

Cite this: *Mol. Omics*, 2024,  
20, 296

# Implementation of multiomic mass spectrometry approaches for the evaluation of human health following environmental exposure

Christina R. Ferreira,<sup>1</sup> Paulo Clairmont F. de Lima Gomes,<sup>2</sup>  
Kiley Marie Robison,<sup>3</sup> Bruce R. Cooper<sup>3</sup> and Jonathan H. Shannahan<sup>3</sup>

Omics analyses collectively refer to the possibility of profiling genetic variants, RNA, epigenetic markers, proteins, lipids, and metabolites. The most common analytical approaches used for detecting molecules present within biofluids related to metabolism are vibrational spectroscopy techniques, represented by infrared, Raman, and nuclear magnetic resonance (NMR) spectroscopies and mass spectrometry (MS). Omics-based assessments utilizing MS are rapidly expanding and being applied to various scientific disciplines and clinical settings. Most of the omics instruments are operated by specialists in dedicated laboratories; however, the development of miniature portable omics has made the technology more available to users for field applications. Variations in molecular information gained from omics approaches are useful for evaluating human health following environmental exposure and the development and progression of numerous diseases. As MS technology develops so do statistical and machine learning methods for the detection of molecular deviations from personalized metabolism, which are correlated to altered health conditions, and they are intended to provide a multi-disciplinary overview for researchers interested in adding multiomic analysis to their current efforts. This includes an introduction to mass spectrometry-based omics technologies, current state-of-the-art capabilities and their respective strengths and limitations for surveying molecular information. Furthermore, we describe how knowledge gained from these assessments can be applied to personalized medicine and diagnostic strategies.

Received 19th October 2023,  
Accepted 22nd March 2024

DOI: 10.1039/d3mo00214d

rsc.li/molomics

## 1 Introduction

This review is intended to provide a description of sample types and techniques used for multiomics studies. It also features research, instrumentation and perspectives on personalized medicine. The authors' goal is to showcase the power of multiomic analyses and provide background materials for researchers interested in diving into combining datasets related to the different levels of biological information present in cells, tissues and organisms. Biofluids used for the molecular analysis of metabolism and for diagnostic purposes include blood (as well as serum and plasma), urine, saliva, skin sebum, cerebrospinal fluid, gut aspirate, bile, amniotic fluid, synovial fluid, exhaled breath or breath condensate, nasal secretions, intact tissue, and tissue extracts. For the purpose of this review,

we will focus on the usage of readily available and easily accessible biofluids, including blood, serum, plasma, saliva, sweat, skin sebum, and urine, for molecular phenotyping and health monitoring. Even though next generation DNA sequencing allows for the detection of genetic conditions in biofluids, the chemical composition of biofluids can be used as a “real-time” molecular phenotypical baseline due to its dynamic changes. Changes in the metabolite composition of diverse biofluids can potentially be utilized to detect health issues or a variety of exposure-related health consequences.<sup>1,2</sup> Information from existing omics databases, which contain molecular features associated with disease conditions and toxicity responses, can be used to understand metabolic mechanisms of diseases and interpret deviations from established baselines.<sup>3,4</sup> Currently, the greatest challenge in employing metabolic phenotyping is that large portions of the human metabolome composition are unknown.<sup>5,6</sup> The most utilized omics approaches are genomics, transcriptomics, proteomics, and metabolomics. The integration of multiple approaches can provide deeper and more comprehensive insight into complex biological processes. Multi-omics approaches fit well into the concept of precision medicine and mass spectrometry (MS) is the dominating analytical technology

<sup>a</sup> *Purdue Metabolite Profiling Facility, Purdue University, West Lafayette, IN 47907, USA. E-mail: cferrei@purdue.edu*<sup>b</sup> *Sao Paulo State University Julio de Mesquita Filho, Institute of Chemistry, Araraquara, Sao Paulo, Brazil*<sup>c</sup> *School of Health Sciences, College of Health and Human Sciences, Purdue University, West Lafayette, IN 47907, USA*

for the omics approaches that can monitor metabolic phenotypes, namely proteomics and metabolomics.<sup>7,8</sup> MS systems can also be applied to the screening of biological reactions, as well as to the establishment of the metabolic transformations of drugs and chemicals. Currently, drug development is largely based on omics approaches and pharmacokinetics studies are the basis of biosafety and efficacy studies. We discuss the portability aspect of MS, which is expected to become the implementation strategy for precision medicine. Lastly, we survey efforts and methods for dealing with data interrelationships among existing omics that are packaged as data analysis workflows and in dedicated software programs.

## 2 Multi-omics approaches

### 2.1 Genomics

The first human genome sequencing was reported in 2003. It was a 13 year-long project and approximately 21 300 genes were detected. Next-generation sequencing (NGS) has only been commercially available for about 12 years. Nonetheless, the meteoric increase in sequencing throughput with NGS has dramatically changed our understanding of our genome and ourselves. NGS has also reduced the cost of generating sequence data and a plethora of sequence-based methods for probing a genome have emerged using NGS as the readout and have been applied to many species. The price of genome sequencing significantly dropped from \$100 000 000 to \$1000. Gene mutations are not the only root causes identified for a disease. Multiple environmental factors that directly influence the metabolism have been found to play a crucial role in health. NGS methods have also entered the medical realm driven by short-read generation (150 bp), but new platforms have emerged and are now capable of generating long multi-kilobase reads. The latter platforms enable reference-independent genome assemblies and long-range haplotype generation. Rapid DNA and RNA sequencing is a mainstream technology in personalized medicine and will continue to have an increasing impact on biology and medicine.<sup>9</sup> NGS is currently established as a test method for germline (inherited) and somatic (acquired) genetic mutations in many clinical laboratories. For inherited diseases, testing for germline mutations may include targeted panel, whole exome, whole genome, or mitochondrial DNA sequencing.<sup>10,11</sup> Targeted panel testing, which varies between laboratories, is possible for a wide variety of inherited disorders such as immune deficiencies, bone marrow failure syndromes, blindness, deafness, mitochondrial disorders, renal disorders, neurologic disorders, connective tissue disorders, cardiomyopathies, and cancer predisposition syndromes, among others.<sup>12–19</sup> Targeted panels for genes associated with a clinical phenotype are usually the first line of testing for inherited disorders, while whole exome sequencing is reserved for cases in which targeted testing has been uninformative.<sup>20,21</sup> Targeted panels for cancer testing also vary between laboratories. Targeted panels may be broad, including genes responsible for both solid and hematologic malignancies, or may be more focused for a particular type of malignancy (such as myeloid

neoplasms).<sup>22</sup> Any given gene within a panel may be completely sequenced or only partially sequenced (*e.g.*, hotspot regions). For both germline and somatic testing, it is important to know the content of the targeted panels when deciding on using a test. Whole exome and whole genome sequencing are not currently used clinically for oncology testing. Several new applications for NGS have more recently moved into the clinical arena or are being actively researched for clinical use, including circulating tumor DNA testing, human leukocyte antigen (HLA) typing, microbial analysis, RNA sequencing and expression, and methylation. Some of these new uses of NGS may be facilitated by the unique advantages of new instruments.

### 2.2 Transcriptomics

In the last few decades, transcriptome profiling has been one of the most utilized approaches to investigate human diseases at the molecular level. Molecular biomarkers and therapeutic targets have been found for several human pathologies through the quantification of gene expression levels and allele-specific expression. Large scale transcriptomics can be performed in a single experiment and can be used to identify novel genes, splice isoforms, and fusion transcripts and to investigate the world of non-coding RNAs at an unprecedented level. RNA sequencing has also been employed in important projects, like ENCODE (Encyclopedia of DNA Elements) and TCGA (The Cancer Genome Atlas), to provide a snapshot of the transcriptome of dozens of cell lines and thousands of primary tumor specimens. Moreover, transcriptomics studies have also paved the way for the development of data integration approaches.<sup>23</sup> However, like any other experimental approach, transcriptomics has its limitations: it is an inappropriate method to identify genes with large impacts on adaptive responses to the environment because: (i) genes with large impacts on fitness are rare; (ii) a large change in gene expression does not necessarily equate to a large effect on fitness; and (iii) protein activity is most relevant to fitness, and mRNA abundance is an unreliable indicator of protein activity.<sup>24</sup>

### 2.3 Proteomics

Proteomics is the study of the interactions, function, composition, and structures of proteins and their cellular activities.<sup>25</sup> Proteomics provides a better understanding of the structure and function of the organism than genomics. However, it is much more complicated than genomics because the protein expression is altered according to time and environmental conditions.<sup>26</sup> It is estimated that there are almost one million human proteins, many of which contain some modifications such as post-translational modifications (PTMs). However, it is also estimated that the human genome codes for about 26 000–31 000 proteins.<sup>27</sup> There are a variety of proteomics techniques including one-dimensional (1D) and two-dimensional (2D) gel electrophoresis (2-DE), as well as gel-free high-throughput screening technologies such as multidimensional protein identification technology, stable isotope labeling with amino acids in cell culture, isotope-coded affinity tag, and isobaric tagging for relative and absolute quantitation. Shotgun



proteomics, 2D difference gel electrophoresis (2D-DIGE), and protein microarrays can be used to investigate tissues, organelles, and cells. Large-scale western blot assays, multiple reaction monitoring assays, and label-free quantification of high mass resolution liquid chromatography (LC)-tandem mass spectrometry (MS) are commonly used for high-throughput processing.<sup>28</sup> Limitations related to the use of proteomics for diagnosis are related to the fact the disease-related proteins are often present at low concentrations mixed with various other proteins of much higher abundance, which makes it more difficult to identify them. Another common drawback is nonspecific adsorption of non-target proteins onto the surface of biosensors. Enzyme-linked immunosorbent assay (ELISA) is commonly used for the detection of specific proteins in biofluids and typically employs antibodies that are raised in animals directed against specific biomarkers. Other technologies used to detect specific proteins are electrochemical immunoassays, surface enhanced Raman spectroscopy (SERS), flow cytometry and protein microarrays.<sup>29</sup> The main limitation of portable systems for specific protein detection is the cost and the difficulty in profiling panels of proteins that can be related to a metabolic baseline.

#### 2.4 Metabolomics

The composition of small molecule metabolites or chemicals that can be found in a cell, a tissue, an organism or even an environmental sample (such as in air or sewage) is defined as the metabolome. Metabolomics, which is the study of the metabolome, is one of the most recent branches of the omics sciences. What makes metabolomics so different is that it focuses on small molecules (*i.e.*, chemicals with a molecular weight less than 1500 Daltons), while the other omics fields focus on big molecules (*i.e.*, DNA, RNA, and proteins). Metabolomics became so interesting because metabolites are the downstream products arising from the collective activities of the genome, the transcriptome and the proteome interacting with their environment. In other words, the metabolome is the closest omics to a molecular phenotype.<sup>30</sup> Applications of metabolomics span a wide range of disciplines including health and various diseases, pharmacology, drug development, toxicology, environment, plants, and food and nutrition. However, most of the studies are focused on improving the mechanistic understanding, along with prevention, early diagnosis, and management of human health and diseases. Metabolism screening is fundamental in interpreting a patient's phenotype. Newborn bloodspot screening (NBS) for phenylketonuria was reported in 1959. In 2012 9.5 million babies were screened for inborn errors of metabolism. Current technology is tandem mass spectrometry due to the short analysis time, high sensitivity, and selectivity to quantify NBS metabolites in dried blood spots of several hundreds of samples per day. In some countries every newborn is screened. Other examples are therapeutic drug monitoring (TDM) for inflammatory bowel diseases (IBD) and synthetic opioids. The most common analytical approaches used for metabolomics are vibrational spectroscopy techniques, represented by infrared, Raman, and nuclear magnetic resonance (NMR) spectroscopies

and mass spectrometry (MS). Among these, NMR spectroscopy and MS are the two most commonly employed methods in the metabolomics field. MS is a highly sensitive method, and it enables the analysis of several hundreds to thousands of metabolites from a single measurement and on a routine basis. In MS analysis, often, metabolites can be directly analyzed using ambient ionization or can be subjected to chromatographic separation using liquid chromatography, gas chromatography, or capillary electrophoresis prior to detection. A variety of MS methods are often used for analysis of different classes of metabolites from the same samples to achieve a wider coverage of the metabolome. NMR spectroscopy is often used without combining with any sample preprocessing or separation techniques and provides data complementary to MS, mostly due to sensitivity issues. Peaks in the NMR spectra can be reliably assigned to specific metabolites especially from pure compounds and peak intensities are directly proportional to the number of contributing nuclei.<sup>31</sup> Vibrational spectroscopy (infrared and Raman spectroscopies) offers rapid, high-throughput, and non-destructive analysis of a wide range of samples through chemical “fingerprinting”. The basis of vibrational spectroscopy is the transitions between quantized vibrational energy states of molecules due to the interaction between the material and the radiation from a light source.<sup>32,33</sup> Vibrational spectroscopy includes infrared and Raman spectroscopies. Even though both the near-IR (12 500–4000  $\text{cm}^{-1}$ ) and mid-IR (4000–400  $\text{cm}^{-1}$ ) are part of the infrared spectrum, most medical researchers focus on the mid-IR part of the spectrum because the fundamental vibrations in the mid-IR region provide sharper bands and more information on disease diagnosis rather than the overtone and harmonic vibrations that are provided by the near-IR region.<sup>34,35</sup> Mid-IR spectroscopy is based on the interaction between the sample and the IR beam, which is absorbed by the functional groups in the sample that vibrate in stretching, bending, deformation modes or their combination, and provides the fingerprint characteristics of the chemical or biochemical substances in the sample.<sup>34</sup> A major hurdle for FT-IR spectroscopy is the interference of the water in the mid-IR region, which masks some key biochemical information, especially in the amide I (1650  $\text{cm}^{-1}$ ) and lipid (3000–3500  $\text{cm}^{-1}$ ) absorption regions, and the water absorption could inhibit the light from penetrating the sample.<sup>36,37</sup> There are several approaches to overcome the water problem including the removal of the pure or scaled water spectrum from the acquired spectrum, dehydrating the sample, using  $\text{D}_2\text{O}$  solution, or lowering the effective path length significantly by using the attenuated total reflectance (ATR) as a sampling technique.<sup>34,36,37</sup> Raman spectroscopy is an inelastic light-scattering phenomenon; the incident photon is irradiated on the sample, and the molecules scatter the light. Although most of the scattered light has the same frequency as the incident light, some of them have different frequencies due to the interaction between the oscillation of light and molecular vibration. This phenomenon is called Raman scattering and, unlike IR spectroscopy, Raman spectroscopy has a very weak water signal and minimal water interference, which is an advantage for the analysis of the biological samples.<sup>38</sup> Raman spectroscopy can offer direct measurements of biofluids and single cells and *in vitro*



or even *in vivo* fiber-optic sampling of bladder and prostate, esophagus, skin, cervix, and arteries.<sup>34</sup> Furthermore, Raman spectroscopy is a non-destructive and non-invasive (wavelength and power-dependent) technique; it requires minimal sample preparation and simultaneous detection of macromolecules suitable for chemical analysis, quantification, classification, and the imaging of biological samples.<sup>39</sup> On the other hand, the Raman effect is very weak, and only 1 in 10<sup>8</sup> photons undergo Raman scattering;<sup>40</sup> to overcome this drawback longer acquisition times could be used, which could cause damage to the sample due to the laser exposure.<sup>39</sup> The other method to amplify the inherent signal weakness of Raman spectroscopy is by using the surface-enhanced Raman scattering (SERS) technique. SERS uses nanoscale roughened metallic surfaces (typically gold or silver), which could greatly enhance the order of the Raman signal (10<sup>8</sup>).<sup>41</sup> The signal enhancement can increase even up to 10<sup>11</sup> with surface-enhanced resonance Raman spectroscopy (SERRS). Advancements in the Raman instrumentation along with the SERS phenomenon have boosted the application of Raman as a diagnosis tool.<sup>42</sup> The other hurdle for Raman spectroscopy is the fluorescence interference, which happens when visible wavelength lasers are used. Especially during *in vivo* analysis, the fluorescence background signal can dominate the fingerprint region of the spectra.<sup>34,39</sup> The fluorescence interference can be removed mathematically or by illuminating the sample with the laser beam for a long time as a pre-treatment (this process is also known as “bleaching” or “photo-bleaching”) or by using longer wavelength lasers (*i.e.*, 1064 nm).<sup>34</sup>

## 2.5 Lipidomics

The same analytical approaches used for metabolomics are also applied to the study of the molecular composition of lipids. Lipidomics is considered a sub-area of metabolomics. Lipids have a variety of cellular functions (fuel for cell growth, signaling molecules, stimulatory agents, and can have an inhibitory effect on enzymes). From a chemical point of view, lipids are a heterogeneous pool of compounds that contain either fatty acyl/alkyl, sphingosine, or isoprene moieties as their hydrophobic building blocks. In 2005, lipids were classified into eight categories: fatty acyls, glycerolipids, glycerophospholipids, sphingolipids, sterols, prenol lipids, saccharolipids, and polyketides.<sup>43</sup> Since lipids play a crucial role in many biological processes, any imbalance in their homeostasis can lead to serious conditions in living organisms, such as chronic inflammation, cardiovascular diseases, diabetes, and neurodegenerative diseases, to name just a few. The method of choice for the analysis of lipid molecules or their huge assemblies (known as the lipidome) is undoubtedly mass spectrometry, due to its sensitivity and specificity.<sup>44</sup> Because of the inherent chemical complexity of the lipidome and the consequent challenges associated with analyzing it, progress in the field of lipidomics lagged behind the progress made in other omics disciplines.<sup>45</sup> Even though chromatographic separation followed by high resolution mass spectrometry is commonly used for lipidomics studies, ‘shotgun lipidomics’, involving direct injection and infusion (*i.e.*, no sample separation), are informative, quantitative, and fast.<sup>46–48</sup> In recent years, desorption electrospray

ionization (DESI) has become a noteworthy option for direct infusion lipidomics. In a comparative study with LC-MS, it was shown that DESI-MS forms different adducts than LC-MS, but when adjusted for these different adducts, the mass spectra show a very high degree of correlation in the determined lipid composition.<sup>49</sup> The major advantage of DESI in lipidomics is the ability to use it in combination with a gas-phase technology named ion mobility.<sup>50</sup>

## 2.6 Exposomics

Health effects of a chemical depend on numerous factors beyond dosage. The concept of exposome was introduced in 2005 to study the health effects of cumulative environmental exposures and concomitant biological responses from conception until death.<sup>51–53</sup> Derived from the term exposure, the exposome is an omic-scale characterization of the nongenetic drivers of health and disease. The exposome represents a shift toward comprehensive exposure assessment by assessing multiple, co-occurring exposures that may be found at low concentrations, like real-life exposure conditions. It also promotes the understanding about the interactions of exposures with endogenous processes influencing their biological effects and enables the identification of critical windows of exposure over the life course. Nonetheless, the study of exposome is challenging due to the low levels of compounds of interest (Fig. 1). Because endo- and exogenous chemicals are simultaneously detected, metabolomics provides an integrated measurement to link exposure to internal dose, biological response, and disease pathobiology.<sup>54–56</sup> By not limiting detected analytes to those selected a priori, untargeted metabolomics greatly expands surveillance of environmental chemicals, detection of new xenobiotic metabolites, and identification of previously uncharacterized pollutants.<sup>57–61</sup> Curation of metabolomics data to provide confirmed identification of the chemicals associated with the mass spectral features represents a critical research need. Despite this limitation, the unbiased and global characterization of metabolic responses enables the generation of new hypotheses for delineating toxicological mechanisms underlying chemical exposure in model systems, as reviewed by Niedzwiecki *et al.*, 2019.<sup>62</sup> In humans, proteomic studies have identified

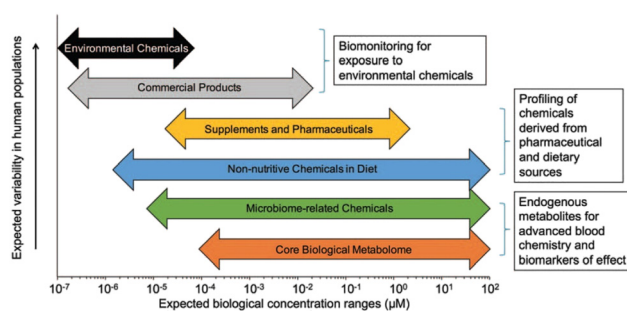


Fig. 1 Analytical sensitivity is important for measurement of endogenous metabolites, especially environmental chemicals, which are often present at four or five orders-of-magnitude lower abundance compared to endogenous metabolites. Published by Douglas *et al.*<sup>54</sup> after adaptation from Walker *et al.*<sup>67</sup> Adapted with copyright permission from Wolters Kluwer Health, Inc.



