ABRF 2017 Annual Meeting

A FORUM FOR ADVANCING TODAY'S CORE TECHNOLOGIES TO ENABLE TOMORROW'S INNOVATIONS

ABRF-MRG2016 Metabolomics Research Group Data Analysis Study

Amrita Cheema PhD
Professor of Oncology
Co-Director: Waters Center of Innovation
Georgetown University Medical Center
Washington DC

March 25-28, 2017

Town and Country Resort & Convention Center
San Diego, California

Metabolomics Research Group

Current Members:

Chris Turck: Max Planck Institute (Chair)

Amrita K Cheema: Georgetown University

Christopher Colangelo: Primary Ion (ED Liaison)

Maryam Goudarzi: Cleveland Clinic; Lerner Research Institute

Tytus Mak: NIST

Andrew Patterson: Pennsylvania State University

Reza Salek: European Bioinformatics Institute

Metabolomics is Concerned with the Simultaneous, Comprehensive Measurements of Small Molecules

Metabolomics is the comparative analysis of endogenous metabolites found in biological samples:

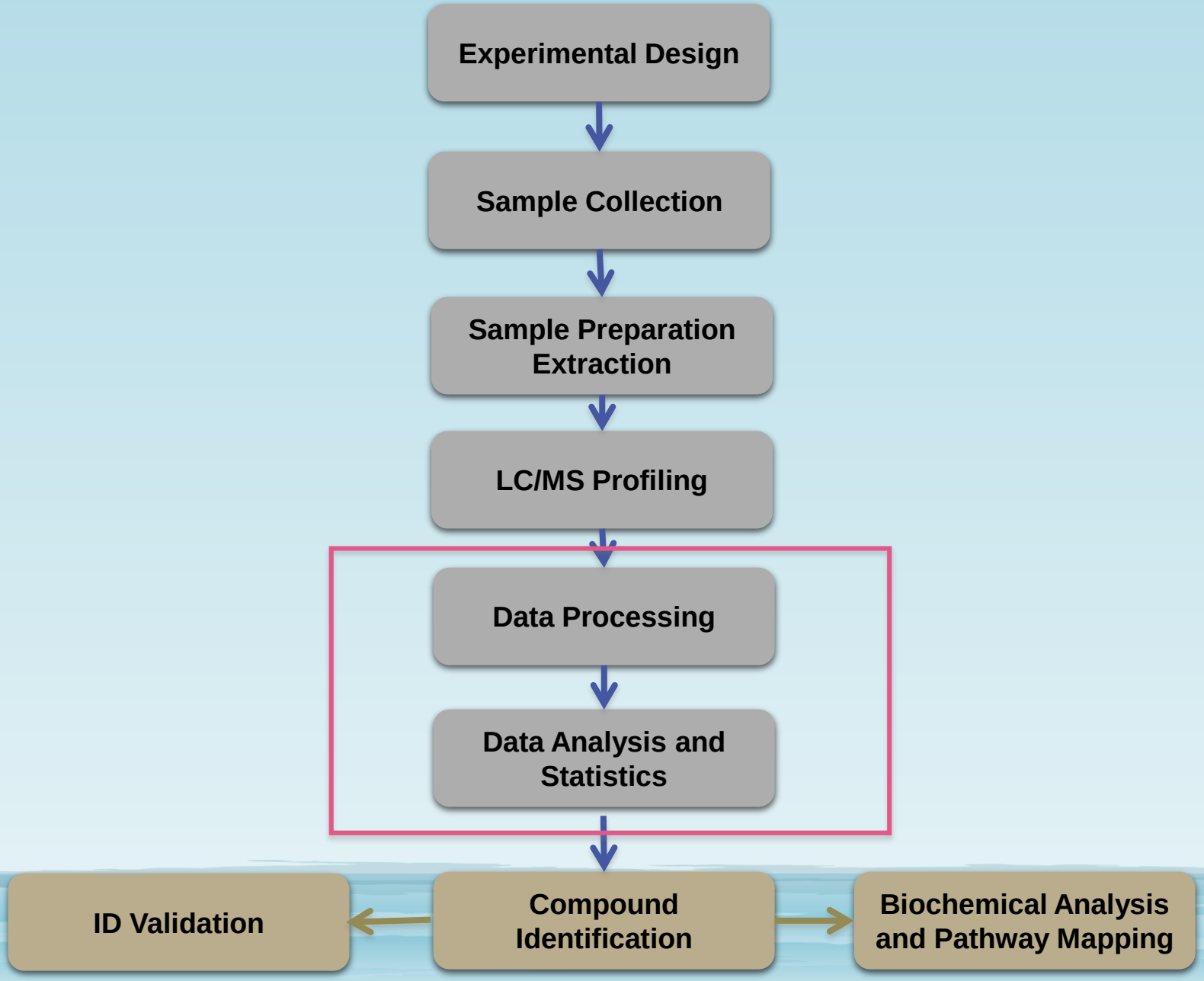
- ***Compare*** two or more biological groups
- Find and identify potential biomarkers
- Look for biomarkers of toxicology
- Understand biological pathways
- Discover new metabolites

