Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Chemical Biology

Recent advances in mass spectrometry-based computational metabolomics

Timothy M. D. Ebbels[1], Justin J. J. van der Hooft[2,3], Haley Chatelaine[5], Corey Broeckling[4], Nicola Zamboni[8], Soha Hassoun[6,7] and Ewy A. Mathé[5]

Abstract

The computational metabolomics field brings together computer scientists, bioinformaticians, chemists, clinicians, and biologists to maximize the impact of metabolomics across a wide array of scientific and medical disciplines. The field continues to expand as modern instrumentation produces datasets with increasing complexity, resolution, and sensitivity. These datasets must be processed, annotated, modeled, and interpreted to enable biological insight. Techniques for visualization, integration (within or between omics), and interpretation of metabolomics data have evolved along with innovation in the databases and knowledge resources required to aid understanding. In this review, we highlight recent advances in the field and reflect on opportunities and innovations in response to the most pressing challenges. This review was compiled from discussions from the 2022 Dagstuhl seminar entitled "Computational Metabolomics: From Spectra to Knowledge".

Addresses

[1] Section of Bioinformatics, Department of Metabolism, Digestion & Reproduction, Imperial College London, Burlington Danes Building, Hammersmith Hospital, Du Cane Road, London W12 0NN, UK
[2] Bioinformatics Group, Wageningen University & Research, Wageningen 6708 PB, the Netherlands
[3] Department of Biochemistry, University of Johannesburg, Auckland Park, Johannesburg 2006, South Africa
[4] Bioanalysis and Omics Center, Analytical Resources Core, Colorado State University, Fort Collins, CO, USA
[5] Informatics Core, Division of Preclinical Innovation, National Center for Advancing Translational Sciences, Rockville, MD, USA
[6] Department of Computer Science, Tufts University, Medford, MA, USA
[7] Department of Chemical and Biological Engineering, Tufts University, Medford, MA, USA
[8] Institute of Molecular Systems Biology, ETH Zurich, Zurich, Switzerland

Corresponding authors: Ebbels, Timothy MD. (t.ebbels@imperial.ac.uk); Mathé, Ewy A. (ewy.mathe@nih.gov)

Current Opinion in Chemical Biology 2023, **74**:102288

This review comes from a themed issue on **Omics - Metabolomics (2023)**

Edited by **James McCullagh** and **Hector Keun**

For complete overview of the section, please refer the article collection - Omics - Metabolomics (2023)

Available online 24 March 2023

https://doi.org/10.1016/j.cbpa.2023.102288

Keywords

Metabolomics, Multi-omics, Machine learning, Metabolite identification, Cheminformatics, Chemometrics, Visualization, Benchmarking, Small molecules.

Introduction

As the field of metabolomics continues to grow, computational methods that enable analysis and interpretation of these complex data are of paramount importance. This area, now known as computational metabolomics, is a highly interdisciplinary science lying at the intersection of computer science, bioinformatics, chemistry, medicine, and biology. It focuses on applying computational, statistical, and machine learning methods to analyze and interpret metabolomic data and its integration with other datasets, such as other omics or clinical data. Computational metabolomics is swiftly evolving, making reviews of the field quickly out of date. Nonetheless, there have been several excellent reviews of tools and resources in the last two years [1–3], serving as important guides to both experienced scientists and trainees. Program and community efforts also exist that consolidate information on the availability and usage of data and tools [4,5] and systematically identify challenges in the use of metabolomics data in multi-omics research [6]. Our intent here is i) to provide a current, concise overview of the latest advances in mass spectrometry-based computational metabolomics and ii) to reflect on current challenges and proposed solutions (Figure 1). This review stems from the most recent seminar in the Dagstuhl Computational Metabolomics series, held in Schloss Dagstuhl, Germany in May 2022, entitled "Computational Metabolomics: From Spectra to Knowledge".

Data acquisition and processing techniques

Recent developments in LC-MS metabolomic data processing include the latest releases of MZmine 3 [7], MS-DIAL 5 [8], and XCMS [9], which provide raw data

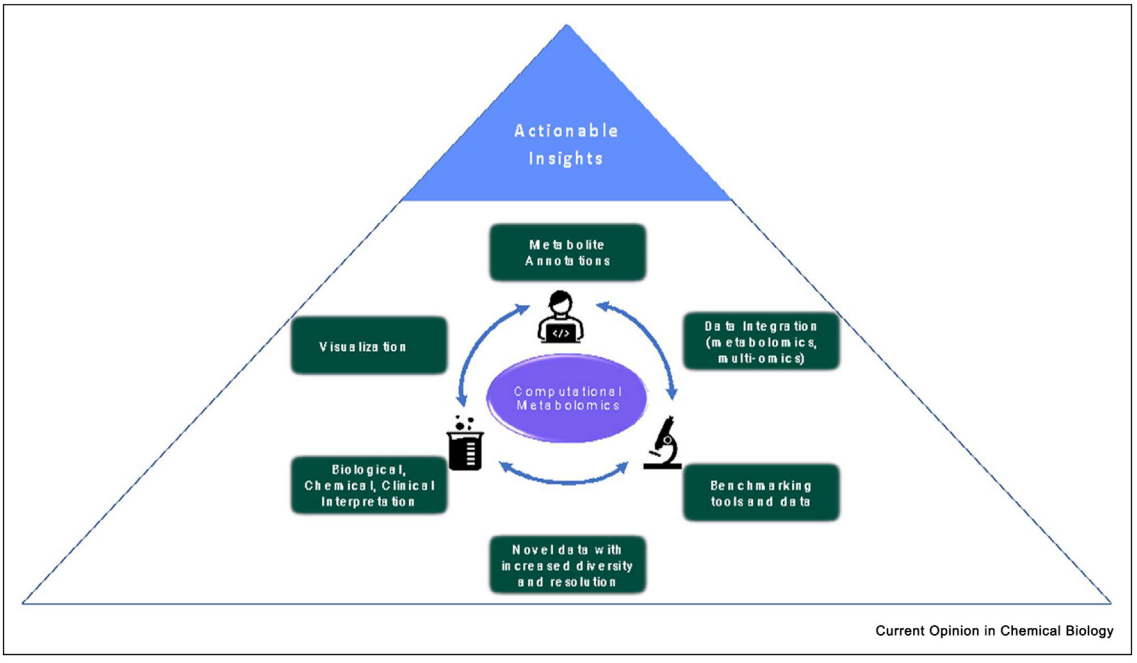
Abbreviations

API application programming interface
CANOPUS clear assignment and ontology prediction using mass spectrometry
CASMI Critical Assessment of Small Molecule Identification
CFM-ID competitive fragmentation modeling identification
COMETS Consortium for Metabolomics Studies
DDA data dependent acquisition
DIA data independent acquisition
EMM extended metabolic model
GNPS Global Natural Product Social Molecular Networking
HMDB Human Metabolome Database
KEGG Kyoto Encyclopedia of Genes and Genomes
LASSO least absolute shrinkage and selection operator
LC liquid chromatography
MassQL mass spectrometry query language
mQACC Metabolomics Quality Assurance and Quality Control Consortium
MS mass spectrometry
MS/MS tandem mass spectrometry
NLP natural language processing
NMDR National Metabolomics Data Repository
PCA principal components analysis
PUMA Probabilistic modeling for Untargeted Metabolomics Analysis
RaMP DB: relational database of metabolic pathways
SODA Software Data Exchange (part of the Metabolomics Association of North America)
t-SNE t-distributed stochastic neighbor embedding
TMAP tree manifold approximation and projection
UMAP uniform manifold approximation and projection

exploration and processing capabilities using state-of-the art approaches. In addition there have been new methods addressing problems such as scalability [10] and even peak detection-free processing [11]. Both the R and Python environments have seen considerable development for versatile MS and MS/MS data processing, with major efforts to modernize legacy code as well as adding new functionality [12–15].

