MMEASE: enhanced analytical workflow for single-cell metabolomics

Qingxia Yang [1],*,†, Yangbo Dai[2,3],†, Shijie Huang[2],†, Bing Liu[1], Huaicheng Sun[2], Yuan Zhou[2], Yaguo Gong[4],*, Feng Zhu [2],*

[1]Zhejiang Provincial Key Laboratory of Precision Diagnosis and Therapy for Major Gynecological Diseases, Women's Hospital, Zhejiang University School of Medicine, Hangzhou 310058, China
[2]College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, China
[3]State Key Laboratory for Organic Electronics and Information Displays, Institute of Advanced Materials (IAM), Nanjing University of Posts and Telecommunications, Nanjing 210023, China
[4]State Key Laboratory of Quality Research in Chinese Medicine, School of Pharmacy, Macau University of Science and Technology, Macao 999078, China

*To whom correspondence should be addressed. Email: zhufeng@zju.edu.cn
Correspondence may also be addressed to Yaguo Gong. Email: gongyglab@gmail.com
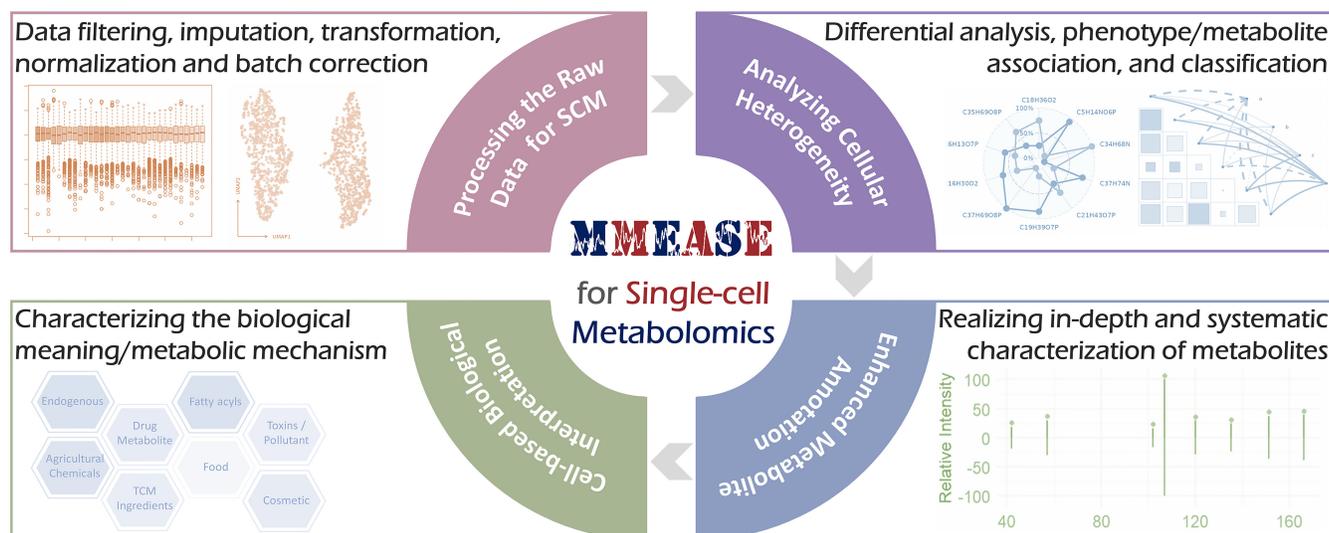Correspondence may also be addressed to Qingxia Yang. Email: yangqx@zju.edu.cn
†The first three authors should be regarded as Joint First Authors.

Abstract

Metabolomics is essential for providing an overview of what chemical processes are taking place. A clear shift from bulk metabolomics to single-cell metabolomics (SCM) is observed in current research, and an integral workflow enabling the analysis of SCM data is therefore in great demand. However, no such workflow has been available to date. Herein, MMEASE, previously designed for analyzing bulk metabolomic data, was therefore updated to its 2.0 version by developing the first comprehensive and in-depth workflow analyzing SCM data. First, it provided all sequential steps of modern SCM research (from SCM data processing, to cellular heterogeneity analysis, then to high-resolution metabolite annotation, and finally to cell-based biological interpretation). Second, compared with the existing tools, MMEASE 2.0 was superior by incorporating the widest variety of methods at every step of the SCM analyses. The originality and functionality of our MMEASE were extensively validated and explicitly described by case studies on benchmark data. All in all, MMEASE 2.0 was unique in accomplishing comprehensive and in-depth analyses of SCM data, which could be considered as an indispensable complement to the existing tools. Now, the latest version of MMEASE is freely accessible by all users at: https://idrblab.org/mmease/

Graphical abstract

Introduction

Metabolomics, as the youngest discipline of all OMICs studies, offers the readout closest to phenotypes and gives a view of what chemical processes are taking place in cells [1, 2]. A workflow of metabolomic analysis includes many crucial steps: data processing, data integration, marker identification, and function annotation [3]. To ensure the production of an accurate and reliable analytical result, it is critical to have a tool which provides a comprehensive workflow for these processes [4, 5]. Therefore, MMEASE 1.0 was designed, and is unique in (i) integrating multiple analytical blocks, (ii) providing enriched annotation for >330 000 metabolites, and (iii) conducting enrichment analysis using various categories/subcategories [6]. Due to these functions, MMEASE has become popular for analyzing bulk metabolomic data and has emerged as a good complement to existing tools [7–9].

A clear shift from bulk metabolomics to single-cell metabolomics (SCM; recognized in 2023 as one of the seven technologies to watch) has recently been observed, and integral workflows for analyzing SCM data are required [10, 11]. Such an analysis consists of four sequential steps: (i) SCM data processing; (ii) cellular heterogeneity analysis; (iii) high-resolution metabolite annotation; and (iv) cell-based biological interpretation [12]. SCM data processing helps to remove the systematic bias/technical variation that arises from the instrumental/sampling issues within SCM studies, which consists of filtering, imputation, transformation, normalization, and batch correction of raw metabolomic data [13]; cellular heterogeneity analysis enables the discovery of metabolic variation among cells (metabolic heterogeneity) and functional change among phenotypes (functional heterogeneity), which includes differential analysis, phenotype/metabolite association, and cell subpopulation classification using metabolic abundance data [14, 15]; high-resolution metabolite annotation accomplishes the in-depth and systematic characterization of metabolic markers at the cellular level using spectra matching [16]; and cell-based biological interpretation aims at characterizing the biological meaning/metabolic mechanism of SCM studies [17]. Due to the extreme complexity of SCM data analysis, as discussed above, a transparent workflow is reported to be greatly needed for non-bioinformaticians in modern SCM-based research [18].