Metabolites are the by-products of metabolism

- Range of physico-chemical properties
- Classes: Amino acids, Sugars, organic acids, fatty acids, lipids…

What are the chemical differences that result in the observable difference

Untargeted Metabolomics Workflow



Bottlenecks in the Metabolomics Field

- Lack of standardization for the entire metabolomics workflow
- Exhaustive and reliable data analysis and interpretation
- Identification of metabolites
- Peak picking/identification
- Time consuming data analysis and identification process (informatics process)
- Spectral databases for metabolite identification are not free
- Availability of MS instruments
- Cost of MS-based metabolomic analysis
- Accessibility of cores/labs to metabolomics experts particularly for data analysis

MRG Inter-Laboratory Metabolomics Study 2016

- Design a study that resembles an untargeted metabolomics profiling experiment comparing biological changes under different conditions (exposure to Sham vs 5 Gy)

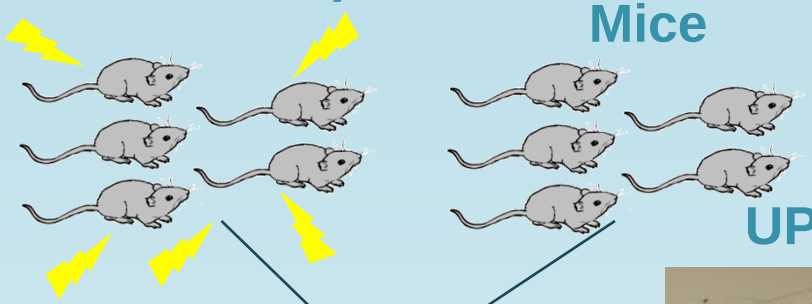
- Participants will identify statistically significant differences between groups A & B of samples in order to compare findings with varied methodologies

- Goal:
  Examine challenges, overlap and variability in results between approaches to LC-MS metabolomics data analysis

Study Design

Irradiated Mice (5 Gy Gamma Radiation)

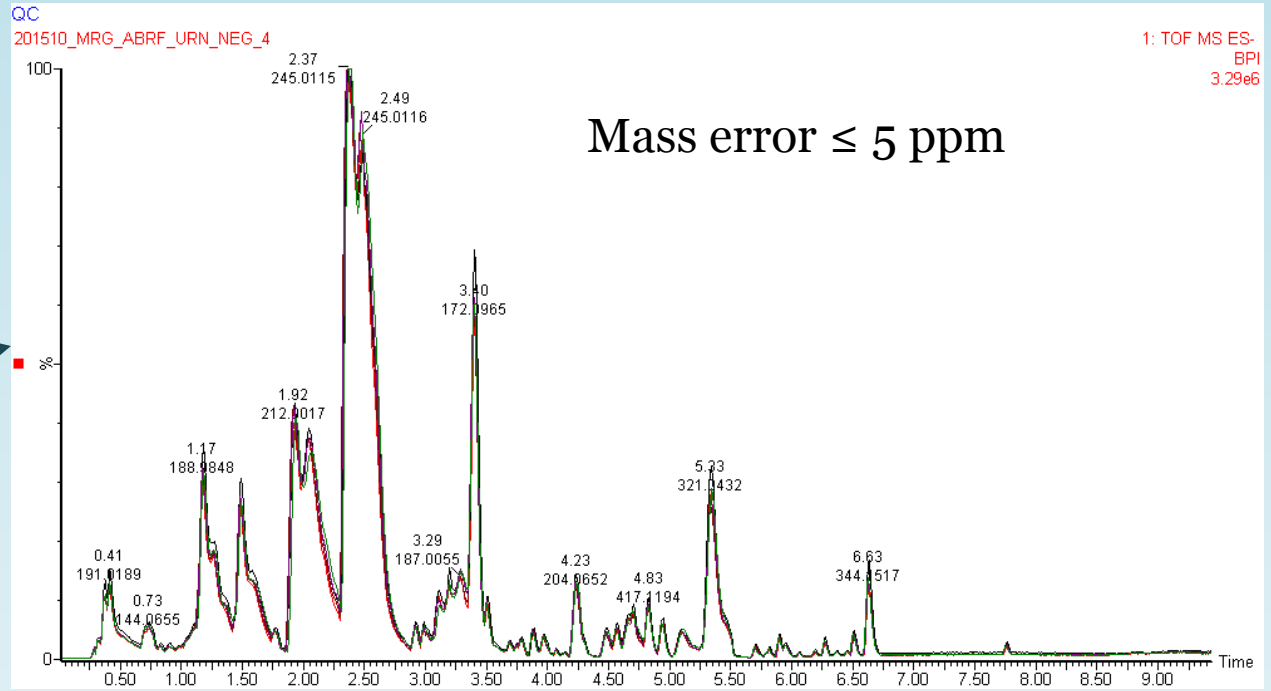
Control Mice



Urine Samples

UPLC-Q-TOF

QC Overlay



MRG 2016 Inter-Laboratory Study

Goal: Design a study to assess the impact of various bioinformatics and statistical approaches on metabolomics data analysis results

