Unsupervised Clustering for Mass Cytometry Data

Xuhong Zhang
School of Public Health
CU Anshutz

Flow Cytometry / Mass Cytometry

- In biotechnology, **flow cytometry** is a laser- or impedance-based, biophysical technology employed in cell counting, cell sorting, biomarker detection and protein engineering, by suspending cells in a stream of fluid and passing them through an electronic detection apparatus. A flow cytometer allows simultaneous **multiparametric** analysis of the physical and chemical characteristics of up to thousands of particles per second.
- Flow cytometry is routinely used in the diagnosis of health disorders, especially blood cancers, but has many other applications in basic research, clinical practice and clinical trials.
- More recently, **mass cytometry** is a mass spectrometry technique based on inductively coupled plasma mass spectrometry and time of flight mass spectrometry used for the determination of the properties of cells (cytometry).
  - The practical flow rate is around 500 cells per second versus several thousand in flow cytometry.
  - Current chemical methods limits cytometer use to around 40 parameters per cell

Unsupervised Clustering

- We have hundreds of samples (people) for disease/control groups. Each sample, we have millions of single cells. For each cell, we have 38 parameters/bio makers.
- Unsupervised clustering will help us to learn and identify unknown cell subpopulations.
- BUT, we tried:
  - Classical unsupervised clustering, like K-means ---- not satisfying our requirement
  - MCMC ---- never converge
  - Bayesian Variational Inference ---- out of memory problem