Single-cell RNA-sequencing

Stefania Giacomello

UC Davis, 22 June 2018

SciLifeLab

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.)

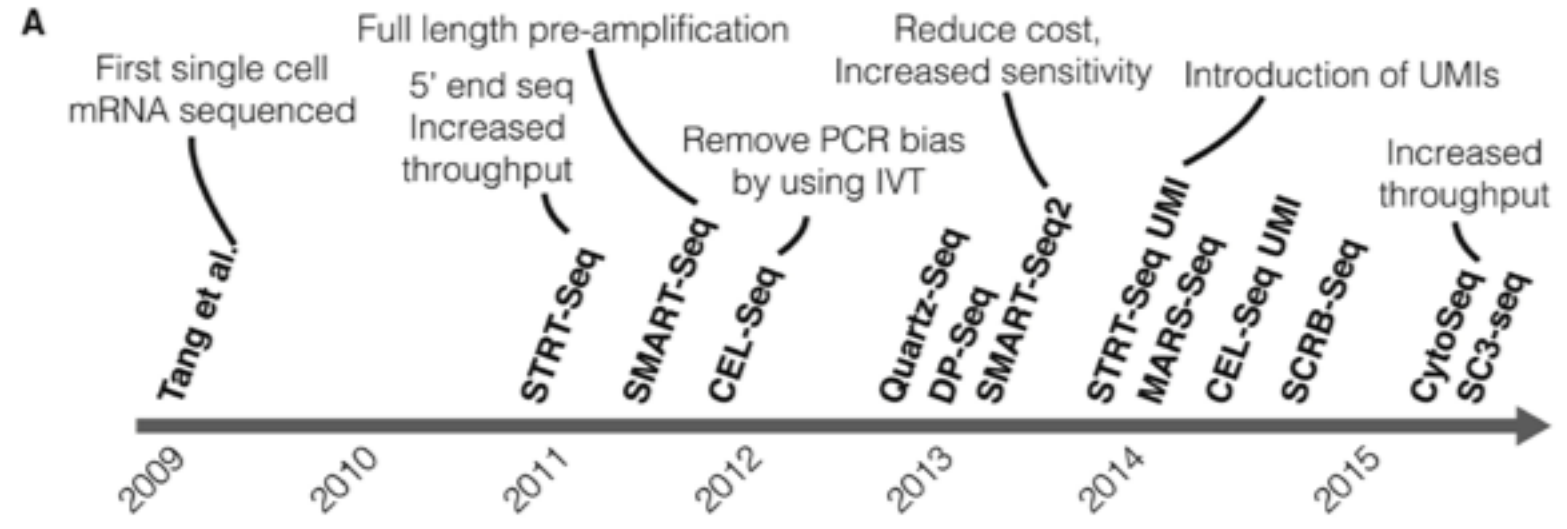
Academic single-cell methods

Improving throughput (n. of cells)

Robustness (varying quality of samples)

Complexity (n. of unique transcripts per cell)

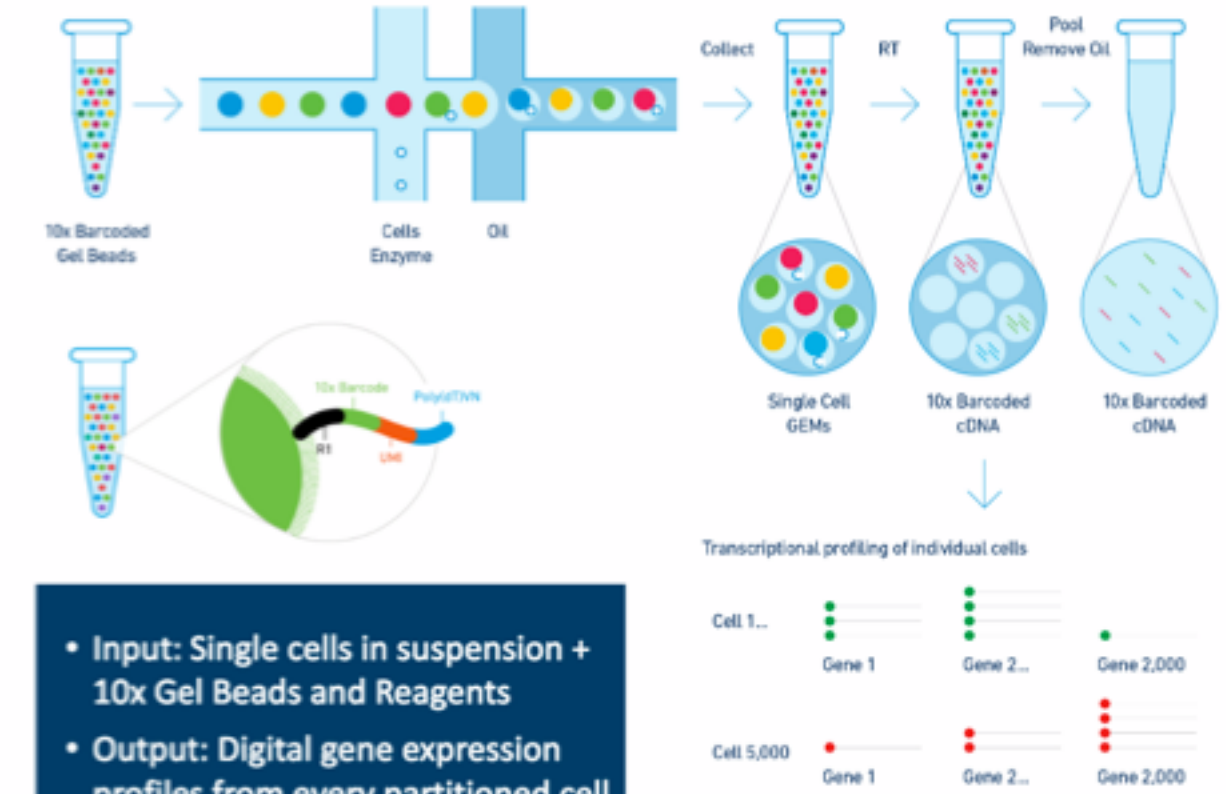
Accuracy (low technical noise; many cells – shallow sequencing)



