Molecular Omics

REVIEW

Check for updates

Single-platform 'multi-omic' profiling: unified mass spectrometry and computational workflows for integrative proteomics–metabolomics analysis

Benjamin C. Blum, [ID] †[ab] Fatemeh Mousavi[c] and Andrew Emili [ID] *[abcd]

The objective of omics studies is to globally measure the different classes of cellular biomolecules present in a biological specimen (*e.g.* proteins, metabolites) as accurately as possible in order to investigate the corresponding 'states' of biological systems. High throughput omics technologies are emerging as an increasingly powerful toolkit in the rapidly advancing field of systems biology, enabling the systematic study of dynamic molecular processes that drive core cell functions like growth, sensing, and environmental adaptation. Advances in high resolution mass spectrometry, in particular, now allow for the near comprehensive study of cellular proteins and metabolites that underlie physiological homeostasis and disease pathogenesis. Yet while the expression levels, modification states, and functional associations of diverse molecular species are now measurable, existing proteomic and metabolomic data generation and analysis workflows are often specialized and incompatible. Hence, while there are now many reports of *ad hoc* combinations of unimolecular proteomic and metabolomic workflows, only a limited number of multi-omic profiling approaches have been reported for obtaining different molecular measurements (proteins, metabolites, nucleic acids) in parallel from a single biological sample. Moreover, elucidating how the myriad of measured cellular components are linked together functionally within the metabolic processes, signal transduction pathways, and macromolecular interaction networks central to living systems remains a massive, complicated, and uncertain endeavor. Presented here is a review of convergent mass spectrometry-based multi-omic methodologies, with a focus on notable recent advances and remaining challenges in terms of efficient sample preparation, biochemical separations, data acquisition, and integrative computational strategies. We outline a unifying network-based integrative framework to better derive biological knowledge from integrated profiling studies with the goal of realizing the full potential of multi-omic data sets.

Introduction

The omics revolution is helping to advance discovery science and basic mechanistic understanding of biological systems, starting with the major sequencing-based omics technologies, genomics and transcriptomics, and continuing with proteomics and metabolomics. These include steady advances in precision mass spectrometry that allow for highly sensitive, rapid, and accurate identification and quantification of thousands of cellular proteins, post-translational modifications (*e.g.* phosphorylation

sites, glycosylation) and metabolites in a single analysis. Yet while progress continues inexorably in each of these domains, the throughput, community adoption, and utility of proteomics and metabolomics techniques has lagged behind that of nucleic-acid-based molecular profiling methods. In part, this reflects the inability to amplify relevant molecules in a manner analogous to polymerase chain reaction amplification, but also to experimental inefficiencies and the lack of standardization of extant analytical workflows.

In principle, mass spectrometry-based proteomics and metabolomics studies are capable of generating near-comprehensive measurements of proteins and metabolites in parallel, providing valuable insights into the molecular mechanisms and dynamics of biological processes that could only otherwise be inferred in an indirect and error-prone manner from mRNA transcript levels.[1] Multi-omics have the potential to reveal the fundamental molecular logic coupling signal transduction, cellular metabolism, and phenotype determination.[2] Mass spectrometry is exceptionally

[a] *Center for Network Systems Biology, Boston University School of Medicine, Boston, Massachusetts, USA. E-mail: aemili@bu.edu; Tel: +1-617-358-4511*

[b] *Department of Biochemistry, Boston University School of Medicine, Boston, Massachusetts, USA*

[c] *Donnelly Centre, Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada*

[d] *Department of Biology, Boston University, Boston, Massachusetts, USA*

† These authors contributed equally.

Table 1 Summary of current major omics techniques, their advantages, limitations, and ability to support multi-omics analysis. Next generation sequencing is well suited for multi-omics analysis and large sample sizes but utility is limited to nucleic acids. HPLC-MS is suitable for a wider range of biomolecular types but consumes large sample quantities. GC-MS, CE-MS, and NMR are excellent for characterization of a wide range of metabolites, but throughput is limited and it is costly to run these three instruments in parallel. Finally, nLC-MS is well suited for multiple major omics-analysis, has moderate throughput, and is well suited for applications where sample is limited. (MS: mass spectrometry, HPLC: high pressure liquid chromatography, GC: gas chromatography, CE: capillary electrophoresis, NMR: nuclear magnetic resonance, nLC: nano-liquid chromatography)

Omics technique	Advantages	Limitations	Practicality for multi-omic profiling
Sequencing	• Massively parallel, high throughput. • Amplification of low abundance features possible.	• Low molecular diversity: nucleic acids only.	High: • Genomics • Transcriptomics • Epigenomics
HPLC-MS	• Wide molecular diversity. • Relatively robust.	• Relatively high sample material required. • Low sensitivity.	Moderate: • Proteomics • Metabolomics
GC-MS CE-MS	• Capable of characterizing volatile analytes. • Lower cost compared to LC-MS.	• Limited molecular diversity.	Low: • Metabolomics
NMR	• Non-destructive. • Simplified sample preparation.	• Lower sensitivity. • Low molecular diversity: metabolites only.	Low: • Metabolomics
nLC-MS	• High sensitivity. • Relatively high molecular diversity. • Low sample material requirements.	• Lower throughput than next generation sequencing.	High: • Proteomics • Metabolomics

well suited for the parallel analysis of protein and metabolite biomolecules, especially in light of recent engineering advances in high performance instrumentation and the accompanying biochemical sample separation techniques (*e.g.* chromatography) that work consistently well across diverse molecular types (Table 1).[3–5] Proteomics research laboratories have largely moved to apply tandem mass spectrometry together with nanoflow liquid chromatography (nLC-MS) as the primary analytical platform of choice for global protein, peptide, and post-translation modification analysis to achieve optimal detection sensitivity.[6] In contrast, metabolomics laboratories are largely wedded to traditional higher analytical flow rates for LC-MS, along with a constellation of alternate robust technologies such as nuclear magnetic resonance spectroscopy (NMR), gas chromatography mass spectrometry (GC-MS), and capillary electrophoresis mass spectrometry (CE-MS), in part due to the greater chemical diversity and biophysical properties of metabolites as compared to polypeptides.[7–10] Utilization of multipronged approaches allows metabolomics specialists to obtain more inclusive information about the metabolome over a wide range of biological samples. However, operating many instruments in parallel adds to expense and among all of these competing methods, LC-MS is the most versatile technique in terms of sensitivity, selectivity, speed and reproducibility, capable of providing routine and widespread metabolite coverage.[11–13]

The appeal of multi-omics studies is compelling: a deeper and better informed understanding of important biological processes, pathways, and functional modules can only be obtained through the combined investigation of the genome, transcriptome, proteome, and metabolome.[14] In this review, we summarize the application of combined LC-MS-based proteomics and metabolomics workflows, covering sample preparation, data acquisition and instrumentation, statistically-driven data processing, and integrative multi-omics computational data analysis using pathway and network-based modelling to further understanding of the network systems underlying biological processes and disease phenotypes (Fig. 1).

Sample preparation

Most existing LC-MS based strategies reported to date for the parallel analysis of the proteome and metabolome are unimolecular approaches.[6–10] In these studies, molecular extraction and analysis workflows are performed independently for both proteomics and metabolomics measurements on unique replicate samples using different LC-MS techniques. However, utilizing methods that perform simultaneous molecular extractions from a single sample minimizes experimental variation and reduces the amount of sample consumed and the number of handling steps performed during sample preparation.[15] This is especially valuable in cases where the biological samples being study are patient-derived, extremely limited, or otherwise hard to generate or obtain.

Proteomics and metabolomics sample extraction

All methods of sample preparation and extraction have significant downstream effects for omics studies.[16] Preparation methods and omics techniques have inherent biases and each additional step introduces possible sources of variation. A sample preparation method for extracting proteins and metabolites from the same biological sample must be compatible with both classes of biomolecules and produce sufficient yield to be practical.[17] During sampling and sample preparation, the state of the biological sample must be fixed or
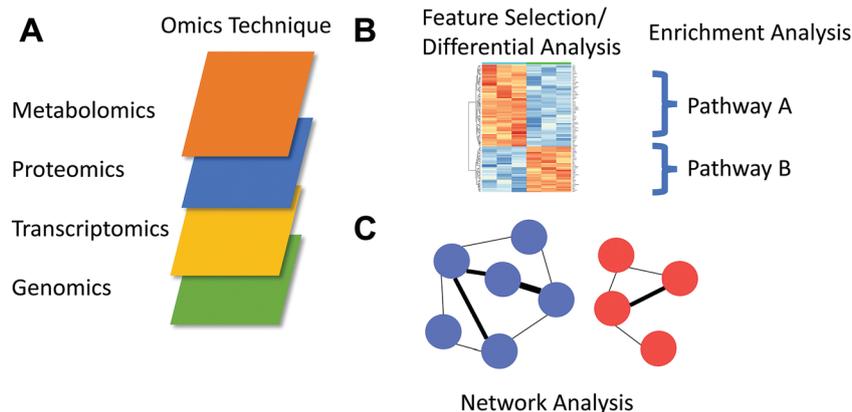
Fig. 1 Multi-omics analysis. (A) Different omics techniques allow for the simultaneous measurement of many features of a certain molecular type within biological samples. (B) Established computational tools perform differential analysis or feature selection, which may provide useful information about individual features, and is often combined with enrichment analysis to find over-representation of pathways or processes. (C) Network analysis takes advantage of known molecular interactions to model dynamic biological systems.

preserved as much as possible in order to maintain the information contained within a sample.[18] This is a challenge due to rapid induction of metabolic stress, degradation, and dynamics of protein modification.

