

Statistical and Computational Methods in Single-cell Analysis

Zhicheng Ji

Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health

Overview

1 Introductions to Single-cell Sequencing

2 Methods for scRNA-seq

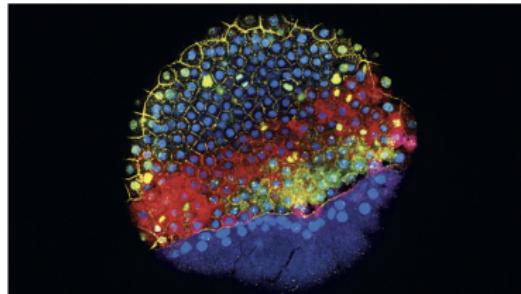
3 Methods for scATAC-seq

Introductions

BREAKTHROUGH OF THE YEAR

Development cell by cell

With a trio of techniques, scientists are tracking embryo development in stunning detail



A zebrafish embryo at an early stage of development. Fluorescent markers highlight cells expressing genes that help determine the type of cell they will become. (JEFFREY FARRELL, SCHIER LAB/HARVARD UNIVERSITY)

Editorial | Published: 30 December 2013

Method of the Year 2013

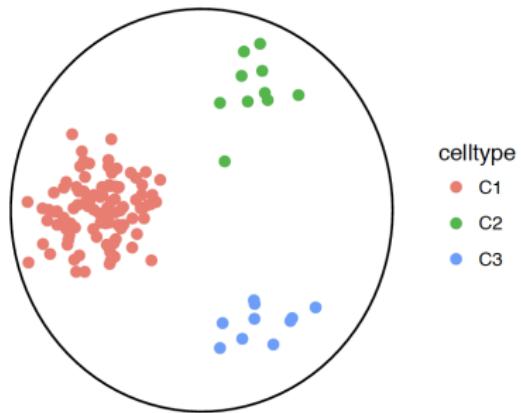
Nature Methods 11, 1 (2014) | Download Citation ↗

Methods to sequence the DNA and RNA of single cells are poised to transform many areas of biology and medicine.

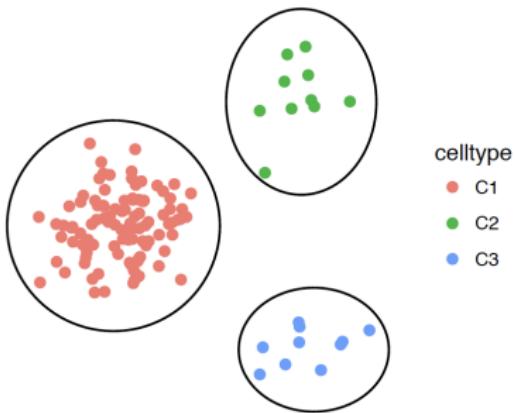
Once considered a technical challenge reserved for a few specialized labs, single-cell transcriptome and genome sequencing is becoming robust and broadly accessible. Exciting insights from recent studies are revealing the potential to understand biology at the unitary resolution of life, and last year marked a turning point in the widespread adoption of these methods to address a variety of research questions. For these reasons, single-cell sequencing is our choice of Method of the Year for 2013.

Introductions

Bulk RNA-seq: averaged transcriptomes



Single-cell RNA-seq: individual transcriptomes



Technologies

single-cell DNA/CNV sequencing: DNA sequence information in single cells.

single-cell RNA-seq (scRNA-seq): Gene expressions in single cells.

single-cell ATAC-seq (scATAC-seq): Open chromatin activities in single cells.

single-cell Bisulfite sequencing: DNA methylation in single cells.