immune- and inflammation-related proteins associated with exposure to diesel exhaust<sup>63</sup> and polycyclic aromatic hydrocarbons (PAHs).<sup>64</sup> Continued development of multiplexed proteomics has considerable potential for characterizing biological responses, though traditional untargeted proteomics is challenging due to the difficulties in detecting low-abundance proteins in serum. Epigenomics is a key approach to evaluate exposure history and allostatic load.<sup>65,66</sup> In human cells, methylation of DNA occurs at the CpG dinucleotides in the cytosine C5 position. While tens of millions of CpG sites are present within the human genome, current high-throughput assays based on massively parallel sequencing of DNA with bisulfite conversions provide measures of up to 850 000 CpG sites. Epigenome-wide association studies have found distinct methylation patterns associated with chemical exposure, providing insight into mechanisms underlying biological responses and diseases.<sup>62</sup>

## 2.7 Microbiome

The gut microbiota is very diverse and contains many culturable and unculturable members that play critical roles in host health and disease. The members of the gut microbiota include archaea, bacteria, viruses, and fungi<sup>68</sup> and these organisms interact with each other and with the host. Metagenomic sequencing techniques have made it possible to study the microbial communities in the gut under different conditions and they help to detect alterations that occur during disease conditions. These techniques have been helpful in distinguishing healthy subjects from cancer,<sup>69</sup> inflammatory bowel disease,<sup>70</sup> and autism<sup>71</sup> patients. However, the presence of a microbe does not give any indication of its role in the gut. Also, the metabolic potentials of uncultured microbes are unknown, and this makes metagenomics data alone inadequate in providing information about the gut microbial ecology.<sup>72</sup> Meanwhile, as only the DNA of live and active microbes is transcribed into RNA, analyzing gut microbial mRNA (metatranscriptomics) has become a robust technique for detecting and quantifying transcribed mRNAs to predict their metabolic potential.<sup>73</sup> However, since not all mRNAs are translated into proteins, metaproteomics, an analytical technique that can analyze gut microbial proteins in samples, is usually used to detect and quantify such proteins.<sup>74</sup> Other microbial metabolites such as lipids, carbohydrates, and some other biomolecules have also been shown to be essential for microbe–host interaction.<sup>75</sup> For this reason, multi-omics approaches are increasingly being applied to identify gut microbial metabolites and host–microbe cometabolites, which may help unravelling the complex interaction between host and gut microbes.<sup>75,76</sup>

## 3 Examples of omics and multiomics studies focused on health and disease

### 3.1 Baseline vs. follow-up monitoring

The iPOP (integrated Personal Omics Profiling) study is an example of a current effort to establish a phenotypical baseline and use it for health status evaluation. The effort includes a longitudinal study of approximately 100 individuals meant to

help lay a foundation for precision personalized medicine through the unprecedented deep biochemical profiling of generally healthy individuals and understanding environmental conditions, such as seasonal influences (Fig. 2). It is designed to understand what “healthy” biochemical and physiological profiles look like at a personal level and what happens when people become ill. The study was designed and performed at Stanford University.<sup>54</sup>

Over the course of several years samples were collected from participants at regular intervals, both while they were in good health and in times of illness or significant stress. Whole genome sequencing was performed on all participants, and other omics data collected include information on how the genome is expressed (transcriptome, proteome, methylome), bacteria and other microorganisms in the gut and on the skin (microbiome), and the intermediate products of metabolism (metabolome). Data are also collected on participants’ diets, stress levels, activity levels, and personal and family medical history. Wearable devices enable tracking of participants physiology and activity. Altogether, billions of measurements were made every time someone was sampled.

Another effort, named the The Pioneer 100 Wellness Project (P100),<sup>78</sup> is presented in Fig. 3, which computed thousands of statistically significant inter-omic correlations using personal, dense, dynamic data clouds to identify many associations that

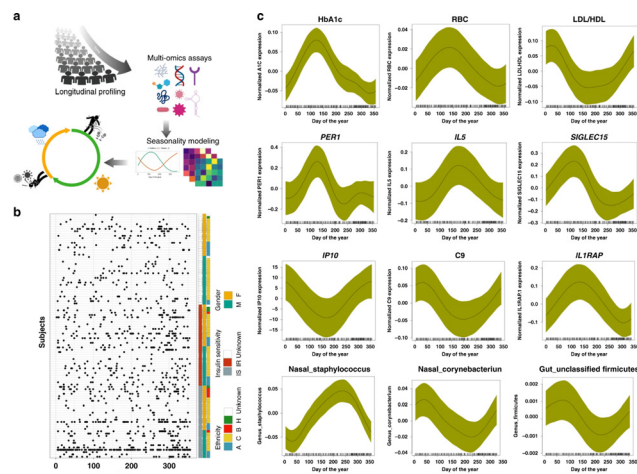
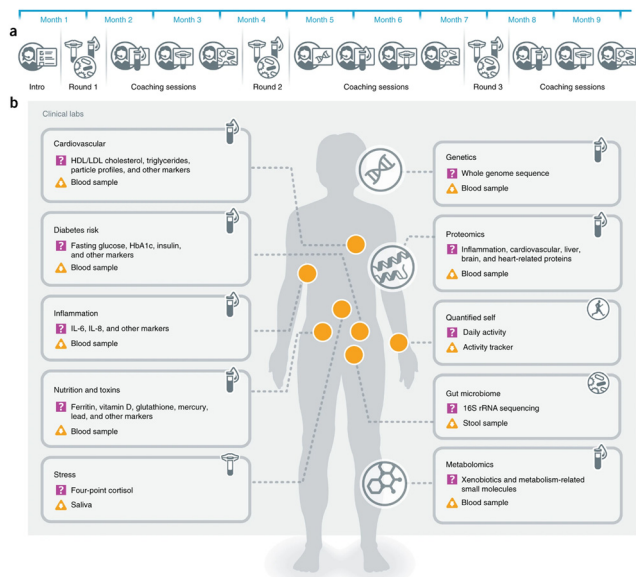


Fig. 2 Schematic view of a multi-omics effort named integrative personal omics profiling (iPOP) to study seasonal influences on the human body. (a) The omics assays included immune molecule profiling, proteomics, metabolomics, transcriptomics, and microbial profiling (gut and nasal), in conjunction with clinical lab tests and meteorological measurements. (b) Subjects and sampling timepoints for everyone, as well as ethnicity (A: Asian, B: black, C: Caucasian), insulin sensitivity (IS) and insulin resistance (IR), and gender information (M: male, F: female). (c) Examples of omics analytes with seasonal patterns (transcripts, cytokines, metabolites, proteins, clinical lab tests, gut, and nasal microbiome). The x-axis shows the days of the year (1–365 days) and y-axis shows the normalized expression/abundance values. The samples were collected up to 4 years and aggregated and mapped to 1-year-long time frame. The shaded area represents 95% confidence bounds computed as  $\pm 1.96$  standard deviation of model coefficients.<sup>54</sup> Adapted with copyright permission from Springer Nature Publishing.





**Fig. 3** (a) Timeline of important events and (b) schematic of the data collected for the generation and analysis of personal, dense, dynamic data clouds called the Pioneer 100 Wellness Project (P100). Personal data for 108 individuals were collected during a 9-month period, including whole genome sequences; clinical tests, metabolomes, proteomes, and microbiomes at three time points; and daily activity tracking. Using these data, a correlation network that revealed communities of related analytes associated with physiology and disease was generated. Connectivity within analyte communities enabled the identification of known and candidate biomarkers (e.g., gamma-glutamyltyrosine was densely interconnected with clinical analytes for cardiometabolic disease). Polygenic scores from genome-wide association studies (GWAS) for 127 traits and diseases were used to discover molecular correlates of polygenic risk (e.g., genetic risk for inflammatory bowel disease was negatively correlated with plasma cystine).<sup>78</sup> Adapted with copyright permission from Springer Nature Publishing.

could be followed up with perturbation experiments. The correlations were partitioned into data communities to establish biomarkers in context within biological networks. This approach led to the identification of putative biomarkers such as gamma-glutamyltyrosine for cardiometabolic disease. The clinical biomarker of many participants significantly changed regarding the disease background during the study (e.g., type 2 diabetes and cardiovascular risk factors). Together this study indicates that personal, dense, dynamic data clouds embody the essence of precision medicine and present possibilities for the discovery of important medical applications.<sup>78</sup>

### 3.2 Environmental exposure

In the United States, over 85 000 chemicals are registered with the EPA for manufacture, import, and use in commercial products. Additionally, approximately 40 000 pesticide formulations, 100 000 dietary phytochemicals, and 5000 other chemicals are approved for use as inert ingredients. Also 7500 compounds are registered by the US Food and Drug Administration as drugs or food additives. An individual's history of these exposures over a lifetime—that is, their chemical experience—may contribute directly to phenotype and health. In almost all cases, limited information is available about these

chemicals in terms of their distributions across populations, the health effects of low-level exposures, and the influence of complex mixtures encountered in real-world scenarios. The adequate characterization of an individual's chemical burden will require the ability to measure upwards of 1 million chemicals routinely across the lifespan in a cost-effective and efficient manner.<sup>79</sup> Exposure to environmental chemicals can initiate local and global changes in gene transcription, enzyme activity, metabolite pathway alterations, and protein synthesis/folding. As a result, micro- and macroscale interactions occur among these systems that can be characterized to study dose-response relationships. In 2011, the United States Institute of Medicine (IOM) recommended that the Department of Defense (DoD) collect individual breathing zone samples and conduct long-term studies of troop health outcomes to address concerns about perceived health risks resulting from exposure during deployment.<sup>80–82</sup> Realistically, there are inherent limits to exposure assessment in deployed settings. For example, the use of personal monitoring equipment limits mobility in active combat situations, logistics of sampler collection is challenging with large-scale troop movements, and assessment for biologically relevant dose requires additional molecular measurements. Furthermore, the post-exposure window of opportunity for measuring exposures or immediate consequences may range from hours to days for some agents. Therefore, valid and reliable measures are needed to characterize exposures that do not disrupt effective operation during deployment. Retrospective profiling of biological specimens collected pre- and post-deployment for biomarkers of exposure, effect, and susceptibility provides a means of assessing the occurrence of chemical exposure related to poor health outcomes. Through the DoD Serum Repository (DoDSR), an extensive system exists for collection, cataloguing, and storing of serum samples collected pre- and post-deployment from armed forces personnel.<sup>83,84</sup> Incorporating chemical screening measures using serum samples collected under the current DoDSR framework could therefore be completed with minimum disruption to military operations.<sup>54</sup>

Gas or liquid chromatography with ultrahigh-accuracy mass spectrometry is the most promising analytical technology for an exposome platform for precision medicine.<sup>54,85–88</sup> Due to increases in scan speed and data extraction algorithms, modern instruments can detect 20 000–100 000 unique chemical signals in small volumes (< 150  $\mu$ L). Including triplicate injections improves reliability of peak detection when studying exposures that occur in a small subset of the population. Combined with a technique known as reference standardization, MS can determine absolute concentrations of biomarkers for the assessment of potential risks from exposures.<sup>88</sup> Additionally, MS is cost-effective relative to other biomonitoring platforms.<sup>54</sup> Further cost reduction is possible through focused analysis of high-abundance metabolites and exposure markers. It reliably detects approximately 1000 common endogenous metabolites, commercial products, and drug metabolites with coefficient of variation (CV) less than 10%.<sup>85,87–89</sup> By limiting detection to chemical signals with low CVs, reducing runtimes, and employing automation, samples could theoretically



be processed with a throughput of 500 samples/day (125 000 samples/instrument-year) at a cost of \$5 per sample. Thus, sufficient chemical coverage for the purposes of precision medicine and the detection of environmental exposures and related bioeffects could be obtained at a low cost with available technology.

Transcriptomics, proteomics, metabolomics, and lipidomics data revealed cigarette smoke induced inflammatory and oxidative stress response, as well as lipid/surfactant alterations. However, at matched nicotine concentrations, aerosol exposure from carbon heated tobacco products and tobacco heating systems, these effects were either limited or absent as described by Titz *et al.*<sup>90</sup>

Herron *et al.*<sup>91</sup> used lipidomics and transcriptomics to demonstrate that benzalkonium chlorides (BACs) can cross the blood–placental barrier and embryonic blood–brain barrier, resulting in altered sterol and lipid homeostasis. When fetuses are exposed to BACs *in utero*, signaling pathways important for neuronal development, such as LXR/RXR and glutamatergic signaling, are negatively affected.

### 3.3 Metabolic diseases

Metabolic diseases including type 2 diabetes mellitus (T2DM), non-alcoholic fatty liver disease (NAFLD), and metabolic syndrome (MetS) are alarming health burdens around the world and examples of multi-omics conditions since they are multifactorial metabolic disorders based on the interactions between genetics and environment. The multiple components of these diseases have been recently reviewed by Hu and Jia.<sup>92</sup> Familial aggregation,<sup>93</sup> ethnic differences,<sup>94</sup> and higher concordance rate of T2DM in monozygotic than in dizygotic twins<sup>95</sup> indicate genetic contribution to T2DM. For NAFLD, there have been biochemical, imaging, genetic, and other omics biomarkers for its staging and progression.<sup>96</sup> A strong heritability of NAFLD susceptibility has been identified in epidemiological, family, and twin studies.<sup>97,98</sup> Studies on T2DM patients have identified transcriptional differences in islets, liver, muscle, adipose tissue, and peripheral blood using RNA sequencing transcriptomics studies on NAFLD mainly focusing on the liver. A differential expression analysis in severe *vs.* non-severe NAFLD and normal liver<sup>99</sup> showed 320 genes differentially expressed in severe NAFLD. Also, several studies identified epigenetic changes in T2DM patients, and the regions were also associated with differential expression of genes. High-fat diet (HFD) can induce modifications in the chromatin structure, thereby contributing to metabolic disease.<sup>100</sup> FAIRE-seq (a method in molecular biology used for determining the sequences of DNA regions in the genome associated with regulatory activity) was performed in the livers of C57BL/6J mice induced by HFD and control diet, which identified 28 484 open chromatin sites in control and 28 253 sites in high-fat livers. There are several proteins associated with incidence and progression of T2DM. The approach of constructing a model comprising a multiple serum biomarker seemed to be promising and critical for the detection, diagnosis, and prognosis of T2DM;<sup>101</sup> however, relevant findings have not been routinely

used in clinical laboratory tests. Moreover, it is challenging to characterize the broad and dynamic spectrum of serum proteins, especially in the case of low-abundance proteins. In the case of NAFLD, ApoE and lymphocyte cytosolic protein 1 (LCP1) were significantly upregulated, while IGFBP3 and vitamin D-binding protein were downregulated in patients with NASH compared with healthy subjects.<sup>102</sup> In the exploration of the mechanism underlying MetS, proteomics has enabled significant advances. Evidence indicates that hyperglycemia induces metabolic changes in  $\beta$  cells that markedly reduce mitochondrial metabolism and adenosine triphosphate (ATP) synthesis.<sup>103</sup> A study using phospho-proteomics revealed the glycogen synthase kinase 3-pancreatic and duodenal homeobox 1 axis as a key pathogenic signaling node in insulin secretion.<sup>104</sup> The adipose tissue proteins identified in proteomic studies addressing diabetes and insulin resistance mainly participate in energy and metabolism, immune response/inflammation, oxidative stress, cytoskeleton, and apoptosis/cell cycle.<sup>105–109</sup> Metabolic diseases including T2DM, NAFLD, and MetS comprise a series of pathway disturbances in carbohydrates, lipids, and proteins; therefore, metabolomics is quite feasible for studying these disorders 68 (Fig. 4).

Bowes *et al.*<sup>110</sup> used metabolomics and genomics to better understand the relationship between near real-time population dietary assessment and wastewater-based epidemiology. Community-scale datasets were made, displaying the relationship between human behavior and dietary indicators that are measurable in municipal wastewater and allowing for the association between wastewater-borne levels of phytoestrogens and composition of gut microbiota to be established. Metabolomics, proteomics, glycomics, and microbiomics are being

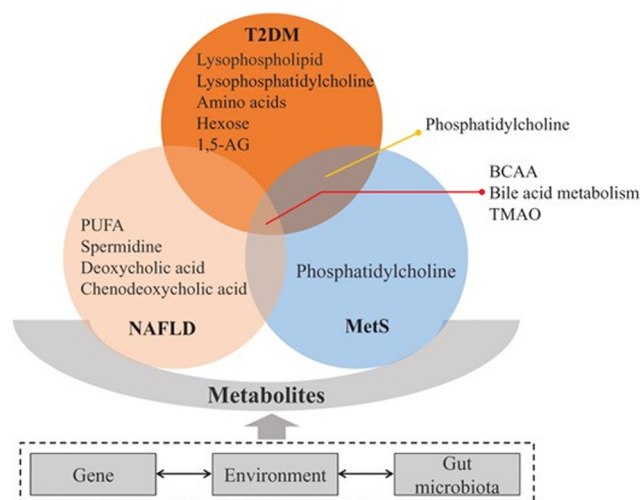


Fig. 4 Metabolites tightly connected with type 2 diabetes mellitus (T2DM), non-alcoholic fatty liver disease (NAFLD), and metabolic syndrome (MetS). These conditions are associated with insulin resistance, bile acid and lipid metabolism changes. Among these pathways, branched-chain amino acids (BCAA), bile acid metabolism and trimethylamine *N*-oxide (TMAO) are implicated in T2DM, NAFLD, and MetS. Phosphatidylcholine is associated with T2DM and MetS. Other metabolites have been related specifically to one of these diseases.<sup>92</sup> Adapted with copyright permission from Oxford University Press.



combined to better understand the infant gut microbiota and the metabolic impact of human milk on infants' health.<sup>111</sup>

### 3.4 Nutrition and microbiome

Dietary factors are the contributors to many diseases. This fact suggests that the personalization of dietary habits may have an impact on changing behavior and ultimate health outcomes.<sup>112</sup> Transcript profiling has been extensively used to evaluate the possible effects of anthocyanins on obesity related gene expression in adipocytes,<sup>113</sup> for biomarker identification<sup>114</sup> and also for designing precise mitigation strategies especially for ready-to-eat food products.<sup>115</sup> Several robust and nutrient-specific microRNAs (miRNAs) as indicators of nutritional stress have been reported in plants.<sup>116</sup>

Nutriproteomics is still a nascent research area, which exploits the proteomic tools to characterize molecular and cellular changes in expression of proteins and their interaction with other nutrients, as the bioavailability and functions of each nutrient including bioactive peptides and proteins can be influenced by the presence of other nutrients/compounds. Bioactive peptides and proteins derived from food in general exert multiple responses such as growth and homeostatic regulation and can even cause adverse allergic reactions in some cases.<sup>117</sup> Proteomics in nutrition can identify and quantify bioactive proteins and peptides and addresses their nutritional bioefficacy.<sup>118</sup> Application of proteomic techniques for determining food quality especially with respect to personalized nutrition is mainly done by analyzing the complete proteome or metabolome of food.<sup>119</sup> The proteomics approach may even be used in the post-marketing surveillance of foods derived from genetically modified crops<sup>120</sup> and in identification of bioactive compounds in nutraceuticals and functional foods,<sup>121</sup> apart from diagnosis and vaccine/drug development.<sup>122</sup> Metabolomics application to studies on dietary interventions allows a greater understanding of the effect of diet on metabolic changes, one's health and related disorders along with the relationship between the genotype and phenotype. For example, metabolomics has been used in different studies for evaluating metabolite profiles as a result of consuming fiber,<sup>123</sup> tea,<sup>124</sup> coffee,<sup>125</sup> fish oil,<sup>126</sup> and high-fat diet<sup>127</sup> and a large number of metabolic perturbations have been revealed. The metabolomics approach can be used for nutritional interventions, to identify dietary biomarkers, and for the development of personalized nutrition or medicine.<sup>128–130</sup>

Transcriptomics, non-serum metabolomics, and genomics were tools applied by Li *et al.*<sup>131</sup> to explore the mechanism and toxicological effect of inorganic arsenic exposure on the liver-microbiota-gut axis in chickens. Inorganic arsenic exposure was found to damage hepatic function-related serum biochemical indicators and alter liver transcription factors, resulting in the development of fibrosis and negatively altering the biodiversity of ileal microbiota. Liu *et al.*<sup>132</sup> utilized metabolomics and microbiomics to better understand the relationship between white matter structure, gut microbiota, and metabolites in infants born with low birth weight and white matter injury. It was found that infants of this group had significant

downregulation of metabolic pathways such as biosynthesis of arginine and primary bile acid, which results in white matter damage in the brain. The microbiota were found to be dysregulated, with an increase in *Klebsiella* sp. These specific bacteria are associated with pro-inflammatory responses within the gastrointestinal tract.