Quenching cell metabolism is critical to capture metabolic pathways in a state reflective of biology rather than the sample preparation method. Low temperature steps (e.g. cold solvent addition, and freezing in liquid nitrogen), addition of acid, or rapid heating are frequently used to accomplish this goal. Reports show some loss or degradation of metabolites with freezing but this phenomenon is more pronounced in protocols with the addition of acid or fast heating.[19–23] High temperature methods also degrade proteins and are not compatible with proteome analysis.[24]

Experimental procedures have been reported that discretely target proteins and metabolites through sequential or parallel experiments.[14] These reports follow standard protocols introduced for the individual extraction of the proteins or metabolites.[16,25] However, for metabolomics sample preparation, there is the additional challenge created by the greater diversity of physical and chemical properties of molecules, which are not easily isolated with a single method.[18,26] Protocols used for metabolite extraction from complex biological matrixes, such as biological fluids, tissue, cell pellets, and cell media, separate metabolites and perform pre-concentration prior to analysis. In order to reduce or eliminate matrix effect, metabolites are separated from salts and macromolecules, such as nucleic acids and proteins.[26–29] Liquid–liquid extraction (LLE), solid phase extraction (SPE), protein precipitation, dialysis, and ultrafiltration are common methods used for metabolomics sample preparation.[30–33]

LLE is the most widespread extraction procedure in integrated one-pot methods of proteomics and metabolomics analysis. Cold extraction (−20 °C), using solvent, combines the quenching and extraction steps, in addition to facilitating cell lysis. In the case of tissue extraction, the frozen tissue preserved in liquid nitrogen, must be homogenized to a fine powder by grinding using vibration mills, vortexing, mixing

with beads made of an inert material, or homogenization of the tissue in a solvent using blenders.[34–37] In order to extract hydrophilic metabolites, isopropanol, ethanol, methanol, acetonitrile, water, or a mixture of these solvents can be used. Non-polar metabolites are mostly extracted by ethyl acetate or chloroform.[38] When a mixture of polar and non-polar extraction solvents is used for separation, biphasic fractionation of metabolites into aqueous and lipophilic layers is possible, and the extracted metabolites can be analyzed discretely.[37,39,40] For integrative proteomics and metabolomics from a single sample, a good starting point is protein precipitation and metabolite depletion from the whole sample using organic solvent.[16] After addition of the extraction solvent, metabolites may be isolated in the supernatant, while the remaining cell pellet is used for protein extraction.[41] Sample preparation for bottom up proteomics must include chemical or enzymatic protein digestion to peptides prior to LC-MS analysis.[42]

Coman et al. presented a simultaneous metabolite, protein, and lipid extraction protocol, referred to as SIMPLEX, as a combinatorial multi-molecular omics approach for the quantitative investigation of metabolites, lipids, and proteins from a single mesenchymal cell line. This study subsequently involved LC-MS analysis and resulted in novel biomarkers associated with Huntington's disease.[15] In this study, separation of each molecular fraction started with metabolite extraction by cycles of methanol suspension and freeze–thaw. Then, sonication steps were used for separation of lipids and protein molecules. Established mass spectrometry techniques resulted in identification of 75 metabolites, 360 lipids and 3327 proteins from 10^6 mesenchymal stem cells. Importantly, results obtained from the optimized protocol were comparable with previously reported unimolecular methods for metabolomics, proteomics, and phospho-proteomics analysis with regards to method sensitivity, reproducibility, and performance.

There are few methods reported for simultaneous extraction of DNA, RNA, proteins and metabolites from a single biological sample.[16] Weckwerth et al. optimized a novel extraction

protocol for sequential extraction of metabolites, proteins, and RNA from the same plant tissue for subsequent GC-ToFMS analysis for metabolites, and LC × LC coupled to MS for protein analysis. Plant tissue was homogenized under liquid nitrogen by grinding. Metabolites were extracted with a cold solvent mixture of methanol, water, and chloroform, which resulted in identification of 652 metabolites after precipitating DNA/RNA, and 297 proteins.[35]

Vorreiter *et al.* optimized a method for simultaneous extraction of nucleic acids, proteins, and metabolites from Jurkat E6-1 cells and Hepa 1c1c7 cells and compared the results with those obtained with unimolecular extraction methods.[43] The study examined DNA and RNA quality and quantity, proteome coverage, and identification of hydrophilic metabolites involved in central carbon and nitrogen metabolism. The sample procedure that provided the best coverage for all four omics studies involved extraction of mainly polar metabolites using a solvent mixture of water, methanol, and chloroform. The resulting pellet was treated by water saturated phenol for RNA, DNA, and protein extraction. The RNA and DNA phases were separated using chloroform and protein precipitation was performed. Proteins were re-suspended in urea and enzymatic digestion was used for subsequent peptide analysis. The best results showed 35 metabolites and 870 proteins from 10^7 Jurkat E6-1 cells, and 23 metabolites and 1023 proteins detected for Hepa 1c1c7 cells. One of the important advantages of this method was the RNA and DNA separation from proteins that reduced complexity of protein mass spectra and permitted integration with transcriptomics and genomics studies.

Sample enrichment and fractionation

Selective protein enrichment from complex proteomics fractions is possible and particularly useful depending on the questions being addressed in a study. For example, affinity capture of phosphoproteins is commonly performed using metal chelate (*e.g.* TiO_2) beads to selectively explore the differential activation state of protein kinases, enzymes and other regulatory proteins in signaling pathways.[44,45] Biochemical fractionation methods have also been devised for the enrichment of protein glycosylation, ubiquitination, and other post-translational modifications, or the isolation of protein subcellular compartments to define organelle functional associations.[46,47] All of these layers are important determinants of cell signaling, disease pathogenesis, and other biological processes. In some cases, advances in high resolution mass-spectrometry are reducing the need for further extensive enrichment to achieve deep proteome coverage.[48]

One of the most important considerations for an integrated workflow based on nLC-MS is susceptibility of microbore columns to blockage, particularly during metabolomics analysis, which can degrade column performance and lifetime due to presence of particulates and insoluble salts.[41] Mobile phase optimization to account for hydrophobic molecules may be beneficial and high concentration sample extracts show increased reproducibility and wider metabolome coverage when extensive sample cleanup and fractionation is performed prior to

nLC-MS analysis. For example, solid-phase microextraction (SPME) can help remove particularly challenging components, such as phospholipids. Lyophilization conjugated with SPE has been applied for yeast metabolomics analysis.[49] Commercial lipid depletion kits for the selective removal of phospholipids can also increase chromatographic column lifetime while reducing ion suppression prior to nLC.[41,50] Column trapping is another method applied for metabolomics analysis in order to load a larger volume of sample while removing salts and other contaminants.[51]

Sample pre-fractionation and labeling

Proteomics studies frequently involve sample pre-fractionation steps to increase proteome coverage by nLC-MS.[45] Protein separation and/or enrichment using offline fractionation techniques, such as gel electrophoresis, HPLC, and affinity capture are common. Where analytical chromatography is used, an orthogonal method is usually performed to complement the standard (reverse phase) separations used in LC-MS. While high resolution can be achieved under denaturing conditions using hydrophilic interaction chromatography (HILIC) and high pH reverse phase chromatography, non-denaturing ion-exchange and mixed mode chromatography can also be effective depending on study goals, such as the isolation and characterization of intact native macromolecular complexes.[42] To improve instrument time efficiency to offset the added number of samples originating from pre-fractionation, stable isotope chemical labeling may be used for fraction multiplexing prior to LC-MS. Metabolic labeling can also be performed to improve relative proteome quantification and multiplexing analysis. It is also possible to perform flux analysis using metabolic labels, for example in a pulse-chase experimental design.[51,52]

Data acquisition and feature identification

Mass spectrometry data acquisition

Given nLC-MS is widely and successfully applied in proteomics research labs, there is significant interest in adapting the same technology for metabolomics analyses to enable combined proteomic and metabolomic studies using the same platform.[17] Recently, nLC-MS workflows have been reported for use in more targeted metabolite and small molecule analyses, including the detection and quantification of amino acids, fatty acids, lipids, prostaglandins, di/tripeptides, steroids, vitamins, nucleic acids, xenobiotics and, less frequently, in global metabolomics surveys of urine, plasma and cell samples.[53,54] Nanoflow techniques do pose challenges beyond column blockage; for example, electrospray ionization in negative mode is complicated by reduced solvent desolvation and electrical discharge.[11]

While HPLC-MS is widely used for high-throughput untargeted metabolomics analysis, matrix effects and poor ionization efficiency due to ion suppression may produce lower sensitivity and restrict detection to highly abundant metabolites.[55,56]

This limitation has motivated the use of nLC-MS for metabolomics, too. Nanoflow HPLC columns with internal diameters between 10 to 75 μm and flow rates of 10 to 500 nL min^{-1} offer lower chromatographic dilution, enhanced peak capacity, resolution and detection sensitivity.[57,58] Nanoflow also provides improved analyte pre-concentration and more efficient introduction of biomolecules *via* electrospray ionization (ESI) to the MS system as compared to conventional high flow LC-MS systems.[59–62] Compared to conventional HPLC-MS, nLC-MS systems are up to 2000 times more sensitive, while achieving up to a 300-fold lower limit of detection and limit of quantification.[63,64] In cases where specimens are limited, nLC-MS is also beneficial because it requires less sample, though reproducibility and speed may be compromised as compared to HPLC. It is possible to use both HILIC and RP chromatography for nLC-MS based metabolomics profiling; however, running samples in negative ionization mode is more difficult for metabolomics studies using nanoflow ESI.[11,65]