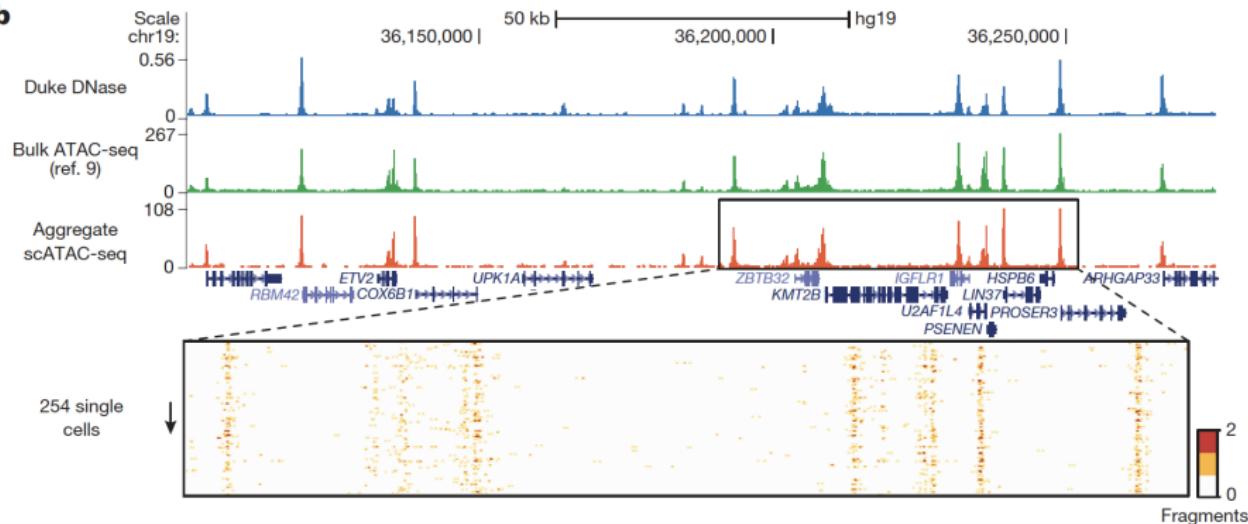
single-cell VDJ-seq: TCR/Ig sequence information in single T/B cells.

CITE-seq: cell surface protein expressions in single cells.

Others: scDNase-seq, scChIP-seq, scHi-C

Sparsity

b



Jason D. Buenrostro, et al. Single-cell chromatin accessibility reveals principles of regulatory variation. (2015) Nature.

Analysis Tasks

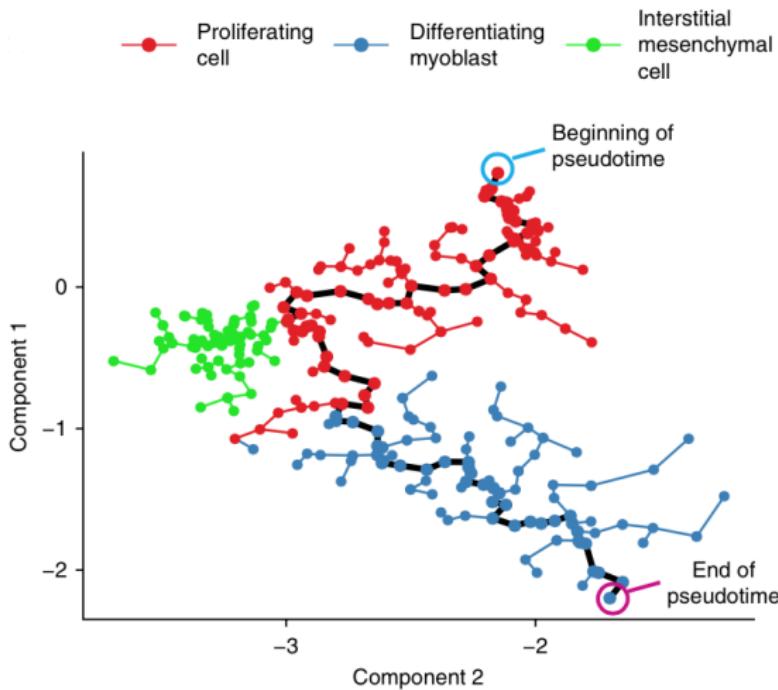
scRNA-seq

- Imputation
- Cell Clustering
- Dimension Reduction
- Differential Gene
- Sample alignment/batch effect removal
- Pseudotime Analysis

