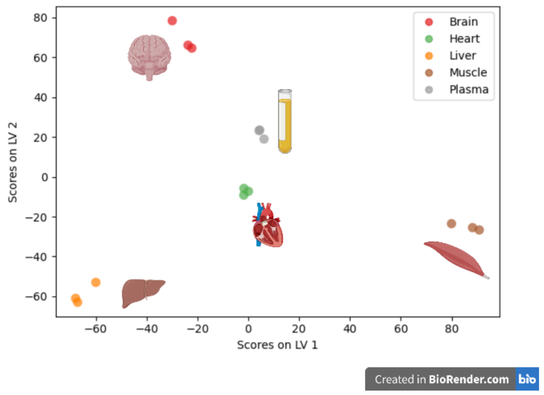
# Extraction of Metabolic Signatures from Untargeted Metabolomics Data in Public Data Repositories

Yunfei Liao; Aleksandr Smirnov; Xiuxia Du
University of North Carolina at Charlotte, Charlotte, NC

UNIVERSITY OF NORTH CAROLINA CHARLOTTE

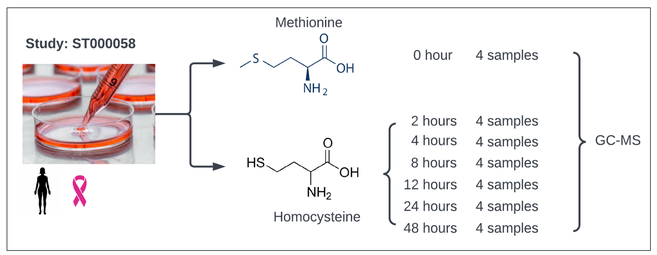
## Introduction

Biomarkers are indicative of diseases or abnormalities in human bodies and are invaluable in early disease diagnosis. Mass spectrometry-based untargeted metabolomics is a discovery tool and allows researchers to look for a set of metabolites that together can serve as disease biomarkers that are generally more applicable than a single metabolite in diagnosing diseases.




## Data

- Study ST000058 in the Metabolomics Workbench shows metabolites changes associated with methionine stress sensitivity of cancer using GC-MS analysis.




- Study MTBLS1033[1] of metabolic pathways and biomarkers associated with pelvic organ prolapse from the Metabolights repository. (Control samples: 59, POP disease samples: 45, and Pooled Quality Control samples: 14)

## Methods

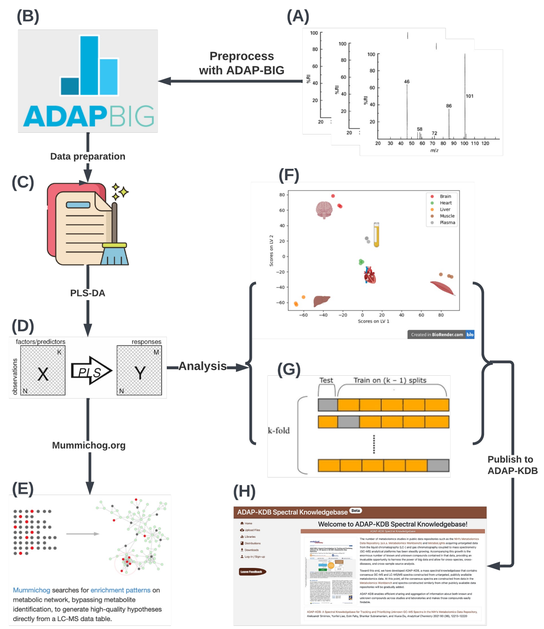



**Figure.1** Informatics workflow for extracting metabolic signatures. (A) Raw mass spec data; (B) Data preprocessing using ADAP-BIG software tool; (C) Data preparation—data cleaning such as handling null values and targets identification; (D) PLS-DA analysis; (E) Pathway analysis using Mummichog [2]; (F) PLS-DA scores plot; (G) Leave-one-out cross validation; (H) Publish results to the cloud resource ADAP-KDB at https://www.adap.cloud/ [3].

## Acknowledgement

We thank the funding support under the National Institutes of Health/National Cancer Institute grant U01CA235507 (PI: Xiuxia Du).