Tandem mass spectrometry (MS/MS) supports confident annotation of metabolites and can, through improved selectivity and sensitivity, enhance quantitation accuracy. Data independent acquisition MS/MS (DIA) has emerged as an enticing alternative to traditional data dependent (DDA) methods. One clear advantage of DIA is the potential to provide complete MS/MS coverage, an area where DDA has generally

Figure 1



A conceptual overview of computational metabolomics.

struggled due to duty cycle limitations. DIA has recently been used to support stable isotope resolved metabolomics and to provide constant sampling of variably labeled precursors, with fragment ion spectra enabling position specific labeling assignments [16].

Fully untargeted precursor-product assignment using DIA increases coverage and MS duty cycle at the expense of decreasing (or even eliminating) precursor mass selectivity. As a result of the decreased selectivity, DIA spectra are more highly convoluted than DDA spectra. DIA is implemented on instruments with diverse architecture, where precursor selection can be performed by quadrupoles or ion mobility cells (or both), with discrete or overlapping precursor windows that can be wide or narrow. However, testing all fragmentation strategies that encompass the diversity and complexity of MS/MS spectra is impractical. A recent in silico framework addresses this issue by evaluating data acquisition strategies on their coverage and mass spectral quality, thus reducing machine time [17,18]. A particularly powerful approach to analyzing mass fragmentation spectra, including those from DIA, is molecular networking [19], enabled within the Global Natural Product Social Molecular Networking (GNPS) resource. This technique embeds spectra according to similarity within a large network of public data, enabling grouping of structurally related molecules and greatly aiding metabolite annotation. Additionally, both spectrum- [20] and chromatogram-centric [21] approaches are under active development in proteomics data processing, which can serve as inspiration for continued development of metabolomics DIA data processing for both qualitative and quantitative aims.

Several software tools have been released recently that enable guided interpretation of DIA metabolomics data. DIAMetAlyzer utilizes a DDA-guided approach to reap the quantitative benefits of DIA, while supporting confident annotation, flexibility in DDA library generation, and estimating annotation false discovery rates [22]. Another approach, DecoID, uses a LASSO regression model to estimate the proportions of the DDA MS/MS spectra that contribute to the DIA spectrum, thereby deconvolving chimeric DIA spectra [23]. MetaboAnnotatoR also addresses deconvolution by using peak shape correlation in DIA spectra, as well as exploiting highly configurable spectral libraries [24]. Despite this progress, there remains opportunity to more fully exploit DIA-based MS/MS to improve quantitative and qualitative metabolomics data.

Adductomics

Adductomics is an approach to studying the chemical modification of biological macromolecules by environmental exposures. A key premise of this approach is that highly abundant biological macromolecules, such as proteins or DNA, can serve as probes that reflect long-term trends in potentially short-lived or low concentration small molecules. As an example, 2-Amino-1-methyl-6-phenylimidazo [4,5-b]pyridine (PhIP), a small molecule formed in cooked meat, is a potential carcinogen and can covalently modify DNA. PhIP, lipid peroxidation, and other DNA adduct types were profiled and associated with the prostate cancer Gleason score [25]. Dedicated methods for detecting and annotating DNA adducts, guided by user supplied accurate mass fragment ions and neutral losses, have been incorporated into MZmine to facilitate wider adoption of adductomics [26]. The resulting adducts can serve as potential biomarkers, for example, serum albumin Cys34 as a probe for monitoring air pollution [27] or albumin adducts as probes for prenatal exposure to airborne pollution [28]. Together, these analytical and computational approaches facilitate discovery of exposure markers that are otherwise inaccessible to traditional metabolomics.

Structural annotation of metabolites

In untargeted metabolomics studies, which aim to capture the broadest range of metabolites in a biospecimen, only a subset of observed signals (often called features) are structurally annotated. The matching of experimental MS/MS spectra to mass spectral libraries is generally recognized as the first step in the structural annotation process of MS/MS spectra that do not match to in-house standards [29]. It is important to note that many false positives are typically returned and that manual validation of key metabolite identities remains important. Since the annotation coverage of library matching remains low, recent annotation advances make use of machine and deep learning models, giving rise to conceptual frameworks that either map molecules to spectra, thus mimicking the physical processes such as fragmentation, or spectra to molecules, which is an inverse problem. Machine learning methods may represent molecular structures using fingerprints — vectors that capture the presence or absence of various structural properties. However, graph neural networks have emerged as strong competitors to fingerprint-based methods for representing molecules [30–32]. An exciting new development is the use of transformers and recursive neural networks that can generate molecular structures directly from spectra [33,34]. Learned spectral representations using unsupervised machine learning can enhance downstream annotation [32,35].

As structural annotation often remains ambiguous, partial annotation techniques have also gained prominence. For example, CANOPUS [36] uses deep learning to predict ClassyFire taxonomies from MS/MS spectra through an intermediate molecular fingerprint step. Another recent development is NPClassifier [37], which uses a deep neural network to predict the structural class of natural products. Molecular networking methods continue to develop, especially those within the GNPS platform, allowing propagation of structural information [38] and

can be combined with chromatographic peak shape correlations to recognize metabolite features [39]. MS2DeepScore, which trains a deep learning model on two spectra to predict their molecular similarity, shows promise in downstream annotation tasks such as library matching and analogue searches [40].

Interrogation of MS data is becoming easier with the introduction of MassQL, which enables queries based on precursor isotopic patterns, MS/MS spectra, drift time, and other parameters [41]. We note here that, whilst powerful, MS/MS spectra cannot discriminate between all molecules. For example, stereoisomers may have almost the same fragmentation spectrum. Here, complementary information is required, such as retention time information, although this is typically platform-dependent. Despite this, recent work that includes information from both MS/MS spectra as well as the platform-independent retention time order shows promise in further improvement of annotation performance [42]. A crucial remaining question is how machine learning methods can leverage the wealth of unpaired molecular and spectral data to better learn molecular and spectral representations and serve downstream annotation and classification tasks. It will be exciting to see how repository-scale analyses will assist in metabolite origin, structure, function, and novelty prediction [43,44].

Visualization

Visualization is an important component throughout the data analytical workflow, from quality control assessment to interpretation. To interpret metabolomics data, it is important to visualize it, perhaps in combination with other omics data, and in conjunction with additional information such as biological processes (e.g. diseases) and/or pathways. Many visualizations give insight into how metabolites relate to each other, for example through structural or spectral similarity or enzymatic reactions connections. Often the data are said to be visualized in a (bio)chemical space. However, this concept is context-dependent [45], as it could include for example, all known biomolecules (metabolites, genes, proteins, etc.), or those identified in one experiment, in a number of samples, or in a specific sample type.