Kolodziejczyi A et al., Molecular Cell, 2015

Single-cell sequencing technologies

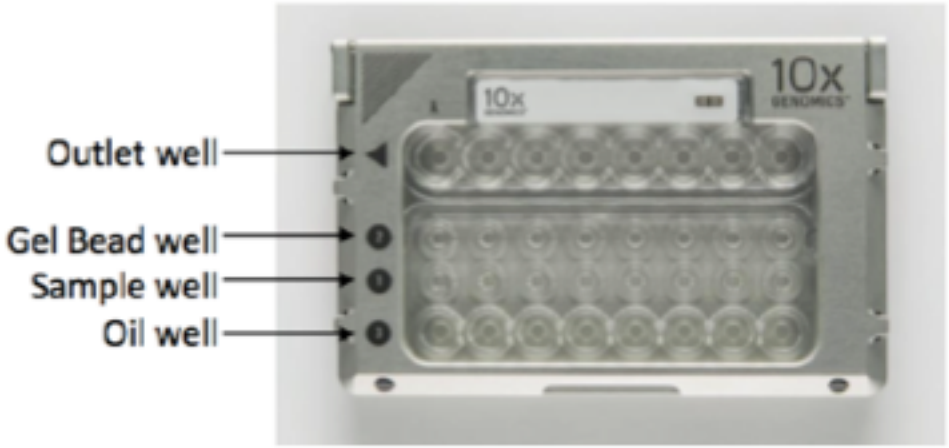
10x GENOMICS™

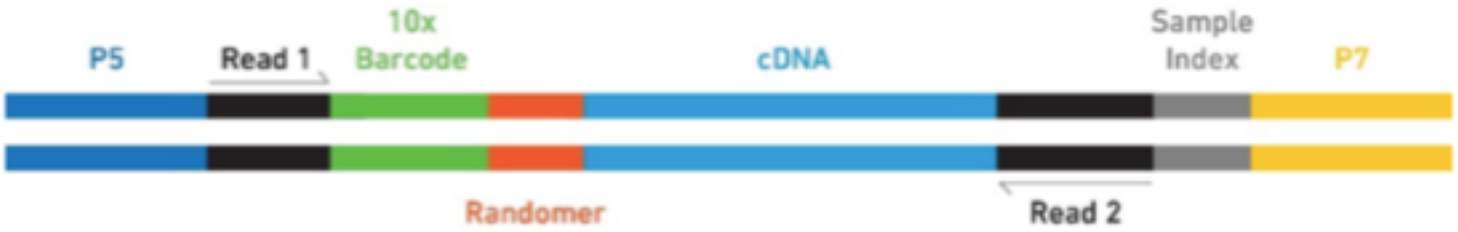


- Input: Single cells in suspension + 10x Gel Beads and Reagents
- Output: Digital gene expression profiles from every partitioned cell

