Differential Expression Analysis Techniques for Single-Cell RNA-seq Experiments

for the Computational Biology Doctoral Seminar (CMPBIO 293),
organized by N. Yosef & T. Ashuach, Spring 2018, UC Berkeley

Kevin Benac and Nima Hejazi

Group in Biostatistics,
University of California, Berkeley

11 April 2018

Outline

The Data: Single-Cell RNA-seq

- scRNA-seq fast growing approach to measure the genome-wide transcriptome of many individual cells in parallel (Kolodziejczyk et al., 2015).
- Major advance compared to standard "bulk" RNA sequencing to investigate complex heterogeneous tissues,
- Access to cell-to-cell variability: better accuracy.

The Data: Single-Cell RNA-seq

- However, analysis of scRNA-seq data challenging.
- In one cell, only a tiny amount of RNA is present and large fraction of polyadenylated RNA can be stochastically lost during sample preparation steps (cell lysis, reverse transcription or amplification).
  $\Longrightarrow$ Many genes fail to be detected although they are expressed!
- In practice, not uncommon to end up with a matrix of read counts where about 80% of the coefficients are zeros.
- This zeros are called *dropouts*.

The Data: Single-Cell RNA-seq

	Cell1	Cell 2	Cell 3	Cell 4	Cell 5	Cell 6	Cell 7
Xkr4	0	0	0	14	0	0	0
Syt11	1	9	2	2	0	0	0
Cpe	0	0	16	0	0	0	0
Rp1	0	0	0	0	0	0	0
Gm73	0	0	0	0	0	0	0
Gm79	0	0	0	0	0	0	0
Mpl15	8	8	6	1	0	0	0
Gm61	0	0	0	0	0	3	0
Lypla1	1	23	266	1	0	1	0
Tcea1	63	101	18	29	2	34	0

The Data: Single-Cell RNA-seq

- Raises modelling and computational issues.
- Need to detect a signal when most of the values are zeros only because they are missing.
- Traditional methods used for bulk RNA-seq data might not be sensible anymore.

Outline

The Objective: Differential Expression

- ...

The Objective: Differential Expression

- ...

The Objective: Differential Expression

- ...

Outline

ZINB-WaVE

- Method that leads to low-dimensional representations of the data the same way PCA or tSNE does.

ZINB-WaVE

- Method that leads to low-dimensional representations of the data the same way PCA or tSNE does.
- However accounts for zero inflation (dropouts), over-dispersion, and the count nature of the data.
- No need for normalization.

ZINB-WaVE

Mathematical set-up:

- n samples (single-cells),
- J genes,
- Y_{ij} read counts for gene j in cell i, $1 \leq \ldots \leq n, \quad 1 \leq j \leq J.$,
- π_{ij}: probability of dropout,
- μ: mean expression level.

ZINB-WaVE

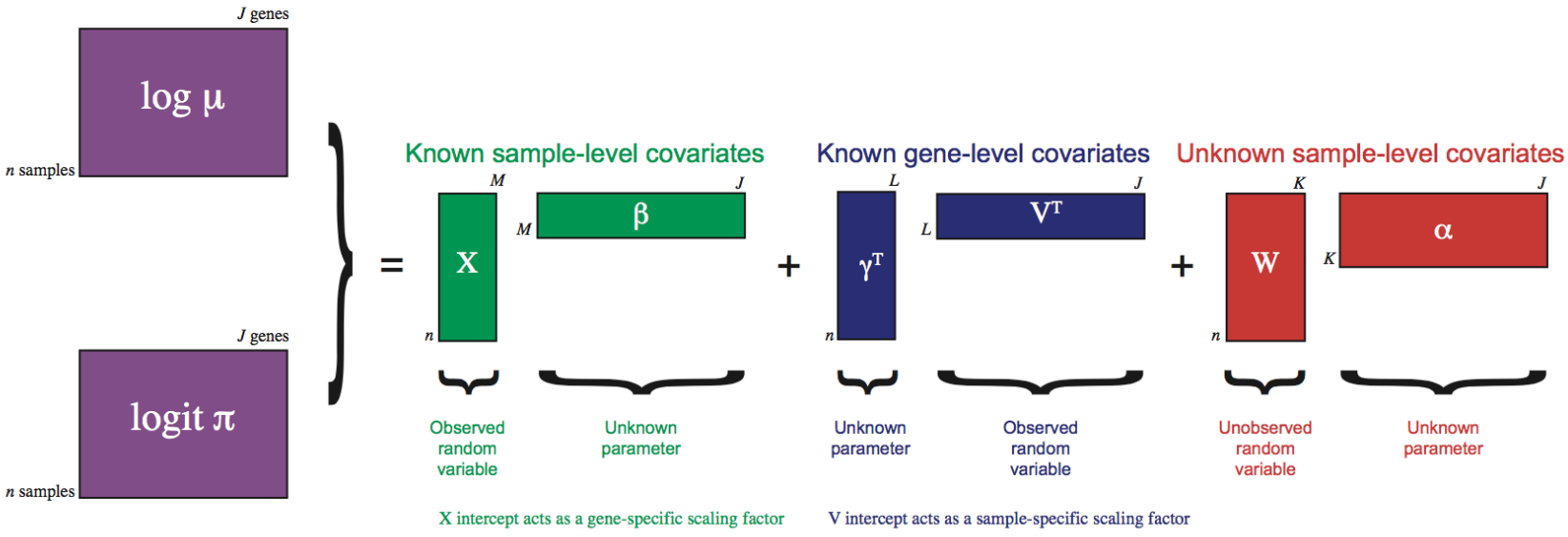

Figure 1: The ZINB-WaVE model

ZINB-WaVE

- ZINB-WaVE mainly used for normalization and dimensionality reduction but can also be used for DE analysis.
- Compute weights from the estimated π using Bayes formula.
- If the observed counts are positive, $w = 1$, otherwise, $0 < w < 1$.
- The higher π, the lower w

ZINB-WaVE

- Once we have the weights, fit a negative binomial glm using the weights.
- End-up with a matrix of fitted values.
- Not sparse anymore, look more like bulk RNA-seq data.
  $\Longrightarrow$ We can use classical tools for differential expression analysis (ex. edgeR or DESeq2 packages in R).

Outline

DropLasso I

- ...
- ...

DropLasso II

- ...
- ...

DropLasso III

- ...
- ...

DropLasso IV

- ...
- ...

DropLasso V

- ...
- ...

Outline

ZINB-WaVE v. DropLasso I

- ...
- ...

ZINB-WaVE v. DropLasso II

- ...
- ...

ZINB-WaVE v. DropLasso III

- ...
- ...

Outline

Review

- ...
- ...

References I