Short Course: Statistical methods for single-cell RNA sequencing analysis

Joint Statistical Meetings, Vancouver, August 2018

Rhonda Bacher, PhD, Assistant Professor
Department of Biostatistics, University of Florida

Mingyao Li, PhD, Professor
Department of Biostatistics and Epidemiology, University of Pennsylvania

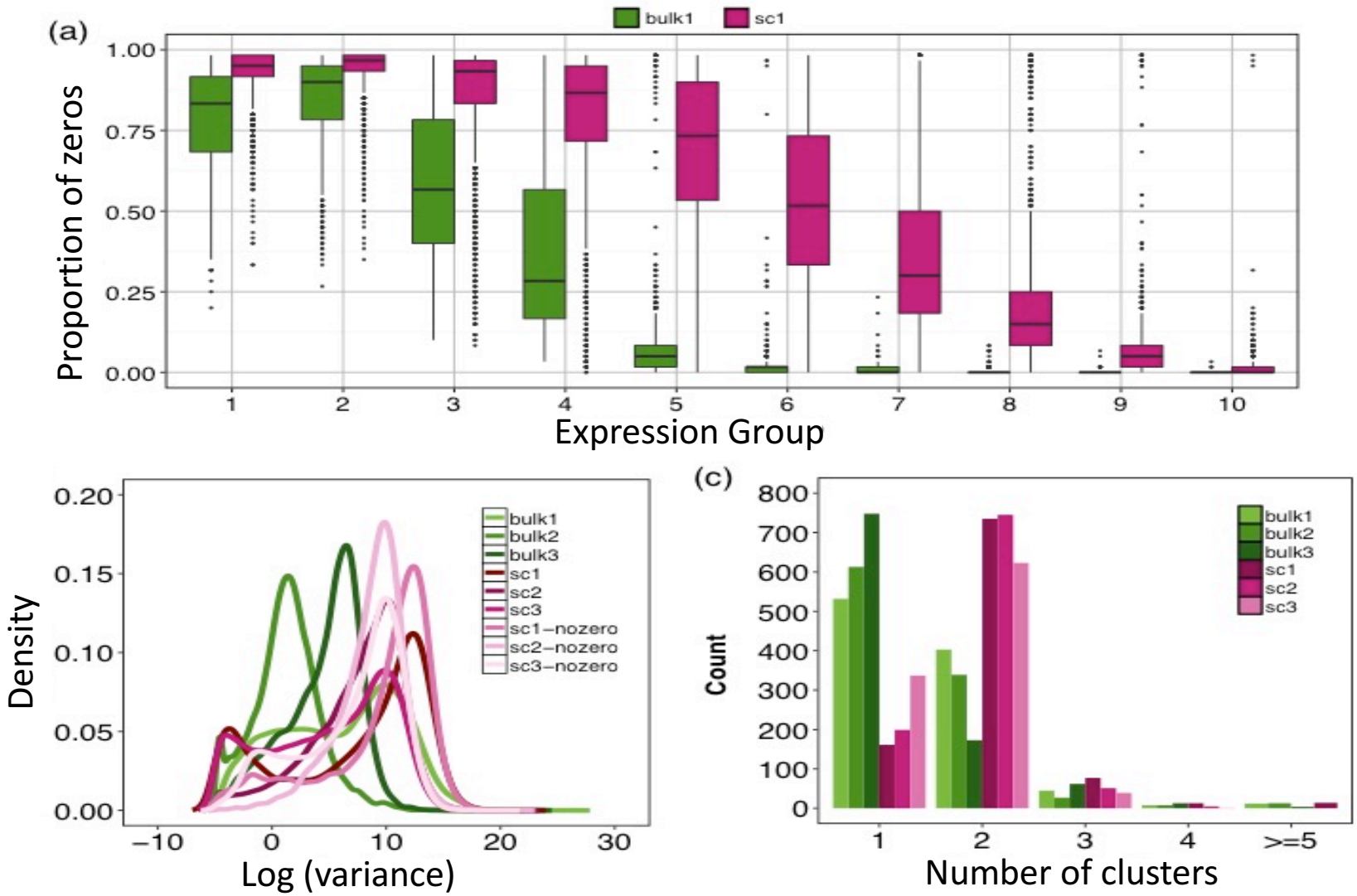
Nancy Zhang, PhD, Professor
Departments of Statistics, University of Pennsylvania

Christina Kendziorski, PhD, Professor
Department of Biostatistics and Medical Informatics, University of Wisconsin- Madison