So far, some professional tools, such as SCMeTA [19] and SinCHet-MS [20], have been designed for processing and analyzing SCM data. SCMeTA is a pipeline specifically designed for processing the SCM data, which covers one, one, one, and five methods for data imputation, data transformation, batch correction, and data normalization, respectively, resulting in the most comprehensive pipeline for SCM data processing[19]. SinCHet-MS, on the other hand, focuses primarily on quantifying cell heterogeneity and subpopulations, making it a popular tool enabling *c*ellular heterogeneity analysis for SCM studies [20]. In other words, existing tools are designed to target one particular step in SCM data analysis and, to the best of our knowledge, an integral workflow for the entire analysis chain is not available to date [19, 20]. Furthermore, there are extensive depths that have not been achieved by the existing tools in their specialized steps. Taking the data imputation in the step of SCM data processing as an example, it is essential for SCM data analysis, since far more missing values are found in the raw data of SCM, when compared with bulk metabolomics [21–23]. Nevertheless, such a critical imputation process has not been provided by any of the existing tools. Therefore, there is a great demand, especially from non-bioinformaticians, for a comprehensive and in-depth workflow of successful SCM data analysis. However, no such workflow has been developed yet.

Herein, MMEASE 1.0 designed for analyzing bulk metabolomics data [6] was therefore updated to the latest 2.0 version by developing the first comprehensive and in-depth workflow analyzing SCM data (demonstrated in Supplementary Table S1). On the one hand, this newly constructed workflow covered all four sequential steps in modern SCM research (from SCM data processing, to cellular heterogeneity analysis, then to high-resolution metabolite annotation, and finally to cell-based biological interpretation; illustrated in Fig. 1). On the other hand, the depth of MMEASE was enhanced by offering the most diverse methods for each step when compared with existing tools. Three case studies further validated the functionality of this newly updated server. All in all, MMEASE 2.0 provides the first comprehensive and in-depth workflow for analysis of SCM data, making it an indispensable complement to the existing tools related to modern SCM research. MMEASE 2.0 is now freely accessible at: https://idrblab.org/mmease/

Key features of MMEASE 2.0

A comprehensive and in-depth procedure promoting SCM data processing

A comprehensive pipeline for SCM data processing was described in MMEASE, including the filtering, imputation, transformation, normalization, and batch correction of raw metabolomic data. SCM data could be filtered when a tolerable percentage of missing values for a metabolite exceeded the user-defined threshold [23, 24]. After filtering, the k-nearest neighbor (KNN) imputation method [23] and 1/5 of the minimum positive value [13] were then utilized to impute the remaining missing values. To reduce the data range and stabilize variance, three data transformation methods were provided, namely G-log [13], log2 [20], and log10 transformation [12]. To remove unwanted variations, five normalization methods were also incorporated into MMEASE 2.0, i.e. as auto-scaling [12], mean [16], median [12], mass spectrometry MS total useful signal (MSTUS) [24], and internal standard-based normalization [12]. Moreover, batch effects in the large-scale dataset existed because of the differences during sample preparation and other biases. Thus, data integration after batch correction using ComBat [20] or Limma [25] was carried out. The description of each method in SCM data processing is shown in the Supplementary Methods.