scATAC-seq

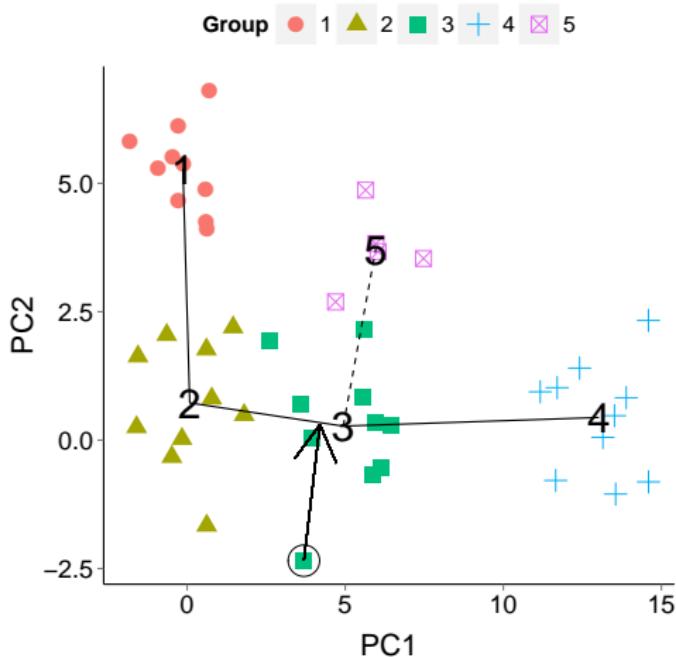
- Signal Aggregation/Enhancing
- Cell Clustering
- Inferring Regulatory Programs
- Prediction/Aligning with scRNA-seq

TSCAN: Pseudotime Reconstruction

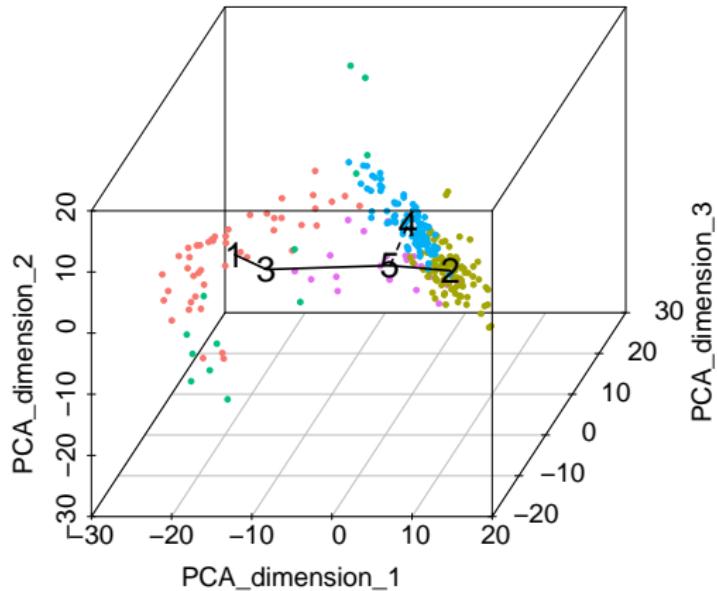


Cole Trapnell, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. (2014) Nat Biotechnol.