https://github.com/rhondabacher/JSM2018_ShortCourseMaterials

Features of single-cell RNA-seq data

- Abundance of zeros, increased variability, complex distributions

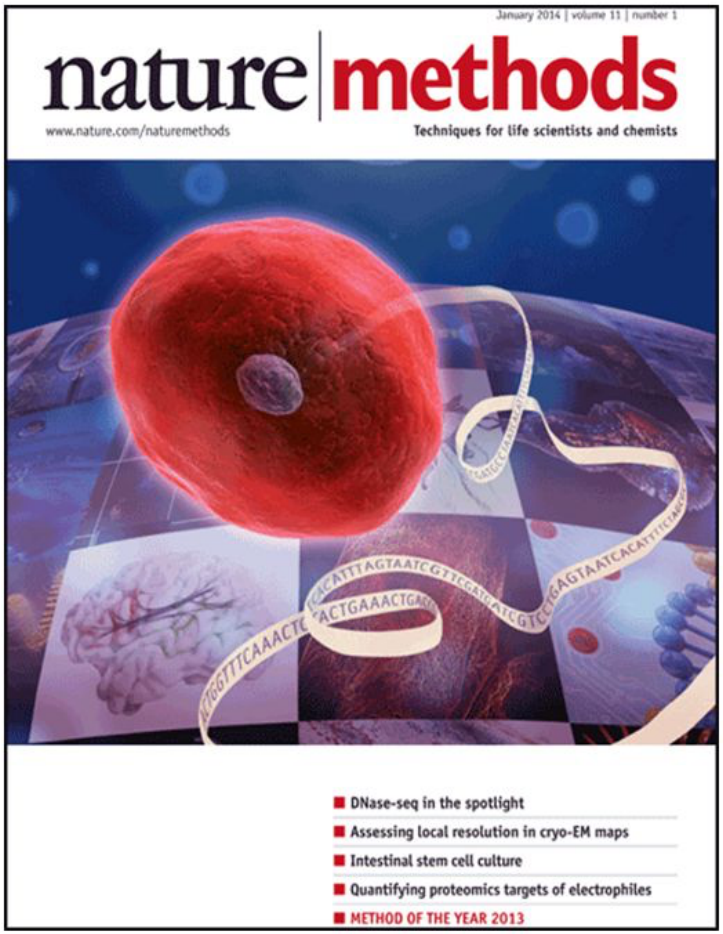


Bacher and Kendziorski, *Genome Biology*, 2016.

Challenges in scRNA-seq

- Normalization
- Technical vs. biological zeros
- Gene expression distributions
- Estimating expression and allele-specific expression
- Clustering; Identifying sub-populations
- De-noising (adjusting for technical and biological variability)
- Identifying and characterizing differences in gene-specific expression distributions (aka. identifying differential distributions)
- Pseudotime reordering
- Network reconstruction
- Developmental timing

Early press: exciting and informative



"Methods to sequence the DNA and RNA of single cells are poised to transform many areas of biology and medicine". -Editorial, Nature Methods

Impact of single-cell technologies: A few examples

- Cancer:
  - Identify rare mutations
  - Profile rare cells
  - Better characterize heterogeneity
  - Describe clonal structure and trace tumor evolution/metastases
  - Personalized therapeutics

- Development:
  - Identify genomic basis of developmental transitions
  - Characterize similarities across species

- Immunology:
  - Characterize heterogeneity among seemingly uniform populations (e.g. cells purified on the basis of cell surface markers).
  - Identify differences in cells not distinguishable by marker genes and/or cell morphology.
  - Group cells in an unbiased way.

Single-cell analysis: Critical challenges remain

More and more scientists are jumping into single-cell analysis, which spans classical cell biology, developmental biology, genomics and computational biology. And as the technologies to study single cells expand, they will require sophisticated analytical tools to tame and make sense of results.

News Feature, *Nature* 547 (19): July 2017

Single-cell analysis reviews (mostly statistical/computational methods)

- Angerer *et al*. Single cells make big data: New challenges and opportunities in transcriptomics. *Current Opinion in Systems Biology*, 2017.
- Bacher and Kendziorski. Design and computational analysis of single-cell RNA-sequencing experiments. *Genome Biology*, 2016.
- Conesa *et al*. A survey of best practices for RNA-seq data analysis. *Genome Biology*, 2016.
- Hon *et al*. The Human Cell Atlas: Technical approaches and challenges. *Briefings in Functional Genomics*, 2017.
- Vallejos *et al*. Normalizing single-cell RNA sequencing data: Challenges and opportunities. *Nature Methods*, 2017.
- Yuan *et al*. Challenges and emerging directions in single-cell analysis. *Genome Biology*, 2017.

Single-cell analysis: Consortia and data

In addition to Gene Expression Omnibus (GEO),

- Human Cell Atlas Data Portal
  - https://preview.data.humancellatlas.org/
- Single-cell Portal at the Broad
  - https://portals.broadinstitute.org/single_cell
- Conquer
  - http://imlspenticton.uzh.ch:3838/conquer/
- scRNASeqDB
  - Cao et al., scRNASeqDB: A Database for RNA-Seq Based Gene Expression Profiles in Human Single Cells. Genes (Basel), 2017.
  - https://bioinfo.uth.edu/scrnaseqdb/
- SCPortalen: Single-cell Centric Database
  - Abugessaisa *et al*. SCPortalen: human and mouse single-cell centric database, Nucleic Acids Research, 2018.
  - http://single-cell.clst.riken.jp/
- Single Cell Analysis Program – Transcriptome Project (SCAP-T)
  - https://www.scap-t.org/
  - Requires application

Short Course: Statistical Methods for Single-cell RNA-seq Analysis
Joint Statistical Meetings, Vancouver, 2018

1:00-1:10: Logistics and Outline (CKendziorski)

1:10-1:40: Overview of scRNA-seq technologies and QC (RBacher)

1:40-2:10 Single-cell expression distributions (NZhang)

2:10-2:40: Normalization methods including SCnorm (RBacher)

2:40-3:10: Analysis of allele-specific gene expression (MLi)

3:10-3:25: Break

3:25-3:55: Single-cell expression denoising and imputation (NZhang)

3:55-4:25: Bulk tissue cell type deconvolution with scRNA-seq gene expression reference (MLi)

4:25-4:55: Identifying differential distributions and pseudotime reordering (CKendziorski)

4:55: Closing comments/ questions/ discussion

https://github.com/rhondabacher/JSM2018_ShortCourseMaterials