Master's Thesis of

Comparisons of Various Metabolomic Data Analyzing Tools

다양한 대사체학 데이터 분석 도구의 비교

February 2023

Graduate School of Life Sciences
Seoul National University
Genetic Engineering Major

이서연

Comparisons of Various Metabolomic Data Analyzing Tools

Under the guidance of professor Sung Won Kwon

Submitting a master's thesis of
Life Sciences

February 2023

Graduate School of Life Sciences
Seoul National University
Genetic Engineering Major

이서연

Confirming the master's thesis written by
이서연
February 2023

Chair _____(Seal)
Vice Chair _____(Seal)
Examiner _____(Seal)

Abstract

Lee, Seo Yeon

Graduate School of Life Sciences

Genetic Engineering Major

Seoul National University

Analyzing metabolomics data comes with a step of inputting the raw data into metabolomics data analyzing tools or software, in which the tools would analyze the data to produce the most optimal results of one's choice. But considering the fact that there are countless number of metabolomics data analysis tools out there, which ones are the most suitable tools to use for different cases of metabolomics data analysis, and which ones are the easiest tools to work with? This study focuses on compiling the names of the various types of metabolomics data analysis tools used in other dissertations, trimming down the list of tools to figure out which of them were used the most, and analyzing the strengths and limitations of each of the tools have, finally to evaluate which one of them is the "best" tool for different cases. Total of 47 papers were compiled initially, in which 8 of them were removed that had the Impact Factor (IF) of less than three, and the remaining 39 papers were sorted out by the metabolomics data analysis tools that the authors used. Out of the list of tools, two of the tools that were the most frequently cited were noted (i.e., MetaboAnalyst and SIMCA), as well as the third tool that was recently developed (i.e., Metaboseek) and their advantages and disadvantages were analyzed

thoroughly by manually inputting data and observing the results that those tools produced. This study was performed in the hopes that researchers who wish to analyze their metabolomics data would understand which tools are the most optimized tools for their research.

Keywords: Metabolomic data analyzing tool, MetaboAnalyst, SIMCA, METABOseek, PCA, PLS-DA

Student Number: 2019-20005

List of Contents

List of Figures

List of Tables

Chapter 1. Introduction

1.1. Study Background

One of the highly crucial steps in research is not only obtaining data from experiments, but also to be able to analyze data to confirm they are feasible to be used and produce results and conclusions out of them. Although acquiring data from experiments is the job to be done by researchers, inputting and producing results with those data is done with data analysis tools. It is up to the researchers to decide whether the analyzed data from the tools is reasonable and is parallel to their thesis. However, it is highly crucial to use the "best" and the most "suitable" tool for analyzing metabolomics data, as each and every data are all different in the sense that some data may produce better results with one tool, and other data may produce better results with a different tool. Therefore, it is bottom-line necessary and imperative to understand the strengths and limitations of each of the metabolomics data analysis tools have and to recognize which tools are the best ones to work with when it comes to analyzing particular data.

1.2. Purpose of Research

Many of the tools are able to perform most of metabolomic data

analysis methods out there, because the results that every researcher needs may be done by different analyses. For the most part, analyses are carried out to achieve the goal of establishing a conclusion of whether the sample is any significantly different from the control or not, as well as to justify this hypothesis. It is up to the researcher on which analysis to use and thus how to display their results from those analyses, but two of the most extensively used basic multivariate analyses used in metabolomics and proteomics are principal component analysis (PCA) and partial least-squares discriminant analysis (PLS-DA) [1, 2]. Both PCA and PLS-DA are analyses for multivariate data in that as the dimensional space becomes lower, they capture as much of the information in the data [2]. However, the main difference is that PCA is an unsupervised method of analysis where the information of each group of samples is unknown. PLS-DA, however, differs from PCA in that it is a supervised method of analysis that the information of each group of samples is supplied [3]. PCA is a convenient analysis when it comes to grouping data points with unknown information, whereas PLS-DA is more efficient when segregating data points under the circumstance that the group information is given. Therefore, this study focused primarily on easiness of the procedure done by the tools for acquiring the PCA and PLS-DA results from data, any limitations one may face when attempting to use other analyses, and lastly any practical complications that one may come across when installing or using such tools. This is carried out by first obtaining various papers from

PubMed Advanced with multiple relevant keywords such as "metabolomics," "metabolites," "PCA," "PLS-DA," and so on. Some of the searched papers that are to be removed are due to the fact that they did not fit into our criteria of selecting suitable papers – those are, firstly, the papers must be in English language; secondly, the link to a paper must be available (i.e., if the page from PubMed did not contain the link to its paper, then it is not readable); thirdly, if the same paper gets searched up twice or more, then it is only counted as once; fourthly, the papers must be research papers, meaning review papers are out of the question; and lastly, the journal impact factor, or impact factor (JIF, IF) of the papers must be at least three.

PubMed Advanced Search was chosen as the main reliable resource as a search engine for screening for research papers among other search engines. The reasons as to why it was chosen are because first of all, Advanced Search for any search tool is a good choice in finding dissertations of your own taste only using few keywords. Among many search engines, PubMed is one of engines that provides the most numbers of papers with the vastest types of categories [4]. Secondly, PubMed is a free and publicly available resource that covers biomedical literature and highly authoritative. Printed literatures are readily and continuously updated, as well as online literatures that are in a early version before print publication by various journals.

Organizing the compilation of all the journals into categories of the types of metabolomic data analysis tools they used to

produce a table, would aid in understanding which types of tools have been used the most widely, and be able to compare the most frequently used tools for their strengths and weaknesses in terms of the easiness of creating PCA and PLS-DA plots from data, capabilities of each tools to perform any other analyses, and lastly, any practical limitations in installment or usages of such tools.

Chapter 2. Methods

2.1. PubMed Advanced Search

PubMed Advanced Search (https://pubmed.ncbi.nlm.nih.gov/advanced/) was used to search for literatures related to metabolomic data analysis and tools for analyzing those data. This was done by looking up several specific keywords on PubMed Advanced Search, placing "Date – Publication: 2018/01/01 to present," and "Article Type: Books and Documents, Clinical Trial, Meta-Analysis, Randomized Controlled Trials" as defaults and adding various other keywords such as "Metabolomics" or "Metabolites" on "Title" query, and "metabolomics data," "metabolites," "principal component analysis (PCA)," or "partial least squares-discriminant analysis (PLS-DA)" on the "Title/Abstract" query, or in combination of these keywords. The number of dissertations that results from entering these keywords is sixty-two. To further trim down the number of results, papers with impact factor (IF) lower than three have been removed, which in final came down to forty-one papers.

2.2. Metabolomics Data on MetaboAnalyst for Analysis

The website that leads to Metaboanalyst is https://metaboanalyst.ca/, and once entered, the button "Click here

to start" leads to the main modules to data analysis. There are thirteen modules in total, but in our lab, we primarily use "statistical analysis [one factor]" module that offers various commonly used statistical and machine learning methods including PCA and PLS-DA. It also provides clustering and visualization tools to create dendrograms and heatmaps as well as to classify data based on random forests and SVM. Clicking the "statistical analysis [one factor]" module we are introduced with the data upload page, and once uploading the data (or alternatively "test data" can be inputted; usually in our lab we obtain MS peak intensity data, so data of "MS peak intensities" can be selected for data analysis), click "submit" on the bottom of the page. This leads to "data integrity check" page where we are able to check for integrity of our data, as well as edit label for our groups. When all the groups were edited, continue on by clicking "proceed" button which leads us to "data filtering" page. This page identifies and removes variables that are unlikely to be of use when modeling the data. After reading how to filtering data works and which data filter should be chosen, click "submit" and then "proceed." Now we are shown with "normalization overview" page, and once sample/data normalization with data scaling is applied after reading their definitions, click "normalize" (in our lab, we use "normalization by median," "log transformation," and "Pareto scaling"). You can view result after these normalizations by clicking "view result." Once you are sure your data have been normalized and scaled to your taste, click

“proceed” to get to the final page where we choose which analysis methods we want to apply to our normalized data. There are various analysis methods to use, but our lab mainly focuses on PCA and PLS-DA. By clicking PCA, we are led with the main overview visualization of the data, scree plot, 2D scores plot, loadings plot, synchronized 3D plots, and biplot. Each plot we can tweak our visualization of data for our research. The same goes for PLS-DA, where it shows the same kinds of plots except there are three more plots, which are importance measures, cross validation, and permutation.

2.3. Metabolomics Data on SIMCA for Analysis

SIMCA can be downloaded from the website "https://www.sartorius.com/" either by trying their free trial (“SIMCA Free Trial Download”) or buying their program right away (“Buy Now”). By entering your information on the bottom of the download page, you are led to the program download page, and once you click the “Download” button, the program starts downloading. When finished, the SIMCA 64-bit zip file first has to be unzipped to install the “SIMCA_17_0_2_x64_Setup” program. Installing this program opens up the SIMCA tool, and clicking “Regular Project” on the “Start” menu would guide you into another windows that asks you to upload your data. SIMCA can only understand when the sample names and labels are on columns, and

variables on rows. If the data table is written the other way around, “transpose” under the “Edit” tab would transpose the raw data table. Sample names on the first column and variables on the first row are marked as “Primary ID,” and the labels on the second row are identified as “Secondary ID.” After assigning Primary ID and Secondary ID, and once no issues were found, clicking “Finish Import” button would import the data into SIMCA program which leads to the imported file saving window into the designated Desktop file. Once the data has been pasted onto the SIMCA program, data could be modified using the tabs on the top row. This automatically opens a new window called “Project Window” that already produced a PCA-X <Unfitted> model, and clicking “Autofit” under “Home” tab would “fit” this PCA model. This opens up “Summary of Fit” window that shows the number of components with R2 and Q2 as bar graphs. Right-clicking the active model in the Project Window and then clicking “Edit Model #” leads to “Dataset” window where log transform can be applied under “Log” tab after selecting all the variables and clicking “Set” button, and pareto scaling (Par) in “Set Scaling” after selecting all the variables and clicking “Set all” which is found under “Scale” tab. This needs data fitting, and this can be done again by clicking “Autofit” under the “Home” tab. PLS-DA can be obtained by right-clicking the PCA-X model and then “New as Model #.” On the “Observations” tab, designate class to each of the samples according to different labels (i.e., secondary ID), and on the bottom of the window “Model type” can be

chosen such as PLS-DA. This PLS-DA also need to be fitted by clicking "Autofit".

For both PCA and PLS-DA created with SIMCA can be displayed as basic 2D scatter plots, loading line plots, score column plots, as well as score scatter 3D plots. The same goes for loadings plots, and the variable importance in projection (VIP) plot (only for PLS-DA), which are all found in "Scores," "Loadings," and "VIP" all under the "Home" tab.

2.4. Metabolomics Data on Metaboseek for Analysis

Metaboseek can be either downloaded from or exists in web-based version on https://metaboseek.com, and once installed, it opens up the Metaboseek window. However, unlike online version of SIMCA, Metaboseek online only computes sample dataset for practice. Installing Metaboseek opens up a new window where on the "Start" menu, you can upload your data either as a feature table, or MS data. To note, Metaboseek is only capable of reading .mzXML, .mzML, .cdf, .nc, and .mzData file formats for MS data files. For feature table data, the data table needs to be tabulated in such a way that samples and labels have to be written in rows, and variables have to be written in columns. Once data has been uploaded, the software directs you to "Data Explorer" menu, which displays "Options," "Data viewer," "Feature table," and "Feature table actions" sections. Each section allows you to

modify uploaded data such as "Sirius options," "Molecular formula prediction," "RT correction," "Mass shifts," and "EIC options" for "Options" section; "Data viewer" section shows various results from data analysis as well as changing them with "MS2 browser," "PCA viewer," "Venn diagrams," "MS browser," "Quickplots," "Ratio plot," "Grouped EICs," and "Regroup MS data"; feature table can be viewed from the "Feature table" section, and in addition to this, columns from the table can be sorted out to one's preference; lastly, there is "Feature table actions" where applying filters, analyzing tables, and regrouping tables are all done under this section. Under the "Filter table," all of the samples were chosen for the "Sample intensities," "then clocked "Apply filters." Changes were also made under the "Analyze table" section, where "Use normalized/inputted data" was check-marked, and under the settings, "Normalize data" as well as "Apply log10" were also included. "Select control group" was WT, and "Basic analysis," "PCA features," as well as "PCA samples" were added on the "Basic analysis" section. Clicking "Run selected analysis" would update the feature table. Finally, by clicking "PCA viewer" under the "Data viewer" tab, selecting "Group" under "Color by" would display PCA plot for WT and KO.

