Problem-based Learning for Bioinformatics

Tao Huan, PhD

Mar 18 and Mar 23, 2020

Biological question

Metabolomics is the study of the entire set of small molecules in biofluids, cells, and tissues. One of the major bioloigcal applications of metabolomics is to identify dysregulated metabolites that correlate with disease conditions and can potentially be used for disease prognostics or diagnostics. In some respects, metabolomics has already been remarkably successful in discovering biomarkers that translate to the clinic. Lu et al. recently developed a MS-based urine metabolomic diagnostic assay that can distinguish between patients with adenom/atous polyps and healthy controls with good sensitivity and specificity. This metabolomic assay is the first and only urine based test and has been tested in a large-scale clinical trials with nearly 1000 Canadian patients. Since metabolites are the downstream products of gene and protein activities, investiagting metabolic changes allows for a better mechanistic understanding of a given biological question. For example, cancer cell metabolism is known to be perturbed as the result of gene and enzyme mutation in cancer cells. Recent metabolomics studies revealed that sarcosine significantly increased during prostate cancer cell lines relative to benign prostate epithelial cells. Guided by the metabolomics discovery, further mechanistic study through knockdown of sarcosine dehydrogenase, the enzyme that degrades sarcosine, induced benign prostate epithelial cells into an invasive phenotype. This study revealed sarcosine as a potentially critical metabolic intermediate of cancer cell invasion and aggressiveness.

Among women, breast cancer (BC) is the most common type of cancer and also the 2nd leading cause of death worldwide1. BC can be divided into several major subtypes based on conventional immunohistochemistry detection of hormone receptors, including human estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth receptor-2 (HER2). As a highly heterogeneous disease, patients with BC have various morphological spectrum, clinical presentation, and prognostic outcomes. Among the various subtypes, triple negative breast cancer (TNBC) uniquely lacks expression of all three hormone receptors and accounts for 15–20% of the BC cases. TNBC has attracted more attention compared with other BC subtypes as it is typically associated with high aggression, poor prognosis and a high risk of disease relapse within 5 years following diagnosis. Women with TNBC have a high frequency of metastasis to the lung, liver and brain, and survival is generally poor. It is thus of great demand to study the molecular basis of TNBC in order to guide the development of promising drugs and therapies for treatment

In this study, metabolomic data were generated from the serum sample of 5 TNBC patients and 5 heathy controls. Please process these raw LC-MS data and extract statistically significant metabolites that can differentiate TNBC from healthy controls.

Learning objectives

- Familiar with the instrumentations for collecting metabolomics data and understand their inherited pros and cons.
- Know how to use the well-established metabolomics software for processing metabolomic raw data
- Understand the key steps in the processing of metabolomic raw data
- Familiar with how to perform univariate and multivariate statistical analysis in metabolomics.

Guiding discussion questions:

1. What are the advantages and disadvantages of NMR and MS-based metabolomics?
2. Describe full-scan, data-dependent, and data-independent MS modes for metabolomics data collection.
3. What is the general workflow for metabolomics data processing?
4. What are the key steps in extracting metabolic features from raw LC-MS data?
5. List ten statistical tools or algorithms for metabolomics statistical analysis.
6. How to perform metabolite identification?
7. List three commonly used metabolic spectral libraries?