Bekiares *et al.*<sup>133</sup> used proteomics and microbiomics to examine the effects of sweetened dried cranberries on the urinary proteome and fecal microbiome. While there was not a statistically significant change in fecal microbiome, 22 proteins were found to have differences between pre- and post-treatment.

Also, proteomics, phosphoproteomics, and transcriptomics data demonstrated by Arumugam *et al.*<sup>134</sup> revealed the mechanisms of intermittent fasting and its beneficial effects on cardiac health and disease prevention. Intermittent fasting regimens modify cyclic GMP signaling, lipid and amino acid metabolism, cell adhesion, cell death, and inflammation. It was shown that shorter intermittent fasting regimens had a larger effect on pathway alteration in comparison to longer intermittent fasting regimens.

### 3.5 Infectious diseases and sepsis

Infection is defined as a pathologic process caused by the invasion of normally sterile tissue or fluid or body cavity by pathogenic or potentially pathogenic microorganisms. Biomarkers play a role in helping to identify—or perhaps more importantly rule out—an infection. Infection is not an all-or-none phenomenon, and there are “gray areas” where one can never really be certain that an infection was present or absent. Sepsis is defined as the presence of organ dysfunction occurring as the result of a dysregulated host response to an infection. Sepsis markers such as chemokines, coagulation system markers, endotoxin, lactate, and procalcitonin are usually more helpful in ruling out than ruling in an infection. This is particularly true in critically ill patients, who often have some inflammatory response, but do not always have infection or require antibiotic administration.<sup>135</sup> The application of metabolomics in infectious disease diagnostics is an evolving area of science that was boosted by the urgency of COVID-19 pandemic. Metabolomics approaches that rely on the analysis of volatile organic compounds exhaled by COVID-19 patients hold promise for applications involving a large-scale screening of population in point-of-care (POC) settings. On the other hand, successful application of mass-spectrometry to detect specific spectral signatures associated with COVID-19 in nasopharyngeal swab specimens may significantly save the cost and turnaround time of COVID-19 testing in the diagnostic microbiology and virology laboratories. Active research is also ongoing on the discovery of potential metabolomics-based prognostic markers for the disease that can be applied to serum or plasma specimens. Several metabolic pathways related to amino acid, lipid and energy metabolism were found to be affected by severe COVID-19. Tryptophan metabolism *via* the kynurenine pathway was persistently dysregulated in several independent studies, suggesting the roles of several



metabolites of this pathway such as tryptophan, kynurenine and 3-hydroxykynurenine as potential prognostic markers of the disease.<sup>136</sup> COVID-19 encompasses a spectrum of varying phenotypes where customized therapy may help more and harm less. Rello *et al.* (2020) have described 5 phenotypes ranging from the most benign (phenotype 1) to increasing respiratory distress and hypoxemia (phenotypes 2 and 3) and acute respiratory distress syndrome (ARDS) (phenotypes 4 and 5).<sup>137</sup> IL-6 has been suggested as a differentiating feature between phenotypes 2 and 3, and procalcitonin as a characteristic feature of phenotype 5. Defining phenotypes based on underlying risk factors, clinical and radiological features and biomarkers may help predicting the need for ICU admission and optimizing therapy. One study assessed the changes in biomarkers with supportive therapy and a variable combination of abidol, lopinavir/ritonavir and methylprednisolone.<sup>138</sup> After treatment, IL-2R, IL-6, TNF- $\alpha$ , and CRP levels decreased significantly, followed by IL-8, IL-10, and PCT. CD4+ and CD8+ T lymphocytes increased significantly but B lymphocytes and natural killer cells showed no changes. Serum ferritin also did not decrease significantly. D-dimer levels have been recommended as a part of the risk stratification criteria to decide anticoagulation.<sup>139</sup> Treatment with low molecular weight heparin (LMWH) is associated with reduction in levels of d-dimer and fibrin degradation products and also in IL-6 levels suggesting a potential anti-inflammatory effect.<sup>140</sup>

The lipidomics data revealed how the lipidome controls the immune response and its effect on sepsis severity depending on the COVID-19 status. Expression of inflammatory hubs responsible for restricting inflammation, such as ChoE-18:3, LPC-O-16:0, and PC-O-30:0, is decreased in patients with sepsis from COVID-19, while expression of inflammatory hubs responsible for enhancing inflammation, such as sPLA2, PGD2, and 12-HETE, is increased in patients with sepsis from COVID-19 as reported by Meng *et al.*<sup>141</sup> Data from a multi-omics approach using proteomics, metabolomics, and lipidomics were applied to build a workflow to predict aggravation of COVID-19 symptoms of patients in the ICU. Two proteins (CCL7 and CA14), as well as one lipid (HexCer 18:2; O2/20:0), were identified as short-term predictors of worsening COVID-19 progression in ICU patients as described by Kugler *et al.*<sup>142</sup>

Also, metabolomics and microbiomics profiling has been utilized recently by Bosnjak *et al.*<sup>143</sup> to characterize the fecal environment of patients with hospital acquired diarrhea before and after receiving antibiotic treatment in relation to the *C. difficile* infection status. It was found that *C. difficile* infection alters metabolic markers that result in antibiotic-associated dysbiosis and proliferation of opportunistic bacteria. Transcriptomics and proteomics were utilized by Noszka *et al.*<sup>144</sup> to better understand the HP1021 regulon and its relationship with *H. pylori*. HP1021 controls the response of *H. pylori* to oxidative stress, as well as DNA uptake and carbohydrate metabolism, directly.

Zeng *et al.*<sup>145</sup> reported that post analysis integration of transcriptomics and proteomics data of synovial cells and fibroblasts has been performed to better understand inflammation

pathways in which the drug celestrol impacts rheumatic arthritis cells by modulating inflammation, inhibiting chemokine pathways and osteoclast differentiation, and promoting synovial cell apoptosis.

### 3.6 Neurodegenerative diseases and inflammation

Biomarkers for neurodegenerative diseases are needed to improve the diagnostic workup in the clinic but also to facilitate the development and monitoring of effective disease-modifying therapies. Positron emission tomography methods for detecting amyloid- $\beta$  and tau pathology in Alzheimer's disease have been increasingly used to improve the design of clinical trials and observational studies. In recent years, easily accessible and cost-effective blood-based biomarkers used for detecting the same Alzheimer's disease pathologies have been developed, which might revolutionize the diagnostic workup of Alzheimer's disease globally. Relevant biomarkers for  $\alpha$ -synuclein pathology in Parkinson's disease are also emerging, as well as blood-based markers of general neurodegeneration and glial activation.<sup>146</sup> Several fluid biomarkers of neurodegeneration have recently emerged. Blood-based assays reveal brain  $\alpha$ -synuclein pathology and would considerably facilitate studies in larger populations, but skin biopsies might also provide an effective alternative. Analysis of CSF is likely to be central in this process because the levels of brain-derived proteins are much higher in CSF than in blood, where brain-derived molecules are diluted in a complex matrix of abundant plasma proteins, such as albumin and immunoglobulins. Currently, the most promising marker of neurodegenerative disease is the neurofilament light (NfL), which can be measured in both CSF and blood. This biomarker reflects axonal degeneration and injury, irrespective of cause, and the levels are especially increased in amyotrophic lateral sclerosis, frontotemporal dementia, and atypical parkinsonian disorders (that is, progressive supranuclear palsy, multiple system atrophy (MSA) and corticobasal degeneration).<sup>147,148</sup> However, NfL levels are also increased in Alzheimer's disease, and studies on autosomal dominant Alzheimer's disease show that the rate of change in blood NfL increased already about 15 years before symptom onset.<sup>149</sup> Importantly, higher levels of NfL are associated with faster disease progression and higher brain atrophy rates in most neurodegenerative disorders.<sup>147,149</sup> As a result, NfL can be regarded as a measure of the intensity of ongoing neurodegeneration. In several brain diseases, including multiple sclerosis and spinal muscular atrophy, effective disease-modifying treatments can normalize NfL levels, and reduction in NfL levels is associated with the clinical effectiveness of the treatment.<sup>150,151</sup>

Regarding inflammatory diseases, proteomics of synovial fluid and plasma revealed that there is a mild time series pattern of expression during osteoarthritis progression as reported by Anderson *et al.*<sup>152</sup> At the initial induction of osteoarthritis, there was a decrease in proteins responsible for signal transduction and regulation of signaling events at day 10, followed by an increase at day 63. On day 10, there was also an initial response in proteins that are responsible for conducting immune system responses. Overall, proteomics



data showed that an EV protein cargo is more important than a small non-coding RNA cargo during osteoarthritis progression.

Transcriptomics, proteomics, and phosphoproteomics methods were applied by Wei *et al.*<sup>153</sup> to better understand the inflammatory and metabolic pathways involved in diabetic kidney disease progression. Pathways found to have a significant effect on the disease include lipid metabolism, fatty acid metabolism, glycolysis, cell cycle regulation, phagocytosis and apoptosis regulation, and inflammatory response regulation. Genes such as ALOX15, known to enhance hypertrophy, fibrosis, and pro-inflammatory gene formation, and SERPINA1E, known to inhibit liver gluconeogenesis, were found to be downregulated, negatively affecting the identified pathways. Integrated transcriptomics and proteomics data were utilized by Pascual-Alonso *et al.*<sup>154</sup> to analyze human fibroblasts to study the molecular consequences of mutated MECP2 in individuals with Rett Syndrome. It was revealed that due to a loss in function of the mutated MECP2 protein, other genes and proteins responsible for neuronal development are downregulated, resulting in neuronal dysregulations, such as cytoskeletal organization, vesicular activity, and mRNA processing.

### 3.7 Respiratory diseases and cancer

Asthma is one of the most prevalent chronic airway diseases characterized by airway hyper-responsiveness, inflammation, and mucus secretion. It is one of the most prevalent chronic airway diseases, affecting approximately 339 million individuals worldwide, globally killing more than 1000 daily, and its incidence rises each year.<sup>155</sup> Among the largest investigations is a European genome-wide association (GWAS) study on allergic disease susceptibility conducted in 360 838 subjects, which identified 136 genetic variants to be associated with allergic disorders, including asthma, implicating 132 nearby genes from 99 loci.<sup>156</sup> Large-scale epigenome-wide association (EWAS) studies indicated that environmental exposures such as prenatal smoking and air pollution were associated with changes in DNA methylation patterns of several asthma-related genes.<sup>157,158</sup> The transcriptome profiles of different tissues/cells have provided significant insights into the role of gene expression in the asthma disease process. A study in healthy controls and mild, moderate, and severe asthmatics showed that CD3+ T cells isolated from sputum and bronchoalveolar lavage fluid (BALF) had a distinct transcriptome profile from endobronchial epithelial brushings.<sup>159</sup> Airway proteome profiles in asthma patients can be associated with different phenotypes. The importance of proteomics is that it can represent the active cellular state of different tissues/cells. However, adequate attention should be directed to tissue-associated proteome differences in comparative and possibly longitudinal studies. Sputum proteomics showed that 10 out of 1129 proteins were significantly different between four previously established clinical asthma clusters.<sup>160</sup> Moreover, the sputum proteome profiles of adult asthmatics were distinguished between current, ex-smokers, and nonsmokers.<sup>161</sup> Most conducted metabolomics studies in asthmatics focused on investigating metabolomics profiles in comparison to healthy controls and other respiratory disorders such as chronic

obstructive pulmonary disease (COPD) or distinguishing different asthma phenotypes. The main identified metabolites across different studies from various body compartments were related to immune reactions, inflammatory processes, tricarboxylic acid cycle, oxidative stress, hypoxia, and lipid metabolism pathways.<sup>162</sup> One of the emerging metabolomics techniques in asthma research is the measurement of volatile organic compounds (VOCs) in exhaled breath (breathomics).<sup>163,164</sup>

Regarding multi-omics analysis Li *et al.*<sup>165</sup> demonstrated that serum proteomics and metabolomics revealed the underlying pathogenesis of severe community-acquired pneumonia, which involves enhancing inflammatory pathways including Hippo and PI3K/Akt, leading to tissue damage resulting from overactivation of inflammatory signals, while decreasing immunoglobulins and suppressing the overall function of humoral immunity. Metabolomics revealed suppression of lipid metabolism with enhanced glycolysis and lactate production, resulting in disordered lipid metabolism. Qu *et al.*<sup>166</sup> integrated transcriptomics and genomics data of lung tissues to better understand the impact and underlying mechanisms of CD93 on prognosis of lung squamous cell carcinoma. Increased expression of CD93 has shown to result in upregulation of cell adhesion and angiogenesis pathways, suggesting that CD93 plays a significant role in the formation of the capillary network of primary tumors and increasing the likelihood of lymph node and distant metastasis. Increased expression of CD93 in endothelial cells also resulted in T cell dysfunction, inducing the local immune tolerance. Also Wang *et al.*<sup>167</sup> integrated transcriptomics to map the cellular subpopulations within the immune microenvironment of the brain to build a predictive model for intracranial aneurysms. M1/M2 type macrophages were found to play a critical role in intracranial aneurysm development, along with the presence of RGS1, which activates and progresses inflammatory signaling. Xie *et al.*<sup>168</sup> used transcriptomics and genomics methods to develop a greater understanding of the effects of G protein-coupled receptors on lung adenocarcinoma and create a prognostic model that tests responses of patients to immunotherapy and sensitivity to first-line drugs.

Table 1 presents a summary of representative multi-omics studies described above.

## 4 Diagnostic value of biofluids

### 4.1 Overview

Blood is a commonly used biofluid for omics analyses. It is composed of two parts: a cellular component consisting of red and white blood cells and platelets, and a liquid carrier, called plasma. Plasma accounts for approximately 50–55% of blood volume, with blood cells (erythrocytes, leukocytes and platelets) accounting for the remaining portion. Plasma is obtained from a blood sample, if anti-coagulants are introduced, by simply centrifuging the sample and removing or decanting the most buoyant (non-cellular) portion. If no anticoagulant is added and the blood is allowed to clot, the supernatant fluid is called the serum, which is less viscous than plasma and lacks



Table 1 Primary studies regarding multi-omics analysis

Focus	Type of omics	Biofluid	Analytical approach	Biomarkers	Species sample size	Ref.
Baseline health monitoring	Peripheral blood, plasma, and serum	Transcriptomics, proteomics, and metabolomics	RNAseq	Primary circadian pacemaker (PER1), inflammatory molecules (C2, C9, IL5), pro-inflammatory cytokines (IP10, IL1, IL1R1)	Human, n = 105	77
Baseline health monitoring	Blood, saliva, stool, and urine	Metabolomics, proteomics, and microbiomics	RNAseq, SRM, and use of public databases	Gamma-glutamyltyrosine	Human, n = 108	78
Baseline health monitoring	Heart tissues for intermittent fasting	Proteomics, phosphoproteomics, and transcriptomics	Use of public databases, LC-MS, immunoblot assay	Metabolic enzymes (PFKFB1/2, HK1, PKC2, AKT1, mTOR, MAPK9), metabolic proteins (AHSG, SRSF1, APOE, ACAD8, PKM), phosphoproteins (CDK, PDK, AMPK, PKA, EGFR), and transcription factors	Mouse	134
Environmental exposure	Blood serum and fecal matter	Transcriptomics, serum non-targeted metabolomics, and genomics	ELISA, colorimetry, western blot assay, RNAseq, and LC-MS	Arsenic-induced liver fibrosis genes (TGFB1, ACTA2, COL1A1), vital oxidative stress response genes (SOD1, GPX4, GST gene family), steatosis-related genes (PPARG, FABP5, ACOX1) liver apoptosis markers (BCL-2, COG gene family), inflammation response genes (NF-kB, IL-1 $\beta$ ) mRNA (C1qtnf4), miRNA (mmu-mir-21a), polyamines (putrescine, N-acetyl-putrescine, acetyl-spermidine), proteins (Sod2, cat, Txnrd1, Atox1, G6pdx), and enzymes	Chickens, n = 40	131
Environmental exposure	Urine and left lung tissue for cigarette smoke exposure	Transcriptomics, proteomics, lipidomics, metabolomics	HRMS/MS, UHPLC-MS/MS, and use of public databases	miR02137, mmu-mir-21a, polyamines (putrescine, N-acetyl-putrescine, acetyl-spermidine), proteins (Sod2, cat, Txnrd1, Atox1, G6pdx), and enzymes	Mouse	90
Environmental exposure	Blood, liver and brain tissues for benzalkonium chloride exposure	Lipidomics and transcriptomics	HILIC-IM-MS, UHPLC-MS/MS, RNAseq, and use of public databases	Sterol precursors (7- and 8-DHC, 7-DHD, desmosterol, lanosterol), lipids (DGs, TGs, HexCers, and Cers), proteins (SREPBs), and sterol biosynthesis genes (Hmgcs2, Idit, Cyp51)	Mouse	91
Metabolic and inflammatory diseases	Rheumatoid arthritis fibroblasts and synovial fibroblasts	Proteomics, transcriptomics and single-cell transcriptomics	Use of public databases and network construction analysis	Multi-pathway (P13K/AKT, Th17, MAPK, TNF, JAK-STAT)	Cultured cells	145
Metabolic and inflammatory diseases	Synovial fluid and plasma for osteoarthritis	Proteomics	Reporter gene assay	Synovial fluid-derived EVs (stabilin 1, perilipin 4, apolipoprotein C-IV)	Horse, n = 4	152
Metabolic and inflammatory diseases	Urine, blood, and kidney tissues for diabetic kidney disease	Transcriptomics, proteomics, and phosphoproteomics	LC-MS/MS, RNAseq, and HPLC	Genes (ALOX15, COL19A1, CXCL3, GKN3, UGT1A2), proteins (ALOX15, SERPINA1E, AKR1C18, SLC01A1, UGT1A1), phosphorylated proteins (ENO1, MIOX, NQO1, LRRFP1)	Mouse	153
Nutrition	Water samples for wastewater-based epidemiology	Metabolomics and genomics	LC-MS/MS	Phytoestrogen metabolites (genistein, daidzein, enterolactone), bacterial taxa ( <i>Bifidobacterium</i> , <i>Blautia</i> , <i>Roboutsia</i> , <i>Clostridium</i> , <i>Dorea</i> , <i>Subdoligranulum</i> , <i>Intestinibacter</i> , <i>Eubacterium</i> , <i>Bacteroides</i> , <i>Prevotella</i> , <i>Sensu-galimassilia</i> , <i>Roseburia</i> , <i>Slackia</i> )	Phytoestrogen	110
Nutrition	Blood, breast milk, fecal matter, urine, and saliva	Metabolomics, proteomics, and peptidomics	NMR, Qubit dsDNA high-sensitivity assay, and LC-MS/MS	Analysis not completed	Human, n = 200 mother-infant dyads	111
Nutrition	Urine and fecal matter	Proteomics and microbiomics	LC-MS, use of public databases, and DMAC assay	Urinary proteins (VIP36, PIK3IP1, uromodulin), microbiota ( <i>Bacteroidetes</i> , <i>Firmicutes</i> , <i>Akkermansia</i> )	Human, n = 10	133
Infectious diseases and sepsis	Fecal matter for Hospital-acquired diarrhea	Metabolomics and microbiomics	RNAseq, GC-MS, SCFA analysis, and use of public databases	Proline Stickland fermentation by-products (5-aminovaleric acid, isovalerate, isobutyrate), L-leucine and L-valine Stickland fermentation by-products (4-MPA/L-leucine ratio)	Human, n = 143	143
Infectious diseases and sepsis	Blood and urine for COVID-19	Lipidomics and proteomics	LC-MS, DESI-MS, ELISA assay, and O-link proximity extension assay	Phospholipases (sPLA2-IIA, PLA2G2D, PGE2, 5(6)-DHET), eicosanoids (5- and 15-HETE, DHETS, DHOMES), and phospholipids (ChoE-18:3, LPC-O-16:0, and PC-O-30:0)	Human	141
Infectious diseases and sepsis	Not described	Transcriptomics and proteomics	LC-MS/MS, CHIP-seq analysis, and EMSA	HP1021	Cultured <i>Helicobacter pylori</i>	144





Table 1 (continued)

