days.[125] In addition to being characteristic of each individual at any time point, metabotypes in urine are stable over a time scale of 8–10 years. Over this time-frame, significant metabotype drifts were observed only after the onset of important pathophysiological conditions; the "normal" individual metabotype was regained if the condition was reversed (Figure 8b).[122] These results clearly underline the enormous value of this approach to monitor subjects during their life-span, in order to detect early disease onset, its progression, response to therapy or dietary intervention, etc., with obvious applications in the context of precision medicine.

A strong signature of the individual phenotypes has been also characterized in blood.[120] Due to its nature, blood is free from most of the daily variations observable in urine, and the intra-individual variability is very low. Thus, good discrimination among individuals can be obtained using fewer samples per person. A high degree of stability of the individual phenotype in serum over 7 years has been also demonstrated.[126]

A clear individual metabolic phenotype exists also in saliva[33] and human breath,[127] although it is slightly less strong than the urinary or blood phenotype.

The "individual metabolic phenotype" of urine and blood, and possibly that of other biofluids, possesses suitable characteristics for personalized healthcare solutions, being typical for each subject, stable over time and essentially independent of lifestyle and dietary intake,[119, 121, 122, 125, 128–130] thus allowing us to monitor individual status and response to different stimuli.[131–133]

4.2. Metabolomics and Diseases

For metabolomics applications in medicine, the ability to monitor the individual metabolic fingerprint parallels the ability to identify clear disease signatures. Classically, metabolomics has been used to characterize metabolic profiles of diseases, with the intent of discovering new biomarkers and identifying biochemical pathways involved in disease pathogenesis. NMR metabolomics has already increased our understanding of cellular and physiological metabolism, helping to identify many unexpected biochemical causes for several important chronic and complex diseases,[134, 135] such as cancer,[23, 136–141] cardiovascular diseases,[115, 142–146] diabetes,[147–150] and obesity.[151–153] One of the most striking aspects of metabolomics is the ability to detect at the systemic level alterations in the metabolome, which correlate with pathological states even for those diseases that are not immediately associated to metabolism. Such sensitivity, most probably involving immune mechanisms, is particularly promising to monitor the individual response to illness. NMR metabolomics has the most ambitious objective of detecting early metabolic perturbations even before the manifestation of disease symptoms. Indeed, the goal of precision medicine is to customize subjects' therapeutical treatments according to

their specific omic profiles/fingerprints.[13] NMR metabolic profiling in large-scale epidemiologic studies, though is still recent, has uncovered novel biomarkers for various diseases, has contribute to their etiologic understanding, and has shown the ability to predict disease risks and the effects of drug administration, with the potential to translate into multiple clinical settings.[140,154–157]

The literature available concerning metabolomics applied to clinical examples is extremely wide; we are presenting here a limited number of examples of application to a few diseases using different types of samples.

Coeliac disease (CD) is a multifactorial and complex disorder involving genetic and environmental factors that induce autoimmune response and nutrients malabsorption, thus having great potential impact on metabolism. Our group has extensively described the NMR metabolic signature of CD in serum and urine by identifying important alterations of the metabolic profiles of patients with respect to healthy controls, especially for what concerns energy and ketone body metabolism, and gut microbiota alterations.[158,162] Furthermore, our statistical model has demonstrated to be effective in monitoring the adherence and efficacy of the gluten free diet in CD patients (Figure 9a).

A striking finding was that potential coeliac people, that is, those with typical antibody signature of CD but without symptoms and without or slight intestinal damage, are already classified as coeliac from a metabolomic point of view, suggesting that the CD metabotype precedes the manifestation of the disease.[2]

Cancer is probably the most studied pathology so far via NMR metabolomics;[163–168] Cancer patients present metabolic profiles that are different from those of healthy controls and patients with benign diseases; moreover, the site, the stage, and the location of the tumors may differently affect the metabolome.[169] The most fascinating aspect of metabolomic research is the challenge of translating these evidences into diagnostic or prognostic models capable of ensuring early diagnosis or prediction of recurrence and of the final outcome of the patients with results comparable or even better with respect to classical methodologies. Metabolomic analysis of intact tissues via HR-MAS offers the possibility to obtain information directly at the level of the organ affected by the tumor, and these ones could have prognostic values. HR-MAS NMR analysis of pancreatic intact tissue can provide important information for the characterization of pancreatic adenocarcinomas (PA) and could discriminate PA from healthy pancreatic parenchyma (PP) (Figure 9 b1). Ethanolamine resulted as a possible single metabolic biomarker (Figure 9 b2) for the prediction of long-term survival.[26] The analysis of systemic biofluids such as blood can give fundamental information regarding the patient prognosis. Metastatic colorectal cancer patients (mCRC) and healthy controls were clearly discriminated by multivariate statistical analysis of the serum NMR data; mCRC patients presented various metabolic alterations regarding energy metabolism and inflammatory response.[159] Furthermore, mCRC patients clustered according to their overall survival (Figure 9 c1) with a better performance with respect to traditional biomarkers (hazard ratio = 3.37).

Breast cancer (BC) is a complex and heterogeneous disease which has been extensively characterized by many platforms such as clinicopathological risk factors and various -omic techniques including metabolomics, which has shown significant promises in diagnosis, prognosis, and patient management.[170–175] In this precision medicine era, however, development of tailored oncological treatments for BC and accurate instruments to identify patients at high risk of disease recurrence are lagging.[176] Our group has demonstrated that NMR serum metabolomics could contribute significantly to this aim: in the first single center pilot study, relapse was predicted with quite good accuracy in both training set (90% sensitivity, 67% specificity, and 73% predictive accuracy, AUC: 0.863), and validation set (82% sensitivity, 72% specificity, and 75% predictive accuracy, AUC: 0.824).[177] The results have been reproduced in a multicenter study by analyzing 699 serum samples collected in the framework of an international phase III clinical trial (Figure 9, d1, d2).[160]

Cardiovascular diseases provide another excellent target for metabolomics, especially for what concerns early diagnosis. For example, heart failure (HF) is a complex, chronic, progressive syndrome in which the heart muscle is unable to pump enough blood to cope with the body needs for blood and oxygen. Unfortunately, HF is asymptomatic in its first stages, when medical intervention would still be effective; therefore, early assessment of this disease is a crucial and challenging task.[178] We have shown that NMR serum metabolomics is an excellent instrument for the discrimination between HF patients and healthy controls (Figure 9e). Even more important, the metabolomic fingerprint does not change with disease progression; therefore, it could really represent a useful tool for the purpose of early diagnosis.[143]

Metabolomics also offers a cost-effective and productive route to uncover new targets for drug discovery, and to predict and monitor individual response to drug treatment (pharmacometabolomics).[13,58,179–182] For these purposes, animal models represent the preferred choice, especially for first explorative studies. Since animal models ensure a high rate of standardization, they enable a clearer identification of the intervention effects using a reduced number of subjects with higher discrimination power (often even unsupervised analysis is effective in identifying clusters of interest). As an example, Clayton and co-workers in their study on paracetamol (acetaminophen) administration showed that NMR pharmacometabolomic phenotyping could be successfully used.[161] They analyzed pre- and post-dose urine samples from 65 rats given a single toxic-threshold dose of paracetamol, observing that pre-dose NMR discrimination is related to the post-dose variation in histopathology (Figure 9 f).

For the sake of completeness, it is important to underline that the use of the metabolomic approach is not confined to human medicine. Several potential applications in the veterinary industry are possible, in particular in the framework of disease diagnosis and investigation, optimization of health and production, drug discovery and animal welfare.[3,183–187]
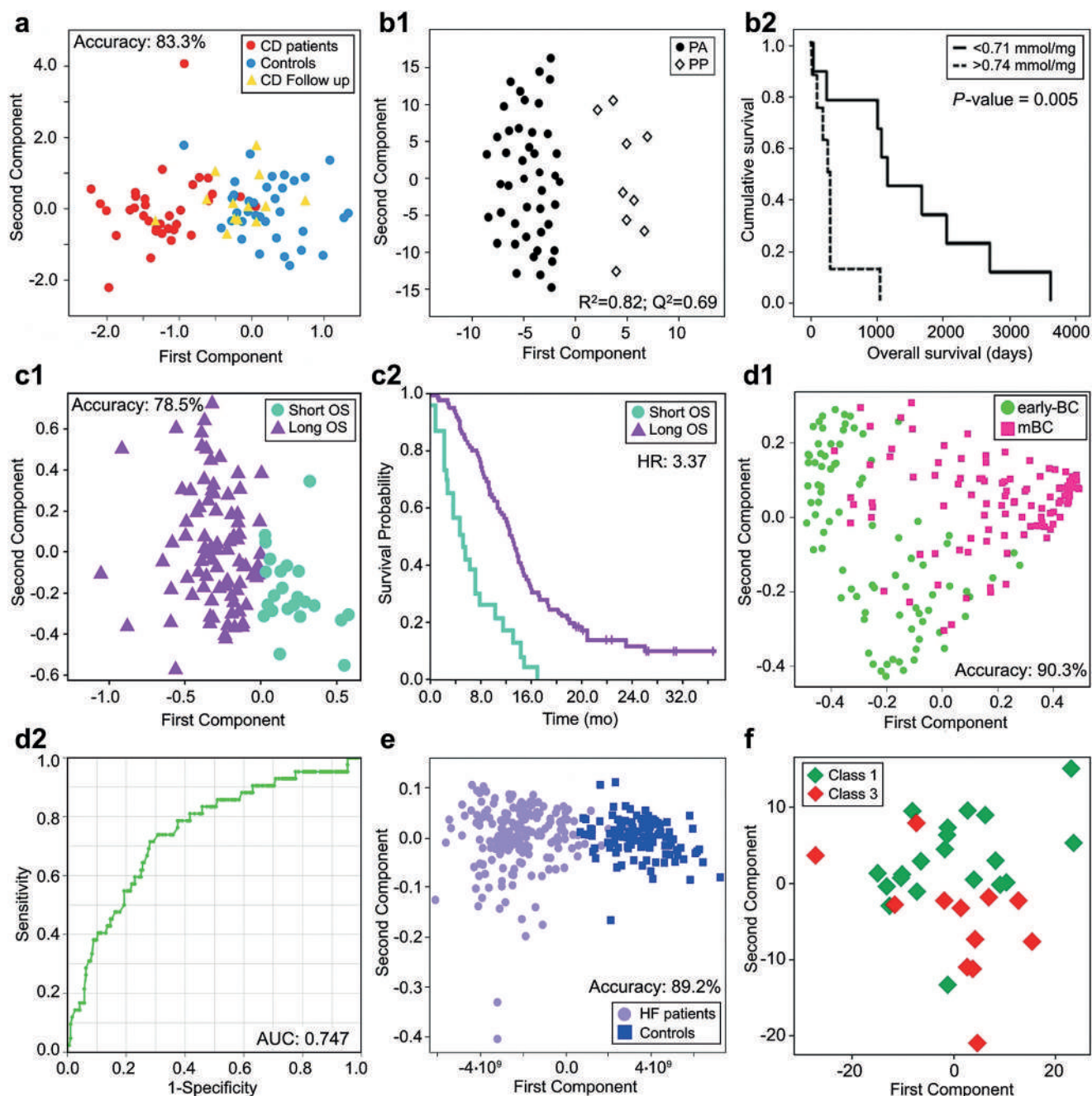
Figure 9. Applications in the biomedical field. a) Monitoring the efficacy of the gluten free diet in CD patients: predictive clustering of CD patients after 12 months on a gluten-free diet (yellow triangles) using a PLS-RCC model built on CD patients (red circles) and healthy controls (light blue circles).[158] b) Metabolomic profiling of tumor tissues predicting clinical outcome of pancreatic adenocarcinoma patients.[26] b1) OPLS-DA model discriminates pancreatic adenocarcinoma tissues (black circles) with respect to pancreatic healthy tissues (white squares), R^2 and Q^2 were used to measure model quality: $R^2 > 0.7$ and $Q^2 > 0.5$ can be considered as a good predictor. b2) Ethanolamine concentration (the threshold value was 0.740 nmol mg^{-1}) as a single metabolic biomarker for the prediction of overall survival in patients with PA. Kaplan–Meier curves show differences between long-term (black line) and short-term (segmented line) survival patients. c) Prediction of overall survival in patients with mCRC. c1) PLS-CA clustering for long OS (purple triangles) and short OS (aquamarine circles) mCRC patients. c2) Kaplan–Meier curves showing survival probability based on the ^{1}H-NMR metabolomic model.[159] d) Identification of early-BC patients at increased risk of disease recurrence via serum metabolomics. d1) Clustering of serum metabolomic profiles between early-BC (green circles) and mBC (pink squares) patients using a Random Forest (RF) classifier in the training set. d2) Prediction of relapse in the test set containing 192 relapse early-BC patients and 42 early-BC patients free from disease up to 6 years (ROC curve).[160] e) Discrimination between HF patients (lilac circles) and healthy subjects (dark blue squares) using an OPLS-DA.[143] f) Pharmaco-metabonomic phenotyping: a scores plot from PCA of the pre-dose of paracetamol urine spectra. NMR data discriminates low histology damage (Class 1, green squares) and severe histology damage (Class 3, red squares).[161]

