Learning outcomes for

Course in Computational Metabolomics for Clinical Research

(**Feb 16 – 20, 2026 from 2pm to 5pm EST**)

Instructor: Dinesh Barupal, PhD
Icahn School of Medicine at Mount Sinai, New York, USA

Day	Topic	Description
1	Introduction	Human metabolism Metabolomics – relevance in clinical research
		RStudio and the server organization
	Peak detection and annotation	Formula, isotopes, ions, structures
		High Resolution Mass Spectrometry basics, TIC, EIC, peak-detection
		IDSL-IPA – automated peak detection, targeted peak analysis
		Fully automated untargeted data processing by a workflow of IDSL-IPA, IDSL-UFA, and IDSL-CSA software.
		IDSL-UFA – automated molecular formula annotation using databases (Metabolon and blood exposome)
		DDA, DIA and CSA spectra, spectra similarity and libraries IDSL-CSA – automated extraction annotation of DDA and CSA spectra using mass spectral libraries.
2	Data filtering and preparing	Merge the results for peak detection, formula annotation and spectra annotation and apply a detection frequency threshold, diagnostic PCAs
		Missing value handling, data normalization, scaling, transformation, batch effects
		data normalization for studies with many batches.
3	Statistics and pathway analysis	Study design and statistical tests, Linear regression (continuous outcomes), Logistic regression (categorical outcomes), Conditional logistic model (nested case control designs), practice these tests for several human datasets. Pathway Analysis and Metabolic Network Mapping of the regression results
		Multi-variate regression models for confounding/covariate variables
4	Metabolic Bioinformatics	Chemical Set Enrichment Analysis (ChemRICH), Gene Ontology Analysis (IDSL-GOA), Co-regulatory Set Enrichment Analysis, Inter-Chemical Correlations
5	Machine Learning and Databases	Introduction to machine learning, method types, errors, overfitting, data splitting, linear and non-linear models, feature-engineering
		Supervised machine learning using random forest to predict continuous outcomes (regression) and binary outcomes (categorical) How to fit an overfitted, training and test (split) and selected feature model.
		Un-supervised machine learning, Clustering, Dimensionality reduction
		Metabolite clusters (data driven, structure (function-driven), Sample clusters (sub-types), HCL and UMAP-HDBSCAN for cluster detection in metabolites and samples
		Biochemical databases Chemical and biomedical text mining Large Language models
5	Self-exercise	Recap, additional functions and self-exercises