Mini Review

Bioinformatics and data science for mass spectrometry data analysis

Kei Zaitsu[1]*, Seiichiro Eguchi[2], Akira Iguchi[3, 4]

[1]Multimodal Informatics and Wide-data Analytics Laboratory (MiWA-Lab.), Department of Computational Systems Biology,
Faculty of Biology-Oriented Science and Technology, Kindai University, 930 Nishi Mitani, Kinokawa, Wakayama 649−6493, Japan
[2]Department of Neurosurgery, Tokyo Women's Medical University, 8−1 Kawada-cho Shinjuku-ku, Tokyo 162−8666, Japan
[3]Geological Survey of Japan, National Institute of Advanced Industrial Science and Technology (AIST),
AIST Tsukuba Central 7, 1−1−1 Higashi, Tsukuba, Ibaraki 305−8567, Japan
[4]Research laboratory on environmentally-conscious developments and technologies [E-code],
National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki 305−8567, Japan

Abstract Mass spectrometry data contributing to proteome and metabolome analyses play an important role in omics sciences. Moreover, the integration of multilayer data called trans-omics has steadily become popular in life sciences, increasing the significance of bioinformatics and data science. This minireview, therefore, outlines commonly used bioinformatics, and data science tools, mainly based on our previous studies using mass spectrometry data. Here, we not only introduce some bioinformatics platforms and data pipelines but also provide a concise explanation of multivariate correlation network analyses and machine learning as well as time-series data analysis. Finally, future perspectives on applying bioinformatics and data sciences to mass spectrometry data are outlined.

Key words: bioinformatics, data science, omics, data-driven analysis

Introduction

Recent developments in omics technologies have made the acquisition of comprehensive bioinformation from various biological samples and/or organisms much easier than ever[1)]. Transcriptome analysis via RNA sequencing (RNA-seq) has been widely used in various scientific fields, allowing researchers to profile expression patterns of tens of thousands of genes[2)]. Moreover, this technology has been extended to single-cell RNA-seq (sc-RNA-seq) for elucidating the heterogeneity of cells, and studies involving sc-RNA-seq have dramatically increased due to its innovativeness in cell sorting and high-throughput sequencing[3–5)]. However, mass spectrometry still plays a pivotal role in omics techniques given its ability for proteome and/or metabolome profiling. In addition, the integration of multilayer omics data called trans-omics analyses has gradually become popular in life sciences, providing researchers access to composite multivariate data[1)].

To unveil the latent relationship among such multivariate data or grasp data structures, implementing bioinformatics, and data sciences to your own data is mandatory, for which numerous visualizing, and/or efficient bioinformatics tools have been available for free[6–8)]. Owing to the widespread use of open-source coding languages, such as R, and Python, implementing, and/or customizing such valuable bioinformatics tools has never been more easier. In particular, data-driven data analyses, such as multivariate and network analyses, are considerably helpful for discovering significant explanatory variables from complicated omics data without special presumptions. For instance, correlation network analysis can successfully estimate the relationship among explanatory variables in omics data[9–12)] and detect

*Corresponding author
Kei Zaitsu
Multimodal Informatics and Wide-data Analytics Laboratory (MiWA-Lab.), Department of Computational Systems Biology, Faculty of Biology-Oriented Science and Technology, Kindai University, 930 Nishi Mitani, Kinokawa, Wakayama 649−6493, Japan
Tel: +81−736−77−0345 (ext. 4232)
E-mail: kzaitsu@waka.kindai.ac.jp

Received: January 19, 2023. Accepted: March 7, 2023.
Epub April 7, 2023.
DOI: 10.24508/mms.2023.06.006

communities in the targeted biological network. Moreover, centrality analyses, such as a betweenness centrality analysis, can elucidate hub molecules in a biological network, which are bottlenecks in the targeted network[9]. Machine learning techniques, such as random forest (RF) and support vector machine (SVM), have also become popular for constructing discriminant or regression models of multivariate data, and these techniques are often used for selecting significant molecules (i.e., potential biomarkers) for discriminating groups or mechanisms[13].

Mathematical and computational modeling techniques have been used for estimating credible intervals (CI) instead of confidence intervals through classical statistical analyses[14,15]. For instance, time-series data analyses, including Bayesian state space modeling, are applied to time-series data obtained through mass spectrometry given that it is fundamentally inappropriate to apply classical statistical analysis to time-series data[14].

As described earlier, the significance of bioinformatics, and data sciences on mass spectrometry multivariate data is expected to increase more than ever, making the outlining of commonly used bioinformatics tools for mass spectrometry data highly meaningful. This minireview, therefore, introduces representative bioinformatics tools that have been applied to mass spectrometry data, especially metabolome data, mainly based on our previous studies[10–12,14,16,17]. We also address bioinformatics platforms and data pipelines, the application of correlation network analysis and machine learning to metabolome data, and mathematical, and computational modeling for time-series data. Finally, future perspectives on applying bioinformatics and data sciences to mass spectrometry data are outlined.