4.3. Pathway Analysis

Once the biomarkers, or metabolic profiles, whose concentrations are altered due to the biological processes under investigation have been identified, pathway analysis can be performed with the final aim of obtaining a mechanistic explanation of the changes observed. Pathway analysis strengthens the information generated by metabolomic analysis. In the framework of diseases, for instance, the identification of the altered metabolites has been successfully correlated with the biological pathways and processes involved in pathogenesis and disease progression.[149,158,188,189] In particular, pathway analysis has been applied in several cancer studies to unravel the processes at the basis of the oncogenic phenotype.[137,190]

Online biological databases, such as KEGG[191] (Kyoto Encyclopedia of Genes and Genomes) Pathway Database, provide information of a large number of metabolic pathways and can be easily used to identify and visualize the metabolites involved in several biological processes. Furthermore, the development of metabolite set enrichment analysis (MSEA) methods[192] strongly helps the identification and the functional and/or biological interpretation of patterns of metabolite concentration changes in a biologically meaningful context. Publicly available tools implementing these methods are available.[193] Among these tools, MetaboAnalyst—a comprehensive tool for NMR- and MS- based metabolomic analysis and interpretation[194] includes sections for metabolic pathway analysis and metabolite set enrichment analysis. These approaches examine the metabolites present in the biological matrix at a particular time point, thus providing steady state levels of metabolites.

Stable isotope-resolved metabolomics (SIRM) provides a more dynamic assessment of specific metabolic pathways using stable isotopes (^{13}C, ^{15}N or ^{2}H).[195] In this approach, isotopically enriched precursors such as 1,2-^{13}C2-glucose or ^{13}C5, ^{15}N2- glutamine, are administered to a biological system, enabling the analysis of the fate of individual atoms of a particular metabolite, and thereby a robust tracing of metabolic pathways and of their fluxes in cells, animals and humans (fluxomics). During the last years, the use of isotopic labelling tracers has been successfully applied to unveil the different metabolic pathways activated under different conditions in cancer, providing important information for the development of new therapeutic strategies.[196–198]

According to the type of samples, the biological interpretation of changes in concentration levels of metabolites through pathway analysis may be tangled. The simpler the system studied, the easier the reconstruction of the biological pathways involved in the investigated process. For instance, in cultured cell models the biological explanation is the most immediate. Our research group investigated the metabolic changes induced by the oncogenic enzyme Sphingosine Kinase 1 (SK1) in ovarian cancer.[23] Using an NMR-untargeted approach, we demonstrated that SK1 expression is sufficient to alter radically the metabolomic profile of ovarian cancer cells influencing both the glycolytic pathway and the tricarboxylic acid cycle (TCA) (Figure 10). With this approach we have highlighted the occurrence of SK1-induced
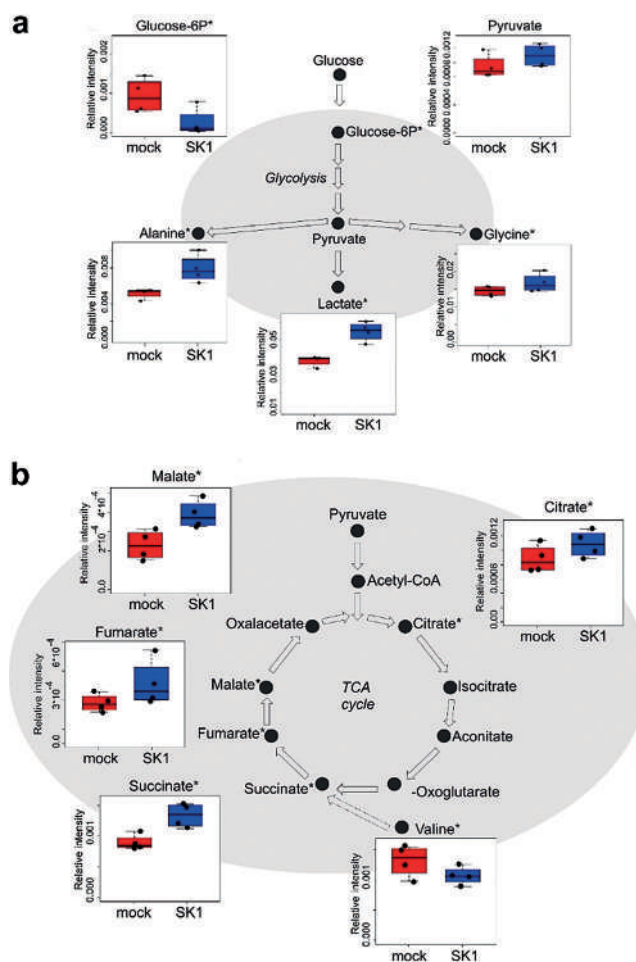

Figure 10. SK1 expression induces a metabolic switch known as the Warburg effect in A2780 ovarian cancer cells.[23] From the comparison between A2780 mock (blue) and SK1-expressing cells (red) emerges that expression of SK1 induced a high glycolytic rate (a), characterized by increased levels of lactate, and decreased oxidative metabolism, associated with the accumulation of intermediates of the TCA (b). Changes in metabolite levels caused by SK1 overexpression were statistically significant by paired Wilcoxon test, *P-value < 0.05.

Warburg effect, and pointed out that SK1 expression also modulates the pentose phosphate pathway and nucleotide biosynthesis, both fundamental to satisfy the anabolic demand of high proliferating cancer cells.

5. General Considerations on Experimental Design and Significance of the Results

As in any other scientific field, before initiating a metabolomic study it is of primary importance to set up a correct experimental design, especially regarding the total number of samples to collect, the type of samples to be analyzed, the number of different groups to be examined, the inclusion and exclusion criteria, the choice of the control group, the kind of clinical design (prospective, retrospective, paired/unpaired). Well planned studies ensure that the experiments are efficient and help to avoid biases in the statistical analysis.[199]

Having in mind a clear object of the study and a well posed biological question is of course of great help in deciding an appropriate experimental plan. However, metabolomics, in its untargeted incarnation, does not require any initial hypothesis; therefore, it may happen that it is not the original needs that drive the experimental steps, but, conversely, metabolomics is applied to samples already collected, and often for different purposes. For this reason, researchers in metabolomics are faced with challenges posed, for instance, by samples not properly collected and/or stored (e.g. according to SOPs designed for different purposes), causing unwanted sources of variation, lack of correct statistical power analysis or small number of samples for statistical analysis. Power analysis[200] is sometimes forgotten by metabolomic practitioners, but it is a pillar of any good experimental design. Roughly speaking, statistical power is a measure of the chance to detect in the experiment statistically significant effects provided that the said effects really exist. Power is (also) a function of the sample size, so the most common aim of a power analysis is to determine the minimum number of samples needed for the study.[201] Human clinical studies often require large cohorts, due to the high interindividual variability[202] that could obliterate the investigated effects. The most common clinical design for human metabolomic studies is the simple comparison of samples collected from two groups of donors (e.g. a diseased group vs. healthy controls). This makes sense and can works reasonably well; however, interindividual variations are an unavoidable source of noise. A better approach would be to compare the same healthy individual with its diseased counterpart (Figure 11), by planning a long-lasting cohort study where a number of healthy individuals are recruited and followed for many years, ideally collecting repeated samples until at least some of them develop the disease of interest (or any disease, see below). This approach has the great advantage to eliminate interindividual variation. Further, using repeated samples, it is not only possible to define the disease fingerprint, but also its dynamic evolution in time, allowing us to trace the path of each individual from healthiness to pathology, thereby opening the door to a predictive medicine strategy able to detect a disease before the appearance of the symptoms.

