2020

Development of an R script for simple lipidomic and metabolomic data analysis

Suad Alshammari
Virginia Commonwealth University

Joshua M. Morriss
Virginia Commonwealth University

Daniel Contaifer
Virginia Commonwealth University

Dayanjan S. Wijesinghe
Virginia Commonwealth University

Monther Alsultan
Virginia Commonwealth University

Development of an R script for simple lipidomic and metabolomic data analysis

[1]Suad Alshammari, [1]Monther Alsultan, [1]Joshua M. Morriss, [1]Daniel Contaifer, [1]Dayanjan S. Wijesinghe

[1]Department of Pharmacotherapy and Outcomes Sciences, VCU School of Pharmacy, Richmond VA

VCU School of Pharmacy

Laboratory of Pharmacometabolomics & Companion Diagnostics

Introduction

Metabolomics and lipidomics studies result in the generation of vast quantities of data. Often the analysis of this data is undertaken in closed software environment with little to no access to enable the assessment of the accuracy of the underlying algorithms. As a result, data processed via different software pipelines often provide slightly different results further aggravating the existing problem of low reproducibility of published results. As a first step towards addressing this challenge, we began the development of LipidAnalyst, an R based lipidomics software pipeline. As a part of this project, we are creating a simple statistical analysis and graphing module in R called LipidAnalyst-Stat module to generate statistically accurate high resolution figures for use in publications. The results from the code developed was independently validated using several commercial software and also against a standard lipidomic and metabolomic data analysis package, namely Metaboanalyst, that is largely used in this domain.

Methods

LipidAnalyst-Stat module is being developed using R version 3.5.3 and currently includes the capability to undertake statistical analyses (e.g. ANOVA) and post-hoc tests (e.g. Tukey). In addition, the module can also graph resultant information as high resolution, publication quality, violin and box plots with automated notation of statistical significance. This module was tested for lipidomic and metabolomic data and the results were compared against commercial software and Metaboanalyst, a primary data analysis software tool used in metabolomic and lipidomic research. The data were log transformed and Pareto centered for statistical analysis and comparison of the two statistical algorithms. We used the same data generated for HFrEF Anakinra study and PCOS study study to test the accuracy of the LipidAnalyst-Stat module currently being developed. Final results were validated by using JMP.

Results

Code generated in house demonstrated the same results as those generated using commercial software (e.g. JMP 15.0) but was significantly different from results obtained by using the same data through the MetaboAnalyst pipeline.

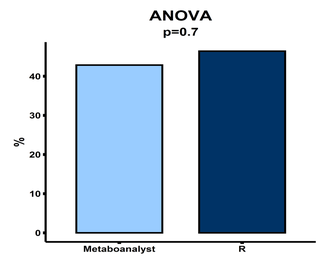

Figure 1: percentage of significant ANOVA for metabolites and lipids using R script compared to Metaboanalyst.(46.4% in R and 42.8% in Metaboanalyst). p-value calculated by using chi-square.

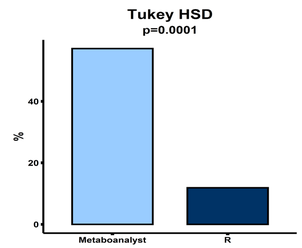

Figure 2: percentage of significant Tukey HSD for metabolites and lipids using R script compared to Metaboanalyst.(11.9% in R and 57.14% in Metaboanalyst). p-value calculated by using chi-square.

```
percentplot<- ggbarplot ( poster, x= "prog", y= "Percentage",color = "black", title = "ANOVA" ,ylab = "%", rug = TRUE,
                          font.label = list(size = 40, face = "bold"), linetype = "solid", size = 1.1,
                          palette = c("#99ccff", "#003366"),fill= "prog", order = c("R", "Metaboanalyst")+
                          border(color = "black", size= 2))+
  theme (plot.title = element_text(hjust=0.5 , face = "bold", size = 20 ),axis.text = element_text(face = "bold"))+
  theme (axis.title.x=element_blank(),element_line(size = 2))+
  theme (axis.title.y = element_text(face = "bold", size= 16 ))+
  theme(legend.position = "none")+ theme (axis.line= element_line (size= 1.2))+
  labs( subtitle = "p=0.7")+
  theme ( plot.subtitle = element_text(hjust=0.5 , face = "bold", size = 16))
percentplot
ggsave("percentplot.png",  height=5, width=5,units="in",dpi=1200,percentplot)
```

Figure 3: This figure represents the code that we used to plot figure 1 by using multiple packages and functions in R.