1. Overview of Bioinformatics Platforms and Data Pipelines

To date, numerous bioinformatics tools for mass spectrometry multivariate data have been reported, and previous review articles are helpful in outlining such tools[17–28]. Prior to bioinformatics analyses, normalization is generally performed for mass spectrometry data because matrix effects (i.e., ionization enhancement or ionization suppression) can disrupt the quantitativity in mass spectrometry. There are various normalization techniques such as internal standard (IS) normalization and locally weighted scatterplot smoothing (LOESS)-based normalization methods for mass spectrometry data[10]. As far as we know, IS normalization method using stable isotopes of targeted compounds has still been commonly used for MS-based metabolomics. Our group has also reported that a total intensity-based normalization method was the optimal inter-batch normalization technique for gas chromatography/tandem mass spectrometry-based metabolomics[10]. For scaling of mass spectrometry data, unit variance scaling is commonly used, while pareto-scaling is often used for orthogonal projection to latent structures discriminant analysis (OPLS-DA)[16,37].

Among bioinformatics platforms and data pipelines for metabolome data, MetaboAnalyst (the latest version 5.0), a web-based comprehensive platform developed by Wishart et al.[28], might be the most popular. Similar to other bioinformatics platforms, MetaboAnalyst is user-friendly, and can perform multivariate analyses, such as principal component analysis (PCA), projection to latent structures discriminant analysis (PLS-DA), and OPLS-DA. This platform also allows for analysis of variance (ANOVA), metabolite set enrichment analysis, metabolic pathway analysis, and multivariate receiver operating characteristic (ROC) curve analysis based on PLS-DA, SVM, or RF.

In omics studies, multiple comparisons commonly cause problems with the analysis of statistical significance[29]. Therefore, appropriate correction methods, such as Bonferroni and false discovery rate (FDR) correction methods[30,31], are generally required for multiple comparisons. Bonferroni correction method adjust familywise error rate, while FDR correction method literally control false discovery rate, where q-values are generally used instead of p-values with FDR correction. Our previous study demonstrated that the newly developed R-based platform called PiTMaP[17] can automatically perform two- or multi-pair statistical analyses with FDR correction for multivariate data. In the PiTMaP algorithm, PLS-DA is used for selecting variables that are subjected to significant analyses, allowing for the objective selection of variables for significant analysis. PiTMaP also automatically generates box-and-whisker plots for all explanatory variables and score and loading plots for PCA and PLS-DA, in addition to automatically drawing box-and-whisker plots for significantly altered variables. Fig. 1 shows some parts of the results obtained automatically within 30 s via PiTMaP when applied to hepatic metabolome data from mouse models of acetaminophen-induced liver-injury and control mice. Such bioinformatics data pipelines can dramatically improve the efficiency of multivariate data analysis considering their ability to auto-

