MetaboAnalyst 6.0: towards a unified platform for metabolomics data processing, analysis and interpretation

Zhiqiang Pang[1], Yao Lu[2], Guangyan Zhou[1], Fiona Hui[1], Lei Xu[1], Charles Viau[1], Aliya F. Spigelman[3], Patrick E. MacDonald[3], David S. Wishart[4], Shuzhao Li[5,6] and Jianguo Xia ⬤[1,2,*]

[1]Institute of Parasitology, McGill University, Sainte-Anne-de-Bellevue, Quebec, Canada
[2]Department of Microbiology and Immunology, McGill University, Montreal, Quebec, Canada
[3]Department of Pharmacology and Alberta Diabetes Institute, University of Alberta, Edmonton, Alberta, Canada
[4]Departments of Biological Sciences and Computing Science, University of Alberta, Edmonton, Alberta, Canada
[5]The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA
[6]University of Connecticut School of Medicine, Farmington, CT, USA

*To whom correspondence should be addressed. Tel: +1 514 398 8668; Fax: +1 514 398 7857; Email: jeff.xia@mcgill.ca

Abstract

We introduce MetaboAnalyst version 6.0 as a unified platform for processing, analyzing, and interpreting data from targeted as well as untargeted metabolomics studies using liquid chromatography - mass spectrometry (LC–MS). The two main objectives in developing version 6.0 are to support tandem MS (MS2) data processing and annotation, as well as to support the analysis of data from exposomics studies and related experiments. Key features of MetaboAnalyst 6.0 include: (i) a significantly enhanced Spectra Processing module with support for MS2 data and the asari algorithm; (ii) a MS2 Peak Annotation module based on comprehensive MS2 reference databases with fragment-level annotation; (iii) a new Statistical Analysis module dedicated for handling complex study design with multiple factors or phenotypic descriptors; (iv) a Causal Analysis module for estimating metabolite - phenotype causal relations based on two-sample Mendelian randomization, and (v) a Dose-Response Analysis module for benchmark dose calculations. In addition, we have also improved MetaboAnalyst's visualization functions, updated its compound database and metabolite sets, and significantly expanded its pathway analysis support to around 130 species. MetaboAnalyst 6.0 is freely available at https://www.metaboanalyst.ca.

Graphical abstract
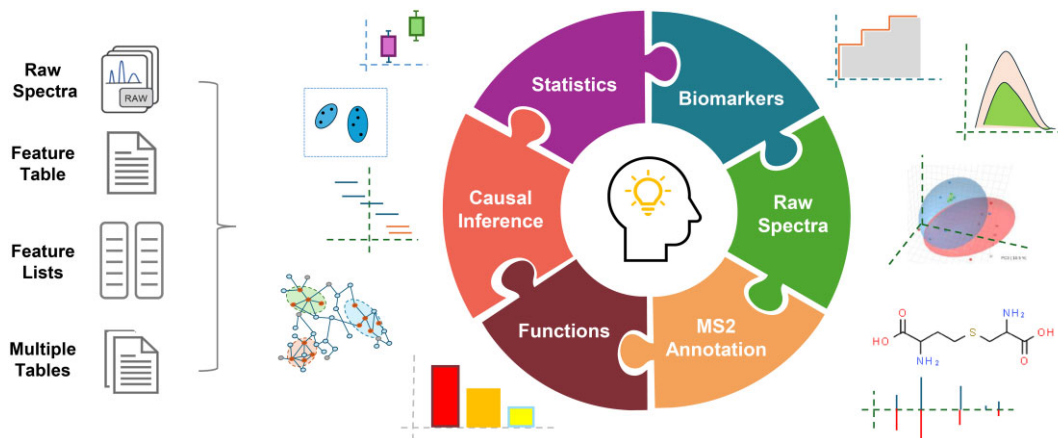
Introduction

Metabolomics involves the comprehensive study of all small molecules in a biological system. It has diverse applications ranging from basic biochemical research to clinical investigation of diseases, food safety assessment, environmental monitoring, etc. (1–5). User-friendly and easily accessible bioinformatics tools are essential to deal with the complex data produced from metabolomics studies. MetaboAnalyst is a user-friendly, web-based platform developed to provide compre-hensive support for metabolomics data analysis (6–10). The early versions (1.0–3.0) focused primarily on supporting statistical and functional analysis of targeted metabolomics data. Increasing support for untargeted metabolomics data from liquid chromatography–mass spectrometry (LC–MS) experiments have been gradually introduced in more recent versions of MetaboAnalyst. For instance, version 4.0 implemented a new module to support functional analysis directly from LC–MS peaks, while version 5.0 added an auto-optimized

LC–MS spectral processing module that works seamlessly with the functional analysis module. A detailed protocol on how to use different modules for comprehensive analysis of untargeted metabolomics data was published in 2022 (11). According to Google Analytics, the MetaboAnalyst web server has processed over 2 million jobs, including 33 000 spectral processing jobs over the past 12 months. Many of these jobs are associated with untargeted metabolomics and exposomics studies.

Untargeted metabolomics data generated from high-resolution LC–MS instruments are typically characterized by thousands of peaks with unknown chemical identities. To assist with compound identification, tandem MS (called MS/MS or MS2) spectra are often collected from pooled QC samples during the experiments (12). The two commonly used MS2 methods are data-dependent acquisition (DDA) and data-independent acquisition (DIA), with sequential window acquisition of all theoretical mass spectra (SWATH) being a promising special case of the latter. DDA data usually have clear associations between the precursor ions and the corresponding MS2 spectra, while DIA data generally require deconvolution of the MS2 data to reconstruct associations with their precursor ions (13). Incorporating MS2 processing and annotation into untargeted metabolomics workflows can greatly improve compound annotations and functional interpretation.

Exposomics is an emerging field centered on profiling the complete set of exposures individuals encounter across their lifespan, which often involves MS analysis of chemical mixtures traditionally rooted in toxicology and public health (4). Untargeted LC–MS based metabolomics is increasingly applied to exposomics and toxicology studies. Exposomics data from human cohorts is often associated with complex phenotypic data due to their observational nature. This requires more sophisticated data analysis and visualization methods that can take into consideration of multiple factors or covariates. Exposomics studies typically produce long lists of potential biomarkers that are significantly associated with phenotypes of interest. Identification of causal links from this large number of metabolite-phenotype relations is a natural next step. It has become possible recently with the availability of many metabolomic genome-wide association studies (mGWAS) that link metabolites and genotypes (14–16). By integrating mGWAS data with comparable GWAS data that associate genotypes with various phenotypes (17), we can now estimate causal relationships between a metabolite and a phenotype of interest through Mendelian randomization (MR) (18). Dose-response experiments are often performed to further quantify cause-and-effect relationships. The experiments are often conducted at multiple dose levels using *in vitro* assays or animal models to calculate dose-response curves for risk assessment of chemical exposures (19–21).