A further point speaks in favor of metabolomics in this respect. When used for epidemiological purposes, it will permit the buildup of fingerprints for many different diseases (especially the most common) as long as they appear in the cohort. Therefore, subsequent samples, either from the same or from different individuals, could be compared with an increasing number of fingerprints, and therefore allow for early diagnosis of an increasing number of diseases.

Another comment is due to the concept itself of univariate analysis applied to untargeted metabolomics: in essence, the aim of univariate analysis is that of finding a specific biomarker. Indeed, as metabolomics can quantitate many metabolites at once, it has been sometimes termed a "biomarker discovery" tool. However, as such, the results of metabolomics have been often disappointing. In fact, a statistically valid criterion such as a low *P*-value or a large effect size can only tell that the concentration of a certain metabolite is statistically different between two groups. Assuming
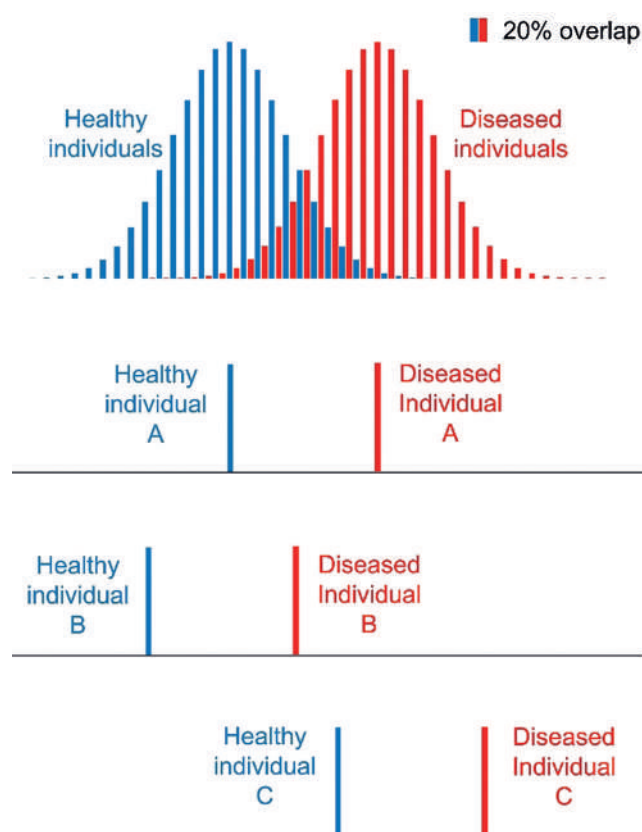

Figure 11. A single metabolite that is significant for the discrimination of healthy and diseased populations still may not provide a clear-cut answer (20% false positives and false negatives in this example, upper panel). The ambiguity originates from inter-individual variability, represented by the width of the two Gaussian distributions. Conversely, it is more likely that an individual can be correctly classified if the level of the metabolite is quantified before and after the onset of the disease. While individual A is correctly classified as healthy and subsequently as diseased also according to population statistics, individuals B and C would not be classified correctly if evaluated only by population statistics.

two Gaussian concentration distributions for the two groups, the region of overlap between the two curves is very often too large (Figure 11), that is, the accuracy in distinguishing if a sample belongs to one group or the other is often modest (in the example, 80%, that is, 20% of false positives or false negatives). A good biomarker should provide a much higher accuracy. Unfortunately, it is becoming evident that there are not so many metabolites that are as good as, for instance, glucose for diabetes. However, a much more robust separation could be made if a panel of all metabolites that are statistically significant between the two groups is considered to be a single biomarker. In fact, two biomarkers that fail in 20% of the cases, when taken together, would only fail in 4% of the cases, and three of them would bring the failures down to less than 1%. Basically, this reasoning brings us back to profiling and multivariate analysis (Section 3.3). Taken to the extreme, fingerprinting and multivariate analysis would always be the most efficient method, that is, a fingerprint would be, by definition, the best biomarker. While the above considerations indicate fingerprinting as the way to go, it

should be considered that any method to be used in clinics will need extensive validation and ultimately approval by regulatory agencies. Up to now, researchers have experienced difficulties in having an analytical method considered by regulatory agencies if not based on the quantitation of one or more well identified metabolites. This is an attitude that may change in the future, under the growing evidence that fingerprints, once obtained under rigorous SOPs, are by themselves reliable analytical quantities.

6. Other Fields of Application: Foodomics as an Example

The food and beverage industry is constantly growing, and the global expenditure on this sector is twelve times bigger than the global expenditure on the health care sector.[203, 204] It is therefore not surprising that the traditional use of NMR as an analytical tool for quality control of food stuff has recently evolved into a key tool for the new field of "foodomics", a discipline that studies food and nutrition in relation to consumer's well-being.[205] Foodomics can be applied to study and characterize each step of the production/consumption chain in a comprehensive, integrated and high-throughput approach, to improve also consumer's well-being, health, and confidence.

The applications of metabolomics in foodomics can be summarized in three main areas (Figure 12): i) in the field of human health, which can be divided in food consumption monitoring studies, where, given a particular diet, the consequent metabolome changes are investigated,[8, 206–209] and in treating/preventing diseases by improving and monitoring patient diet;[158] ii) in the food resources area, which can be investigated through the analysis of the composition of food from animal and plant origin and defined also on the basis of climate, land and cultural practices ("terroir") that

contribute to the traceability and thus to the characterization of the genuine product; iii) in food processing, which includes the characterization of the most influencing pre- and post-production manipulations on the original product, such as altered growing conditions (e.g. feed, GMOs, chemicals, pesticides etc.), the effect of packaging and storage,[210,211] and, last but not least, safety and authenticity control. Post-harvest manipulations, such as storage, can widely affect food profiles. For example, the molecular effect of controlled atmosphere, a common practice applied to extend the storage life of fruit, was studied on apples (*Malus domestica*).[210]

Despite the fact that conventional quality, safety and authenticity control in food is based on targeted strategies, high resolution ^{1}H-NMR has found its way into routine food analysis,[214] offering several advantages if performed under well-defined instrumental specifications and SOPs, generating an extremely reproducible food fingerprint and fully quantitative data by means of a single experiment with minimal or no sample preparation. Thus, the fingerprint of a food matrix, analyzed with statistical methods, can reveal some latent information such as the provenance of milk samples from different farms even when all located in the same small geographical area,[215] the origin of fruit juice[212] and the grape variety for wine[213] (Figure 13).
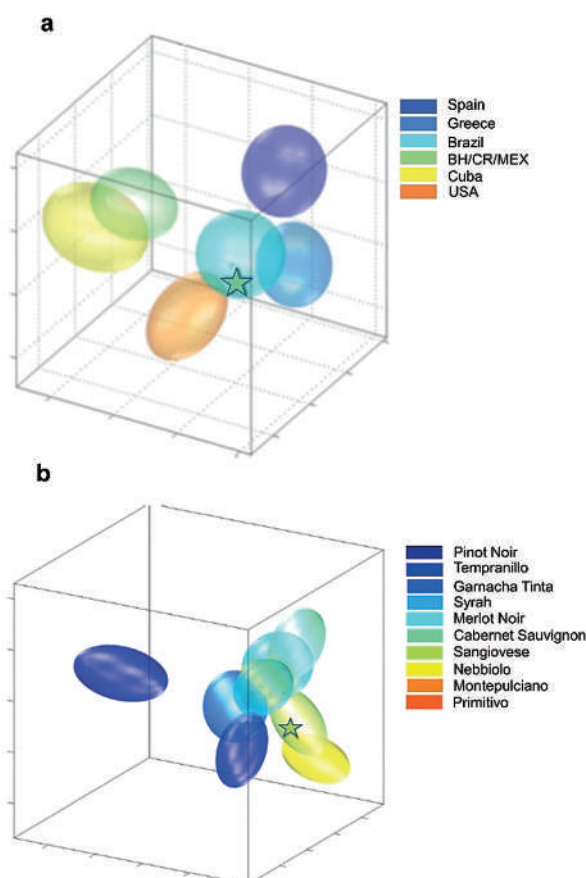
Figure 13. 3D projections of the model space with the ellipsoids of possible groups available in the databases. a) Extract from JuiceScreener™, estimation of the origin of an orange juice.[212] b) Extract from Winescreener™ report of a Sangiovese sample (star) with respect to other Italian wines. Classification model for the wine type assignment, representing the probability of classification for every group.[213]
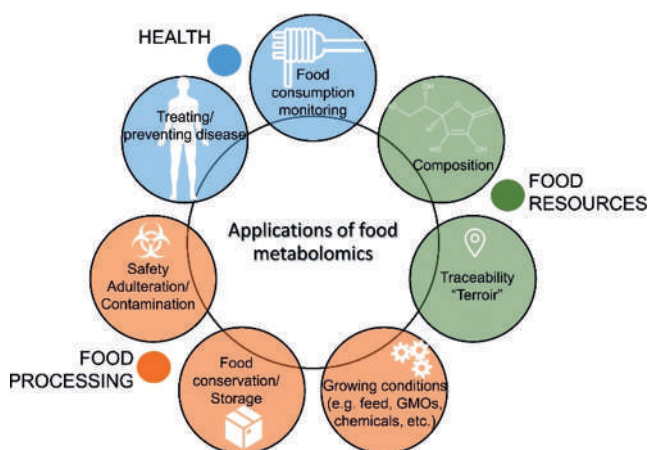

Figure 12. The main applications of food metabolomics. Food consumption monitoring and treating/preventing diseases, for the human health field (blue); chemical and molecular analyses of food composition and its characteristic ecosystem that contributes to the definition of traceability/authenticity of food resource (green); food processing area for the characterization of the effect of pre- and post-harvest manipulations on food (orange).

A platform for ^{1}H-NMR based food quality and authenticity control, the FoodScreenerTM platform, has already been introduced by Bruker BioSpin based on an Avance 400 MHz spectrometer, for the analysis and characterization of fruit juices and wines, and with application to honey samples being currently in development.[214]

7. Technical Improvements

As for other omics, the future of metabolomics is driven by technology developments. Herein, we will summarize a few recent technical advances that could expand the common metabolomic routine analyses by NMR, by providing alternative protocols for sample preparation and methods for facilitating metabolite assignment in the NMR spectra of biofluids.