A wide range of dimension reduction techniques have been applied to embed biomolecules [46,47], such as PCA, t-SNE, UMAP, and TMAP [47]. Figure 2 shows how TMAP highlights the chemical relationships between structures in GNPS mass spectral libraries [48], and how Treemap visualizes frequently occurring ClassyFire compound classes. Combining different visualization methods can provide insight into various aspects of the metabolomics pipeline, including assessment of biases in training data used for machine learning models.

Network-based visualizations are increasingly applied to connect structures or analytical features [49]. At a basic level, networks comprise nodes, which could represent metabolites, other biomolecules, or various annotations, and edges, which represent relationships between nodes. Examples of edge types include mass spectral similarity (quantified by e.g., cosine score), structure similarity (e.g., Tanimoto similarity), abundance correlation, or chemical relatedness [50]. Recent advances in scalable network visualization techniques simplify the complexity of networks and help identify areas for further study [51].

It is important to recognize that visualization techniques can be biased and may erroneously display artifacts or redundant information [52]. The use of robust statistical and heuristic techniques is critical to ensure appropriate interpretation and reproducibility of findings.

Data interpretation

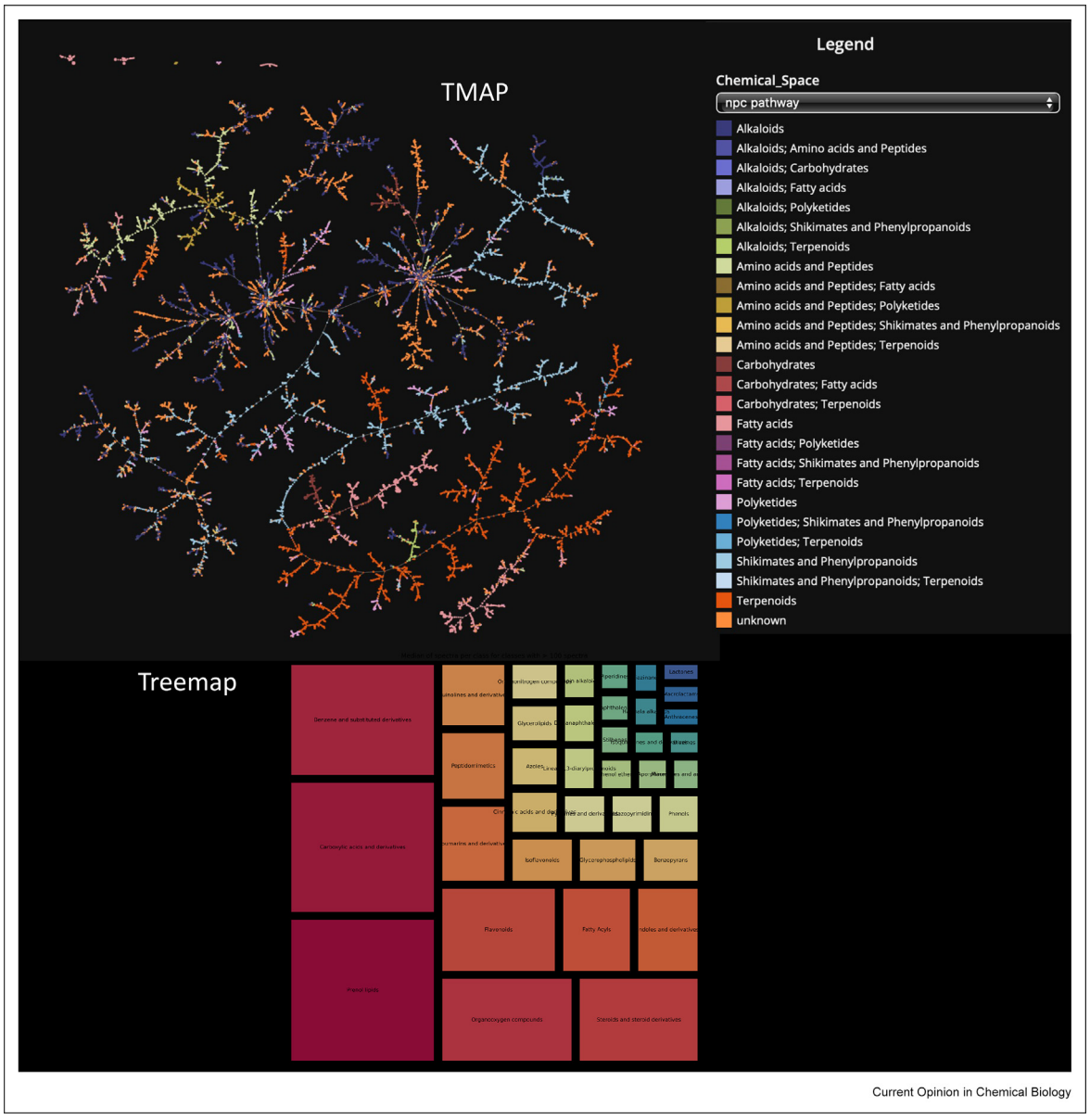
Databases and knowledge resources

Databases and knowledge graphs form critical components of any computational workflow, allowing users to compare experimental data to existing knowledge. For example, LOTUS reports structures of most plant natural products along with their taxonomy [53]. Similarly, the NP Atlas provides curated information on microbial metabolites with taxa descriptions [54]. Another large curation effort generated annotations and bioactivities of exposome-relevant molecules in PubChemLite [55]. HMDB 5.0 provides an increasing amount of biochemical, analytical, and pathway information on the human metabolome, with currently >200 K metabolite entries [56].

A noticeable recent trend is the use of natural language processing (NLP) to annotate metabolomic data by mining the literature. NJC19, for example, maps metabolites to microbes that produce or consume them [57], while FORUM [58] maps metabolites to diseases, enabling users to associate metabolites with disease risk. Deep learning algorithms are also becoming more prominent in this area [59]. Based on these developments, we anticipate an increase in this use of NLP in computational metabolomics.

Integrative platforms, such as MetaboAnalyst [60] streamline data analysis and interpretation using biochemical and clinical annotations through a web-based environment and are of great help to labs with little computational support. Importantly, there are multiple source databases that provide annotations for interpretation of metabolomic and multi-omic data. Recently updated resources include KEGG [61], HMDB [56], Reactome [62], WikiPathways [63] and LIPIDMAPS [64]. To expand the breadth of useful and up-to-date annotations, RaMP-DB [65] aggregates

Figure 2



Visualization of biochemical space. Top - TMAP visualization of ~24 K unique GNPS mass spectral library molecules colored by the NPClassifier pathway level. Various branches of specialized metabolism become visible. Bottom - Treemap visualization of a subset (with >100 spectra in the library) of unique GNPS mass spectral molecules tiled by ClassyFire chemical compound classes.

biological and chemical annotations on human metabolites and proteins/genes from multiple sources. Users can interact with RaMP-DB to perform batch queries and enrichment analyses through a web application, APIs, and an R package.