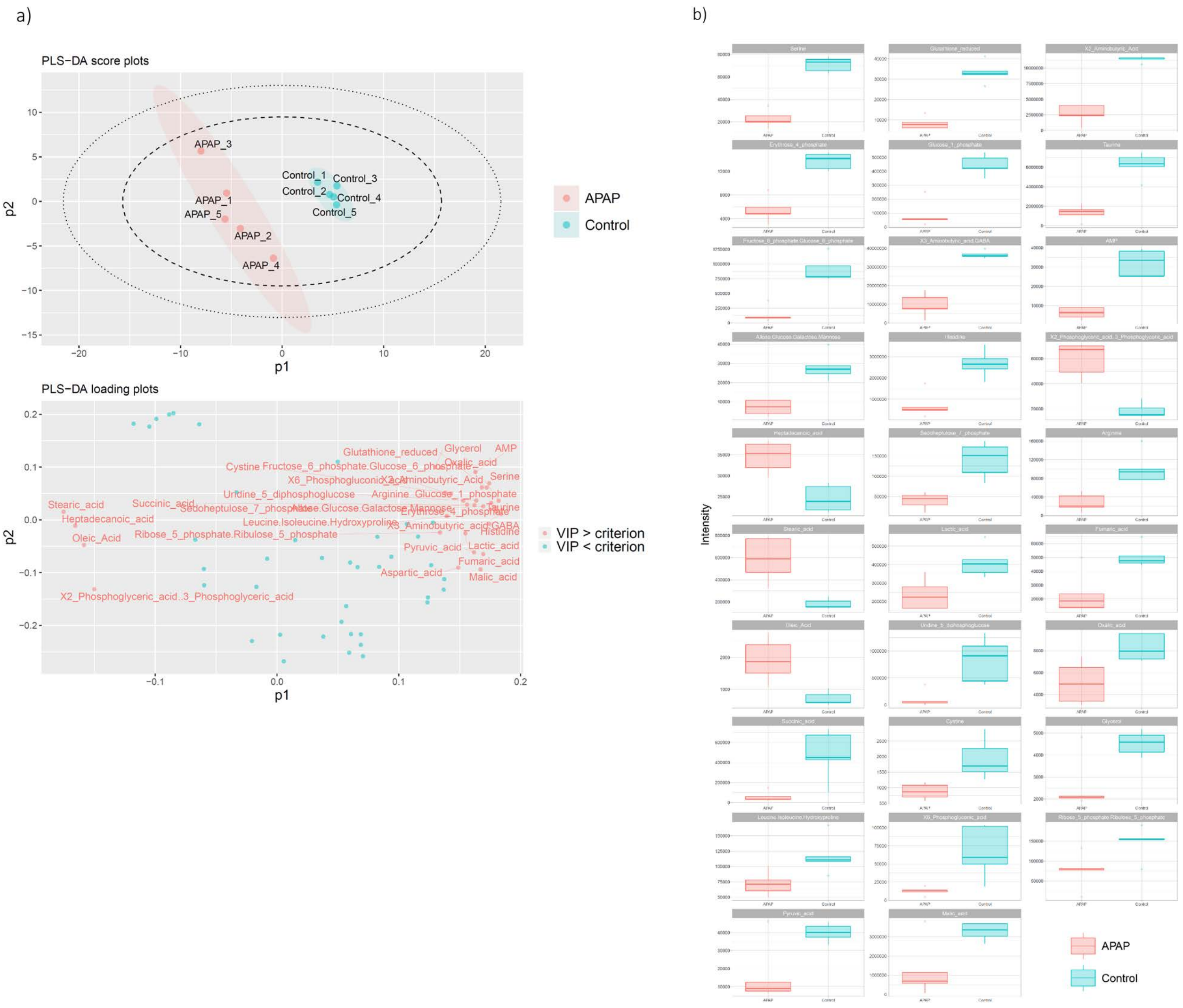

Fig. 1. Results automatically created by PiTMaP.
(a) PLS-DA score and loading plots for the control and acetaminophen (APAP)-induced liver injury model mice with a VIP criterion of 1.0. Red: APAP model mice and blue green: control mice in PLS-DA score plots. Dotted and solid circle in the score plots show 95 and 99% confidence intervals of all plots, respectively. Colored circles show 95% confidence intervals of each cohort, and (b) box-and-whisker plots for significantly altered metabolites. Reprinted with permission from *Anal Chem* 92(12): 8514–8522, 2020. Copyright @ 2020 American Chemical Society.

matically and objectively provide results. Our team has been improving PiTMaP to develop a web application that implements other bioinformatics tools to achieve high-efficiency bioinformatics data analyses.

2. Data-driven Bioinformatics

2.1. Correlation network analysis

As mentioned earlier, data-driven bioinformatics is essential for identifying potential biomarkers and/or influential hub molecules in targeted biological networks. Correlation network analysis is based on pairwise Peason's or Spearman's correlation coefficients, and it is widely applied not only to transcriptome, proteome, and metabolome data but also to metagenome data[9,11,32,33]. Moreover, the trans-omics approach strongly depends on data-driven network analysis. Zhou et al. reported a comprehensive web-based platform for multi-omics called OmicsAnalyst, where a correlation network analysis is applicable for multi-omics data[7]. Correlation network analysis can extract hub molecules, which are the most influential variables in the targeted biological networks. To identify such hub molecules, centrality analysis of network has been commonly used[9–12]. There are several types of centrality indices, namely betweenness centrality (BC), degree centrality (DC), closeness centrality (CC), and eigenvector centrality (EC). DC is the simplest centrality index, and the high DC value simply means that

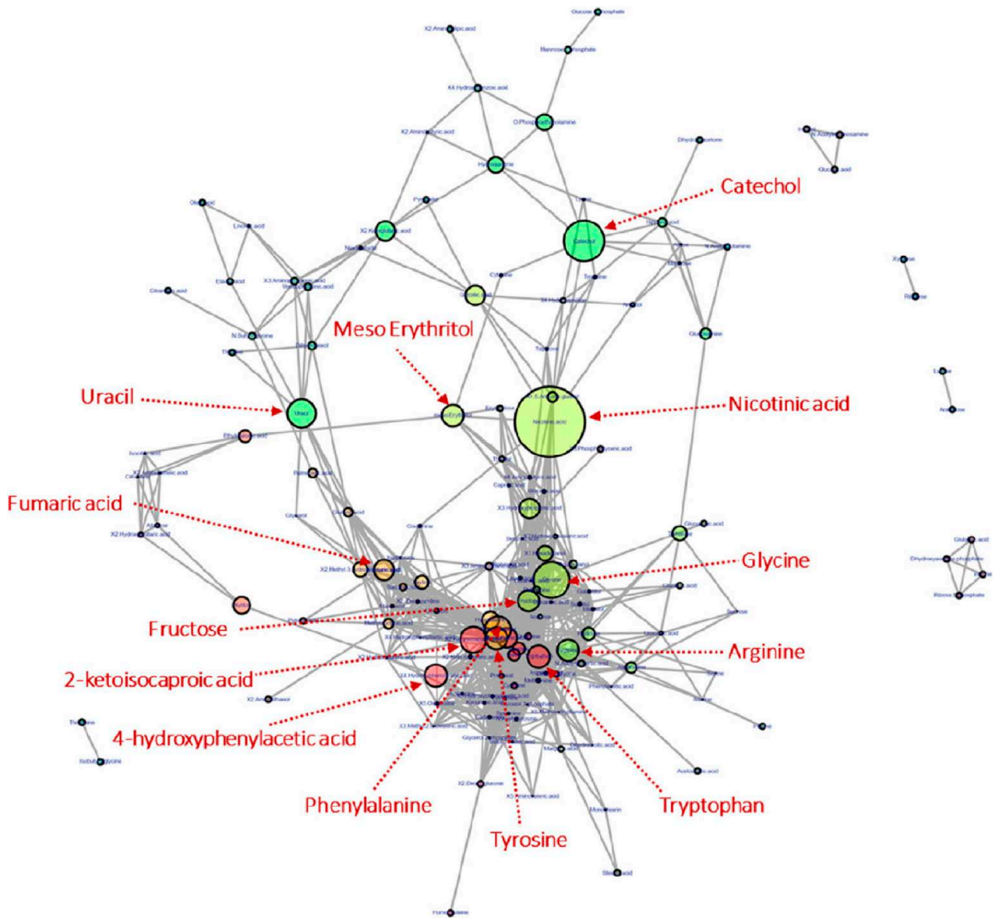

Fig. 2. Network analysis results for the blood metabolome of maternal mice exposed to a phthalate during pregnancy. The criterion was set at $R>0.75$, and the size of each node (circle) was proposed to the corresponding betweenness centrality (BC) values. Each color (i.e., red, green, yellow and etc.) indicates the clusters determined by the hierarchical cluster analysis (Reprinted from *ACS Omega* 7(27): 23717–23726, 2022).

