BST281: Genomic Data Manipulation, Spring 2019

Wednesday 13: Metabolomics

**Metabolomics**

* Seeks to identify and quantify small molecules produced via biological processes in a biosample
  + “Small” typically defined as <1 kda (1 kilodalton)
    - E.g. amino acids, nucleotides, sugars, small lipids, etc.
  + Only loosely connected to the central dogma (sequence analysis can’t help us for once)
* Two major philosophies
  + Targeted - identify and quantify a predefined subset of metabolites with very high accuracy
    - Usually limited to 10s of metabolites
  + Untargeted - identify as many unique metabolites as possible based on clustering of their physical and chemical properties, then worry about identification later
    - Many metabolites will not be identifiable (some will be non-biological)

**Technologies**

* Nuclear magnetic resonance (NMR) spectroscopy
  + Relies on the tendency of certain isotopes (e.g. 1H) to resonate at characteristic frequencies according to their environment (e.g. the metabolite they are present in)
  + Provides accurate quantification with minimal sample prep and doesn’t perturb the sample
  + Low sensitivity (# of molecules, min. abundance) and machines are very expensive
* Mass spectrometry (MS)
  + Relies on the tendency of charged particles to move in magnetic fields according to their mass (*m*), charge (*z*), and, velocity (fixing velocity reveals *m*/*z*)
  + A second round of MS (a.k.a. tandem MS = MS/MS = MS2) helps to identify parent molecules based on the masses of their breakdown products
  + Very sensitive and flexible; can differentiate 1,000s of metabolites
  + Provides relative quantification only; sample destroyed; complex preps
* Liquid chromatography (LC) separates complex mixtures according to physicochemical properties
  + Limits the number of distinct species entering an instrument at a given time
  + Retention Time (RT) can help with metabolite identification, but is instrument/lab/day-specific

**Downstream steps**

* Quality control
  + Performed with spiked-in metabolite standards and by analyzing constant pooled samples
* Metabolite identification
  + Match *m*/*z* + RT + spectra to in-house standards or databases (e.g. HMDB and Metlin)
  + Data-driven approaches like metabolite co-variation and “metabolic networking”
    - i.e. guilt-by-association on abundance or spectra
* Metabolite abundance tables are amenable to standard genomics analyses, including phenotype association, phenotype prediction, enrichment testing, clustering, and data integration

**Suggested reading**

* (*Review*) Johnson, CH et al. "[Metabolomics: beyond biomarkers and towards mechanisms](https://www.nature.com/articles/nrm.2016.25)." *Nature reviews Molecular cell biology* 17.7 (2016): 451.
* Franzosa, Eric A., et al. "[Gut microbiome structure and metabolic activity in inflammatory bowel disease](https://www.nature.com/articles/s41564-018-0306-4)." *Nature microbiology* 4.2 (2019): 293.