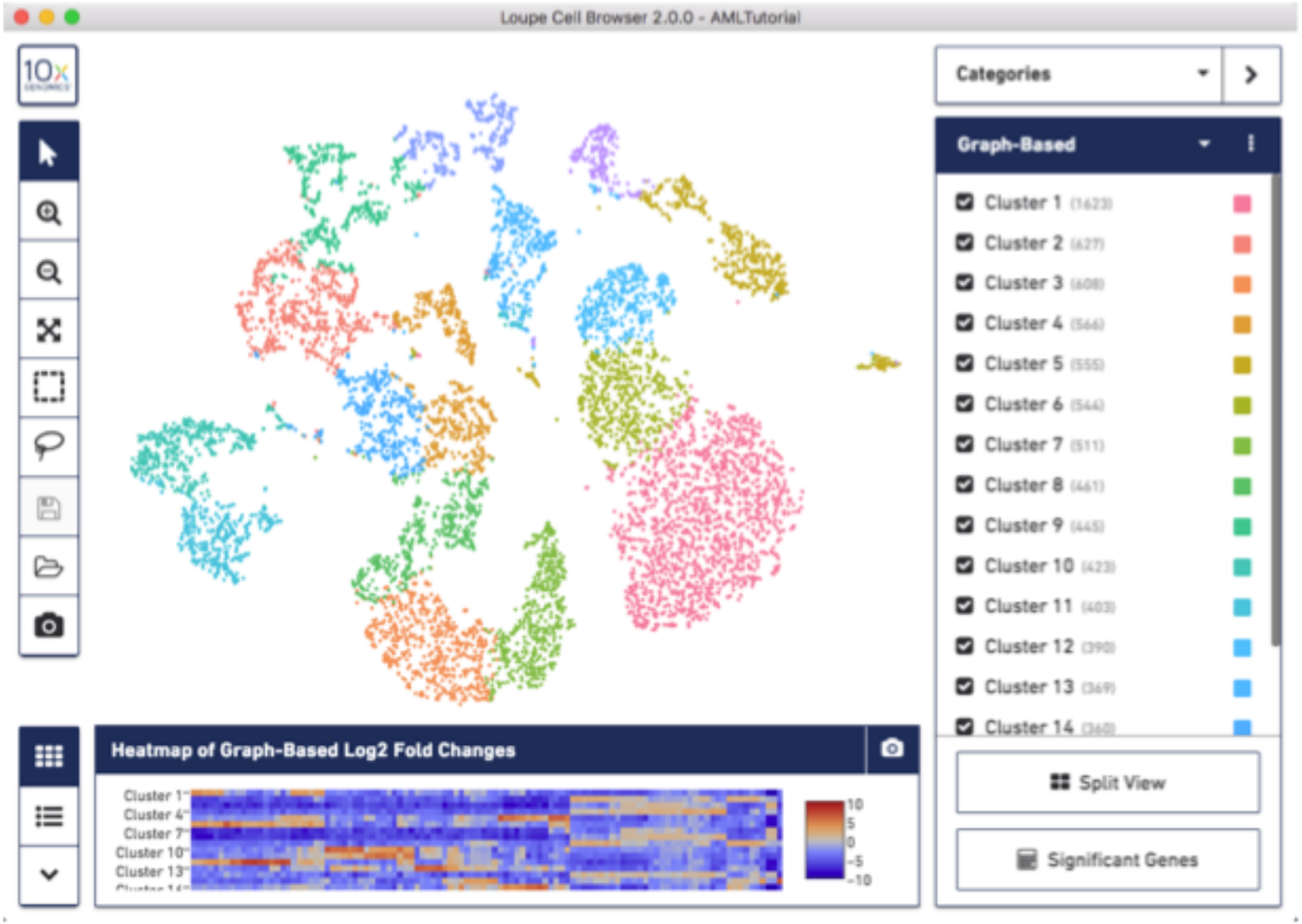
Single-cell sequencing technologies

10x GENOMICS™





Single-cell sequencing technologies

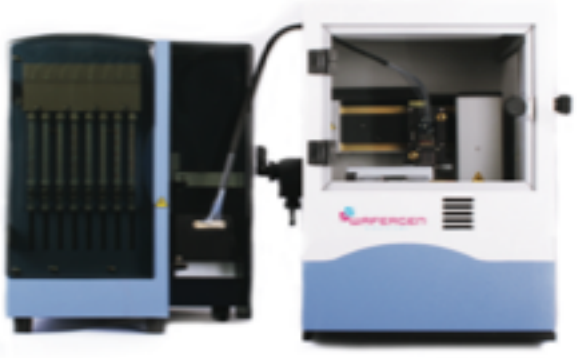


Single-cell sequencing technologies

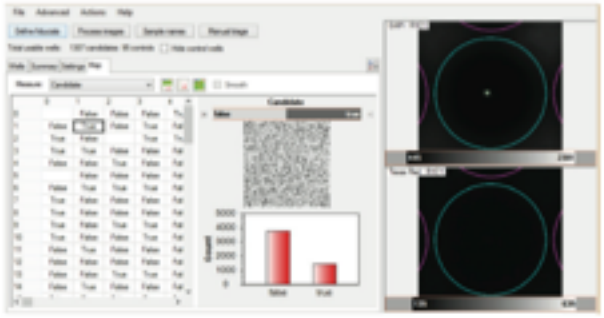
WAFERGEN BIOSYSTEMS

ICELL8 Chips and Reagents

Imaging Station



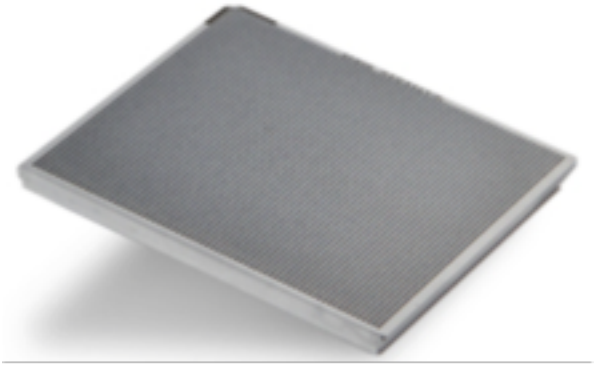
MultiSample NanoDispenser



CellSelect Software

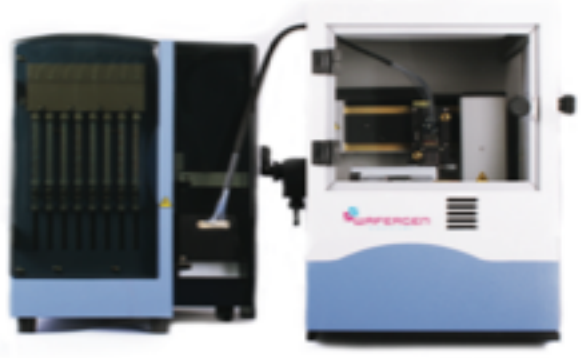
Single-cell sequencing technologies

WAFERGEN BIOSYSTEMS

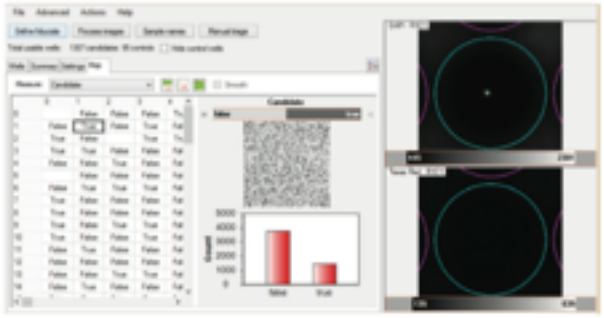


ICELL8 Chips and Reagents

Imaging Station



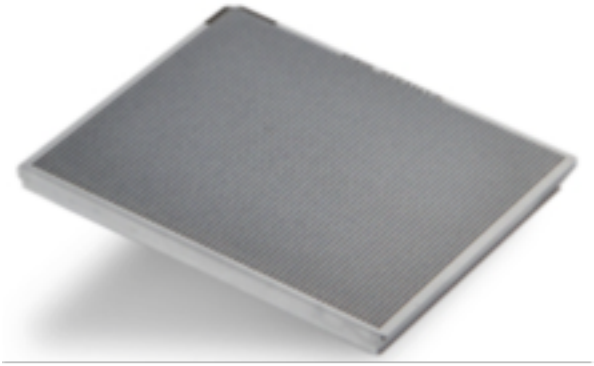
MultiSample NanoDispenser



CellSelect Software

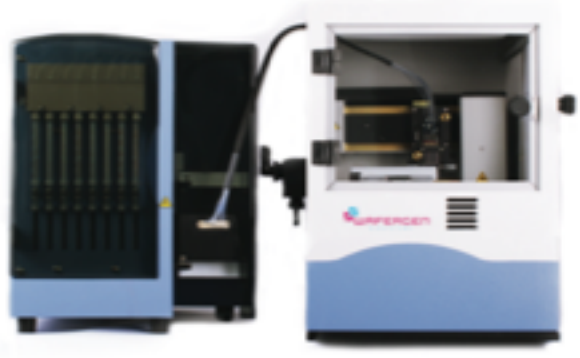
Single-cell sequencing technologies