Participants were asked to:

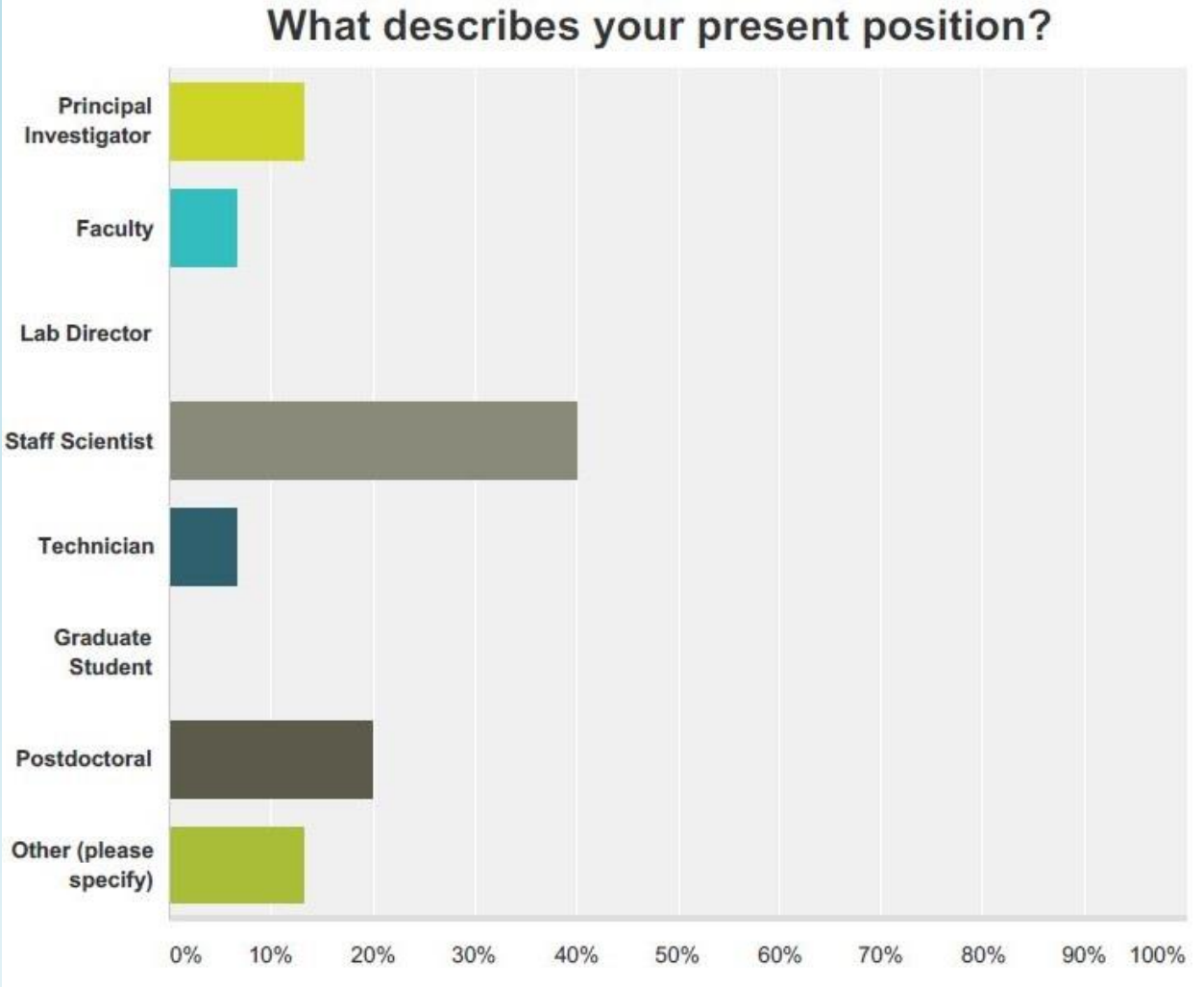
- **Pre-process raw data using a pre-processing software of their choice and provide a data matrix consisting of m/z, retention time, and ion intensity**
- **Post-process the data using statistical tools and determine the top 50 spectral features that were significantly dysregulated**
- **Assign putative identification to these urinary spectral features using various online databases**