Chapter 3. Results

By searching for dissertations that include metabolomic data analysis and the tools that were used to analyze them on PubMed Advanced Search and removing the ones with impact factor (IF) lower than three, 39 papers came up in total. All of these papers used different kinds of metabolomic data analysis tools, which were MetaboAnalyst, SIMCA, R, MATLAB, Cytoscape, SPSS, Mestrenova, Compound Discoverer, BioEStat, STATA, and Unscrambler X. Most papers used one analysis tool where some of them used two of these tools. The frequency of these tools being cited in the searched dissertations were thirteen papers for MetaboAnalyst, eleven papers for SIMCA, eight papers for R, four papers for MATLAB, two papers for Cytoscape, two papers for SPSS, one paper for Mestrenova, one paper for Compound Discoverer, one paper for BioEStat, one paper for STATA, and lastly, one paper for Unscrambler X. The list of metabolomics data analysis tools and the papers that cited them are shown in Table 1.

Tool Name	Tool Description	Samples and Analytes Used	JIF (2022)	Keywords (PubMed Advanced)
MetaboAnalyst	A web-based tool that supports not only the analysis of metabolomic data but also its interpretation and integration with other omics data	Plasma [5]	4.36	Date - Publication: 2018/01/01 to present Title/Abstract: metabolomics data Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Urine [6]	3.411	Date - Publication: 2018/01/01 to present Title: metabolites Title/Abstract: Principal Component Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial

		Urine [7]	5.429	Date – Publication: 2018/01/01 to present Title: metabolites Title/Abstract: Principal Component Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Urine [8]	4.15	Date – Publication: 2018/01/01 to present Title/Abstract: metabolites Title/Abstract: Principal Component Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Plasma [9]	4.379	Date – Publication: 2018/01/01

				to present Title/Abstract: metabolomics data Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Plasma and fecal samples [10]	4.389	Date - Publication: 2018/01/01 to present Title: metabolites Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Plasma [11]	5.279	Date - Publication: 2018/01/01 to present Title: metabolomics Title/Abstract: Partial Least Squares-Discriminant Analysis

				Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Exhaled breath condensate [12]	6.01	Date - Publication: 2018/01/01 to present Title/Abstract: metabolites Title/Abstract: Partial Least Squares-Discriminant Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Serum [13]	6.055	Date - Publication: 2018/01/01 to present Title: metabolomics Title/Abstract: Principal Component Analysis Article Type: books and

				documents, clinical trial, meta-analysis, randomized controlled trial
		Urine [14]	5.279	Date - Publication: 2018/01/01 to present Title: metabolomics Title/Abstract: Partial Least Squares-Discriminant Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Serum from blood and tumor tissue [15]	4.638	Date - Publication: 2018/01/01 to present Title/Abstract: metabolites Title/Abstract: Partial Least Squares-Discriminant Analysis Article Type: books and documents, clinical trial, meta-

				analysis, randomized controlled trial
		cerebrospinal fluid [16]	4.996	Date – Publication: 2018/01/01 to present Title/Abstract: metabolites Title/Abstract: Partial Least Squares–Discriminant Analysis Article Type: books and documents, clinical trial, meta–analysis, randomized controlled trial
		Urine [17]	3.24	Date – Publication: 2018/01/01 to present Title/Abstract: metabolomics Title/Abstract: Principal Component Analysis Article Type: books and documents, clinical trial, meta–analysis, randomized controlled

				trial
SIMCA	User-friendly software developed by Umetrics chiefly for the analyses of PCA and PLS regression	Plasma [5]	5.09	Date - Publication: 2018/01/01 to present Title/Abstract: metabolomics data Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Urine [6]	3.411	Date - Publication: 2018/01/01 to present Title: metabolites Title/Abstract: Principal Component Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Urine [7]	6.706	Date - Publication: 2018/01/01

				to present Title: metabolites Title/Abstract: Principal Component Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Plasma [18]	3.52	Date – Publication: 2018/01/01 to present Title/Abstract: metabolomics Title/Abstract: Principal Component Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Plasma from the venous blood [19]	4.996	Date – Publication: 2018/01/01 to present

				Title/Abstract: metabolomics data Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Fasting blood samples [20]	5.914	Date - Publication: 2018/01/01 to present Title: metabolomics Title/Abstract: Partial Least Squares-Discriminant Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Pleural effusion [21]	4.996	Date - Publication: 2018/01/01 to present Title: metabolites Title/Abstract: Partial Least

				Squares Discriminant Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Serum [22]	3.14	Date - Publication: 2018/01/01 to present Title: metabolomics Title/Abstract: Principal Component Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Serum [23]	4.93	Date - Publication: 2018/01/01 to present Title/Abstract: metabolites Title/Abstract: Partial Least Squares-Discriminant Analysis

				Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Serum [24]	3.828	Date - Publication: 2018/01/01 to present Title/Abstract: metabolites Title/Abstract: Partial Least Squares-Discriminant Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Plasma [25]	5.914	Date - Publication: 2018/01/01 to present Title/Abstract: metabolites Title/Abstract: Partial Least Squares-Discriminant Analysis Article Type: books and

				documents, clinical trial, meta-analysis, randomized controlled trial
		Serum [26]	3.364	Date - Publication: 2018/01/01 to present Title/Abstract: metabolites Title/Abstract: Partial Least Squares-Discriminant Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
Cytoscape	An open-source tool that integrates high-throughput expression data with biomolecular interaction networks	Human platelets [27]	10.787	Date - Publication: 2018/01/01 to present Title/Abstract: metabolomics data Article Type: books and documents, clinical trial, meta-analysis, randomized controlled

				trial
		Cerebrospinal fluid [28]	7.598	Date – Publication: 2018/01/01 to present Title/Abstract: metabolomics Title/Abstract: Principal Component Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
SPSS	A statistical software that allows a user to dig deeper into their data through intuitive user interface, advanced data visualizations, automated data preparation, and more	Serum [13]	6.055	Date – Publication: 2018/01/01 to present Title: metabolomics Title/Abstract: Principal Component Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial

		Plasma [18]	3.52	Date – Publication: 2018/01/01 to present Title/Abstract: metabolomics Title/Abstract: Principal Component Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
Mestrenova	A spectral data analyzing software, that is able to analyze various data such as 1H, 13C or any other 1D NMR as well as any 2D correlations, such as HSQC, HMBC, NOESY, COSY, TOCSY, etc.	Serum [29]	6.706	Date – Publication: 2018/01/01 to present Title: metabolites Title/Abstract: Principal Component Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
R	A statistical programming that is	Urine [14]	5.279	Date – Publication: 2018/01/01

	uniquely able to handle lots of data, main function being linear and nonlinear modelling, classical statistical tests, time−series analysis, classification, clustering, as well as graphical techniques, and more			to present Title: metabolomics Title/Abstract: Partial Least Squares−Discriminant Analysis Article Type: books and documents, clinical trial, meta−analysis, randomized controlled trial
		Human peripheral blood [30]	8.469	Date − Publication: 2018/01/01 to present Title/Abstract: metabolites Article Type: books and documents, clinical trial, meta−analysis, randomized controlled trial
		Urine [31]	5.914	Date − Publication: 2018/01/01 to present Title: metabolites Title/Abstract: Partial Least

				Squares Discriminant Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Pleural effusion [21]	4.996	Date - Publication: 2018/01/01 to present Title: metabolites Title/Abstract: Partial Least Squares Discriminant Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Plasma [32]	5.914	Date - Publication: 2018/01/01 to present Title: metabolites Title/Abstract: Partial Least Squares Discriminant Analysis

				Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Serum [33]	4.142	Date - Publication: 2018/01/01 to present Title: metabolomics Title/Abstract: Principal Component Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Plasma [34]	4.29	Date - Publication: 2018/01/01 to present Title/Abstract: metabolites Title/Abstract: Principal Component Analysis Article Type: books and

				documents, clinical trial, meta-analysis, randomized controlled trial
		Plasma [35]	7.514	Date - Publication: 2018/01/01 to present Title/Abstract: metabolites Title/Abstract: Partial Least Squares-Discriminant Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Serum [36]	7.045	Date - Publication: 2018/01/01 to present Title/Abstract: metabolomics Title/Abstract: Partial Least Squares-Discriminant Analysis Article Type: books and documents, clinical trial, meta-

				analysis, randomized controlled trial
Compound Discoverer	Developed by Thermo Fisher Scientific, uses chromatographic and mass spectra (MS) data and streamlines compound identification as well as comparative analyses, and provides extensive filtering and data visualization capabilities	Human peripheral blood [30]	8.469	Date – Publication: 2018/01/01 to present Title/Abstract: metabolites Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
MATLAB	A high-performance language combining computation, visualization, and programming for technical computing	Urine [37]	3.26	Date – Publication: 2018/01/01 to present Title: metabolites Title/Abstract: Principal Component Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Tissue samples	4.466	Date – Publication: 2018/01/01

		[38]		to present Title: metabolomics Title/Abstract: Partial Least Squares-Discriminant Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Serum [39]	5.614	Date - Publication: 2018/01/01 to present Title/Abstract: metabolites Title/Abstract: Principal Component Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Urine [40]	5.23	Date - Publication: 2018/01/01 to present

				Title: metabolomics Title/Abstract: Partial Least Squares−Discriminant Analysis Article Type: books and documents, clinical trial, meta−analysis, randomized controlled trial
STATA	An integrated software that allows data manipulation, visualization, statistics, and automated reporting	Plasma known lipid metabolites a [41]	4.614	Date − Publication: 2018/01/01 to present Title/Abstract: metabolites Title/Abstract: Principal Component Analysis Article Type: books and documents, clinical trial, meta−analysis, randomized controlled trial
BioEStat	Free software developed for undergraduates and graduates, and with easy−to−perform procedures,	Site−specific supragingival plaque samples	6.116	Date − Publication: 2018/01/01 to present Title/Abstract: metabolites

	able to carry out various statistical and graphical analysis	[42]		Title/Abstract: Principal Component Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
Unscrambler X	Software for multivariate data analysis, which frequently uses data calibration in the application of analytical data and the building of predictive models for use in real-time spectroscopic material analysis	Peritoneal dialysis effluent [43]	4.569	Date – Publication: 2018/01/01 to present Title/Abstract: metabolites Title/Abstract: Principal Component Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial

Table 1. List of Metabolomic Data Analysis Tools in Dissertations. The dissertations with the same colors represent identical ones, describing a single paper has used multiple metabolomic data analysis tools. A total of 47 papers, in which 8 papers overlapped, thereby resulting with 39 non-overlapping papers in total.