Cellular heterogeneity analysis from the metabolic and functional perspectives

In cellular heterogeneity analyses, metabolic heterogeneity and functional heterogeneity were key for discovering metabolic variation among cells and functional change among phenotypes, respectively. To perform differential analysis, eight methods could be used to discover metabolic markers, namely fold change [26], one-way analysis of variance (ANOVA) [27], orthogonal partial least squares discriminant analysis (OPLS-DA) [28], partial least squares discriminant analysis (PLS-DA) [29], variable selection using random
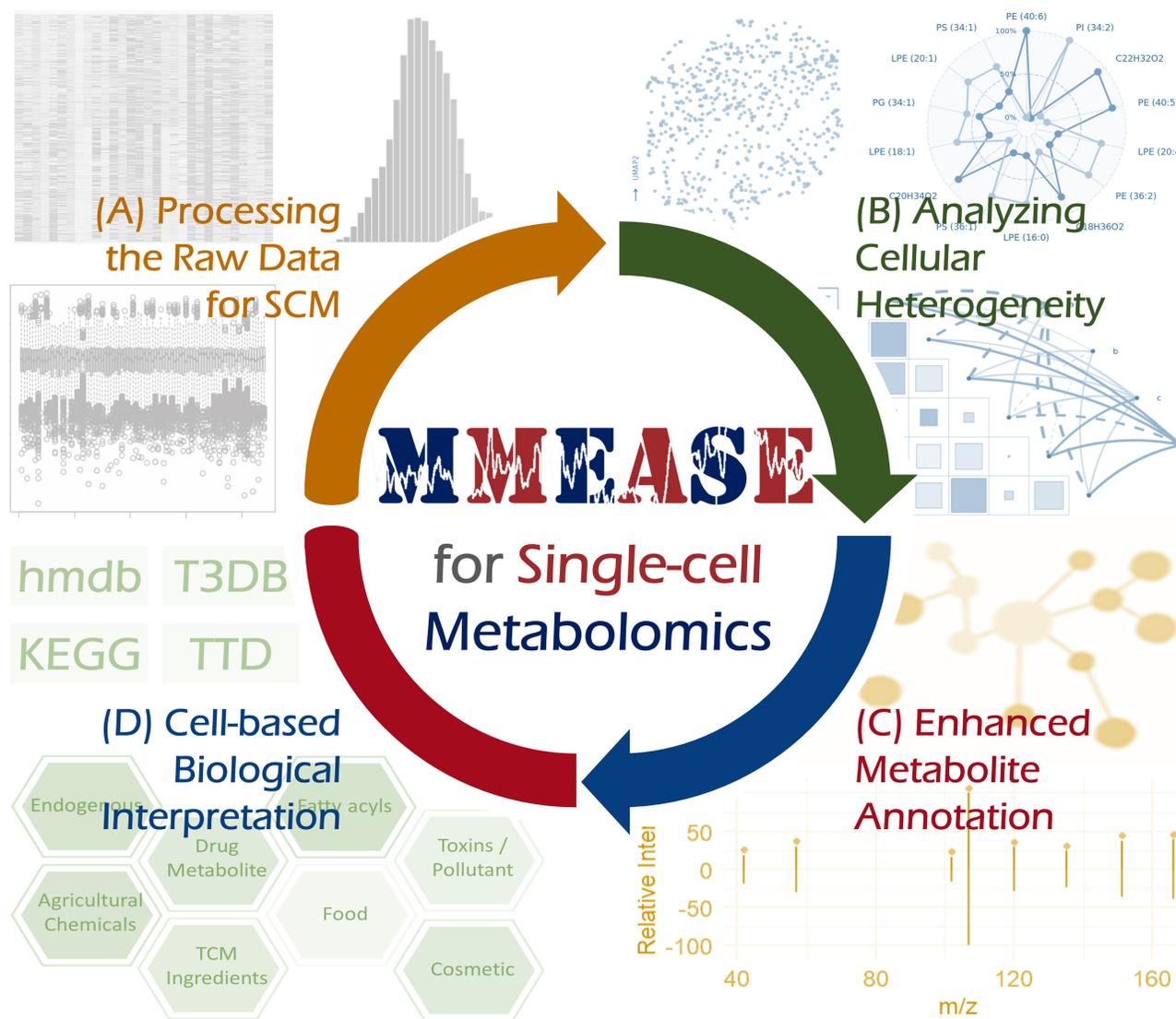
Figure 1. Four key features of MMEASE 2.0. (**A**) SCM data processing comprising filtering, imputation, transformation, normalization, and batch correction of raw metabolomic data. (**B**) Cellular heterogeneity analysis for identifying metabolic and functional heterogeneities by differential analysis, phenotype/metabolite association, and cell subpopulation classification. (**C**) High-resolution metabolite annotation to accomplish in-depth and systematic characterization of metabolic markers at the cellular level using spectra matching. (**D**) Cell-based biological interpretation aimed at characterizing the biological meaning and metabolic mechanism of SCM studies.

forests (VSRF) [30], Student's *t*-test [31], Kruskal–Wallis test (KWT) [32], and support vector machine (SVM)-recursive feature elimination (RFE) [33]. KWT, ANOVA, PLS-DA, VSRF, and SVM-RFE were also designed to discover key metabolites in multiclass ($n \geq 3$) SCM. To visualize the differential metabolites, relative intensity, boxplot, heatmap, and radargrams were also described. In the analyses of phenotype and metabolite association, a correlation scatter matrix represented the correlation among metabolites [34], and the correlation between metabolite and phenotype could be analyzed via Mantel test correlation heatmaps [35]. Moreover, eight classification methods could be selected to construct a model of cell subpopulation classification using the data of metabolite abundance, such as AdaBoost [36], Bagging [37], decision trees [38], KNN [39], linear discriminate analysis [40], naive Bayes (NB), PLS [41], random forest [42], and SVM [43]. For cellular heterogeneity analysis, UMAP (uniform manifold approximation and projection) [44] and *t*-SNE (*t*-distributed stochastic neighbor embedding) [24] were used to visualize di-

mensionality reduction of cells. Explicit descriptions of each method in the cellular heterogeneity analysis are also specifically provided in the Supplementary Methods.

Integrating the new algorithm for enabling high-resolution metabolite annotation

High-resolution metabolite annotation was integrated to facilitate in-depth and systematic characterization of metabolites at the cellular level. To ensure multidimensional metabolite characterization, a high-resolution metabolite annotation was realized using a systematic reference spectra database. First, this database was organized into several libraries of biology, lipid, and exposome. The data of reference spectra databases were derived from public sources, including HMDB [45], MoNA [46], LipidBllast [47], GNPS [48], and KEGG [49].Tandem spectra fragments were then annotated into the molecular formulas using BUDDY with maximum completeness [50]. Finally, two well-established matching approaches

(dot product and spectral entropy) were implemented for assessing the matching similarity between reference databases and the input data [51].

Cell-based biological interpretation through multidimensional characterization

After metabolite annotation, a cell-based biological interpretation was constructed to explain the biological meaning or metabolic mechanism of the corresponding SCM research. To provide the in-depth metabolic characterization, enhanced coverage of biological interpretations was offered in MMEASE. These metabolites were classified into different biological or functional groups through literature reviews and information from public sources, such as HMDB, T3DB [52], KEGG, DrugBank [53], and TTD [54]. To enhance cell-based biological interpretation, MMEASE provided insights into both exogenous factors and biological functions, such as food, drug, microbial, cosmetic, ingredients of traditional medicines, toxins, and pollutants.

Benchmark datasets collected to facilitate case evaluations

Three benchmark datasets were collected to test the utility of MMEASE 2.0. In the first case, the single-cell model of co-cultured human epithelial cells (HeLa) and mouse fibroblast cells (NIH3T3) was used [12]. A total of 88 metabolites were detected from HeLa and NIH3T3 cells, which were analyzed in two replicates. In the second case, hematopoietic stem cells (HSCs) were detected using a platform of high-throughput SCM, and were divided into four subpopulations with 111 features based on signal intensity, such as 39 HSC^a, 40 HSC^b, 39 HSC^c, and 42 HSC^d cells [14]. In the third case, an untargeted metabolomics analysis was conducted in two pairs of colorectal cancer cell lines with various metastatic abilities [15]. Each pair of cell lines comprised primary and metastatic colorectal cancer cell lines (SW480 versus SW620 and HT-29 versus COLO 205). These metabolic features were annotated to 42 and 41 metabolites (SW480 versus SW620 and HT-29 versus COLO 205), respectively, which were then used for enhanced biological interpretation.