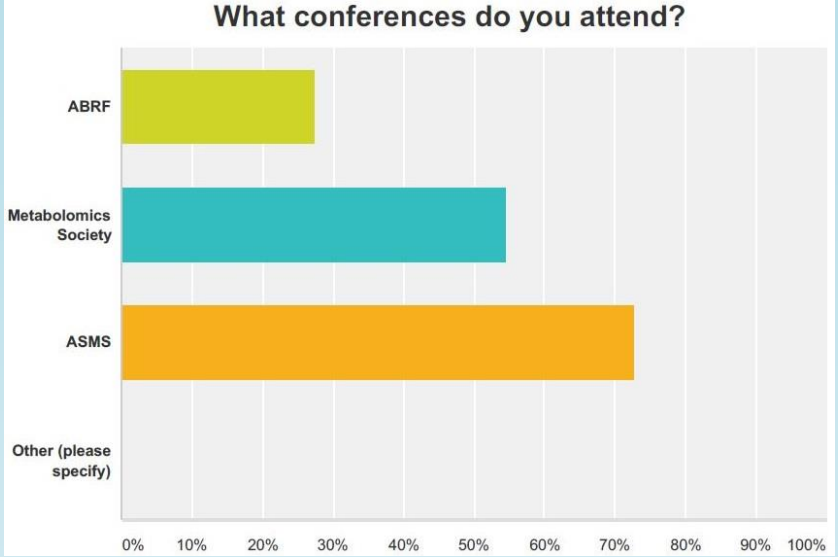
Expected Outcomes

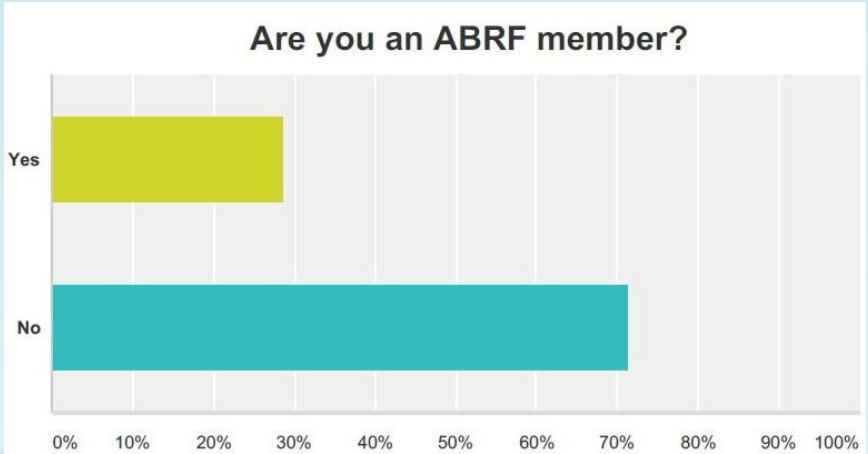
The primary objective of this study was to examine reproducibility and optimal data analysis strategies for metabolomics studies:

- Compare the relative quantitative metabolite differences across two sample types reported by participants
- Examine effects of different computational techniques on the determination of significantly altered metabolites in the two groups.
- Assess the level of confidence and consistency in the results obtained from unique computational and chemometric approaches.
- Compare ability of software to determine differences across samples or help analyze data from metabolomics experiments
- Compare databases used for assigning metabolite ID