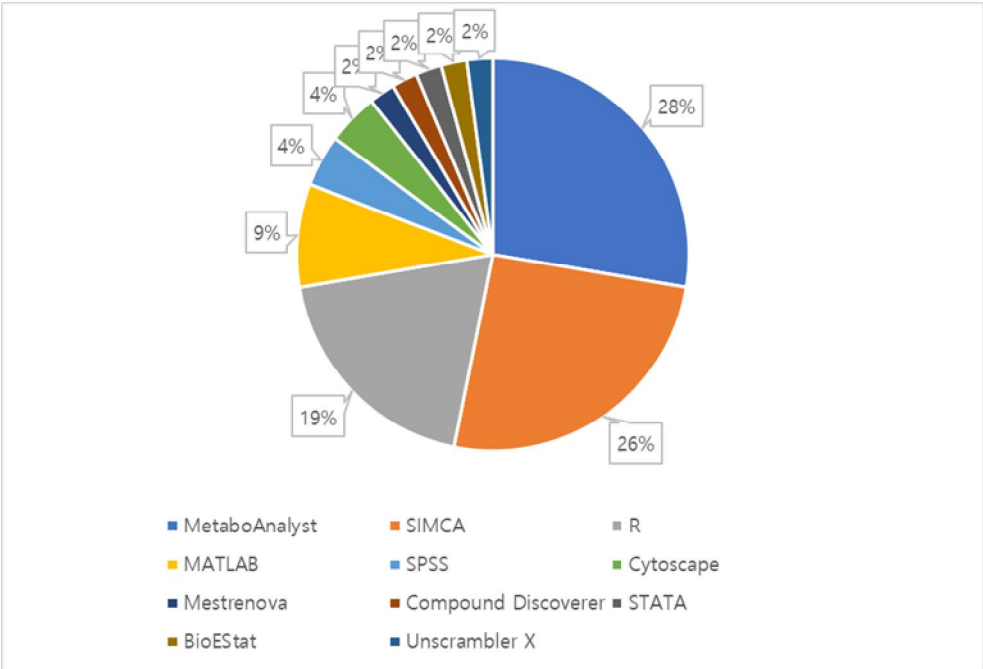


Figure 1. Pie chart depicting each metabolomic data analyzing tools and the numbers of them being cited in 39 dissertations listed above on Table 1.

The two most frequently cited metabolomic data analysis tools were MetaboAnalyst and SIMCA. A third tool had been added to the list, Metaboseek, as this tool was developed relatively recently in 2017 by Max Helf and his colleagues from Cornell University, to overcome the limitations of contemporary metabolomics data analysis tools have when it comes to analyzing comparative metabolomics data. These three tools were then used to analyze via PCA and PLS-DA a set of data obtained from an experiment by Saghatelian *et al.* Both PCA and PLS-DA were carried out with MetaboAnalyst, SIMCA, or Metaboseek, in which the plots are depicted on Figure 1. for MetaboAnalyst, Figure 2. for SIMCA, and Figure 3. for Metaboseek.

(a) (b)

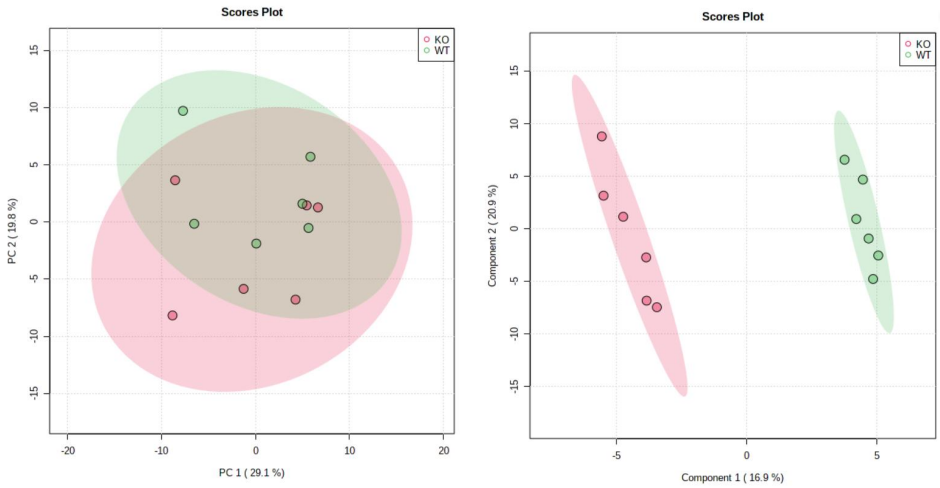


Figure 2. PCA and PLS−DA using MetaboAnalyst in 2D. (a) PCA plot by MetaboAnalyst for data from Saghatelian *et al*. Wildtype (WT) mice for fatty acid amide hydrolase, FAAH(+/+), compared with knockout (KO) mice, FAAH(−/−), is shown [50]. In 2D, no significant difference is shown between the two groups. PCA was done with the module "statistical analysis [one factor]" and used the test data for MS peak intensities from Saghatelian *et al*. Data was not filtered as the features were less than 5000, but several normalizations were done to the data – those were, normalization by median, log transformation, and pareto scaling. Applying these normalizations led to the 2D PCA plot shown above. (b) PLS−DA plot for data from Saghatelian *et al*. WT FAAH(+/+) compared with FAAH(−/−) is shown. In 2D, significant difference is shown between the two groups. PLS−DA plot was obtained similar to how PCA plot was acquired, but instead of choosing PCA at the last step, PLS−DA was chosen to obtain the above 2D PLS−DA plot.

(a) (b)

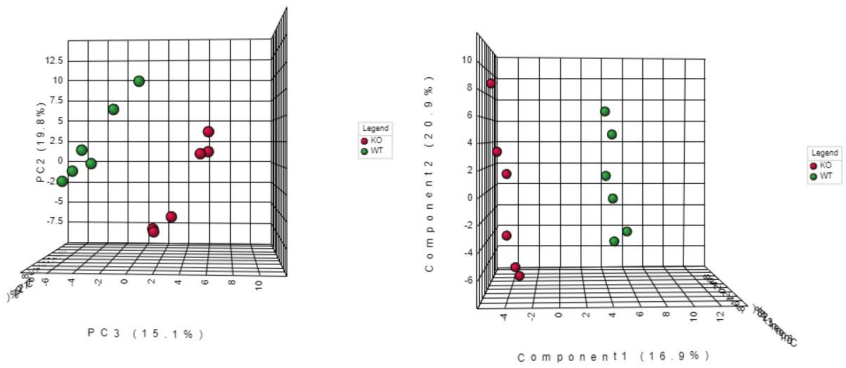


Figure 3. PCA and PLS-DA using MetaboAnalyst in 3D. (a) MetaboAnalyst's PCA plot using information from Saghatelian et al. It is proven that wildtype (WT) mice for fatty acid amide hydrolase, FAAH(+/+), are different from knockout (KO) mice for FAAH(−/−). Significant differences between the two groups are seen in 3D. Data from Saghatelian et al. are shown in (b) a PLS-DA plot. It is illustrated how WT FAAH(+/+) compares to FAAH(−/−). Significant differences between the two groups are seen in 3D.

(a) (b)

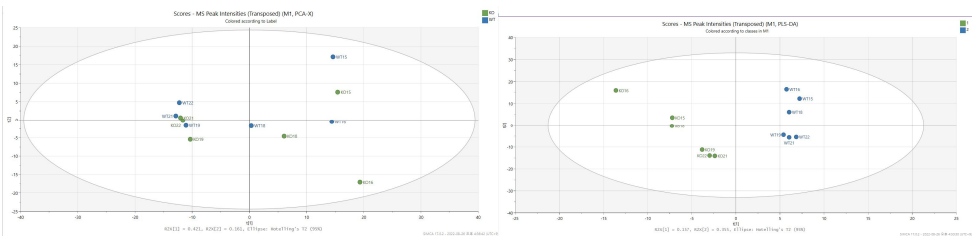
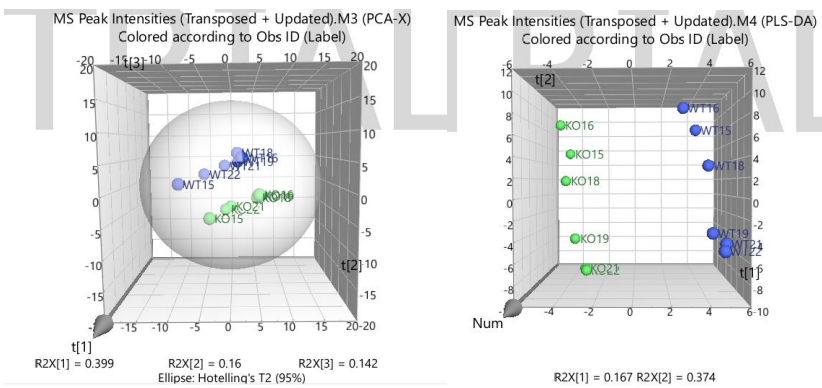


Figure 4. PCA and PLS-DA using SIMCA in 2D. (a) PCA plot by SIMCA for data from Saghatelian *et al.* Wildtype (WT) mice for fatty acid amide hydrolase, FAAH(+/+), compared with knockout

(KO) mice, FAAH(−/−), is shown. In 2D, no significant difference is shown between the two groups. The data table was uploaded onto SIMCA first, which by default gave PCA model, and by performing log transformation and pareto scaling on the "Dataset" window, then clicking on the "Scatter plot" gives 2D PCA plot. (b) PLS−DA plot for data from Saghatelian *et al.* WT FAAH(+/+) compared with FAAH(−/−) is shown. In 2D, significant difference is shown between the two groups. 2D PLS−DA plot is similarly produced as the PCA plot, except in PLS−DA, classes are labeled under the "Dataset" window and then model type is changed to "PLS−DA."

a) b)



<u>Figure 5.</u> **PCA and PLS−DA using SIMCA in 3D.** (a) PCA plot by SIMCA for data from Saghatelian *et al.* Wildtype (WT) mice for fatty acid amide hydrolase, FAAH(+/+), compared with knockout (KO) mice, FAAH(−/−), is shown. In 3D, significant difference is shown between the two groups. (b) PLS−DA plot for data from Saghatelian *et al.* WT FAAH(+/+) compared with FAAH(−/−) is shown. In 3D, significant difference is shown between the two

groups.

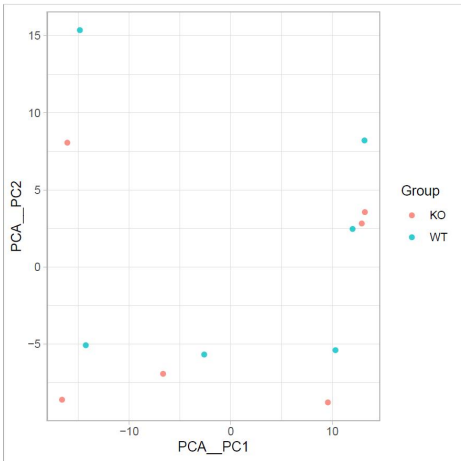


Figure 6. PCA using Metaboseek. PCA plot by Metaboseek for data from Saghatelian *et al*. Wildtype (WT) mice for fatty acid amide hydrolase, FAAH(+/+), compared with knockout (KO) mice, FAAH(−/−), is shown. In 2D, no significant difference is shown between the two groups.

Although the same data have been inputted into the tools, each of the tool outputted results in different formats. Fundamentally, all three tools generated the same results from PCA and PLS-DA, however the way the data points are depicted on plots is different by each tool. By performing both analyses with MetaboAnalyst, SIMCA, and Metaboseek, these tools were compared for their advantages and disadvantages as to how convenient and easily these tools could perform data analysis, how simple and visually the plots are neat to look at, as well as how efficiently and effortlessly data manipulation could take place. Table 2. represents brief descriptions of these tools as well as highlighting the

strengths and limitations of these three tools have.