TSCAN: Pseudotime Reconstruction

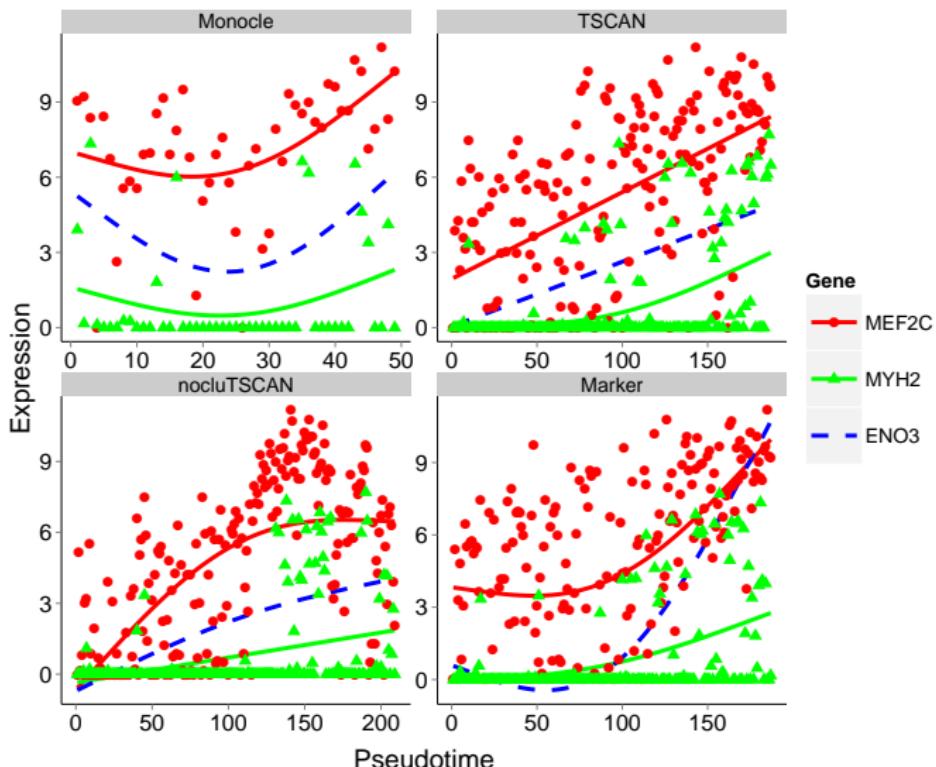


TSCAN: Pseudotime Reconstruction



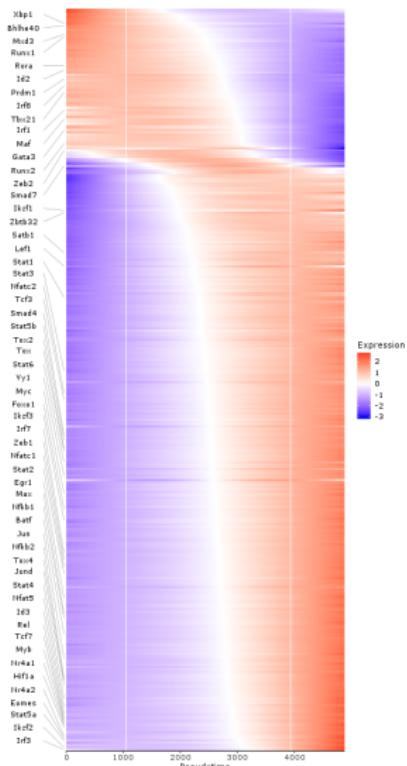
Zhicheng Ji, et al. TSCAN: Pseudo-time reconstruction and evaluation in single-cell RNA-seq analysis. (2016) Nucleic Acids Res.

TSCAN: Pseudotime Reconstruction



Zhicheng Ji, et al. TSCAN: Pseudo-time reconstruction and evaluation in single-cell RNA-seq analysis. (2016) Nucleic Acids Res.

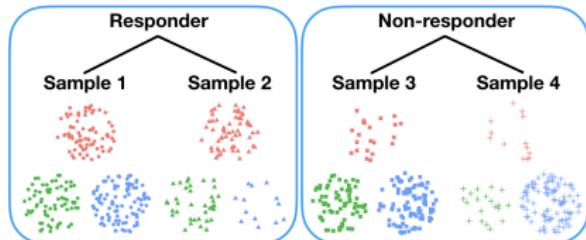
STIP: State Transition Inference Prediction



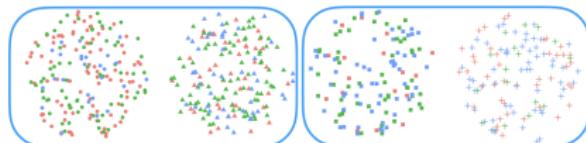
Zeyu Chen*, Zhicheng Ji*, et al. A TCF-1-centered transcriptional network drives an early binary effector versus exhausted CD8+ T cell fate decision . In revision Immunity.