We note two major opportunities for maximizing the utility of these knowledge sources. First, there are varying degrees of structural resolution represented in metabolite annotations, such that mapping of analytes across data resources or annotation types (e.g., ontologies, diseases) is not always one-to-one. RefMet [66] aims to address this issue by mapping metabolite names and IDs to a common lowest denominator structure. Second, large-scale efforts, such as the Metabolomics Standards Initiative [67] and the Lipidomics Standards Initiative [68], are making strides to create ontologies for standardized reporting of metabolomics data. However, there is currently no widely used and mature ontology for the metabolomics community, which leads to the need for extensive manual curation to match database IDs across resources.

Pathway and chemical class analysis

It is widely acknowledged that molecular biology is well described as a network of interacting molecules, where molecules in one neighborhood belong to a 'pathway'. Such pathways may correspond to biological functions (e.g., glycolysis), chemical classes (e.g., triglycerides) or other meaningful categories. Metabolomic data can be interpreted by finding which pathways are enriched for differentially abundant metabolites [69]. Many methods, such as over-representation analysis, have been developed for transcriptomic data, and a cautious approach should be taken when applying them to metabolomics. The major difficulties relate to the low coverage and inexact structural identification present in most metabolomics datasets [70]. Metabolites can also be interpreted in light of their enriched chemical classes, as supported by ChemRICH and RaMP-DB, which, to some extent, is less affected by these drawbacks primarily because all metabolites can be mapped to a chemical class [71]. A promising alternative to traditional pathway methods is Pathway-Activity Likelihood Analysis (PUMA), a machine learning model that infers the likelihood of pathway activities and the activity of metabolites within pathways [72]. Another new development is the application of single sample pathway approaches, which allow transformation of metabolomic data into a 'pathway space' [73], facilitating a wide array of downstream analyses.

Data integration

There are many reasons to integrate data, and clarifying the intended purpose will dictate which computational methods are appropriate. For example, the focus could be to improve predictions or to find new connections between data sources. Integrative models should not only be able to match and harmonize data from different sources, but they should develop a unified view of a biological system, allowing the different sources to contribute according to the information present [74–76].

Integration within metabolomics

Metabolomics experiments often generate multiple data blocks, for example when more than one assay (e.g., ionization mode or chromatographic method) is used on the same samples. Within-metabolomics integration can yield chemical information, such as connecting signals from the same metabolite in different assays, which might help with annotating unknown metabolites and interpretation [51]. Further, biological insights can be made by integrating metabolomics data recorded on different sample types from the same individual [77]. A further dimension to within-metabolomics integration is the combination of multiple sample sets measured with the same assay, for example when several human cohorts are combined in large epidemiological studies. For untargeted metabolomics, matching metabolites, including unannotated features, between datasets is required [78], as is efficiently running error-free models for multi-cohort meta-analyses, as supported by the COMETS Analytics software [79].

Metabolomics-focused multi-omic data integration

Methods for combining metabolomics with other omics modalities enable discovery of multi-omic biomarkers, linking genes to metabolites for pathway prediction, or elucidating interactions between multiple omics layers. These wide-ranging aims and diverse data types lead to a range of challenges, from sourcing and matching datasets, to developing statistical and machine learning models that accommodate high degrees of heterogeneity.

Multi-omic datasets are typically available in omic-specific repositories, for metabolomics, including Metabolights, Metabobank, and the National Metabolomics Data Repository (NMDR, formerly Metabolomics Workbench), making it difficult to discover datasets amenable to integration. Further, inconsistent use of standard ontologies across resources make it time- and resource-intensive to identify publicly available multi-omic datasets [6,80]. Recently, the Paired Omics Data Platform addressed some of these challenges for natural products researchers by using a controlled vocabulary to make links between biomolecules and omics data from public repositories [81].

We also highlight a significant strand of recent work focused on modeling of metabolic flux to integrate data in a systems biology context [82]. For example, estimation of whether metabolic or transcriptional effects are dominant in controlling a given flux were made by applying a constraint based stoichiometric model that integrates metabolomic and gene expression data [83]. Neural network and deep learning approaches to integration are also gradually appearing in the literature, and recent applications have uncovered novel metabolite-microbe connections and biomarkers of cervicovaginal phenotypes [84]. However, interpretability of these highly complex models is of paramount importance if they are to yield actionable biological insights [85].

Augmenting these models can shed light on metabolite identities and uncover unknown metabolic pathways. Creating extended metabolic models (EMMs) [86], based on metabolic models consisting of known metabolites and enzymes for the sample under study and cataloging putative metabolic products in databases [87], is a promising approach for identifying putative cellular products. Underlying this idea is the ability to predict the promiscuity of enzymes on a large number of substrates [88] and the feasibility of the relevant biochemical reactions [89]. Using known metabolism to interpret metabolomics data is a promising direction that deserves further research.

Benchmarking tools and datasets

As computational metabolomics methods continue to develop along with analytical instrumentation, benchmarking datasets for testing new tools and methods are essential. The community has developed some key standards to facilitate the generation of benchmarking datasets. These include the NIST 1950 samples [90], COMETS reference samples [91], and efforts from mQACC and MetQual [92] that evaluate comparability of metabolite measurements across different labs. Further, NMDR and CASMI [93] provide datasets that can be used to develop algorithms. Yet, there are still challenges to finding appropriate benchmarking datasets, and data used for developing new algorithms are not consistently made available. Possible solutions for increasing the findability and utility of benchmarking data include tagging benchmarking datasets in repositories and clearly defining use cases. The SODA (SOftware and DAta exchange) interest group from the Metabolomics Association of North America provides a forum for exchanging information on current software, datasets, and data analysis results [94]. An important point is the potential bias training datasets may introduce into bioinformatics tools. Recent developments in the mass spectral prediction software CFM-ID 4.0 [95] showed how spectral prediction performance of specific chemical classes were improved based on available training data and fragmentation rules. Notably, a recent benchmarking study demonstrated clear performance differences of CFM-ID for various chemical classes [96]. Application of best practices for tool development and benchmarking is also key in ensuring appropriate usage, reliability, and long term maintenance of tools [15,97].

Conclusions

Computational metabolomics will continue to grow as metabolomics is increasingly applied across the scientific community and new instrumentation modalities are developed. In the last two years, there have been highly impactful advances in machine learning techniques for annotating metabolites, visualization, data interpretation and integration, as well as benchmarking software. Other areas of metabolomic technology such as single cell techniques and imaging mass spectrometry are rapidly expanding and will no doubt lead to new computational challenges. We also observe increasing awareness of best practice and standardization in how data and computational methods/tools are developed, reported and made available [15]. Overall, it is important to emphasize the value of continued communication and collaborations among scientists working on data analysis, knowledge sources, and tools, and most importantly among biologists, chemists, clinicians and informaticians. This interdisciplinary conversation will bring clarity in which methods and tools should be developed and, importantly, how they should be used correctly.

