bioinformatics.ca

Canadian Bioinformatics Workshops

www.bioinformatics.ca

Module 5
MetaboAnalyst – Lab

bioinformatics.ca

Jeff Xia and David Wishart
Informatics and Statistics for Metabolomics
June 15-16, 2015

UNIVERSITY OF ALBERTA
QUAECUMQUE VERA

Schedule

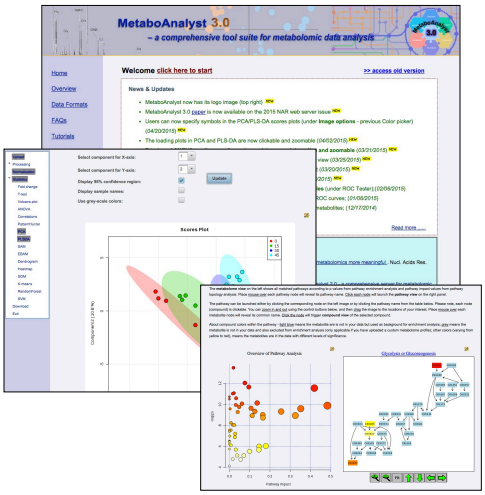
CANADIAN BIOINFORMATICS WORKSHOPS
Informatics and Statistics for Metabolomics

	Day 1	Day 2
	Monday June 15	**Tuesday June 16**
8:00	*Registration & Breakfast*	*Breakfast*
8:30	Welcome (Michelle Brazas) Module 1: Introduction to Metabolomics (David Wishart)	Module 4: Backgrounder in Statistics
10:30	*Coffee Break*	*Coffee Break*
11:00	Module 2: Metabolite Identification	Module 5: MetaboAnalyst
12:30	*Lunch - on your own*	*Lunch - on your own*
1:30	Module 2 Lab: Compound ID & Quantification	Module 5 Lab: Metabolomic Data Analysis using MetaboAnalyst 3.0
3:00	*Coffee Break*	*Coffee Break*
3:30	Module 3 Lecture: Databases for Chemical, Spectral and Biological Data	Module 5 Lab: Continued
5:00	*Dinner - on your own*	Survey & Closing Remarks
8:00	Compound ID until 8pm	

Learning Objectives

- Get Familiar with MetaboAnalyst on your own (answer some questions, explore)
- Analyze NMR-based metabolomic data (from Lab 2) using MetaboAnalyst
- Analyze GC-MS-based metabolomic data (from Lab 2) using MetaboAnalyst
- Analyze LC-MS/MS-based metabolomic data (from Lab 2) using MetaboAnalyst

MetaboAnalyst



http://www.metaboanalyst.ca

- Web server designed to handle large sets of LC-MS, GC-MS or NMR-based metabolomic data
- Supports both univariate and multivariate data processing, including t-tests, ANOVA, PCA, PLS-DA
- Identifies significantly altered metabolites, produces colorful plots, provides detailed explanations & summaries
- Links sig. metabolites to pathways via KEGG & SMPDB

Metabolomics Data Workflow

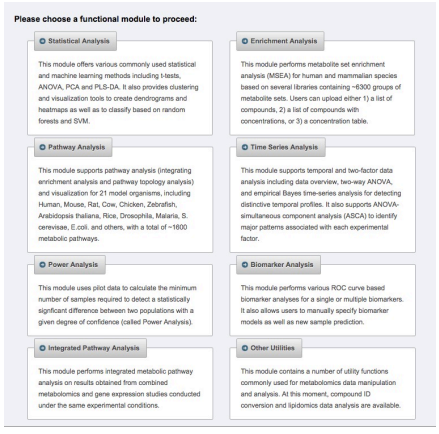
Chemometric Methods

- Data Integrity Check
- Spectral alignment or binning
- Data normalization
- Data QC/outlier removal
- Data reduction & analysis
- Compound ID

Targeted Methods

- Data Integrity Check
- Compound ID and quantification
- Data normalization
- Data QC/outlier removal
- Data reduction & analysis

MetaboAnalyst Modules



Data pre-processing → Data normalization → Data analysis → Data interpretation

8

Suggested Analysis Ideas

- Look through the data for outliers, errors or mis-annotated compounds (QC analysis)
- Play around with different scaling and normalization parameters, see what they do and why (for each data set)
- Try performing PCA, look through all the different tabs to see what is revealed

Suggested Analysis Ideas

- Try performing PLS-DA, look through all the different tabs to see what is revealed
- What does the VIP data tell you at a biological level?
- Try the permutation test, how significant are the clusters?
- Try different clustering methods, look to see what clusters form and why

Suggested Analysis Ideas

- From your PLS-DA data try to develop some hypotheses regarding the causes or consequences of the metabolic changes
- Use the pathway analysis modules to identify key pathways
- Look through the databases (HMDB, SMPDB, PubMed, others) to learn more about the significant metabolites or significant pathways you've found

Suggested Analysis Ideas

- Try the MSEA analysis and see if that helps with understanding the biology
- Try to develop a set of robust biomarkers for these different conditions
- Look at different methods (SVM, PLS-DA, random forest), try things manually
- Are the biomarkers the same (or different) from the most significant metabolites from your PLS-DA analysis?

For Those Who Are Interested…

Some Challenging Questions That Require Using MetaboAnalyst

Questions on ANOVA (Bovine Feeding Data)

- Q: Which compounds show significant difference among all the neighboring groups (0-15, 15-30, and 30-45)?
- Q: For *Uracil*, are groups 15, 30, 45 significantly different from each other?

ANOVA Correlation (Bovine Feeding Data)

- Q: In untargeted metabolomics using NMR, researchers often look for region(s) in their spectra showing large changes in their correlation patterns under different conditions. Can you do that in MetaboAnalyst?
- Hint: check the available parameters in Correlation analysis

Pattern Matching (Bovine Feeding Data)

- Q: Can you identify compounds that decrease in the first three groups (0%, 15%, 30%) but increase in the last group (45%)?

PCA Analysis (Bovine Feeding Data)

Q: Identify compounds that contribute most to the separation between group 15% and 45%

PLS-DA Model Validation

- Q: What does p < 0.01 mean?
- Q: How many permutations need to be performed if you want to claim p value < 0.0001?

MSEA/QEA Matched Metabolite Set (Cachexia Group)

- Q: Are these metabolites increased or decreased in the cachexia group?

Pathway Visualization (Cachexia Group)

- Q: Which pathway do you think is likely to be affected the most? Why?

Biomarker Analysis

- Compare the biomarkers automatically selected via the multivariate ROC tool and those top ranked metabolites generated via the univariate ROC tool. How much do they overlap? Can you manually create a biomarker model that performs better?
  - Hint: Try ROC-tester

Power Analysis

- Based on the power analysis curve, can you identify the "optimal" sample size for detecting an effect?