the molecule with high DC value have many interactions with other molecules. EC is an expansion of DC, and it is defined by total sum of neighbor molecules' centralities. BC and CC are based on the shortest paths. CC considers the molecule with the shortest paths as the hub molecule, while BC considers the molecule that is the most frequently passed through in the shortest paths as the hub molecule. Centrality analysis can rank explanatory variables in a targeted network according to importance via the centrality index, with variables having high centrality values literally being at the center of the targeted network.

Our previous studies showed that correlation network analyses could successfully extract hub molecules in the metabolome network (Figs. 2 and 3). In our studies, betweenness centrality was commonly used for extracting the hub molecules. In Figs. 2 and 3, the size of each node (circle) was proposed to the corresponding betweenness centrality values, and thus, bigger size of circle visually demonstrates the importance of hub molecules in the targeted network. Each color of node also indicates the clusters determined by the hierarchical cluster analysis. Fig. 2 shows the hub metabolites in the blood metabolome of maternal mice exposed to a phthalate during pregnancy, and these metabolites could be associated with fetal lethality via phthalate exposure[11]. Fig. 3 also shows hub molecules in HepG2 cells exposed to some mitochondrial toxicants, and these hub molecules could be molecular indicators for discriminating mitochondrial toxicity mechanisms[12]. These results demonstrated that a correlation network analysis could be a powerful tool for unveiling hidden key factors in a targeted biological network. However, given that pseudo-correlations could potentially exist during correlation network analysis, we need to pay close attention to the biological interpretation of the results.

To draw such correlation-based networks, setting a threshold for the selection of correlated variables is neces-