## Results

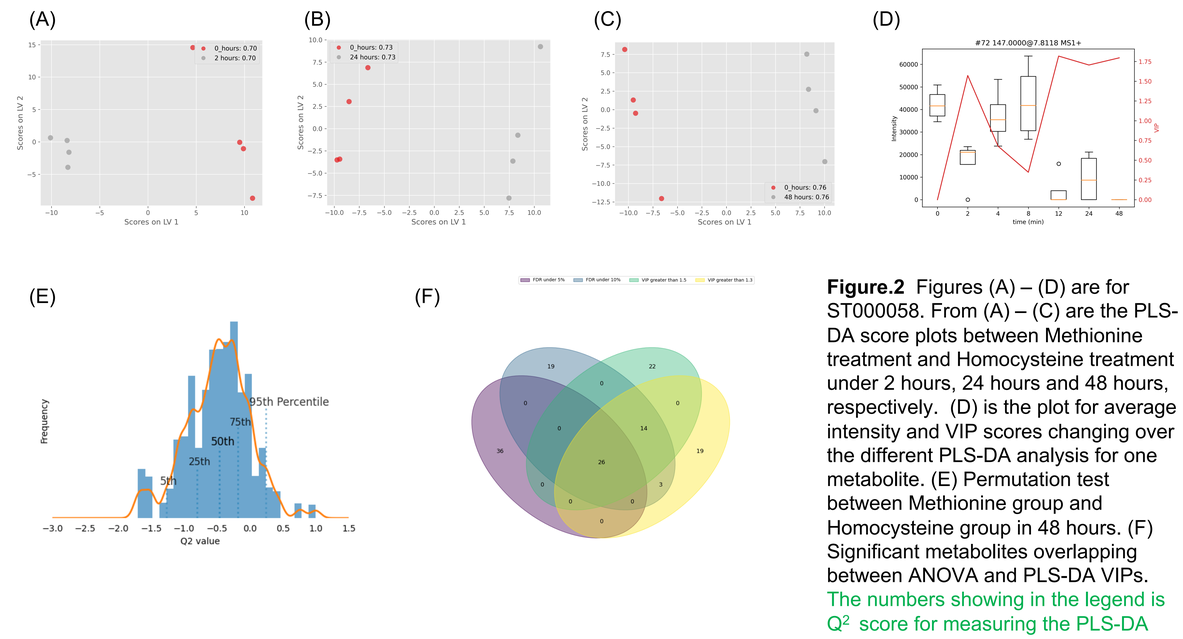
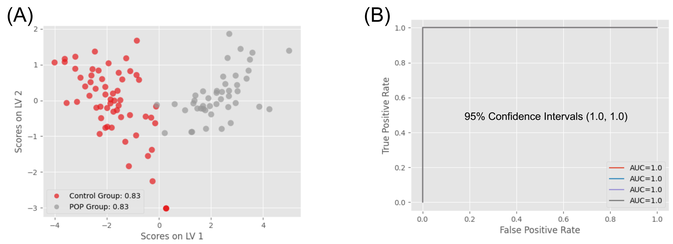



**Figure.2** Figures (A) – (D) are for ST000058. From (A) – (C) are the PLS-DA score plots between Methionine treatment and Homocysteine treatment under 2 hours, 24 hours and 48 hours, respectively. (D) is the plot for average intensity and VIP scores changing over the different PLS-DA analysis for one metabolite. (E) Permutation test between Methionine group and Homocysteine group in 48 hours. (F) Significant metabolites overlapping between ANOVA and PLS-DA VIPs. The numbers showing in the legend is $Q^2$ score for measuring the PLS-DA models.




(C)

| Pathways | overlap_size | pathway_size | p-value | overlap_EmpiricalCompounds |
|---|---|---|---|---|
| Biopterin metabolism | 3 | 3 | 0.00218 | E70,E72,E66 |
| Purine metabolism | 4 | 5 | 0.00395 | E70,E68,E56,E18 |
| Fructose and mannose metabolism | 2 | 2 | 0.00907 | E70,E54 |
| TCA cycle | 3 | 4 | 0.0142 | E35,E18,E70 |
| Methionine and cysteine metabolism | 2 | 3 | 0.0568 | E22,E25 |

**Figure.3** Figures (A) – (C) are for MTBLS1033. (A) is the PLS-DA score plots between Control group and POP Disease group; (B) is the ROC-AUC result of k-folder validation (k=4) using 20% of the whole samples as testing data and 80% for training and feature selections; (C) The pathway enrichment analysis results from Mummichog.

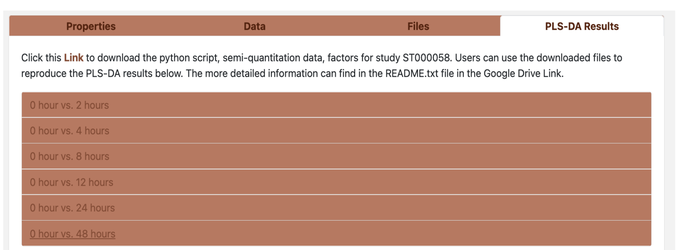
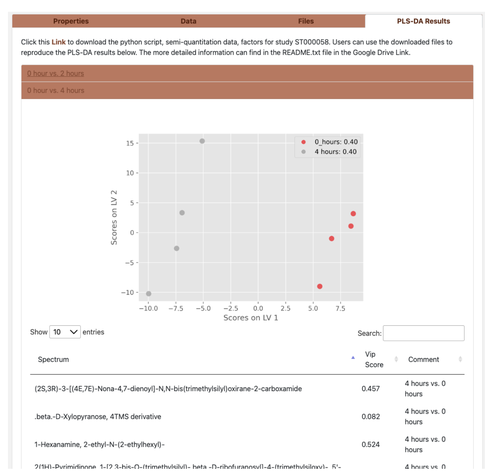



**Figure.4** ADAP-KDB design to display metabolic signatures.




## Conclusion

- A total of 12 studies from Metabolomics Workbench have been processed using this pipeline and the results will be uploaded to ADAP-KDB. Among these 12 studies, 2 are LC-MS studies and the rest are GC-MS studies.
- The link to use ADAP-BIG and ADAP-KDB are in the QR code.

## References

[1] Deng W et al. Metabolomics study of serum and urine samples reveals metabolic pathways and biomarkers associated with pelvic organ prolapse. J Chromatogr B Analyt Technol Biomed Life Sci. 2020 Jan.

[2] Li et al. Predicting Network Activity from High Throughput Metabolomics. PLoS Computational Biology 9.7 (2013): e1003123.

[3] Aleksandr Smirnov, Yunfei Liao, et al. ADAP-KDB: A Spectral Knowledgebase for Tracking and Prioritizing Unknown GC–MS Spectra in the NIH's Metabolomics Data Repository, Analytical Chemistry 2021 93 (36), 12213-12220.

ADAP-KDB

ADAP-BIG