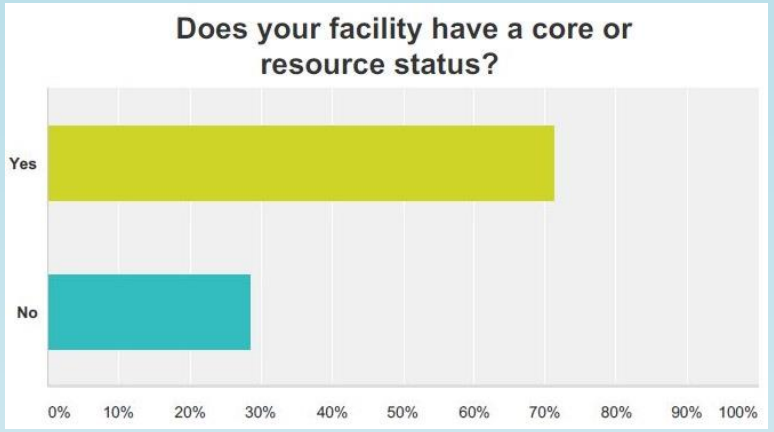
Participants

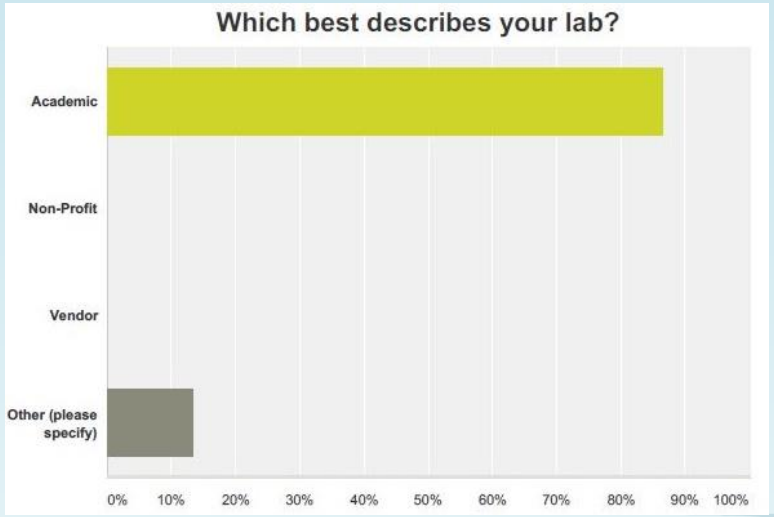


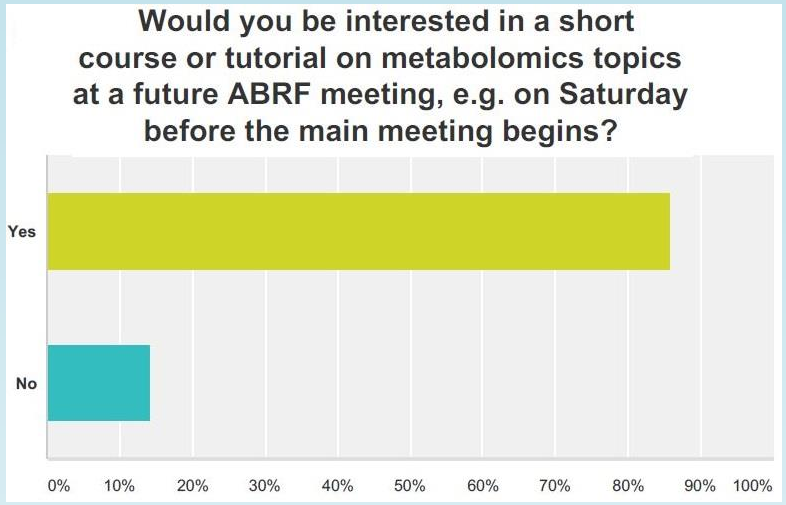