WAFERGEN BIOSYSTEMS

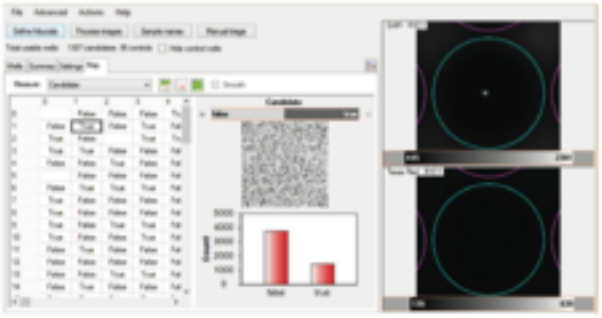


ICELL8 Chips and Reagents

Imaging Station



MultiSample NanoDispenser



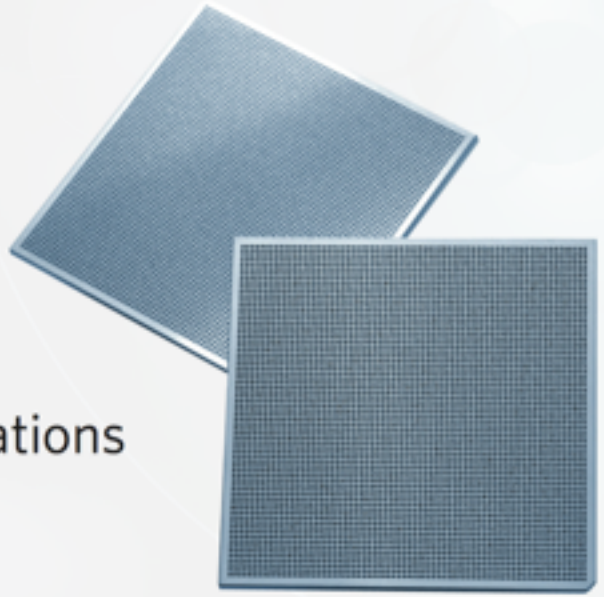
CellSelect Software

Single-cell sequencing technologies

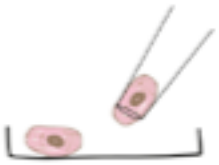
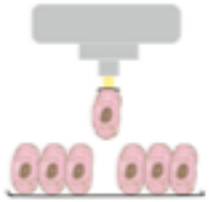
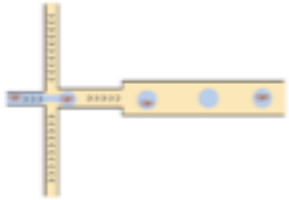
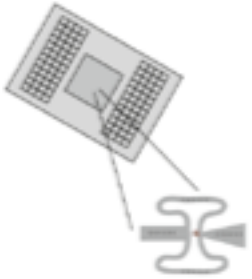
WAFERGEN BIOSYSTEMS

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

MICROPIPETTING MICROMANIPULATION	LASER CAPTURE MICRODISSECTION	FACS	MICRODROPLETS	MICROFLUIDICS e.g. FLUIDIGM C1
				
low number of cells	low number of cells	hundreds of cells	large number of cells	hundreds of cells
any tissue	any tissue	dissociated cells	dissociated cells	dissociated cells
enables selection of cells based on morphology or fluorescent markers	enables selection of cells based on morphology or fluorescent markers	enables selection of cells based on size or fluorescent markers	no selection of cells (can presort with FACS)	no selection of cells (can presort with FACS)
visualisation of cells	visualisation of cells	fluorescence and light scattering measurements	fluorescence detection	visualisation of cells
time consuming	time consuming	fast	fast	fast
reaction in microliter volumes	reaction in microliter volumes	reaction in microliter volumes	reaction in nanoliter volumes	reaction in nanoliter volumes

Kolodziejczyi A et al., Molecular Cell, 2015

Cell suspensions

Mechanical/enzymatic dissociation:

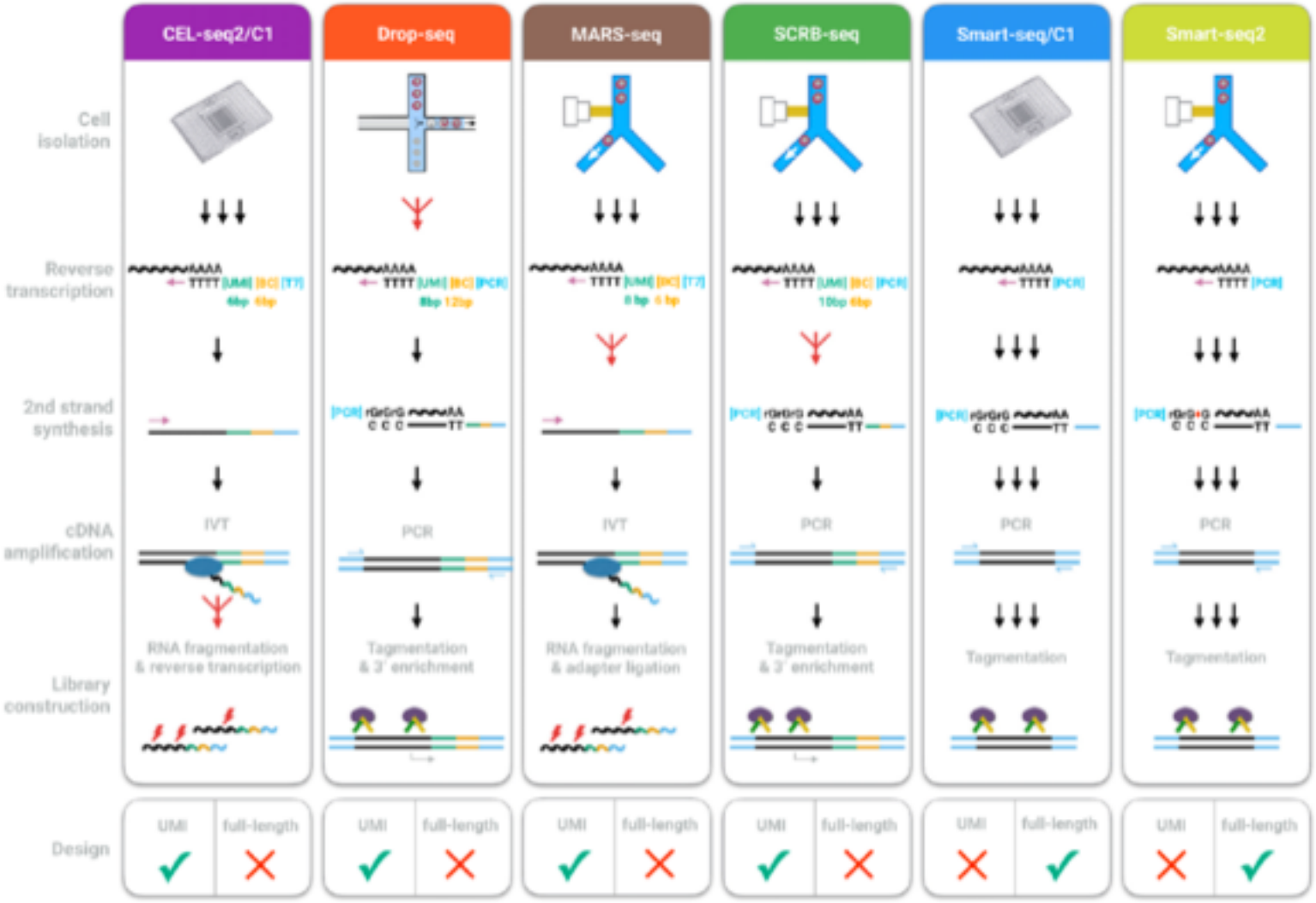
- bias for specific subpopulations
- Different dissociation kinetics compared to their normal counterparts or between samples of the same disease
- No duplets → microscopy

Affects:

Robustness (varying quality of samples)

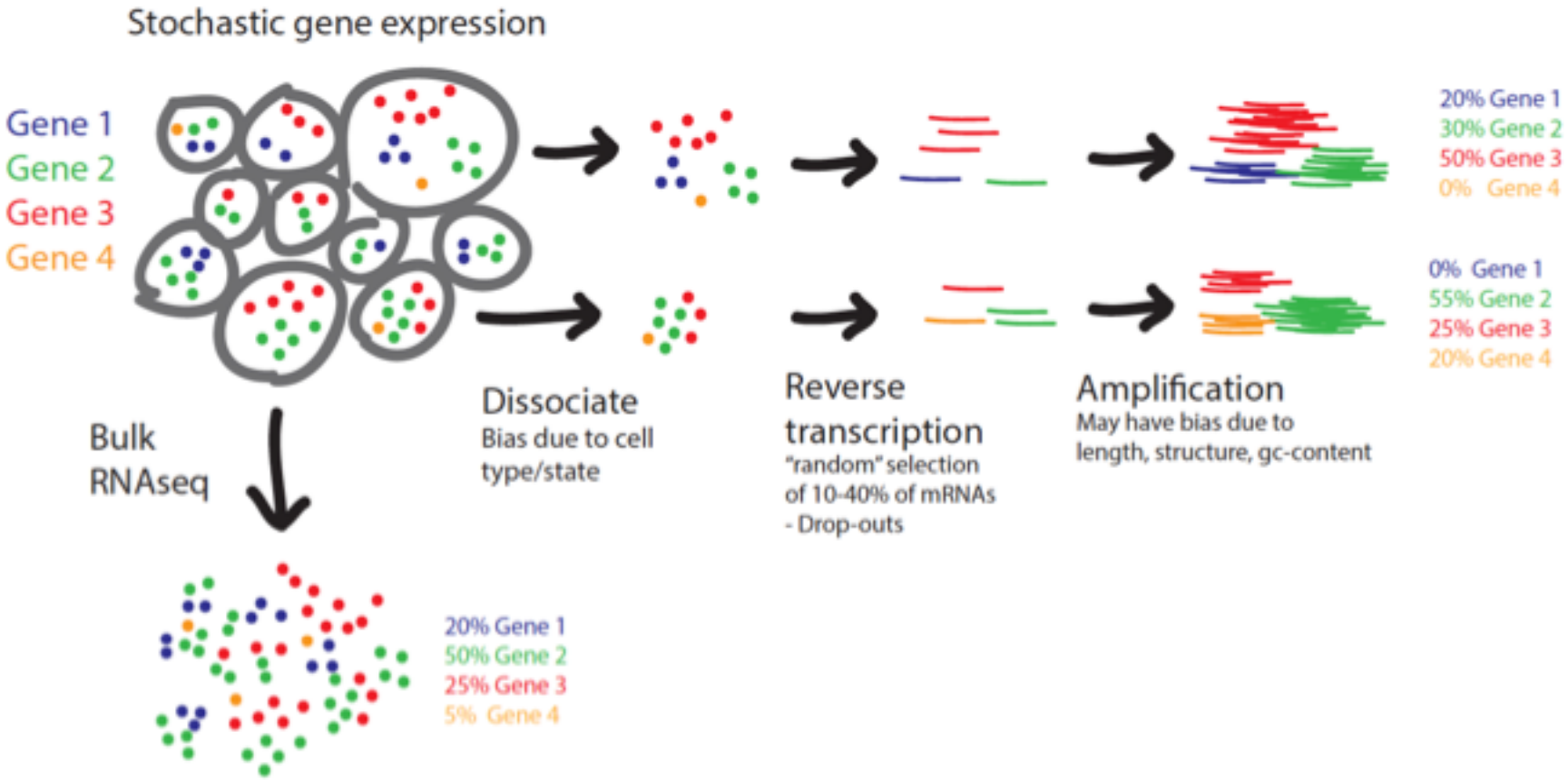
Accuracy (high technical noise)

