Mathematical Frameworks for Integrative Analysis of Multi-omics Biological Data

This manuscript (permalink) was automatically generated from BIRSBiointegration/whitePaper@6b895b4 on April 22, 2020.

Authors

- **Kim-Anh Lê Cao**
  0000-0003-3923-1116 · mixOmicsTeam · mixOmicsTeam
  Melbourne Integrative Genomics, School of Mathematics and Statistics, University of Melbourne, Australia · Funded by Grant National Health and Medical Research Council Career Development fellowship (GNT1159458).

- **Jane Roe**
  XXXX-XXXX-XXXX-XXXX · janeroe
  Department of Something, University of Whatever; Department of Whatever, University of Something

Abstract

Introduction to single cell and imaging multi-omics

Testing github pages.

Current multi-omic technologies

Challenges for interpretation

Need for technology-specific questions and analysis methods vs one size fits all data blender

Case studies

scNMT-seq as a case-study for epigenetic regulation

Overview and biological question

Computational challenges

- Identification of multi-omics signatures that characterise lineage, stage or both.
- Handling missing values
- Do epigenetic changes in some genomic contexts affect cell fate decision more than others? If so, how?

Methods for stats/maths analyses and results summary

scRNA-seq + FISH as a case study for spatial transcriptomics

Overview and biological question

Computational challenges

- Can scRNA-seq data be overlaid onto seqFISH for resolution enhancement
- What is the minimal number of genes needed for data integration?
- Are there signatures of cellular co-localization or spatial coordinates in the non-spatial scRNA-seq data?

Methods for stats/maths analyses and results summary

Spatial proteomics and cross-study analysis

Overview and biological question

Computational challenges

- Integrating partially-overlapping proteomic data collected on different patients with similar phenotypes
- Integration of spatial x-y coordinate co-location and co-expression
- Integration with other 'omics datasets (e.g., scRNA-seq) to support the results of these proteomic analyses
- Can we predict the spatial expression patterns of proteins measured on mass-tag but not measured in the MIBI-TOF data?

- What additional information can we learn about the different macrophage and immune populations in breast cancer by conducting integrated analyses of these datasets?

Methods for stats/maths analyses and results summary

Overview of common analytical methods spanning technologies / case studies

- matrix factorization
- neural network / autoencoders

Data structures and software packages to enable analyses

- Bioc/multiAssayExperiment for single cell
- new classes for proteomics
- PyTorch

Techniques and challenges for benchmarking methods

- realistic simulation studies
- cross-study validation
- benchmark datasets

Discussion

Emerging analytical methods and technologies

Community needs for data structures, analysis methods, etc

References