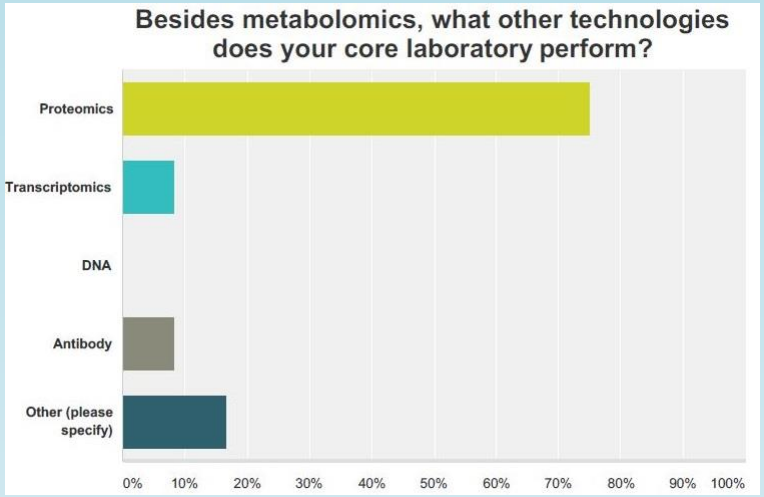


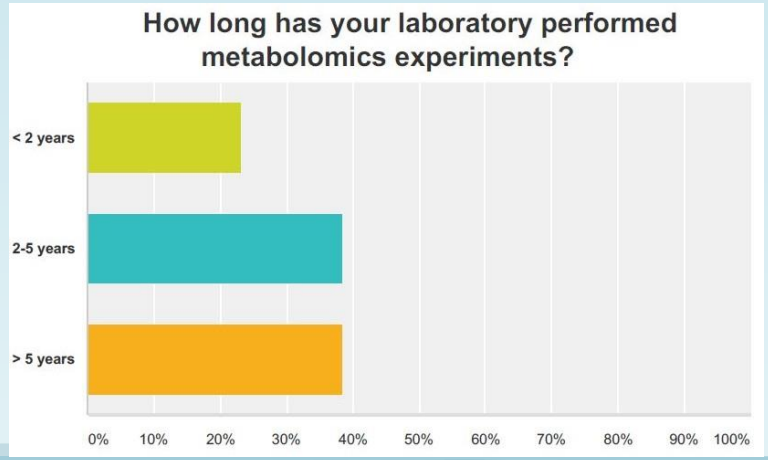
Participant Resources



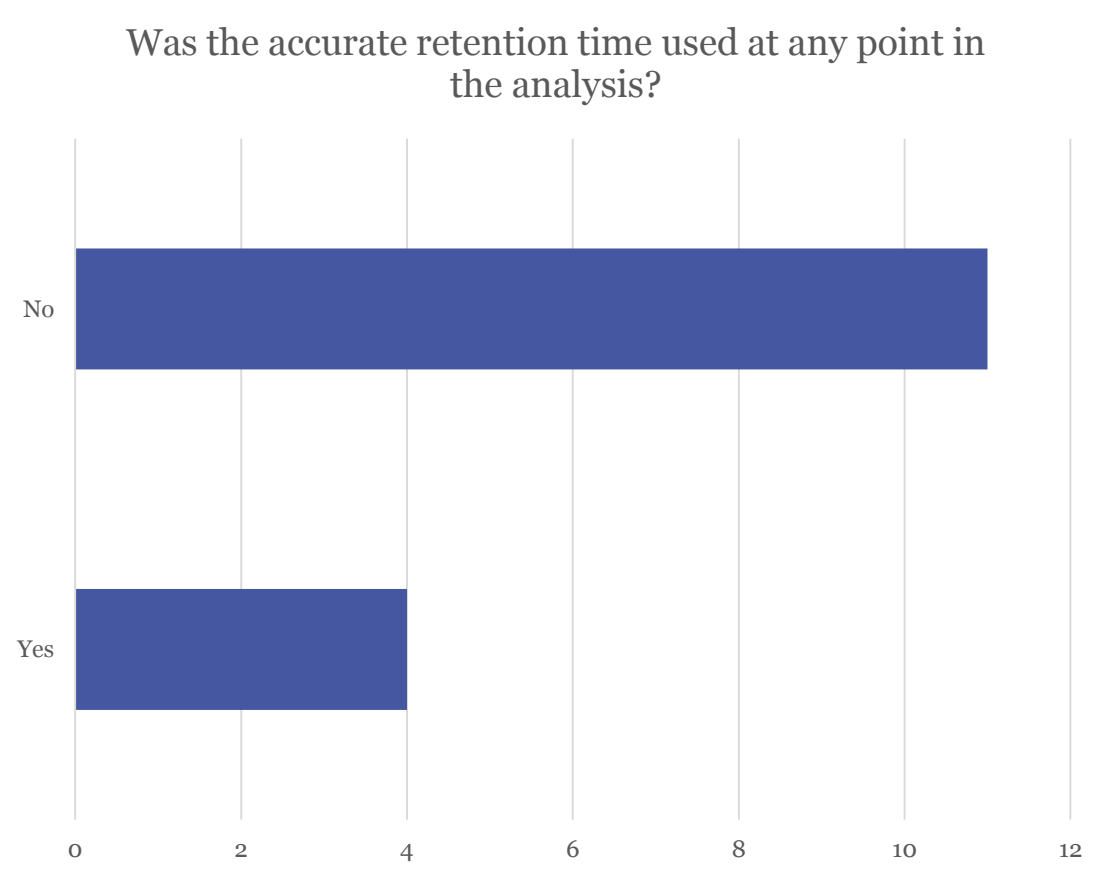