There are several methods used to produce consistent nanoflow rates, including split flow, direct infusion, and direct flow. Split flow was the first introduction of nLC, which utilized dual HPLC pumps with a diverter that split the higher flow fluidics to deliver nano-scale flow rates to the mass spectrometer.[56] This results in about 99% of the sample and most of solvent being lost to waste.[66]

Nanoflow direct infusion mass spectrometry has long been successfully applied for metabolomics sample analysis from many sample types, such as liver extracts, plasma, urine, embryos, and *Daphnia*.[57,67–74] These studies take advantage of the high-throughput offered by direct infusion. However, the absence of chromatography before MS enhances the matrix effect and makes differentiation between isobars or isomers impossible using this method.[75–77]

Nanoflow direct flow mass spectrometry analysis methods may use nanoflow reciprocal or syringe pumps and microfluidic flow.[78,79] These methods improve compound retention and spray characteristics of the nanoflow platform. Application has been reported in metabolomics analysis of complex samples such as plasma, tissue, urine, breath, sweat, cerebrospinal fluids and cell extracts.[49,50,52,80–86] Deviation in back pressure and surface tension due to variability in mobile phase viscosities can create challenges in obtaining stable analyte retention times, especially with gradient separations.[87,88] Direct flow is not as high-throughput as direct infusion, however, direct flow does provide increased resolution *via* pre-mass spectrometry separations without waste inherent with sample split-flow. Nano-columns, up to 0.01 mm internal diameter, can provide stable low flow gradients, reducing chromatographic dilution while requiring low sample and mobile phase consumption.[89–91]

There are many different types of commercial nano-columns, including BEH C18, HSS T3, Acclaim PepMax C18, and ZIC-HILIC. Different columns are often better suited to specific molecular classes and column optimization is an important step. In order to provide consistent separations of polar metabolites, HILIC chromatography is often most effective.[85]

Different case studies evaluating distinct nano-columns for metabolomics surveys have shown that the performance of the analytical system increases by decreasing pore size even more than is evident in proteomics applications. In comparison to conventional HPLC with the same stationary phase, nLC shows higher sensitivity and enhanced chromatographic resolution for a wider range of small molecule metabolites, while co-elution and ion suppression drop significantly.[92,93]

Two other parameters influencing the quality of the recorded metabolomics data are retention time drift and peak intensity. Methods and techniques should be optimized to improve retention time stability and minimize the coefficient of variation (CV) between replicate analyses. Variation in peak intensity by replicate nLC-MS analyses may be as high as 30%,[94] and CVs consistently less than 20% for targeted and untargeted metabolomics analyses are difficult but not impossible to achieve by nLC-MS.[95] For example, a study using nLC-MS for yeast metabolomics reported an average CV of less than 14%.[69] The use of internal standards may decrease CVs below 0.2%.[41,50]

Electrospray ionization (ESI) is the most common soft ionization technique employed for proteomics.[96] Nano ESI (nESI) produces droplets up to 1000 times smaller than generated by high flow ESI, which leads to improved evaporation, more efficient sample transfer and potentially more ions detected.[75] The best performing nESI sources use a smaller internal diameter for higher ionization efficiency. Utilization of non-tapered emitters with an internal diameter in the range of 10–30 μm provides an adequate compromise between sensitivity, system robustness, and sample consumption.[97–101] Wider internal diameters reduce potential clogging and facilitate more stable ionization spray, resulting in more reproducible mass spectra and ion chromatograms.

Both proteomics and metabolomics benefit greatly from continued progress in LC-MS instrument engineering. High resolution mass-spectrometers, with higher mass accuracy and faster scan speeds provide more accurate, in-depth raw data. Improving ion resolution and throughput facilitates better feature identification and makes the analysis larger numbers of sample feasible.[102,103] While there remains a fundamental trade-off between throughput and analysis depth and coverage, high-resolutions systems enable acquisition of high quality data from complex mixtures with even relatively short gradient separations.

The sensitivity and throughput of mass-spectrometry-based omics studies does lag considerably behind genomic and transcriptomic sequencing technologies. This challenge makes studies with large sample sizes difficult. While some large-scale efforts, such as the Clinical Proteomic Tumor Analysis Consortium (CPTAC), are addressing this shortcoming, proteomics and metabolomics is still catching up to the massively parallel nucleic acid profiling approaches in terms of ease of data generation, standardization, and accessibility.[104–106] Further improvements in multiplexing, more rapid and standardized separations, and improved mass-spectrometry scan speed and automation will be important criteria for continuing to advance the field.

Proteomics and metabolomics raw data processing and feature identification

Identifying features (cognate proteins/peptides and metabolites) and abundance (relative intensities) from raw mass spectrometry data files are fundamental steps in the analysis pipeline. The process of converting MS precursor ion scans and associated MS/MS fragmentation spectra into identified proteins and metabolites is different for each data type and is addressed using different software tools. For proteomics, the peptide/protein search space is inherently limited by biology (genome of the particular species) and chemistry (peptide bond constraints). Peptide database files for many species are available for download from sources such as Uniprot, and have well defined rules for polymeric structure and fragmentation patterns.[107] MaxQuant is a popular and relatively easy to use full-featured tool, which runs the Andromeda search engine as well as associated quantification tools. Other powerful software tools for searching MS/MS spectra data to identify peptides include the popular search engines MS-GF+, Comet, SpectraST, Mascot, XTandem, MyriMatch, Sequest, and Tide.[108–114] Validation of peptide identification is often estimated through empirical false discovery rate (FDR) using a reverse 'decoy' database search.[115,116] Many additional tools are used to interface with these search engines, to assign intensity values, and to format data for subsequent statistical analysis.[117–121] Several community tools, such as the Trans-Proteomic Pipeline, Peptide Shaker, and OpenMS, offer added functionality that improve overall performance, including permitting running multiple search engines to compare and compile results.[114] Open-source software, such as OpenMS, also provide powerful configurable features, while retaining a simplified user interface even when running complicated analysis pipelines.[118]

Metabolomics feature identification is a more difficult challenge in that the search space for identifications is much greater due to the increased structural diversity. This is further confounded by the detection of multiple alternately adducted ion species by LC-MS, and the lack of constrained MS/MS fragmentation rules, compared to polypeptides, which tend to cleave in a more predictable manner. To address this gap, there are efforts to compile compound spectral libraries containing empirically derived or simulated metabolite spectra, along with matching information regarding chemical composition, structure, and other annotations. Most notable of these databases are PubChem, HMDB, and METLIN.[122,123] METLIN has an extensive library of experimentally defined MS2 spectra, while HMDB features extensive compound annotation information and, as of the most recent release, has made all curated data, including spectral libraries, downloadable in open-source formats. Unfortunately, the community still lacks a common identification nomenclature for conversion between resources, limiting their utility.

In principle, metabolite identifications may also be performed through comparing MS/MS fragmentation patterns, isotope patterns and retention times for reference standard compounds, though this can be expensive and time-consuming. Surprisingly, many putative human metabolites are not readily verifiable using synthetic standards, suggesting a potentially major gap in LC-MS assay design and detection reliability. Metabolomics data can be noisy, leading to spurious compound assignments, which heightens the need for confident metabolite identifications from features of interest.[124,125] In order to achieve reliable metabolite feature matching, planning the overall experimental design, data analysis pipeline, and validation strategy in a stringent manner early on in the project lifecycle is crucial.

Many open-source and commercial software tools are available for performing metabolomics spectral filtering, peak detection, retention time alignment, and normalization. These include MZmine, MS-DIAL, and R packages and associated with the mzR parser, and other tools, like MAIT, that provide additional functionality for routine detection of differential metabolites from LC/MS-based metabolomics data.[126–129] XCMS and MetaboAnalyst are also popular, user friendly tools associated with the METLIN and HMDB databases, respectively. Although available as stand-alone R packages, powerful features like raw data file processing and pathway enrichment analysis are only or most easily accessed via online web-based versions of these tools, providing a hurdle for integration into more advanced computational workflows.[130–132] OpenMS likewise offers plugins for metabolite feature identification that use advanced signal processing techniques to identify features in raw data and permit the querying of public databases to find matches based on exact mass and structural fragmentation patterns.[133]

Despite these powerful resources, metabolomics feature identification and quantification remain a challenge. Compared to the thriving ecosystem of robust open-source sequencing analysis tools, metabolomics and proteomics search tools tend to be more closed and different algorithms often do not converge to similar results with the same degree of confidence. Hopefully, more robust, user-friendly, and open-source solutions will continue to advance this area of research.

Integrative multi-omics analysis

Normalization and feature selection

Proteomics and metabolomics generate a large amount of complementary data that is potentially informative of the activity of biological pathways but requires specialized mathematical, statistical, and bioinformatics analysis strategies.[134] In order to best leverage these overlapping signals to achieve a more comprehensive understanding of biological systems, integration of the omics data is an important and often underappreciated step. While there are now numerous omics data analysis and integration tools for this purpose, each has various strengths and limitations.