Details of server implementation and the required formats of input files

MMEASE 2.0 was deployed on a server installing the Ubuntu Linux v20.04.6, Apache HTTP v2.2.15, and Apache Tomcat servlet container. Its web interface was developed by R v4.4.2 and R package shiny v1.10.0 running on Shiny-server v1.5.16.958. Various R packages were utilized in the background processes. MMEASE 2.0 can be accessed by all users without a login requirement, and by popular web browsers, such as Google Chrome, Mozilla Firefox, Safari, and Internet Explorer. For the input file, a sample-by-feature matrix (cells in rows and features in columns) in csv format was required. For the SCM data analysis, the first row of the first four columns should be sequentially labeled as 'cell', 'class', 'cell type', and 'batch', which denoted the cell ID, functional class, cell types, and batch ID, respectively. The cell ID should be unique; the functional class and cell type of each cell denotes different functional groups and cell types; the batch ID gives different analytical blocks/cell batches. In the following columns, the raw peak intensities across various cells were further provided, and unique metabolites or peaks are listed

in the first row of the input file. Moreover, as regards metabolite annotation, precursor ion mass and tandem spectra should be properly provided. For tandem spectra, the m/z values of spectra were in the first column and the relative abundances of m/z values were in the second column; the two columns were separated using a space and comma in the paste box and uploaded csv file, respectively. The example file strictly following the requirements is provided and can be fully downloaded from the MMEASE website.

Results and discussion

Investigating the functional heterogeneity based on the SCM benchmark

To assess the performance of MMEASE in identifying functional heterogeneity, one benchmark for two independent replicates of co-cultured HeLa and NIH3T3 cells was collected [12]. The SCM data processing pipeline was implemented using the following steps: missing values were imputed by 1/5 of the minimum positive value and mean normalization was used to normalize SCM data. After data processing (left part of Fig. 2A), distinct separation between batches was observed in UMAP before batch corrections. As described on the right-hand side of Fig. 2A, cells from two batches were thoroughly mixed, and ComBat was applied for correcting batch effects. This showed that interbatch variation was effectively removed, resulting in well-integrated data.

As shown in Fig. 2B, after batch correction, clear separation was demonstrated for visualizing the differences between HeLa and NIH3T3 cells, and differential analysis between HeLa and NIH3T3 cells was performed using PLS-DA to identify key metabolites. Among all metabolites, 19 key metabolites found in the original publication were analyzed. As presented in Fig. 2C, differential abundance of 19 metabolites was observed between cell groups, and an inverse correlation relationship between these metabolites and HeLa or NIH3T3 cells was revealed (shown in Fig. 2D). Metabolites positively correlated with HeLa cells were identified as negatively correlated with NIH3T3 cells, and vice versa. Additionally, UMAPs and boxplots were applied to visualize the abundances of specific metabolites, which are provided in Supplementary Figs S1 and S2, respectively. Notably, the differences in two critical metabolites in boxplots are shown in Fig. 2E. It was observed that phosphatidylethanolamine (PE) (40:6), a metabolite marker of NIH3T3 cells, was significantly more highly expressed in NIH3T3 cells, while phosphatidylinositol (PI) (34:2), a marker of HeLa cells, was predominantly expressed in HeLa cells, which was consistent with the results of the original publication.