To address these emerging needs from both the metabolomics and exposomics communities, we have developed MetaboAnalyst version 6.0. This version includes many key features:

— A significantly enhanced spectra processing workflow with the addition of asari algorithm for LC–MS spectra processing (22), as well as support for MS2 (DDA or SWATH-DIA) data processing.

— A new module for MS2 spectral database searching for compound identification and results visualization.
— A new module for causal analysis between metabolites and phenotypes of interest based on two-sample MR (2SMR).
— A new module for dose-response analysis including dose-response curve fitting and benchmark dose (BMD) calculation.
— A new module for statistical analysis with complex metadata;
— A number of other important updates including: improved functional analysis of untargeted metabolomics data by integrating MS2-based compound identification; updated compound database, pathways and metabolite sets; as well as improved data visualization support across multiple modules.

MetaboAnalyst 6.0 is feely accessible at https://www.metaboanalyst.ca, with comprehensive documentations and updated tutorials. To better engage with our users, a dedicated user forum (https://omicsforum.ca) has been operational since May 2022. To dates, this forum contains >4000 posts on ∼700 topics related to different aspects of using MetaboAnalyst.

Overall design and workflow of MetaboAnalyst 6.0

MetaboAnalyst 6.0 accepts a total of five different data types across various modules encompassing spectra processing, statistical analysis, functional analysis, meta-analysis, and integration with other omics data. Once the data are uploaded, all analysis steps are conducted within a consistent framework including data integrity checks, parameter customization, and results visualization (Figure 1). Some of the key features in MetaboAnalyst 6.0 are described below.

Supporting asari and MS2 spectra in LC–MS spectra processing workflow

LC–MS spectra processing remains an active research topic in the field of untargeted metabolomics. Many powerful tools have been developed over time, including XCMS (23), MZmine (24), MS-DIAL (13) and asari (22). In addition to using different peak detection algorithms, most tools require manual parameter tuning to ensure good results. Such practice often leads to results that vary significantly (25). To mitigate this issue, MetaboAnalyst 5.0 introduced an auto-optimized LC–MS processing pipeline to minimize the parameter-related effects (10,26). The asari software has introduced a set of quality metrics, concepts of mass tracks and composite mass tracks and new algorithmic design to minimize errors in feature correspondence. It requires minimal parameter tuning while achieving much faster computational performance (22). The asari algorithm is now available in the LC–MS spectra processing options, alongside the traditional approaches.

MS2 spectra processing and metabolite identification are important components of untargeted metabolomics. It is now recognized that MS2 spectral deconvolution is necessary to achieve high-quality compound identification results for both DDA and SWATH-DIA data (27–29). MetaboAnalyst 6.0 offers an efficient, auto-optimized pipeline for MS2
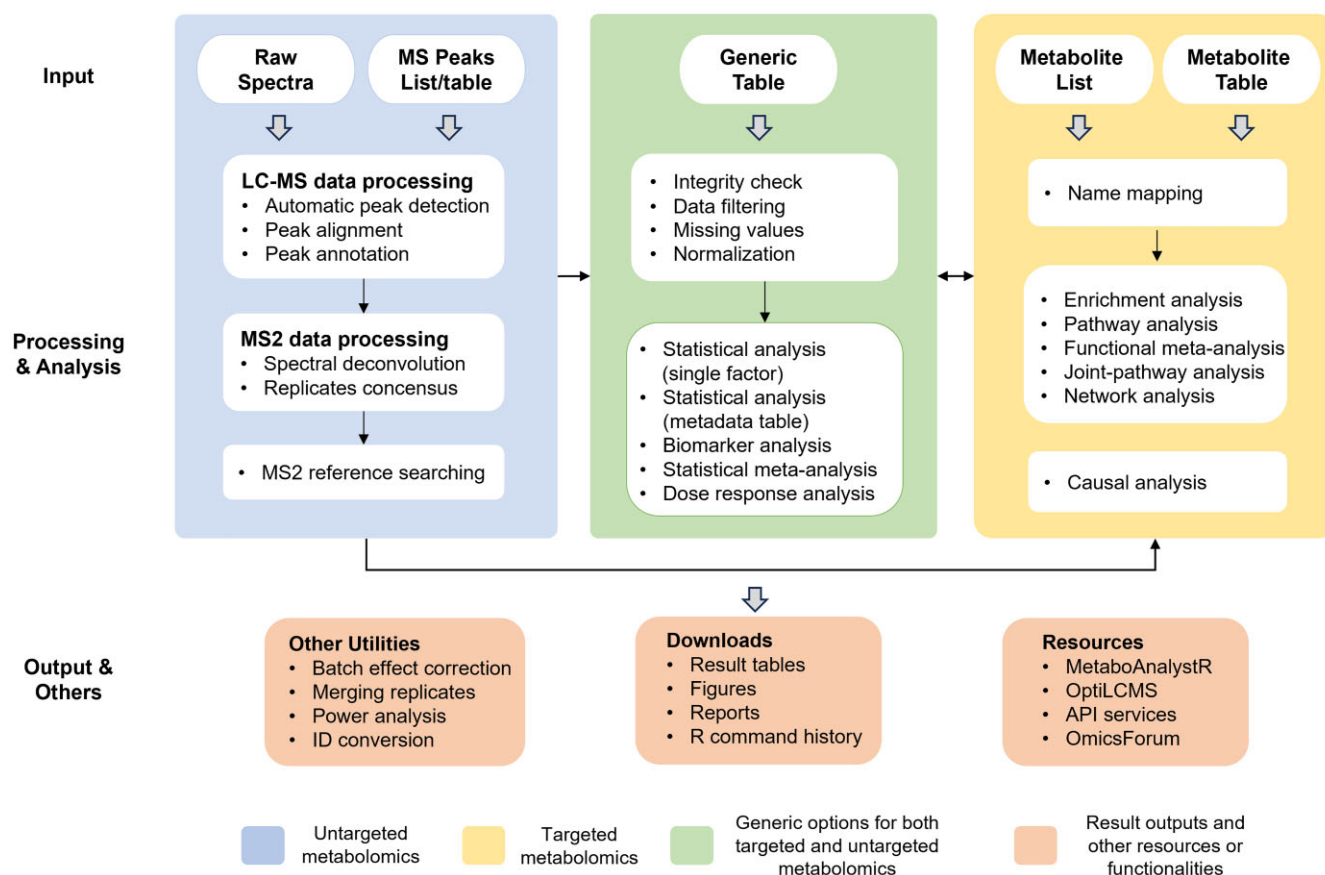
Figure 1. MetaboAnalyst 6.0 workflow for targeted and untargeted metabolomics data. Multiple data input types are accepted. Untargeted metabolomics inputs require extra steps for spectra processing and peak annotation. The result table can be used for statistical and functional analysis within a consistent workflow in the same manner as for targeted metabolomics data.

spectral deconvolution. The DDA data deconvolution method is derived from the DecoID algorithm (28), which employs a database-dependent regression model to deconvolve contaminated spectra. The SWATH-DIA data deconvolution algorithm is based on the DecoMetDIA method (29), with the core algorithm re-implemented using a Rcpp/C++ framework to achieve high performance. When MS2 spectra replicates are provided, an extra step will be performed to generate consensus spectra across replicates. The consensus spectra are searched against MetaboAnalyst's curated MS2 reference databases for compound identification based on dot product (28) or spectral entropy (30) similarity scores. The complete pipelines for DDA and SWATH-DIA are available from the *Spectra Processing [LC–MS w/wo MS2]* module.