As far as sample collection is concerned, one of the main problems in following SOPs is associated to situations where sample collection at home is needed: samples like urine and saliva can be collected directly by the donor, but problems may arise if filtering or keeping the samples at the right temperature is needed. Recently, we proposed a complementary pre-analytical method relying on the gelification of biofluids and subsequent analysis by HR-MAS NMR spectroscopy.[216] Urine gelification is rapid and reproducible and permits preservation of the sample for 24 h without the need of refrigeration (Figure 14a).

Moreover, this approach has a substantial advantage in cases of low biofluids abundance (e.g. newborns, small animals), since the HR-MAS sample only requires ≈ 40 μL of gelified urine. In order to improve the identification of new metabolites in biological matrixes, a separation protocol has been proposed,[217] where the solution of an NMR metabolomic mixture is passed through a set of filters, loaded with either anionic or cationic ion exchange resin beads (IERB). This method provides a separation of the metabolites depending on their specific physical-chemical properties.[217] As a result, the ^{1}H-^{13}C HSQC 2D spectra of the flow through and the eluate could be used for a better assignment of the metabolites.

A novel approach for the automated and accurate identification of urine metabolites was recently introduced by our group. This has been achieved by constructing a large number (ca. 4000) of artificial urine samples where the concentrations of the most abundant metabolites and of several inorganic ions were varied to sparsely but efficiently populate an N-dimensional concentration matrix. By the use of statistical machine learning tools, we constructed an algorithm that, based upon the chemical shifts of only five signals, can accurately predict the chemical shifts of many other metabolites in the urine sample (so far 90 spin systems chemical shifts of 63 metabolites) (Figure 7). This predictor, besides allowing for easy quantitation of many metabolites whose signals are visible in the spectrum, provides estimates of the concentrations of 11 NMR "invisible" inorganic ions (Figure 14b). Our software estimates the inorganic ions (including trace metal ions) in one single, inexpensive experiment, thus making NMR a significant tool in the recently introduced field of ionomics.[218] This strategy could be expanded to other metabolomic fluids, such as juices, soft drinks etc.

8. Summary and Outlook

This Review aimed at highlighting the many chemical aspects of metabolomics. This was not meant to imply that the ultimate goal of metabolomics research is chemistry, but rather that chemistry is needed for a better understanding of the biomedical goals and can even be inspirational for defining the strategy of how to best achieve these goals. By restricting ourselves to metabolomics by NMR we hope we have shown that the horizon of metabolomics—in particular of its untargeted fingerprinting version, which is best performed by NMR—can be actually enlarged towards population-wide metabolomic screening, especially as we expect further technological advancements, of which Section 7 provides some examples. As highlighted in Section 4, metabolomics is being gradually appreciated as the science that is closest to provide early diagnosis of diseases. The medical community is progressively appreciating that searching for disease-specific biomarkers is probably no longer the way to go: single biomarkers seldom have enough discriminating power, while a panel of biomarkers taken together has a better chance to provide higher discriminating power. By extrapolating this concept (Section 5), a whole fingerprint, that virtually contains "all" detectable metabolites, will have the highest possible discriminating power. This extrapolation requires a leap of faith, that is, that an NMR spectrum is equivalent to a complete panel of perfectly quantitated metabolites. Only a chemist with enough NMR knowledge could convincingly make the case in front of non-chemists, that presently would rather trust a table of chemical substances with the associated concentrations.

But how far are we from actually adopting metabolomics by NMR as a population-wide screening method? From the examples reported in Section 4, and from many other examples in the recent literature, it appears that i) for several diseases there are already NMR fingerprints that are characteristic of each of these diseases; ii) by comparing the NMR spectrum of a "healthy" subject with these fingerprints, an early diagnosis can be made with an accuracy of 70–85 %; iii) recurrences in some types of cancers can be predicted earlier than by any other method, again with ca. 80 % accuracy. The conceptual distance from the present situation to the adoption of metabolomics at the level of national health systems is given by the need to move further in two directions: towards defining fingerprints for a larger number of diseases, and towards increasing the number of subjects from which the fingerprint is built, to increase robustness. It should be noted that, as the population screening will progress, the system will become more and more informative (more diseases will be fingerprinted, and therefore more could be detected from the same NMR spectrum) and more reliable (accuracy will further increase). It should also be noted that even an accuracy of 80 % would be more than enough to prescribe, to individuals with a tentative early diagnosis of a given
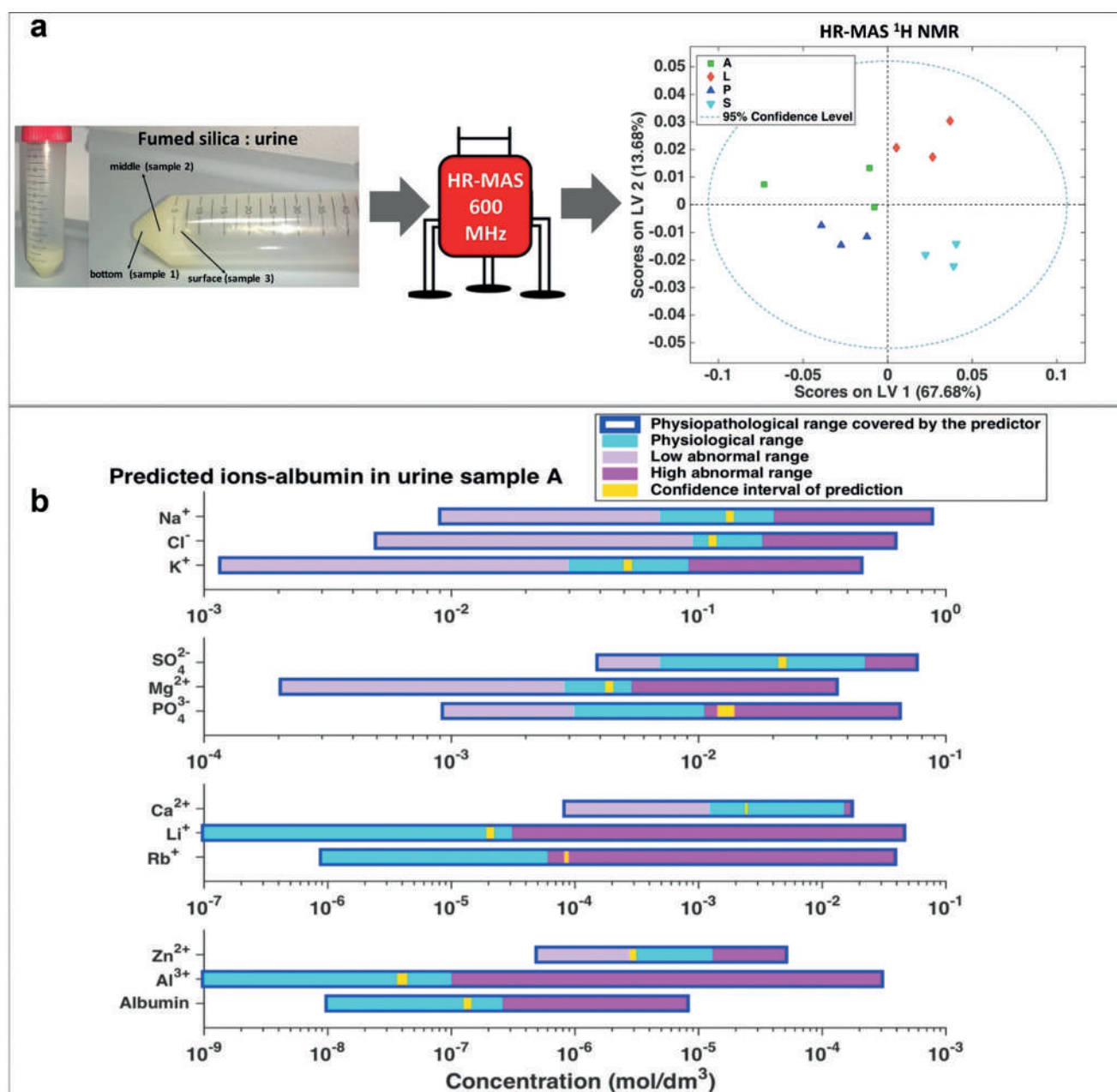
Figure 14. Examples of technical improvement. a) Gelification of urine by silica particles provides a ready-to-use HR-MAS sample that could be exploited for metabolomic routine studies. b) Prediction of 11 inorganic ions and albumin by urine shift predictor.[14]

pathology, a specific check-up for that disease (e.g. echocardiography for heart failure, or antibody tests for celiac disease, etc.).

The next question to answer is the cost of adopting such a population-wide NMR screening: although NMR instruments are expensive (a 600 MHz dedicated instrument costs more than 1 million € + VAT), they are long-lived and can work almost continuously for months, provided the automatic sample changer is refilled constantly. They do not need to be recalibrated nearly as often as MS instruments, and sample preparation is extremely simple. Overall, the cost per sample when all operations are optimized, including initial investment, maintenance and upgrading, personnel and consum-

ables, would probably be lower than 15 €. Therefore, a metabolomic NMR analysis on, say, a serum sample would add only a modest amount to the cost of a blood test for a panel of analytes as it is currently done for periodic check-ups.

The last question is the feasibility: for instance, to screen 400 million Europeans once every two years, 200 million spectra per year would need to be recorded. A single instrument can run about 120 spectra per day, that is, around 40 000 spectra per year, so ca. 5000 NMR instruments will be needed. They will have to be spread across the continent, in multiple copies in major and middle-sized hospital, as well as in public and private clinical analysis

laboratories, because clinical analyses are typically performed locally to avoid shipping of massive numbers of samples. Five thousand spectrometers may seem as a very large number but is about the same as the number of MRI scanners in use in Europe, each of which costing significantly more than one million €. When MRI started to appear as a potentially useful technique, it took only a few years for MRI scanners to spread and grow to the levels of today.

Acknowledgements

Conflict of interest

The authors declare no conflict of interest.