Funding

This work was supported in part by the Intramural Research Program of the National Center for Advancing Translational Sciences, National Institutes of Health (1ZICTR000410-02), and by the National Institute of General Medical Sciences, National Institutes of Health (1GM132391). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. TE acknowledges support from UKRI BBSRC grants BB/T007974/1 and BB/W002345/1. CB acknowledges support from NSF grant 2117943, NIFA grant 2021-67019-33726 and NIH Common Fund grant 5U01CA235507-03-s. SH acknowledges support from NIGMS grant R01GM132391.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:
J.J.J.v.d.H. is a member of the Scientific Advisory Board of NAICONS Srl., Milano, Italy.

Data availability

No data was used for the research described in the article.

Acknowledgements

The authors thank all Dagstuhl 2022 participants for contributing to vivid discussions on the above topics that served as input for this perspective. We also thank the Schloss Dagstuhl - Leibniz Center for Informatics for hosting the "Computational Metabolomics: From Spectra to Knowledge" Seminar and providing an international forum to advance informatics research. The authors also thank Prof. Florian Huber for assistance with the tmap and treemap visualization figure.

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

1. ** Misra BB: **New software tools, databases, and resources in metabolomics: updates from 2020**. *Metabolomics* 2021, **17**:1–24.
This paper provides a comprehensive list of tools and resources for computational metabolomics.

2. Beniddir MA, Kang K Bin, Genta-Jouve G, Huber F, Rogers S, Van Der Hooft JJJ: **Advances in decomposing complex metabolite mixtures using substructure- and network-based computational metabolomics approaches**. *Nat Prod Rep* 2021, **38**:1967–1993.

3. ** Li S (Ed). *Computational methods and data analysis for metabolomics*. Humana Press Inc.; 2020.
Up-to-date guide for metabolomics data analysis, targeting both experienced scientists and students.

4. Jarmusch SA, Van Der Hooft JJJ, Dorrestein PC, Jarmusch AK: **Advancements in capturing and mining mass spectrometry data are transforming natural products research**. *Nat Prod Rep* 2021, **38**:2066–2082.

5. Dekermanjian J, Labeikovsky W, Ghosh D, Kechris K: **MSCAT : a machine learning assisted catalog of metabolomics software tools**. *Metabolites* 2021, **11**:678.

6. ** Yu CT, Chao BN, Barajas R, Haznadar M, Maruvada P, Nicastro HL, Ross SA, Verma M, Rogers S, Zanetti KA: **An**

evaluation of the National Institutes of Health grants portfolio: identifying opportunities and challenges for multi-omics research that leverage metabolomics data. *Metabolomics* 2022, **18**:29.

This paper provides an excellent overview of NIH-funded Metabolomics driven multi-omics studies, highlighting challenges and needs.

7. Schmid R, Heuckeroth S, Korf A, Smirnov A, Myers O, Dyrlund TS, Bushuiev R, Murray KJ, Hoffmann N, Lu M, Sarvepalli A, Zhang Z, Fleischauer M, Dührkop K, Wesner M, Hoogstra SJ, Rudt E, Mokshyna O, Brungs C, Ponomarov K, Mutabdžija L, Damiani T, Pudney CJ, Earll M, Helmer PO, Fallon TR, Schulze T, Rivas-Ubach A, Bilbao A, Richter H, Nothias L-F, Wang M, Orešič M, Weng J-K, Böcker S, Jeibmann A, Hayen H, Karst U, Dorrestein PC, Petras D, Du X, Pluskal T: **Integrative analysis of multimodal mass spectrometry data in MZmine 3**. *Nat Biotechnol* 2023, https://doi.org/10.1038/s41587-023-01690-2.

8. Tsugawa H, Matsuzawa Y, Tada I, Takahashi M, Pedrosa D, Cajka T, Uchino H, Wohlgemuth G: MS-DIAL 5. http://prime.psc.riken.jp/compms/msdial/main.html. (accessed: February 6, 2023).

9. Domingo-Almenara X, Siuzdak G: **Metabolomics data processing using XCMS**. In *Computational methods and data analysis for metabolomics*. Edited by Li S, Humana Press; 2020.

10. Delabriere A, Warmer P, Brennsteiner V, Zamboni N: **SLAW: a scalable and self-optimizing processing workflow for untargeted LC-MS**. *Anal Chem* 2021, **93**:15024–15032.

11. Giné R, Capellades J, Badia JM, Vughs D, Schwaiger-Haber M, Alexandrov T, Vinaixa M, Brunner AM, Patti GJ, Yanes O: **HERMES: a molecular-formula-oriented method to target the metabolome**. *Nat Methods* 2021, **18**:1370–1376.

12. Rainer J, Vicini A, Salzer L, Stanstrup J, Badia JM, Neumann S, Stravs MA, Hernandes VV, Gatto L, Gibb S, *et al.*: **A modular and expandable ecosystem for metabolomics data annotation in R**. *Metabolites* 2022, **12**:173.

13. Bittremieux W, Levitsky L, Pilz M, Sachsenberg T, Huber F, Wang M, Dorrestein PC: **Unified and standardized mass spectrometry data processing in Python using spectrum_utils**. *J Proteome Res* 2023, **22**:625–631.

14. Riquelme G, Zabalegui N, Marchi P, Jones CM, Monge ME: **A python-based pipeline for preprocessing lc–ms data for untargeted metabolomics workflows**. *Metabolites* 2020, **10**: 1–14.

15. Chang H, Colby SM, Du X, Gomez JD, Helf MJ, Kechris K, Kirkpatrick CR, Li S, Patti GJ, Renslow RS, *et al.*: **A practical guide to metabolomics software development**. *Anal Chem* 2021, **93**:1912–1923.

16. Sun Q, Fan TWM, Lane AN, Higashi RM: **Applications of chromatography-ultra high-resolution MS for stable isotope-resolved metabolomics (SIRM) reconstruction of metabolic networks**. *TrAC, Trends Anal Chem* 2020, **123**, 115676.

17. Wandy J, Davies V, McBride R, Weidt S, Rogers S, Daly R: **ViMMS 2.0: a framework to develop, test and optimise fragmentation strategies in LC-MS metabolomics**. *J Open Source Softw* 2022, **7**:3990.

18. Wandy J, McBride R, Rogers S, Terzis N, Weidt S, van der Hooft JJJ, Bryson K, Daly R, Davies V: **Simulated-to-real benchmarking of acquisition methods in untargeted metabolomics**. *Front Mol Biosc* 2023, **10**, https://doi.org/10.3389/fmolb.2023.1130781.

19. Aron AT, Gentry EC, McPhail KL, Nothias LF, Nothias-Esposito M, Bouslimani A, Petras D, Gauglitz JM, Sikora N, Vargas F, *et al.*: **Reproducible molecular networking of untargeted mass spectrometry data using GNPS**. *Nat Protoc* 2020, **15**:1954–1991.

20. Wang JH, Choong WK, Chen CT, Sung TY: **Calibr improves spectral library search for spectrum-centric analysis of data independent acquisition proteomics**. *Sci Rep* 2022, **12**:1–13.

21. Messner CB, Demichev V, Bloomfield N, Yu JSL, White M, Kreidl M, Egger AS, Freiwald A, Ivosev G, Wasim F, *et al.*: **Ultrafast proteomics with scanning SWATH**. *Nat Biotechnol* 2021, **39**:846–854.