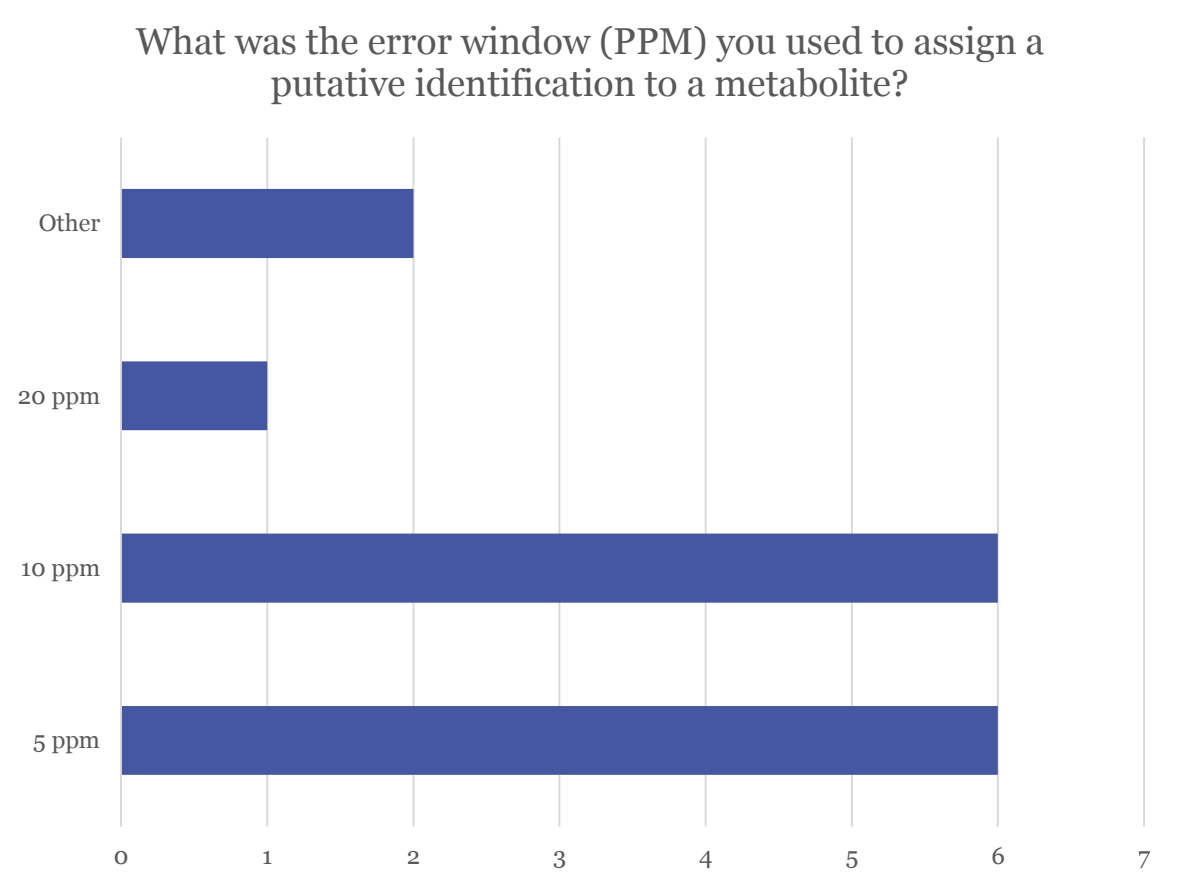




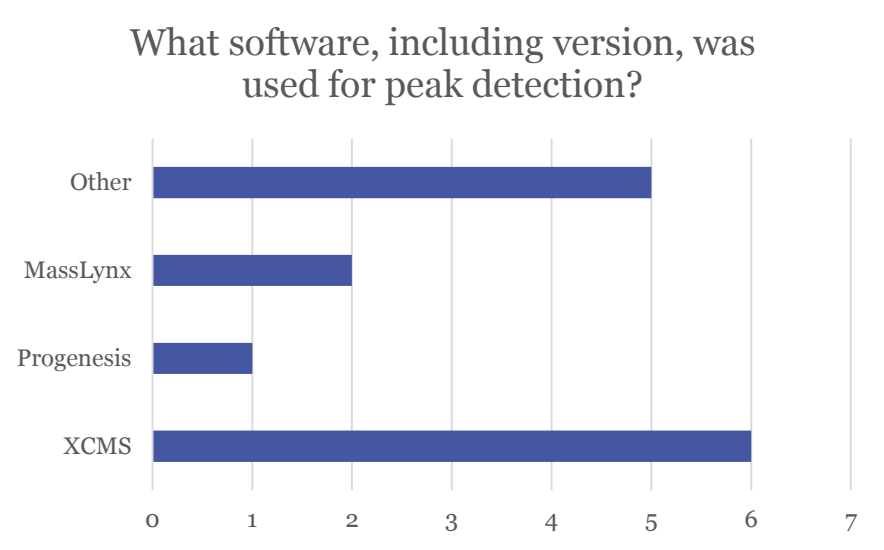


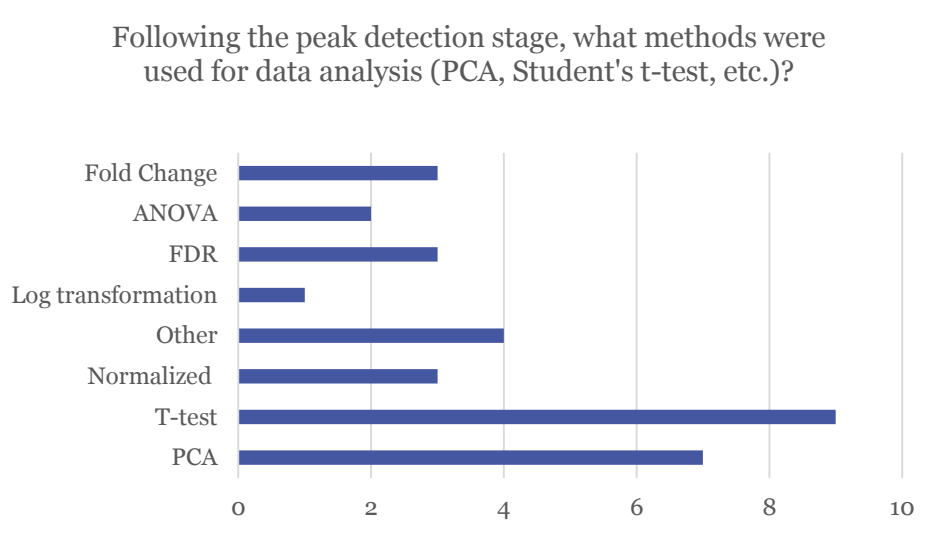
Analysis Differences Continued..