Focus	Type of omics	Biofluid	Analytical approach	Biomarkers	Species sample size	Ref.
Infectious diseases and sepsis	Blood for SARS-CoV-2	Proteomics, metabolomics, and lipidomics	LC-HRMS, LC-MS/MS, and FIA-MS/MS	Lipid metabolites (thiopyran, tryptophan, and kynurenine), lipids (LPC, LPE, PC, and hexosylceramide classes), and proteins (CCL7, CAI4, and HexCer)	Human patients, $n = 32$	142
Neurodegenerative diseases and inflammation	Intracranial aneurysm fibroblasts and macrophages	Transcriptomics	Use of public databases, network construction analysis, western blot assay, and RNAseq	Cel5, Ptgds, Spp1, Fos, Cfb, Rgs6, Slamf8, Nbl1, Ndrfg2, Tmem158, RGS1, CD68, CXCL1, MCP-1, TNF- $\alpha$	Cultured cells, $n = 97$	167
Neurodegenerative diseases and inflammation	Fecal matter for white matter disease	Metabolomics and microbiomics	Use of public databases, LC-MS/MS, and MRI	Microbiota	Human	132
Neurodegenerative diseases and inflammation	Rett syndrome fibroblasts	Transcriptomics and proteomics	RNAseq, TMT-MS, and public databases	Transcription factors (CREB1, SRF, HEYL, GLIS2, NFATC4, JUN), DE genes (MYO1C, HARS2, MECP2), and DEP genes (APPL2, CNPY4, CTSC, REPS1, CNRN1)	Human, $n = 49$	154
Respiratory diseases and cancer	Blood serum for community-acquired pneumonia	Proteomics and metabolomics	Use of public databases, ELISA, and LC-MS/MS	Hippo signaling pathway (ASS1, SAA2 SFTPB, GRP, LBP, CETP, ITIH1, and LDHA), metabolic pathways (TKT, GPI, 161 LCAT, APOA1, APOA2, APOA4, APOC1, APOI1), and inflammation-related pathways (IGHG1, IGHG2, IGHM, TLR, and PI3K/Akt)	Human, $n = 161$	165
Respiratory diseases and cancer	Lung tissue for lung squamous cell carcinoma	Transcriptomics and genomics	Use of public databases and network construction analysis	CD93	Human, $n = 61$	166
Respiratory diseases and cancer	Lung adenocarcinoma	Transcriptomics (single-cell and bulk) and genomics	Use of public databases and network construction analysis	G protein-coupled receptor-related genes (CCL20, DDIT4, GPX3, BEX5, AKAP12, DSG2, SERPINH1, LDHA, DNAJB4, and DOCK4)	Cultured cells	168

fibrinogen, prothrombin, and other clotting proteins. Both plasma and serum are aqueous solutions (about 95% water) containing a variety of substances including proteins and peptides (such as albumins, globulins, lipoproteins, enzymes, and hormones), nutrients (such as carbohydrates, lipids and amino acids), electrolytes, organic wastes and a variety of other small organic molecules suspended or dissolved in them. Based on current analytical techniques the primary difference between serum and plasma appears to lie in the compounds involved in the clotting process, although modest discrepancies in the relative distribution of some compounds between these pools have also been reported. Serum is a primary carrier of small molecules in the body and its chemical composition has been studied in the last 70 years.<sup>169–171</sup> The successful identification of biomarkers in blood, serum or plasma has a long history, which started with blood typing to guide blood transfusions,<sup>172</sup> followed by newborn metabolic screening for the early detection of metabolic diseases,<sup>173</sup> analysis of serum prostate-specific antigen for the early detection of prostate cancer<sup>174</sup> and dozens of other applications. Dozens of proteins have received FDA approval for clinical practice and most of them are for cancer diagnosis.<sup>175</sup> Human metabolites, mostly present in blood, plasma, or serum, serve as valid clinical markers and are currently compiled in the Mayo Clinic test catalog (<https://www.mayocliniclabs.com/test-catalog/>), LabCorp test menu (<https://www.labcorp.com/test-menu/search>), and Quest Diagnostics Test Directory (<https://testdirectory.questdiagnostics.com/test/home>). The word metabolomics in the U.S. National Library of Medicine at Clinical Trials retrieved 962 studies (<https://clinicaltrials.gov/>) and an almost equal number of completed or recruiting clinical studies. Saliva is a complex bodily fluid consisting of *ca.* 99% water, inorganic and organic substances and a variety of proteins such as enzymes, mucus and glycoproteins.<sup>176</sup> The potential of salivary inorganic constituents, antioxidants, hormones, antibodies and antigens as biomarkers in the diagnosis of several oral and systemic diseases is expanding and has been recently reviewed.<sup>177</sup> In oral diseases, saliva has been used to detect oral cavity cancer, dental caries, periodontal disease, and oral dryness. In systemic diseases (*i.e.*, diseases that affect the entire body), saliva has been demonstrated to have strong correlations with plasma or serum for many cell components. Examples include the monitoring of hormone levels, pregnancy, risks for preterm labor, psychological disorders, neurological disorders, immune system status, smoking status, virus infections, malaria and nutritional status.<sup>177–179</sup> Sweat is a slightly acidic biofluid composed mainly of water (99%), electrolytes (*e.g.*, sodium, chloride, and potassium), urea, pyruvate, and lactate. Proteins, peptides, amines, amino acids, and metal ions in smaller concentrations are also found in this biofluid in addition to inhibitors, antigens, antibodies, and a variety of xenobiotics such as drugs, cosmetics, and ethanol. These substances are stored in the sweat glands and secreted into the sweat. At the epidermis surface, partial selective reabsorption of sodium and chloride takes place during transportation, which results in hypotonicity of the secreted sweat in healthy individuals.<sup>180</sup> Diseases can change sweat

composition either by altering the concentration of common components or by forming new components that act as biomarkers for diseases.<sup>181</sup> Except for the case of some high molecular weight proteins, which reach sweat by different intracellular storage mechanisms in particular situations, most sweat components are small molecules resulting from metabolic pathways; therefore, biomarkers characterized so far are mostly metabolites. Besides sweat, the human skin has sebaceous glands responsible for the continuous production of sebum. Human sebum consists of squalene, esters of glycerol, wax, and cholesterol, as well as free cholesterol and fatty acids. Triacylglycerides and fatty acids, taken together, account for the predominant proportion (57.5%), followed by wax esters (26%) and squalene (12%). The metabolic pathways regulating its composition and secretion rate are far from complete understanding.<sup>182</sup> Recently, skin sebum has shown promising results as a sample for the diagnosis of neurodegenerative diseases (Fig. 5).

Urine is the most widely used biological specimen, apart from blood. Interest in using urine for diagnostic purposes arises from the fact that it is a rich source of disease biomarkers, and the sample can be obtained noninvasively. Unlike blood, urine has a relatively low concentration of proteins and many low molecular weight compounds (metabolites); hence metabolomics studies of urine are relatively simple in terms of both sample preparation and analysis. Metabolomics studies with nuclear magnetic resonance and mass spectrometry can measure the concentration of more than 3 000 chemical compounds in the urine, providing possible chemical signatures of different diseases and health.<sup>186</sup> Urinary proteomics and metabolomics studies show altered signatures in patients with gastrointestinal disorders and cancer compared to healthy controls. All of these disorders may include the alteration of urinary metabolites in association with the gastrointestinal microbiota

and possibly dysbiosis, especially in chronic conditions.<sup>187</sup> Nasal secretions originate mostly from submucosal glands and goblet cells. Mucus is composed of water (95%); glycoproteins (2%); albumin, immunoglobulins, lysozyme, lactoferrin and other proteins (1%); inorganic salts (1%); and lipids (<1%).<sup>188</sup> In recent years, investigations on upper airway mucosa inflammation in response to inoculation with bacterial or viral pathogens,<sup>189,190</sup> allergen challenge,<sup>191–193</sup> or exposure to environmental pollutants<sup>194–196</sup> have focused on the detection of minute amounts of cytokines and inflammatory mediators. Recently, COVID-19 diagnosis in nasal secretion based on small molecules has been reported by more than one mass spectrometry method,<sup>185,197</sup> as described in Fig. 5.

Exhaled breath and breath condensate contain inorganic and organic compounds, as well as aerosols in the form of water vapor and particles. Focusing on the gas phase, breath contains diverse inorganic species and several hundred volatile organic compounds (VOCs) of diverse chemical nature, the latter being present only in trace quantities.<sup>198</sup> Breath analysis has received unprecedented attention recently in relation to the severe acute respiratory syndrome associated with the COVID-19 pandemic.<sup>199</sup> The potential for exhaled breath to either detect this airborne virus directly or to diagnose infection is currently being investigated as a comfortable alternative to existing approaches that collect mucus secretions. While no breath test has yet been translated to patients for the reliable detection of the infection, studies have reported potential breath-borne VOC biomarkers detected *via* gas chromatography ion mobility spectrometry (GC-IMS) or specific breathprints using proton transfer reaction time-of-flight mass spectrometry (PTR-TOF-MS).<sup>200</sup> Several breath tests exploit exogenous compounds, such as the well-known and widely implemented breath alcohol 'breathalyser' test, as used in law enforcement to identify drink-drivers,<sup>201</sup> breath tests for hypolactasia,<sup>202</sup> and *Helicobacter pylori* infection.<sup>203</sup> In the latest Breath Biopsy Conference, advancements in the collection and analysis of volatile organic compounds in exhaled breath for the diagnosis of asthma, cirrhosis, cancer, and tuberculosis have been documented.<sup>204</sup>

## 4.2 Limitations of using biofluids for molecular phenotyping and declining health

In blood and related biofluids (serum and plasma) the concentration of the average metabolite normally can vary by about  $\pm 50\%$ , with some metabolites varying by as much as  $\pm 100\%$  (such as L-lactic acid, L-glutamine, glycine). These relatively large ranges of metabolite concentrations are challenging to baseline measurements and are due to several factors, including age, gender, genetic background, diurnal variation, health status, activity level, and diet.<sup>169,205</sup> Saliva is a viscous fluid due to the presence of mucopolysaccharides and mucoproteins. The difficulty in measuring precise volumes of viscous saliva samples can decrease the accuracy of analytical measurements. Due to the low concentrations of compounds in saliva, assays with high sensitivity and low limit of detection are required. The components of saliva may depend on the area in the oral cavity in which the sample is collected as well as the collection

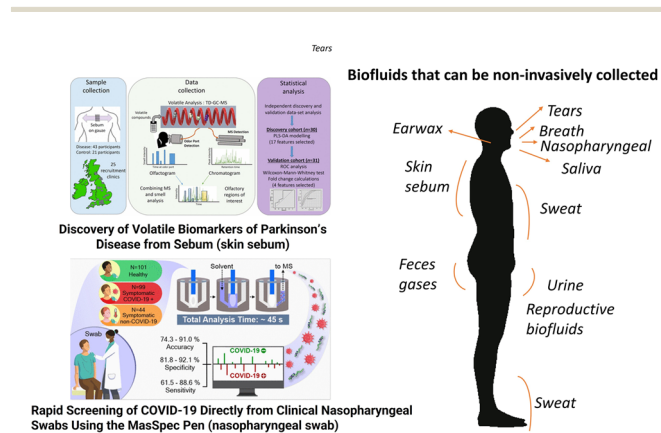


Fig. 5 The composition and diagnostic value of blood and some biofluids that can be collected non-invasively are indicated in the right side of the figure; saliva, sweat, urine, nasopharyngeal mucus, breath, and skin sebum and urine are explored in this review. On the left side of the figure, illustrative examples show that skin sebum obtained from t-shirts was successfully used to diagnose Parkinson's disease,<sup>183</sup> and nasopharyngeal mucus sampled directly from the swab and analyzed using a mass spectrometer was used to screen for COVID-19 infection.<sup>185</sup> Adapted with copyright permission from ACS Publications.



methods (unstimulated vs. stimulated). Also, oral diseases such as gingivitis, change in pH and other potential interferences such as smoking, fasting and hydration level can also influence the levels of the biomarkers of interest.<sup>206</sup> There are relatively large variations in metabolite concentrations between different individuals because many salivary components play multifunctional roles and, in some cases, have overlapping functions. For example, glutathione, ascorbic acid, and uric acid are all antioxidants. Amylases, cystatins, histatins, mucins and peroxidases are all involved in antibacterial activity in the oral cavity. Diurnal fluctuation, seasons (temperature) and other intra-individual inconsistency may also have to be considered.<sup>175</sup> Regarding sweat tests, the main limitations are related to the high cost and special care required for the collection patches and the lack of automation.<sup>207</sup> There is also the need to normalize sampled volume by reference compounds such as sodium.<sup>208</sup>

Urine has long been a “favored” biofluid among metabolomics researchers. It is easy to obtain in large volumes and chemically complex. However, this chemical complexity has also made urine a particularly difficult substrate to fully understand. As a biological waste material, urine typically contains metabolic breakdown products resulting from a wide range of foods, drinks, drugs, environmental contaminants, endogenous waste metabolites and bacterial by-products. Many of these compounds are poorly characterized and understood. Besides this, diagnostic compounds in urine present diverse confounding effects, such as diet variations, large inter- and intra-individual variations, variations induced by sample collection, handling and storage and inconsistency in data extraction, interpretation, and analytical methods. Also, effects of the kidney function and the metabolic function of the body, which may affect secretion and reabsorption of the circulating metabolites may confound the final results.<sup>209,210</sup> Biological markers in nasal secretions provide valuable information on nasal pathophysiology and as a health indicator, this biofluid has so far mainly been explored for infectious disease diagnostics. Published data on biomarker concentrations in nasal fluids are inconsistent mainly due to different sampling techniques.<sup>211</sup> Regarding breath analysis, a major challenge is the sensitive detection of individual compounds. There are also technological limitations associated with reliably capturing breath and the analytical intricacy of extracting potential biomarkers from complex datasets.<sup>198,212</sup> The lack of standardization in breath analysis has led to a limited alignment of results between independent studies employing different approaches.<sup>213</sup> The successful development and implementation of the nitric oxide breath test for asthma represents an example of diagnosis using breath analysis.<sup>214</sup>

## 5 Examples of databases focused on human disease and exposure

Currently, 500 000 chemicals (synthetic and endogenous small molecules) can be found in our bodies, in our food or in the environment. This information is catalogued into publicly accessible databases, and those listed below include molecular

features related to diseases. Libraries can be commercial but there are extensive open-access databases, such as *m/z* Cloud (<https://www.mzcloud.org/>) and the Human Metabolome Database (described below in more detail), containing hundreds of thousands of mass spectra that provide chemical structural information of endogenous and exogenous (synthetic) compounds.

Marker DB: this database is freely available and attempts to consolidate information on all known clinical and a selected set of pre-clinical biomarkers into a single resource. MarkerDB includes five major types of biomarkers (condition-specific, protein, chemical, karyotypic and genetic) and four biomarker categories (diagnostic, predictive, prognostic and exposure). Information compiled include biomarker names and synonyms, associated conditions or pathologies, detailed disease descriptions, detailed biomarker descriptions, biomarker specificity, sensitivity and ROC curves, standard reference values (for protein and chemical markers), variants for SNP (single nucleotide polymorphisms) or genetic markers, sequence information (for genetic and protein markers), molecular structures (for protein and chemical markers), tissue or biofluid sources (for protein and chemical markers), chromosomal location and structure (for genetic and karyotype markers), clinical approval status and relevant literature references. Users can browse the data by conditions, condition categories, biomarker types, biomarker categories or search by sequence similarity through the advanced search function. Currently, the database contains 142 protein biomarkers, 1089 chemical biomarkers, 154 karyotype biomarkers, and 26 374 genetic markers. These are categorized into 25 560 diagnostic biomarkers, 102 prognostic biomarkers, 265 exposure biomarkers and 6746 predictive biomarkers. Collectively, these markers can be used to detect, monitor, or predict 670 specific human conditions, which are grouped into 27 broad condition categories (<https://markerdb.ca>).

Online Mendelian Inheritance in Man (OMIM) is a comprehensive compendium of human genes and genetic phenotypes. The full text and referenced overviews in OMIM contain information on all known Mendelian disorders and over 15 000 genes. OMIM focuses on the relationship between phenotype and genotype. It is updated daily, and the entries contain many links to other genetics resources. OMIM contains nearly 4800 single gene disorders and traits as well as 5800 phenotypes for which the molecular basis is known. <https://www.omim.org/>.

OMMBID or the On-Line Metabolic and Molecular Basis to Inherited Disease is a subscription-based online book/encyclopedia describing the genetics, metabolism, diagnosis, and treatment of hundreds of metabolic disorders contributed by hundreds of experts. It also contains extensive reviews, detailed pathways, chemical structures, physiological data, and tables that are particularly useful for clinical biochemists (<https://ommbid.mhmedical.com/book.aspx?bookID=2709#225069394>).

METAGENE is a knowledge base for genetic metabolic disorders providing information about the disease, genetic cause, treatment and the characteristic metabolite concentrations or clinical tests that may be used to diagnose or monitor the condition. It has data on 1150 genetic diseases (<https://www.metagene.de/>).



The PharmGKB database is a central repository for genetic, genomic, molecular, and cellular phenotype data and clinical information about people who have participated in pharmacogenomics research studies. The data include, but not limited to, clinical and basic pharmacokinetic and pharmacogenomic research in the cardiovascular, pulmonary, cancer, pathway, metabolic and transporter domains. Its aim is to aid researchers in understanding how genetic variation among individuals contributes to differences in reactions to drugs. PharmGKB contains searchable data on genes (>20 000), diseases (>3000), drugs (>2500) and pathways (235). It also has detailed information on 470 genetic variants (SNP data) affecting drug metabolism (<https://www.pharmgkb.org/>).

SMPDB (The Small Molecule Pathway Database) is an interactive, visual database containing more than 30 000 small molecule pathways found in humans only. The majority of these pathways are not found in any other pathway database. The pathways include metabolic, drug, and disease pathways (<https://smpdb.ca/>).

Serum metabolome database: the serum metabolome database is integrated into the Human Metabolome Database (HMDB), allowing users to browse the data in different views, metabolites, concentrations, and diseases. It is a freely available electronic database containing detailed information about 4651 small molecule metabolites found in human serum along with 10 895 concentration values. The data tables may be sorted and searched by concentration values and ranges. The information includes literature and experimentally derived chemical data, clinical data, and molecular/biochemistry data. Each MetaboCard entry contains more than 110 data fields and many of them are hyperlinked to other databases (KEGG, PubChem, MetaCyc, ChEBI, PDB, Swiss-Prot, and GenBank) (<https://www.serummetabolome.ca>).

Exposome-explorer is the first database dedicated to biomarkers of exposure to environmental risk factors for diseases. It contains detailed information on the nature of biomarkers, the population studied, biospecimen(s) used, biomarker concentrations and the reference publication(s) (<https://exposome-explorer.iarc.fr/>).

The METLIN metabolite database is a repository with full access by subscription for mass spectral metabolite data. METLIN is searchable by compound name, mass, formula, or structure. It contains over 930 000 molecular standards as of December 2023. METLIN contains MS/MS, LC/MS and FTMS data that can be searched by peak lists, mass range, biological source, or disease (<https://metlin.scripps.edu/>).

FoodDB is the world's largest and most comprehensive resource on food constituents, chemistry, and biology. Each chemical entry in the FoodDB contains more than 100 separate data fields covering detailed compositional, biochemical, and physiological information (obtained from the literature). Users can browse or search FoodDB by food source, name, descriptors, function, or concentrations (<https://foodb.ca/>).

DrugBank: the latest release of the database (version 5.0) contains 9591 drug entries, including 2037 FDA-approved small molecule drugs, 241 FDA-approved biotech (protein/peptide)

drugs, 96 nutraceuticals and over 6000 experimental drugs. Additionally, 4270 non-redundant protein (*i.e.*, drug target/enzyme/transporter/carrier) sequences are linked to these drug entries. Each DrugCard entry contains more than 200 data fields with half of the information being devoted to drug/chemical data and the other half devoted to drug target or protein data. The DrugBank Online website is available to the public as a free-to-access resource. However, use and redistribution of content from DrugBank Online or the underlying DrugBank Data, in whole or part, and for any purpose requires a license. Academic users can apply for a free license for certain use cases while all other users require a paid license (<https://www.drugbank.com/>).

T3DB: the Toxin and Toxin Target Database (T3DB) or soon to be referred as the Toxic Exposome Database is a unique bioinformatics resource that combines detailed toxin data with comprehensive toxin target information. The database currently houses 3678 toxins described by 41 602 synonyms, including pollutants, pesticides, drugs, and food toxins, which are linked to 2073 corresponding toxin target records. Altogether there are 42 374 toxin and toxin target associations. Each toxin record (ToxCARD) contains over 90 data fields and holds information such as chemical properties and descriptors, toxicity values, molecular and cellular interactions, and medical information. This information has been extracted from over 18 143 sources, which include other databases, government documents, books, and scientific literature. The focus of the T3DB is on providing mechanisms of toxicity and target proteins for each toxin (<https://www.t3db.ca/>).