Raw spectra must be saved in common open formats and uploaded individually as separate zip files. LC–MS spectra data is mandatory, while MS2 is optional. Upon data uploading, MetaboAnalyst 6.0 first validates the status of the MS files. For SWATH-DIA data, the SWATH window design is automatically extracted from the spectra. If the related information is missing, users will be prompted to manually enter the window design. On the parameters setting page, users can choose the auto-optimized *centWave* algorithm (26) or the *asari* algorithm for LC–MS data processing. If MS2 data is included, spectra deconvolution, consensus, and database searching will be performed using the identified MS features as target list. Once the spectra processing is complete, users can explore both MS and MS2 data processing results

(Figure 2A-B) and download the files or directly go to the *Functional Analysis* module.

MS2 peak annotation

MS2 data could be acquired independently from MS data acquisition. To accommodate this scenario and offer compatibility with MS2 spectra results from other popular tools such as MS-DIAL, we have added a *Peak Annotation [MS2-DDA/DIA]* module to allow users to directly upload MS2 spectra for database searching. Users can enter a single MS2 spectrum or upload an MSP or MGF file containing multiple MS2 spectra. For single spectrum searching, users must specify the m/z value of the precursor ion. However, for batch searching based on an MSP file, users do not need to specify the precursors' m/z values. To ensure timely completion of database searching, the public server processes only 20 spectra for each submission (the first 20 spectra by default). Users can manually specify spectra for searching. After conducting this pilot analysis with 20 spectra, users can download the R command history and use our MetaboAnalystR package to annotate all MS2 spectra (26).

Multiple databases are available for compound identification. Database searching can be performed based on regular reference MS2 spectra and/or their corresponding neutral loss spectra. The results are visually summarized as mirror plots based on the matching scores (Figure 2B). Users can interactively explore the MS2 database matching results. The
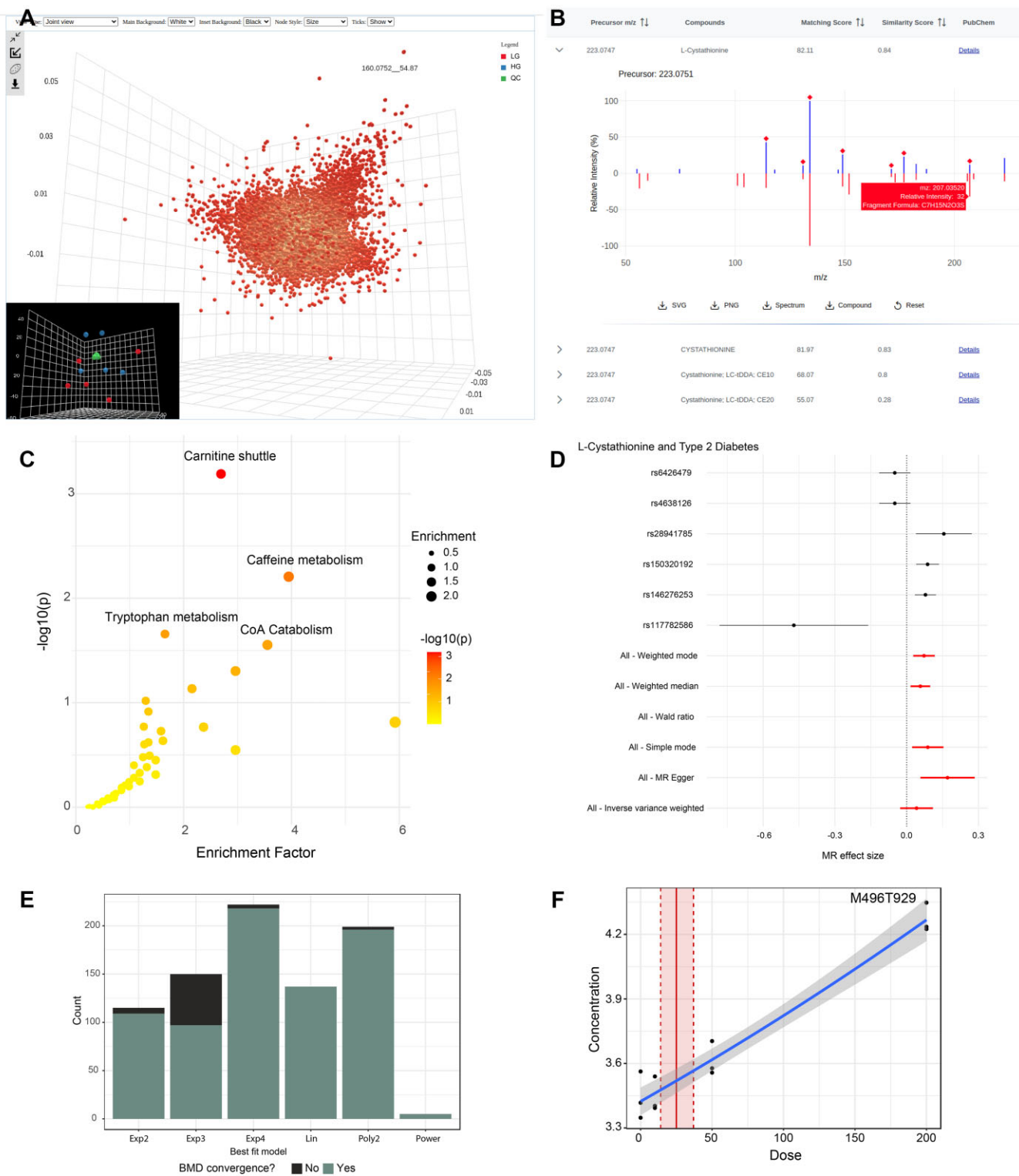
Figure 2. Example outputs from MetaboAnalyst 6.0. (**A**) Integrated 3D PCA score and loading plots summarizing the raw spectra processing results. (**B**) An interactive mirror plot showing the MS2 matching result. Matched fragments are marked with a red diamond. (**C**) Functional analysis results with the top four significant pathways labelled. (**D**) A forest plot comparing the effect sizes calculated based on individual SNPs (black) or using all SNPs by different MR methods (red). (**E**) Bar plots of the dose response curve fitting results showing how many times each model type was identified as the best fit. (**F**) A dose-response curve fitting result showing each of the concentration values (black points), the fitted curve (solid blue line), and the estimated benchmark dose (solid red line) with its lower and upper 95% confidence intervals (dashed red lines), respectively.

molecular formulas for the MS2 peaks in the reference database spectra are predicted using the BUDDY program (31). Users can download the complete compound identification table together with the mirror plots.