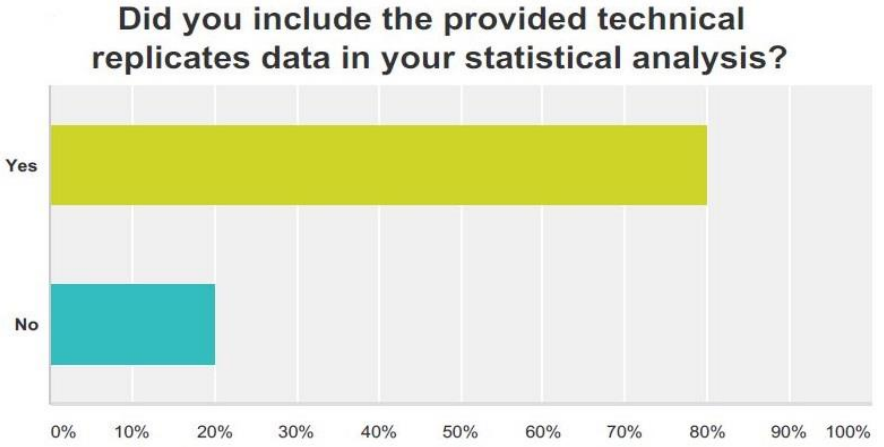


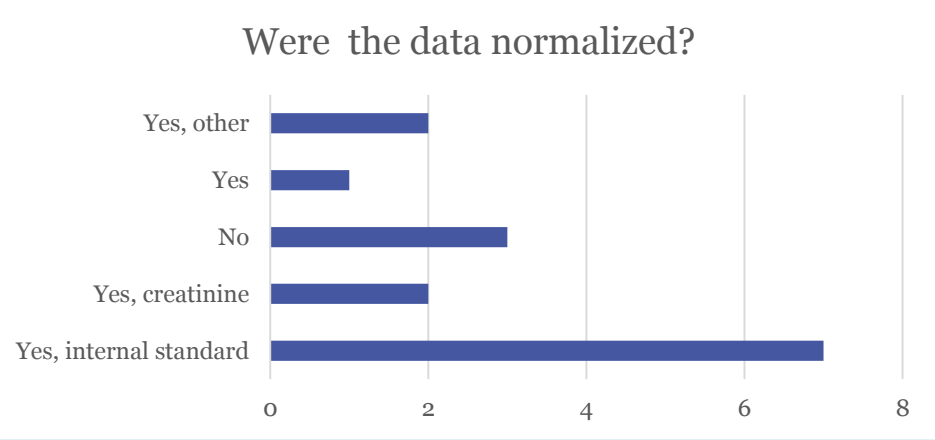


Analysis Differences



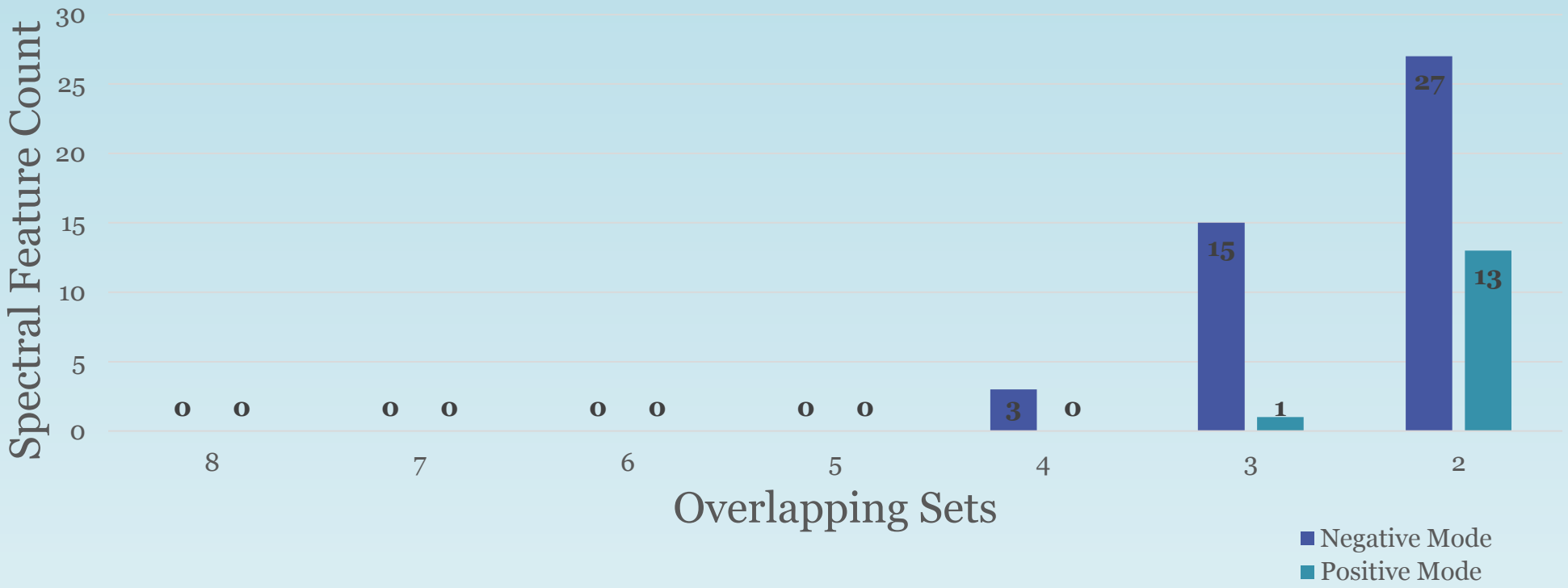






Results

Degree of Overlap for Top 50 Spectral Features



- **Parameters for identifying the same spectral feature across multiple data sets:**
  - **50 ppm**
  - **+/- 20% of the retention time**
  - **Candidate spectral feature must match 1 or more features in a group to be considered a member**

Less than 10% overlap in pre-processed output files that were subsequently used for statistical analysis

Conclusions:

- Data pre-processing has a significant impact on resultant data
- This is particularly concerning for untargeted metabolomics
- Standardization of peak picking methods is urgently needed

Next Steps

- Evaluate impact of pre-processing:
  - ask participants for complete peak list and the XCMS parameters used for preprocessing.
  - ask participants to repeat their analysis using MRG XCMS files.

Acknowledgements

- Participants
- ABRF

CCSG support to the Metabolomics Shared Resource at Georgetown University

NIH and NSBRI grant support