Systematic Data Processing of LC-MS Untargeted Metabolomics of Common Variable Immunodeficiency Disease

Darshak Gadara(1), Kateřina Coufalíková(1), Juraj Bosak(2), David Smajs(2), Zdeněk Spáčil(1)

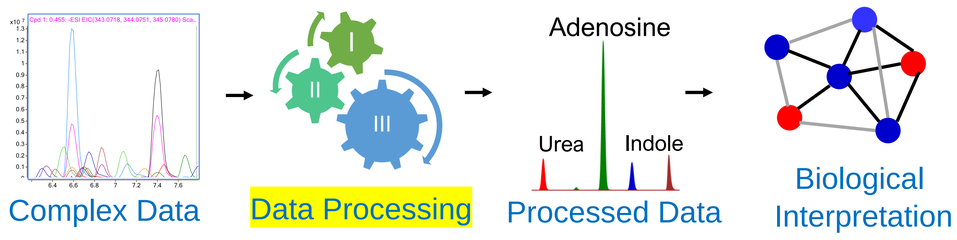
(1) Masaryk University, Faculty of Science, The RECETOX Centre, Brno, Czech republic

(2) Masaryk University, Faculty of Medicine, Department of Biology, Brno, Czech Republic

INTRODUCTION

Background

- Unlike genome and proteome, metabolome accounts for the genotype and phenotype[1]
- LC-MS generates complex dataset (thousands of peaks) because of high sensitivity and specificity[2]
- Therefore, data processing is considered as a major bottleneck to unravel the translational potential of metabolomics[2]



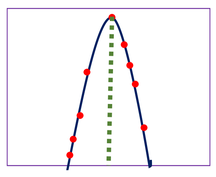
Aims

- The present study has proposed a systematic data processing workflow for the untargeted metabolomics
- Developed workflow tested on common variable immunodeficiency (CVID) case/control fecal samples .

METHOD

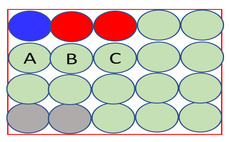
Data Acquisition

2 µL fecal extract injected on Thermo Orbitrap Fusion interfaced with Shimadzu UHPLC, Acquity CSH (100 x 2.1 mm, 1.7 µm)



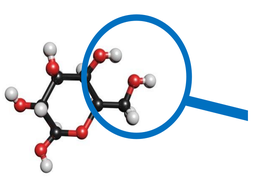
Peak Peaking

XCMS online was used. Data were exported as CSV file contains m/z, retention time, and peak area for each peak.



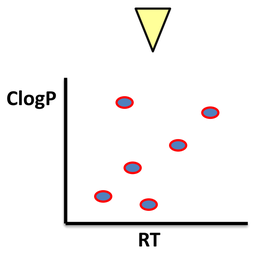
Blank Subtraction & QC Filtration

Blank and pooled QC samples were employed for the removal of background noise and quantitatively unreliable peaks



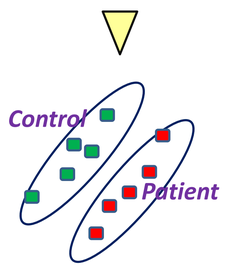
Database Search

Peaks were matched against HMDB and KEGG to retain only biological relevant metabolites (mass accuracy <3 ppm)



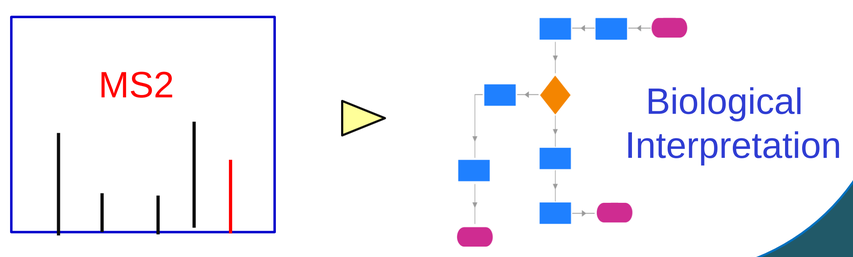
Chemistry Based Filtration

To filter peaks, The RPLC gradient was segmented based on elution order of the 36 standard compounds (Clog P range -6 to 11).