Raisin: Differential Gene

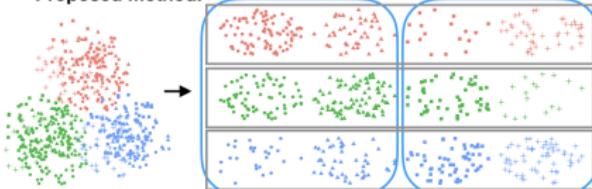
Differential between groups of samples



Current Methods (limma, Wilcoxon, MAST, etc.):



Proposed Method:



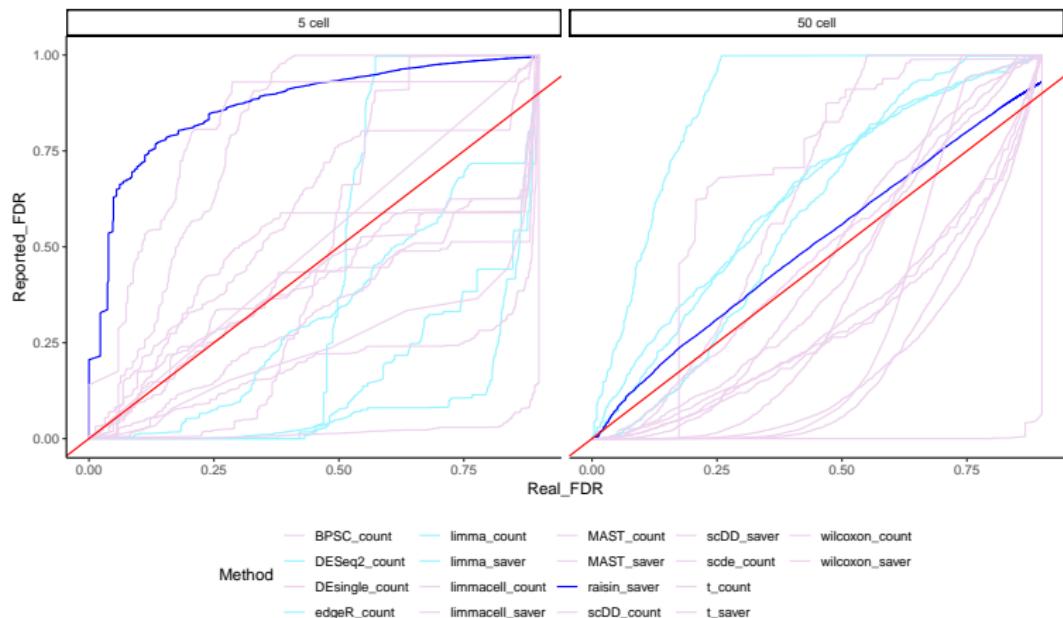
Treat cells as samples (Wilcoxon test, MAST, etc.)
Falsey assume independency of cells within samples

Pool cells as bulk for each sample (Limma, DEseq, etc.)
Falsey ignore the variability by single-cell measurements