22. Alka O, Shanthamoorthy P, Witting M, Kleigrewe K, Kohlbacher O, Röst HL: **DIAMetAlyzer allows automated false-discovery rate-controlled analysis for data-independent acquisition in metabolomics**. *Nat Commun* 2022, **13**:1347.

23. Stancliffe E, Schwaiger-Haber M, Sindelar M, Patti GJ: **DecoID improves identification rates in metabolomics through database-assisted MS/MS deconvolution**. *Nat Methods* 2021, **18**:779–787.

24. Graça G, Cai Y, Lau CHE, Vorkas PA, Lewis MR, Want EJ, Herrington D, Ebbels TMD: **Automated annotation of untargeted all-ion fragmentation LC-MS metabolomics data with MetaboAnnotatoR**. *Anal Chem* 2022, **94**:3446–3455.

25. Guo J, Koopmeiners JS, Walmsley SJ, Villalta PW, Yao L, Murugan P, Tejpaul R, Weight CJ, Turesky RJ: **The cooked meat carcinogen 2-Amino-1-methyl-6-phenylimidazo[4,5- b]pyridine hair dosimeter, DNA adductomics discovery, and associations with prostate cancer pathology biomarkers**. *Chem Res Toxicol* 2022, **35**:703–730.

26. Murray KJ, Carlson ES, Stornetta A, Balskus EP, Villalta PW, Balbo S: **Extension of diagnostic fragmentation filtering for automated discovery in DNA adductomics**. *Anal Chem* 2021, **93**:5754–5762.

27. Smith JW, O'Meally RN, Ng DK, Chen JG, Kensler TW, Cole RN, Groopman JD: **Biomonitoring of ambient outdoor air pollutant exposure in humans using targeted serum albumin adductomics**. *Chem Res Toxicol* 2021, **34**:1183–1196.

28. Funk WE, Montgomery N, Bae Y, Chen J, Chow T, Martinez MP, Lurmann F, Eckel SP, McConnell R, Xiang AH: **Human serum albumin Cys34 adducts in newborn dried blood spots: associations with air pollution exposure during pregnancy**. *Front Public Health* 2021, **9**:1–9.

29. Bittremieux W, Wang M, Dorrestein PC: **The critical role that spectral libraries play in capturing the metabolomics community knowledge**. *Metabolomics* 2022, **18**:1–16.

30. Zhu H, Liu L, Hassoun S: **Using graph neural networks for mass spectrometry prediction**. *arXiv Prepr arXiv201004661* 2020, https://doi.org/10.48550/arXiv.2010.04661.

31. Young A, Wang B, Röst H: **MassFormer: tandem mass spectrum prediction with graph transformers**. *arXiv Prepr arXiv211104824* 2021, https://doi.org/10.48550/arXiv.2111.04824.

32. Li X, Zhu H, Liu L, Hassoun S: **Ensemble spectral prediction (ESP) model for metabolite annotation**. *arXiv Prepr arXiv220313783* 2022, https://doi.org/10.48550/arXiv.2203.13783.

33. Shrivastava AD, Swainston N, Samanta S, Roberts I, Wright Muelas M, Kell DB, MassGenie: **A transformer-based deep learning method for identifying small molecules from their mass spectra**. *Biomolecules* 2021, **11**:1793.

34. Stravs MA, Dührkop K, Böcker S, Zamboni N: **MSNovelist: de novo structure generation from mass spectra**. *Nat Methods* 2022, **19**:1–6.

35. * Huber F, Ridder L, Verhoeven S, Spaaks JH, Diblen F, Rogers S, *et al.*: **Spec2Vec: Improved mass spectral similarity scoring through learning of structural relationships**. *PLoS Comput Biol* 2021, **17**(2), e1008724.

This paper introduces the first proposed machine learning-based mass spectral similarity score that shows substantial improvements in mass spectral library matching performance compared to widely-used cosine-based scores.

36. Dührkop K, Nothias LF, Fleischauer M, Reher R, Ludwig M, Hoffmann MA, Petras D, Gerwick WH, Rousu J, Dorrestein PC, *et al.*: **Systematic classification of unknown metabolites using high-resolution fragmentation mass spectra**. *Nat Biotechnol* 2021, **39**:462–471.

37. Kim HW, Wang M, Leber CA, Nothias L-F, Reher R, Kang K Bin, van der Hooft JJJ, Dorrestein PC, Gerwick WH, Cottrell GW, NPClassifier: **A deep neural network-based structural classification tool for natural products**. *J Nat Prod* 2021, **84**: 2795–2807.

38. Bittremieux W, Avalon NE, Thomas SP, Kakhkhorov SA, Gauglitz JM, Gerwick WH, Jarmusch AK, Rima F: **Open access**

repository-scale propagated nearest neighbor suspect spectral library for untargeted metabolomics. *BioRxiv Prepr* 2022.

39. * Schmid R, Petras D, Nothias LF, Wang M, Aron AT, Jagels A, Tsugawa H, Rainer J, Garcia-Aloy M, Dührkop K, *et al.*: **Ion identity molecular networking for mass spectrometry-based metabolomics in the GNPS environment**. *Nat Commun* 2021, **12**:1–12.

This paper presents a key development in the molecular networking technique overcoming the problem of multiple ion species with different fragmentation behavior.

40. Huber F, van der Burg S, van der Hooft JJJ, Ridder L: **MS2DeepScore: a novel deep learning similarity measure to compare tandem mass spectra**. *J Cheminf* 2021, **13**:84.

41. * Jarmusch AK, Aron AT, Petras D, Phelan VV, Bittremieux W, Acharya DD, Ahmed MMA, Bauermeister A, Bertin MJ, Boudreau PD, *et al.*: **A universal language for finding mass spectrometry data patterns**. *bioRxiv* 2022.

This paper highlights a generalized, flexible, and scalable mass spectral query language, MassQL. Use of such a language enables querying and using mass spectral data across any type/manufacturer of mass spectrometers and thus across any data source with raw spectral data.

42. Bach E, Schymanski EL, Rousu J: **Joint structural annotation of small molecules using liquid chromatography retention order and tandem mass spectrometry data**. *Nat Mach Intell* 2022, **4**:1224–1237.

43. Wang M, Jarmusch AK, Vargas F, Aksenov AA, Gauglitz JM, Weldon K, Petras D, da Silva R, Quinn R, Melnik AV, *et al.*: **Mass spectrometry searches using MASST**. *Nat Biotechnol* 2020, **38**: 23–26.

44. Jarmusch AK, Wang M, Aceves CM, Advani RS, Aguirre S, Aksenov AA, Aleti G, Aron AT, Bauermeister A, Bolleddu S, *et al.*: **ReDU: a framework to find and reanalyze public mass spectrometry data**. *Nat Methods* 2020, **17**:901–904.

45. Medina-Franco JL, Sánchez-Cruz N, López-López E, Díaz-Eufracio BI: **Progress on open chemoinformatic tools for expanding and exploring the chemical space**. *J Comput Aided Mol Des* 2022, **36**:341–354.