[1] J. K. Nicholson, J. C. Lindon, *Nature* **2008**, *455*, 1054–1056.
[2] A. Calabrò, E. Gralka, C. Luchinat, E. Saccenti, L. Tenori, *J. Autoimmune Dis.* **2014**, *2014*, e756138.
[3] J. Kirwan, *In Practice* **2013**, *35*, 438–445.
[4] J. G. M. Pontes, A. J. M. Brasil, G. C. F. Cruz, R. N. de Souza, L. Tasic, *Anal. Methods* **2017**, *9*, 1078–1096.
[5] J. Tang, *Curr. Genomics* **2011**, *12*, 391–403.
[6] M. Ramirez-Gaona, A. Marcu, A. Pon, A. C. Guo, T. Sajed, N. A. Wishart, N. Karu, Y. Djoumbou Feunang, D. Arndt, D. S. Wishart, *Nucleic Acids Res.* **2017**, *45*, D440–D445.
[7] R. D. Hall, I. D. Brouwer, M. A. Fitzgerald, *Physiol. Plant.* **2008**, *132*, 162–175.
[8] L. H. Münger, A. Trimigno, G. Picone, C. Freiburghaus, G. Pimentel, K. J. Burton, F. P. Pralong, N. Vionnet, F. Capozzi, R. Badertscher, et al., *J. Proteome Res.* **2017**, *16*, 3321–3335.
[9] S. Kim, J. Kim, E. J. Yun, K. H. Kim, *Curr. Opin. Biotechnol.* **2016**, *37*, 16–23.
[10] J. L. Markley, R. Brüschweiler, A. S. Edison, H. R. Eghbalnia, R. Powers, D. Raftery, D. S. Wishart, *Curr. Opin. Biotechnol.* **2017**, *43*, 34–40.
[11] A. Klupczyńska, P. Dereziński, Z. J. Kokot, *Acta Pol. Pharm.* **2015**, *72*, 629–641.
[12] J. C. Lindon, J. K. Nicholson, *Annu. Rev. Anal. Chem.* **2008**, *1*, 45–69.
[13] D. S. Wishart, *Nat. Rev. Drug Discovery* **2016**, *15*, 473–484.
[14] P. G. Takis, H. Schäfer, M. Spraul, C. Luchinat, *Nat. Commun.* **2017**, *8*, 1662.
[15] A. Scalbert, L. Brennan, O. Fiehn, T. Hankemeier, B. S. Kristal, B. van Ommen, E. Pujos-Guillot, E. Verheij, D. Wishart, S. Wopereis, *Metabolomics* **2009**, *5*, 435–458.
[16] B. S. Barbosa, L. G. Martins, T. B. B. C. Costa, G. Cruz, L. Tasic, *Methods Mol. Biol.* **2018**, *1735*, 365–379.
[17] J. Li, T. Vosegaard, Z. Guo, *Prog. Lipid Res.* **2017**, *68*, 37–56.
[18] S. Bouatra, F. Aziat, R. Mandal, A. C. Guo, M. R. Wilson, C. Knox, T. C. Bjorndahl, R. Krishnamurthy, F. Saleem, P. Liu, et al., *PloS One* **2013**, *8*, e73076.
[19] N. Psychogios, D. D. Hau, J. Peng, A. C. Guo, R. Mandal, S. Bouatra, I. Sinelnikov, R. Krishnamurthy, R. Eisner, B. Gautam, et al., *PLoS ONE* **2011**, *6*, e16957.
[20] Z. T. Dame, F. Aziat, R. Mandal, R. Krishnamurthy, S. Bouatra, S. Borzouie, A. C. Guo, T. Sajed, L. Deng, H. Lin, et al., *Metabolomics* **2015**, *11*, 1864–1883.
[21] D. S. Wishart, M. J. Lewis, J. A. Morrissey, M. D. Flegel, K. Jeroncic, Y. Xiong, D. Cheng, R. Eisner, B. Gautam, D. Tzur, et al., *J. Chromatogr. B* **2008**, *871*, 164–173.
[22] C. Airoldi, C. Ciaramelli, M. Fumagalli, R. Bussei, V. Mazzoni, S. Viglio, P. Iadarola, J. Stolk, *J. Proteome Res.* **2016**, *15*, 4569–4578.
[23] C. Bernacchioni, V. Ghini, F. Cencetti, L. Japtok, C. Donati, P. Bruni, P. Turano, *Mol. Oncol.* **2017**, *11*, 517–533.
[24] M. Cuperlovic-Culf, D. Ferguson, A. Culf, P. Morin, M. Touaibia, *J. Biol. Chem.* **2012**, *287*, 20164–20175.
[25] B. Martínez-Granados, D. Monleón, M. C. Martínez-Bisbal, J. M. Rodrigo, J. del Olmo, P. Lluch, A. Ferrández, L. Martí-Bonmatí, B. Celda, *NMR Biomed.* **2006**, *19*, 90–100.
[26] S. Battini, F. Faitot, A. Imperiale, A. E. Cicek, C. Heimburger, G. Averous, P. Bachellier, I. J. Namer, *BMC Med.* **2017**, *15*, 56.
[27] T. L. Palama, I. Canard, G. J. P. Rautureau, C. Mirande, S. Chatellier, B. Elena-Herrmann, *Analyst* **2016**, *141*, 4558–4561.
[28] F. Romano, G. Meoni, V. Manavella, G. Baima, L. Tenori, S. Cacciatore, M. Aimetti, *J. Periodontol.* **2018**, https://doi.org/10.1002/JPER.18-0097.
[29] CEN/TS 16945, Molecular in vitro diagnostic examinations. Specifications for pre-examination processes for metabolomics in urine, venous blood serum and plasma, **2016**.
[30] P. Bernini, I. Bertini, C. Luchinat, P. Nincheri, S. Staderini, P. Turano, *J. Biomol. NMR* **2011**, *49*, 231–243.
[31] O. Beckonert, H. C. Keun, T. M. D. Ebbels, J. Bundy, E. Holmes, J. C. Lindon, J. K. Nicholson, *Nat. Protoc.* **2007**, *2*, 2692–2703.
[32] M. Navazesh, *Ann. N. Y. Acad. Sci.* **1993**, *694*, 72–77.
[33] S. Wallner-Liebmann, L. Tenori, A. Mazzoleni, M. Dieber-Rotheneder, M. Konrad, P. Hofmann, C. Luchinat, P. Turano, K. Zatloukal, *J. Proteome Res.* **2016**, *15*, 1787–1793.
[34] C. E. Teunissen, A. Petzold, J. L. Bennett, F. S. Berven, L. Brundin, M. Comabella, D. Franciotta, J. L. Frederiksen, J. O. Fleming, R. Furlan, et al., *Neurology* **2009**, *73*, 1914–1922.
[35] G. de Laurentiis, D. Paris, D. Melck, M. Maniscalco, S. Marsico, G. Corso, A. Motta, M. Sofia, *Eur. Respir. J.* **2008**, *32*, 1175–1183.
[36] S. Lamichhane, C. C. Yde, M. S. Schmedes, H. M. Jensen, S. Meier, H. C. Bertram, *Anal. Chem.* **2015**, *87*, 5930–5937.
[37] O. Deda, H. G. Gika, I. D. Wilson, G. A. Theodoridis, *J. Pharm. Biomed. Anal.* **2015**, *113*, 137–150.
[38] O. Beckonert, M. Coen, H. C. Keun, Y. Wang, T. M. D. Ebbels, E. Holmes, J. C. Lindon, J. K. Nicholson, *Nat. Protoc.* **2010**, *5*, 1019–1032.
[39] A. Gardner, H. G. Parkes, G. H. Carpenter, P.-W. So, *J. Proteome Res.* **2018**, *17*, 1521–1531.
[40] C. Brunius, A. Pedersen, D. Malmodin, B. G. Karlsson, L. I. Andersson, G. Tybring, R. Landberg, *Bioinformatics* **2017**, *33*, 3567–3574.
[41] A.-H. Emwas, C. Luchinat, P. Turano, L. Tenori, R. Roy, R. M. Salek, D. Ryan, J. S. Merzaban, R. Kaddurah-Daouk, A. C. Zeri, et al., *Metabolomics* **2015**, 11, 872—894.

[42] S. Cacciatore, X. Hu, C. Viertler, M. Kap, G. A. Bernhardt, H.-J. Mischinger, P. Riegman, K. Zatloukal, C. Luchinat, P. Turano, *J. Proteome Res.* **2013**, *12*, 5723–5729.

[43] V. Ghini, F. T. Unger, L. Tenori, P. Turano, H. Juhl, K. A. David, *Metabolomics* **2015**, *11*, 1769–1778.

[44] S. Ravanbakhsh, P. Liu, T. Bjorndahl, R. Mandal, J. R. Grant, M. Wilson, R. Eisner, I. Sinelnikov, X. Hu, C. Luchinat, et al., *ArXiv14091456 Cs Q-Bio*, **2014**.

[45] G. A. N. Gowda, D. Raftery, *Anal. Chem.* **2014**, *86*, 5433–5440.

[46] D. S. Wishart, Y. D. Feunang, A. Marcu, A. C. Guo, K. Liang, R. Vázquez-Fresno, T. Sajed, D. Johnson, C. Li, N. Karu, et al., *Nucleic Acids Res.* **2018**, *46*, D608–D617.

[47] J. T. M. Pearce, T. J. Athersuch, T. M. D. Ebbels, J. C. Lindon, J. K. Nicholson, H. C. Keun, *Anal. Chem.* **2008**, *80*, 7158–7162.

[48] G. Wider, L. Dreier, *J. Am. Chem. Soc.* **2006**, *128*, 2571–2576.

[49] S. Akoka, L. Barantin, M. Trierweiler, *Anal. Chem.* **1999**, *71*, 2554–2557.

[50] Y. Wu, L. Li, *J. Chromatogr. A* **2016**, *1430*, 80–95.

[51] J. Hochrein, H. U. Zacharias, F. Taruttis, C. Samol, J. C. Engelmann, R. Spang, P. J. Oefner, W. Gronwald, *J. Proteome Res.* **2015**, *14*, 3217–3228.

[52] F. Dieterle, A. Ross, G. Schlotterbeck, H. Senn, *Anal. Chem.* **2006**, *78*, 4281–4290.

[53] J. Lee, J. Park, M. Lim, S. J. Seong, J. J. Seo, S. M. Park, H. W. Lee, Y.-R. Yoon, *Anal. Sci.* **2012**, *28*, 801–805.

[54] M. Åstrand, *J. Comput. Biol.* **2003**, *10*, 95–102.

[55] P. Filzmoser, B. Walczak, *J. Chromatogr. A* **2014**, *1362*, 194–205.

[56] S. M. Kohl, M. S. Klein, J. Hochrein, P. J. Oefner, R. Spang, W. Gronwald, *Metabolomics* **2012**, *8*, 146–160.