Causal analysis via two-sample Mendelian randomization

Understanding the causal relationships between metabolites and phenotypes is of great interest in both metabolomics and exposomics. GWAS have established links between genetic variants (e.g. single nucleotide polymorphism, or SNPs) and various phenotypes (32), while recent mGWAS provide connections between genotypes with metabolites or metabolite concentration changes. It becomes possible to estimate causal relationships between metabolites and a phenotype of interest. If a metabolite is causal for a given disease, genetic variants which influence the levels of that metabolite, either directly through affecting related enzymes or indirectly through influencing lifestyle choices (such as dietary habits), should result in a higher risk of the disease. These causal effects can be estimated through Mendelian randomization (MR) analysis (18). MR relies on the principle that genetic variants are randomly distributed across populations, similar to how treatments are randomly assigned in clinical trials. By leveraging this random allocation, MR can evaluate whether a relationship between a metabolite and a phenotype is causal, while reducing the impact of confounding factors and reverse causality that often plague observational studies.

MR analysis in MetaboAnalyst is based on the 2SMR approach (using the *TwoSampleMR* and *MRInstruments* R packages) which enables application of MR methods using summary statistics from non-overlapping individuals (17,33). Users should first select an exposure (i.e. a metabolite) and an outcome (i.e. a disease) of interest. Based on the selections, the program searches for potential instrumental variables (i.e. SNPs) that are associated with both the metabolite from our large collections of the recent mGWAS studies (14) and the disease from the OpenGWAS database (17). The next step is to perform SNP filtering and harmonization to identify independent SNPs through linkage disequilibrium (LD) clumping (34). When SNPs are absent in the GWAS database, proxy SNPs are identified using LD. In addition, it is critical to harmonize SNPs to make sure effect sizes for the SNPs on both exposures and the outcomes are for the same reference alleles. The last step before conducting MR analysis is to exclude SNPs affecting multiple metabolites to reduce horizontal pleiotropy which occurs when a genetic variant influences the outcome through pathways other than the exposure of interest (35). MetaboAnalyst's MR analysis page provides diverse statistical methods (currently 12), each of which has its own strengths and limitations. For instance, the weighted median method is robust to the violation of MR assumptions by some of the genetic variants, while Egger regression method is more robust to horizontal pleiotropy. Users can point their mouse over the corresponding question marks beside each method to learn more details.

Dose–response analysis

Dose–response analysis is commonly used in toxicology and pharmacology for understanding how varying concentrations of a chemical can impact a biological system. It plays a pivotal role in risk assessment of chemical exposures (36). A key output of dose-response analysis is the benchmark dose (BMD), the minimum dose of a substance that produces a clear, low level health risk relative to the control group (37). Chemicals identified from exposomics are often followed up by dose–response studies to understand their mechanism of action or adverse outcome pathways (21,38,39).

Dose–response experiment design includes a control group (dose = 0) and at least three different dose groups, typically with the same number of replicates in each group. The data should be formatted as a csv file with their dose information included as the second row or column. The analysis workflow consists of four main steps: (i) data upload, integrity checking, processing and normalization; (ii) differential analysis to select features that vary with dose levels; (iii) curve fitting on the intensity or concentration values of those selected features against a suite of linear and non-linear models, and (iv) computing BMD values for each feature. The algorithm for dose–response analysis was adapted from the algorithm we developed for transcriptomics BMD analysis (40,41).

Updated compound database and knowledge libraries

Compound database

The compound database has been updated based on HMDB 5.0 (42), with particular efforts made to synchronize with the IDs of other databases such as KEGG (43) and PubChem (44) to improve cross-references during compound mapping and pathway analysis. The compound database was expanded by ~4000 compounds (after removing ~10 000 deprecated HMDB entries and adding ~14 000 new entries).

MS2 reference spectra database.

A total of 12 MS2 reference databases were collected and curated from public resources, including the HMDB experimental MS2 database (42), the HMDB predicted MS2 database (42), Global Natural Product Social Molecular Networking (GNPS) database (45), MoNA (46), MassBank (46), MINEs (47), LipidBlast (48), RIKEN (49), ReSpect (50), BMDMS (51), VaniyaNP (46) and the MS-DIAL database (v4.90) (52). The complete MS2 reference database currently comprises 10 420 215 MS2 records from 1 551 012 unique compounds. We also created a neutral loss spectra database calculated based on the algorithm implemented by the METLIN neutral loss database (53). The molecular formula of all MS2 fragments were pre-calculated using BUDDY (31).

Pathway and metabolite set libraries

The KEGG pathway libraries have been updated to their recent version (12/20/2023) via KEGG API. Based on user feedback, the pathway analysis for both targeted and untargeted metabolomics data now supports ~130 species (up from 28 species in version 5.0), including many new mammals, plants, insects, fungi, and bacteria, etc. We also updated the metabolite set libraries based on HMDB 5.0, MarkerDB (54), as well as manual curation. For instance, a total of 62 metabolite sets associated with dietary and chemical exposures were added during this process. The metabolite set library also incorporated ~3700 pathways downloaded from the RaMP-DB (55).

Other features

Statistical analysis with complex metadata

The *Statistical Analysis [metadata table]* module in Metabo-Analyst 6.0 now provides a comprehensive suite of methods for analyzing and visualizing metabolomics data in relation to various metadata, be it discrete or continuous. Users can quickly assess the correlation patterns among different experimental factors using the metadata overview heatmaps or interactive PCA visualization. The interactive heatmap visualization coupled with hierarchical clustering allows users to easily explore feature abundance variations across different samples and metadata variables. The statistical methods in this module include both univariate linear models with covariate adjustment as well as multivariate methods such as ANOVA Simultaneous Component Analysis (56,57). Random forest is offered for classification with consideration of different metadata variables of interest. More details about this module can be found in our recently published protocol (11).