ContaminantDB: the contaminantDB is a unique bioinformatics resource that combines detailed contaminant data from different online references and databases on contaminants. The database currently houses 54 249 compounds. It is both modelled after and closely linked to the Human Metabolome Database (HMDB) and DrugBank. The databases and sources used to gather contaminant data include IARC carcinogens group 1, 2A, 2B, 3 and 4, DrugBank drugs and metabolites, disinfection by-products, MyExposome Chemicals, ToxCast and Tox21 chemicals, EPA high production volume chemicals, OSHA hazardous chemicals, Clean Air Act Chemicals, T3DB toxins, ECHA substances of high concern, DEA chemicals, EPA endocrine screening chemicals, EAFUS chemicals, and OECD high production volume chemicals (<https://contaminantdb.ca/>).

BioTransformer: BioTransformer 3.0 is a software tool that predicts small molecule metabolism in mammals, their gut microbiota, and the soil/aquatic microbiota. Moreover, it can also assist in the identification of metabolites, which is based on the metabolism prediction (<https://biotransformer.ca/>).

### 5.1 Metabolome annotation in databases

Metabolomics too often relies on library matches to data from previously identified standards for compound identification and leaves many, potentially important, compounds in datasets unidentified. The “dark matter” metabolome can be defined as the metabolites not yet curated in available databases. These can be either not extracted and/or not identified using standard



analytical methods or are lost/transformed during extraction. Even though the latest version of the Human Metabolome Database – HMDB (version 5.0) lists 220 945 metabolite entries (almost two-fold increase from version 4.0), the actual number of human metabolites is expected to be about 10 times higher. Note that the HMDB does not include anthropogenic compounds (*e. g.* pollutants), but it harbors large quantities of predicted MS/MS and GC-MS reference spectral data and predicted (physiologically feasible) metabolite structures.<sup>215</sup>

Developing new analytical methods and technology could provide new insights into the dark metabolome, reduce information loss, and generate fresh insights and new knowledge in many fields, including animal health, biomedicine, disease diagnostics, environmental and food science, physiology, pharmacology, toxicology and zoology. In short, illuminating the dark metabolome could be a leap forward in the molecular characterization of biological systems.<sup>30,215</sup>

One of the approaches to tackle the dark metabolome is the use of *in silico* predictive tools and mining existing data for building evidence for the presence of small molecules in complex samples using libraries of calculated chemical properties and associated matching to experimental data using multiple molecular attributes (*i.e.*, multiattribute matching). Therefore, libraries of chemical properties derived computationally can replace libraries derived from authentic standards for compound identification under specific conditions, for example, when the evidence of presence is strong enough and confidence is high enough to support the intended application.<sup>216</sup>

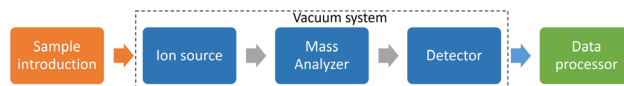
## 6 Mass spectrometry instrumentation for omics analyses

### 6.1 Benchtop MS instrumentation

Even though we may not notice, mass spectrometry analysis is embedded in diverse aspects of day-to-day life. Bodily fluids drawn for medical testing may be analyzed by MS. The doped semiconductors that form the basis of all our electronic devices use MS as part of the quality control process. The use of MS helps identifying potential explosives before they make it on board an airplane and can also assist in finding toxins in our food supply. Mass spectrometry is a form of chemical analysis used to measure the mass-to-charge ratio ( $m/z$ ) of atoms and/or molecules in a sample and it can be accomplished using a larger variety of platforms (Fig. 6). It is also capable of distinguishing between different isotopes of the same element. Depending upon the type of mass spectrometer, these measurements can often be used to determine the exact molecular weight of the sample components and to identify unknown compounds.

### 6.2 Miniature mass spectrometers

The majority of research in the biomedical science field is carried out using high mass resolution instruments, such as orbitraps and QTOF mass analyzers. However, in the clinic, instrument footprinting, quantification needs and cost constraints necessitate the use of low-resolution devices and some

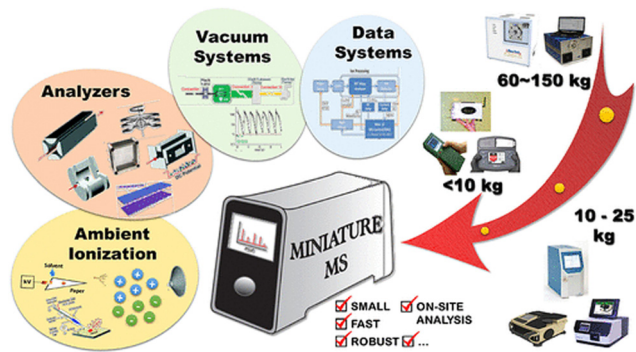


**Fig. 6** Schematic of the main steps commonly used in lab-based mass spectrometry analyses. There are many different types of mass spectrometers, but they all have three features in common. The first is some means by which atoms or molecules from the sample can be ionized. Neutral species cannot be steered by electric fields used in mass spectrometers, and thus it is necessary to produce ions. There are many different means by which this can be accomplished, and they are collectively referred to as ion sources. The second component of all mass spectrometers is the mass analyzer itself. There are several different means by which the  $m/z$  ratio of ions can be measured. Time-of-flight (ToF), orbitraps and quadrupole mass analyzers are commonly used, each with its own set of strengths and limitations. The final component common to all mass spectrometer systems is a detector for detecting or counting the number of ions with a specific  $m/z$  value. A final factor that needs consideration is how to couple the ion source to the sample to produce the ions for measurement, especially because all mass spectrometers must be operated under vacuum. In some cases, the sample will also be housed under vacuum, while in other cases the sample will be at atmospheric pressure and some may incorporate some other form of separation technology prior to introduction into the ionization chamber.

of them are miniature mass spectrometers. Areas such as emergency response, military, and law enforcement, but also agriculture, home health care, archaeology, and pollution control are being benefited using portable mass spectrometers, which are having a fast pace of miniaturization (Fig. 7). Portability is a primary driver for miniaturization efforts, but factors such as ease of use are also strongly considered since users are expected to have little or no technical training in analytical chemistry. Although performance may be reduced from laboratory-scale instruments, the value of the instrumentation is only as good as the quality of information provided to allow critical decisions to be made.

A compilation of 35 lightweight mass spectrometers was reported by Snyder *et al.*<sup>217</sup> The Mini lineage of instruments developed at Purdue University exemplifies well the concept of portability. The Mini 10<sup>218</sup> and 11<sup>219</sup> mass spectrometers are hand-held instruments of total weights 10 and 4 kg, respectively, which operate under low-power conditions (<70 W). They contain identical rectilinear ion traps that can isolate compounds from complex mixtures and perform structural analysis with unit resolution across a mass range of up to  $m/z = 700$ . While the Mini 10 used an internal electron impact source, the Mini 11 was designed with a discontinuous atmospheric pressure interface (DAPI) to take advantage of internal, external, and ambient sources. The Mini 12 is the newest member of Purdues lineup of miniature mass spectrometers and is designed for point-of-care analysis.<sup>220</sup> Power consumption is similar for the Mini 10 and 11, 50 W, because the same trap is used. The Mini 12 has an integrated solvent pump with solvent containers, sample cassette, and novice user interface for direct analysis of blood and other biofluids. The Mini S is an alternative backpack configuration that has similar performance characteristics but is designed for in-field applications (*e.g.*, pesticides, narcotics, and explosives detection).<sup>221</sup> The





**Fig. 7** Graphical abstract of a review about miniature and fieldable mass spectrometers adapted from ref. 217. Some miniature mass spectrometer analyzers are currently providing unit resolution up to an upper mass of  $m/z = 1000$ . These systems have been continuously shrinking in size with adoption of microelectromechanical system technologies. These mass analyzers can provide compound identification/confirmation and quantitation limits close to benchtop instruments for some warfare and environmental compounds operating on battery power for a few hours. Many ionization techniques are available, especially ambient methods, to tailor the analysis to a specific application, and the entire system is expected to be light and small. A long development period preceded these advances. The development began with the first hand-held mass analyzer reported in 1991, proceeding to the first arrays of quadrupoles, as well as serial and parallel ion trap arrays along with the first micro-ion trap arrays. The introduction of ambient ionization and the first discontinuous atmospheric pressure interface resulted in fully portable systems coupled to ambient ionization sources. Miniature mass spectrometers are expected to become a major tool in the analytical sciences, especially given the increasing interest in *in situ*, point-of-care, online, and on-site measurements. Adapted with copyright permission from ACS Publications.

usual ion source is a low-temperature plasma probe that operates independent of geometry and is integrated into a 2 kg hand-held unit. This unit also contains the ion transfer capillary, DAPI valve, rectilinear ion trap, and detector. The electronics, battery, vacuum system, and plasma gas, the heaviest components, are in the 10 kg backpack unit. Two miniaturized mass spectrometry systems have recently been developed by the Beijing Institute of Technology. The first is a typical DAPI/rectilinear ion trap instrument that couples capillary electrophoresis (CE) with nanoelectrospray.<sup>222</sup> The system was able to separate a peptide (MRFA) from doubly charged angiotensin II, which has the same nominal  $m/z$ . A second instrument used a continuous atmospheric pressure interface rather than DAPI.<sup>223</sup> This was achieved by differential pumping, where the first low-pressure region (1–6.6 torr) was separated from ambient pressure by the ion transfer capillary and separated from the mass analyzer region (1–6.6 mTorr) by a small aperture. The vacuum system was standard: a diaphragm pump from Scroll Tech combined with a Pfeiffer HiPace 10 turbo pump. The instrument showed excellent reproducibility (RSD < 7%), low ppm limits of detection, and unit resolution. Other trap-based configurations with external/ambient ionization are MassTechs MT Explorer 50, which has a 3D ion trap and can be interfaced with ESI, atmospheric pressure matrix-assisted laser desorption/ionization (AP MALDI), atmospheric pressure chemical ionization (APCI), and direct analysis in real time (DART) sources, and an instrument

with a low-pressure dielectric barrier discharge ionization source coupled to a linear ion trap.<sup>224</sup> In this instrument, a diaphragm pump is used to pump down the sample container to transfer vapors through a pinch valve into the ion source. An AC voltage is applied to a dielectric, resulting in ionization of analytes. A reflectron-TOF developed by Shanghai University, Guangzhou Hexin Analytical Instrument Co., Ltd. and the National University of Defense Technology of the Peoples Liberation Army of China uses dimethylsiloxane membrane introduction and single photon ionization (UV) for detection of volatile components in air. A second TOF, the Suitcase TOF, was developed by the Johns Hopkins Applied Physics Laboratory.<sup>225</sup> This breakthrough instrument is pumped by a standard diaphragm/turbo pump combination. Matrix assisted laser desorption/ionization is used as the source. The mass analyzer, a reflectron TOF, was able to detect 4 pmol bovine serum albumin, a 66 kDa protein, and showed performance similar to a commercial TOF in terms of resolution and sensitivity. Gas/air sampling is a popular target for portable analysis. The Sam Yang Chemical Company of Korea has developed a palm portable trap-based mass spectrometer for chemical warfare agent (CWA) determination weighing only 1.48 kg but pumped only by an ion getter (and thus requiring frequent recharge).<sup>226</sup> The IonCam from OI Analytical, an instrument no longer commercially available but in the process of being updated, is built using Mattauch–Herzog sector geometry with gas chromatographic (GC) separation, allowing for simultaneous detection of ions of a range of masses. Other instruments for sampling gaseous analytes include the M908 from 908 Devices, which operates at high pressures (> 1 torr) with an ion trap array and which can perform continuous vapor analysis or solid/liquid analysis with thermal desorption swabs. The MS-200 from Kore Technologies, Ltd has a membrane inlet, TOF analyzer, and nonmechanical pumps. Another instrument, the HAP-SITE (Inficon), allows the detection of volatile and semivolatile organic compounds, while the ruggedized GUARDION-7 and TRIDION-9 GC/MS instruments from Torion Technologies (PerkinElmer) use toroidal ion traps.<sup>227,228</sup> Other instruments to mention are a GC-quadrupole ion trap from the California Institute of Technology and Thorleaf Research;<sup>229</sup> and a double focusing (ExB) instrument from the University of Costa Rica and University of Minnesota.<sup>230</sup> Field-based chemical analysis in harsh environments presents challenges for the operator and instrumentation alike. To address the detection needs in such environments, ruggedization is a key attribute of portable chemical detection systems. Electrical systems and low-level sensors have been designed and reported to allow measurement at variable temperature and pressure, under toxic/caustic conditions, and at high radiation and electromagnetic pulse levels.<sup>231–233</sup> Complex modern instruments, such as mass spectrometer devices, have components that can be susceptible to variable operating conditions. In recent years, miniaturization of all the components of mass spectrometers, from pumps and power supplies to mass analyzers and ionization sources, is rapidly advancing. Also, ambient ionization methods that allow for the direct, rapid, and high-throughput analysis of samples under open air conditions with minimal sample preparation are



proliferating. Ambient ionization was introduced by Graham Cooks in 2004, when desorption electrospray (DESI) was reported.<sup>234</sup> Over 80 ambient ionization strategies for the direct analysis of samples have been reported,<sup>235</sup> including paper spray,<sup>236</sup> DART,<sup>237</sup> and rapid evaporative ionization mass spectrometry (REIMS).<sup>238</sup> This has resulted in lowering the power requirements and allowing battery operation for a variety of applications such as rapidly identifying poisoning of patients in ambulances or emergency rooms, monitoring transoceanic shipping containers, and ensuring safe food and water supplies throughout the world. Such a device could accompany medical staff assisting patients in regions of widespread pandemic, where both time and laboratory access become critical constraints. Burns *et al.*<sup>239</sup> have assessed the use of a deployable, single quadrupole mass spectrometer for its ability to detect explosives from glass fiber swabs using two different ambient ionization techniques. Both secondary electrospray ionization and pressure chemical ionization were found to be able to detect the five explosives studied from swabs at levels suitable for trace screening. Recently, researchers from the Naval Air Warfare Center, Weapons Division (NAWCWD) developed a 3D printed cone that can be used as a collection device and an ambient ionization source in field environments. It has been applied for the detection of chemical warfare agent simulants, environmental chemicals (perfluoroalkyl and polyfluoroalkyl substances – PFAS), and synthetic cannabinoids.<sup>240</sup> The detection and identification of chemical warfare agents (CWAs) is a paradigm of health risk assessment in the field. Current methods of detection include color-changing paper, vibrational spectroscopy (infrared and Raman), ion mobility spectrometry and mass spectrometry.<sup>241–244</sup> The detection, identification and quantification of volatile CWAs are well established in laboratory analysis *via* gas chromatography mass spectrometry (GC-MS) and this technology has been adapted for portable on-site use.<sup>217,244</sup>

## 7 Conclusions

Substantial advances have recently been made in the application of mass spectrometry-based approaches for evaluation of human health. Our improved ability to detect compositional variations in biofluids will continue to benefit diagnostic approaches as well as the development and implementation of personalized therapeutic strategies. The increased utilization of multiple omics platforms often being complementarily employed necessitates the development and optimization of integrative frameworks to create synergistic data sets. The need for method and metabolite identification standardization, quality control, and preanalytical procedures is key to data quality and the translation of biomarkers into diagnostic tools. These factors have been recently reported in a number of publications.<sup>245–248</sup> The generation of these integrative data sets from multiple omics platforms should improve our understanding of complex biological processes often induced following exposure and during disease progression. Although often utilized and an extremely valuable component in the

elucidation of biological responses, metabolomics currently is only capable of identifying a fraction of the metabolites within a biological sample. The expansion of metabolite information would unveil new targets and diagnostic markers. Lastly, the production of new mass spectrometry instruments to be utilized in the field or in clinical settings should enable the ability to make data-driven decisions in real-time for healthcare providers. Critical to the everyday use of these tools is the development of instrumentation and software that allows for minimum training and expertise in mass spectrometry. These devices will fundamentally alter medicine and field applications through the ability of novice users from a variety of disciplines to apply this powerful and precise technology.

## Author contributions

All authors mentioned in this article contributed to different stages of this study. The study design, materials preparation, data collection, and analyses were performed by Christina R Ferreira. The first draft of the manuscript was written by Christina R Ferreira. All authors contributed to interpreting the results and provided critical feedback, commenting on previous versions of the manuscript. The final version of the manuscript, the figures and the main text, was written and revised by Christina R. Ferreira, Paulo Clairmont F. de Lima Gomes, Bruce R. Cooper and Jonathan H. Shannahan. All authors read and approved the final manuscript.

## Conflicts of interest

The authors declare that there are no conflicts of interest.

## Acknowledgements

This study was supported by the National Institute of Environmental Health Sciences (NIEHS) grant R01ES033173, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP grants 2022/02991-3 and 2014/50945-4).

## References

- 1 D. Hornburg, S. Wu, M. Moqri, X. Zhou, K. Contrepolis, N. Bararpour, G. M. Traber, B. Su, A. A. Metwally, M. Avina, W. Zhou, J. M. Ubellacker, T. Mishra, S. M. S.-F. Rose, P. B. Kavathas, K. J. Williams and M. P. Snyder, *Nat. Metab.*, 2023, 1578–1594.
- 2 A. N. M. Araujo, I. N. Leroux, D. Z. S. Furtado, A. P. S. da Silva Ferreira, B. L. Batista, H. D. T. Silva, E. Handakas, N. A. Assunção and K. P. K. Olympio, *Front. Public Health*, 2023, **10**, 1–29.
- 3 U. Raudvere, L. Kolberg, I. Kuzmin, T. Arak, P. Adler, H. Peterson and J. Vilo, *Nucleic Acids Res.*, 2019, **47**, W191–W198.
- 4 M. Eskola, C. T. Elliott, J. Hajšlová, D. Steiner and R. Krska, *Crit. Rev. Food Sci. Nutr.*, 2020, **60**, 1890–1911.