Identification of the metabolic heterogeneity using the SCM benchmark

A benchmark dataset was collected from a previous publication [14], and was utilized for assessing the capability of analyzing metabolic heterogeneity in MMEASE 2.0, where four subpopulations of cells (HSC^a, HSC^b, HSC^c, and HSC^d) were grouped. Using MMEASE, the metabolites with >20% missing values were filtered out, followed by a missing value imputation based on 1/5 of the minimum positive value. MS-TUS was applied for data normalization. As shown in Fig. 3A, t-SNE visualization of metabolomic differences among the four subpopulations by SVM analysis illustrated that HSC^a
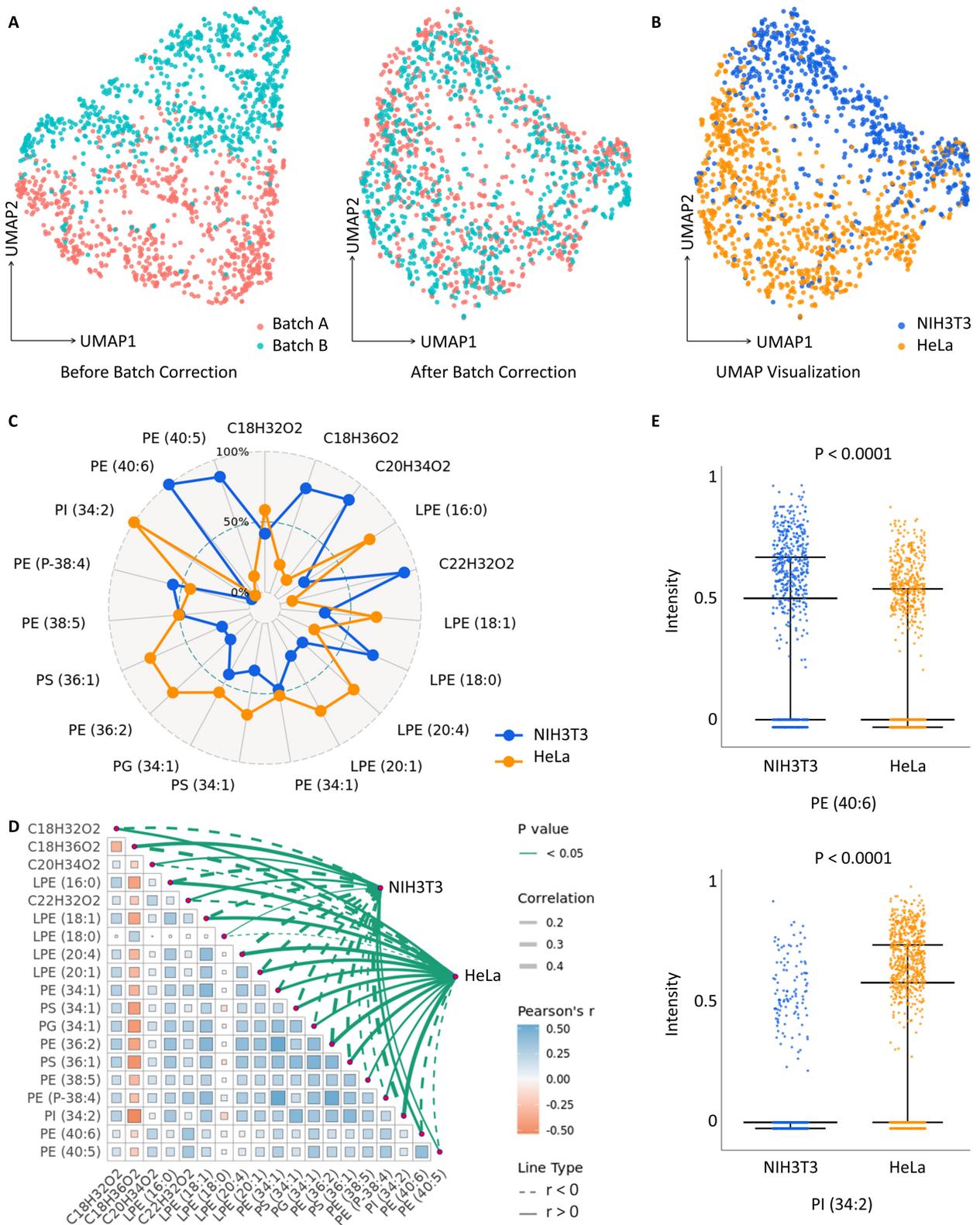
Figure 2. Case study for evaluating the performance of MMEASE 2.0 in revealing functional heterogeneity [12]. (**A**) UMAP of two batches before and after batch correction using ComBat. (**B**) Dimensionality reduction using UMAP for HeLa and NIH3T3 cells. (**C**) Radargram of 19 key metabolites in HeLa and NIH3T3 cells. (**D**) Mantel test correlations between key metabolites and two cell types. (**E**) Differential expression patterns of two key metabolites: phosphatidylethanolamine (PE) (40:6) and phosphatidylinositol (PI) (34:2).
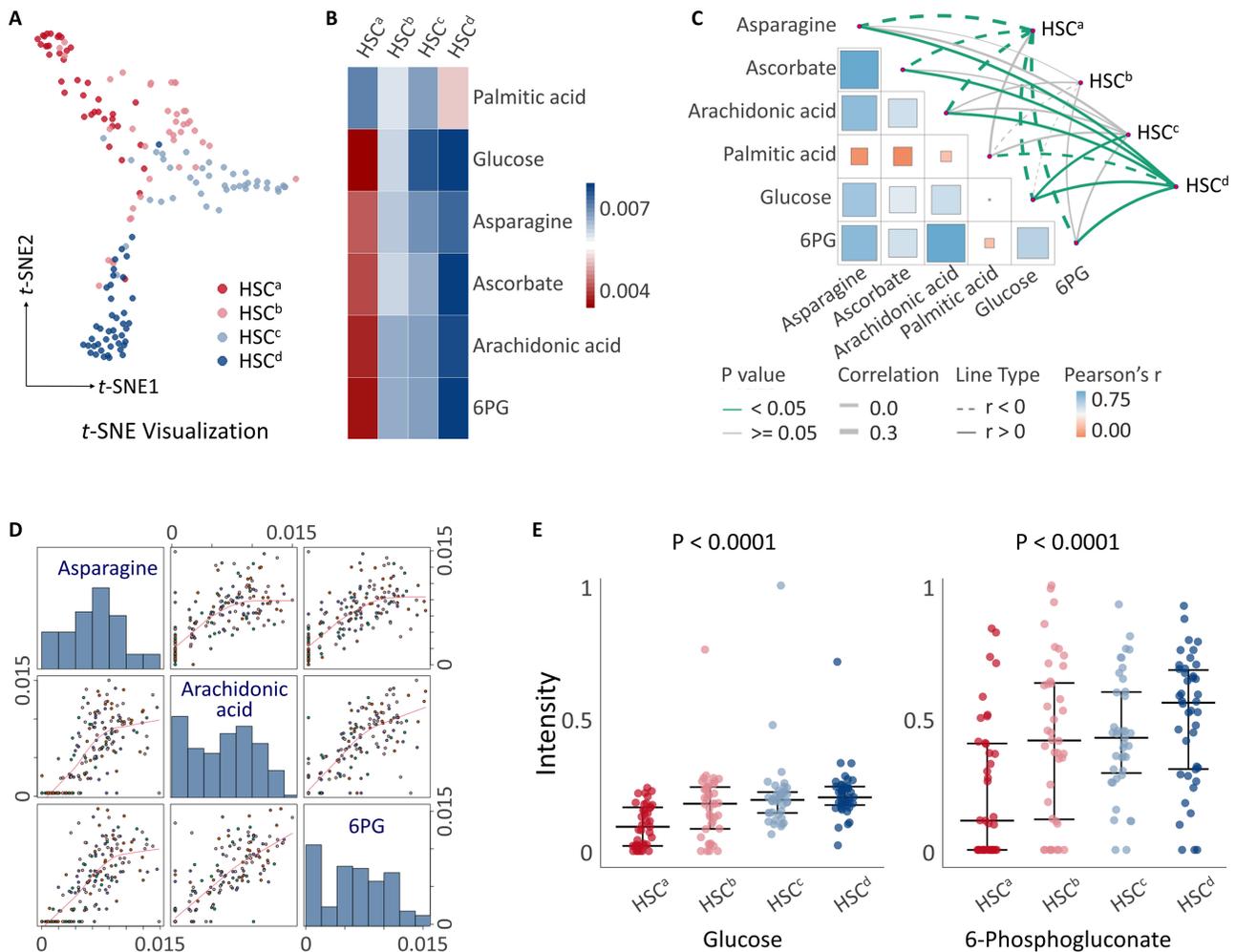
Figure 3. Case study for assessing the performance of MMEASE 2.0 in identifying metabolic heterogeneity [14]. (**A**) Dimensionality reduction using *t*-SNE across subpopulations (HSC[a], HSC[b], HSC[c], and HSC[d]). (**B**) Mean of six key metabolites across subpopulations. (**C**) A Mantel test correlation heatmap was adopted for correction analyses between six metabolites and four subpopulations (HSC[a], HSC[b], HSC[c,] and HSC[d]). (**D**) A correlation scatter matrix was used to find significant positive correlation among three metabolites. (**E**) Two key metabolites [glucose and 6-phosphogluconate (6PG)] were found to be significantly increased from HSC[a] to HSC[d].

