Single-cell RNA-sequencing

Stefania Giacomello

UC Davis, 22 June 2018
Applications

- Heterogeneity analysis
- Cell-type identification
- Cellular states in differentiation and developmental processes
- Splicing patterns
Cell types

- Function
- Function at specific tissue
- Populations and subpopulations
- Cell states (cell cycles, active/inactive, apoptosis, etc.)
Academic single-cell methods

Improving throughput (n. of cells)
Robustness (varying quality of samples)
**Complexity** (n. of unique transcripts per cell)
**Accuracy** (low technical noise; many cells – shallow sequencing)

Kolodziejczyi A et al., Molecular Cell, 2015
Single-cell sequencing technologies

- Input: Single cells in suspension + 10x Gel Beads and Reagents
- Output: Digital gene expression profiles from every partitioned cell
Single-cell sequencing technologies
Single-cell sequencing technologies
Single-cell sequencing technologies

ICELL8 Chips and Reagents

Imaging Station

MultiSample NanoDispenser

CellSelect Software
Single-cell sequencing technologies

ICELL8 Chips and Reagents

Imaging Station

MultiSample NanoDispenser

CellSelect Software
Single-cell sequencing technologies

ICELL8 Chips and Reagents

Imaging Station

MultiSample NanoDispenser

CellSelect Software
Single-cell sequencing technologies

REVOLUTIONARY NEW SINGLE-CELL PLATFORM

1. Isolate up-to 1,800 cells per chip
2. Evaluate cells from 5-100 μm per sample
3. Select specific cells for downstream applications
4. Discover unique populations of cells
Cell isolation

Kolodziejczyi A et al., Molecular Cell, 2015
Cell suspensions

Mechanical/enzymatic dissociation:
- bias for specific subpopulations
- Different dissociation kinetics compared to their normal counterparts or between samples of the same disease
- No duplets → microscopy

Affects:
Robustness (varying quality of samples)

Accuracy (high technical noise)
Protocols

Zieghain et al., Molecular Cell, 2017
Cell isolation

Stochastic gene expression

Gene 1
Gene 2
Gene 3
Gene 4

Bulk RNAseq

Dissociate
Bias due to cell type/state

Reverse transcription
“random” selection of 10-40% of mRNAs - Drop-outs

Amplification
May have bias due to length, structure, gc-content

20% Gene 1
30% Gene 2
50% Gene 3
0% Gene 4

0% Gene 1
55% Gene 2
25% Gene 3
20% Gene 4

Åsa Björklund - NBIS
**Some theory**

**Drop-out** = the transcript is present in the cell but not detected due to missed conservation to cDNA

**Transcriptional bursting** = the transcript is present in most cells of a specific cell-type but not in every cell

**Lowly expressed transcript** = drop-out or low bursting?
The sensitivity depends on the efficiency of the reverse transcription reaction → drop-out rate between 90 to 60% depending on methods

Kolodziejczyi A et al., Molecular Cell, 2015
Amplification steps introduce bias in the data

UMIs allows to avoid PCR duplicates

Kolodziejczyi A et al., Molecular Cell, 2015
Differences between single-cell and bulk RNA-seq

- Amplification bias
- Drop-out
- Transcriptional bursting
- Background noise
- Bias due to cell-cycle, cell size
- Clear batch effects
How to analyze the data

- Mapping - STAR
- QC analysis – number of genes
- Filtering
- Normalization – SC Ran (Aaron et al., Genome Biology, 2016)
- Dimensionality reduction
- Clustering, marker genes, annotation
- Differential gene expression
- Trajectory
Useful tools - Seurat

R toolkit for single-cell genomics

https://satijalab.org/seurat/
Useful tools – SC3

Single-cell Consensus Clustering

Kiselev et al., Nature Methods, 2017
Other tools

- Pagoda ([http://hms-dbmi.github.io/scde/pagoda.html](http://hms-dbmi.github.io/scde/pagoda.html))
Other tools

- Pagoda ([http://hms-dbmi.github.io/scde/pagoda.html](http://hms-dbmi.github.io/scde/pagoda.html))
- Graphs ([http://igraph.org/r/](http://igraph.org/r/))
Dimensionality reduction step

Convert high-dimensional data to a more simplified representation, while maintaining the main characteristics of the data in the original space.