Processed raw LC-MS data usually is structured as a matrix of features (i.e. proteins, peptides, or metabolites) and biological samples, with their corresponding intensity values. Tools and processes developed for interpreting other complex data types, such as transcriptomics, have been adapted

successfully for transformation and normalization of raw intensity values.[135] Feature selection is an important early step and most often is performed by analyzing differences between two or more groups (*e.g.* case *versus* control samples). R packages, such as limma, have been well established for the use of differential analysis for datasets with known phenotypes based on linear models and other statistical criteria.[136] Innovative new tools are also rapidly advancing that are based on the application of more advanced machine learning algorithms, which are effective in cases with sufficient sample size, with the caret R package and scikit-learn Python modules being among the most commonly used.[137,138]

Enrichment analysis and pathway-centered visualization

More automated tools are also emerging that can integrate feature selection with subsequent functional enrichment analysis. Gene Set Enrichment Analysis (GSEA), the Database for Annotation, Visualization, and Integrated Discovery (DAVID), and Enrichr are three widely used computational tools for performing routine pathway enrichment analysis on omics datasets. In the case of proteomics data, protein accessions must be mapped to corresponding gene symbols.[139,140] GSEA requires java and is downloadable as both a desktop application and for command line usage, DAVID is accessible as a web-based tool, while Enrichr is available both as an online tool and through an R interface package and python module to interface with public pathway databases. Tools like the EnrichmentMap app for Cytoscape can help simplify and visualize overall trends among occasionally overwhelming enrichment results.[141]

Other computational tools have been developed to integrate metabolomics data into the enrichment analysis. MetaboAnalyst and XCMS Online both incorporate basic enrichment analysis and omics integration tools along with their metabolomics search software. The online versions of the software are not open-source and pose challenges to incorporate into more extensive data mining pipelines. MixOmics is an R package that includes functions expressly optimized for the statistical analysis and visualization of multi-omics data sets.[142] 3Omics is a web-based platform that performs integration and joint visualization of parallel transcriptomics, proteomics, and metabolomics datasets. Preprocessed and normalized data can be submitted using a standard comma-delimited.csv file format. Kuo *et al.* showed the utility of 3Omics for the integrative analysis of multi-omic human data.[143] InCroMAP is another tool for the integration of disparate omics-data sets[144] that was successfully used for characterization of biological pathways perturbed in a mouse liver tumor model setting.[145]

Interaction-based network meta-analysis

While grouping features together based on available functional annotations is a valuable and important analysis step, it does not consider the relationships, or interactions, between different molecules that are often necessary for executing biological processes. Many public curation databases aggregate information reported in the literature on molecular interactions, particularly protein–protein and protein–metabolite interactions.[123,146,147]

iRefindex, for example, provides a comprehensive database of protein–protein interactions aggregated from a number of primary databases.[148] Likewise, HMDB, STITCH, and Recon all strive to provide protein–metabolite interaction information in addition to charting metabolic pathways.

Omics integrator is an open source python tool that enables the selection of 'active' network modules that show differential features in overlays of omics data onto curated interactions from public databases.[149] The modules are selected from the larger interaction network using a machine learning algorithm. In this way, the tool accounts for changes in sets of interacting molecules with a reasonable level of confidence even if not all of the partner components are detected in the omics data sets. Pirhaji *et al.* described a more recent extension called PIUmet, which uses protein–protein and protein–metabolite interaction networks gleaned from public database to inform the process of feature identification from metabolomics data, in addition to the identification of relevant molecular interaction subnetworks.[150] Unfortunately, PIUmet is proprietary and access is currently only available online.

Currently, most enrichment and interaction-based analyses heavily rely on annotation databases, which comprise another area where advances may be made. The most complete and best-annotated databases are determined based on genes, which are useful in proteomics only after mapping identified protein products to their corresponding cognate genes. As a result, the underlying models tend to lose information regarding specific protein isoforms, post-translational modification sites, and molecular functions and associations. Continuing to build out and improve the quality and breadth of integrative omics analysis tools that exploit user defined protein–protein interactions, protein modification sites, and protein–metabolite interaction information will be an important endeavor to enhance the informativeness of integrative studies. Where some analysis tools are currently capable of incorporating known (curated) interactions into the multi-omic pipeline, these models still represent an over-simplification of complex biological network topologies. Details known to be important for driving biological processes, such as macromolecular structure, interaction binding kinetics, equilibrium states, and spatiotemporal control, are not incorporated into most analyses. Borrowing from the fields of Bioengineering and Physics, there is hope for more advanced models of dynamical systems and their application to omics data analysis to drive better understanding of the molecular mechanisms driving cell biology.[151,152] Developing innovative computational tools to make the application of such integrative models more accessible to more biological researchers and robust enough for a variety of use cases will remain a challenge. However, the potential exists to be able to more accurately describe changes in biological systems and to better predict cellular responses to genetic and interventional perturbations, for example in the case of combinatorial therapeutics.

Conclusion

Advances in sample preparation techniques, instrumentation, analysis software, and public annotation databases have

succeeded in permitting researchers to gain important insights into biological processes from combined proteomics and metabolomics studies. The closer integration of proteomics and metabolomics workflows, from sample preparation to data acquisition and network-based computational analysis, offers the potential of elucidating a more comprehensive understanding of complex biological systems and their dynamical responses to diverse stimuli. Using a unified nLC-MS-based platform facilitates joint multi-omic analyses of complex biological specimens with a streamlined instrumentation requirement, as well as more consistent sample preparation, quantitative and qualitative identification of biomolecules, and integration of the resulting data using pathway and network-based methods. Parallel proteomic and metabolomic-based multi-omic studies hold exciting promise for the near future of network systems biology.

Conflicts of interest

There are no conflicts of interest to declare.

Acknowledgements

References

1 F. Edfors, F. Danielsson and B. M. Hallström, *et al.*, Gene-specific correlation of RNA and protein levels in human cells and tissues, *Mol. Syst. Biol.*, 2016, **12**(10), 883, DOI: 10.15252/msb.20167144.

2 M. Mayr, B. Madhu and Q. Xu, Proteomics and metabolomics combined in cardiovascular research, *Trends Cardiovasc. Med.*, 2007, **17**(2), 43–48, DOI: 10.1016/j.tcm.2006.11.004.

3 J. R. Yates, C. I. Ruse and A. Nakorchevsky, Proteomics by mass spectrometry: approaches, advances, and applications, *Annu. Rev. Biomed. Eng.*, 2009, **11**(1), 49–79, DOI: 10.1146/annurev-bioeng-061008-124934.

4 A. Bensimon, A. J. R. Heck and R. Aebersold, Mass spectrometry-based proteomics and network biology, *Annu. Rev. Biochem.*, 2012, **81**(1), 379–405, DOI: 10.1146/annurev-biochem-072909-100424.

5 Y. V. Karpievitch, A. D. Polpitiya, G. A. Anderson, R. D. Smith and A. R. Dabney, Liquid Chromatography Mass Spectrometry-Based Proteomics: Biological and Technological Aspects, *Ann. Appl. Stat.*, 2010, **4**(4), 1797–1823, DOI: 10.1214/10-AOAS341.

6 Y. Zhao and O. N. Jensen, Modification-specific proteomics: strategies for characterization of post-translational modifications using enrichment techniques, *Proteomics*, 2009, **9**(20), 4632–4641, DOI: 10.1002/pmic.200900398.

7 J. E. Peironcely, T. Reijmers, L. Coulier, A. Bender and T. Hankemeier, Understanding and classifying metabolite space and metabolite-likeness, *PLoS One*, 2011, **6**(12), e28966, DOI: 10.1371/journal.pone.0028966.

8 M. Schrader and P. Schulz-Knappe, Peptidomics technologies for human body fluids, *Trends Biotechnol.*, 2001, **19**(10 suppl), S55–S60, DOI: 10.1016/S0167-7799(01)01800-5.

9 P. Schulz-Knappe, M. Schrader and H.-D. Zucht, The peptidomics concept, *Comb. Chem. High Throughput Screening*, 2005, **8**(8), 697–704.

10 M. M. Jost, P. Budde and H. Tammen, *et al.*, The concept of functional peptidomics for the discovery of bioactive peptides in cell culture models, *Comb. Chem. High Throughput Screening*, 2005, **8**(8), 767–773.

11 B. Zhou, J. F. Xiao, L. Tuli and H. W. Ressom, LC-MS-based metabolomics, *Mol. BioSyst.*, 2012, **8**(2), 470–481, DOI: 10.1039/c1mb05350g.

12 G. A. Theodoridis, H. G. Gika, E. J. Want and I. D. Wilson, Liquid chromatography-mass spectrometry based global metabolite profiling: a review, *Anal. Chim. Acta*, 2012, **711**, 7–16, DOI: 10.1016/j.aca.2011.09.042.

13 D. Vuckovic, Current trends and challenges in sample preparation for global metabolomics using liquid chromatography-mass spectrometry, *Anal. Bioanal. Chem.*, 2012, **403**(6), 1523–1548, DOI: 10.1007/s00216-012-6039-y.

14 B.-S. Chen and C.-C. Wu, Systems biology as an integrated platform for bioinformatics, systems synthetic biology, and systems metabolic engineering, *Cells*, 2013, **2**(4), 635–688, DOI: 10.3390/cells2040635.

15 C. Coman, F. A. Solari, A. Hentschel, A. Sickmann, R. P. Zahedi and R. Ahrends, Simultaneous Metabolite, Protein, Lipid Extraction (SIMPLEX): A Combinatorial Multimolecular Omics Approach for Systems Biology, *Mol. Cell. Proteomics*, 2016, **15**(4), 1453–1466, DOI: 10.1074/mcp.M115.053702.

16 S. C. Sapcariu, T. Kanashova, D. Weindl, J. Ghelfi, G. Dittmar and K. Hiller, Simultaneous extraction of proteins and metabolites from cells in culture, *MethodsX*, 2014, **1**, 74–80, DOI: 10.1016/j.mex.2014.07.002.

17 R. Fischer, P. Bowness and B. M. Kessler, Two birds with one stone: doing metabolomics with your proteomics kit, *Proteomics*, 2013, **13**(23–24), 3371–3386, DOI: 10.1002/pmic.201300192.