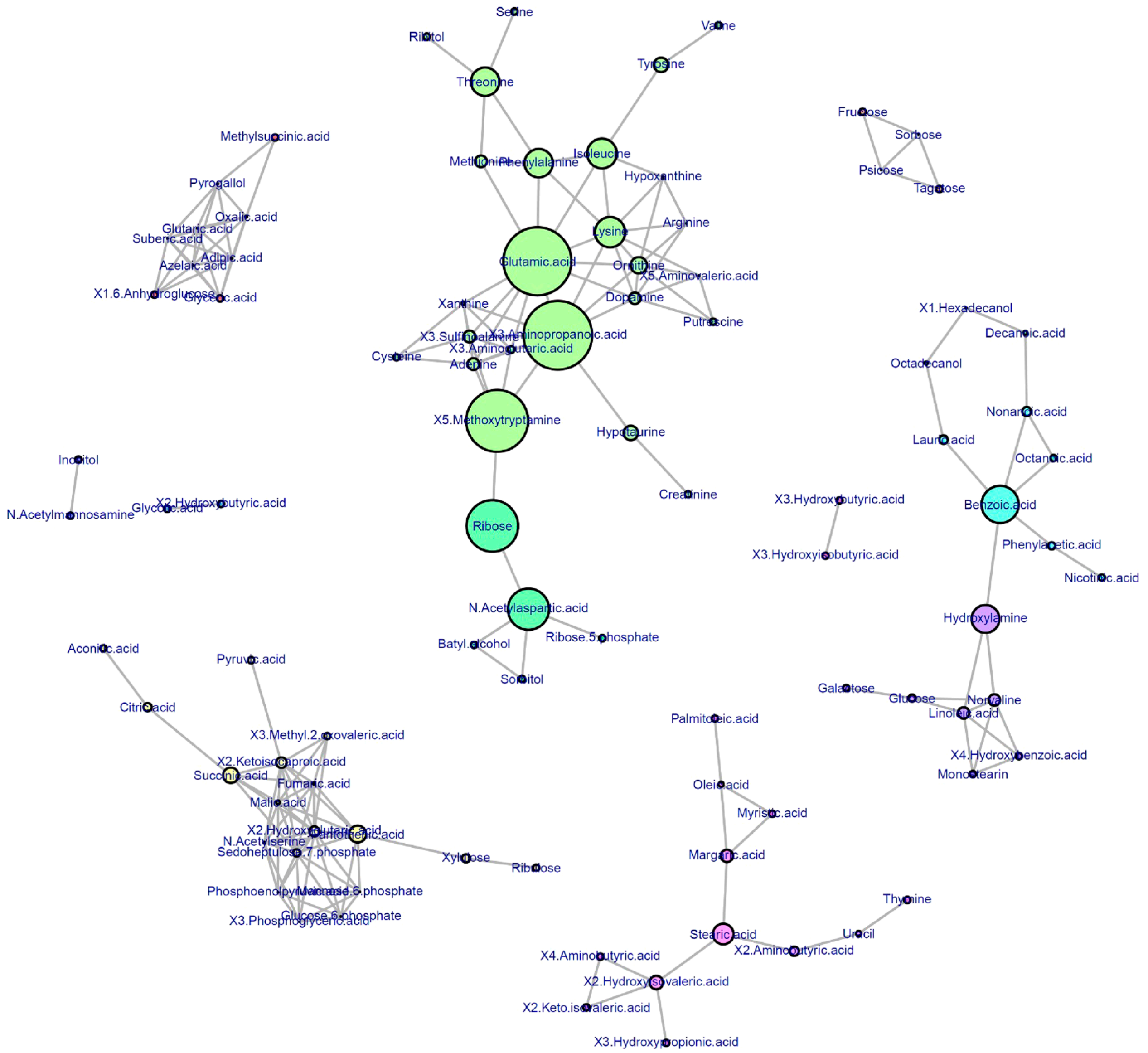

Fig. 3. Network analysis results for metabolome data obtained from HepG2 cells exposed to some mitochondrial toxicants. The criterion was set at $R>0.75$, and the size of each node (circle) was proposed to the corresponding betweenness centrality (BC) values Hub-metabolites were determined by BC values. Each color (i.e., purple, yellow green and etc.) indicates the clusters determined by the hierarchical cluster analysis (Reprinted with permission from *Toxicol Appl Pharmacol* 457(15): 116316, 2022).

sary. To date, $r>0.75$–0.80 has been the commonly accepted threshold for correlation network analysis. Though this threshold has been experimentally accepted by many researchers, the use of such values still remains arbitrary. In fact, correlation coefficient-based thresholds could be theoretically determined. In line with this, Langfelder and Horvath demonstrated weighted correlation network analysis (WGCNA)[34], which can keep overall information of correlation coefficients while reducing false positives; however, given its somewhat tedious implementation, its applications to omics data have been limited at this time.

2.2. *Machine learning*

In medical and medicinal sciences, one of the most interesting things for omics data is searching for potential biomarkers that can discriminate disease groups from healthy control groups or predict diseases and/or their levels. For this purpose, differentially expressed gene (DEG) analysis had generally been performed for transcriptome data.[2,35] For instance, Robinson et al. reported that the R-based bioinformatics tool called edgeR can create MA-plot, an application of a Bland–Altman plot for gene data, and visualize DEGs[35]. For metabolome data, multivariate analyses, such as PLS-DA, and OPLS-DA, as well as PCA-regression and