Protocols



Zieghain et al., Molecular Cell, 2017

Cell isolation



Some theory

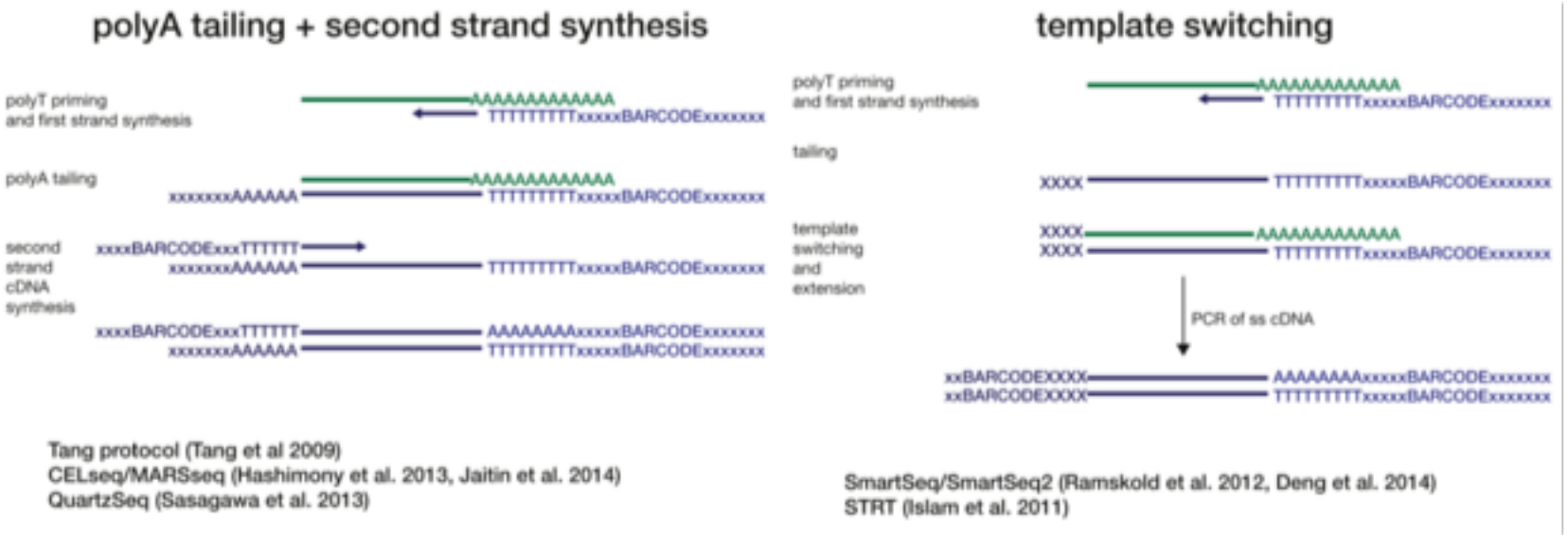
Drop-out = the transcript is present in the cell but not detected due to missed conservation to cDNA

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?