was at the top of the map, while HSC[d] was at the bottom of the map. This finding suggested the gradual change through a series of intermediate metabolic states, rather than a binary switching on/off pattern. Differential metabolites among the four cell subpopulations were revealed using PLS-DA. In particular, six metabolites among all the differential metabolites were validated, i.e. palmitic acid, glucose, asparagine, ascorbate, arachidonic acid, and 6-phosphogluconate (6PG). As presented in Fig. 3B, the mean levels of all six metabolites showed great differences across the four subpopulations. Among them, palmitic acid demonstrated a decreasing trend from HSC[a] to HSC[d], while the remaining five metabolites presented an increasing trend from HSC[a] to HSC[d].

As illustrated in Fig. 3C, the correlation of these six metabolites with four subpopulations was analyzed. The metabolites (palmitic acid, asparagine, ascorbate, arachidonic acid, glucose, and 6PG) that were negatively correlated with HSC[a] were found to be positively correlated with HSC[d]. As shown in Fig. 3D, asparagine, arachidonic acid, and 6PG showed a significant positive correlation with each other in the correlation scatter matrix. As shown in Supplementary Fig. S3A, the radargram illustrated the obvious trend of six key metabolites from HSC[a] to HSC[d]. Moreover, it has been validated

that the expression of oxidative pentose phosphate pathway (OxiPPP)-related genes gradually increases during HSC proliferation, and glucose and 6PG (metabolites in the OxiPPP) progressively increased their levels from HSC[a] to HSC[d] [14]. As shown in Fig. 3E, two metabolites (glucose and 6PG) were significantly increased from HSC[a] to HSC[d]. The boxplot of palmitic acid showed a significant decrease from HSC[a] to HSC[d] (described in Supplementary Fig. S3B). However, the boxplots of three metabolites (asparagine, ascorbate, and arachidonic acid) showed a large increase from HSC[a] to HSC[d], as seen in Supplementary Fig. S3C, D, and E, respectively. The statistical significance of the lineage trend among subpopulations was assessed by one-way ANOVA. These results were in good agreement with the conclusions of the original study.

Biological interpretation through multidimensional characterization

Based on the metabolic profiling of low- and high-metastatic human colorectal cancer cells (SW480 versus SW620 and HT-29 versus COLO 205), 42 and 41 metabolites, respectively, were annotated on the basis of tandem spectra [15]. MMEASE

could use the built-in database for annotating metabolites with enhanced biological or functional interpretations. The annotated results with enhanced interpretations of 42 and 41 metabolites are described in Supplementary Tables S2 and S3, respectively. Compared with HMDB [45], MMEASE 2.0 demonstrated significantly enhanced interpretability by enabling multidimensional metabolic characterization using a variety of metabolite-related databases (HMDB, T3DB, KEGG, DrugBank, TTD, etc.), resulting in in-depth biological interpretation from multiple perspectives of food, drug, microbial, cosmetic, traditional medicine, toxin, pollutant, and so on. Taking the interpretation of 'lactic acid' as an example, its interpretation results based on HMDB were 'endogenous' and 'food', while those based on MMEASE were much more versatile, including 'yeast metabolite', 'agricultural chemical', 'toxin/pollutant', 'endogenous', and 'food'. Another example could be 'aspartic acid', whose interpretation result based on HMDB was also 'endogenous' and 'food', while that based on MMEASE was extensively enriched to 'pharmaceutical ingredient', 'cosmetic compound', 'food additive', 'endogenous', and 'food'.

Taken together, such significant enrichments in metabolite biological interpretation highlighted the good performance of MMEASE in processing, analyzing, and interpreting SCM data. MMEASE exhibits extensive applicability in SCM research, demonstrating capabilities in resolving metabolic heterogeneity across cellular subpopulations, decoding immune cell metabolic reprogramming dynamics, and mapping intracellular drug metabolism pathways. To further enhance its utility, future development should prioritize two objectives: (i) developing adaptive computational pipelines that dynamically optimize the analytical workflow to ensure maximized analytical precision, and (ii) incorporating spatial metabolomic data to enable cellular-resolution mapping of metabolic spatial patterns. These advancements would substantially strengthen the ability of MMEASE to address fundamental mechanistic questions. A local version of MMEASE was further constructed, which can be downloaded to and run on a user's computer. To install this local version, three sequential steps should be followed: first, install the R and RStudio environment; second, install the MMEASE package from GitHub; third, run the package in RStudio via executing the R commands described in the User Manual. Exemplar inputs and output files can be downloaded directly from the MMEASE package at: https://idrblab.org/mmease/

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

None declared.

Data availability

MMEASE is open to all users without a login requirement and is freely available at https://idrblab.org/mmease/