- 5 D. A. Dias, O. A. Jones, D. J. Beale, B. A. Boughton, D. Benheim, K. A. Kouremenos, J.-L. Wolfender and D. S. Wishart, *Metabolites*, 2016, **6**, 1–29.
- 6 R. R. D. Silva, P. C. Dorrestein and R. A. Quinn, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 12549.
- 7 B. L. Peisl, E. L. Schymanski and P. Wilmes, *Anal. Chim. Acta*, 2018, **1037**, 13–27.
- 8 Z. Zhou, M. Luo, H. Zhang, Y. Yin, Y. Cai and Z. J. Zhu, *Nat. Commun.*, 2022, **13**, 1–15.
- 9 S. Yohe and B. Thyagarajan, *Arch. Pathol. Lab. Med.*, 2017, **141**, 1544–1557.
- 10 J. Platt, R. Cox and G. M. Enns, *J. Genet. Couns.*, 2014, **23**, 594–603.
- 11 R. A. Pagon, *Bull. Med. Libr. Assoc.*, 2006, **94**, 343–348.
- 12 I. Audo, K. M. Bujakowska, T. Léveillard, S. Mohand-Sad, M. E. Lancelot, A. Germain, A. Antonio, C. Michiels, J. P. Saraiva, M. Letexier, J. A. Sahel, S. S. Bhattacharya and C. Zeitz, *Orphanet J. Rare Dis.*, 2012, **7**, 1–17.
- 13 I. J. Nijman, J. M. van Montfrans, M. Hoogstraat, M. L. Boes, L. van de Corput, E. D. Renner, P. van Zon, S. van Lieshout, M. G. Elferink, M. van der Burg, C. L. Vermont, B. van der Zwaag, E. Janson, E. Cuppen, J. K. Ploos van Amstel and M. E. van Gijn, *J. Allergy Clin. Immunol.*, 2014, **133**, 529–534.
- 14 G. E. Nia, M. Fadaee, R. Royer, K. Dehghan, H. Najmabadi and M. R. Akbari, *Arch. Iran. Med.*, 2016, **19**, 236–240.
- 15 R. Artuso, C. Fallerini, L. Dosa, F. Scionti, M. Clementi, G. Garosi, L. Massella, M. C. Epistolato, R. Mancini, F. Mari, I. Longo, F. Ariani, A. Renieri and M. Bruttini, *Eur. J. Hum. Genet.*, 2012, **20**, 50–57.
- 16 R. A. Weerakkody, J. Vandrovцова, C. Kanonidou, M. Mueller, P. Gampawar, Y. Ibrahim, P. Norsworthy, J. Biggs, A. Abdullah, D. Ross, H. A. Black, D. Ferguson, N. J. Cheshire, H. Kazkaz, R. Grahame, N. Ghali, A. Vandersteen, F. M. Pope and T. J. Aitman, *Genet. Med.*, 2016, **18**, 1119–1127.
- 17 Y.-C. Lee, Y.-T. Liu and B.-W. Soong, *J. Neurogenet.*, 2015, **29**, 103–112.
- 18 K. V. Voelkerding, S. Dames and J. D. Durtschi, *J. Mol. Diagn.*, 2010, **12**, 539–551.
- 19 S. D. Keulenaer, J. Hellemans, S. Lefever, J. P. Renard, J. D. Schrijver, H. V. D. Voorde, M. A. Tabatabaiefar, F. V. Nieuwerburgh, D. Flamez, F. Pattyn, B. Scharlaken, D. Deforce, S. Bekaert, W. V. Crielinge, J. Vandesompele, G. V. Camp and P. Coucke, *BMC Med. Genomics*, 2012, **5**, 1–8.
- 20 K. K. Kim, B. J. Seung, D. Kim, H. M. Park, S. Lee, D. W. Song, G. Lee, J. H. Cheong, H. Nam, J. H. Sur and S. Kim, *Sci. Data*, 2019, **6**, 1–9.
- 21 Y. Yang, D. M. Muzny, J. G. Reid, M. N. Bainbridge, A. Willis, P. A. Ward, A. Braxton, J. Beuten, F. Xia, Z. Niu, M. Hardison, R. Person, M. R. Bekheirnia, M. S. Leduc, A. Kirby, P. Pham, J. Scull, M. Wang, Y. Ding, S. E. Plon, J. R. Lupski, A. L. Beaudet, R. A. Gibbs and C. M. Eng, *N. Engl. J. Med.*, 2013, **369**, 1502–1511.
- 22 M. M. Zutter, K. J. Bloom, L. Cheng, I. S. Hagemann, J. H. Kaufman, A. M. Krasinskas, A. J. Lazar, D. G. B. Leonard, N. I. Lindeman, A. M. Moyer, M. N. Nikiforova, J. A. Nowak, J. D. Pfeifer, A. R. Sepulveda, J. E. Willis and S. L. Yohe, *Arch. Pathol. Lab. Med.*, 2014, **139**, 989–1008.
- 23 A. Casamassimi, A. Federico, M. Rienzo, S. Esposito and A. Ciccodicola, *Int. J. Mol. Sci.*, 2017, **18**, 1–15.
- 24 T. G. Evans, J. E. Podrabsky, J. H. Stillman and L. Tomanek, *J. Exp. Biol.*, 2015, **218**, 1925–1935.
- 25 M. R. Wilkins, J.-C. Sanchez, A. A. Gooley, R. D. Appel, I. Humphery-Smith, D. F. Hochstrasser and K. L. Williams, *Biotechnol. Genet. Eng. Rev.*, 1996, **13**, 19–50.
- 26 J. D. Holman, S. Dasari and D. L. Tabb, in *Informatics of Protein and Posttranslational Modification Detection via Shotgun Proteomics*, ed. M. Zhou and T. Veenstra, Humana Press, Totowa, NJ, 2013, pp. 167–179.
- 27 K. Chandramouli and P.-Y. Qian, *Hum. Genomics Proteomics*, 2009, **1**, 1–22.
- 28 S. Al-Amrani, Z. Al-Jabri, A. Al-Zaabi, J. Alshekaili and M. Al-Khabori, *World J. Biol. Chem.*, 2021, **12**, 57–69.
- 29 S. B. Nimse, M. D. Sonawane, K.-S. Song and T. Kim, *Analyst*, 2016, **141**, 740–755.
- 30 S. Hazlegreaves, Putting a Spotlight on the Dark Metabolome, <https://www.openaccessgovernment.org/putting-a-spotlight-on-the-dark-metabolome/82464/>.
- 31 G. A. Nagana Gowda and D. Raftery, in *NMR-Based Metabolomics*, ed. S. Hu, Springer International Publishing, Cham, 2021, pp. 19–37.
- 32 E. C. Li-Chan, *Introduction to Vibrational Spectroscopy in Food Science*, John Wiley & Sons, Ltd, 2010, ch. 5.
- 33 H. H. Mantsch, *Analyst*, 2013, **138**, 3863–3870.
- 34 D. I. Ellis and R. Goodacre, *Analyst*, 2006, **131**, 875–885.
- 35 S. Türker-Kaya and C. W. Huck, *Molecules*, 2017, **22**, 1–20.
- 36 S. Yu and J. Kong, *Acta Biochim. Biophys. Sin.*, 2007, **39**, 549–559.
- 37 J. Doherty, Z. Zhang, K. Wehbe, G. Cinque, P. Gardner and J. Denbigh, *Anal. Bioanal. Chem.*, 2018, **410**, 5779–5789.
- 38 M. Mehta, R. Naffa, C. Maidment, G. Holmes and M. Waterland, *J. Leather Sci. Eng.*, 2020, **2**, 1–15.
- 39 K. Eberhardt, C. Stiebing, M. S. Christian Matthäus and J. Popp, *Expert Rev. Mol. Diagn.*, 2015, **15**, 773–787.
- 40 R. R. Jones, D. C. Hooper, L. Zhang, D. Wolverson and V. K. Valev, *Nanoscale Res. Lett.*, 2019, **14**, 1–34.
- 41 J. Langer, D. Jimenez de Aberasturi, J. Aizpurua, R. A. Alvarez-Puebla, B. Auguie, J. J. Baumberg, G. C. Bazan, S. E. J. Bell, A. Boisen, A. G. Brolo, J. Choo, D. Cialla-May, V. Deckert, L. Fabris, K. Faulds, F. J. Garcia de Abajo, R. Goodacre, D. Graham, A. J. Haes, C. L. Haynes, C. Huck, T. Itoh, M. Käll, J. Kneipp, N. A. Kotov, H. Kuang, E. C. Le Ru, H. K. Lee, J.-F. Li, X. Y. Ling, S. A. Maier, T. Mayerhöfer, M. Moskovits, K. Murakoshi, J.-M. Nam, S. Nie, Y. Ozaki, I. Pastoriza-Santos, J. Perez-Juste, J. Popp, A. Pucci, S. Reich, B. Ren, G. C. Schatz, T. Shegai, S. Schlücker, L.-L. Tay, K. G. Thomas, Z.-Q. Tian, R. P. Van Duyne, T. Vo-Dinh, Y. Wang, K. A. Willets, C. Xu, H. Xu, Y. Xu, Y. S. Yamamoto, B. Zhao and L. M. Liz-Marzán, *ACS Nano*, 2020, **14**, 28–117.



- 42 R. Pilot, R. Signorini, C. Durante, L. Orian, M. Bhamidipati and L. Fabris, *Biosensors*, 2019, **9**, 1–99.
- 43 E. Fahy, S. Subramaniam, H. A. Brown, C. K. Glass, A. H. Merrill, R. C. Murphy, C. R. Raetz, D. W. Russell, Y. Seyama, W. Shaw, T. Shimizu, F. Spener, G. van Meer, M. S. VanNieuwenhze, S. H. White, J. L. Witztum and E. A. Dennis, *J. Lipid Res.*, 2005, **46**, 839–861.
- 44 Y. H. Rustam and G. E. Reid, *Anal. Chem.*, 2018, **90**, 374–397.
- 45 T. Züllig, M. Trötz Müller and H. C. Köfeler, *Anal. Bioanal. Chem.*, 2020, **412**, 2191–2209.
- 46 X. Han and R. W. Gross, *Mass Spectrom. Rev.*, 2005, **24**, 367–412.
- 47 G. Liebisch, W. Drobnik, M. Reil, B. Trümbach, R. Arnecke, B. Olgemöller, A. Roscher and G. Schmitz, *J. Lipid Res.*, 1999, **40**, 1539–1546.
- 48 Z. Xie, C. R. Ferreira, A. A. Virequ and R. G. Cooks, *Chem. Phys. Lipids*, 2021, **235**, 105048.
- 49 N. Abbassi-Ghadi, E. A. Jones, M. Gomez-Romero, O. Golf, S. Kumar, J. Huang, H. Kudo, R. D. Goldin, G. B. Hanna and Z. Takats, *J. Am. Soc. Mass Spectrom.*, 2016, **27**, 255–264.
- 50 D. R. Klein, C. L. Feider, K. Y. Garza, J. Q. Lin, L. S. Eberlin and J. S. Brodbelt, *Anal. Chem.*, 2018, **90**, 10100–10104.
- 51 C. P. Wild, *Cancer Epidemiol., Biomarkers Prev.*, 2005, **14**, 1847–1850.
- 52 G. W. Miller and D. P. Jones, *Toxicol. Sci.*, 2013, **137**, 1–2.
- 53 G. W. Miller, *The Exposome*, Academic Press, San Diego, 2014, pp. 1–12.
- 54 D. I. Walker, T. M. Mallon, P. K. Hopke, K. Uppal, Y. M. Go, P. Rohrbeck, K. D. Pennell and D. P. Jones, *J. Occup. Environ. Med.*, 2016, **58**, S12–S21.
- 55 Y. H. Park, K. Lee, Q. A. Soltow, F. H. Strobel, K. L. Brigham, R. E. Parker, M. E. Wilson, R. L. Sutliff, K. G. Mansfield, L. M. Wachtman, T. R. Ziegler and D. P. Jones, *Toxicology*, 2012, **295**, 47–55.
- 56 N. Bonvallot, M. Tremblay-Franco, C. Chevrier, C. Canlet, L. Debrauwer, J.-P. Cravedi and S. Cordier, *J. Toxicol. Environ. Health, Part B*, 2014, **17**, 21–44.
- 57 J. E. Rager, M. J. Strynar, S. Liang, R. L. McMahan, A. M. Richard, C. M. Grulke, J. F. Wambaugh, K. K. Isaacs, R. Judson, A. J. Williams and J. R. Sobus, *Environ. Int.*, 2016, **88**, 269–280.
- 58 E. L. Jamin, N. Bonvallot, M. Tremblay-Franco, J. P. Cravedi, C. Chevrier, S. Cordier and L. Debrauwer, *Anal. Bioanal. Chem.*, 2014, **406**, 1149–1161.
- 59 M. Roca, N. Leon, A. Pastor and V. Yusà, *J. Chromatogr. A*, 2014, **1374**, 66–76.
- 60 V. Bessonneau, J. Pawliszyn and S. M. Rappaport, *Environ. Health Perspect.*, 2017, **125**, 077014.
- 61 N. Bonvallot, M. Tremblay-Franco, C. Chevrier, C. Canlet, C. Warembourg, J.-P. Cravedi and S. Cordier, *PLoS One*, 2013, **8**, 1–9.
- 62 M. M. Niedzwiecki, D. I. Walker, R. Vermeulen, M. Chadeau-Hyam, D. P. Jones and G. W. Miller, *Annu. Rev. Pharmacol. Toxicol.*, 2019, **59**, 107–127.
- 63 M. S. Shiels, X. O. Shu, A. K. Chaturvedi, Y. T. Gao, Y. B. Xiang, Q. Cai, W. Hu, G. Shelton, B. T. Ji, L. A. Pinto, T. J. Kemp, N. Rothman, W. Zheng, A. Hildesheim and Q. Lan, *Carcinogenesis*, 2017, **38**, 1004–1010.
- 64 C. F. Woeller, T. H. Thatcher, D. V. Twisk, S. J. Pollock, A. Croasdell, N. Kim, P. K. Hopke, X. Xia, J. Thakar, C. T. M. Mallon, M. J. Utell and R. P. Phipps, *J. Occup. Environ. Med.*, 2016, **58**, 1–26.
- 65 Y. M. Go and D. P. Jones, *Transatlantic Airway Conference*, 2016, pp. S451–S461.
- 66 Y.-M. Go and D. P. Jones, *Redox Biol.*, 2014, **2**, 358–360.
- 67 D. I. Walker, Y.-M. Go, K. Liu, K. D. Pennell and D. P. Jones, *Metabolic Phenotyping in Personalized and Public Healthcare*, Academic Press, Boston, 2016, pp. 167–211.
- 68 E. B.-M. Daliri, F. K. Ofofu, R. Chelliah, B. H. Lee and D.-H. Oh, *High-Throughput*, 2020, **9**, 1–22.
- 69 J. Fessler, V. Matson and T. F. Gajewski, *J. Immuno Ther. Cancer*, 2019, **7**, 1–15.
- 70 E. A. Franzosa, A. Sirota-Madi, J. Avila-Pacheco, N. Fornelos, H. J. Haiser, S. Reinker, T. Vatanen, A. B. Hall, H. Mallick, L. J. McIver, J. S. Sauk, R. G. Wilson, B. W. Stevens, J. M. Scott, K. Pierce, A. A. Deik, K. Bullock, F. Imhann, J. A. Porter, A. Zhernakova, J. Fu, R. K. Weersma, C. Wijmenga, C. B. Clish, H. Vlamakis, C. Huttenhower and R. J. Xavier, *Nat. Microbiol.*, 2019, **4**, 1–31.
- 71 V. Saurman, K. G. Margolis and R. A. Luna, *Dig. Dis. Sci.*, 2020, **65**, 818–828.
- 72 Z. S. Hua, Y. J. Han, L. X. Chen, J. Liu, M. Hu, S. J. Li, J. L. Kuang, P. S. Chain, L. N. Huang and W. S. Shu, *ISME J.*, 2015, **9**, 1280–1294.
- 73 D. Galambos, R. E. Anderson, J. Reveillaud and J. A. Huber, *Environ. Microbiol.*, 2019, **21**, 4395–4410.
- 74 X. Zhang and D. Figeys, *J. Proteome Res.*, 2019, **18**, 2370–2380.
- 75 E. E. Canfora, R. C. Meex, K. Venema and E. E. Blaak, *Nat. Rev. Endocrinol.*, 2019, **15**, 261–273.
- 76 E. B.-M. Daliri, F. K. Ofofu, R. Chelliah, B. H. Lee and D.-H. Oh, *Biomolecules*, 2021, **11**, 1–10.
- 77 M. R. Sailani, A. A. Metwally, W. Zhou, S. M. S.-F. Rose, S. Ahadi, K. Contrepois, T. Mishra, M. J. Zhang Kidzinski, T. J. Chu and M. P. Snyder, *Nat. Commun.*, 2020, **11**, 4933.
- 78 N. D. Price, A. T. Magis, J. C. Earls, G. Glusman, R. Levy, C. Lausted, D. T. McDonald, U. Kusebauch, C. L. Moss, Y. Zhou, S. Qin, R. L. Moritz, K. Brogaard, G. S. Omenn, J. C. Lovejoy and L. Hood, *Nat. Biotechnol.*, 2017, **35**, 747–756.
- 79 K. Uppal, D. I. Walker, K. Liu, S. Li, Y.-M. Go and D. P. Jones, *Chem. Res. Toxicol.*, 2016, **29**, 1956–1975.
- 80 L. E. Lindler, *Mil. Med.*, 2015, **180**, 5–9.
- 81 I of medicine of the National Academies, *Mil. Med.*, 2015, **180**, 601–603.
- 82 C. P. Baird, *U.S. Army Med. Dep. J.*, 2012, 43–47.
- 83 M. V. Rubertone and J. F. Brundage, *Am. J. Public Health*, 2002, **92**, 1900–1904.



- 84 C. L. Perdue, A. A. Eick-Cost and M. V. Rubertone, *Mil. Med.*, 2015, **180**, 10–12.
- 85 D. P. Jones, *Toxicol. Rep.*, 2016, **3**, 29–45.
- 86 B. Warth, S. Spangler, M. Fang, C. H. Johnson, E. M. Forsberg, A. Granados, R. L. Martin, X. Domingo-Almenara, T. Huan, D. Rinehart, J. R. Montenegro-Burke, B. Hilmers, A. Aisporna, L. T. Hoang, W. Uritboonthai, H. P. Benton, S. D. Richardson, A. J. Williams and G. Siuzdak, *Anal. Chem.*, 2017, **89**, 11505–11513.
- 87 D. P. Jones, Y. Park and T. R. Ziegler, *Annu. Rev. Nutr.*, 2012, **32**, 183–202.
- 88 Y.-M. Go, D. I. Walker, Y. Liang, K. Uppal, Q. A. Soltow, V. Tran, F. Strobel, A. A. Quyyumi, T. R. Ziegler, K. D. Pennell, G. W. Miller and D. P. Jones, *Toxicol. Sci.*, 2015, **148**, 531–543.
- 89 K. Uppal, Q. A. Soltow, F. H. Strobel, W. S. Pittard, K. M. Gernert, T. Yu and D. P. Jones, *BMC Bioinf.*, 2013, **14**, 1–12.
- 90 B. Titz, J. Szostak, A. Sewer, B. Phillips, C. Nury, T. Schneider, S. Dijon, O. Lavrynenko, A. Elamin, E. Guedj, E. T. Wong, S. Lebrun, G. Vuillaume, A. Kondylis, S. Gubian, S. Cano, P. Leroy, B. Keppler, N. V. Ivanov, P. Vanscheeuwijck, F. Martin, M. C. Peitsch and J. Hoeng, *Comput. Struct. Biotechnol. J.*, 2020, **18**, 1056–1073.
- 91 J. M. Herron, K. M. Hines, H. Tomita, R. P. Seguin, J. Y. Cui and L. Xu, *Toxicol. Sci.*, 2019, **171**, 32–45.
- 92 C. Hu and W. Jia, *J. Mol. Cell Biol.*, 2021, **13**, 576–593.
- 93 J. B. Meigs, L. A. Cupples and P. W. Wilson, *Diabetes*, 2000, **49**, 2201–2207.
- 94 K. Kodama, D. Tojjar, S. Yamada, K. Toda, C. J. Patel and A. J. Butte, *Diabetes Care*, 2013, **36**, 1789–1796.
- 95 P. Poulsen, K. O. Kyvik, A. Vaag and H. Beck-Nielsen, *Diabetologia*, 1999, **42**, 139–145.
- 96 V. W. S. Wong, L. A. Adams, V. de Lédighen, G. L. H. Wong and S. Sookoian, *Nat. Rev. Gastroenterol. Hepatol.*, 2018, **15**, 461–478.
- 97 P. Dongiovanni, S. Petta, V. Mannisto, R. M. Mancina, R. Pipitone, V. Karja, M. Maggioni, P. Kakela, O. Wiklund, E. Mozzi, S. Grimaudo, D. Kaminska, R. Rametta, A. Craxi, S. Fargion, V. Nobili, S. Romeo, J. Pihlajamaki and L. Valenti, *J. Hepatol.*, 2015, **63**, 705–712.
- 98 P. Dongiovanni and L. Valenti, *Metabolism*, 2016, **65**, 1026–1037.
- 99 G. A. Baselli, P. Dongiovanni, R. Rametta, M. Meroni, S. Pelusi, M. Maggioni, S. Badiali, P. Pingitore, S. Maurotti, T. Montalcini, A. E. Taliento, D. Prati, G. Rossi, A. L. Fracanzani, R. M. Mancina, S. Romeo and L. Valenti, *Gut*, 2020, **69**, 1855–1866.
- 100 A. Leung, B. W. Parks, J. Du, C. Trac, R. Setten, Y. Chen, K. Brown, A. J. Lusic, R. Natarajan and D. E. Schones, *J. Biol. Chem.*, 2014, **289**, 23557–23567.
- 101 J. A. Kolberg, T. Jørgensen, R. W. Gerwien, S. Hamren, M. P. McKenna, E. Moler, M. W. Rowe, M. S. Urdea, X. M. Xu, T. Hansen, O. Pedersen and K. Borch-Johnsen, *Diabetes Care*, 2009, **32**, 1207–1212.
- 102 M. H. Miller, S. V. Walsh, A. Atrih, J. T.-J. Huang, M. A. J. Ferguson and J. F. Dillon, *J. Gastroenterol. Hepatol.*, 2014, **29**, 1839–1847.
- 103 E. Haythorne, M. Rohm, M. van de Bunt, M. F. Brereton, A. I. Tarasov, T. S. Blacker, G. Sachse, M. Silva dos Santos, R. Terron Exposito, S. Davis, O. Baba, R. Fischer, M. R. Duchon, P. Rorsman, J. I. MacRae and F. M. Ashcroft, *Nat. Commun.*, 2019, **10**, 2474.
- 104 F. Sacco, A. Seelig, S. J. Humphrey, N. Krahmer, F. Volta, A. Reggion, P. Marchetti, J. Gerdes and M. Mann, *Cell Metab.*, 2019, **29**, 1422–1432.e3.
- 105 M. Murri, M. Insenser, M. R. Bernal-Lopez, P. Perez-Martinez, H. F. Escobar-Morreale and F. J. Tinahones, *Mol. Cell. Endocrinol.*, 2013, **376**, 99–106.
- 106 M. Murri, M. Insenser, M. Luque, F. J. Tinahones and H. F. Escobar-Morreale, *Expert Rev. Proteomics*, 2014, **11**, 491–502.
- 107 S.-J. Kim, S. Chae, H. Kim, D.-G. Mun, S. Back, H. Y. Choi, K. S. Park, D. Hwang, S. H. Choi and S.-W. Lee, *Mol. Cell. Proteomics*, 2014, **13**, 811–822.
- 108 M. Gómez-Serrano, E. Camafeita, E. Garca-Santos, J. A. López, M. A. Rubio, A. Sánchez-Pernaute, A. Torres, J. Vázquez and B. Peral, *Sci. Rep.*, 2016, **6**, 25756.
- 109 A. A. Alfadda, A. Masood and M. Y. Al-Naami, Pierre Chaurand and Hicham Benabdelkamel, *Mol. Cells*, 2017, **40**, 685–695.
- 110 D. A. Bowes, E. M. Driver, S. Savic, Q. Cheng, C. M. Whisner, R. Krajmalnik-Brown and R. U. Halden, *Nat. Food*, 2023, **4**, 257–266.
- 111 K. O. Poulsen, J. Astono, R. R. Jakobsen, N. Uldbjer, J. Fuglsang, D. S. Nielsen and U. K. Sundekilde, *BMJ Open*, 2022, **12**, 1–10.
- 112 J. M. Ordovas, L. R. Ferguson, E. S. Tai and J. C. Mathers, *BMJ*, 2018, **361**, 1–7.
- 113 T. Tsuda, Y. Ueno, T. Yoshikawa, H. Kojo and T. Osawa, *Biochem. Pharmacol.*, 2006, **71**, 1184–1197.
- 114 C. Picó, F. Serra, A. M. Rodríguez, J. Keijer and A. Palou, *Nutrients*, 2019, **11**, 1–30.
- 115 K. A. Soni, R. Nannapaneni and T. Tasara, *Foodborne Pathog. Dis.*, 2011, **8**, 843–852.
- 116 U. Chand Jha, H. Nayyar, N. Mantri and K. H. M. Siddique, *Cells*, 2021, **10**, 1674.
- 117 S. Sauer and T. Luge, *Proteomics*, 2015, **15**, 997–1013.
- 118 M. Kussmann, *Curr. Pharmacogenomics Pers. Med.*, 2010, **8**, 245–256.
- 119 M. Van Der Werf, F. Schuren, S. Bijlsma, A. Tas and B. V. Ommen, *J. Food Sci.*, 2001, **66**, 772–780.
- 120 H. A. Kuiper, G. A. Kleter, H. P. J. M. Noteborn and E. J. Kok, *Plant J.*, 2001, **27**, 503–528.
- 121 M. Galvani, M. Hamdan and P. G. Righetti, *Rapid Commun. Mass Spectrom.*, 2001, **15**, 258–264.
- 122 J. E. Plowman, W. G. Bryson and T. W. Jordan, *Electrophoresis*, 2000, **21**, 1899–1906.
- 123 A. Johansson-Persson, T. Barri, M. Ulmius, G. Önning and L. O. Dragsted, *Anal. Bioanal. Chem.*, 2013, **405**, 4799–4809.
- 124 E. J. van Velzen, J. A. Westerhuis, J. P. M. van Duynhoven, F. A. van Dorsten, C. H. Grün, D. M. Jacobs, G. S. M. J.