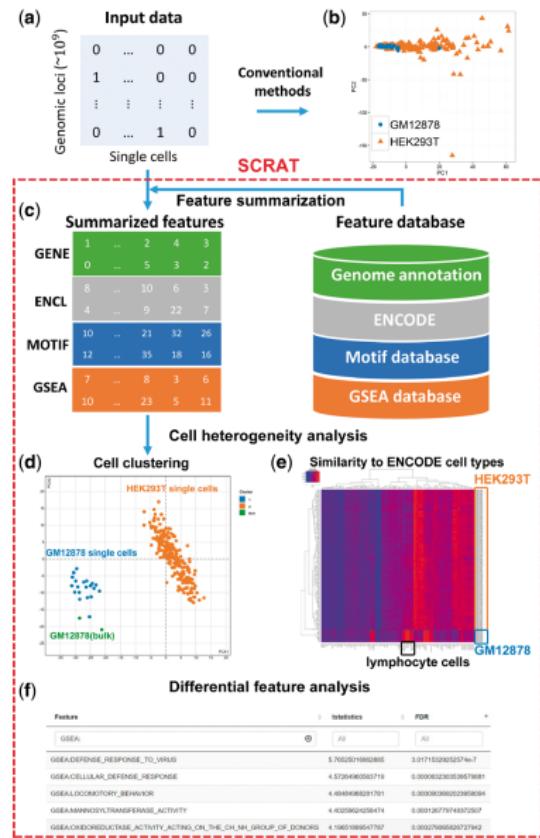
Hierarchical variance structure, wrong assumption
(Limma duplicateCorrelation)
Incorrect characterization of variance

Proposed method: Raisin
Hierarchical variance structure, Correct characterization of variance

Raisin: Differential Gene



SCRAT: scATAC-seq Signal Aggregation

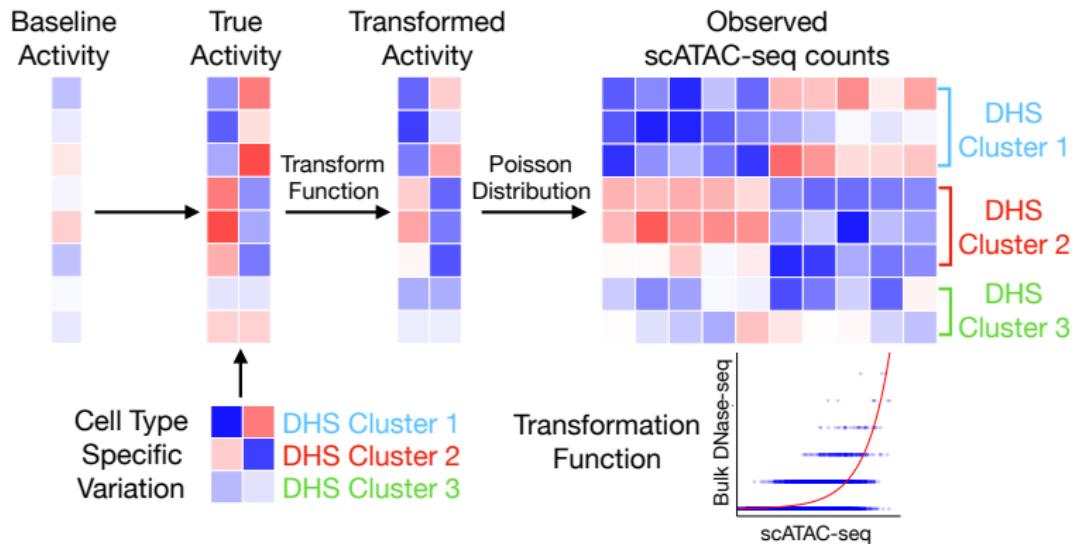


Zhicheng Ji*, Weiqiang Zhou*, et al. Single-cell regulome data analysis by SCRAT. (2017)

Zhicheng Ji (Department of Biostatistics, Jol

Single-cell Analysis

SCATE: scATAC-seq Signal Enhancement



SCATE: scATAC-seq Signal Enhancement

$$y_{i,j} \sim \text{Poisson}(L_j \mu_{i,j}^{sc})$$

$$\log \mu_{i,j}^{sc} = h(\log(\mu_{i,j}))$$

$$\log(\mu_{i,j}) = a_i + b_i * \delta_{c(i),j}$$

Objective:

Learn $\delta_{c(i),j}$ with MLE

$y_{i,j}$: observed scATAC-seq read count for DHS i and cell j

$\mu_{i,j}^{sc}$: Transformed activity

$\mu_{i,j}$: True activity

h : transformation function

a_i, b_i : Baseline mean and standard deviation

$\delta_{c(i),j}$: Cell type specific bias

SCATE: scATAC-seq Signal Enhancement