[57] I. Bertini, C. Luchinat, M. Miniati, S. Monti, L. Tenori, *Metabolomics* **2014**, *10*, 302–311.

[58] P. Montuschi, G. Santini, N. Mores, A. Vignoli, F. Macagno, R. Shohreh, L. Tenori, G. Zini, L. Fuso, C. Mondino, et al., *Front. Pharmacol.* **2018**, https://doi.org/10.3389/fphar.2018.00258.

[59] A. Vignoli, D. M. Rodio, A. Bellizzi, A. P. Sobolev, E. Anzivino, M. Mischitelli, L. Tenori, F. Marini, R. Priori, R. Scrivo, et al., *Anal. Bioanal. Chem.* **2017**, *409*, 1405–1413.

[60] Z. León, J. C. García-Cañaveras, M. T. Donato, A. Lahoz, *Electrophoresis* **2013**, *34*, 2762–2775.

[61] C. Muschet, G. Möller, C. Prehn, M. H. de Angelis, J. Adamski, J. Tokarz, *Metabolomics* **2016**, *12*, 151.

[62] X. Xiao, M. Hu, M. Liu, J. Hu, *Metabolomics* **2016**, *6*, 1–11.

[63] L. D. Roberts, A. L. Souza, R. E. Gerszten, C. B. Clish, *Curr. Protoc. Mol. Biol.* **2012**, *CHAPTER*, Unit30.2.

[64] W. J. Griffiths, *Metabolomics, Metabonomics and Metabolite Profiling*, Royal Society of Chemistry, London, **2008**.

[65] A. Smolinska, L. Blanchet, L. M. C. Buydens, S. S. Wijmenga, *Anal. Chim. Acta* **2012**, *750*, 82–97.

[66] A.-H. Emwas, E. Saccenti, X. Gao, R. T. McKay, V. A. P. M. dos Santos, R. Roy, D. S. Wishart, *Metabolomics* **2018**, *14*, 31.

[67] T. N. Vu, K. Laukens, *Metabolites* **2013**, *3*, 259–276.

[68] F. Savorani, G. Tomasi, S. B. Engelsen, *J. Magn. Reson.* **2010**, *202*, 190–202.

[69] C. J. Lindon, J. K. Nicholson, E. Holmes, *The Handbook of Metabonomics and Metabolomics*, Elsevier, Amsterdam, **2007**.

[70] S. Ren, A. A. Hinzman, E. L. Kang, R. D. Szczesniak, L. J. Lu, *Metabolomics* **2015**, *11*, 1492–1513.

[71] S. Wold, K. Esbensen, P. Geladi, *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37–52.

[72] H. Abdi, L. J. Williams, *Wiley Interdiscip. Rev. Comput. Stat.* **2010**, *2*, 433–459.

[73] R. Bujak, M. J. Markuszewski, in *Identif. Data Process. Methods Metabolomics*, Future Science Ltd, London, **2015**, pp. 82–95.

[74] M. Ringnér, *Nat. Biotechnol.* **2008**, *26*, 303–304.

[75] A. Hyvärinen, E. Oja, *Neural Netw.* **2000**, *13*, 411–430.

[76] M. V. Baptiste, F. R. Rousseau, P. de Tullio, B. Govaerts, *J. Biom. Biostat.* **2017**, 1–8.

[77] J. Camacho, R. A. Rodríguez-Gómez, E. Saccenti, *J. Comput. Graph. Stat.* **2017**, *26*, 501–512.

[78] J. A. Hartigan, M. A. Wong, *J. R. Stat. Soc. Ser. C* **1979**, *28*, 100–108.

[79] A. Srinivasan, C. J. Galbán, T. D. Johnson, T. L. Chenevert, B. D. Ross, S. K. Mukherji, *Am. J. Neuroradiol.* **2010**, *31*, 736–740.

[80] M. Cuperlović-Culf, N. Belacel, A. S. Culf, I. C. Chute, R. J. Ouellette, I. W. Burton, T. K. Karakach, J. A. Walter, *Magn. Reson. Chem.* **2009**, *47*, S96–104.

[81] L. Kaufman, P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley, Hoboken, **2009**.

[82] J. Shi, J. Malik, *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 888–905.

[83] P. Tiwari, M. Rosen, A. Madabhushi, *Med. Phys.* **2009**, *36*, 3927–3939.

[84] S. Cacciatore, C. Luchinat, L. Tenori, *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 5117–5122.

[85] S. Cacciatore, L. Tenori, C. Luchinat, P. R. Bennett, D. A. MacIntyre, *Bioinformatics* **2017**, *33*, 621–623.

[86] T. Kohonen, *Neural Netw.* **2013**, *37*, 52–65.

[87] H. Zheng, J. Ji, L. Zhao, M. Chen, A. Shi, L. Pan, Y. Huang, H. Zhang, B. Dong, H. Gao, *Oncotarget* **2016**, *7*, 59189–59198.

[88] P. Geladi, *Chemom. Intell. Lab. Syst.* **1992**, *15*, vii–viii.

[89] S. Wold, L. Eriksson, *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130.

[90] H. Abdi, *Wiley Interdiscip. Rev. Comput. Stat.* **2010**, *2*, 97–106.

[91] P. S. Gromski, H. Muhamadali, D. I. Ellis, Y. Xu, E. Correa, M. L. Turner, R. Goodacre, *Anal. Chim. Acta* **2015**, *879*, 10–23.

[92] J. Trygg, S. Wold, *J. Chemom.* **2002**, *16*, 119–128.

[93] T. Cover, P. Hart, *IEEE Trans. Inf. Theory* **1967**, *13*, 21–27.

[94] O. Beckonert, E. Bollard, T. M. D. Ebbels, H. C. Keun, H. Antti, E. Holmes, J. C. Lindon, J. K. Nicholson, *Anal. Chim. Acta* **2003**, *490*, 3–15.

[95] J. Schmidhuber, *Neural Netw.* **2015**, *61*, 85–117.

[96] F. M. Alakwaa, K. Chaudhary, L. X. Garmire, *J. Proteome Res.* **2018**, *17*, 337–347.

[97] J. Westerhuis, H. Hoefsloot, S. Smit, D. Vis, A. Smilde, E. van Velzen, J. van Duijnhoven, F. van Dorsten, *Metabolomics* **2008**, *4*, 81–89.

[98] D. S. Wishart, D. Tzur, C. Knox, R. Eisner, A. C. Guo, N. Young, D. Cheng, K. Jewell, D. Arndt, S. Sawhney, et al., *Nucleic Acids Res.* **2007**, *35*, D521–D526.

[99] D. S. Wishart, C. Knox, A. C. Guo, R. Eisner, N. Young, B. Gautam, D. D. Hau, N. Psychogios, E. Dong, S. Bouatra, et al., *Nucleic Acids Res.* **2009**, *37*, D603–D610.

[100] D. S. Wishart, T. Jewison, A. C. Guo, M. Wilson, C. Knox, Y. Liu, Y. Djoumbou, R. Mandal, F. Aziat, E. Dong, et al., *Nucleic Acids Res.* **2013**, *41*, D801–D807.

[101] A. C. Dona, M. Kyriakides, F. Scott, E. A. Shephard, D. Varshavi, K. Veselkov, J. R. Everett, *Comput. Struct. Biotechnol. J.* **2016**, *14*, 135–153.

[102] K. Bingol, L. Bruschweiler-Li, C. Yu, A. Somogyi, F. Zhang, R. Brüschweiler, *Anal. Chem.* **2015**, *87*, 3864–3870.

[103] R. M. Boiteau, D. W. Hoyt, C. D. Nicora, H. A. Kinmonth-Schultz, J. K. Ward, K. Bingol, *Metabolites* **2018**, https://doi.org/10.3390/metabo8010008.

[104] J. Hao, M. Liebeke, W. Astle, M. De Iorio, J. G. Bundy, T. M. D. Ebbels, *Nat. Protoc.* **2014**, *9*, 1416–1427.

[105] P. J. C. Tardivel, C. Canlet, G. Lefort, M. Tremblay-Franco, L. Debrauwer, D. Concordet, R. Servien, *Metabolomics* **2017**, *13*, 109.

[106] M. Neuhäuser, in *Int. Encycl. Stat. Sci.*, Springer, Berlin, Heidelberg, **2011**, pp. 1656–1658.

[107] G. W. Snedecor, W. G. Cochran, *Statistical Methods*, Iowa State University Press, Ames, Iowa, **1989**.

[108] W. H. Kruskal, W. A. Wallis, *J. Am. Stat. Assoc.* **1952**, *47*, 583–621.

[109] W. J. Conover, *Practical Nonparametric Statistics*, 3rd ed., Wiley, New York, **1999**.

[110] C. E. Bonferroni, in *Studi Onore Profr. Salvatore Ortu Carboni*, Bardi, Rome, **1935**, pp. 13–60.

[111] Y. Benjamini, Y. Hochberg, *J. R. Stat. Soc. Ser. B* **1995**, *57*, 289–300.

[112] B. Vidgen, T. Yasseri, *Front. Phys.* **2016**, https://doi.org/10.3389/fphy.2016.00006.

[113] G. M. Sullivan, R. Feinn, *J. Grad. Med. Educ.* **2012**, *4*, 279–282.

[114] E. Saccenti, G. Menichetti, V. Ghini, D. Remondini, L. Tenori, C. Luchinat, *J. Proteome Res.* **2016**, *15*, 3298–3307.

[115] E. Saccenti, M. Suarez-Diez, C. Luchinat, C. Santucci, L. Tenori, *J. Proteome Res.* **2015**, *14*, 1101–1111.

[116] M. Suarez-Diez, J. Adam, J. Adamski, S. A. Chasapi, C. Luchinat, A. Peters, C. Prehn, C. Santucci, A. Spyridonidis, G. A. Spyroulias, et al., *J. Proteome Res.* **2017**, *16*, 2547–2559.

[117] A. Rosato, L. Tenori, M. Cascante, P. R. De Atauri Carulla, V. A. P. Martins Dos Santos, E. Saccenti, *Metabolomics* **2018**, *14*, 37.

[118] G. J. Patti, O. Yanes, G. Siuzdak, *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 263–269.

[119] M. Assfalg, I. Bertini, D. Colangiuli, C. Luchinat, H. Schäfer, B. Schütz, M. Spraul, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 1420–1424.