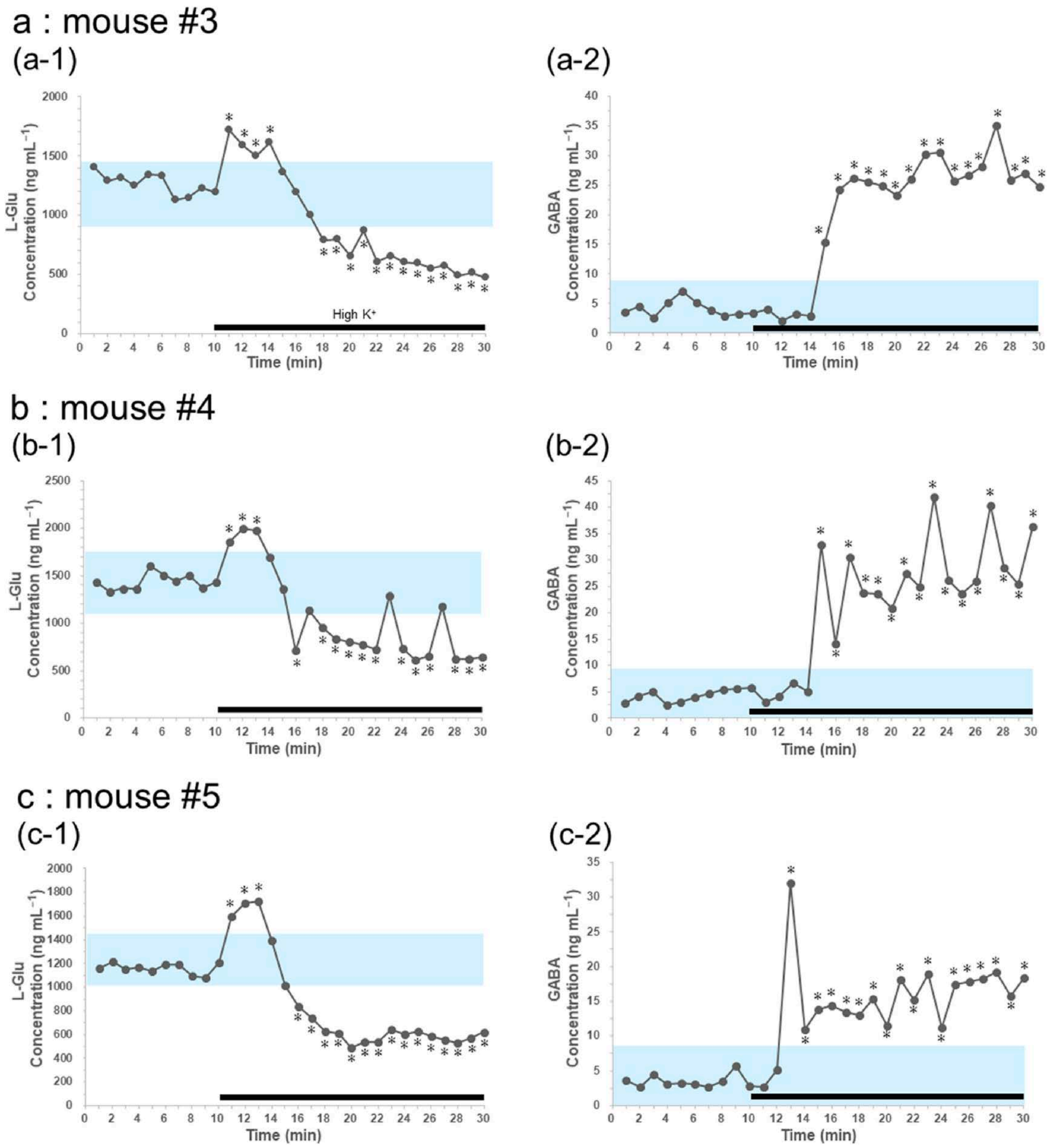

Fig. 4. Time-series changes in (a-1, b-1, and c-1) L-glutamic acid (L-Glu) and (a-2, b-2, and c-2) gamma-aminobutyric acid (GABA) levels in the microdialysate every 1 min, and their credible intervals (blue band) estimated by Bayesian state-space model using the steady state levels.

The model could detect significant deviation (*) of the neurotransmitters via depolarization induced by replacing high-potassium containing artificial cerebrospinal fluid (Reprinted with permission from *Talanta* 234(1): 122620, 2021).

PLS-regression (PLSR), have been commonly used[36]. In our previous studies, PLSR was used for predicting biological states using metabolome data[16, 37].

In addition to these commonly-used approaches, machine learning, such as RF, and SVM, have recently been applied to omics-data to construct discriminant and prediction models[38,39] given that these methods can be easily implemented by R or Python more than initially expected. In our previous studies, RF, and ROC curves were used for selecting important metabolites that can discriminate groups, where we programmed that hyperparameters of RF (e.g., mtry) were automatically tuned and area under curve by each ROC curve were automatically calculated, determining the most important variables for discriminating groups automatically[11,12]. This will facilitate the application of machine learning to omics data.

2.3. *Time-series data analysis*

Mathematical and computational modeling techniques have been used for understanding biological data[6,14,15,40].

Sriyudthsak et al. reviewed mathematical modeling and dynamic simulation of metabolic reaction systems for time series data of metabolome[15]. Hirai and Shiraishi also reported on the mathematical modeling of plant metabolic systems for time-series data[40].

Unlike the aforementioned cross-sectional data (i.e., one time-point data), time-series data contain trends, cycle, irregularity, and seasonality; thus, appropriate modeling methods, such as autoregressive, autoregressive moving average, and state space models, are generally required for understanding time-series data. In our previous report, we applied a Bayesian state-space model to time-series data obtained from combinational use of in vivo microdialysis and ambient ionization mass spectrometry[14]. In this study, quantitative time-series data of neurotransmitters (GABA and glutamate) were obtained from each living mouse brain via an in vivo microdialysis technique. Conventionally, time-series data obtained from microdialysis were statistically analyzed by applying the t-test and/or ANOVA to each time point, although such statistical tests are fundamentally inappropriate for analyzing time-series data given their autocovariance (autocorrelation) and periodicity. In our study, therefore, we implemented the Bayesian state-space model using the R and Stan software and applied the model to estimate the CI based on the steady state (initial state) levels of the neurotransmitters[14]. Here, we used the following equation 1 to estimate the CI range.

$$\text{pred}\ (t\text{+i})\text{=Nomal}\ [2^{*}\mu_{\text{pred}}\ (t\text{+i} - 1)\ \text{-}\ \mu_{\text{pred}}\ (t\text{+}i - 2), \sigma] \qquad \text{(Eq. 1)}$$

In this equation, t is time, i is difference, pred(t+i) is a prediction value, μ_{pred} (t) is level component at t, and σ is standard deviation of process error. We ran 4 chains of 8,000 from the posterior distribution and discarded the first 2,000 ones before inference. Model diagnostics were also confirmed using Rhat values. As shown in Fig. 4, the Bayesian state-space model successfully estimated the CI range of the neurotransmitters from their steady state, and could detect significant deviations in the neurotransmitters via depolarization induced by replacing the high-potassium-containing artificial cerebrospinal fluid. This approach enables us to evaluate variations in the target biomolecules within the time-series data without testing for statistical significance. In addition, our group has developed in vivo real-time monitoring techniques for metabolites in mouse tissues using ambient ionization mass spectrometry[41–43], which can more easily provide time-series data. Now, we apply a time-series analysis to real-time monitoring data obtained by this technique. Our successful demonstration of a platform for such time-series data would certainly increase the value of ambient ionization mass spectrometry.

3. Future Perspectives

As mentioned earlier, researchers have already been able to access big omics data, and the significance of bioinformatics and data science has been steadily increasing. Moreover, diverse data obtained via multimodal analytical methods will be combined in near future. For instance, Zhu et al. reported on the application of single-cell multimodal omics[44], where they emphasized "the power of many." Bredikhin et al. also reported on the multimodal omics analysis framework called MUON[8]. Thus, mass spectrometric information, which is positioned at an intermediate level of genomic and phenotypic information in multi-omics, is expected to become more a significant part of multimodality in terms of bridging each level. In addition to analytical multimodality, spatiotemporal information is expected to become more valuable in the future[45, 46]. Thus, more focus will be placed on single cell mass spectrometry and spatiotemporal mass spectrometry[47, 48]. Multimodality and spatiotemporal information will be a key factor in elucidating hidden molecular mechanisms, with such "wide data" requiring further improvements in bioinformatics and data science tools.