- E. Duchateau, D. J. Vis and A. K. Smilde, *J. Proteome Res.*, 2009, **8**, 3317–3330.
- 125 K. Redeuil, C. Smarrito-Menozi, P. Guy, S. Rezzi, F. Dionisi, G. Williamson, K. Nagy and M. Renouf, *J. Chromatogr. A*, 2011, **1218**, 4678–4688.
- 126 G. McCombie, L. M. Browning, C. M. Titman, M. Song, J. Shockcor, S. A. Jebb and J. L. Griffin, *Metabolomics*, 2009, **5**, 363–374.
- 127 I. Rubio-Aliaga, B. d Roos, M. Sailer, G. A. McLoughlin, M. V. Boekschoten, M. van Erk, E.-M. Bachmair, E. M. van Schothorst, J. Keijer, S. L. Coort, C. Evelo, M. J. Gibney, H. Daniel, M. Muller, R. Kleemann and L. Brennan, *Physiol. Genomics*, 2011, **43**, 408–416.
- 128 L. Brennan and F. B. Hu, *Mol. Nutr. Food Res.*, 2019, **63**, 1701064.
- 129 C. Collins, A. E. McNamara and L. Brennan, *Proc. Nutr. Soc.*, 2019, **78**, 189–196.
- 130 D. González-Peña and L. Brennan, *Annu. Rev. Food Sci. Technol.*, 2019, **10**, 479–519.
- 131 J. Li, C. Guo, Y. Liu, B. Han, Z. Lv, H. Jiang, S. Li and Z. Zhang, *J. Adv. Res.*, 2024, DOI: [10.1016/j.jare.2024.01.019](https://doi.org/10.1016/j.jare.2024.01.019).
- 132 L. Liu, M. Xiang, X. Cai, B. Wu, C. Chen, N. Cai and D. Ao, *BMC Microbiol.*, 2023, **23**, 387.
- 133 N. Bekiares, C. G. Krueger, J. J. Meudt, D. Shanmuganayagam and J. D. Reed, *OMICS: J. Integr. Biol.*, 2018, **22**, 145–153.
- 134 T. V. Arumugam, A. Alli-Shaik, E. A. Liehn, S. Selvaraji, L. Poh, V. Rajeev, Y. Cho, Y. Cho, J. Kim, J. Kim, H. L. Swa, D. T. Z. Hao, C. Rattanasopa, D. Y.-W. Fann, D. C. Mayan, G. Y.-Q. Ng, S.-H. Baik, K. Mallilankaraman, M. Gelderblom, G. R. Drummond, C. G. Sobey, B. K. Kennedy, R. R. Singaraja, M. P. Mattson, D.-G. Jo and J. Gunaratne, *eLife*, 2023, **12**, 1–31.
- 135 K. K. Jain, *Biomarkers of Infectious Diseases*, Springer New York, New York, NY, 2017, pp. 219–238.
- 136 M. R. Hasan, M. Suleiman and A. Pérez-López, *Front. Genet.*, 2021, **12**, 1–14.
- 137 J. Rello, E. Storti, M. Belliato and R. Serrano, *Eur. Respir. J.*, 2020, **55**, 2001028.
- 138 M. Ni, F.-B. Tian, D.-D. Xiang and B. Yu, *J. Med. Virol.*, 2020, **92**, 2600–2606.
- 139 M. Aggarwal, J. Dass and M. Mahapatra, *Indian J. Hematol. Blood Transfusion*, 2020, **36**, 616–662.
- 140 C. Shi, C. Wang, H. Wang, C. Yang, F. Cai, F. Zeng, F. Cheng, Y. Liu, T. Zhou, B. Deng, I. Vlodaysky, J.-P. Li and Y. Zhang, *Clin. Transl. Sci.*, 2020, **13**, 1087–1095.
- 141 H. Meng, A. Sengupta, E. Ricciotti, A. Mrčela, D. Mathew, L. L. Mazaleuskaya, S. Ghosh, T. G. Brooks, A. P. Turner, A. S. Schanoski, N. F. Lahens, A. W. Tan, A. Woolfork, G. Grant, K. Susztak, A. G. Letizia, S. C. Sealfon, E. J. Wherry, K. Laudanski, A. M. Weljie, N. J. Meyer and G. A. FitzGerald, *Clin. Transl. Med.*, 2023, **13**, e1440.
- 142 S. Kugler, L. Hahnefeld, J. A. Kloka, S. Ginzel, E. Nürenberg-Goloub, S. Zinn, M. J. Vehreschild, K. Zacharowski, S. Lindau, E. Ullrich, J. Burmeister, J. Kohlhammer, J. Schwäble, R. Gurke, E. Dorocho, A. Bennett, S. Dauth, J. Campe, T. Knape, V. Laux, A. Kannt, M. Köhm, G. Geisslinger, E. Resch and F. Behrens, *Talanta*, 2024, **268**, 125295.
- 143 M. Bosnjak, A. V. Karpe, T. T. H. Van, D. Kotsanas, G. A. Jenkin, S. P. Costello, P. Johannesen, R. J. Moore, D. J. Beale, Y. N. Srikhanta, E. A. Palombo, S. Larcombe and D. Lyras, *Nat. Commun.*, 2023, **14**, 7737.
- 144 M. Noszka, A. Strzalka, J. Muraszko, R. Kolenda, C. Meng, C. Ludwig, K. Stingl and A. Zawilak-Pawlik, *Nat. Commun.*, 2023, **14**, 6715.
- 145 L. Zeng, G. Yu, K. Yang, Q. He, W. Hao, W. Xiang, Z. Long, H. Chen, X. Tang and L. Sun, *Sci. Rep.*, 2024, **14**, 1604.
- 146 O. Hansson, *Nat. Med.*, 2021, **27**, 954–963.
- 147 M. Khalil, C. E. Teunissen, M. Otto, F. Piehl, M. P. Sormani, T. Gatringer, C. Barro, L. Kappos, M. Comabella, F. Fazekas, A. Petzold, K. Blennow, H. Zetterberg and J. Kuhle, *Nat. Rev. Neurol.*, 2018, **14**, 577–589.
- 148 N. J. Ashton, S. Janelidze, A. A. Khleifat, A. Leuzy, E. L. van der Ende, T. K. Karikari, A. L. Benedet, T. A. Pascoal, A. Lleó, L. Parnetti, D. Galimberti, L. Bonanni, A. Pilotto, A. Padovani, J. Lycke, L. Novakova, M. Axelsson, L. Velayudhan, G. D. Rabinovici, B. Miller, C. Pariante, N. Nikkheslat, S. M. Resnick, M. Thambisetty, M. Schöll, G. Fernández-Eulate, F. J. Gil-Bea, A. L. de Munain, A. Al-Chalabi, P. Rosa-Neto, A. Strydom, P. Svenningsson, E. Stomrud, A. Santillo, D. Aarsland, J. C. van Swieten, S. Palmqvist, H. Zetterberg, K. Blennow, A. Hye and O. Hansson, *Nat. Commun.*, 2021, **12**, 1–12.
- 149 O. Preische, S. A. Schultz, A. Apel, J. Kuhle, S. A. Kaeser, C. Barro, S. Gräber, E. Kuder-Buletta, C. LaFougere, C. Laske, J. Vöglein, J. Levin, C. Masters, R. Martins, P. Schofield, M. N. Rossor, N. Graff-Radford, S. Salloway, B. Ghetti, J. Ringman, J. Noble, J. Chhatwal, A. Goate, T. Benzinger, J. Morris, R. Bateman, G. Wang, A. Fagan, E. McDade, B. Gordon, M. Jucker, R. Allegri, F. Amtashar, S. Berman, C. Bodge, S. Brandon, W. Brooks, J. Buck, V. Buckles, S. Chea, P. Chrem, H. Chui, J. Cinco, J. Clifford, C. Cruchaga, M. DMello, T. Donahue, J. Douglas, N. Edigo, N. Erekin-Taner, M. Farlow, A. Farrar, H. Feldman, G. Flynn, N. Fox, E. Franklin, H. Fujii, C. Gant, S. Gardener, J. Goldman, J. Gray, J. Gurney, J. Hassenstab, M. Hirohara, D. Holtzman, R. Hornbeck, S. H. DiBari, T. Ikeuchi, S. Ikonovic, G. Jerome, C. Karch, K. Kasuga, T. Kawarabayashi, W. Klunk, R. Koeppe, J. H. Lee, D. Marcus, N. S. Mason, D. Maue-Dreyfus, L. Montoya, H. Mori, A. Nagamatsu, K. Neimeyer, J. Norton, R. Perrin, M. Raichle, J. H. Roh, H. Shimada, T. Shiroto, M. Shoji, W. Sigurdson, H. Sohrabi, P. Sparks, K. Suzuki, L. Swisher, K. Taddei, J. Wang, P. Wang, M. Weiner, M. Wolfsberger, C. Xiong and X. Xu, *Nat. Med.*, 2019, **25**, 277–283.
- 150 B. Delcoigne, A. Manouchehrinia, C. Barro, P. Benkert, Z. Michalak, L. Kappos, D. Leppert, J. A. Tsai, T. Plavina, B. C. Kieseier, J. Lycke, L. Alfredsson, I. Kockum, J. Kuhle, T. Olsson and F. Piehl, *Neurology*, 2020, **94**, e1201–e1212.



- 151 B. Olsson, L. Alberg, N. C. Cullen, E. Michael, L. Wahlgren, A. K. Kroksmark, K. Rostasy, K. Blennow, H. Zetterberg and M. Tullinius, *J. Neurol.*, 2019, **266**, 2129–2136.
- 152 J. R. Anderson, E. Johnson, R. Jenkins, S. Jacobsen, D. Green, M. Walters, L. Bundgaard, B. A. C. Hausmans, G. van den Akker, T. J. M. Welting, A. Chabronova, Y. A. Kharaz, E. J. Clarke, V. James and M. J. Peffers, *Int. J. Mol. Sci.*, 2023, **24**, 1–24.
- 153 L. Wei, J. Gao, L. Wang, Q. Tao and C. Tu, *Hum. Mol. Genet.*, 2023, **33**, 122–137.
- 154 A. Pascual-Alonso, C. Xiol, D. Smirnov, R. Kopajtich, H. Prokisch and J. Armstrong, *Hum. Genomics*, 2023, **17**, 85.
- 155 M. I. Abdel-Aziz, A. H. Neerincx, S. J. Vijverberg, A. D. Kraneveld and A. H. M. van der Zee, *Semin. Immunopathol.*, 2020, **42**, 111–126.
- 156 M. A. Ferreira, J. M. Vonk, H. Baurecht, I. Marenholz, C. Tian, J. D. Hoffman, Q. Helmer, A. Tillander, V. Ullemar, J. V. Dongen, Y. Lu, F. Rüschenhoff, J. Esparza-Gordillo, C. W. Medway, E. Mountjoy, K. Burrows, O. Hummel, S. Grosche, B. M. Brumpton, J. S. Witte, J. J. Hottenga, G. Willemsen, J. Zheng, E. Rodriguez, M. Hotze, A. Franke, J. A. Revez, J. Beesley, M. C. Matheson, S. C. Dharmage, L. M. Bain, L. G. Fritsche, M. E. Gabrielsen, B. Balliu, J. B. Nielsen, W. Zhou, K. Hveem, A. Langhammer, O. L. Holmen, M. Løset, G. R. Abecasis, C. J. Willer, A. Arnold, G. Homuth, C. O. Schmidt, P. J. Thompson, N. G. Martin, D. L. Duffy, N. Novak, H. Schulz, S. Karrasch, C. Gieger, K. Strauch, R. B. Melles, D. A. Hinds, N. Hübner, S. Weidinger, P. K. Magnusson, R. Jansen, E. Jorgenson, Y. A. Lee, D. I. Boomsma, C. Almqvist, R. Karlsson, G. H. Koppelman and L. Paternoster, *Nat. Genet.*, 2017, **49**, 1–26.
- 157 B. R. Joubert, J. F. Felix, P. Yousefi, K. M. Bakulski, A. C. Just, C. Breton, S. E. Reese, C. A. Markunas, R. C. Richmond, C.-J. Xu, L. K. Küpers, S. S. Oh, C. Hoyo, O. Gruziova, C. Söderhäll, L. A. Salas, N. Baiz, H. Zhang, J. Lepeule, C. Ruiz, S. Lighthart, T. Wang, J. A. Taylor, L. Duijts, G. C. Sharp, S. A. Jankipersadsing, R. M. Nilsen, A. Vaez, M. D. Fallin, D. Hu, A. A. Litonjua, B. F. Fuemmeler, K. Huen, J. Kere, I. Kull, M. C. Munthe-Kaas, U. Gehring, M. Bustamante, M. J. Saurel-Coubizolles, B. M. Quraishi, J. Ren, J. Tost, J. R. Gonzalez, M. J. Peters, S. E. Håberg, Z. Xu, J. B. van Meurs, T. R. Gaunt, M. Kerkhof, E. Corpeleijn, A. P. Feinberg, C. Eng, A. A. Baccarelli, S. B. Neelon, A. Bradman, S. K. Merid, A. Bergström, Z. Hereceg, H. Hernandez-Vargas, B. Brunekreef, M. Pinart, B. Heude, S. Ewart, J. Yao, N. Lemonnier, O. H. Franco, M. C. Wu, A. Hofman, W. McArdle, P. V. der Vlies, F. Falahi, M. W. Gillman, L. F. Barcellos, A. Kumar, M. Wickman, S. Guerra, M.-A. Charles, J. Holloway, C. Auffray, H. W. Tiemeier, G. D. Smith, D. Postma, M.-F. Hivert, B. Eskenazi, M. Vrijheid, H. Arshad, J. M. Antó, A. Dehghan, W. Karmaus, I. Annesi-Maesano, J. Sunyer, A. Ghantous, G. Pershagen, N. Holland, S. K. Murphy, D. L. DeMeo, E. G. Burchard, C. Ladd-Acosta, H. Snieder, W. Nystad, G. H. Koppelman, C. L. Relton, V. W. Jaddoe, A. Wilcox, E. Melén and S. J. London, *Am. J. Hum. Genet.*, 2016, **98**, 680–696.
- 158 O. Gruziova, C.-J. Xu, C. V. Breton, I. Annesi-Maesano, J. M. Antó, C. Auffray, S. Ballereau, T. Bellander, J. Bousquet, M. Bustamante, M.-A. Charles, Y. de Kluizenaar, H. T. den Dekker, L. Duijts, J. F. Felix, U. Gehring, M. Guxens, V. W. Jaddoe, S. A. Jankipersadsing, S. K. Merid, J. Kere, A. Kumar, N. Lemonnier, J. Lepeule, W. Nystad, C. M. Page, S. Panasevich, D. Postma, R. Slama, J. Sunyer, C. Söderhäll, J. Yao, S. J. London, G. Pershagen, G. H. Koppelman and E. Melén, *Environ. Health Perspect.*, 2017, **125**, 104–110.
- 159 A. Singhanian, J. C. Wallington, C. G. Smith, D. Horowitz, K. J. Staples, P. H. Howarth, S. D. Gadola, R. Djukanovi, C. H. Woelk and T. S. Hinks, *Am. J. Respir. Cell Mol. Biol.*, 2018, **58**, 261–270.
- 160 D. Lefaudeux, B. De Meulder, M. J. Loza, N. Peffer, A. Rowe, F. Baribaud, A. T. Bansal, R. Lutter, A. R. Sousa, J. Corfield, I. Pandis, P. S. Bakke, M. Caruso, P. Chanez, S.-E. Dahlén, L. J. Fleming, S. J. Fowler, I. Horvath, N. Krug, P. Montuschi, M. Sanak, T. Sandstrom, D. E. Shaw, F. Singer, P. J. Sterk, G. Roberts, I. M. Adcock, R. Djukanovic, C. Auffray, K. F. Chung, N. Adriaens, H. Ahmed, A. Aliprantis, K. Alving, P. Badorek, D. Balgoma, C. Barber, A. Bautmans, A. F. Behndig, E. Bel, J. Beleta, A. Berglind, A. Berton, J. Bigler, H. Bisgaard, G. Bochenek, M. J. Boedigheimer, K. Bønnelykke, J. Brandsma, A. Braun, P. Brinkman, D. Burg, D. Campagna, L. Carayannopoulos, J. P. Carvalho da Purificação Rocha, A. Chaiboonchoe, R. Chaleckis, C. Coleman, C. Compton, A. D'Amico, B. Dahlén, J. De Alba, P. de Boer, I. De Lepeleire, T. Dekker, I. Delin, P. Dennison, A. Dijkhuis, A. Draper, J. Edwards, R. Emma, M. Ericsson, V. Erpenbeck, D. Erzen, C. Faulenbach, K. Fichtner, N. Fitch, B. Flood, U. Frey, M. Gahlemann, G. Galfy, H. Gallart, T. Garret, T. Geiser, J. Gent, M. Gerhardsson de Verdier, D. Gibeon, C. Gomez, K. Gove, N. Gozzard, Y.-K. Guo, S. Hashimoto, J. Haughney, G. Hedlin, P.-P. Hekking, E. Henriksson, L. Hewitt, T. Higgenbottam, U. Hoda, J. Hohlfield, C. Holweg, P. Howarth, R. Hu, S. Hu, X. Hu, V. Hudson, A. J. James, J. Kamphuis, E. J. Kennington, D. Kerry, M. Klüglich, H. Knobel, R. Knowles, A. Knox, J. Kolmert, J. Konradsen, M. Kots, L. Krueger, S. Kuo, M. Kupczyk, B. Lambrecht, A.-S. Lantz, L. Larsson, N. Lazarinis, S. Lone-Satif, L. Marouzet, J. Martin, S. Masefield, C. Mathon, J. G. Matthews, A. Mazein, S. Meah, A. Maiser, A. Menzies-Gow, L. Metcalf, R. Middelveld, M. Mikus, M. Miralpeix, P. Monk, N. Mores, C. S. Murray, J. Musial, D. Myles, S. Naz, K. Nething, B. Nicholas, U. Nihlen, P. Nilsson, B. Nordlund, J. Östling, A. Pacino, L. Pahuš, S. Palkonnen, S. Pavlidis, G. Pennazza, A. Petrén, S. Pink, A. Postle, P. Powel, M. Rahman-Amin, N. Rao, L. Ravanetti, E. Ray, S. Reinke, L. Reynolds, K. Riemann, J. Riley,