[120] E. Holmes, R. L. Loo, J. Stamler, M. Bictash, I. K. S. Yap, Q. Chan, T. Ebbels, M. De Iorio, I. J. Brown, K. A. Veselkov, et al., *Nature* **2008**, *453*, 396–400.

[121] P. Bernini, I. Bertini, C. Luchinat, S. Nepi, E. Saccenti, H. Schäfer, B. Schütz, M. Spraul, L. Tenori, *J. Proteome Res.* **2009**, *8*, 4264–4271.

[122] V. Ghini, E. Saccenti, L. Tenori, M. Assfalg, C. Luchinat, *J. Proteome Res.* **2015**, *14*, 2951–2962.

[123] E. Saccenti, G. Menichetti, V. Ghini, D. Remondini, L. Tenori, C. Luchinat, *J. Proteome Res.* **2016**, *15*, 3298–3307.

[124] E. Saccenti, L. Tenori, P. Verbruggen, M. E. Timmerman, J. Bouwman, J. van der Greef, C. Luchinat, A. K. Smilde, *PLOS ONE* **2014**, *9*, e106077.

[125] S. Wallner-Liebmann, E. Gralka, L. Tenori, M. Konrad, P. Hofmann, M. Dieber-Rotheneder, P. Turano, C. Luchinat, K. Zatloukal, *Genes Nutr.* **2015**, *10*, 441.

[126] N. A. Yousri, G. Kastenmüller, C. Gieger, S.-Y. Shin, I. Erte, C. Menni, A. Peters, C. Meisinger, R. P. Mohney, T. Illig, et al., *Metabolomics* **2014**, *10*, 1005–1017.

[127] P. Martinez-Lozano Sinues, M. Kohler, R. Zenobi, *PLoS ONE* **2013**, *8*, e59909.

[128] S. Krug, G. Kastenmüller, F. Stückler, M. J. Rist, T. Skurk, M. Sailer, J. Raffler, W. Römisch-Margl, J. Adamski, C. Prehn, et al., *FASEB J.* **2012**, *26*, 2607–2619.

[129] G. Nicholson, M. Rantalainen, A. D. Maher, J. V. Li, D. Malmodin, K. R. Ahmadi, J. H. Faber, I. B. Hallgrímsdóttir, A. Barrett, H. Toft, et al., *Mol. Syst. Biol.* **2011**, *7*, 525.

[130] C. Stella, B. Beckwith-Hall, O. Cloarec, E. Holmes, J. C. Lindon, J. Powell, F. van der Ouderaa, S. Bingham, A. J. Cross, J. K. Nicholson, *J. Proteome Res.* **2006**, *5*, 2780–2788.

[131] J. Pinto, A. S. Barros, M. R. M. Domingues, B. J. Goodfellow, E. Galhano, C. Pita, M. D. C. Almeida, I. M. Carreira, A. M. Gil, *J. Proteome Res.* **2015**, *14*, 1263–1274.

[132] S. O. Diaz, J. Pinto, A. S. Barros, E. Morais, D. Duarte, F. Negrao, C. Pita, M. D. C. Almeida, I. M. Carreira, M. Spraul, et al., *J. Proteome Res.* **2016**, *15*, 311–325.

[133] A. Scalabre, E. Jobard, D. Demède, S. Gaillard, C. Pontoizeau, P. Mouriquand, B. Elena-Herrmann, P.-Y. Mure, *J. Proteome Res.* **2017**, *16*, 3732–3740.

[134] D. A. Rodríguez, G. Alcarraz-Vizán, S. Díaz-Moralli, M. Reed, F. P. Gómez, F. Falciani, U. Günther, J. Roca, M. Cascante, *Metabolomics* **2012**, *8*, 508–516.

[135] J. T. Bjerrum, Y. Wang, F. Hao, M. Coskun, C. Ludwig, U. Günther, O. H. Nielsen, *Metabolomics* **2015**, *11*, 122–133.

[136] W. M. Claudino, A. Quattrone, L. Biganzoli, M. Pestrin, I. Bertini, L. A. Di, *J. Clin. Oncol.* **2007**, *25*, 2840–2846.

[137] J. L. Spratlin, N. J. Serkova, S. G. Eckhardt, *Clin. Cancer Res.* **2009**, *15*, 431–440.

[138] J. Carrola, C. M. Rocha, A. S. Barros, A. M. Gil, B. J. Goodfellow, I. M. Carreira, J. Bernardo, A. Gomes, V. Sousa, L. Carvalho, et al., *J. Proteome Res.* **2011**, *10*, 221–230.

[139] A. Hasim, H. Ma, B. Mamtimin, A. Abudula, M. Niyaz, L.-W. Zhang, J. Anwer, I. Sheyhidin, *Mol. Biol. Rep.* **2012**, *39*, 8955–8964.

[140] S. Tiziani, A. Lodi, F. L. Khanim, M. R. Viant, C. M. Bunce, U. L. Günther, *PLoS ONE* **2009**, https://doi.org/10.1371/journal.pone.0004251.

[141] Y. Yang, Y. Liu, L. Zheng, Q. Zhang, Q. Gu, L. Wang, L. Wang, *Int. J. Biol. Sci.* **2015**, *11*, 595–603.

[142] M. Deidda, C. Piras, P. P. Bassareo, G. Cadeddu Dessalvi, G. Mercuro, *IJC Metab. Endocr.* **2015**, *9*, 31–38.

[143] L. Tenori, X. Hu, P. Pantaleo, B. Alterini, G. Castelli, I. Olivotto, I. Bertini, C. Luchinat, G. F. Gensini, *Int. J. Cardiol.* **2013**, *168*, e113–e115.

[144] J. T. Brindle, J. K. Nicholson, P. M. Schofield, D. J. Grainger, E. Holmes, *Analyst* **2003**, *128*, 32–36.

[145] J. T. Brindle, H. Antti, E. Holmes, G. Tranter, J. K. Nicholson, H. W. L. Bethell, S. Clarke, P. M. Schofield, E. McKilligin, D. E. Mosedale, et al., *Nat. Med.* **2002**, *8*, 1439–1444.

[146] P. Bernini, I. Bertini, C. Luchinat, L. Tenori, A. Tognaccini, *J. Proteome Res.* **2011**, *10*, 4983–4992.

[147] V.-P. Mäkinen, T. Tynkkynen, P. Soininen, T. Peltola, A. J. Kangas, C. Forsblom, L. M. Thorn, K. Kaski, R. Laatikainen, M. Ala-Korpela, et al., *J. Proteome Res.* **2012**, *11*, 1782–1790.

[148] S. Deja, E. Barg, P. Młynarz, A. Basiak, E. Willak-Janc, *J. Pharm. Biomed. Anal.* **2013**, *83*, 43–48.

[149] X. Liu, J. Gao, J. Chen, Z. Wang, Q. Shi, H. Man, S. Guo, Y. Wang, Z. Li, W. Wang, *Sci. Rep.* **2016**, *6*, 30785.

[150] C. Dani, C. Bresci, E. Berti, S. Ottanelli, G. Mello, F. Mecacci, R. Breschi, X. Hu, L. Tenori, C. Luchinat, *J. Matern.-Fetal Neonatal Med.* **2014**, *27*, 537–542.

[151] E. Gralka, C. Luchinat, L. Tenori, B. Ernst, M. Thurnheer, B. Schultes, *Am. J. Clin. Nutr.* **2015**, *102*, 1313–1322.

[152] R. Calvani, E. Brasili, G. Praticò, F. Sciubba, M. Roselli, A. Finamore, F. Marini, E. Marzetti, A. Miccheli, *J. Clin. Gastroenterol.* **2014**, *48*, S5–7.

[153] P. Elliott, J. M. Posma, Q. Chan, I. Garcia-Perez, A. Wijeyesekera, M. Bictash, T. M. D. Ebbels, H. Ueshima, L. Zhao, L. van Horn, et al., *Sci. Transl. Med.* **2015**, *7*, 285ra62.

[154] P. Würtz, A. J. Kangas, P. Soininen, D. A. Lawlor, G. Davey Smith, M. Ala-Korpela, *Am. J. Epidemiol.* **2017**, *186*, 1084–1096.

[155] N. J. Rankin, D. Preiss, P. Welsh, K. E. V. Burgess, S. M. Nelson, D. A. Lawlor, N. Sattar, *Atherosclerosis* **2014**, *237*, 287–300.

[156] G. F. Giskeødegård, T. S. Madssen, L. R. Euceda, M.-B. Tessem, S. A. Moestue, T. F. Bathen, *NMR Biomed.* **2018**, e3927.

[157] J. R. Everett, *Prog. Nucl. Magn. Reson. Spectrosc.* **2017**, *102–103*, 1–14.

[158] I. Bertini, A. Calabrò, V. De Carli, C. Luchinat, S. Nepi, B. Porfirio, D. Renzi, E. Saccenti, L. Tenori, *J. Proteome Res.* **2009**, *8*, 170–177.

[159] I. Bertini, S. Cacciatore, B. V. Jensen, J. V. Schou, J. S. Johansen, M. Kruhoffer, C. Luchinat, D. L. Nielsen, P. Turano, *Cancer Res.* **2012**, *72*, 356–364.

[160] C. D. Hart, A. Vignoli, L. Tenori, G. L. Uy, T. Van To, C. Adebamowo, S. M. Hossain, L. Biganzoli, E. Risi, R. R. Love, et al., *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **2017**, *23*, 1422–1431.

[161] T. Andrew Clayton, J. C. Lindon, O. Cloarec, H. Antti, C. Charuel, G. Hanton, J.-P. Provost, J.-L. Le Net, D. Baker, R. J. Walley, et al., *Nature* **2006**, *440*, 1073–1077.

[162] P. Bernini, I. Bertini, A. Calabrò, G. la Marca, G. Lami, C. Luchinat, D. Renzi, L. Tenori, *J. Proteome Res.* **2011**, *10*, 714–721.

[163] M. Čuperlović-Culf, *NMR Metabolomics in Cancer Research*, Elsevier, Amsterdam, **2012**.

[164] M. A. C. Reed, R. Singhal, C. Ludwig, J. B. Carrigan, D. G. Ward, P. Taniere, D. Alderson, U. L. Günther, *Neoplasia* **2017**, *19*, 165–174.

[165] M. S. Monteiro, A. S. Barros, J. Pinto, M. Carvalho, A. S. Pires-Luís, R. Henrique, C. Jerónimo, M. D. L. Bastos, A. M. Gil, D. P. Guedes, *Sci. Rep.* **2016**, https://doi.org/10.1038/srep37275.