Table 1: Significant metabolites and lipids results using Mytaboanalyst.

metabolites and lipids	p-values	Tukey HSD
glycine	6.57E-12	PLACEBO-BASELINE; ANAK02-BASELINE; ANAK12-BASELINE; ANAK12-ANAK02
glutamine	1.37E-0.5	PLACEBO-BASELINE; ANAK02-BASELINE; ANAK12-BASELINE; ANAK12-PLACEBO; ANAK12-ANAK02
histidine	1.96E-0.5	PLACEBO-BASELINE; ANAK02-BASELINE; ANAK12-BASELINE; ANAK02-PLACEBO; ANAK12-PLACEBO
CRP	5.42E-0.5	ANAK12-BASELINE; ANAK12-PLACEBO; ANAK12-ANAK02
uracil	0.0014648	PLACEBO-BASELINE; ANAK02-BASELINE; ANAK12-BASELINE; ANAK02-PLACEBO; ANAK12-PLACEBO

Table2: significant metabolites and lipids results LipidAnalyst-Stat module.

metabolites and lipids	p-values	Tukey HSD
glycine	6.57E-12	ANAK02-BASELINE; ANAK12-BASELINE; BASELINE-PLACEBO
glutamine	1.37E-0.5	ANAK02-BASELINE; BASELINE-PLACEBO
histidine	1.96E-0.5	ANAK02 -BASELINE; ANAK12 - BASELINE; BASELINE-PLACEBO
CRP	5.42E-0.5	ANAK02-ANAK12; ANAK12-BASELINE; ANAK12 PLACEBO
uracil	0.0014648	ANAK02-BASELINE; ANAK12-BASELINE

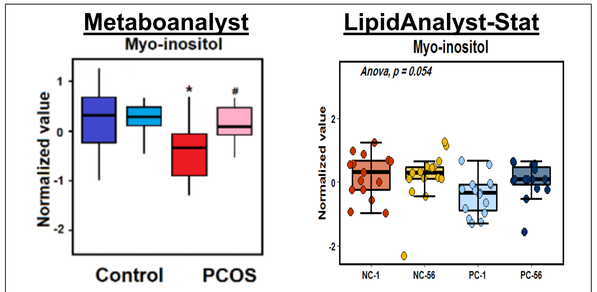

Figure 4: Comparison of ANOVA results for myo-inositol metabolite between Metaboanalyst and LipidAnalyst-Stat. Metaboanalyst show statistical significant while in our separately validated LipidAnalyst-Stat, the same level of significance was not observed.

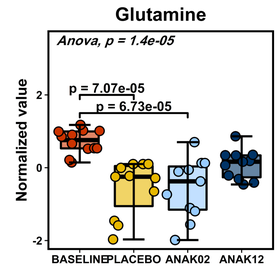

Figure 5: An example of a high-resolution figure (dpi=1200) showing a pairwise comparison to display significant levels.

Conclusions

This study demonstrates the danger of blindly using closed source software pipelines for the analysis of lipidomic and metabolomic data without also independently validating the analysis outcomes via separate software. Alternatively, we propose the use of open-source software such as what has been created by us that is also independently validated using multiple data such as JMP, wich 100% agree with what we have. Such code can then be published with the results to better enable transparency of data analysis and greatly increasing the replicability of results across different labs. Another limitation in metaboanalyst is the limited availability of graphs, graph designs and layouts, challenges that are being overcome in LipidAnalyst-Stat.

References

-Pinu FR, Goldansaz SA, Jaine J. Translational Metabolomics: Current Challenges and Future Opportunities. *Metabolites*. 2019;9(6). doi:10.3390/metabo9060108

Chong J, Wishart DS, Xia J. Using MetaboAnalyst 4.0 for Comprehensive and Integrative - Metabolomics Data Analysis. *Current Protocols in Bioinformatics*. 2019;68(1):e86. doi:10.1002/cpbi.86