Complete Annotation of the Untargeted, LC/MS Based Metabolomic Analysis of Escherichia coli

Nathaniel G. Mahieu, Amanda Chen, Xiaojing Huang, Gary J. Patti

Department of Chemistry, Washington University, St. Louis, MO 63110, United States

Introduction

LC/MS based untargeted metabolomics is driven by the of thousands of signals detected in single biological samples. The large number of signals commonly detected has led to widely varying estimates of the number of metabolites being assayed in untargeted metabolomic experiments. Understanding the composition of these datasets is critical to making informed decisions during experimental design and data interpretation.

In this work we thoroughly annotate the peaks detected in an LC/MS based untargeted metabolomic analysis of E. coli metabolic extract. Specifically we classify peaks as biological or artifactual, isotopes, and adducts. MS/MS analysis of peaks is used to empirically annotate fragment peaks. Further an estimate of known and unknown compounds is made by matching fragmentation spectra to various metabolite databases.

Methods

Credentialed E. coli (Strain K12 MG1655) standard samples were prepared by batch culture in M9 minimal media to OD600 1.0 in both uniformly enriched and natural abundance glucose. Metabolism was quenched with -70C, 60% Methanol and cells were harvested by centrifugation. Credentialed samples were extracted. Experiments were performed on Thermo Q-Exactive Plus with Dionex RSLCnano using the Phenomenex Luna NH2 150mmx1mm 3um column. Peak finding was performed using XCMS. Biological peaks were annotated using our previously published credentialing algorithm. MS/MS was performed and peaks were empirically classified as fragments, isotopes, adducts, knowns and unknowns.

Results

Analysis of E. coli credentialing extract produced 20,609 features while the corresponding natural abundance sample produced 14,061 features. Credentialing of the standard extract annotated 1,511 peaks as biological in origin after removal of isotopes. MS/MS fragmentation data is collected on these 1,511 peaks in order to empirically annotate fragments and knowns.

Conclusions

Utilization of credentialing allows the complete annotation of biological peaks, previously impractical due to the large number of signals. This approach provides information valuable to planning and interpretation of metabolomic data.