Enhanced functional analysis for untargeted metabolomics

Functional analysis of untargeted metabolomics was initially established based on *mummichog* and Gene Set Enrichment Analysis (GSEA) since MetaboAnalyst 4.0 (58). It was further enhanced in MetaboAnalyst 5.0 by incorporating retention time into calculating empirical compounds. MetaboAnalyst 6.0 now allows users to upload an LC–MS peak list along with a corresponding MS2-based compound list to filter out unrealistic empirical compounds to further improve the accuracy in functional analysis (59).

Enhanced data visualization support

We have enhanced the quality of the interactive and synchronized 3D plots across the dimensionality reduction methods (PCA, PLS-DA, sPLS-DA) used in MetaboAnalyst based on the powerful three.js library (https://threejs.org/). New features include customizable backgrounds, data point annotations and confidence ellipsoids (Figure 2A). We have also implemented interactive plots for clustering heatmaps in the Statistical Analysis modules to better support visual exploration of large data matrices typical in untargeted metabolomics. Both mouse-over and zoom-in functionalities are supported to allow users to examine specific features or patterns of interest. In addition to these enhancements, we also updated the visualization for KEGG's global metabolic network (43).

Case study

To illustrate the utility of the new features of MetaboAnalyst 6.0, we used a metabolomics dataset collected in-house that aimed at studying glucose-induced insulin secretion in isolated human islets. The dataset contains five samples of high-glucose (16.7 mM) exposures, five samples of low-glucose (2.8 mM) exposures, both for 30 min, and five quality control (QC) samples. The LC-MS spectra were collected using our Q-Exactive Orbitrap platform (Thermo Scientific, Waltham, MA USA), together with three SWATH-DIA acquisitions from the pooled QC. The spectra were first centroided and converted into mzML format using ProteoWizard (60,61) and uploaded to MetaboAnalyst 6.0. LC–MS spectra processing was performed using the *asari* algorithm. All detected MS1 features

were used as a target list for MS2 deconvolution and database searching. A total of 27 209 MS1 features were detected, with 4959 of them identified with at least one potential named chemical identity. Functional analysis using the *mummichog* algorithm indicated compounds showing significant changes between the high-glucose and low-glucose groups were involved in the *Carnitine shuttle*, *Caffeine*, *Tryptophan,* and *Coenzyme A metabolism* pathways (Figure 2C). These pathways have been consistently identified in previous studies (62–65). Finally, we performed a causal analysis on the associations between one of the significant metabolites identified, *L-Cystathionine* and type 2 diabetes (GWAS ID: finn-b-E4_DM2). The default parameters were used for both SNP filtering and harmonization, as well as MR analysis. Based on these results, a significantly altered cystathionine level was found to have a causal effect on type 2 diabetes (Figure 2D), which aligns well with a study published recently (66). This case study highlights how MetaboAnalyst 6.0 allows users to investigate the chemical identities of MS peaks, elucidate associations between metabolites and phenotypes to unveil previously unknown functional insights. To showcase the dose-response analysis module, we utilized a published data collected from BT549 breast cancer cells treated with four different doses of etomoxir (21). Figure 2E summarizes the results from dose-response modeling. Figure 2F shows an example feature-level BMD calculated based on the fitted curve. The workflow is included as a series of tutorials on our website.

Comparison with other tools

Several web-based tools have been developed to address various aspects of metabolomics data processing, statistical analysis, functional interpretation, and results visualization. Table 1 compares the main features of MetaboAnalyst 6.0 with other popular tools including the previous version, XCMS online (23), GNPS (45), Workflow4Metabolomics (W4M) (67) and MetExplore (68). For raw data processing, MetaboAnalyst primarily focuses on supporting LC–MS data, whereas W4M also supports GC–MS and NMR raw data processing, and GNPS emphasizes MS2-based compound identification via molecular networks. In comparison, MetaboAnalyst provides an auto-optimized workflow along with an additional algorithm (asari) for efficient LC–MS spectra processing, together with more extensive MS2 spectra libraries for compound identification. In terms of statistical analysis, MetaboAnalyst 6.0 has introduced new modules for dealing with complex metadata, causal analysis and dose–response analysis, while maintaining all other functionalities. MetaboAnalyst contains unique features for enrichment and pathway analysis, and these strengths were further improved in version 6.0, with the addition of unique functions and supports for more species. For network analysis and integration, MetExplore specializes in metabolic network visualization and integration with other omics. These features are addressed by our companion tool, OmicsNet (69). Overall, MetaboAnalyst 6.0 continues to be the most comprehensive tool for metabolomics data processing, analysis and interpretation.

Conclusion

By incorporating a new MS2 data processing workflow, MetaboAnalyst 6.0 now offers a web-based, end-to-end platform for metabolomics data analysis. The workflow spans

Table 1. Comparison of MetaboAnalyst 6.0 with its previous version and other common web-based metabolomics tools. Symbols used for feature evaluations with '√' for present, '-' for absent, and '+' for a more quantitative assessment (more '+' indicate better support)

Tools	MetaboAnalyst 6.0	MetaboAnalyst 5.0	XCMS Online	GNPS	W4M	MetExplore
Raw spectra processing						
MS1 feature detection	+++	++	+	−	+	−
MS1 feature annotation	++	++	++	−	+	−
MS2 spectra deconvolution	√	−	−	−	−	−
MS2 compound identification	+++	−	+++	+++	−	−
Raw spectra results visualization	++	++	+	+	+	−
GC-MS/NMR spectra processing	−	−	−	−	√	−
Statistical Analysis						
Univariate	+++	++	+	−	++	−
Multivariate	+++	++	+	−	++	−
Clustering	+++	+++	+	−	+	−
Classification	√	√	−	−	−	−
Complex metadata support	√	−	−	−	−	−
Biomarker analysis	√	√	−	−	−	−
Power analysis	√	√	−	−	−	−
Meta-analysis	√	√	−	−	−	−
Dose-response analysis	√	−	−	−	−	−
Causal analysis	√	−	−	−	−	−
Functional analysis						
Functional analysis (MS peaks)	+++	++	++	−	−	−
Enrichment analysis (compounds)	+++	++	−	−	−	+
Functional meta-analysis	√	√	−	−	−	−
Network analysis	++	++	−	+	−	++

- XCMS online: https://xcmsonline.scripps.edu/.
- GNPS: https://gnps.ucsd.edu/.
- Workflow4Metabolomics (W4M): https://workflow4metabolomics.org/.
- MetExplore: https://metexplore.toulouse.inra.fr/metexplore2/

from raw MS spectra processing to compound identification to functional analysis. A key motivation in developing version 6.0 was to support the data analysis needs emerging from exposomics and follow-up validation studies. The new statistical analysis module specifically takes into account of complex metadata to better identify robust associations. From these associations, users can perform causal analysis based on 2SMR to narrow down candidate compounds. The remaining compounds can be validated through dose-response studies based on *in vitro* or animal models. Our case study highlights the streamlined analysis workflow from raw spectra processing to compound annotation, to functional interpretation, and finally to causal insights. In conclusion, MetaboAnalyst 6.0 is a user-friendly platform for comprehensive analysis of metabolomics data and help address emerging needs from recent exposomics research. For future directions, we will continue to improve metabolome annotations, better integrate with other omics data, and explore new ways to interact with users via generative artificial intelligence technologies (70–73).