18 K. Dettmer, P. A. Aronov and B. D. Hammock, Mass spectrometry-based metabolomics, *Mass Spectrom. Rev.*, 2007, **26**(1), 51–78, DOI: 10.1002/mas.20108.

19 W. B. Dunn and D. I. Ellis, Metabolomics: current analytical platforms and methodologies, *TrAC, Trends Anal. Chem.*, 2005, **24**(4), 285–294, DOI: 10.1016/j.trac.2004.11.021.

20 H. Cheng, G. Sun, K. Yang, R. W. Gross and X. Han, Selective desorption/ionization of sulfatides by MALDI-MS

facilitated using 9-aminoacridine as matrix, *J. Lipid Res.*, 2010, **51**(6), 1599–1609, DOI: 10.1194/jlr.D004077.

21 J. M. Büscher, D. Czernik, J. C. Ewald, U. Sauer and N. Zamboni, Cross-platform comparison of methods for quantitative metabolomics of primary metabolism, *Anal. Chem.*, 2009, **81**(6), 2135–2143, DOI: 10.1021/ac8022857.

22 H. John, F. Breyer, J. O. Thumfart, H. Höchstetter and H. Thiermann, Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) for detection and identification of albumin phosphylation by organophosphorus pesticides and G- and V-type nerve agents, *Anal. Bioanal. Chem.*, 2010, **398**(6), 2677–2691, DOI: 10.1007/s00216-010-4076-y.

23 C. J. Bolten, P. Kiefer, F. Letisse, J.-C. Portais and C. Wittmann, Sampling for metabolome analysis of microorganisms, *Anal. Chem.*, 2007, **79**(10), 3843–3849, DOI: 10.1021/ac0623888.

24 M. Rosa, C. J. Roberts and M. A. Rodrigues, Connecting high-temperature and low-temperature protein stability and aggregation, *PLoS One*, 2017, **12**(5), e0176748, DOI: 10.1371/journal.pone.0176748.

25 R. L. Gundry, M. Y. White and C. I. Murray, *et al.*, Preparation of proteins and peptides for mass spectrometry analysis in a bottom-up proteomics workflow, *Curr. Protoc. Mol. Biol.*, 2009, ch. 10(1), Unit 10.25–10.25.23, DOI: 10.1002/0471142727.mb1025s88.

26 D. G. Sitnikov, C. S. Monnin and D. Vuckovic, Systematic Assessment of Seven Solvent and Solid-Phase Extraction Methods for Metabolomics Analysis of Human Plasma by LC-MS, *Sci. Rep.*, 2016, **6**(1), 38885, DOI: 10.1038/srep38885.

27 E. Rogatsky and D. Stein, Evaluation of matrix effect and chromatography efficiency: new parameters for validation of method development, *J. Am. Soc. Mass Spectrom*, 2005, **16**(11), 1757–1759, DOI: 10.1016/j.jasms.2005.07.012.

28 B. K. Matuszewski, M. L. Constanzer and C. M. Chavez-Eng, Matrix effect in quantitative LC/MS/MS analyses of biological fluids: a method for determination of finasteride in human plasma at picogram per milliliter concentrations, *Anal. Chem.*, 1998, **70**(5), 882–889.

29 M. S. Chang, Q. Ji, J. Zhang and T. A. El-Shourbagy, Historical review of sample preparation for chromatographic bioanalysis: pros and cons, *Drug Dev. Res.*, 2007, **68**(3), 107–133, DOI: 10.1002/ddr.20173.

30 H. Stahnke, S. Kittlaus, G. Kempe and L. Alder, Reduction of matrix effects in liquid chromatography-electrospray ionization-mass spectrometry by dilution of the sample extracts: how much dilution is needed?, *Anal. Chem.*, 2012, **84**(3), 1474–1482, DOI: 10.1021/ac202661j.

31 G. Theodoridis, H. G. Gika and I. D. Wilson, LC-MS-based methodology for global metabolite profiling in metabonomics/metabolomics, *TrAC, Trends Anal. Chem.*, 2008, **27**(3), 251–260, DOI: 10.1016/j.trac.2008.01.008.

32 L. Wang, Z. Pi, S. Liu, Z. Liu and F. Song, Targeted metabolome profiling by dual-probe microdialysis sampling and treatment using Gardenia jasminoides for rats with type 2 diabetes, *Sci. Rep.*, 2017, **7**(1), 10105, DOI: 10.1038/s41598-017-10172-w.

33 S. Tiziani, A.-H. Emwas and A. Lodi, *et al.*, Optimized metabolite extraction from blood serum for 1H nuclear magnetic resonance spectroscopy, *Anal. Biochem.*, 2008, **377**(1), 16–23, DOI: 10.1016/j.ab.2008.01.037.

34 P. Jonsson, J. Gullberg and A. Nordström, *et al.*, A strategy for identifying differences in large series of metabolomic samples analyzed by GC/MS, *Anal. Chem.*, 2004, **76**(6), 1738–1745, DOI: 10.1021/ac0352427.

35 W. Weckwerth, K. Wenzel and O. Fiehn, Process for the integrated extraction identification, and quantification of metabolites, proteins and RNA to reveal their co-regulation in biochemical networks, *Proteomics*, 2004, **4**(1), 78–83, DOI: 10.1002/pmic.200200500.

36 U. Roessner, C. Wagner, J. Kopka, R. N. Trethewey and L. Willmitzer, Technical advance: simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry, *Plant J.*, 2000, **23**(1), 131–142.

37 G. Colebatch, G. Desbrosses and T. Ott, *et al.*, Global changes in transcription orchestrate metabolic differentiation during symbiotic nitrogen fixation in Lotus japonicus, *Plant J.*, 2004, **39**(4), 487–512, DOI: 10.1111/j.1365-313X.2004.02150.x.

38 A. Aharoni, C. H. Ric de Vos and H. A. Verhoeven, *et al.*, Nontargeted Metabolome Analysis by use of Fourier Transform Ion Cyclotron Mass Spectrometry, *OMICS*, 2002, **6**(3), 217–234, DOI: 10.1089/15362310260256882.

39 O. Fiehn, J. Kopka, R. N. Trethewey and L. Willmitzer, Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry, *Anal. Chem.*, 2000, **72**(15), 3573–3580.

40 J. Gullberg, P. Jonsson, A. Nordström, M. Sjöström and T. Moritz, Design of experiments: an efficient strategy to identify factors influencing extraction and derivatization of Arabidopsis thaliana samples in metabolomic studies with gas chromatography/mass spectrometry, *Anal. Biochem.*, 2004, **331**(2), 283–295, DOI: 10.1016/j.ab.2004.04.037.

41 A. David, A. Abdul-Sada, A. Lange, C. R. Tyler and E. M. Hill, A new approach for plasma (xeno)metabolomics based on solid-phase extraction and nanoflow liquid chromatography-nanoelectrospray ionisation mass spectrometry, *J. Chromatogr. A*, 2014, **1365**, 72–85, DOI: 10.1016/j.chroma.2014.09.001.

42 Y. Zhang, B. R. Fonslow, B. Shan, M.-C. Baek and J. R. Yates, Protein analysis by shotgun/bottom-up proteomics, *Chem. Rev.*, 2013, **113**(4), 2343–2394, DOI: 10.1021/cr3003533.

43 F. Vorreiter, S. Richter, M. Peter, S. Baumann, M. Bergen von and J. M. Tomm, Comparison and optimization of methods for the simultaneous extraction of DNA, RNA, proteins, and metabolites, *Anal. Biochem.*, 2016, **508**, 25–33, DOI: 10.1016/j.ab.2016.05.011.

44 T. S. Batth, M. Papetti, A. Pfeiffer, M. A. X. Tollenaere, C. Francavilla and J. V. Olsen, Large-Scale Phosphoproteomics Reveals Shp-2 Phosphatase-Dependent Regulators of Pdgf Receptor Signaling, *Cell Rep.*, 2018, **22**(10), 2784–2796, DOI: 10.1016/j.celrep.2018.02.038.

45 B. Bai, H. Tan and V. R. Pagala, et al., Deep Profiling of Proteome and Phosphoproteome by Isobaric Labeling, Extensive Liquid Chromatography, and Mass Spectrometry, Elsevier Inc., 1st edn, 2017, vol. 585, pp. 377–395, DOI: 10.1016/bs.mie.2016.10.007.

46 J.-M. Ahn, H.-J. Sung and Y.-H. Yoon, et al., Integrated glycoproteomics demonstrates fucosylated serum paraoxonase 1 alterations in small cell lung cancer, Mol. Cell. Proteomics, 2014, 13(1), 30–48, DOI: 10.1074/mcp.M113.028621.

47 W. Kim, E. J. Bennett and E. L. Huttlin, et al., Systematic and Quantitative Assessment of the Ubiquitin-Modified Proteome, Mol. Cell, 2011, 44(2), 325–340, DOI: 10.1016/j.molcel.2011.08.025.

48 Y. Hu, P. Shah, D. J. Clark, M. Ao and H. Zhang, Reanalysis of global proteomic and phosphoproteomic data identified a large number of glycopeptides, Anal. Chem., 2018, 90(13), 8065–8071, DOI: 10.1101/233247.

49 D. R. Jones, Z. Wu, D. Chauhan, K. C. Anderson and J. Peng, A nano ultra-performance liquid chromatography-high resolution mass spectrometry approach for global metabolomic profiling and case study on drug-resistant multiple myeloma, Anal. Chem., 2014, 86(7), 3667–3675, DOI: 10.1021/ac500476a.