4. Conclusions

In this mini review, we outlined the commonly-used bioinformatics and data science tools for mass spectrometry data, mainly based on our previous studies. To date, several user-friendly platforms/data pipelines have been reported, with their usability being continuously improved. We demonstrated that multivariate and correlation network analyses have been useful for understanding omics data and that machine learning, such as RF, and SVM, is expected to become more popular for discriminant and/or prediction analyses of omics data. Also, mathematical, and computational modeling techniques are required for analyzing time-series data. Given the feasibility of implementation by freely available coding languages such as R and Python, bioinformatics and data sciences have already been mandatory for any scientists, and these techniques can provide valuable information to elucidate hidden molecular mechanisms in organisms. Finally, future perspectives suggest that mass spectrometric multivariate data will play a more important role in multimodal and spatiotemporal information in the future.

Acknowledgements

We are very grateful to Dr. Tomomi Asano, Dr. Masaru Taniguchi, Dr. Daisuke Kawakami, Dr. Yui Hibino, and Mr. Kazuaki Hisatsune for their contribution to our research team. We would like to thank anonymous reviewers who strongly improved this manuscript with the feedback and comments.

Conflicts of Interest

There are no conflicts of interest to declare.

References

1) Karczewski KJ, Snyder MP: Integrative omics for health and disease. *Nat Rev Genet* 19: 299–310, 2018.

2) Wang Z, Gerstein M, Snyder M: RNA-Seq: A revolutionary tool for transcriptomics. *Nat Rev Genet* 10: 57–63, 2009.

3) Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, et al: Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* 344: 1396–1401, 2014.

4) Kolodziejczyk AA, Kim JK, Svensson V, Marioni JC, Teichmann SA: The technology and biology of single-cell RNA sequencing. *Mol Cell* 58: 610–620, 2015.

5) Hwang B, Lee JH, Bang D: Single-cell RNA sequencing technologies and bioinformaticspipelines. *Exp Mol Med* 50: 1–14, 2018.

6) Bersanelli M, Mosca E, Remondini D, Giampieri E, Sala C, et al: Methods for the integration of multi-omics data: Mathematical aspects. *BMC Bioinformatics* 17: S15, 2016.

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

8) Bredikhin D, Kats I, Stegle O: MUON: multimodal omics analysis framework. *Genome Biol* 23: 42, 2022.

9) Yu H, Kim PM, Sprecher E, Trifonov V, Gerstein M: The importance of bottlenecks in protein networks: Correlation with gene essentiality and expression dynamics. *PLoS Comput Biol* 3: e59, 2007.

10) Zaitsu K, Noda S, Iguchi A, Hayashi Y, Ohara T, et al: Metabolome analysis of the serotonin syndrome rat model: Abnormal muscular contraction is related to metabolic alterations and hyper-thermogenesis. *Life Sci* 207: 550–561, 2018.

11) Zaitsu K, Asano T, Kawakami D, Chang J, Hisatsune K, et al: Metabolomics and data-driven bioinformatics revealed key maternal metabolites related to fetal lethality via Di(2-ethylhexyl)phthalate exposure in pregnant mice. *ACS Omega* 7: 23717–23726, 2022.

12) Hibino Y, Iguchi A, Zaitsu K: Preliminary study to classify mechanisms of mitochondrial toxicity by in vitro metabolomics and bioinformatics. *Toxicol Appl Pharmacol* 457: 116316, 2022.

13) Reel PS, Reel S, Pearson E, Trucco E, Jefferson E: Using machine learning approaches for multi-omics data analysis: A review. *Biotechnol Adv* 49: 107739, 2021.

14) Kawakami D, Tsuchiya M, Murata T, Iguchi A, Zaitsu K: Rapid quantification of extracellular neurotransmitters in mouse brain by PESI/MS/MS and longitudinal data analysis using the R and Stan-based Bayesian state-space model. *Talanta* 234: 122620, 2021.

15) Sriyudthsak K, Shiraishi F, Hirai MY: Mathematical modeling and dynamic simulation of metabolic reaction systems using metabolome time series data. *Front Mol Biosci* 3: 15, 2016.

16) Zaitsu K, Miyawaki I, Bando K, Horie H, Shima N, et al: Metabolic profiling of urine and blood plasma in rat models of drug addiction on the basis of morphine, methamphetamine, and cocaine-induced conditioned place preference. *Anal Bioanal Chem* 406: 1339–1354, 2014.

17) Zaitsu K, Eguchi S, Ohara T, Kondo K, Ishii A, et al: PiTMaP: A new analytical platform for high-throughput direct metabolome analysis by probe electrospray ionization/tandem mass spectrometry using an R software-based data pipeline. *Anal Chem* 92: 8514–8522, 2020.

18) Xia J, Wishart DS: MSEA: A web-based tool to identify biologically meaningful patterns in quantitative metabolomic data. *Nucleic Acids Res* 38: W71-W77, 2010.

19) Aggio R, Villas-Bôas SG, Ruggiero K: Metab: An R package for high-throughput analysis of metabolomics data generated by GC-MS. *Bioinformatics* 27: 2316–2318, 2011.

20) Uppal K, Soltow QA, Strobel FH, Pittard WS, Gernert KM, et al: xMSanalyzer: Automated pipeline for improved feature detection and downstream analysis of large-scale, non-targeted metabolomics data. *BMC Bioinform* 14: 15, 2013.

21) Edmands WMB, Barupal DK, Scalbert A: MetMSLine: An automated and fully integrated pipeline for rapid processing of high-resolution LC–MS metabolomic datasets. *Bioinformatics* 31: 788–790, 2015.