Data availability

MetaboAnalyst 6.0 is freely available at https://www.metaboanalyst.ca. No log in required.

Acknowledgements

Human islets for research were provided by the Alberta Diabetes Institute IsletCore at the University of Alberta in Edmonton with the assistance of the Human Organ Procurement and Exchange (HOPE) program, Trillium Gift of Life Network (TGLN), and other Canadian organ procurement organizations. Islet isolation was approved by the Human Research Ethics Board at the University of Alberta (Pro00013094). All donors' families gave informed consent for the use of pancreatic tissue in research.

Conflict of interest statement

J. Xia is the founder of XiaLab Analytics.

References

1. Lloyd-Price,J., Arze,C., Ananthakrishnan,A.N., Schirmer,M., Avila-Pacheco,J., Poon,T.W., Andrews,E., Ajami,N.J., Bonham,K.S., Brislawn,C.J., *et al.* (2019) Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature*, **569**, 655–662.
2. Utpott,M., Rodrigues,E., Rios,A.O., Mercali,G.D. and Flores,S.H. (2022) Metabolomics: an analytical technique for food processing evaluation. *Food Chem.*, **366**, 130685.
3. Wishart,D.S. (2019) Metabolomics for investigating physiological and pathophysiological processes. *Physiol. Rev.*, **99**, 1819–1875.

4. Vermeulen,R., Schymanski,E.L., Barabasi,A.L. and Miller,G.W. (2020) The exposome and health: where chemistry meets biology. *Science*, **367**, 392–396.

5. Danzi,F., Pacchiana,R., Mafficini,A., Scupoli,M.T., Scarpa,A., Donadelli,M. and Fiore,A. (2023) To metabolomics and beyond: a technological portfolio to investigate cancer metabolism. *Signal. Transduct. Target. Ther.*, **8**, 137.

6. Xia,J., Psychogios,N., Young,N. and Wishart,D.S. (2009) MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Res.*, **37**, W652–W660.

7. Xia,J., Mandal,R., Sinelnikov,I.V., Broadhurst,D. and Wishart,D.S. (2012) MetaboAnalyst 2.0—A comprehensive server for metabolomic data analysis. *Nucleic Acids Res.*, **40**, W127–W133.

8. Xia,J., Sinelnikov,I.V., Han,B. and Wishart,D.S. (2015) MetaboAnalyst 3.0—Making metabolomics more meaningful. *Nucleic Acids Res.*, **43**, W251–W257.

9. Chong,J., Soufan,O., Li,C., Caraus,I., Li,S., Bourque,G., Wishart,D.S. and Xia,J. (2018) MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res.*, **46**, W486–W494.

10. Pang,Z., Chong,J., Zhou,G., de Lima Morais,D.A., Chang,L., Barrette,M., Gauthier,C., Jacques,P.-É., Li,S. and Xia,J. (2021) MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res.*, **49**, W388–W396.

11. Pang,Z., Zhou,G., Ewald,J., Chang,L., Hacariz,O., Basu,N. and Xia,J. (2022) Using MetaboAnalyst 5.0 for LC–HRMS spectra processing, multi-omics integration and covariate adjustment of global metabolomics data. *Nat. Protoc.*, **17**, 1735–1761.

12. Frigerio,G., Moruzzi,C., Mercadante,R., Schymanski,E.L. and Fustinoni,S. (2022) Development and application of an LC–MS/MS untargeted exposomics method with a separated pooled quality control strategy. *Molecules*, **27**, 2580.

13. Tsugawa,H., Cajka,T., Kind,T., Ma,Y., Higgins,B., Ikeda,K., Kanazawa,M., VanderGheynst,J., Fiehn,O. and Arita,M. (2015) MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat. Methods*, **12**, 523–526.

14. Chang,L., Zhou,G. and Xia,J. (2023) mGWAS-Explorer 2.0: causal analysis and interpretation of metabolite-phenotype associations. *Metabolites*, **13**, 826.

15. Shin,S.Y., Fauman,E.B., Petersen,A.K., Krumsiek,J., Santos,R., Huang,J., Arnold,M., Erte,I., Forgetta,V., Yang,T.P., *et al.* (2014) An atlas of genetic influences on human blood metabolites. *Nat. Genet.*, **46**, 543–550.

16. Chen,Y., Lu,T., Pettersson-Kymmer,U., Stewart,I.D., Butler-Laporte,G., Nakanishi,T., Cerani,A., Liang,K.Y.H., Yoshiji,S., Willett,J.D.S., *et al.* (2023) Genomic atlas of the plasma metabolome prioritizes metabolites implicated in human diseases. *Nat. Genet.*, **55**, 44–53.

17. Hemani,G., Zheng,J., Elsworth,B., Wade,K.H., Haberland,V., Baird,D., Laurin,C., Burgess,S., Bowden,J., Langdon,R., *et al.* (2018) The MR-Base platform supports systematic causal inference across the human phenome. *eLife*, **7**, e34408.

18. Sanderson,E., Glymour,M.M., Holmes,M.V., Kang,H., Morrison,J., Munafò,M.R., Palmer,T., Schooling,C.M., Wallace,C., Zhao,Q., *et al.* (2022) Mendelian randomization. *Nat. Rev. Methods Primers*, **2**, 6.

19. Zhao,H., Liu,M., Lv,Y. and Fang,M. (2022) Dose-response metabolomics and pathway sensitivity to map molecular cartography of bisphenol A exposure. *Environ. Int.*, **158**, 106893.

20. Thomas,R.S., Wesselkamper,S.C., Wang,N.C.Y., Zhao,Q.J., Petersen,D.D., Lambert,J.C., Cote,I., Yang,L., Healy,E., Black,M.B., *et al.* (2013) Temporal concordance between apical and transcriptional points of departure for chemical risk assessment. *Toxicol. Sci.*, **134**, 180–194.

21. Yao,C.-H., Wang,L., Stancliffe,E., Sindelar,M., Cho,K., Yin,W., Wang,Y. and Patti,G.J. (2020) Dose-response metabolomics to understand biochemical mechanisms and off-target drug effects with the TOXcms software. *Anal. Chem.*, **92**, 1856–1864.

22. Li,S., Siddiqa,A., Thapa,M., Chi,Y. and Zheng,S. (2023) Trackable and scalable LC–MS metabolomics data processing using asari. *Nat. Commun.*, **14**, 4113.