Table 2. Strengths and Limitations of MetaboAnalyst, SIMCA, and Metaboseek

METABOANALYST

WEB-BASED, SIMPLE TO USE, AND IDEALLY SUITED FOR IN-DEPTH STUDIES, INTERPRETATION, AND INTEGRATION OF METABOLOMIC DATA WITH OTHER OMICS DATA	
PROS	**CONS**
WEB-BASED PROGRAM	LIMITED TO BASIC MANIPULATIONS OF DATA
EASY TO WORK WITH TO THOSE WHO ARE UNFAMILIAR WITH CODING	
COMPATIBLE WITH MULTIPLE FILE TYPES	
ABLE TO PERFORM VARIOUS ANALYSES	

SIMCA

USER-FRIENDLY SOFTWARE DEVELOPED BY UMETRICS CHIEFLY FOR THE ANALYSES OF PCA AND PLS REGRESSION	
PROS	**CONS**
FORMAT SIMILAR TO MICROSOFT EXCEL	LIMITED SUPPORT FOR GLOBAL METABOLOMICS ESPECIALLY WITH REGARDS TO RAW DATA PROCESSING
MODIFYING AND FILTERING EACH INDIVIDUAL DATA	MORE COMPLICATED TO WORK WITH THAN

OR THE ENTIRE DATASET CAN BE ACCOMPLISHED	METABOANALYST
ABLE TO OBSERVE MULTIPLE PLOTS AND GRAPHS CONSTRUCTED FROM THE SAME DATASET AT THE SAME TIME IN ONE SCREEN AND ABLE TO LOCATE EACH DATA POINTS FROM ONE PLOT TO ANOTHER	PRIMARILY MADE FOE PERFORMING PCA, PLS-DA, PLS DERIVATIVES

METABOSEEK

WEB-BASED TOOL THAT IS PRIMARILY USED FOR COMPARATIVE METABOLOMICS DATA ANALYSIS	
PROS	CONS
EXTREMELY WELL-SUITED PLATFORM FOR COMPREHENSIVE DATA ANALYSIS WORKFLOW, SUCH AS FEATURE DETECTION TO COMPOUND IDENTIFICATION, PARTICULARLY DESIGNED TO FACILITATE UNTARGETED METABOLOMICS	NOT WEB-BASED
	TAKES EXTENSIVE AMOUNT OF TIME TO GET FAMILIARIZED WITH THE TOOL
	ONLY PCA IS ENABLED FOR ANY MULTIVARIATE ANALYSES

Chapter 4. Discussion

Dissertations that cited tools for metabolomic data analysis were searched on PubMed Advanced (https://pubmed.ncbi.nlm.nih.gov/advanced/) with multiple filters such as date of publication, article type, and multiple keywords being "Metabolomics" or "Metabolites" on "Title" query, and "metabolomics data," "metabolites," "principal component analysis (PCA)," or "partial least squares-discriminant analysis (PLS-DA)" on the "Title/Abstract" query, or in combination of these keywords. Firstly, the date of publication was set to 2018/01/01 as journals and papers that were no more than five years old from now (i.e., year 2022) since publication would be up-to-date with the technology and instruments thus the software and tools that are being used today. Otherwise, the papers that were published earlier than 2018/01/01 would have used outdated software and tools that researchers no longer would be able to perform our metabolomic data analyses with.

Secondly, the title and/or the abstract section should contain the words "metabolomics data" to specify on finding journals and papers that studied and underwent experiments regarding on metabolomics and obtained data from them. This cut down the number of journals and papers to 860 papers, however the 860 papers that have come up were still too vague and too innumerable to be used for this research paper.

This led to the third rationale to finding journals and papers, which was applying the filter called "Article Type" that is located on the left side of the PubMed Advanced Search. Under the "Article Type" section, there are six types of articles that can be chosen and filter out the rest of the articles, and the filters that were chosen were "Books and Documents," "Clinical Trial," "Meta-Analysis," and "Randomized Controlled Trial." The two articles that have been filtered out were "Review" and "Systematic Review," as these articles are just the reviews on metabolomics and do not consist of Figures of metabolomics data and Methods of how the researchers used various software or tools to analyze the metabolomics data. Applying these types of articles onto the filter gave 15 results.

The number of articles that have come up is few enough, but out of these, the articles that were chosen were only the articles that contain Figures of the metabolomic data analyses, and whether the articles contain metabolomics data analysis software and tools on the Methods section. The reason as to why only these types of articles were chosen is because the articles must be research articles that truly completed research experiments to obtain such data, thus, the articles represented above did not fit into our criteria.

Further selecting down the number of articles using these keywords and filters came up with five papers, which are Liu X *et al* [5], Yang M *et al* [9], Huang L *et al* [19], Lindqvist HM *et al* [46], and Kirchberg FF *et al* [47].

On a similar level, making a slight change on the words for

the "Title/Abstract" query box and searching for "metabolites" but keeping other keywords and filters the same, we get 1,475 results. On page 3, we find a clinical trial paper by Danlos FX *et al* [30]. The exceptional thing about this paper over other papers is that the main research of this paper is on the metabolomic analyses of COVID-19 patients which unravel stage-dependent and prognostic biomarkers. This paper focuses on the ever-growing patients of current pandemic disease COVID-19, as well as concerns on them. To top it off, this article uses two different metabolomic data analyzing tools which are R software and Thermo Compound Discoverer.

When analyzing metabolomics data, the fundamental analyses that are performed are principal component analysis (PCA) and partial least squares-discriminant analysis (PLS-DA) by default. Therefore, the preferred metabolomics data analyzing tools must be able to perform these analyses and thus to find any journals and papers on PubMed Advanced Search tool that included the data of these analyses, the keywords that were added were "principal component analysis" on the "Title/Abstract" query and "metabolites" on the "Title" query. Still keeping the publication date and the article type filters the same, four results come up. all of these four papers (i.e., Esquivel A *et al* [37], Dos Santos Fechine CPN *et al* [29], Cao B *et al* [6], Dickson L *et al* [7]) were cited on this research paper as all of these papers have used various tools to perform PCA.

If PCA was used as a keyword to search for papers, then

PLS-DA should also be used as a keyword. Now to look for papers that implemented PLS-DA data into their research paper, changing words only from "principal component analysis" on the "Title/Abstract" query to "partial least squares-discriminant analysis." This entry gave three results (i.e., Kim H *et al* [31], Chen KC *et al* [21], Kim H *et al* [32]), which were all added to the citation at the end. Needless to say, these three papers were chosen because they contain the data from PLS-DA.

Using "metabolomics" instead of "metabolites" as the keyword for the "Title" query and using "principal component analysis" in the "Title/Abstract" query but keeping other filters the same, three results from Peng ML *et al* [22], Renai L *et al* [33], and Lijing W *et al* [13]. come out. As mentioned before, "metabolomics" as the keyword was entered since this research paper is dealing with metabolomics and the data from experiments on metabolomics. The three papers that came up have also been added to the list.

Like so, only modifying the keyword from "principal component analysis" to "partial least squares-discriminant analysis" for the "Title/Abstract" query came up with seven papers, in which one of them is redundant (i.e., Peng ML *et al* [22]). Therefore, six papers (i.e., Bejder J *et al* [40], Wang PS *et al* [11], Debik J *et al* [38], Xu D *et al* [20], Quartieri E *et al* [44], Madrid-Gambin F *et al* [14]) counted into the citation.

More papers were to be found if we put the keyword "metabolites" into the "Title/Abstract" query instead of the

"Title" query, as well as "principal component analysis" in the "Title/Abstract" query. Other filters were again kept the same, and now we are introduced with sixteen research papers. Out of these sixteen results, five of them have no articles attached on the website, therefore they were discarded. Similarly, out of the rest eleven results, five same results also come up when other keywords, and they are again disregarded as they are redundant. Final six results have been added onto my citation list (i.e., Nascimento MM *et al* [42], Grunert T *et al* [43], Romo-Hualde A *et al* [8], Pigsborg K *et al* [39], Razquin C *et al* [41], Chatterjee R *et al* [34]).

Like so, entering the keywords "metabolites" and "partial least squares-discriminant analysis" onto the "Title/Abstract" query but not changing any other filters gives twenty-three results. From the twenty-three articles, one article did not contain the relevant article, and thirteen articles were again repetitive, and could be discarded. Remaining eight article papers in total were added onto the citation, which were Metwaly S *et al* [23], Paris D *et al* [12], Marques JG *et al* [48], Lende TH *et al* [15], Jang HH *et al* [35], Pan X *et al* [16], Zhao S *et al* [25], and Park SA *et al* [26].

Now, switching the keyword from "metabolites" to "metabolomics" onto the "Title/Abstract" query and adding "principal component analysis" onto the same query, again keeping other filters the same, fourteen papers were obtained. Out of these, three of the papers again do not have the relevant articles attached, and eight of them also come up again when other keywords are looked up. Therefore, only three articles are new

articles, which are Hagos FT *et al* [28], McNairn M *et al* [17], and Luo Y *et al* [18].

Similarly, keeping the "metabolomics" on the "Title/Abstract" query but changing the "principal component analysis" to "partial least squares–discriminant analysis," we are shown with twenty–four resulting papers. One paper has no related article paper attached, and twenty papers were repeated papers. The rest three articles (i.e., Ulven SM *et al* [36], Jangsiripornpakorn J *et al* [49], Braga DPAF *et al* [45].) were relevant to the research.

Relevant papers were also looked up by switching "principal component analysis" or "partial least squares–discriminant analysis" as the "Title" query and either "metabolomics" or "metabolites" as the "Title/Abstract" query but not rendering any other filters, however no article papers have come up at all, which finalizes the list of relevant papers.

This results to 45 papers in total in the citation list, which is too many thus it must be shortened down by removing the papers with impact factors (IF) less than three. The reasons as to why IF less than three was chosen were because firstly, it was concluded that the journals Metabolites and Metabolomics must be kept on my list as they contain all the metabolites and metabolomics data, and are crucial to my research. The IF of Metabolites is 4.754, and that of Metabolomics is 4.29. Secondly, If the cut–off score was chosen for IF less than four (to retain the journals from Metabolites and Metabolomics) then most of the journals must be removed as many

of them have IF less than four. However, journals with IF two have too little significance, and thus journals with higher than IF three have been chosen to be kept in the citation list. This leaves with thirty-nine papers.

The 39 papers were tabulated according to the data analysis tools they used, which reveals the number of times each tool has been cited; out of the eleven tools, two tools that were used the most came out to be MetaboAnalyst and SIMCA. Both of the tools were very different as to how they were installed, where MetaboAnalyst was web-based as opposed to SIMCA which was an installable software from the website of the company. Besides this, MetaboAnalyst guides step-by-step as to how to perform analyses of one's preference, whereas SIMCA allows freely to manipulate data - much like Microsoft Excel. Therefore, MetaboAnalyst is extremely easy and simple to produce results from one's data just by following the instructions on the website. SIMCA is slightly more complicated to work with, but allows a more fluent manipulation of both individual data and entire dataset than MetaboAnalyst. It is able to display multiple plots and graphs all in one screen as well as able to locate each data point from one plot or graph to another, which allows users to easily pinpoint which data point from one plot appears on other plots and/or graphs.

MetaboAnalyst also covers extremely vast numbers of types of analyses other than PCA and PLS-DA, and since manipulation of data is designed as step-by-step approach, there is no difficulty following them - however because MetaboAnalyst only allows basic

data manipulation performance, this limits from accomplishing more complicated processing of data. Usually, these basic manipulations of data are sufficient for data analysis results, but if a user wish to carry out more complex process of data, then they may find it hard to do so. MetaboAnalyst typically permits a user to perform various types of analyses, but with basic data manipulation options.

SIMCA, on the other hand, is designed to perform multivariate data analyses especially PCA, PLS-DA, oPLS-DA, o2PLS, PLS, and PLS-Tree. It is specific to these analyses such as PCA and PLS-DA derivatives, therefore there are more you can do in terms of data manipulation. Data processing is more detailed although one needs to get familiarized with the tool first before actually carrying it out, because SIMCA is formatted in such a way that is somewhat difficult to locate the data processing toolbox. SIMCA is opposite to MetaboAnalyst in that it only covers a small number of types of analyses, but the data processing and manipulation parts are highly sophisticated and very detailed.

Metaboseek is the third tool added to the list of the most widely used tools after MetaboAnalyst and SIMCA. This tool had been also considered upon with the two previous tools because although it has not been cited on the dissertations, it is relatively recently developed and is still currently being updated thus Metaboseek has not been widely applied on data analyses by researchers yet, therefore there are very few papers on the usage of Metaboseek. Metaboseek seems to be more of a tool that analyzes mass spectra peaks, but also be able to generate PCA (but

not PLS-DA) plot using raw data. Because it is specialized in decrypting mass spectra and their corresponding molecular formulae, data manipulation and processing with Metaboseek is very limited, more limited than MetaboAnalyst is. It does produce PCA plots, but very little a user is able to tweak with data.