46. Cihan Sorkun M, Mullaj D, Koelman JMVA, Er S: **ChemPlot, a Python library for chemical space visualization**. *Chem Mater* 2022, **2**, e202200005.

47. Probst D, Reymond J-L: **Visualization of very large high-dimensional data sets as minimum spanning trees**. *J Cheminf* 2020, **12**:12.

48. * Nothias L-F, Petras D, Schmid R, Dührkop K, Rainer J, Sarvepalli A, Protsyuk I, Ernst M, Tsugawa H, Fleischauer M, *et al.*: **Feature-based molecular networking in the GNPS analysis environment**. *Nat Methods* 2020, **17**:905–908.

This paper introduces a workflow that combines qualitative and quantitative information from metabolomics profiles that has already been demonstrated to generate novel hypotheses and results across many different scientific disciplines.

49. ** Amara A, Frainay C, Jourdan F, Naake T, Neumann S, Novoa-del-toro EM, Salek RM, Salzer L, Scharfenberg S: **Networks and graphs discovery in metabolomics data analysis and interpretation**. *Front Mol Biosci* 2022, **9**:1–15.

This paper formalizes the nature of and language used to describe networks used in metabolomics, and highlights recent advances in using these networks, individually or combined, in research efforts.

50. Redžepović I, Furtula B: **Chemical similarity of molecules with physiological response**. *Mol Divers* 2022, https://doi.org/10.1007/s11030-022-10514-5.

51. Barupal DK, Mahajan P, Fakouri-Baygi S, Wright RO, Arora M, Teitelbaum SL: **CCDB: a database for exploring inter-chemical correlations in metabolomics and exposomics datasets**. *Environ Int* 2022, **164**, 107240.

52. Silverman EK, Schmidt HHHW, Anastasiadou E, Altucci L, Angelini M, Badimon L, Balligand J-L, Benincasa G, Capasso G, Conte F, *et al.*: **Molecular networks in network medicine: development and applications**. *Wiley Interdiscip Rev Syst Biol Med* 2020, **12**:e1489.

53. * Rutz A, Sorokina M, Galgonek J, Mietchen D, Willighagen E, Gaudry A, Graham JG, Stephan R, Page R, Vondrášek J, *et al.*: **The LOTUS initiative for open knowledge management in natural products research**. *Elife* 2022, **11**, e70780.

This paper is the first example of FAIR data sharing of metabolite structural and taxonomic information relevant for natural product discovery with an automated curation pipeline and integration into WikiData.

54. van Santen JA, Poynton EF, Iskakova D, McMann E, Alsup TA, Clark TN, Fergusson CH, Fewer DP, Hughes AH, McCadden CA, *et al.*: **The Natural Products Atlas 2.0: a database of microbially-derived natural products**. *Nucleic Acids Res* 2022, **50**:D1317–D1323.

55. Schymanski EL, Kondić T, Neumann S, Thiessen PA, Zhang J, Bolton EE: **Empowering large chemical knowledge bases for exposomics: PubChemLite meets MetFrag**. *J Cheminf* 2021, **13**:1–15.

56. Wishart DS, Guo A, Oler E, Wang F, Anjum A, Peters H, Dizon R, Sayeeda Z, Tian S, Lee BL, *et al.*: **HMDB 5.0: the human metabolome database for 2022**. *Nucleic Acids Res* 2022, **50**: D622–D631.

57. Lim R, JJT Cabatbat, Martin TLP, Kim H, Kim S, Sung J, Ghim C-M, Kim P-J: **Large-scale metabolic interaction network of the mouse and human gut microbiota**. *Sci Data* 2020, **7**:1–8.

58. Delmas M, Filangi O, Paulhe N, Vinson F, Duperier C, Garrier W, Saunier P-E, Pitarch Y, Jourdan F, Giacomoni F, *et al.*: **FORUM: building a Knowledge Graph from public databases and scientific literature to extract associations between chemicals and diseases**. *Bioinformatics* 2021, **37**:3896–3904.

59. Yeung CS, Beck T, Posma JM: **MetaboListem and TABoLiSTM: two deep learning algorithms for metabolite named entity recognition**. *Metabolites* 2022, **12**:276.

60. Pang Z, Chong J, Zhou G, de Lima Morais DA, Chang L, Barrette M, Gauthier C, Jacques P-É, Li S, Xia J: **MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights**. *Nucleic Acids Res* 2021, https://doi.org/10.1093/nar/gkab382.

61. Kanehisa M, Sato Y, Kawashima M: **KEGG mapping tools for uncovering hidden features in biological data**. *Protein Sci* 2022, **31**:47–53.

62. Gillespie M, Jassal B, Stephan R, Milacic M, Rothfels K, Senff-Ribeiro A, Griss J, Sevilla C, Matthews L, Gong C, *et al.*: **The reactome pathway knowledgebase 2022**. *Nucleic Acids Res* 2022, **50**:D687–D692.

63. Martens M, Ammar A, Riutta A, Waagmeester A, Slenter DN, Hanspers K, Miller RA, Digles D, Lopes EN, Ehrhart F, *et al.*: **WikiPathways: connecting communities**. *Nucleic Acids Res* 2021, **49**:D613–D621.

64. Liebisch G, Fahy E, Aoki J, Dennis EA, Durand T, Ejsing CS, Fedorova M, Feussner I, Griffiths WJ, Köfeler H, *et al.*: **Update on LIPID MAPS classification, nomenclature, and shorthand notation for MS-derived lipid structures**. *J Lipid Res* 2020, **61**: 1539–1555.

65. * Braisted J, Patt A, Tindall C, Sheils T, Neyra J, Spencer K, Eicher T, Mathé EA: **RaMP-DB 2.0: a renovated knowledge-base for deriving biological and chemical insight from metabolites, proteins, and genes**. *Bioinformatics* 2023, **39**: btac726.

An open-source, comprehensive resource of chemical and biological annotations of human metabolites, proteins and genes. This paper highlights expansions in the data content and ways users can interact with it (e.g. queries, enrichment analyses, etc.).

66. Fahy E, Subramaniam S: **RefMet: a reference nomenclature for metabolomics**. *Nat Methods* 2020, **17**:1173–1174.

67. Haug K, Cochrane K, Nainala VC, Williams M, Chang J, Jayaseelan KV, O'Donovan C, MetaboLights: **A resource evolving in response to the needs of its scientific community**. *Nucleic Acids Res* 2020, **48**:D440–D444.

68. O'Donnell VB, Ekroos K, Liebisch G, Wakelam M: **Lipidomics: current state of the art in a fast moving field**. *Wiley Interdiscip Rev Syst Biol Med* 2020, **12**:e1466.

69. McLuskey K, Wandy J, Vincent I, Hooft JJJ va der, Rogers S, Burgess K, Daly R: **Ranking metabolite sets by their activity levels**. *Metabolites* 2021, **11**:1–15.

70. * Wieder C, Frainay C, Poupin N, Rodríguez-Mier P, Vinson F, Cooke J, Lai RPJ, Bundy JG, Jourdan F, Ebbels T: **Pathway analysis in metabolomics: recommendations for the use of over-representation analysis**. *PLoS Comput Biol* 2021, **17**, e1009105.

