NCATS
COLLABORATE. INNOVATE. ACCELERATE.

National Center for Advancing Translational Sciences

Reproducibility of metabolomic profiles generated using the TruQuant platform in a multi-lab "round robin" study design

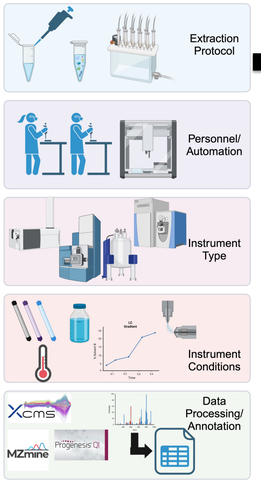
Haley Chatelaine[1], Chris Beecher[2], Ewy Mathé[1]
[1]Informatics Core, Division of Preclinical Innovation, National Center for Advancing Translational Sciences, Rockville, MD, [2]IROA Technologies, Chapel Hill, NC

Goals

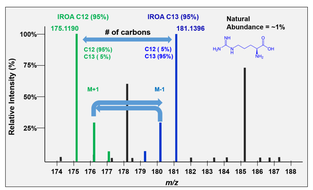
Evaluate the reproducibility of the IROA TruQuant metabolomic profiling platform across multiple labs
Differentiate heterogenous sample mixtures using metabolomic profiles generated using the TruQuant platform

Why Reproducibility?

Sources of Variability in Metabolomics Studies



Inter-lab variability in:
- Real features vs. artifacts
- Metabolite relative abundances
- Metabolite annotations



IROA TruQuant

95/5 C13/C12 labels: spectral pattern only in *biologically derived* features →
- retain real features
- compare against internal standards for relative abundances
- use mass spectral patterns for consistent annotations

Agnostic to LC-MS platform or extraction protocol

Hypothesis

The IROA TruQuant platform will enable reproducible characterization of metabolically distinct mixtures, despite variability in lab-specific analytical strategies.

Round Robin Overview

1 **Sample Preparation** – each sample is prepared in duplicate N=32 total using TruQuant preparation and standards

All samples are run in duplicate

Pure Sample — Mixed Sample

	Concentration *	Whole sample			Mix of 2 samples			Mix of 3 samples
		A	B	C	A + B	A + C	B + C	A + B + C
Low Conc	1/3				x	x	x	x
	1/2	x	x	x	x	x	x	
	1	x	x	x				
High Conc	1 1/2	x	x	x				

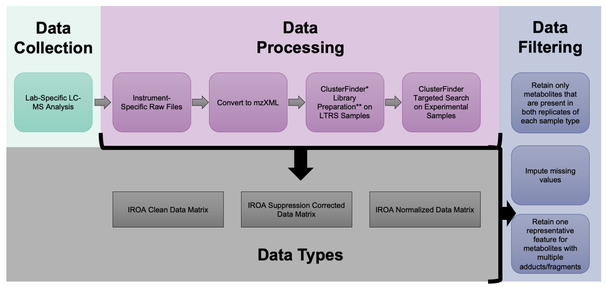
*fractions indicate relative concentrations of A, B, and C each

A B C

2 **Metabolomic Profiling** - each lab runs their own method and sends data centrally

3 **Data Processing and analysis**

Data Anonymized and processed with Cluster Finder → Results sent for quality control, reproducibility, and prediction analyses



Input Data

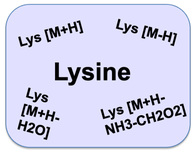
Lab Number	Total Named, High-Quality Features (Original)
Lab 1	347 (1193)
Lab 2	381 (1867)
Lab 3	194 (781)
Lab 4	237 (884)
Lab 5	221 (737)
Lab 6	163 (525)
Lab 7	237 (784)
Lab 8	232 (767)

Analyses

Intra-Lab Assessment	Evaluate expected vs. observed differences in concentration as well as basic reproducibility metrics
Inter-Lab Assessment	• Compare metabolite annotations, relative abundances • Assess and compare reproducibility using raw, suppression corrected, and normalized CF results
Predict/Cluster Groups	Use metabolomic profiles to determine sample group membership using supervised (prediction) and unsupervised (clustering) approaches (PCA, RF, NMF)

Feature Selection

The Puzzle



Different adducts are the most abundant feature for a given metabolite in each lab, making metabolite-to-metabolite comparisons between labs difficult.