Purchasing the tools is another problem. It seems that both MetaboAnalyst and Metaboseek are free to use, however for SIMCA there is only a free 30-day trial until official purchase need to be made in order to continue using it. If one weighs more on the purchasing of tools upon choosing the "best" tool, then either MetaboAnalyst and Metaboseek may be better choices than SIMCA is.

Chapter 5. Conclusion

Researching through the dissertations that have used tools to analyze their metabolomics data have led to the creation of lengthy list of metabolomics data analysis tools, with MetaboAnalyst and SIMCA being two of the most commonly used tools by researchers. Metaboseek had been added to the list since it is a tool that has come out relatively recently and is still currently being developed, therefore it has not been widely known by other researchers yet and it is considered to be rapidly growing to become one of the frequently used tools. Evidently there are strengths as there are weaknesses to each of those three tools, and understanding those points and choosing the most favorable and desirable tool of one's choice for their particular data is intended to produce the best results from one's data. It is never the right thing to say which tool is the right and the best tool for any metabolomics data analysis, but it is certainly safe to say that although there are a number of tools that are popularly used among researchers more than other tools, each and every tool including the tools that were seldomly cited or even the tools that were not cited on this research at all, researchers may find them also useful for their own study which means that these are not tools of a poor quality at the least. Therefore, it can be concluded that each and every metabolomics data analyzing tool, although some of them are preferred by many researchers, has its own strengths and is highly

possible for some researchers would not choose the tools that majority of people use, depending on the types of results they would like to compute out from inputting their data, as well as their preferred tool formats.

Bibliography

1. Saccenti, E., & Timmerman, M. E. (2016, June 20). *Approaches to sample size determination for multivariate data: Applications to PCA and PLS-DA of Omics Data* Retrieved December 20, 2022, from https://pubs.acs.org/doi/pdf/10.1021/acs.jproteome.5b01029
2. Saccenti, E., Hoefsloot, H.C.J., Smilde, A.K. *et al.* Reflections on univariate and multivariate analysis of metabolomics data. *Metabolomics* **10**, 361–374 (2014). https://doi.org/10.1007/s11306-013-0598-6
3. Ruiz-Perez, D., Guan, H., Madhivanan, P. *et al.* So you think you can PLS-DA?. *BMC Bioinformatics* **21** (Suppl 1), 2 (2020). https://doi.org/10.1186/s12859-019-3310-7
4. Ossom Williamson P, Minter CIJ. Exploring PubMed as a reliable resource for scholarly communications services. J Med Libr Assoc. 2019 Jan;107(1):16-29. doi: 10.5195/jmla.2019.433. Epub 2019 Jan 1. PMID: 30598645; PMCID: PMC6300231.
5. Xiaojie Liu, Caichun Liu, Junsheng Tian, Xiaoxia Gao, Ke Li, Guanhua Du, Xuemei Qin, Plasma metabolomics of depressed patients and treatment with Xiaoyaosan based on mass spectrometry technique, Journal of Ethnopharmacology, Volume 246, 2020, 112219, ISSN 0378-8741, https://doi.org/10.1016/j.jep.2019.112219. (https://www.sciencedirect.com/science/article/pii/S0378874119319555)
6. Cao B, Liu S, Yang L, Chi A. Changes of Differential Urinary Metabolites after High-Intensive Training in Teenage Football Players. Biomed Res Int. 2020 Mar 18;2020:2073803. doi: 10.1155/2020/2073803. PMID: 32258106; PMCID: PMC7109581.
7. Dickson L, Tenon M, Svilar L, Fança-Berthon P, Lugan R, Martin JC, Vaillant F, Rogez H. Main Human Urinary Metabolites after Genipap (*Genipa americana* L.) Juice Intake. Nutrients. 2018 Aug 24;10(9):1155. doi: 10.3390/nu10091155. PMID: 30149503; PMCID:

PMC6165415.

8. Romo-Hualde A, Huerta AE, González-Navarro CJ, Ramos-López O, Moreno-Aliaga MJ, Martínez JA. Untargeted metabolomic on urine samples after α-lipoic acid and/or eicosapentaenoic acid supplementation in healthy overweight/obese women. Lipids Health Dis. 2018 May 9;17(1):103. doi: 10.1186/s12944-018-0750-4. PMID: 29743087; PMCID: PMC5941619.

9. Yang M, Yu Z, Chen X, Guo Z, Deng S, Chen L, Wu Q, Liang F. Active Acupoints Differ from Inactive Acupoints in Modulating Key Plasmatic Metabolites of Hypertension: A Targeted Metabolomics Study. Sci Rep. 2018 Dec 13;8(1):17824. doi: 10.1038/s41598-018-36199-1. PMID: 30546033; PMCID: PMC6292875.

10. Kang DW, Adams JB, Vargason T, Santiago M, Hahn J, Krajmalnik-Brown R. Distinct Fecal and Plasma Metabolites in Children with Autism Spectrum Disorders and Their Modulation after Microbiota Transfer Therapy. mSphere. 2020 Oct 21;5(5):e00314-20. doi: 10.1128/mSphere.00314-20. PMID: 33087514; PMCID: PMC7580952.

11. Wang PS, Kuo CH, Yang HC, Liang YJ, Huang CJ, Sheen LY, Pan WH. Postprandial Metabolomics Response to Various Cooking Oils in Humans. J Agric Food Chem. 2018 May 16;66(19):4977-4984. doi: 10.1021/acs.jafc.8b00530. Epub 2018 May 7. PMID: 29716192.

12. Paris D, Palomba L, Mirra V, Borrelli M, Corcione A, Santamaria F, Maniscalco M, Motta A. NMR Profiling of Exhaled Breath Condensate Defines Different Metabolic Phenotypes of Non-Cystic Fibrosis Bronchiectasis. Int J Mol Sci. 2020 Nov 14;21(22):8600. doi: 10.3390/ijms21228600. PMID: 33202684; PMCID: PMC7698311.

13. Lijing W, Sujie K, Linxi W, Lishan H, Liqin Q, Zhidong Z, Kejun W, Mengjun Z, Xiaoying L, Xiaohong L, Libin L. Altered Caffeine Metabolism Is Associated With Recurrent Hypoglycemia in Type 2 Diabetes Mellitus: A UPLC-MS-Based Untargeted Metabolomics Study. Front Endocrinol

(Lausanne). 2022 Jun 17;13:843556. doi: 10.3389/fendo.2022.843556. PMID: 35784552; PMCID: PMC9248032.

14. Madrid-Gambin F, Garcia-Aloy M, Vázquez-Fresno R, Vegas-Lozano E, Sánchez-Pla A, Misawa K, Hase T, Shimotoyodome A, Andres-Lacueva C. Metabolic Signature of a Functional High-Catechin Tea after Acute and Sustained Consumption in Healthy Volunteers through 1H NMR Based Metabolomics Analysis of Urine. J Agric Food Chem. 2019 Mar 20;67(11):3118-3124. doi: 10.1021/acs.jafc.8b04198. Epub 2019 Mar 7. PMID: 30574780.

15. Lende TH, Austdal M, Bathen TF, Varhaugvik AE, Skaland I, Gudlaugsson E, Egeland NG, Lunde S, Akslen LA, Jonsdottir K, Janssen EAM, Søiland H, Baak JPA. Metabolic consequences of perioperative oral carbohydrates in breast cancer patients - an explorative study. BMC Cancer. 2019 Dec 4;19(1):1183. doi: 10.1186/s12885-019-6393-7. PMID: 31801490; PMCID: PMC6894229.

16. Pan X, Cunningham EL, Passmore AP, McGuinness B, McAuley DF, Beverland D, O'Brien S, Mawhinney T, Schott JM, Zetterberg H, Green BD. Cerebrospinal Fluid Spermidine, Glutamine and Putrescine Predict Postoperative Delirium Following Elective Orthopaedic Surgery. Sci Rep. 2019 Mar 12;9(1):4191. doi: 10.1038/s41598-019-40544-3. PMID: 30862889; PMCID: PMC6414730.

17. McNairn M, Brito A, Dillard K, Heath H, Pantaleon M, Fanter R, Pilolla K, Amin S, La Frano MR. Postprandial Dried Blood Spot-Based Nutritional Metabolomic Analysis Discriminates a High-Fat, High-Protein Meat-Based Diet from a High Carbohydrate Vegan Diet: A Randomized Controlled Crossover Trial. J Acad Nutr Diet. 2021 May;121(5):931-941.e2. doi: 10.1016/j.jand.2020.10.024. Epub 2020 Dec 3. PMID: 33279463.

18. Luo Y, Wang L, Liu XY, Chen X, Song YX, Li XH, Jiang C, Peng A, Liu JY. Plasma profiling of amino acids distinguishes acute gout from asymptomatic hyperuricemia. Amino Acids.

2018 Nov;50(11):1539-1548. doi: 10.1007/s00726-018-2627-2. Epub 2018 Aug 2. PMID: 30073607.

19. Huang L, Zhang L, Li T, Liu YW, Wang Y, Liu BJ. Human Plasma Metabolomics Implicates Modified 9-cis-Retinoic Acid in the Phenotype of Left Main Artery Lesions in Acute ST-Segment Elevated Myocardial Infarction. Sci Rep. 2018 Aug 28;8(1):12958. doi: 10.1038/s41598-018-30219-w. PMID: 30154509; PMCID: PMC6113282.

20. Xu D, Wang S, Feng M, Shete V, Chu Y, Kamil A, Yang C, Liu H, Xia H, Wang X, Sun G, Yang Y. Serum Metabolomics Reveals Underlying Mechanisms of Cholesterol-Lowering Effects of Oat Consumption: A Randomized Controlled Trial in a Mildly Hypercholesterolemic Population. Mol Nutr Food Res. 2021 May;65(9):e2001059. doi: 10.1002/mnfr.202001059. Epub 2021 Apr 21. PMID: 33793078.

21. Chen KC, Tsai SW, Zhang X, Zeng C, Yang HY. The investigation of the volatile metabolites of lung cancer from the microenvironment of malignant pleural effusion. Sci Rep. 2021 Jun 30;11(1):13585. doi: 10.1038/s41598-021-93032-y. PMID: 34193905; PMCID: PMC8245642.

22. Peng ML, Han J, Li LL, Ma HT. Metabolomics reveals the mechanism of (-)-hydroxycitric acid promotion of protein synthesis and inhibition of fatty acid synthesis in broiler chickens. Animal. 2018 Apr;12(4):774-783. doi: 10.1017/S175173111700221X. Epub 2017 Sep 7. PMID: 28877777.

23. Metwaly S, Côté A, Donnelly SJ, Banoei MM, Lee CH, Andonegui G, Yipp BG, Vogel HJ, Fiehn O, Winston BW. ARDS metabolic fingerprints: characterization, benchmarking, and potential mechanistic interpretation. Am J Physiol Lung Cell Mol Physiol. 2021 Jul 1;321(1):L79-L90. doi: 10.1152/ajplung.00077.2021. Epub 2021 May 5. PMID: 33949201.

24. Zhang C, Zhou S, Chang H, Zhuang F, Shi Y, Chang L, Ai W, Du J, Liu W, Liu H, Zhou X, Wang Z, Hong T. Metabolomic Profiling Identified Serum Metabolite Biomarkers and Related

Metabolic Pathways of Colorectal Cancer. Dis Markers. 2021 Dec 7;2021:6858809. doi: 10.1155/2021/6858809. PMID: 34917201; PMCID: PMC8670981.