23. Tautenhahn,R., Patti,G.J., Rinehart,D. and Siuzdak,G. (2012) XCMS Online: a web-based platform to process untargeted metabolomic data. *Anal. Chem.*, **84**, 5035–5039.

24. Schmid,R., Heuckeroth,S., Korf,A., Smirnov,A., Myers,O., Dyrlund,T.S., Bushuiev,R., Murray,K.J., Hoffmann,N., Lu,M., *et al.* (2023) Integrative analysis of multimodal mass spectrometry data in MZmine 3. *Nat. Biotechnol.*, **41**, 447–449.

25. Myers,O.D., Sumner,S.J., Li,S., Barnes,S. and Du,X. (2017) Detailed investigation and comparison of the XCMS and MZmine 2 chromatogram construction and chromatographic peak detection methods for preprocessing mass spectrometry metabolomics data. *Anal. Chem.*, **89**, 8689–8695.

26. Pang,Z., Chong,J., Li,S. and Xia,J. (2020) MetaboAnalystR 3.0: toward an optimized workflow for global metabolomics. *Metabolites*, **10**, 186.

27. Xing,S., Yu,H., Liu,M., Jia,Q., Sun,Z., Fang,M. and Huan,T. (2021) Recognizing contamination fragment ions in liquid chromatography–Tandem mass spectrometry data. *J. Am. Soc. Mass. Spectrom.*, **32**, 2296–2305.

28. Stancliffe,E., Schwaiger-Haber,M., Sindelar,M. and Patti,G.J. (2021) DecoID improves identification rates in metabolomics through database-assisted MS/MS deconvolution. *Nat. Methods*, **18**, 779–787.

29. Yin,Y., Wang,R., Cai,Y., Wang,Z. and Zhu,Z.-J. (2019) DecoMetDIA: deconvolution of multiplexed MS/MS spectra for metabolite identification in SWATH-MS-based untargeted metabolomics. *Anal. Chem.*, **91**, 11897–11904.

30. Li,Y., Kind,T., Folz,J., Vaniya,A., Mehta,S.S. and Fiehn,O. (2021) Spectral entropy outperforms MS/MS dot product similarity for small-molecule compound identification. *Nat. Methods*, **18**, 1524–1531.

31. Xing,S., Shen,S., Xu,B., Li,X. and Huan,T. (2023) BUDDY: molecular formula discovery via bottom-up MS/MS interrogation. *Nat. Methods*, **20**, 881–890.

32. Sollis,E., Mosaku,A., Abid,A., Buniello,A., Cerezo,M., Gil,L., Groza,T., Gunes,O., Hall,P., Hayhurst,J., *et al.* (2023) The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res.*, **51**, D977–D985.

33. Hemani,G., Tilling,K. and Davey Smith,G. (2017) Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLos Genet.*, **13**, e1007081.

34. Marees,A.T., de Kluiver,H., Stringer,S., Vorspan,F., Curis,E., Marie-Claire,C. and Derks,E.M. (2018) A tutorial on conducting genome-wide association studies: quality control and statistical analysis. *Int. J. Methods Psychiatr. Res.*, **27**, e1608.

35. de Leeuw,C., Savage,J., Bucur,I.G., Heskes,T. and Posthuma,D. (2022) Understanding the assumptions underlying mendelian randomization. *Eur. J. Hum. Genet.*, **30**, 653–660.

36. Altshuler,B. (1981) Modeling of dose-response relationships. *Environ. Health Perspect.*, **42**, 23–27.

37. Thomas,R.S., Wesselkamper,S.C., Wang,N.C., Zhao,Q.J., Petersen,D.D., Lambert,J.C., Cote,I., Yang,L., Healy,E., Black,M.B., *et al.* (2013) Temporal concordance between apical and transcriptional points of departure for chemical risk assessment. *Toxicol. Sci.*, **134**, 180–194.

38. Zhao,H., Liu,M., Lv,Y. and Fang,M. (2022) Dose-response metabolomics and pathway sensitivity to map molecular cartography of bisphenol A exposure. *Environ. Int.*, **158**, 106893.

39. Kleensang,A., Maertens,A., Rosenberg,M., Fitzpatrick,S., Lamb,J., Auerbach,S., Brennan,R., Crofton,K.M., Gordon,B., Fornace,A.J. Jr, *et al.* (2014) Pathways of toxicity. *ALTEX*, **31**, 53–61.

40. Ewald,J., Soufan,O., Xia,J. and Basu,N. (2020) FastBMD: an online tool for rapid benchmark dose–response analysis of transcriptomics data. *Bioinformatics*, **37**, 1035–1036.

41. Ewald,J., Zhou,G., Lu,Y. and Xia,J. (2023) Using ExpressAnalyst for comprehensive gene expression analysis in model and non-model organisms. *Curr Protoc*, **3**, e922.

42. Wishart,D.S., Guo,A., Oler,E., Wang,F., Anjum,A., Peters,H., Dizon,R., Sayeeda,Z., Tian,S., Lee,B.L., *et al.* (2021) HMDB 5.0: the Human Metabolome Database for 2022. *Nucleic Acids Res.*, **50**, D622–D631.

43. Kanehisa,M., Furumichi,M., Sato,Y., Kawashima,M. and Ishiguro-Watanabe,M. (2022) KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.*, **51**, D587–D592.

44. Kim,S. (2021) Exploring chemical information in PubChem. *Curr. Protoc.*, **1**, e217.

45. Aron,A.T., Gentry,E.C., McPhail,K.L., Nothias,L.-F., Nothias-Esposito,M., Bouslimani,A., Petras,D., Gauglitz,J.M., Sikora,N., Vargas,F., *et al.* (2020) Reproducible molecular networking of untargeted mass spectrometry data using GNPS. *Nat. Protoc.*, **15**, 1954–1991.

46. Horai,H., Arita,M., Kanaya,S., Nihei,Y., Ikeda,T., Suwa,K., Ojima,Y., Tanaka,K., Tanaka,S., Aoshima,K., *et al.* (2010) MassBank: a public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom.*, **45**, 703–714.

47. Jeffryes,J.G., Colastani,R.L., Elbadawi-Sidhu,M., Kind,T., Niehaus,T.D., Broadbelt,L.J., Hanson,A.D., Fiehn,O., Tyo,K.E. and Henry,C.S. (2015) MINEs: open access databases of computationally predicted enzyme promiscuity products for untargeted metabolomics. *J Cheminform*, **7**, 44.

48. Kind,T., Liu,K.-H., Lee,D.Y., DeFelice,B., Meissen,J.K. and Fiehn,O. (2013) LipidBlast in silico tandem mass spectrometry database for lipid identification. *Nat. Methods*, **10**, 755–758.