50 A. J. Chetwynd, A. Abdul-Sada and E. M. Hill, Solid-Phase Extraction and Nanoflow Liquid Chromatography-Nanoelectrospray Ionization Mass Spectrometry for Improved Global Urine Metabolomics, Anal. Chem., 2014, 87(2), 1158–1165, DOI: 10.1021/ac503769q.

51 X. Luo and L. Li, Metabolomics of Small Numbers of Cells: Metabolomic Profiling of 100, 1000, and 10000 Human Breast Cancer Cells, Anal. Chem., 2017, 89(21), 11664–11671, DOI: 10.1021/acs.analchem.7b03100.

52 Z. Li, J. Tatlay and L. Li, Nanoflow LC-MS for High-Performance Chemical Isotope Labeling Quantitative Metabolomics, Anal. Chem., 2015, 87(22), 11468–11474, DOI: 10.1021/acs.analchem.5b03209.

53 R. Fischer, D. C. Trudgian and C. Wright, et al., Discovery of candidate serum proteomic and metabolomic biomarkers in ankylosing spondylitis, Mol. Cell. Proteomics, 2012, 11(2), M111.013904, DOI: 10.1074/mcp.M111.013904.

54 S. Becker, L. Kortz, C. Helmschrodt, J. Thiery and U. Ceglarek, LC-MS-based metabolomics in the clinical laboratory, J. Chromatogr. B: Anal. Technol. Biomed. Life Sci., 2012, 883–884, 68–75, DOI: 10.1016/j.jchromb.2011.10.018.

55 J.-P. Antignac, K. de Wasch, F. Monteau, H. De Brabander, F. Andre and B. Le Bizec, The ion suppression phenomenon in liquid chromatography–mass spectrometry and its consequences in the field of residue analysis, Anal. Chim. Acta, 2005, 529(1–2), 129–136, DOI: 10.1016/j.aca.2004.08.055.

56 T. O. Metz, J. S. Page and E. S. Baker, et al., High-resolution separations and improved ion production and transmission in metabolomics, TrAC, Trends Anal. Chem., 2008, 27(3), 205–214, DOI: 10.1016/j.trac.2007.11.003.

57 E. R. Wickremsinhe, G. Singh, B. L. Ackermann, T. A. Gillespie and A. K. Chaudhary, A review of nanoelectrospray ionization applications for drug metabolism and pharmacokinetics, Curr. Drug Metab., 2006, 7(8), 913–928.

58 J. P. Chervet, M. Ursem and J. P. Salzmann, Instrumental requirements for nanoscale liquid chromatography, Anal. Chem., 1996, 68(9), 1507–1512, DOI: 10.1021/ac9508964.

59 M. Wilm and M. Mann, Analytical properties of the nano-electrospray ion source, Anal. Chem., 1996, 68(1), 1–8.

60 I. Marginean, K. Tang, R. D. Smith and R. T. Kelly, Pico-electrospray ionization mass spectrometry using narrow-bore chemically etched emitters, J. Am. Soc. Mass Spectrom, 2014, 25(1), 30–36, DOI: 10.1007/s13361-013-0749-z.

61 M. Karas, U. Bahr and T. Dülcks, Nano-electrospray ionization mass spectrometry: addressing analytical problems beyond routine, Fresenius' J. Anal. Chem., 2000, 366(6–7), 669–676.

62 I. Marginean, R. T. Kelly, D. C. Prior, B. L. LaMarche, K. Tang and R. D. Smith, Analytical characterization of the electrospray ion source in the nanoflow regime, Anal. Chem., 2008, 80(17), 6573–6579, DOI: 10.1021/ac800683s.

63 A. J. Chetwynd, A. David, E. M. Hill and A. Abdul-Sada, Evaluation of analytical performance and reliability of direct nanoLC-nanoESI-high resolution mass spectrometry for profiling the (xeno)metabolome, J. Mass Spectrom., 2014, 49(10), 1063–1069, DOI: 10.1002/jms.3426.

64 R. García-Villalba, A. Carrasco-Pancorbo and G. Zurek, et al., Nano and rapid resolution liquid chromatography-electrospray ionization-time of flight mass spectrometry to identify and quantify phenolic compounds in olive oil, J. Sep. Sci., 2010, 33(14), 2069–2078, DOI: 10.1002/jssc.201000184.

65 W. Lu, B. D. Bennett and J. D. Rabinowitz, Analytical strategies for LC-MS-based targeted metabolomics, J. Chromatogr. B: Anal. Technol. Biomed. Life Sci., 2008, 871(2), 236–242, DOI: 10.1016/j.jchromb.2008.04.031.

66 R. D. Smith, Y. Shen and K. Tang, Ultrasensitive and quantitative analyses from combined separations-mass spectrometry for the characterization of proteomes, Acc. Chem. Res., 2004, 37(4), 269–278, DOI: 10.1021/ar0301330.

67 J. S. Page, R. T. Kelly, K. Tang and R. D. Smith, Ionization and transmission efficiency in an electrospray ionization-mass spectrometry interface, J. Am. Soc. Mass Spectrom, 2007, 18(9), 1582–1590, DOI: 10.1016/j.jasms.2007.05.018.

68 X. Shi, B. Wahlang and X. Wei, et al., Metabolomic Analysis of the Effects of Polychlorinated Biphenyls in Nonalcoholic Fatty Liver Disease, J. Proteome Res., 2012, 11(7), 3805–3815, DOI: 10.1021/pr300297z.

69 A. D. Southam, T. G. Payne, H. J. Cooper, T. N. Arvanitis and M. R. Viant, Dynamic range and mass accuracy of wide-scan direct infusion nanoelectrospray fourier transform ion cyclotron resonance mass spectrometry-based metabolomics increased by the spectral stitching method, Anal. Chem., 2007, 79(12), 4595–4602, DOI: 10.1021/ac062446p.

70 H. Wu, A. D. Southam, A. Hines and M. R. Viant, High-throughput tissue extraction protocol for NMR- and MS-based metabolomics, Anal. Biochem., 2008, 372(2), 204–212, DOI: 10.1016/j.ab.2007.10.002.

71 K. O. Boernsen, S. Gatzek and G. Imbert, Controlled protein precipitation in combination with chip-based nanospray infusion mass spectrometry. An approach for

metabolomics profiling of plasma, *Anal. Chem.*, 2005, **77**(22), 7255–7264, DOI: 10.1021/ac0508604.

72 E. Chekmeneva, G. D. S. Correia and Q. Chan, *et al.*, Optimization and Application of Direct Infusion Nanoelectrospray HRMS Method for Large-Scale Urinary Metabolic Phenotyping in Molecular Epidemiology, *J. Proteome Res.*, 2017, **16**(4), 1646–1658, DOI: 10.1021/acs.jproteome.6601003.

73 R.-J. Raterink, F. M. van der Kloet and J. Li, *et al.*, Rapid metabolic screening of early zebrafish embryogenesis based on direct infusion-nanoESI-FTMS, *Metabolomics*, 2013, **9**(4), 864–873, DOI: 10.1007/s11306-012-0493-6.

74 N. S. Taylor, R. J. M. Weber and A. D. Southam, *et al.*, A new approach to toxicity testing in *Daphnia magna*: application of high throughput FT-ICR mass spectrometry metabolomics, *Metabolomics*, 2009, **5**(1), 44–58, DOI: 10.1007/s11306-008-0133-3.

75 J.-M. Dethy, B. L. Ackermann, C. Delatour, J. D. Henion and G. A. Schultz, Demonstration of direct bioanalysis of drugs in plasma using nanoelectrospray infusion from a silicon chip coupled with tandem mass spectrometry, *Anal. Chem.*, 2003, **75**(4), 805–811.

76 C. E. C. A. Hop, Y. Chen and L. J. Yu, Uniformity of ionization response of structurally diverse analytes using a chip-based nanoelectrospray ionization source, *Rapid Commun. Mass Spectrom.*, 2005, **19**(21), 3139–3142, DOI: 10.1002/rcm.2182.

77 E. Chekmeneva, G. Correia and J. Denes, *et al.*, Development of nanoelectrospray high resolution isotope dilution mass spectrometry for targeted quantitative analysis of urinary metabolites: application to population profiling and clinical studies, *Anal. Methods*, 2015, **7**(12), 5122–5133, DOI: 10.1039/c5ay00850f.

78 J. Sestak, D. Moravcova and V. Kahle, Instrument platforms for nano liquid chromatography, *J. Chromatogr. A*, 2015, **1421**, 2–17, DOI: 10.1016/j.chroma.2015.07.090.

79 M. Noga, F. Sucharski, P. Suder and J. Silberring, A practical guide to nano-LC troubleshooting, *J. Sep. Sci.*, 2007, **30**(14), 2179–2189, DOI: 10.1002/jssc.200700225.

80 A. J. Chetwynd, A. Abdul-Sada, S. G. Holt and E. M. Hill, Use of a pre-analysis osmolality normalisation method to correct for variable urine concentrations and for improved metabolomic analyses, *J. Chromatogr. A*, 2016, **1431**, 103–110, DOI: 10.1016/j.chroma.2015.12.056.

81 A. J. Chetwynd, A. Samarawickrama and J. H. Vera, *et al.*, Nanoflow-Nanospray Mass Spectrometry Metabolomics Reveals Disruption of the Urinary Metabolite Profiles of HIV-Positive Patients on Combination Antiretroviral Therapy, *J. Acquired Immune Defic. Syndr.*, 2017, **74**(2), e45–e53, DOI: 10.1097/QAI.0000000000001159.

82 S. Medina, F. Ferreres and C. García-Viguera, *et al.*, Non-targeted metabolomic approach reveals urinary metabolites linked to steroid biosynthesis pathway after ingestion of citrus juice, *Food Chem.*, 2013, **136**(2), 938–946, DOI: 10.1016/j.foodchem.2012.09.004.