[166] C. M. Rocha, A. S. Barros, B. J. Goodfellow, I. M. Carreira, A. Gomes, V. Sousa, J. Bernardo, L. Carvalho, A. M. Gil, I. F. Duarte, *Carcinogenesis* **2015**, *36*, 68–75.

[167] A. Fages, T. Duarte-Salles, M. Stepien, P. Ferrari, V. Fedirko, C. Pontoizeau, A. Trichopoulou, K. Aleksandrova, A. Tjønneland, A. Olsen, et al., *BMC Med.* **2015**, *13*, 242.

[168] M. Stenson, A. Pedersen, S. Hasselblom, H. Nilsson-Ehle, B. G. Karlsson, R. Pinto, P.-O. Andersson, *Leuk. Lymphoma* **2016**, *57*, 1814–1822.

[169] M. S. A. Palmnas, H. J. Vogel, *Metabolites* **2013**, *3*, 373–396.

[170] E. Jobard, C. Pontoizeau, B. J. Blaise, T. Bachelot, B. Elena-Herrmann, O. Trédan, *Cancer Lett.* **2014**, *343*, 33–41.

[171] V. M. Asiago, L. Z. Alvarado, N. Shanaiah, G. A. N. Gowda, K. Owusu-Sarfo, R. A. Ballas, D. Raftery, *Cancer Res.* **2010**, *70*, 8309–8318.

[172] J. S. Choi, H.-M. Baek, S. Kim, M. J. Kim, J. H. Youk, H. J. Moon, E.-K. Kim, Y. K. Nam, *PLoS ONE* **2013**, https://doi.org/10.1371/journal.pone.0083866.

[173] C. Oakman, L. Tenori, L. Biganzoli, L. Santarpia, S. Cappadona, C. Luchinat, L. A. Di, *Int. J. Biochem. Cell Biol.* **2011**, *43*, 1010–1020.

[174] C. Oakman, L. Tenori, W. M. Claudino, S. Cappadona, S. Nepi, A. Battaglia, P. Bernini, E. Zafarana, E. Saccenti, M. Fornier, et al., *Ann. Oncol.* **2011**, *22*, 1295–1301.

[175] E. Jobard, O. Trédan, T. Bachelot, A. M. Vigneron, C. M. Aït-Oukhatar, M. Arnedos, M. Rios, J. Bonneterre, V. Diéras, M. Jimenez, et al., *Oncotarget* **2017**, *8*, 83570–83584.

[176] A. McCartney, A. Vignoli, C. Hart, L. Tenori, C. Luchinat, L. Biganzoli, A. Di Leo, *Breast* **2017**, *34*, S13–S18.

[177] L. Tenori, C. Oakman, P. G. Morris, E. Gralka, N. Turner, S. Cappadona, M. Fornier, C. Hudis, L. Norton, C. Luchinat, et al., *Mol. Oncol.* **2015**, *9*, 128–139.

[178] E. E. Tripoliti, T. G. Papadopoulos, G. S. Karanasiou, K. K. Naka, D. I. Fotiadis, *Comput. Struct. Biotechnol. J.* **2017**, *15*, 26–47.

[179] K. A. Mercier, M. Al-Jazrawe, R. Poon, Z. Acuff, B. Alman, *Sci. Rep.* **2018**, *8*, 584.

[180] A. L. Merz, N. J. Serkova, *Biomark. Med.* **2009**, *3*, 289–306.

[181] L. Zhang, E. Hatzakis, A. D. Patterson, *Curr. Pharmacol. Rep.* **2016**, *2*, 231–240.

[182] J. B. Carrigan, M. A. C. Reed, C. Ludwig, F. L. Khanim, C. M. Bunce, U. L. Günther, *ChemPlusChem* **2016**, *81*, 453–459.

[183] A. Basoglu, N. Baspinar, L. Tenori, X. Hu, R. Yildiz, *Metabolomics: open access* **2014**, *4*, 134.

[184] A. Basoglu, N. Baspinar, L. Tenori, A. Vignoli, R. Yildiz, *Metabolomics* **2016**, *12*, 128.

[185] A. Basoglu, N. Baspinar, L. Tenori, A. Vignoli, E. Gulersoy, *Biol. Trace Elem. Res.* **2017**, 1–8.

[186] A. Basoglu, I. Sen, G. Meoni, L. Tenori, A. Naseri, "NMR-Based Plasma Metabolomics at Set Intervals in Newborn Dairy Calves with Severe Sepsis," https://doi.org/10.1155/2018/8016510 can be found under https://www.hindawi.com/journals/mi/2018/8016510/, **2018**.

[187] E. Dervishi, G. Zhang, R. Mandal, D. S. Wishart, B. N. Ametaj, *Anim. Int. J. Anim. Biosci.* **2018**, *12*, 1050–1059.

[188] M. Caracausi, V. Ghini, C. Locatelli, M. Mericio, A. Piovesan, F. Antonaros, M. C. Pelleri, L. Vitale, R. A. Vacca, F. Bedetti, et al., *Sci. Rep.* **2018**, *8*, 2977.

[189] R. Schicho, R. Shaykhutdinov, J. Ngo, A. Nazyrova, C. Schneider, R. Panaccione, G. G. Kaplan, H. J. Vogel, M. Storr, *J. Proteome Res.* **2012**, *11*, 3344–3357.

[190] K. E. Cano, L. Li, S. Bhatia, R. Bhatia, S. J. Forman, Y. Chen, *J. Proteome Res.* **2011**, *10*, 2873–2881.

[191] M. Kanehisa, S. Goto, *Nucleic Acids Res.* **2000**, *28*, 27–30.

[192] J. Xia, D. S. Wishart, *Nucleic Acids Res.* **2010**, *38*, W71–W77.

[193] M. Ramirez-Gaona, A. Marcu, A. Pon, J. Grant, A. Wu, D. S. Wishart, *J. Vis. Exp.* **2017**, https://doi.org/10.3791/54869.

[194] J. Xia, D. S. Wishart, *Curr. Protoc. Bioinforma.* **2016**, *55*, 14.10.1–14.10.91.

[195] T. W.-M. Fan, P. K. Lorkiewicz, K. Sellers, H. N. B. Moseley, R. M. Higashi, A. N. Lane, *Pharmacol. Ther.* **2012**, *133*, 366–391.

[196] A. N. Lane, T. W.-M. Fan, M. Bousamra, R. M. Higashi, J. Yan, D. M. Miller, *OMICS J. Integr. Biol.* **2011**, *15*, 173–182.

[197] R. C. Bruntz, A. N. Lane, R. M. Higashi, T. W.-M. Fan, *J. Biol. Chem.* **2017**, *292*, 11601–11609.

[198] A. Le, A. N. Lane, M. Hamaker, S. Bose, A. Gouw, J. Barbi, T. Tsukamoto, C. J. Rojas, B. S. Slusher, H. Zhang, et al., *Cell Metab.* **2012**, *15*, 110–121.

[199] A. L. Oberg, O. Vitek, *J. Proteome Res.* **2009**, *8*, 2144–2156.

[200] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, L. Erlbaum Associates, Mahwah, NJ, **1988**.

[201] M. McHugh, *Biochem. Medica* **2008**, 263–274.

[202] A. Vignoli, L. Tenori, C. Luchinat, E. Saccenti, *J. Proteome Res.* **2018**, *17*, 97–107.

[203] "Food & Beverages—worldwide | Statista Market Forecast," can be found under https://www.statista.com/outlook/253/100/food-beverages/worldwide.

[204] "Wellness Now a $3.72 Trillion Global Industry," can be found under https://www.globalwellnessinstitute.org/wellness-now-a-372-trillion-global-industry/.

[205] A. Cifuentes, *J. Chromatogr. A* **2009**, *1216*, 7109.

[206] F. Savorani, M. A. Rasmussen, M. S. Mikkelsen, S. B. Engelsen, *Food Res. Int.* **2013**, *54*, 1131–1145.

[207] M. Zotti, L. D. Coco, S. A. D. Pascali, D. Migoni, S. Vizzini, G. Mancinelli, F. P. Fanizzi, *Heliyon* **2016**, *2*, e00075.

[208] Q. Gao, G. Praticò, A. Scalbert, G. Vergères, M. Kolehmainen, C. Manach, L. Brennan, L. A. Afman, D. S. Wishart, C. Andres-Lacueva, et al., *Genes Nutr.* **2017**, *12*, 34.

[209] M. Rådjursöga, G. B. Karlsson, H. M. Lindqvist, A. Pedersen, C. Persson, R. C. Pinto, L. Ellegård, A. Winkvist, *Food Chem.* **2017**, *231*, 267–274.

[210] D. Cukrov, M. Zemiani, S. Brizzolara, A. Cestaro, F. Licausi, C. Luchinat, C. Santucci, C. Tenori, H. Van Veen, A. Zuccolo, et al., *Front. Plant Sci.* **2016**, https://doi.org/10.3389/fpls.2016.00146.

[211] C. Santucci, S. Brizzolara, L. Tenori, *Food Anal. Methods* **2015**, *8*, 2135–2140.

[212] M. Spraul, B. Schütz, E. Humpfer, M. Mörtter, H. Schäfer, S. Koswig, P. Rinke, *Magn. Reson. Chem.* **2009**, *47*, S130–S137.

[213] A. P. Minoja, C. Napoli, *Food Res. Int.* **2014**, *63*, 126–131.

[214] E. Humpfer, B. Schütz, F. Fang, C. Cannet, M. Mörtter, H. Schäfer, M. Spraul, *Magn. Reson. Food Sci.* **2015**, 77–83.

[215] L. Tenori, C. Santucci, G. Meoni, V. Morrocchi, G. Matteucci, C. Luchinat, *Food Res. Int.* **2018**, https://doi.org/10.1016/j.foodres.2018.06.066.

[216] P. G. Takis, L. Tenori, E. Ravera, C. Luchinat, *Anal. Chem.* **2017**, *89*, 1054–1058.

[217] B. Zhang, J. Yuan, R. Brüschweiler, *Chem. Eur. J.* **2017**, *23*, 9239–9243.

[218] T. Konz, E. Migliavacca, L. Dayon, G. Bowman, A. Oikonomidi, J. Popp, S. Rezzi, *J. Proteome Res.* **2017**, *16*, 2080–2090.