49. Tsugawa,H., Nakabayashi,R., Mori,T., Yamada,Y., Takahashi,M., Rai,A., Sugiyama,R., Yamamoto,H., Nakaya,T., Yamazaki,M., *et al.* (2019) A cheminformatics approach to characterize metabolomes in stable-isotope-labeled organisms. *Nat. Methods*, **16**, 295–298.

50. Sawada,Y., Nakabayashi,R., Yamada,Y., Suzuki,M., Sato,M., Sakata,A., Akiyama,K., Sakurai,T., Matsuda,F., Aoki,T., *et al.* (2012) RIKEN tandem mass spectral database (ReSpect) for phytochemicals: a plant-specific MS/MS-based data resource and database. *Phytochemistry*, **82**, 38–45.

51. Lee,S., Hwang,S., Seo,M., Shin,K.B., Kim,K.H., Park,G.W., Kim,J.Y., Yoo,J.S. and No,K.T. (2020) BMDMS-NP: a comprehensive ESI-MS/MS spectral library of natural compounds. *Phytochemistry*, **177**, 112427.

52. Tsugawa,H., Ikeda,K., Takahashi,M., Satoh,A., Mori,Y., Uchino,H., Okahashi,N., Yamada,Y., Tada,I., Bonini,P., *et al.* (2020) A lipidome atlas in MS-DIAL 4. *Nat. Biotechnol.*, **38**, 1159–1163.

53. Aisporna,A., Benton,H.P., Chen,A., Derks,R.J.E., Galano,J.M., Giera,M. and Siuzdak,G. (2022) Neutral loss mass spectral data enhances molecular similarity analysis in METLIN. *J. Am. Soc. Mass. Spectrom.*, **33**, 530–534.

54. Wishart,D.S., Bartok,B., Oler,E., Liang,K.Y.H., Budinski,Z., Berjanskii,M., Guo,A., Cao,X. and Wilson,M. (2021) MarkerDB: an online database of molecular biomarkers. *Nucleic Acids Res.*, **49**, D1259–D1267.

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

56. Smilde,A.K., Jansen,J.J., Hoefsloot,H.C., Lamers,R.J., van der Greef,J. and Timmerman,M.E. (2005) ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data. *Bioinformatics*, **21**, 3043–3048.

57. Ritchie,M.E., Phipson,B., Wu,D., Hu,Y., Law,C.W., Shi,W. and Smyth,G.K. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.*, **43**, e47.

58. Li,S., Park,Y., Duraisingham,S., Strobel,F.H., Khan,N., Soltow,Q.A., Jones,D.P. and Pulendran,B. (2013) Predicting network activity from high throughput metabolomics. *PLoS Comput. Biol.*, **9**, e1003123.

59. Lu,Y., Pang,Z. and Xia,J. (2023) Comprehensive investigation of pathway enrichment methods for functional interpretation of LC–MS global metabolomics data. *Brief. Bioinform.*, **24**, bbac553.

60. Chambers,M.C., Maclean,B., Burke,R., Amodei,D., Ruderman,D.L., Neumann,S., Gatto,L., Fischer,B., Pratt,B., Egertson,J., *et al.* (2012) A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.*, **30**, 918–920.

61. Adusumilli,R. and Mallick,P. (2017) Data conversion with ProteoWizard msConvert. *Methods Mol. Biol.*, **1550**, 339–368.

62. Bene,J., Hadzsiev,K. and Melegh,B. (2018) Role of carnitine and its derivatives in the development and management of type 2 diabetes. *Nutr Diabetes*, **8**, 8.

63. Lane,J.D., Barkauskas,C.E., Surwit,R.S. and Feinglos,M.N. (2004) Caffeine impairs glucose metabolism in type 2 diabetes. *Diabetes Care.*, **27**, 2047–2048.

64. Unluturk,U. and Erbas,T. (2015) In: Engin, A. and Engin,A.B. (eds.) *Tryptophan Metabolism: Implications for Biological Processes, Health and Disease*. Springer International Publishing, Cham, pp. 147–171.

65. Jackowski,S. and Leonardi,R. (2014) Deregulated coenzyme A, loss of metabolic flexibility and diabetes. *Biochem. Soc. Trans.*, **42**, 1118–1122.

66. Cruciani-Guglielmacci,C., Meneyrol,K., Denom,J., Kassis,N., Rachdi,L., Makaci,F., Migrenne-Li,S., Daubigney,F., Georgiadou,E., Denis,R.G., *et al.* (2022) Homocysteine metabolism pathway is involved in the control of glucose homeostasis: a cystathionine beta synthase deficiency study in mouse. *Cells*, **11**, 1737.

67. Giacomoni,F., Le Corguillé,G., Monsoor,M., Landi,M., Pericard,P., Pétéra,M., Duperier,C., Tremblay-Franco,M., Martin,J.-F., Jacob,D., *et al.* (2014) Workflow4Metabolomics: a collaborative research infrastructure for computational metabolomics. *Bioinformatics*, **31**, 1493–1495.

68. Cottret,L., Frainay,C., Chazalviel,M., Cabanettes,F., Gloaguen,Y., Camenen,E., Merlet,B., Heux,S., Portais,J.C., Poupin,N., *et al.* (2018) MetExplore: collaborative edition and exploration of metabolic networks. *Nucleic Acids Res.*, **46**, W495–W502.

69. Zhou,G., Pang,Z., Lu,Y., Ewald,J. and Xia,J. (2022) OmicsNet 2.0: a web-based platform for multi-omics integration and network visual analytics. *Nucleic Acids Res.*, **50**, W527–W533.

70. Lu,Y., Zhou,G., Ewald,J., Pang,Z., Shiri,T. and Xia,J. (2023) MicrobiomeAnalyst 2.0: comprehensive statistical, functional and integrative analysis of microbiome data. *Nucleic Acids Res.*, **51**, W310–W318.

71. Liu,P., Ewald,J., Pang,Z., Legrand,E., Jeon,Y.S., Sangiovanni,J., Hacariz,O., Zhou,G., Head,J.A., Basu,N., *et al.* (2023) ExpressAnalyst: a unified platform for RNA-sequencing analysis in non-model species. *Nat. Commun.*, **14**, 2995.

72. Zhou,G., Ewald,J. and Xia,J. (2021) OmicsAnalyst: a comprehensive web-based platform for visual analytics of multi-omics data. *Nucleic Acids Res.*, **49**, W476–W482.

73. Moor,M., Banerjee,O., Abad,Z.S.H., Krumholz,H.M., Leskovec,J., Topol,E.J. and Rajpurkar,P. (2023) Foundation models for generalist medical artificial intelligence. *Nature*, **616**, 259–265.