- M. Robberechts, A. Roberts, C. Rossios, K. Russell, M. Rutgers, G. Santini, M. Sentoninco, C. Schoelch, J. P. Schofield, W. Seibold, R. Sigmund, M. Sjödin, P. J. Skipp, B. Smids, C. Smith, J. Smith, K. M. Smith, P. Söderman, A. Sogbesan, D. Staykova, K. Strandberg, K. Sun, D. Supple, M. Szentkereszty, L. Tamasi, K. Tariq, J.-O. Thörngren, B. Thornton, J. Thorsen, S. Valente, W. van Aalderen, M. van de Pol, K. van Drunen, M. van Geest, J. Versnel, J. Vestbo, A. Vink, N. Vissing, C. von Garnier, A. Wagerner, S. Wagers, F. Wald, S. Walker, J. Ward, Z. Weiszhart, K. Wetzler, C. E. Wheelock, C. Wiegman, S. Williams, S. J. Wilson, A. Woodcock, X. Yang, E. Yeyashingham, W. Yu, W. Zetterquist and K. Zwinderman, *J. Allergy Clin. Immunol.*, 2017, **139**, 1797–1807.
- 161 K. Takahashi, S. Pavlidis, F. N. K. Kwong, U. Hoda, C. Rossios, K. Sun, M. Loza, F. Baribaud, P. Chanez, S. J. Fowler, I. Horvath, P. Montuschi, F. Singer, J. Musial, B. Dahlen, S.-E. Dahlen, N. Krug, T. Sandstrom, D. E. Shaw, R. Lutter, P. Bakke, L. J. Fleming, P. H. Howarth, M. Caruso, A. R. Sousa, J. Corfield, C. Auffray, B. D. Meulder, D. Lefaudeux, R. Djukanovic, P. J. Sterk, Y. Guo, I. M. Adcock and K. F. Chung, *Eur. Respir. J.*, 2018, **51**, 1–19.
- 162 R. S. Kelly, A. Dahlin, M. J. McGeachie, W. Qiu, J. Sordillo, E. S. Wan, A. C. Wu and J. Lasky-Su, *Chest*, 2017, **151**, 262–277.
- 163 L. D. Bos, P. J. Sterk and S. J. Fowler, *J. Allergy Clin. Immunol.*, 2016, **138**, 970–976.
- 164 A. H. Neerinx, S. J. H. Vijverberg, L. D. J. Bos, P. Brinkman, M. P. van der Schee, R. de Vries, P. J. Sterk and A.-H. Maitland-van der Zee, *Pediatr. Pulmonol.*, 2017, **52**, 1616–1627.
- 165 J. Li, Y. Wang, W. Zhao, T. Yang, Q. Zhang, H. Yang, X. Li and Z. Tong, *Respir. Res.*, 2024, **25**, 1–15.
- 166 J. Qu, L. Lin, G. Fu, M. Zheng, J. Geng, X. Sun and L. Xing, *Life Sci.*, 2024, **339**, 122422.
- 167 K. Wang, Y. Xiao, W. Zhang, H. Yang, C. Li, J. Wang and G. Li, *Environ. Toxicol.*, 2024, DOI: [10.1002/tox.24117](https://doi.org/10.1002/tox.24117).
- 168 Y. Xie, X. Pan, Z. Wang, H. Ma, W. Xu, H. Huang, J. Zhang, X. Wang and C. Lian, *J. Cancer*, 2024, **15**, 776–795.
- 169 N. Psychogios, D. D. Hau, J. Peng, A. C. Guo, R. Mandal, S. Bouatra, I. Sinelnikov, R. Krishnamurthy, R. Eisner, B. Gautam, N. Young, J. Xia, C. Knox, E. Dong, P. Huang, Z. Hollander, T. L. Pedersen, S. R. Smith, F. Bamforth, R. Greiner, B. McManus, J. W. Newman, T. Goodfriend and D. S. Wishart, *PLoS One*, 2011, **6**, 1–23.
- 170 G. H. Grant and W. R. Butt, *Immunochemical Methods in Clinical Chemistry*, Elsevier, 1970, vol. 13, pp. 383–466.
- 171 L. G. Longworth, T. Shedlovsky and D. A. MacInnes, *J. Exp. Med.*, 1939, **70**, 399–413.
- 172 C. Bagnis, *Blood Rev.*, 2015, **29**, 163–169.
- 173 D. Ombrone, E. Giocaliere, G. Forni, S. Malvagìa and G. la Marca, *Mass Spectrom. Rev.*, 2016, **35**, 71–84.
- 174 S. McGrath, D. Christidis, M. Perera, S. K. Hong, T. Manning, I. Vela and N. Lawrentschuk, *Prostate Int.*, 2016, **4**, 130–135.
- 175 M. J. Levine, *Ann. N. Y. Acad. Sci.*, 1993, **694**, 11–16.
- 176 P. D. V. D. Almeida, A. M. T. Grégio, M. Ângela Naval Machado, A. A. S. D. Lima and L. R. Azevedo, *J. Contemp. Dent Pract.*, 2008, **9**, 1–11.
- 177 K. Ngamchuea, K. Chaisiwamongkhol, C. Batchelor-McAuley and R. G. Compton, *Analyst*, 2018, **143**, 81–99.
- 178 M. François, A. Karpe, J.-W. Liu, D. Beale, M. Hor, J. Hecker, J. Faunt, J. Maddison, S. Johns, J. Doecke, S. Rose and W. R. Leifert, *J. Alzheimer's Dis.*, 2021, **82**, 1301–1313.
- 179 D. Tao, B. McGill, T. Hamerly, T. Kobayashi, P. Khare, A. Dzedzic, T. Leski, A. Holtz, B. Shull, A. E. Jedlicka, A. Walzer, P. D. Slowey, C. C. Slowey, S. E. Nsango, D. A. Stenger, M. Chaponda, M. Mulenga, K. H. Jacobsen, D. J. Sullivan, S. J. Ryan, R. Ansumana, W. J. Moss, I. Morlais and R. R. Dinglasan, *Sci. Translat. Med.*, 2019, **11**, eaan4479.
- 180 K. Sato, W. Kang, K. Saga and K. Sato, *J. Am. Acad. Dermatol.*, 1989, **20**, 537–563.
- 181 A. Mena-Bravo and M. Luque de Castro, *J. Pharm. Biomed. Anal.*, 2014, **90**, 139–147.
- 182 M. Picardo, M. Ottaviani, E. Camera and A. Mastrofrancesco, *Dermatoendocrinol.*, 2009, **1**, 68–71.
- 183 D. K. Trivedi, E. Sinclair, Y. Xu, D. Sarkar, C. Walton-Doyle, C. Liscio, P. Banks, J. Milne, M. Silverdale, T. Kunath, R. Goodacre and P. Barran, *ACS Cent. Sci.*, 2019, **5**, 599–606.
- 184 S. Briganti, M. Truglio, A. Angiolillo, S. Lombardo, D. Leccese, E. Camera, M. Picardo and A. Di Costanzo, *Metabolites*, 2021, **11**, 1–18.
- 185 K. Y. Garza, A. A. R. Silva, J. R. Rosa, M. F. Keating, S. C. Povilaitis, M. Spradlin, P. H. G. Sanches, A. Varão Moura, J. Marrero Gutierrez, J. Q. Lin, J. Zhang, R. J. DeHoog, A. Bensussan, S. Badal, D. Cardoso de Oliveira, P. H. Dias Garcia, L. Dias de Oliveira Negrini, M. A. Antonio, T. C. Canevari, M. N. Eberlin, R. Tibshirani, L. S. Eberlin and A. M. Porcari, *Anal. Chem.*, 2021, **93**, 12582–12593.
- 186 S. Bouatra, F. Aziat, R. Mandal, A. C. Guo, M. R. Wilson, C. Knox, T. C. Bjorndahl, R. Krishnamurthy, F. Saleem, P. Liu, Z. T. Dame, J. Poelzer, J. Huynh, F. S. Yallou, N. Psychogios, E. Dong, R. Bogumil, C. Roehring and D. S. Wishart, *PLoS One*, 2013, **8**, 1–28.
- 187 I. Sarosiek, R. Schicho, P. Blandon and M. Bashashati, *World J. Gastrointest. Oncol.*, 2016, **8**, 3674–3685.
- 188 Y. Ozsoy, S. Gungor and E. Cevher, *Molecules*, 2009, **14**, 3754–3779.
- 189 C. Bachert, M. Wagenmann, C. Rudack, K. Höpken, M. Hildebrandt, D. Wang and P. van Cauwenberge, *Allergy*, 1998, **53**, 2–13.
- 190 S. J. Wilson, L. Lau and P. H. Howarth, *Clin. Exp. Allergy*, 1998, **28**, 220–227.
- 191 R. M. Naclerio, D. Proud, A. G. Togias, N. F. Adkinson, D. A. Meyers, A. Kagey-Sobotka, M. Plaut, P. S. Norman and L. M. Lichtenstein, *N. Engl. J. Med.*, 1985, **313**, 65–70.
- 192 M. Raulf-Heimsoth, C. Wirtz, F. Papenfuss and X. Baur, *Clin. Exp. Allergy*, 2000, **30**, 110–121.



- 193 C. Svensson, R. Grönneberg, M. Andersson, U. Alkner, O. Andersson, B. Billing, H. Gilljam, L. Greiff and C. G. A. Persson, *J. Allergy Clin. Immunol.*, 1995, **96**, 239–246.
- 194 R. Hauser, S. Elreedy, J. A. Hoppin and D. C. Christiani, *Occup. Environ. Med.*, 1995, **52**, 353–358.
- 195 H. S. Koren, G. E. Hatch and D. E. Graham, *Toxicology*, 1990, **60**, 15–25.
- 196 A. Muttray, L. Klimek, M. Faas, D. Schäfer, W. Mann and J. Konietzko, *Int. Arch. Occup. Environ. Health*, 1999, **72**, 485–488.
- 197 F. M. Nachtigall, A. Pereira, O. S. Trofymchuk and L. S. Santos, *Nat. Biotechnol.*, 2020, **38**, 1168–1173.
- 198 O. Lawal, W. M. Ahmed, T. M. E. Nijssen, R. Goodacre and S. J. Fowler, *Metabolomics*, 2017, **13**, 110.
- 199 H. Chen, X. Qi, J. Ma, C. Zhang, H. Feng and M. Yao, *medRxiv*, 2020.
- 200 S. Grassin-Delyle, C. Roquencourt, P. Moine, G. Saffroy, S. Carn, N. Heming, J. Fleuriet, H. Salvator, E. Naline, L.-J. Couderc, P. Devillier, E. A. Thévenot and D. Annane, *eBioMedicine*, 2021, **63**, 103154.
- 201 J. Wigmore and R. M. Langille, *J. – Can. Soc. Forensic Sci.*, 2009, **42**, 276–283.
- 202 A. Eisenmann, A. Amann, M. Said, B. Datta and M. Ledochowski, *J. Breath Res.*, 2008, **2**, 046002.
- 203 V. Savarino, S. Vigneri and G. Celle, *Gut*, 1999, **45**, I18–I22.
- 204 J. Lawson, B. Boyle and J. Beauchamp, *J. Breath Res.*, 2020, **14**, 030202.
- 205 Y. Pang, M. V. Holmes, Z. Chen and C. Kartsonaki, *J. Gastroenterol. Hepatol.*, 2019, **34**, 330–345.
- 206 L. von Knorring and H. Mörnstad, *Neuropsychobiology*, 2008, **15**, 146–154.
- 207 D. A. Kidwell and F. P. Smith, *Forensic Sci. Int.*, 2001, **116**, 89–106.
- 208 B. M. Appenzeller, C. Schummer, S. B. Rodrigues and R. Wennig, *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.*, 2007, **852**, 333–337.
- 209 A. Scalbert, L. Brennan, O. Fiehn, T. Hankemeier, B. S. Kristal, B. van Ommen, E. Pujos-Guillot, E. Verheij, D. Wishart and S. Wopereis, *Metabolomics*, 2009, **5**, 435–458.
- 210 R. Poesen, K. Windey, E. Neven, D. Kuypers, V. De Preter, P. Augustijns, P. D'Haese, P. Evenepoel, K. Verbeke and B. Meijers, *J. Am. Soc. Nephrol.*, 2016, **27**, 1389–1399.
- 211 H. Riechelmann, T. Deutschle, E. Friemel, H.-J. Gross and M. Bachem, *Eur. Respir. J.*, 2003, **21**, 600–605.
- 212 N. J. Rattray, Z. Hamrang, D. K. Trivedi, R. Goodacre and S. J. Fowler, *Trends Biotechnol.*, 2014, **32**, 538–548.
- 213 J. D. Beauchamp and J. D. Pleil, *J. Breath Res.*, 2013, **7**, 042001.
- 214 N. M. Grob and R. A. Dweik, *J. Breath Res.*, 2008, **2**, 037002.
- 215 O. A. H. Jones, *Metabolomics*, 2018, **14**, 101.
- 216 J. R. Nuñez, S. M. Colby, D. G. Thomas, M. M. Tfaily, N. Tolic, E. M. Ulrich, J. R. Sobus, T. O. Metz, J. G. Teeguarden and R. S. Renslow, *J. Chem. Inf. Model.*, 2019, **59**, 4052–4060.
- 217 D. T. Snyder, C. J. Pulliam, Z. Ouyang and R. G. Cooks, *Anal. Chem.*, 2016, **88**, 2–29.
- 218 L. Gao, Q. Song, G. E. Patterson, R. G. Cooks and Z. Ouyang, *Anal. Chem.*, 2006, **78**, 5994–6002.
- 219 L. Gao, A. Sugiarto, J. D. Harper, R. G. Cooks and Z. Ouyang, *Anal. Chem.*, 2008, **80**, 7198–7205.
- 220 L. Li, T.-C. Chen, Y. Ren, P. I. Hendricks, R. G. Cooks and Z. Ouyang, *Anal. Chem.*, 2014, **86**, 2909–2916.
- 221 P. I. Hendricks, J. K. Dalgleish, J. T. Shelley, M. A. Kirleis, M. T. McNicholas, L. Li, T. C. Chen, C. H. Chen, J. S. Duncan, F. Boudreau, R. J. Noll, J. P. Denton, T. A. Roach, Z. Ouyang and R. G. Cooks, *Anal. Chem.*, 2014, **86**, 2900–2908.
- 222 M. He, Z. Xue, Y. Zhang, Z. Huang, X. Fang, F. Qu, Z. Ouyang and W. Xu, *Anal. Chem.*, 2015, **87**, 2236–2241.
- 223 Y. Zhai, Y. Feng, Y. Wei, Y. Wang and W. Xu, *Analyst*, 2015, **140**, 3406–3414.
- 224 S. Kumano, M. Sugiyama, M. Yamada, K. Nishimura, H. Hasegawa, H. Morokuma, H. Inoue and Y. Hashimoto, *Anal. Chem.*, 2013, **85**, 5033–5039.
- 225 S. A. Ecelberger, T. J. Cornish, B. F. Collins, D. L. Lewis and W. A. Bryden, *Johns Hopkins APL Tech. Dig.*, 2004, **25**, 14–19.
- 226 M. Yang, T.-Y. Kim, H.-C. Hwang, S.-K. Yi and D.-H. Kim, *J. Am. Soc. Mass Spectrom.*, 2008, **19**, 1442–1448.
- 227 J. A. Contreras, J. A. Murray, S. E. Tolley, J. L. Oliphant, H. D. Tolley, S. A. Lammert, E. D. Lee, D. W. Later and M. L. Lee, *J. Am. Soc. Mass Spectrom.*, 2008, **19**, 1425–1434.
- 228 S. A. Lammert, *Detection of Chemical, Biological, Radiological and Nuclear Agents for the Prevention of Terrorism*, Dordrecht, 2014, pp. 83–98.
- 229 B. J. Shortt, M. R. Darrach, P. M. Holland and A. Chutjian, *J. Mass Spectrom.*, 2005, **40**, 36–42.
- 230 J. A. Diaz, C. F. Giese and W. R. Gentry, *Field Anal Chem Technol*, 2001, **5**, 156–167.
- 231 M. Werner and W. Fahrner, *IEEE Trans. Ind. Electron.*, 2001, **48**, 249–257.
- 232 Y. Wang, Y. Jia, Q. Chen and Y. Wang, *Sensors*, 2008, **8**, 7982–7995.
- 233 M. Rafiqul Islam, R. H. Galib, M. Sarkar and S. Chowdhury, *Wide-Bandgap Semiconductor Device Technologies for High-Temperature and Harsh Environment Applications*, John Wiley & Sons, Ltd, 2019, ch. 1, pp. 1–29.
- 234 Z. Takáts, J. M. Wiseman, B. Gologan and R. G. Cooks, *Science*, 2004, **306**, 471–473.
- 235 C. R. Ferreira, K. E. Yannell, A. K. Jarmusch, V. Pirro, Z. Ouyang and R. G. Cooks, *Clin. Chem.*, 2016, **62**, 99–110.
- 236 H. Wang, N. E. Manicke, Q. Yang, L. Zheng, R. Shi, R. G. Cooks and Z. Ouyang, *Anal. Chem.*, 2011, **83**, 1197–1201.
- 237 R. B. Cody, J. A. Laramée and H. D. Durst, *Anal. Chem.*, 2005, **77**, 2297–2302.
- 238 J. Balog, L. Sasi-Szabó, J. Kinross, M. R. Lewis, L. J. Muirhead, K. Veselkov, R. Mirnezami, B. Dezsó, L. Damjanovich, A. Darzi, J. K. Nicholson and Z. Takáts, *Sci. Translat. Med.*, 2013, **5**, 194ra93.
- 239 D. Burns, S. Mathias, B. J. McCullough, C. J. Hopley, D. Douce, N. Lumley, S. Bajic and P. Sears, *Int. J. Mass Spectrom.*, 2022, **471**, 116735.



- 240 H. M. Brown, T. J. McDaniel, C. P. West, E. H. Bondzie, M. R. Aldeman, B. T. Molnar, C. C. Mulligan and P. W. Fedick, *Int. J. Mass Spectrom.*, 2022, **474**, 116781.
- 241 L. Rittfeldt, *Anal. Chem.*, 2001, **73**, 2405–2411.
- 242 H. Frostling, K. Åse, A. Borgan, T. Østvold, A. Bjørseth and D. L. Powell, *Acta Chem. Scand.*, 1974, **28a**, 83–85.
- 243 R. Maynard and F. W. Beswick, *Clinical and Experimental Toxicology of Organophosphates and Carbamates*, Butterworth-Heinemann, 1992, pp. 373–385.
- 244 Z. Ouyang and R. G. Cooks, *Annu. Rev. Anal. Chem.*, 2009, **2**, 187–214.
- 245 S. Rozanova, J. Uszkoreit, K. Schork, B. Serschnitzki, M. Eisenacher, L. Tönges, K. Barkovits-Boeddinghaus and K. Marcus, *Biomolecules*, 2023, **13**, 1–20.
- 246 R. L. González Domínguez, Á. L. González-Domínguez, A. Sayago and Á. Fernández Recamales, *Metabolites*, 2020, **10**, 1–18.
- 247 A. Sens, S. Rischke, L. Hahnefeld, E. Dorochow, S. Schäfer, D. Thomas, M. Köhm, G. Geisslinger, F. Behrens and R. Gurke, *J. Mass Spectrom. Adv. Clin. Lab.*, 2023, **28**, 35–46.
- 248 J. G. McDonald, C. S. Ejsing, D. Kopczyński, M. Holčapek, J. Aoki, M. Arita, M. Arita, E. S. Baker, J. Bertrand-Michel, J. A. Bowden, B. Brügger, S. R. Ellis, M. Fedorova, W. J. Griffiths, X. Han, J. Hartler, N. Hoffmann, J. P. Koelmel, H. C. Köfeler, T. W. Mitchell, V. B. O'Donnell, D. Saigusa, D. Schwudke, A. Shevchenko, C. Z. Ulmer, M. R. Wenk, M. Witting, D. Wolrab, Y. Xia, R. Ahrends, G. Liebisch and K. Ekroos, *Nat. Metab.*, 2022, **4**, 1086–1088.