References

1. Seydel C. Single-cell metabolomics hits its stride. *Nat Methods* 2021;**18**:1452–6. https://doi.org/10.1038/s41592-021-01333-x
2. Gentry EC, Collins SL, Panitchpakdi M *et al.* Reverse metabolomics for the discovery of chemical structures from humans. *Nature* 2024;**626**:419–26. https://doi.org/10.1038/s41586-023-06906-8
3. Huang H, Chen Y, Xu W *et al.* Decoding aging clocks: new insights from metabolomics. *Cell Metab* 2025;**37**:34–58. https://doi.org/10.1016/j.cmet.2024.11.007
4. Pang Z, Lu Y, Zhou G *et al.* MetaboAnalyst 6.0: towards a unified platform for metabolomics data processing, analysis and interpretation. *Nucleic Acids Res* 2024;**52**:W398–406. https://doi.org/10.1093/nar/gkae253
5. Pang Z, Zhou G, Ewald J *et al.* Using MetaboAnalyst 5.0 for LC-HRMS spectra processing, multi-omics integration and covariate adjustment of global metabolomics data. *Nat Protoc* 2022;**17**:1735–61. https://doi.org/10.1038/s41596-022-00710-w
6. Yang Q, Li B, Chen S *et al.* MMEASE: online meta-analysis of metabolomic data by enhanced metabolite annotation, marker selection and enrichment analysis. *J Proteomics* 2021;**232**:104023. https://doi.org/10.1016/j.jprot.2020.104023
7. Wang Y, Liu X, Dong L *et al.* iMSEA: a novel metabolite set enrichment analysis strategy to decipher drug interactions. *Anal Chem* 2023;**95**:6203–11. https://doi.org/10.1021/acs.analchem.2c04603
8. Zhang Y, Zhou Y, Zhou Y *et al.* TheMarker: a comprehensive database of therapeutic biomarkers. *Nucleic Acids Res* 2024;**52**:D1450–64. https://doi.org/10.1093/nar/gkad862
9. Yang Q, Chen S, Jiang W *et al.* MultiClassMetabo: a superior classification model constructed using metabolic markers in multiclass metabolomics. *Anal Chem* 2024;**96**:1410–8. https://doi.org/10.1021/acs.analchem.3c03212
10. Sheridan C. Can single-cell biology realize the promise of precision medicine? *Nat Biotechnol* 2024;**42**:159–62. https://doi.org/10.1038/s41587-024-02138-x
11. Eisenstein M. Seven technologies to watch in 2023. *Nature* 2023;**613**:794–7. https://doi.org/10.1038/d41586-023-00178-y
12. Rappez L, Stadler M, Triana S *et al.* SpaceM reveals metabolic states of single cells. *Nat Methods* 2021;**18**:799–805. https://doi.org/10.1038/s41592-021-01198-0
13. Li P, Gao S, Qu W *et al.* Chemo-selective single-cell metabolomics reveals the spatiotemporal behavior of exogenous pollutants during *Xenopus laevis* embryogenesis. *Adv Sci* 2024;**11**:e2305401. https://doi.org/10.1002/advs.202305401
14. Cao J, Yao QJ, Wu J *et al.* Deciphering the metabolic heterogeneity of hematopoietic stem cells with single-cell resolution. *Cell Metab* 2024;**36**:209–21. https://doi.org/10.1016/j.cmet.2023.12.005
15. Zhang W, Xu F, Yao J *et al.* Single-cell metabolic fingerprints discover a cluster of circulating tumor cells with distinct metastatic potential. *Nat Commun* 2023;**14**:2485. https://doi.org/10.1038/s41467-023-38009-3
16. Hu T, Allam M, Cai S *et al.* Single-cell spatial metabolomics with cell-type specific protein profiling for tissue systems biology. *Nat*

Commun 2023;**14**:8260.
https://doi.org/10.1038/s41467-023-43917-5

17. Hu R, Li Y, Yang Y *et al.* Mass spectrometry-based strategies for single-cell metabolomics. *Mass Spectrom Rev* 2023;**42**:67–94. https://doi.org/10.1002/mas.21704

18. Zhang C., Le Devedec SE, Ali A *et al.* Single-cell metabolomics by mass spectrometry: ready for primetime? *Curr Opin Biotechnol* 2023;**82**:102963. https://doi.org/10.1016/j.copbio.2023.102963

19. Pan X, Pan S., Du M *et al.* SCMeTA: a pipeline for single-cell metabolic analysis data processing. *Bioinformatics* 2024;**40**:btae545. https://doi.org/10.1093/bioinformatics/btae545

20. Liu R, Li J, Lan Y *et al.* Quantifying cell heterogeneity and subpopulations using single cell metabolomics. *Anal Chem* 2023;**95**:7127–33. https://doi.org/10.1021/acs.analchem.2c05245

21. Eraslan G, Simon LM, Mircea M *et al.* Single-cell RNA-seq denoising using a deep count autoencoder. *Nat Commun* 2019;**10**:390. https://doi.org/10.1038/s41467-018-07931-2

22. Amodio M, van Dijk D, Srinivasan K *et al.* Exploring single-cell data with deep multitasking neural networks. *Nat Methods* 2019;**16**:1139–45. https://doi.org/10.1038/s41592-019-0576-7

23. Liu R, Zhang G, Sun M *et al.* Integrating a generalized data analysis workflow with the single-probe mass spectrometry experiment for single cell metabolomics. *Anal Chim Acta* 2019;**1064**:71–9. https://doi.org/10.1016/j.aca.2019.03.006

24. Xu T, Li H, Dou P *et al.* Concentric hybrid nanoelectrospray ionization–atmospheric pressure chemical ionization source for high-coverage mass spectrometry analysis of single-cell metabolomics. *Adv Sci* 2024;**11**:e2306659. https://doi.org/10.1002/advs.202306659

25. Misra BB. Open-source software tools, databases, and resources for single-cell and single-cell-type metabolomics. *Methods Mol Biol* 2020;**2064**:191–217. https://doi.org/10.1007/978-1-4939-9831-9_15

26. Yu H, Xing S, Nierves L *et al.* Fold-change compression: an unexplored but correctable quantitative bias caused by nonlinear electrospray ionization responses in untargeted metabolomics. *Anal Chem* 2020;**92**:7011–9. https://doi.org/10.1021/acs.analchem.0c00246

27. Chen Y, Li EM, Xu LY. Guide to metabolomics analysis: a bioinformatics workflow. *Metabolites* 2022;**12**:357. https://doi.org/10.3390/metabo12040357

28. Forsgren E, Bjorkblom B, Trygg J *et al.* OPLS-based multiclass classification and data-driven interclass relationship discovery. *J Chem Inf Model* 2025;**65**:1762–70. https://doi.org/10.1021/acs.jcim.4c01799

29. Gromski PS, Muhamadali H, Ellis DI *et al.* A tutorial review: metabolomics and partial least squares-discriminant analysis—a marriage of convenience or a shotgun wedding. *Anal Chim Acta* 2015;**879**:10–23. https://doi.org/10.1016/j.aca.2015.02.012