Kumar et al. Development, 2017
Dimensionality reduction techniques:

- PCA (linear projection of the data such that the variance is preserved in the new space)
- independent component analysis (ICA)
- t-stochastic neighbor embedding (t-SNE)
- diffusion maps

Graph-based techniques
- cells = nodes in a graph
- edges = connect transcriptionally similar cells
- It retains the most important edges in the graph \( \rightarrow \) scales well to large numbers of cells \( (n > 10\,000) \)
Trajectory inference
The basics

Cells display a **continuous spectrum of states** (i.e. activation and/or differentiation process)

Individual cells are executing through a gene expression program in an **unsynchronized** manner → each cell is a **snapshot of the transcriptional program** under study

**sc-omics** technologies allow to **model biological systems**
The basics

Discrete classification of cells is not appropriate

Summary of the continuity of cell states in the data

→ Trajectory Inference (TI) (or pseudotemporal ordering)
What is a trajectory?

Sequence of gene expression changes each cell must go through as part of a dynamic biological process.
What is a trajectory?

Sequence of gene expression changes each cell must go through as part of a dynamic biological process

Track changes in gene expression:
- function of time
- function of progress along the trajectory
What is a trajectory?

Sequence of gene expression changes each cell must go through as part of a dynamic biological process

Track changes in gene expression:
- function of time
- function of progress along the trajectory

Pseudotime → abstract unit of progress: distance between a cell and the start of the trajectory
How do TI tools work?

1. Population of single cells → different stages

2. Computational tools to order cells along a trajectory topology
   Automatic reconstruction of a cellular dynamic process by structuring individual cells sampled and profiled from that process

3. Identify the different stages in the dynamic process and their interrelationships

What TI offers

• Unbiased and transcriptome-wide understanding of a dynamic process

• They allow the objective identification of new subsets of cells

Type of trajectories

Trajectory’s total length: total amount of transcriptional change that a cell undergoes as it moves from the starting to the end state

Linear, branched, or a more complex tree or graph structure
Type of input data

- Transcriptome-wide data
- Starting cell from which the trajectory will originate
- Set of important marker genes, or even a grouping of cells into cell states.
Providing prior information:

- Can help the method to find the correct trajectory among many, equally likely, alternatives

- If available, can bias the trajectory towards current knowledge
How TI tools usually work

1. conversion of data to a simplified representation using:
   - dimensionality reduction
   - clustering
   - graph building

2. ordering the cells along the simplified representation:
   - identify cell states
   - constructing a trajectory through the different states
   - projecting cells back to the trajectory
Tools available

59 methods - unique combination of characteristics:

- required input
- methodology used
- produced outputs (topology fixing and trajectory type)

*Saelens et al., bioRxiv, 2018*
Topology of the trajectory:

- **fixed by design**

  Early methods
  Mainly focused on correctly ordering the cells along the fixed topology

- **inferred computationally**

  Increased difficulty of the problem
  Broadly applicable on more use cases
  Topology inference still in the minority
Tool classification

TI methods classified also on a set of algorithmic components:

- Performance
- Scalability
- Output data structures
Monocle 2

Monocle introduced the concept of pseudotime

Now it has a complete new version - has been rated one of the most performing methods
Trajectory inference workflow:

1. Choosing genes to order the data

2. Reducing dimensionality of the data

3. Ordering cells in pseudotime
Fates of human fetal heart cells

State 1 2 3

Component 1

Component 2

stem-like

differentiated

Branch 1

Branch 2

start-point

SciLifeLab
Fates of human fetal heart cells

Component 1

Component 2

State 1 2 3

Cardiomyocyte-like

Endothelial-like

Branch 1

Branch 2

Low gene expression

High gene expression

TTN

TNNT2

TNNI3

MYL3

ENG

EGFL7

ESAM
Fates of human fetal heart cells

Expression

Pseudotime (stretched)