83 M. L. Stoll, R. Kumar, E. J. Lefkowitz, R. Q. Cron, C. D. Morrow and S. Barnes, Fecal metabolomics in

pediatric spondyloarthritis implicate decreased metabolic diversity and altered tryptophan metabolism as pathogenic factors, *Genes Immun.*, 2016, **17**(7), 400–405, DOI: 10.1038/gene.2016.38.

84 A. S. Kononikhin, N. L. Starodubtseva and V. V. Chagovets, *et al.*, Exhaled breath condensate analysis from intubated newborns by nano-HPLC coupled to high resolution MS, *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.*, 2017, **1047**, 97–105, DOI: 10.1016/j.jchromb.2016.12.036.

85 K. T. Myint, K. Aoshima, S. Tanaka, T. Nakamura and Y. Oda, Quantitative profiling of polar cationic metabolites in human cerebrospinal fluid by reversed-phase nanoliquid chromatography/mass spectrometry, *Anal. Chem.*, 2009, **81**(3), 1121–1129, DOI: 10.1021/ac802259r.

86 A. David, A. Lange, A. Abdul-Sada, C. R. Tyler and E. M. Hill, Disruption of the Prostaglandin Metabolome and Characterization of the Pharmaceutical Exposome in Fish Exposed to Wastewater Treatment Works Effluent As Revealed by Nanoflow-Nanospray Mass Spectrometry-Based Metabolomics, *Environ. Sci. Technol.*, 2017, **51**(1), 616–624, DOI: 10.1021/acs.est.6b04365.

87 J. Hernández-Borges, Z. Aturki and A. Rocco, Fanali S. Recent applications in nanoliquid chromatography, *J. Sep. Sci.*, 2007, **30**(11), 1589–1610, DOI: 10.1002/jssc.200700061.

88 G. A. Valaskovic, J. P. Murphy and M. S. Lee, Automated orthogonal control system for electrospray ionization, *J. Am. Soc. Mass Spectrom*, 2004, **15**(8), 1201–1215, DOI: 10.1016/j.jasms.2004.04.033.

89 C. E. D. Nazario, M. R. Silva, M. S. Franco and F. M. Lanças, Evolution in miniaturized column liquid chromatography instrumentation and applications: An overview, *J. Chromatogr. A*, 2015, **1421**, 18–37, DOI: 10.1016/j.chroma.2015.08.051.

90 Y. Saito, K. Jinno and T. Greibrokk, Capillary columns in liquid chromatography: between conventional columns and microchips, *J. Sep. Sci.*, 2004, **27**(17–18), 1379–1390.

91 D. A. Collins, E. P. Nesterenko and B. Paull, Porous layer open tubular columns in capillary liquid chromatography, *Analyst*, 2014, **139**(6), 1292–1302, DOI: 10.1039/c3an01869e.

92 M. Hilhorst, C. Briscoe and N. van de Merbel, Sense and nonsense of miniaturized LC-MS/MS for bioanalysis, *Bioanalysis*, 2014, **6**(24), 3263–3265, DOI: 10.4155/bio.14.263.

93 M. Motokawa, H. Kobayashi and N. Ishizuka, *et al.*, Monolithic silica columns with various skeleton sizes and through-pore sizes for capillary liquid chromatography, *J. Chromatogr. A*, 2002, **961**(1), 53–63.

94 E. J. Want, I. D. Wilson and H. Gika, *et al.*, Global metabolic profiling procedures for urine using UPLC-MS, *Nat. Protoc.*, 2010, **5**(6), 1005–1018, DOI: 10.1038/nprot.2010.50.

95 P. Kiefer, N. Delmotte and J. A. Vorholt, Nanoscale ion-pair reversed-phase HPLC-MS for sensitive metabolome analysis, *Anal. Chem.*, 2011, **83**(3), 850–855, DOI: 10.1021/ac102445r.

96 L. Imbert, M. Gaudin and D. Libong, *et al.*, Comparison of electrospray ionization, atmospheric pressure chemical ionization and atmospheric pressure photoionization for a lipidomic analysis of Leishmania donovani, *J. Chromatogr. A*, 2012, **1242**, 75–83, DOI: 10.1016/j.chroma.2012.04.035.

97 F. Lopes, D. A. Cowan, M. Thevis, A. Thomas and M. C. Parkin, Quantification of intact human insulin-like growth factor-I in serum by nano-ultrahigh-performance liquid chromatography/tandem mass spectrometry, *Rapid Commun. Mass Spectrom.*, 2014, **28**(13), 1426–1432, DOI: 10.1002/rcm.6908.

98 G. T. T. Gibson, S. M. Mugo and R. D. Oleschuk, Nanoelectrospray emitters: trends and perspective, *Mass Spectrom. Rev.*, 2009, **28**(6), 918–936, DOI: 10.1002/mas.20248.

99 R. T. Kelly, J. S. Page and Q. Luo, *et al.*, Chemically etched open tubular and monolithic emitters for nanoelectrospray ionization mass spectrometry, *Anal. Chem.*, 2006, **78**(22), 7796–7801, DOI: 10.1021/ac061133r.

100 W. Shui, Y. Yu, X. Xu, Z. Huang, G. Xu and P. Yang, Microelectrospray with stainless steel emitters, *Rapid Commun. Mass Spectrom.*, 2003, **17**(14), 1541–1547, DOI: 10.1002/rcm.1083.

101 W. Xiong, J. Glick, Y. Lin and P. Vouros, Separation and sequencing of isomeric oligonucleotide adducts using monolithic columns by ion-pair reversed-phase nano-HPLC coupled to ion trap mass spectrometry, *Anal. Chem.*, 2007, **79**(14), 5312–5321, DOI: 10.1021/ac0701435.

102 C. D. Kelstrup, D. B. Bekker-Jensen, T. N. Arrey, A. Hogrebe, A. Harder and J. V. Olsen, Performance Evaluation of the Q Exactive HF-X for Shotgun Proteomics, *J. Proteome Res.*, 2017, **17**(1), 727–738, DOI: 10.1021/acs.jproteome.7b00602.

103 C. A. Crutchfield, W. Lu, E. Melamud and J. D. Rabinowitz, Mass Spectrometry-Based Metabolomics of Yeast, *Methods Enzymol.*, 2010, **470**, 393–426, DOI: 10.1016/S0076-6879(10)70016-1.

104 P. A. Rudnick, S. P. Markey and J. Roth, *et al.*, A Description of the Clinical Proteomic Tumor Analysis Consortium (CPTAC) Common Data Analysis Pipeline, *J. Proteome Res.*, 2016, **15**(3), 1023–1032, DOI: 10.1021/acs.jproteome.5b01091.

105 J. Barretina, G. Caponigro and N. Stransky, *et al.*, The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity, *Nature*, 2012, **483**(7391), 603–607, DOI: 10.1038/nature11003.

106 Network TCGAR, B. J. Raphael and R. H. Hruban, *et al.*, Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma, *Cancer Cell*, 2017, **32**(2), 185–203, DOI: 10.1016/j.ccell.2017.07.007.

107 T. U. Consortium, UniProt: the universal protein knowledgebase, *Nucleic Acids Res.*, 2016, **45**(D1), D158–D169, DOI: 10.1093/nar/gkw1099.

108 M. Hirosawa, M. Hoshida, M. Ishikawa and T. Toya, MASCOT: multiple alignment system for protein sequences based on three-way dynamic programming, *CABIOS, Comput. Appl. Biosci.*, 1993, **9**(2), 161–167.

109 R. Craig and R. C. Beavis, TANDEM: matching proteins with tandem mass spectra, *Bioinformatics*, 2004, **20**(9), 1466–1467, DOI: 10.1093/bioinformatics/bth092.

110 M. Brosch, S. Swamy, T. Hubbard and J. Choudhary, Comparison of Mascot and X!Tandem performance for low and high accuracy mass spectrometry and the development of an adjusted Mascot threshold, *Mol. Cell. Proteomics*, 2008, **7**(5), 962–970, DOI: 10.1074/mcp.M700293-MCP200.

111 D. L. Tabb, C. G. Fernando and M. C. Chambers, MyriMatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis, *J. Proteome Res.*, 2007, **6**(2), 654–661, DOI: 10.1021/pr0604054.

112 H. Lam, E. W. Deutsch, J. S. Eddes, J. K. Eng, S. E. Stein and R. Aebersold, Building consensus spectral libraries for peptide identification in proteomics, *Nat. Methods*, 2008, **5**(10), 873–875, DOI: 10.1038/nmeth.1254.

113 J. K. Eng, T. A. Jahan and M. R. Hoopmann, Comet: an open-source MS/MS sequence database search tool, *Proteomics*, 2013, **13**(1), 22–24, DOI: 10.1002/pmic.201200439.

114 M. Vaudel, H. Barsnes, F. S. Berven, A. Sickmann and L. Martens, SearchGUI: an open-source graphical user interface for simultaneous OMSSA and X!Tandem searches, *Proteomics*, 2011, **11**(5), 996–999, DOI: 10.1002/pmic.201000595.

115 Z. Zhang, M. Burke and Y. A. Mirokhin, *et al.*, Reverse and Random Decoy Methods for False Discovery Rate Estimation in High Mass Accuracy Peptide Spectral Library Searches, *J. Proteome Res.*, 2018, **17**(2), 846–857, DOI: 10.1021/acs.jproteome.7b00614.