30. Mayer J, Rahman R, Ghosh S *et al.* Sequential feature selection and inference using multi-variate random forests. *Bioinformatics* 2018;**34**:1336–44. https://doi.org/10.1093/bioinformatics/btx784

31. Bahado-Singh RO, Syngelaki A, Akolekar R *et al.* Validation of metabolomic models for prediction of early-onset preeclampsia. *Am J Obstet Gynecol* 2015;**213**:530. https://doi.org/10.1016/j.ajog.2015.06.044

32. Abenavoli A, Pisa S, Maggiani A. A pilot study of jugular compression (Queckenstedt maneuver) for cranial movement perception. *J Am Osteopath Assoc* 2020;**120**:647–54. https://doi.org/10.7556/jaoa.2020.119

33. She H, Tan L, Wang Y *et al.* Integrative single-cell RNA sequencing and metabolomics decipher the imbalanced lipid-metabolism in maladaptive immune responses during sepsis. *Front Immunol* 2023;**14**:1181697. https://doi.org/10.3389/fimmu.2023.1181697

34. Tian H, Ni Z, Lam SM *et al.* Precise metabolomics reveals a diversity of aging-associated metabolic features. *Small Methods* 2022;**6**:e2200130. https://doi.org/10.1002/smtd.202200130

35. Lee HJ, Kremer DM, Sajjakulnukit P *et al.* A large-scale analysis of targeted metabolomics data from heterogeneous biological samples provides insights into metabolite dynamics. *Metabolomics* 2019;**15**:103. https://doi.org/10.1007/s11306-019-1564-8

36. Dou L, Li X, Zhang L *et al.* iGlu_AdaBoost: identification of lysine glutarylation using the AdaBoost classifier. *J Proteome Res* 2021;**20**:191–201. https://doi.org/10.1021/acs.jproteome.0c00314

37. Datta S, Pihur V, Datta S. An adaptive optimal ensemble classifier via bagging and rank aggregation with applications to high dimensional data. *BMC Bioinformatics* 2010;**11**:427. https://doi.org/10.1186/1471-2105-11-427

38. Ye S, Hu W, Li X *et al.* A neural network protocol for electronic excitations of N-methylacetamide. *Proc Natl Acad Sci USA* 2019;**116**:11612–7. https://doi.org/10.1073/pnas.1821044116

39. Wang Y, Pan Z, Pan Y. A training data set cleaning method by classification ability ranking for the k-nearest neighbor classifier. *IEEE Trans Neural Netw Learn Syst* 2020;**31**:1544–56. https://doi.org/10.1109/TNNLS.2019.2920864

40. Ye Q, Fu L, Zhang Z *et al.* Lp- and ls-norm distance based robust linear discriminant analysis. *Neural Netw* 2018;**105**:393–404. https://doi.org/10.1016/j.neunet.2018.05.020

41. Lee SY, Mediani A, Maulidiani M *et al.* Comparison of partial least squares and random forests for evaluating relationship between phenolics and bioactivities of *Neptunia oleracea. J Sci Food Agric* 2018;**98**:240–52. https://doi.org/10.1002/jsfa.8462

42. de Santana FB, Borges Neto W, Poppi RJ. Random forest as one-class classifier and infrared spectroscopy for food adulteration detection. *Food Chem* 2019;**293**:323–32. https://doi.org/10.1016/j.foodchem.2019.04.073

43. Nedaie A, Najafi AA. Support vector machine with Dirichlet feature mapping. *Neural Netw* 2018;**98**:87–101. https://doi.org/10.1016/j.neunet.2017.11.006

44. Qin S, Zhang Y, Shi M *et al.* In-depth organic mass cytometry reveals differential contents of 3-hydroxybutanoic acid at the single-cell level. *Nat Commun* 2024;**15**:4387. https://doi.org/10.1038/s41467-024-48865-2

45. Wishart DS, Guo A, Oler E *et al.* HMDB 5.0: the human metabolome database for 2022. *Nucleic Acids Res* 2022;**50**:D622–31. https://doi.org/10.1093/nar/gkab1062

46. Horai H, Arita M, Kanaya S *et al.* MassBank: a public repository for sharing mass spectral data for life sciences. *J Mass Spectrom* 2010;**45**:703–14. https://doi.org/10.1002/jms.1777

47. Kind T, Liu KH, Lee DY *et al.* LipidBlast in silico tandem mass spectrometry database for lipid identification. *Nat Methods* 2013;**10**:755–8. https://doi.org/10.1038/nmeth.2551

48. Wang M, Carver JJ, Phelan VV *et al.* Sharing and community curation of mass spectrometry data with global natural products social molecular networking. *Nat Biotechnol* 2016;**34**:828–37. https://doi.org/10.1038/nbt.3597

49. Kanehisa M, Furumichi M, Sato Y *et al.* KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res* 2023;**51**:D587–92. https://doi.org/10.1093/nar/gkac963

50. Xing S, Shen S, Xu B *et al.* BUDDY: molecular formula discovery via bottom-up MS/MS interrogation. *Nat Methods* 2023;**20**:881–90. https://doi.org/10.1038/s41592-023-01850-x

51. Pang Z, Xu L, Viau C *et al.* MetaboAnalystR 4.0: a unified LC-MS workflow for global metabolomics. *Nat Commun* 2024;**15**:3675. https://doi.org/10.1038/s41467-024-48009-6

52. Wishart D, Arndt D, Pon A *et al.* T3DB: the toxic exposome database. *Nucleic Acids Res* 2015;**43**:D928–34. https://doi.org/10.1093/nar/gku1004

53. Knox C, Wilson M, Klinger CM *et al.* DrugBank 6.0: the DrugBank knowledgebase for 2024. *Nucleic Acids Res* 2024;**52**:D1265–75. https://doi.org/10.1093/nar/gkad976

54. Zhou Y, Zhang Y, Zhao D *et al.* TTD: therapeutic target database describing target druggability information. *Nucleic Acids Res* 2024;**52**:D1465–77. https://doi.org/10.1093/nar/gkad751