Over-representation analysis is the most widely used method for pathway analysis. This paper lays out the risks of using it in metabolomics and recommends best practice.

71. Barupal DK, Fiehn O: **Chemical Similarity Enrichment Analysis (ChemRICH) as alternative to biochemical pathway mapping for metabolomic datasets**. *Sci Rep* 2017, **7**:1–11.

72. * Hosseini R, Hassanpour N, Liu LP, Hassoun S: **Pathway-activity likelihood analysis and metabolite annotation for untargeted metabolomics using probabilistic modeling**. *Metabolites* 2020, **10**:183.

This paper introduces a promising machine-learning model to infer the likelihood of pathway activities, thus providing an alternative to traditional statistics-based pathway activity metrics.

73. Wieder C, Lai RPJ, Ebbels T: **Single sample pathway analysis in metabolomics: performance evaluation and application**. *BMC Bioinf* 2022, **23**:481.

74. Jendoubi T: **Approaches to integrating metabolomics and multi-omics data: a primer**. *Metabolites* 2021, **11**:184.

75. Eicher T, Kinnebrew G, Patt A, Spencer K, Ying K, Ma Q, Machiraju R, Math EA: **Metabolomics and multi-omics integration: a survey of computational methods and resources**. *Metabolites* 2020, **10**:202.

76. Worheide MA, Krumsiek J, Kastenmüller G, Arnold M: **Multi-omics integration in biomedical research - a metabolomics-centric review**. *Anal Chim Acta* 2021, **1141**:144–162.

77. Torell F, Skotare T, Trygg J: **Application of multiblock analysis on small metabolomic multi-tissue dataset**. *Metabolites* 2020, **10**:1–14.

78. Climaco Pinto R, Karaman I, Lewis MR, Hällqvist J, Kaluarachchi M, Graça G, Chekmeneva E, Durainayagam B, Ghanbari M, Ikram MA, *et al.*: **Finding correspondence between metabolomic features in untargeted liquid chromatography–mass spectrometry metabolomics datasets**. *Anal Chem* 2022, **94**:5493–5503.

79. Temprosa M, Moore SC, Zanetti KA, Appel N, Ruggieri D, Mazzilli KM, Chen K, Kelly RS, Lasky-Su JA, Loftfield E, *et al.*: **COMETS Analytics: an online tool for analyzing and meta-analyzing metabolomics data in large research consortia**. *Am J Epidemiol* 2022, **191**:147–158.

80. Tarazona S, Arzalluz-Luque A, Conesa A: **Undisclosed, unmet and neglected challenges in multi-omics studies**. *Nat Comput Sci* 2021, **1**:395–402.

81. Schorn MA, Verhoeven S, Ridder L, Huber F, Acharya DD, Aksenov AA, Aleti G, Moghaddam JA, Aron AT, Aziz S, *et al.*: **A community resource for paired genomic and metabolomic data mining**. *Nat Chem Biol* 2021, **17**:363–368.

82. Volkova S, Matos MRA, Mattanovich M, de Mas IM: **Metabolic modelling as a framework for metabolomics data integration and analysis**. *Metabolites* 2020, **10**:1–27.

83. Di Filippo M, Pescini D, Galuzzi BG, Bonanomi M, Gaglio D, Mangano E, Consolandi C, Alberghina L, Vanoni M, Damiani C: **INTEGRATE: model-based multi-omics data integration to characterize multi-level metabolic regulation**. *PLoS Comput Biol* 2022, **18**, e1009337.

84. Bokulich NA, Łaniewski P, Adamov A, Chase DM, Caporaso JG, Herbst-Kralovetz MM: **Multi-omics data integration reveals metabolome as the top predictor of the cervicovaginal microenvironment**. *PLoS Comput Biol* 2022, **18**, e1009876.

85. * Le V, Quinn TP, Tran T, Venkatesh S: **Deep in the Bowel: highly interpretable neural encoder-decoder networks predict gut metabolites from gut microbiome**. *BMC Genom* 2020, **21**:256.

Neural network-based methods are often difficult to interpret. This paper presents a novel neural model which emphasizes interpretability to integrate multi-omics data.

86. Hassanpour N, Alden N, Menon R, Jayaraman A, Lee K, Hassoun S: **Biological filtering and substrate promiscuity prediction for annotating untargeted metabolomics**. *Metabolites* 2020, **10**:160.

87. Strutz J, Shebek KM, Broadbelt LJ, Tyo KEJ: **MINE 2.0: enhanced biochemical coverage for peak identification in untargeted metabolomics**. *Bioinformatics* 2022, **38**:3484–3487.

88. Visani GM, Hughes MC, Hassoun S: **Enzyme promiscuity prediction using hierarchy-informed multi-label classification**. *Bioinformatics* 2021, **37**:2017–2024.

89. Kim Y, Ryu JY, Kim HU, Jang WD, Lee SY: **A deep learning approach to evaluate the feasibility of enzymatic reactions generated by retrobiosynthesis**. *Biotechnol J* 2021, **16**, 2000605.

90. Simón-Manso Y, Lowenthal MS, Kilpatrick LE, Sampson ML, Telu KH, Rudnick PA, Mallard WG, Bearden DW, Schock TB, Tchekhovskoi DV, *et al.*: **Metabolite profiling of a NIST standard reference material for human plasma (SRM 1950): GC-MS, LC-MS, NMR, and clinical laboratory analyses, libraries, and web-based resources**. *Anal Chem* 2013, **85**: 11725–11731.

91. Zanetti KA: **Building infrastructure at the National Cancer Institute to support metabolomic analyses in epidemiological studies**. *Metabolomics* 2021, **17**:1–4.

92. Lippa KA, Aristizabal-Henao JJ, Beger RD, Bowden JA, Broeckling C, Beecher C, Clay Davis W, Dunn WB, Flores R, Goodacre R, *et al.*: **Reference materials for MS-based untargeted metabolomics and lipidomics: a review by the metabolomics quality assurance and quality control consortium (mQACC)**. *Metabolomics* 2022, **18**:1–29.

93. Schymanski E, Neumann S: **The critical assessment of small molecule identification (CASMI): challenges and solutions**. *Metabolites* 2013, **3**:517–538.

94. Mathe EA, Mak T, Hitchcock D: Metabolomics Association of North America SOftware DAta Exchange. https://sites.google.com/metabolomicsna.org/soda/home?pli=1. (accessed February 6, 2023)..

95. Wang F, Liigand J, Tian S, Arndt D, Greiner R, Wishart DS: **CFM-ID 4.0: more accurate ESI-MS/MS spectral prediction and compound identification**. *Anal Chem* 2021, **93**: 11692–11700.

96. Bremer PL, Vaniya A, Kind T, Wang S, Fiehn O: **How well can we predict mass spectra from structures? Benchmarking competitive fragmentation modeling for metabolite identification on untrained tandem mass spectra**. *J Chem Inf Model* 2022, **62**:4049–4056.

97. de Jonge NF, Mildau K, Meijer D, Bueschl C, Huber F, Van Der Hooft JJJ: **Good practices and recommendations for using and benchmarking computational metabolomics metabolite annotation tools**. *Metabolomics* 2022, **18**:103.