25. Zhao S, Liu H, Su Z, Khoo C, Gu L. Identifying Cranberry Juice Consumers with Predictive OPLS-DA Models of Plasma Metabolome and Validation of Cranberry Juice Intake Biomarkers in a Double-Blinded, Randomized, Placebo-Controlled, Cross-Over Study. Mol Nutr Food Res. 2020 Jun;64(11):e1901242. doi: 10.1002/mnfr.201901242. Epub 2020 Apr 24. PMID: 32281738.

26. Park SA, Son SY, Lee AY, Park HG, Lee WL, Lee CH. Metabolite Profiling Revealed That a Gardening Activity Program Improves Cognitive Ability Correlated with BDNF Levels and Serotonin Metabolism in the Elderly. Int J Environ Res Public Health. 2020 Jan 15;17(2):541. doi: 10.3390/ijerph17020541. PMID: 31952145; PMCID: PMC7014360.

27. Gardinassi LG, Arévalo-Herrera M, Herrera S, Cordy RJ, Tran V, Smith MR, Johnson MS, Chacko B, Liu KH, Darley-Usmar VM, Go YM; MaHPIC Consortium, Jones DP, Galinski MR, Li S. Integrative metabolomics and transcriptomics signatures of clinical tolerance to Plasmodium vivax reveal activation of innate cell immunity and T cell signaling. Redox Biol. 2018 Jul;17:158-170. doi: 10.1016/j.redox.2018.04.011. Epub 2018 Apr 11. PMID: 29698924; PMCID: PMC6007173.

28. Hagos FT, Empey PE, Wang P, Ma X, Poloyac SM, Bayir H, Kochanek PM, Bell MJ, Clark RSB. Exploratory Application of Neuropharmacometabolomics in Severe Childhood Traumatic Brain Injury. Crit Care Med. 2018 Sep;46(9):1471-1479. doi: 10.1097/CCM.0000000000003203. PMID: 29742587; PMCID: PMC6095742.

29. Dos Santos Fechine CPN, Monteiro MGCA, Tavares JF, Souto AL, Luna RCP, da Silva CSO, da Silva JA, Dos Santos SG, de Carvalho Costa MJ, Persuhn DC. Choline Metabolites, Hydroxybutyrate and HDL after Dietary Fiber

Supplementation in Overweight/Obese Hypertensive Women: A Metabolomic Study. Nutrients. 2021 Apr 24;13(5):1437. doi: 10.3390/nu13051437. PMID: 33923171; PMCID: PMC8146352.

30. Danlos, FX., Grajeda-Iglesias, C., Durand, S. *et al.* Metabolomic analyses of COVID-19 patients unravel stage-dependent and prognostic biomarkers. *Cell Death Dis* **12,** 258 (2021). https://doi.org/10.1038/s41419-021-03540-y

31. Kim H, Lichtenstein AH, Wong KE, Appel LJ, Coresh J, Rebholz CM. Urine Metabolites Associated with the Dietary Approaches to Stop Hypertension (DASH) Diet: Results from the DASH-Sodium Trial. Mol Nutr Food Res. 2021 Feb;65(3):e2000695. doi: 10.1002/mnfr.202000695. Epub 2020 Dec 28. PMID: 33300290; PMCID: PMC7967699.

32. Kim H, Lichtenstein AH, White K, Wong KE, Miller ER 3rd, Coresh J, Appel LJ, Rebholz CM. Plasma Metabolites Associated with a Protein-Rich Dietary Pattern: Results from the OmniHeart Trial. Mol Nutr Food Res. 2022 Mar;66(6):e2100890. doi: 10.1002/mnfr.202100890. Epub 2022 Feb 5. PMID: 35081272; PMCID: PMC8930517.

33. Renai L, Ancillotti C, Ulaszewska M, Garcia-Aloy M, Mattivi F, Bartoletti R, Del Bubba M. Comparison of chemometric strategies for potential exposure marker discovery and false-positive reduction in untargeted metabolomics: application to the serum analysis by LC-HRMS after intake of Vaccinium fruit supplements. Anal Bioanal Chem. 2022 Feb;414(5):1841-1855. doi: 10.1007/s00216-021-03815-5. Epub 2022 Jan 14. PMID: 35028688.

34. Chatterjee R, Davenport CA, Kwee L, D'Alessio D, Svetkey LP, Lin PH, Slentz CA, Ilkayeva O, Johnson J, Edelman D, Shah SH. Preliminary evidence of effects of potassium chloride on a metabolomic path to diabetes and cardiovascular disease. Metabolomics. 2020 Jun 18;16(7):75. doi: 10.1007/s11306-020-01696-w. PMID: 32556595; PMCID: PMC8053254.

35. Jang HH, Noh H, Kim HW, Cho SY, Kim HJ, Lee SH, Lee SH,

Gunter MJ, Ferrari P, Scalbert A, Freisling H, Kim JB, Choe JS, Kwon O. Metabolic tracking of isoflavones in soybean products and biosamples from healthy adults after fermented soybean consumption. Food Chem. 2020 Nov 15;330:127317. doi: 10.1016/j.foodchem.2020.127317. Epub 2020 Jun 11. PMID: 32569934.

36. Ulven SM, Christensen JJ, Nygård O, Svardal A, Leder L, Ottestad I, Lysne V, Laupsa-Borge J, Ueland PM, Midttun Ø, Meyer K, McCann A, Andersen LF, Holven KB. Using metabolic profiling and gene expression analyses to explore molecular effects of replacing saturated fat with polyunsaturated fat-a randomized controlled dietary intervention study. Am J Clin Nutr. 2019 May 1;109(5):1239-1250. doi: 10.1093/ajcn/nqy356. Erratum in: Am J Clin Nutr. 2020 Jan 1;111(1):232-236. PMID: 31051508; PMCID: PMC6499508.

37. Esquivel A, Alechaga É, Monfort N, Ventura R. Sulfate metabolites improve retrospectivity after oral testosterone administration. Drug Test Anal. 2019 Mar;11(3):392-402. doi: 10.1002/dta.2529. Epub 2018 Nov 20. PMID: 30362276.

38. Debik J, Euceda LR, Lundgren S, Gythfeldt HVL, Garred Ø, Borgen E, Engebraaten O, Bathen TF, Giskeødegård GF. Assessing Treatment Response and Prognosis by Serum and Tissue Metabolomics in Breast Cancer Patients. J Proteome Res. 2019 Oct 4;18(10):3649-3660. doi: 10.1021/acs.jproteome.9b00316. Epub 2019 Sep 16. PMID: 31483662.

39. Pigsborg K, Gürdeniz G, Rangel-Huerta OD, Holven KB, Dragsted LO, Ulven SM. Effects of changing from a diet with saturated fat to a diet with n-6 polyunsaturated fat on the serum metabolome in relation to cardiovascular disease risk factors. Eur J Nutr. 2022 Jun;61(4):2079-2089. doi: 10.1007/s00394-021-02796-6. Epub 2022 Jan 9. PMID: 34999928; PMCID: PMC9106625.

40. Bejder J, Gürdeniz G, Cuparencu C, Hall F, Gybel-Brask M, Breenfeldt Andersen A, Dragsted LO, Secher NH, Johansson PI, Nordsborg NB. An Untargeted Urine Metabolomics

Approach for Autologous Blood Transfusion Detection. Med Sci Sports Exerc. 2021 Jan;53(1):236-243. doi: 10.1249/MSS.0000000000002442. PMID: 32694367.

41. Razquin C, Liang L, Toledo E, Clish CB, Ruiz-Canela M, Zheng Y, Wang DD, Corella D, Castaner O, Ros E, Aros F, Gomez-Gracia E, Fiol M, Santos-Lozano JM, Guasch-Ferre M, Serra-Majem L, Sala-Vila A, Buil-Cosiales P, Bullo M, Fito M, Portoles O, Estruch R, Salas-Salvado J, Hu FB, Martinez-Gonzalez MA. Plasma lipidome patterns associated with cardiovascular risk in the PREDIMED trial: A case-cohort study. Int J Cardiol. 2018 Feb 15;253:126-132. doi: 10.1016/j.ijcard.2017.10.026. PMID: 29306452; PMCID: PMC5759348.

42. Nascimento MM, Alvarez AJ, Huang X, Browngardt C, Jenkins R, Sinhoreti MC, Ribeiro APD, Dilbone DA, Richards VP, Garrett TJ, Burne RA. Metabolic Profile of Supragingival Plaque Exposed to Arginine and Fluoride. J Dent Res. 2019 Oct;98(11):1245-1252. doi: 10.1177/0022034519869906. Epub 2019 Aug 27. PMID: 31454264; PMCID: PMC6755720.

43. Grunert T, Herzog R, Wiesenhofer FM, Vychytil A, Ehling-Schulz M, Kratochwill K. Vibrational Spectroscopy of Peritoneal Dialysis Effluent for Rapid Assessment of Patient Characteristics. Biomolecules. 2020 Jun 26;10(6):965. doi: 10.3390/biom10060965. PMID: 32604921; PMCID: PMC7357123.

44. Quartieri E, Marraccini C, Merolle L, Pulcini S, Buzzi M, Guardi M, Schiroli D, Baricchi R, Pertinhez TA. Metabolomics comparison of cord and peripheral blood-derived serum eye drops for the treatment of dry eye disease. Transfus Apher Sci. 2021 Aug;60(4):103155. doi: 10.1016/j.transci.2021.103155. Epub 2021 May 8. PMID: 33975808.

45. Braga DPAF, Montani DA, Setti AS, Turco EGL, Oliveira-Silva D, Borges E Jr. Metabolomic profile as a noninvasive adjunct tool for the diagnosis of Grades III and IV endometriosis-related infertility. Mol Reprod Dev. 2019 Aug;86(8):1044-1052. doi: 10.1002/mrd.23221. Epub 2019

Jun 18. PMID: 31215101.

46. Lindqvist HM, Gjertsson I, Andersson S, Calder PC, Bärebring L. Influence of blue mussel (Mytilus edulis) intake on fatty acid composition in erythrocytes and plasma phospholipids and serum metabolites in women with rheumatoid arthritis. Prostaglandins Leukot Essent Fatty Acids. 2019 Nov;150:7-15. doi: 10.1016/j.plefa.2019.08.004. Epub 2019 Aug 24. PMID: 31491682.

47. Kirchberg, Franca Fabiana*; Grote, Veit*; Gruszfeld, Dariusz †; Socha, Piotr †; Closa-Monasterolo, Ricardo †; Escribano, Joaquin †; Verduci, Elvira §; Mariani, Benedetta §; Langhendries, Jean-Paul||; Poncelet, Pascale ¶; Koletzko, Berthold*; Hellmuth, Christian*; for The European Childhood Obesity Trial Study Group Are All Breast-fed Infants Equal? Clustering Metabolomics Data to Identify Predictive Risk Clusters for Childhood Obesity, Journal of Pediatric Gastroenterology and Nutrition: March 2019 - Volume 68 - Issue 3 - p 408-415 doi: 10.1097/MPG.0000000000002184

48. Marques JG, Shokry E, Frivolt K, Werkstetter KJ, Brückner A, Schwerd T, Koletzko S, Koletzko B. Metabolomic Signatures in Pediatric Crohn's Disease Patients with Mild or Quiescent Disease Treated with Partial Enteral Nutrition: A Feasibility Study. SLAS Technol. 2021 Apr;26(2):165-177. doi: 10.1177/2472630320969147. Epub 2020 Nov 18. PMID: 33207993; PMCID: PMC7985853.

49. Jangsiripornpakorn J, Srisuk S, Chailurkit L, Nimitphong H, Saetung S, Ongphiphadhanakul B. The glucose-lowering effect of low-dose diacerein and its responsiveness metabolic markers in uncontrolled diabetes. BMC Res Notes. 2022 Mar 4;15(1):91. doi: 10.1186/s13104-022-05974-9. PMID: 35246243; PMCID: PMC8896078.

50. Saghatelian A, Trauger SA, Want EJ, Hawkins EG, Siuzdak G, Cravatt BF. Assignment of endogenous substrates to enzymes by global metabolite profiling. Biochemistry. 2004 Nov 16;43(45):14332-9. doi: 10.1021/bi0480335. PMID:

15533037.