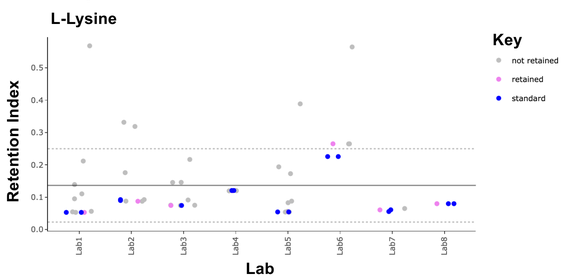
Need to select one representative feature for each metabolite

Selection Criteria

A) Retention time (RT) aligns with standard, if available
B) Highest intensity and retention index* within one standard deviation of the mean retention index across all labs for features annotated as the metabolite.

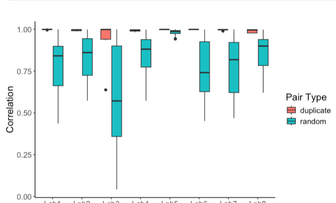
*Retention Index = RT of feature / RT of last eluting feature

Example Metabolite



Features (all points) annotated as L-Lysine in each lab. Pink are retained as representative features, selected by comparing to an authentic standard run using the same LC method (blue); grey are not retained.

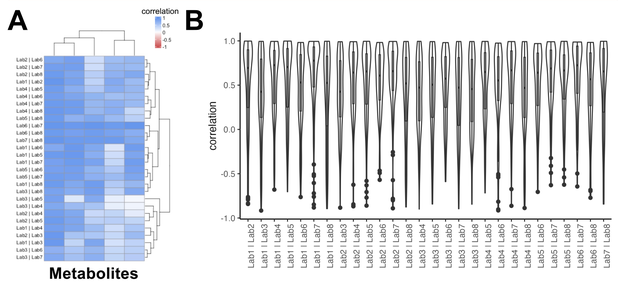
Intra-Lab Quality Control



Correlations of Duplicates

Distribution of correlations of metabolite intensities between duplicate samples (red) and random sample pairs (blue).

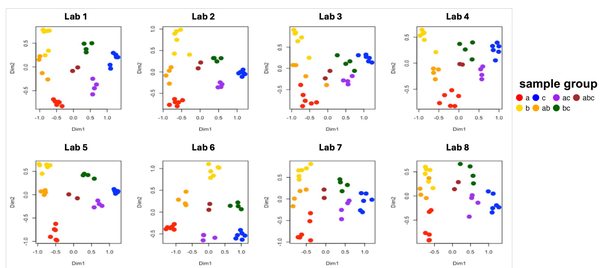
Inter-Lab Reproducibility



Metabolite Correlations by Lab

If a metabolite was detected in a pair of labs, the correlation of normalized intensity values for the representative feature of that metabolite was calculated between the lab pair. A) Correlation values for metabolites (columns) with representative features retained in all 8 labs based on authentic standards. B) Distribution of correlation values for all metabolites detected in lab pairs.

Sample Group Clustering



Unsupervised Clustering of Samples

Unsupervised random forest clustering of normalized metabolite data from all labs reveals anticipated clustering by sample group with pure samples clustering farthest from each other and mixtures clustering toward the middle.

Conclusions & Future Directions

The TruQuant platform enables reproducible metabolomic profiling and characterization of heterogeneous sample mixtures using unsupervised models.

Acknowledgments

Collaborators: Cristina Balcells Nadal, Stephen Barnes, Djawed Bennouna, Justin Cross, Clay Davis, Felice de Jong, Maureen Kachman, Tim Garrett, Hector Keun, Mahmud Iqbal, Wenqian Li, Philip Lorenzi, Robert Powers, Alexander Raskind, Michelle Saoi, Tracey Schock, Landon Wilson
This research was supported in part by the Intramural/Extramural program of the NCATS, NIH (ZICTR000410-03).