116 G. W. Park, H. Hwang and K. H. Kim, *et al.*, Integrated Proteomic Pipeline Using Multiple Search Engines for a Proteogenomic Study with a Controlled Protein False Discovery Rate, *J. Proteome Res.*, 2016, **15**(11), 4082–4090, DOI: 10.1021/acs.jproteome.6b00376.

117 S. Tyanova, T. Temu and J. Cox, The MaxQuant computational platform for mass spectrometry-based shotgun proteomics, *Nat. Protoc.*, 2016, **11**(12), 2301–2319, DOI: 10.1038/nprot.2016.136.

118 J. Pfeuffer, T. Sachsenberg and O. Alka, *et al.*, OpenMS - A platform for reproducible analysis of mass spectrometry data, *J. Biotechnol.*, 2017, **261**, 142–148, DOI: 10.1016/j.jbiotec.2017.05.016.

119 F. Fournier, C. Joly Beauparlant and R. Paradis, Droit A. rTANDEM, an R/Bioconductor package for MS/MS protein identification, *Bioinformatics*, 2014, **30**(15), 2233–2234, DOI: 10.1093/bioinformatics/btu178.

120 E. W. Deutsch, L. Mendoza, D. Shteynberg, J. Slagel, Z. Sun and R. L. Moritz, Trans-Proteomic Pipeline, a standardized data processing pipeline for large-scale reproducible proteomics informatics, *Proteomics Clin Appl*, 2015, **9**(7-8), 745–754, DOI: 10.1002/prca.201400164.

121 S. Kim and P. A. Pevzner, MS-GF+ makes progress towards a universal database search tool for proteomics, *Nat. Commun.*, 2014, **5**, 5277, DOI: 10.1038/ncomms6277.

122 C. A. Smith, G. O'Maille and E. J. Want, *et al.*, METLIN: a metabolite mass spectral database, *Ther. Drug Monit.*, 2005, **27**(6), 747–751.

123 D. S. Wishart, Y. D. Feunang and A. Marcu, *et al.*, HMDB 4.0: the human metabolome database for 2018, *Nucleic Acids Res.*, 2017, **46**(D1), D608–D617, DOI: 10.1093/nar/gkx1089.

124 I. Blaženović, T. Kind, J. Ji and O. Fiehn, Software Tools and Approaches for Compound Identification of LC-MS/MS

Data in Metabolomics, *Metabolites*, 2018, **8**(2), 31, DOI: 10.3390/metabo8020031.

125 N. G. Mahieu and G. J. Patti, Systems-Level Annotation of a Metabolomics Data Set Reduces 25 000 Features to Fewer than 1000 Unique Metabolites, *Anal. Chem.*, 2017, **89**(19), 10397–10406, DOI: 10.1021/acs.analchem.7b02380.

126 M. C. Chambers, B. MacLean and R. Burke, *et al.*, A cross-platform toolkit for mass spectrometry and proteomics, *Nat. Biotechnol.*, 2012, **30**(10), 918–920, DOI: 10.1038/nbt.2377.

127 T. Pluskal, S. Castillo, A. Villar-Briones and M. Oresic, MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data, *BMC Bioinf.*, 2010, **11**(1), 395, DOI: 10.1186/1471-2105-11-395.

128 F. Fernández-Albert, R. Llorach, C. Andrés-Lacueva and A. Perera, An R package to analyse LC/MS metabolomic data: MAIT (Metabolite Automatic Identification Toolkit), *Bioinformatics*, 2014, **30**(13), 1937–1939, DOI: 10.1093/bioinformatics/btu136.

129 H. Tsugawa, T. Cajka and T. Kind, *et al.*, MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis, *Nat. Methods*, 2015, **12**(6), 523–526, DOI: 10.1038/nmeth.3393.

130 C. A. Smith, E. J. Want, G. O'Maille, R. Abagyan and G. Siuzdak, XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification, *Anal. Chem.*, 2006, **78**(3), 779–787, DOI: 10.1021/ac051437y.

131 J. Xia, N. Psychogios, N. Young and D. S. Wishart, Metabo-Analyst: a web server for metabolomic data analysis and interpretation, *Nucleic Acids Res.*, 2009, **37**(Web Server issue), W652–W660, DOI: 10.1093/nar/gkp356.

132 T. Huan, E. M. Forsberg and D. Rinehart, *et al.*, Systems biology guided by XCMS Online metabolomics, *Nat. Methods*, 2017, **14**(5), 461–462, DOI: 10.1038/nmeth.4260.

133 E. Kenar, H. Franken and S. Forcisi, *et al.*, Automated label-free quantification of metabolites from liquid chromatography-mass spectrometry data, *Mol. Cell. Proteomics*, 2014, **13**(1), 348–359, DOI: 10.1074/mcp.M113.031278.

134 A. Cambiaghi, M. Ferrario and M. Masseroli, Analysis of metabolomic data: tools, current strategies and future challenges for omics data integration, *Briefings Bioinf.*, 2017, **18**(3), 498–510, DOI: 10.1093/bib/bbw031.

135 T. Välikangas, T. Suomi and L. L. Elo, A systematic evaluation of normalization methods in quantitative label-free proteomics, *Briefings Bioinf.*, 2018, **19**(1), 1–11, DOI: 10.1093/bib/bbw095.

136 M. E. Ritchie, B. Phipson and D. Wu, *et al.*, limma powers differential expression analyses for RNA-sequencing and microarray studies, *Nucleic Acids Res.*, 2015, **43**(7), e47–e47, DOI: 10.1093/nar/gkv007.

137 M. Kuhn, Building Predictive Models in R Using the caret Package, *J. Stat. Softw.*, 2008, **28**(5), 1–26.

138 F. Pedregosa, G. Varoquaux and A. Gramfort, *et al.*, Scikit-learn: Machine Learning in Python, *J. Mach. Learn. Res.*, 2011, **12**(1), 2825–2830, DOI: 10.18637/jss.v033.i01.

139 A. Subramanian, P. Tamayo and V. K. Mootha, *et al.*, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**(43), 15545–15550, DOI: 10.1073/pnas.0506580102.

140 M. V. Kuleshov, M. R. Jones and A. D. Rouillard, *et al.*, Enrichr: a comprehensive gene set enrichment analysis web server 2016 update, *Nucleic Acids Res.*, 2016, **44**(W1), W90–W97, DOI: 10.1093/nar/gkw377.

141 J. Reimand, R. Isserlin and V. Voisin, *et al.*, Pathway enrichment analysis of -omics data, *Genes*, 2017, **8**(10), 246, DOI: 10.1101/232835.

142 F. Rohart, B. Gautier, A. Singh and K.-A. Lê Cao, mixOmics: An R package for 'omics feature selection and multiple data integration, *PLoS Comput. Biol.*, 2017, **13**(11), e1005752, DOI: 10.1371/journal.pcbi.1005752.

143 T.-C. Kuo, T.-F. Tian and Y. J. Tseng, 3Omics: a web-based systems biology tool for analysis, integration and visuali-zation of human transcriptomic, proteomic and metabo-lomic data, *BMC Syst. Biol.*, 2013, **7**(1), 64, DOI: 10.1186/1752-0509-7-64.

144 J. Eichner, L. Rosenbaum, C. Wrzodek, H.-U. Häring, A. Zell and R. Lehmann, Integrated enrichment analysis and pathway-centered visualization of metabolomics, proteomics, transcriptomics, and genomics data by using the InCroMAP software, *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.*, 2014, **966**, 77–82, DOI: 10.1016/j.jchromb.2014.04.030.

145 E. B. Unterberger, J. Eichner and C. Wrzodek, *et al.*, Ha-ras and beta-catenin oncoproteins orchestrate metabolic pro-grams in mouse liver tumors, *Int. J. Cancer*, 2014, **135**(7), 1574–1585, DOI: 10.1002/ijc.28798.

146 D. Szklarczyk, A. Santos, C. Mering von, L. J. Jensen, P. Bork and M. Kuhn, STITCH 5: augmenting protein-chemical interaction networks with tissue and affinity data, *Nucleic Acids Res.*, 2016, **44**(D1), D380–D384, DOI: 10.1093/nar/gkv1277.

147 E. Brunk, S. Sahoo and D. C. Zielinski, *et al.*, Recon3D enables a three-dimensional view of gene variation in human metabolism, *Nat. Biotechnol.*, 2018, **36**(3), 272–281, DOI: 10.1038/nbt.4072.

148 S. Razick, G. Magklaras and I. M. Donaldson, iRefIndex: a consolidated protein interaction database with provenance, *BMC Bioinf.*, 2008, **9**(1), 405, DOI: 10.1186/1471-2105-9-405.

149 N. Tuncbag, S. J. C. Gosline, A. Kedaigle, A. R. Soltis, A. Gitter and E. Fraenkel, Network-Based Interpretation of Diverse High-Throughput Datasets through the Omics Integrator Software Package, *PLoS Comput. Biol.*, 2016, **12**(4), e1004879, DOI: 10.1371/journal.pcbi.1004879.

150 L. Pirhaji, P. Milani and M. Leidl, *et al.*, Revealing disease-associated pathways by network integration of untargeted metabolomics, *Nat. Methods*, 2016, **13**(9), 770–776, DOI: 10.1038/nmeth.3940.

151 J. Wang, Landscape and flux theory of non-equilibrium dynamical systems with application to biology, *Adv. Phys.*, 2015, **64**(1), 1–137, DOI: 10.1080/00018732.2015.1037068.

152 D. K. Wells, W. L. Kath and A. E. Motter, Control of Stochastic and Induced Switching in Biophysical Networks, *Phys. Rev. X*, 2015, **5**(3), 031036, DOI: 10.1103/PhysRevX.5.031036.