Abstract

대사체학 데이터를 분석하는 것은 그 대사체학 데이터를 분석하는 도구나 소프트웨어에 입력하는 단계로서, 그 도구는 그 데이터를 분석하여 가장 최적의 결과를 생성한다. 하지만 대사체학 데이터 분석 도구가 무수히 많다는 사실을 고려하였을 때, 다양한 대사체학 분석의 성향을 보아 어떤 대사체학 도구가 가장 적합한 도구이고, 작업하기 제일 쉬운 도구는 무엇일까? 본 연구는 다른 논문에서 사용된 다양한 대사체학 데이터 분석 도구를 수집하고, 가장 많이 사용된 도구들을 간추려서 그 간추려진 도구들이 가지고 있는 장단점을 분석하여, 마지막으로 그 도구들 중 각 사례들에 어떤 도구가 “최고의” 도구인지 평가하는 것에 중점을 두고 있다. 총 45 편의 논문이 초기에 편집되었으며, 그 중 impact factor (IF) 가 3 미만인 15 편을 제거하고 나머지 논문을 정리하여 어느 대사체학 데이터 분석 도구를 썼는지 분류하였다. 도구 목록 중 가장 많이 인용된 두 가지 도구 (MetaboAnalyst 와 SIMCA) 와 최근에 개발된 세 번째 도구 (Metaboseek)를 언급하고 장단점을 철저히 분석하고, 마지막으로 직접 데이터를 입력하여 해당 도구가 생성한 결과를 관찰하였다. 이 연구는 대사체학 데이터를 분석하고자 하는 연구자들이 본인의 연구에 가장 최적화되고 “최고인” 도구가 무엇인지 이해하기를 바라는 차원에서 수행되었다.

Supplementary Materials

Tool Name	Tool Description	Samples and Analytes that were Used	JIF (2022)	Keywords (PubMed Advanced)
MetaboAnalyst	A web-based tool that supports not only the analysis of metabolomic data but also its interpretation and integration with other omics data	Plasma [5]	4.36	Date - Publication: 2018/01/01 to present Title/Abstract: metabolomics data Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial

		Urine [6]	3.411	Date – Publication: 2018/01/01 to present Title: metabolites Title/Abstract: Principal Component Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Urine [7]	5.429	Date – Publication: 2018/01/01 to

				present Title: metabolites Title/Abstract: Principal Component Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Urine [8]	4.15	Date - Publication: 2018/01/01 to present Title/Abstract:

				metabolites Title/Abstract: Principal Component Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Plasma [9]	4.379	Date - Publication: 2018/01/01 to present Title/Abstract: metabolomics data

				Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Plasma and fecal samples [10]	4.389	Date – Publication: 2018/01/01 to present Title: metabolites Article Type: books and documents, clinical trial, meta-analysis, randomized

				controlled trial
		Plasma [11]	5.279	Date – Publication: 2018/01/01 to present Title: metabolomics Title/Abstract: Partial Least Squares–Discriminant Analysis Article Type: books and documents, clinical trial, meta–analysis, randomized

				controlled trial
		Exhaled breath condensate [12]	6.01	Date – Publication: 2018/01/01 to present Title/Abstract: metabolites Title/Abstract: Partial Least Squares–Discriminant Analysis Article Type: books and documents, clinical trial, meta–analysis, randomized

				controlled trial
		cord serum samples from newborns at the time of delivery [44]	2.07 (2021)	Date – Publication: 2018/01/01 to present Title: metabolomics Title/Abstract: Partial Least Squares–Discriminant Analysis Article Type: books and documents, clinical trial, meta–analysis, randomized

				controlled trial
		Serum [13]	6.055	Date – Publication: 2018/01/01 to present Title: metabolomics Title/Abstract: Principal Component Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial

		Urine [14]	5.279	Date – Publication: 2018/01/01 to present Title: metabolomics Title/Abstract: Partial Least Squares–Discriminant Analysis Article Type: books and documents, clinical trial, meta–analysis, randomized controlled trial

		Serum from blood and tumor tissue [15]	4.638	Date – Publication: 2018/01/01 to present Title/Abstract: metabolites Title/Abstract: Partial Least Squares–Discriminant Analysis Article Type: books and documents, clinical trial, meta–analysis, randomized controlled trial

		cerebrospinal fluid [16]	4.996	Date – Publication: 2018/01/01 to present Title/Abstract: metabolites Title/Abstract: Partial Least Squares–Discriminant Analysis Article Type: books and documents, clinical trial, meta–analysis, randomized controlled trial

		Urine [17]	3.24	Date – Publication: 2018/01/01 to present Title/Abstract: metabolomics Title/Abstract: Principal Component Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Serum samples [45]	2.56	Date – Publication:

				2018/01/01 to present Title/Abstract: metabolomics Title/Abstract: Partial Least Squares–Discriminant Analysis Article Type: books and documents, clinical trial, meta–analysis, randomized controlled trial
SIMCA	User–friendly	Plasma [5]	5.09	Date – Publication:

	software developed by Umetrics chiefly for the analyses of PCA and PLS regression			2018/01/01 to present Title/Abstract: metabolomics data Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Urine [6]	3.411	Date - Publication: 2018/01/01 to present Title: metabolites Title/Abstract: Principal

				Component Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Urine [7]	6.706	Date - Publication: 2018/01/01 to present Title: metabolites Title/Abstract: Principal Component Analysis

				Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Plasma [18]	3.52	Date - Publication: 2018/01/01 to present Title/Abstract: metabolomics Title/Abstract: Principal Component Analysis Article Type:

				books and documents, clinical trial, meta-analysis, randomized controlled trial
		Plasma from the venous blood [19]	4.996	Date – Publication: 2018/01/01 to present Title/Abstract: metabolomics data Article Type: books and documents, clinical trial, meta-analysis, randomized

				controlled trial
		Fatty acids in erythrocytes and in plasma phospholipids [46]	2.81	Date – Publication: 2018/01/01 to present Title/Abstract: metabolomics data Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Fasting blood samples [20]	5.914	Date – Publication: 2018/01/01 to present Title:

				metabolomics Title/Abstract: Partial Least Squares–Discriminant Analysis Article Type: books and documents, clinical trial, meta–analysis, randomized controlled trial
		Pleural effusion [21]	4.996	Date – Publication: 2018/01/01 to present Title: metabolites

				Title/Abstract: Partial Least Squares Discriminant Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Serum [22]	3.14	Date – Publication: 2018/01/01 to present Title: metabolomics

				Title/Abstract: Principal Component Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Serum [23]	4.93	Date - Publication: 2018/01/01 to present Title/Abstract: metabolites Title/Abstract:

				Partial Least Squares–Discriminant Analysis Article Type: books and documents, clinical trial, meta–analysis, randomized controlled trial
		Serum [24]	3.828	Date – Publication: 2018/01/01 to present Title/Abstract: metabolites Title/Abstract:

				Partial Least Squares–Discriminant Analysis Article Type: books and documents, clinical trial, meta–analysis, randomized controlled trial
		Plasma [25]	5.914	Date – Publication: 2018/01/01 to present Title/Abstract: metabolites Title/Abstract:

				Partial Least Squares–Discriminant Analysis Article Type: books and documents, clinical trial, meta–analysis, randomized controlled trial
		Serum [26]	3.364	Date – Publication: 2018/01/01 to present Title/Abstract: metabolites Title/Abstract:

				Partial Least Squares–Discriminant Analysis Article Type: books and documents, clinical trial, meta–analysis, randomized controlled trial
Cytoscape	The open–source software project Cytoscape combines high–throughput expression data, various molecular	Human platelets [27]	10.787	Date – Publication: 2018/01/01 to present Title/Abstract: metabolomics data Article Type:

	states, and biomolecular interaction networks into a single conceptual framework			books and documents, clinical trial, meta-analysis, randomized controlled trial
	Cytoscape is most effective when combined with the vast databases of protein-protein, protein-DNA, and genetic connections that are becoming more and more accessible to humans and model	Cerebrospinal fluid [28]	7.598	Date - Publication: 2018/01/01 to present Title/Abstract: metabolomics Title/Abstract: Principal Component Analysis Article Type: books and

	organisms The Core program from Cytoscape offers the very minimum capabilities for network creation and querying, visual network integration with expression profiles, phenotypes, and other molecular states, and network connection to functional annotation databases			documents, clinical trial, meta–analysis, randomized controlled trial
SPSS	A statistical	Fatty acids in	2.81	Date –

	software that allows a user to dig deeper into their data through intuitive user interface, advanced data visualizations, automated data preparation, and more	erythrocytes and in plasma phospholipids [46]		Publication: 2018/01/01 to present Title/Abstract: metabolomics data Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Serum [13]	6.055	Date - Publication: 2018/01/01 to present Title: metabolomics

				Title/Abstract: Principal Component Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Plasma [18]	3.52	Date - Publication: 2018/01/01 to present Title/Abstract: metabolomics Title/Abstract:

				Principal Component Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
Mestrenova	A spectral data analyzing software, that is able to analyze various data such as 1H, 13C or any other 1D NMR as well as any 2D	Serum [29]	6.706	Date – Publication: 2018/01/01 to present Title: metabolites Title/Abstract: Principal Component

	correlations, such as HSQC, HMBC, NOESY, COSY, TOCSY, etc.			Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
R	A statistical programming that is uniquely able to handle lots of data, main function being linear and nonlinear modelling, classical statistical tests, time-series	Urine [14]	5.279	Date - Publication: 2018/01/01 to present Title: metabolomics Title/Abstract: Partial Least Squares-Discriminant

	analysis, classification, clustering, as well as graphical techniques, and such more			Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Plasma [47]	2.839	Date - Publication: 2018/01/01 to present Title/Abstract: metabolomics data Article Type: books and documents, clinical trial, meta-

				analysis, randomized controlled trial
		Human peripheral blood [30]	8.469	Date – Publication: 2018/01/01 to present Title/Abstract: metabolites Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Urine [31]	5.914	Date – Publication: 2018/01/01 to

				present Title: metabolites Title/Abstract: Partial Least Squares Discriminant Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Pleural effusion [21]	4.996	Date - Publication: 2018/01/01 to present

			Title: metabolites Title/Abstract: Partial Least Squares Discriminant Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Plasma [32]	5.914	Date - Publication: 2018/01/01 to present Title: metabolites

				Title/Abstract: Partial Least Squares Discriminant Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Serum [33]	4.142	Date – Publication: 2018/01/01 to present Title: metabolomics

				Title/Abstract: Principal Component Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Plasma [34]	4.29	Date - Publication: 2018/01/01 to present Title/Abstract: metabolites Title/Abstract:

				Principal Component Analysis Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
		Plasma [48]	2.67	Date - Publication: 2018/01/01 to present Title/Abstract: metabolites Title/Abstract: Partial Least

				Squares–Discriminant Analysis Article Type: books and documents, clinical trial, meta–analysis, randomized controlled trial
		Plasma [35]	7.514	Date – Publication: 2018/01/01 to present Title/Abstract: metabolites Title/Abstract: Partial Least

				Squares–Discriminant Analysis Article Type: books and documents, clinical trial, meta–analysis, randomized controlled trial
		Plasma [49]	1.66	Date – Publication: 2018/01/01 to present Title/Abstract: metabolomics Title/Abstract: Partial Least

				Squares–Discriminant Analysis Article Type: books and documents, clinical trial, meta–analysis, randomized controlled trial
		Serum [36]	7.045	Date – Publication: 2018/01/01 to present Title/Abstract: metabolomics Title/Abstract: Partial Least