Statistical Analysis

Data was filtered retaining only features with fold change >1.5 or <0.67 and corresponding p-value of < 0.1



RESULTS

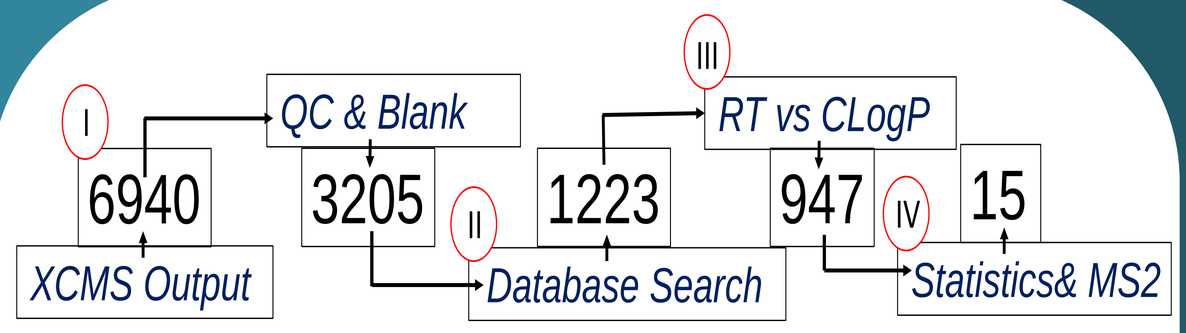


Fig.1. XCMS generated 6940 peaks with corresponding m/z, retention time and peak area. Peak list was submitted to data processing workflow describe in method section. Tentative structure of 15 statistically significant metabolites is confirmed based on the MS2 data

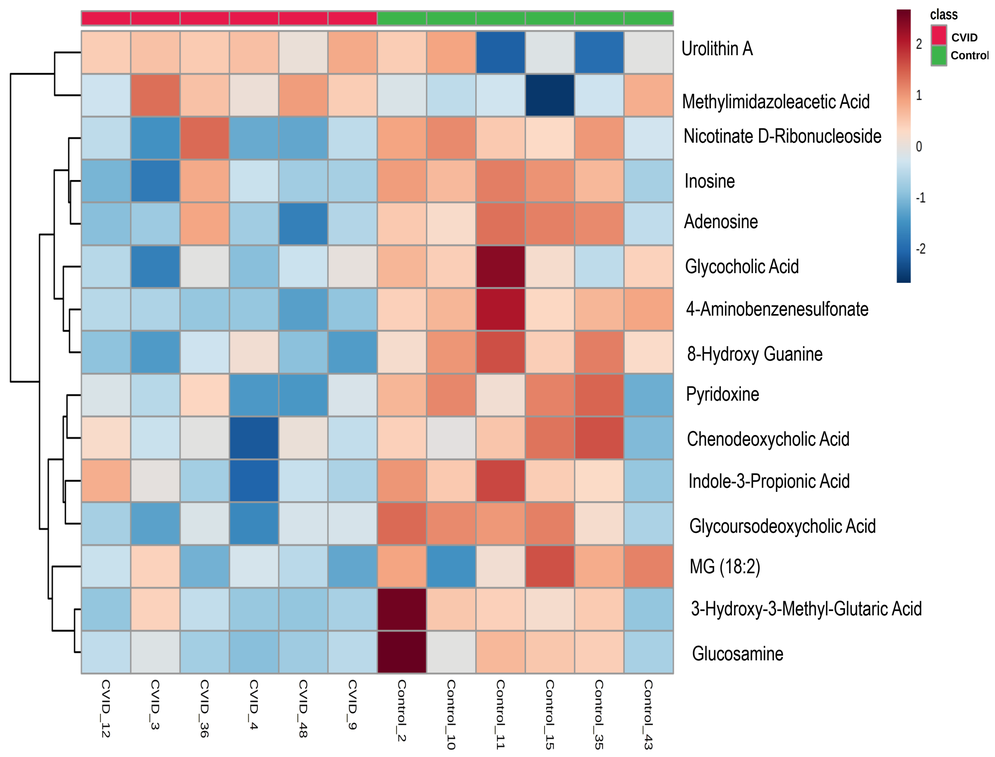


Fig.2. Heat map visualize the gradient changes of 15 differential features between CVID and healthy subject. Red color indicates a high intensity while blue color depicts low intensity. Each column represents individual biological sample

CONCLUSIONS AND FUTURE DIRECTIONS

- Untargeted metabolomics data processing pipeline has effectively removed the noise and produce quantitatively reliable metabolite peaks
- Pipeline tested on CVID case/control samples (6 each) and 15 significant metabolites found which represent the metabolic dysregulation in CVID
- In future, pipeline will be fine tuned to characterized membrane lipids to understand pathophysiology of Alzheimer disease

References

1. *Fiehn, O. Fiehn Plant Mol Biol 2002_review Link between Genotypes to Phenotypes.Pdf. 2002, 155–171.*
2. *Mahieu, N. G.; Patti, G. J. Systems-Level Annotation of a Metabolomics Data Set Reduces 25 000 Features to Fewer than 1000 Unique Metabolites. Anal. Chem. 2017, 89 (19), 10397–10406*

Acknowledgements

This work was supported by the Grant Agency of the Czech Republic (project No. 17-24592Y), the RECETOX research infrastructure (Ministry of Education, Youth, and Sports MEYS,LM2015051) and CETOCOEN PLUS (MEYS, CZ.02.1.01/0.0/0.0/15_003/0000469).