22) Franceschi P, Mylonas R, Shahaf N, Scholz, M, Arapitsas P, et al: MetaDB a data processing workflow in untargeted

MS-based metabolomics experiments. *Front Bioeng Biotechnol* 2: 1−12, 2014.

23) Wehrens R, Weingart G,Mattivi F: metaMS: An open-source pipeline for GC−MS-based untargeted metabolomics. *J Chromatogr* B 966: 109−116, 2014.

24) Wen B, Mei Z, Zeng C, Liu S: metaX: A flexible and comprehensive software for processing metabolomics data. *BMC Bioinform* 18: 183, 2017.

25) Liggi S, Hinz C, Hall Z, Santoru ML, Poddighe S, et al: KniMet: A pipeline for the processing of chromatography—Mass spectrometry metabolomics data. *Metabolomics* 14: 52, 2018.

26) Riquelme G, Zabalegui N, Marchi P, Jones CM, Monge ME: A python-based pipeline for preprocessing LC-MS data for untargeted metabolomics workflows. *Metabolites* 10: 416, 2020.

27) Winkler R: CHAPTER 1 Introduction, in Winkler R (ed): *Processing metabolomics and proteomics data with open software: A practical guide*. pp. 1−25, The Royal Society of Chemistry, Cambridge, 2020.

28) Pang Z, Chong J, Zhou G, de Lima Morais DA, Chang L, et al: MetaboAnalyst 5.0: Narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res* 49: W388−W396, 2021.

29) Ghosh D, Poisson LM: "Omics" data and levels of evidence for biomarker discovery. *Genomics* 93: 13−16, 2009.

30) Dunn OJ: Multiple comparisons among means. *J Am Stat Assoc* 56: 52−64, 1961.

31) Benjamini Y, Hochberg Y: Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol* 57: 289−300, 1995.

32) Xue J, Schmidt S V, Sander J, Draffehn A, Krebs W, et al: Transcriptome-based network analysis reveals a spectrum model of human macrophage activation. *Immunity* 40: 274−288, 2014.

33) Bajaj JS, Sikaroodi M, Shamsaddini A, Henseler Z, Santiago-Rodriguez T, et al: Interaction of bacterial metagenome and virome in patients with cirrhosis and hepatic encephalopathy. *Gut* 70: 1162−1173, 2021.

34) Langfelder P, Horvath S: WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform* 9: 559, 2008.

35) Robinson MD, McCarthy DJ,Smyth GK: edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139−140, 2010.

36) Worley B, Powers R: Multivariate analysis in metabolomics. *Curr Metabolomics* 1: 92−107, 2013.

37) Sato T, Zaitsu K, Tsuboi K, Nomura M, Kusano M, et al: A preliminary study on postmortem interval estimation of suffocated rats by GC-MS/MS-based plasma metabolic profiling. *Anal Bioanal Chem* 407: 3659−3665, 2015.

38) Cuperlovic-Culf M: Machine learning methods for analysis of metabolic data and metabolic pathway modeling. *Metabolites* 8: 4, 2018.

39) Pomyen Y, Wanichthanarak K, Poungsombat P, Fahrmann J, Grapov D, et al: Deep metabolome: Applications of deep learning in metabolomics. *Comput Struct Biotechnol J* 18: 2818−2825, 2020.

40) Hirai MY, Shiraishi F: Using metabolome data for mathematical modeling of plant metabolic systems. *Curr Opin Biotechnol* 54: 138−144, 2018.

41) Zaitsu K, Hayashi Y, Murata T, Ohara T, Nakagiri K, et al: Intact endogenous metabolite analysis of mice liver by probe electrospray ionization/triple quadrupole tandem mass spectrometry and its preliminary application to in vivo real-time analysis. *Anal Chem* 88: 3556−3561, 2016.

42) Zaitsu K, Hayashi Y, Murata T, Yokota K, Ohara T, et al: In vivo real-time monitoring system using probe electrospray ionization/tandem mass spectrometry for metabolites in mouse brain. *Anal Chem* 90: 4695−4701, 2018.

43) Murata T, Zaitsu K: Chapter6 -Probe electrospray ionization/mass spectrometry and its applications to the life sciences, in Zaitsu K (ed): *Ambient ionization mass spectrometry in life sciences*. pp. 171−205, Elsevier, Amsterdam, 2020.

44) Zhu C, Preissl S, Ren B: Single-cell multimodal omics: The power of many. *Nat Methods* 17: 11−14, 2020.

45) Lin S, Liu Y, Zhang M, Xu X, Chen Y, et al: Microfluidic single-cell transcriptomics: Moving towards multimodal and spatiotemporal omics. *Lab Chip* 21: 3829−3849, 2021.

46) Debois D, Jourdan E, Smargiasso N, Thonart P, De Pauw E, et al: Spatiotemporal monitoring of the antibiome secreted by bacillus biofilms on plant roots using MALDI mass spectrometry imaging. *Anal Chem* 86: 4431−4438, 2014.

47) Ali A, Abouleila Y, Shimizu Y, Hiyama E, Emara S, et al: Single-cell metabolomics by mass spectrometry: Advances, challenges, and future applications. *TrAC Trends Anal Chem* 120: 115436, 2019.

48) Taylor MJ, Lukowski JK, Anderton CR: Spatially resolved mass spectrometry at the single cell: Recent innovations in proteomics and metabolomics. *J Am Soc Mass Spectrom* 32: 872−894, 2021.