				Squares–Discriminant Analysis Article Type: books and documents, clinical trial, meta–analysis, randomized controlled trial
Compound Discoverer	Developed by Thermo Fisher Scientific, uses chromatographic and mass spectra (MS) data and streamlines compound	Human peripheral blood [30]	8.469	Date – Publication: 2018/01/01 to present Title/Abstract: metabolites Article Type: books and

	identification as well as comparative analyses, and provides extensive filtering and data visualization capabilities			documents, clinical trial, meta-analysis, randomized controlled trial
MATLAB	A high-performance language for technical computing that combines computation, visualization, and programming	Urine [37]	3.26	Date - Publication: 2018/01/01 to present Title: metabolites Title/Abstract: Principal Component Analysis Article Type: books and

				documents, clinical trial, meta-analysis, randomized controlled trial
		Tissue samples [38]	4.466	Date - Publication: 2018/01/01 to present Title: metabolomics Title/Abstract: Partial Least Squares-Discriminant Analysis Article Type: books and

				documents, clinical trial, meta-analysis, randomized controlled trial
		Serum [39]	5.614	Date - Publication: 2018/01/01 to present Title/Abstract: metabolites Title/Abstract: Principal Component Analysis Article Type: books and documents, clinical

				trial, meta-analysis, randomized controlled trial
		Urine [40]	5.23	Date - Publication: 2018/01/01 to present Title: metabolomics Title/Abstract: Partial Least Squares-Discriminant Analysis Article Type: books and documents, clinical

				trial, meta–analysis, randomized controlled trial
STATA	An integrated software that allows data manipulation, visualization, statistics, and automated reporting	Plasma known lipid metabolites a [41]	4.614	Date – Publication: 2018/01/01 to present Title/Abstract: metabolites Title/Abstract: Principal Component Analysis Article Type: books and documents, clinical trial, meta–

				analysis, randomized controlled trial
BioEStat	Free software developed for undergraduates and graduates, and with easy-to-perform procedures, able to carry out various statistical and graphical analyses	Site-specific supragingival plaque samples [42]	6.116	Date - Publication: 2018/01/01 to present Title/Abstract: metabolites Title/Abstract: Principal Component Analysis Article Type: books and documents, clinical trial, meta-analysis,

				randomized controlled trial
Unscrambler X	Software for multivariate data analyses, in which calibration of data is often used in the application of analytical data and development of predictive models for use in real−time spectroscopic analysis of materials	Peritoneal dialysis effluent [43]	4.569	Date − Publication: 2018/01/01 to present Title/Abstract: metabolites Title/Abstract: Principal Component Analysis Article Type: books and documents, clinical trial, meta−analysis, randomized

				controlled trial

Supplementary Table 1. Table presenting all of the dissertations that were searched up on PubMed using specific keywords. The dissertations with the same colors represent identical ones, describing a single paper has used multiple metabolomic data analysis tools. The number of papers in total came out to be 54 papers, and removing the overlapping 9 papers, the number of papers was 45 papers.

Tool Name	Tool Description	Samples and Analytes that were Used	JIF (2022)	Keywords (PubMed Advanced)
Metaboanalyst	A web-based tool that supports not only the analysis of metabolomic data but also its interpretation and integration with other omics data	cord serum samples at the time of delivery [44]	2.07 (2021)	Date - Publication: 2018/01/01 to present Title: metabolomics Title/Abstract: Partial Least Squares-Discriminant Analysis Article Type: books and documents, clinical

				trial, meta-analysis, randomized controlled trial
		Serum samples [45]	2.56	Date - Publication: 2018/01/01 to present Title/Abstract: metabolomics Title/Abstract: Partial Least Squares-Discriminant Analysis Article Type: books and documents, clinical

				trial, meta-analysis, randomized controlled trial
SIMCA	User-friendly software developed by Umetrics chiefly for the analyses of PCA and PLS regression	Fatty acids in erythrocytes and in plasma phospholipids [46]	2.81	Date - Publication: 2018/01/01 to present Title/Abstract: metabolomics data Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
SPSS	A statistical	Fatty acids in	2.81	Date - Publication:

	software that allows a user to dig deeper into their data through intuitive user interface, advanced data visualizations, automated data preparation, and more	erythrocytes and in plasma phospholipids [46]		2018/01/01 to present Title/Abstract: metabolomics data Article Type: books and documents, clinical trial, meta-analysis, randomized controlled trial
R	A statistical programming that is uniquely able to handle lots of data, main function being linear and	Plasma [47]	2.839	Date – Publication: 2018/01/01 to present Title/Abstract: metabolomics data Article Type:

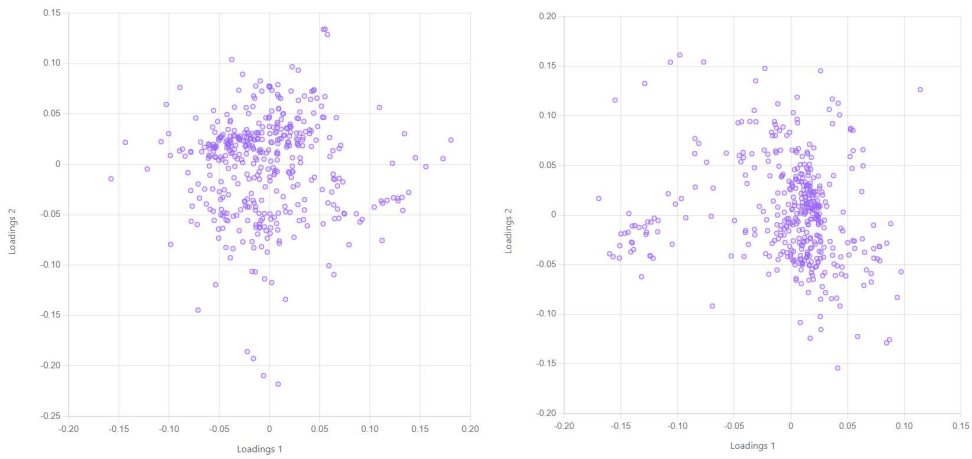
	nonlinear modelling, classical statistical tests, time−series analysis, classification, clustering, as well as graphical techniques, and such more			books and documents, clinical trial, meta−analysis, randomized controlled trial
		Plasma [48]	2.67	Date − Publication: 2018/01/01 to present Title/Abstract: metabolites Title/Abstract: Partial Least Squares−Discriminant Analysis Article Type:

				books and documents, clinical trial, meta-analysis, randomized controlled trial
		Plasma [49]	1.66	Date - Publication: 2018/01/01 to present Title/Abstract: metabolomics Title/Abstract: Partial Least Squares-Discriminant Analysis Article Type:

				books and documents, clinical trial, meta-analysis, randomized controlled trial

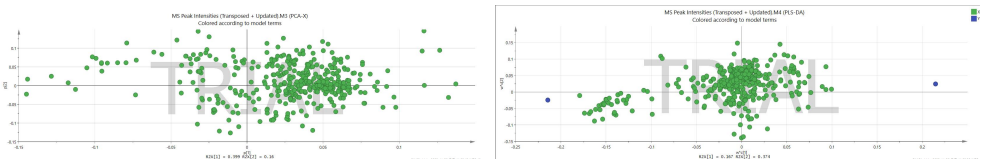
Supplementary Table 2. List of dissertations that have been removed due to their impact factors (IF) being less than three were tabulated. The dissertations with the same colors represent identical ones, describing a single paper has used multiple metabolomic data analyzing tools. A total of 7 papers, in which 1 paper overlapped, thereby resulting in 6 non-overlapping papers that were removed from the initial table (i.e., Supplementary Table 1.) and the final format could be seen on Table 1.

a) b)



Supplementary Figure 1. Loadings plots of PCA and PLS-DA using MetaboAnalyst. (a) PCA loadings plot on MetaboAnalyst for data from Saghatelian *et al.* Wildtype (WT) mice for fatty acid amide hydrolase, FAAH(+/+), compared with knockout (KO) mice, FAAH(−/−), is shown. (b) PLS-DA loadings plot for data from Saghatelian *et al.* WT FAAH(+/+) compared with FAAH(−/−) is shown.

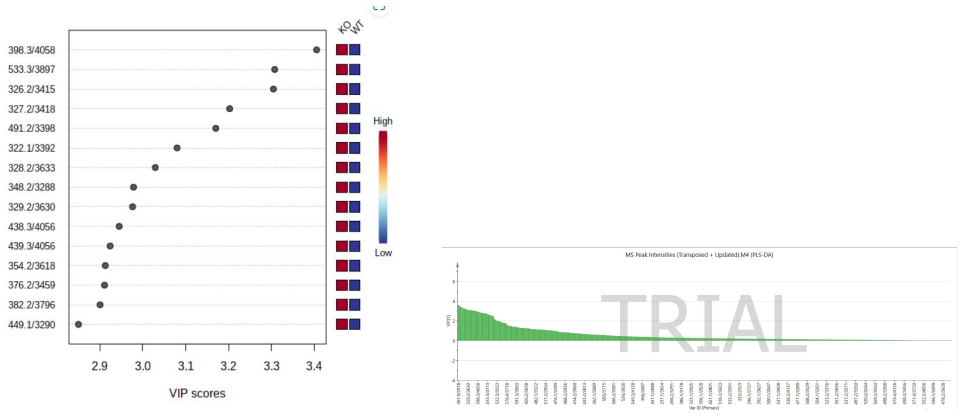
a) b)



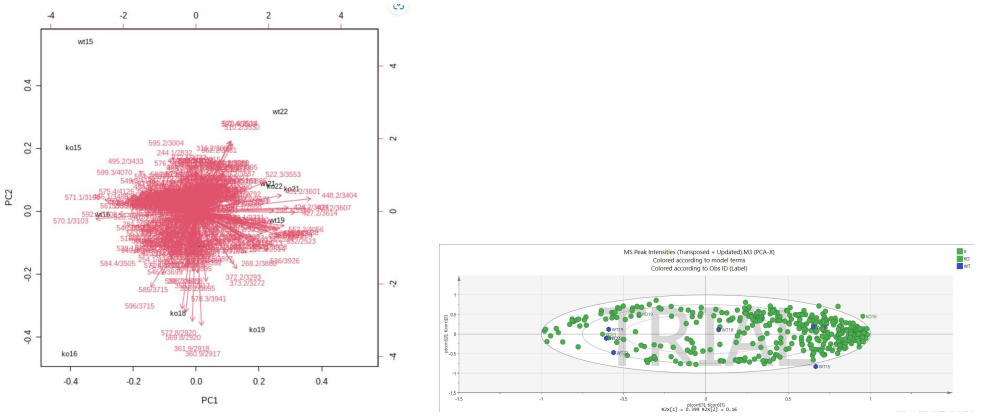
Supplementary Figure 2. Loadings plots of PCA and PLS-DA using SIMCA. Data from Saghatelian et al. are shown in (a) PCA loadings plotted on SIMCA. It is proven that wildtype (WT) mice for fatty acid amide hydrolase, FAAH(+/+), are different from knockout (KO) mice for FAAH(−/−). Data from Saghatelian et al. are shown in (b) a PLS-DA loadings plot. It is illustrated how WT

FAAH(+/+) compares to FAAH(−/−).

a) b)



Supplementary Figure 3. Variable importance in projection (VIP) plots of PLS−DA using MetaboAnalyst and SIMCA. (a) MetaboAnalyst VIP features for information from Saghatelian et al. It is proven that wildtype (WT) mice for fatty acid amide hydrolase, FAAH(+/+), are different from knockout (KO) mice for FAAH(−/−). (b) VIP characteristics for Saghatelian et al. data on SIMCA. It is illustrated how WT FAAH(+/+) compares to FAAH(−/−). It is possible to zoom in to see each feature in detail. b)



Supplementary Figure 4. Biplots of PCA using MetaboAnalyst and

SIMCA. (a) Biplots on MetaboAnalyst for data from Saghatelian *et al*. Wildtype (WT) mice for fatty acid amide hydrolase, FAAH(+/+), compared with knockout (KO) mice, FAAH(−/−), is shown. (b) Biplots on SIMCA for data from Saghatelian *et al*. WT FAAH(+/+) compared with FAAH(−/−) is shown. Zooming in to observe each feature is available.