Reverse Transcription (RT)

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

PCR

- exponential amplification

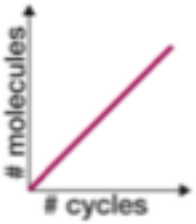
- PCR base specific biases

Tang protocol (Tang et al. 2009)
STRT (Islam et al. 2011)
SmartSeq/SmartSeq2 (Ramskold et al. 2012, Deng et al. 2014)

IVT

- linear amplification

- 3' bias due to two rounds of reverse transcription



CELseq/MARSseq (Hashimony et al. 2013, Jaitin et al. 2014)

Kolodziejczyi A et al., Molecular Cell, 2015

Amplification steps introduce bias in the data

UMIs allows to avoid PCR duplicates

Differences between single-cell and bulk RNA-seq

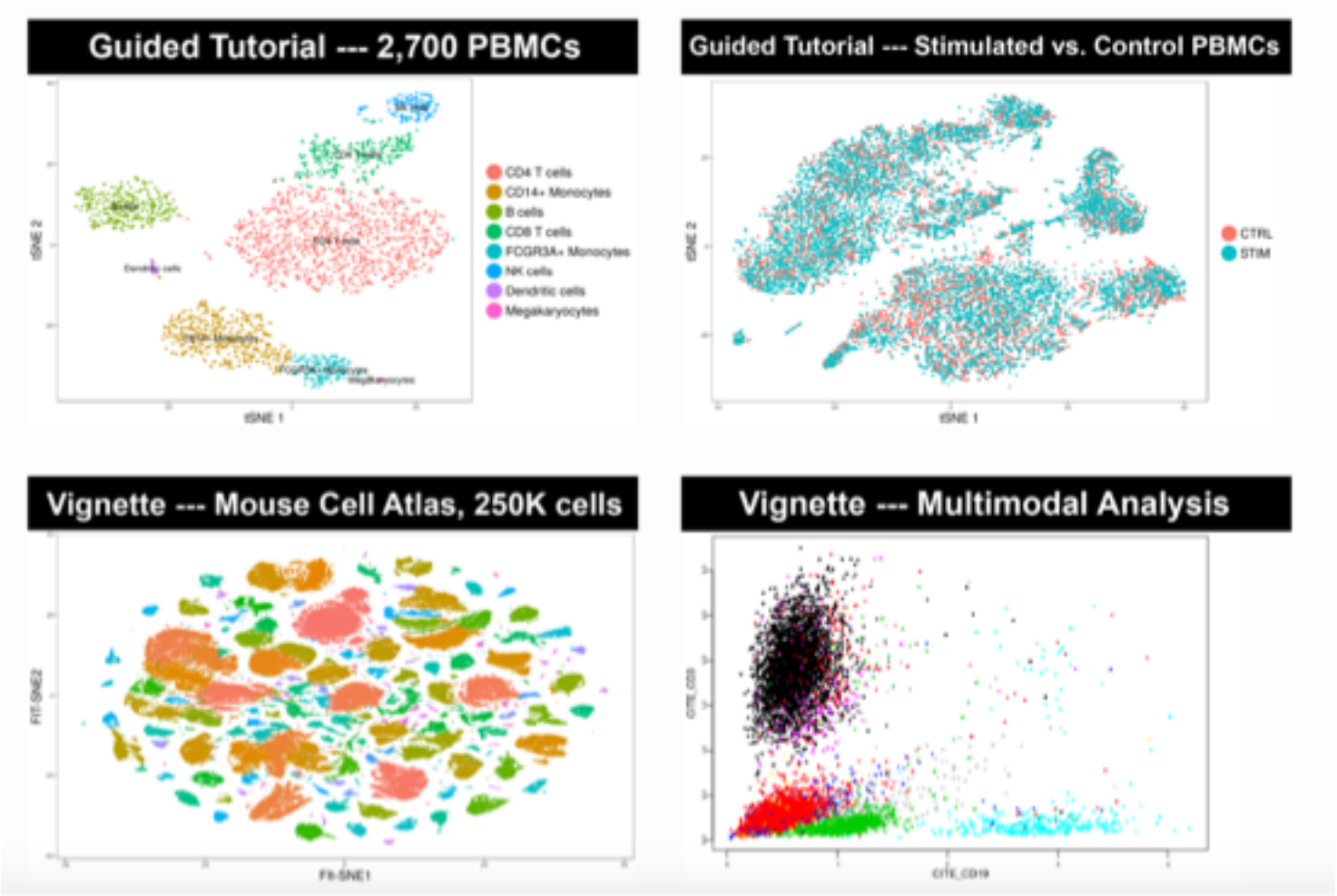
- Amplification bias
- Drop-out
- Transcriptional bursting
- Background noise
- Bias due to cell-cycle, cell size
- Clear batch effects

How to analyze the data

- Mapping - STAR
- QC analysis – number of genes
- Filtering
- Normalization – SCRAN (*Aaron et al., Genoome Biology, 2016*)
- Dimensionality reduction
- Clustering, marker genes, annotation
- Differential gene expression
- Trajectory

Useful tools - Seurat

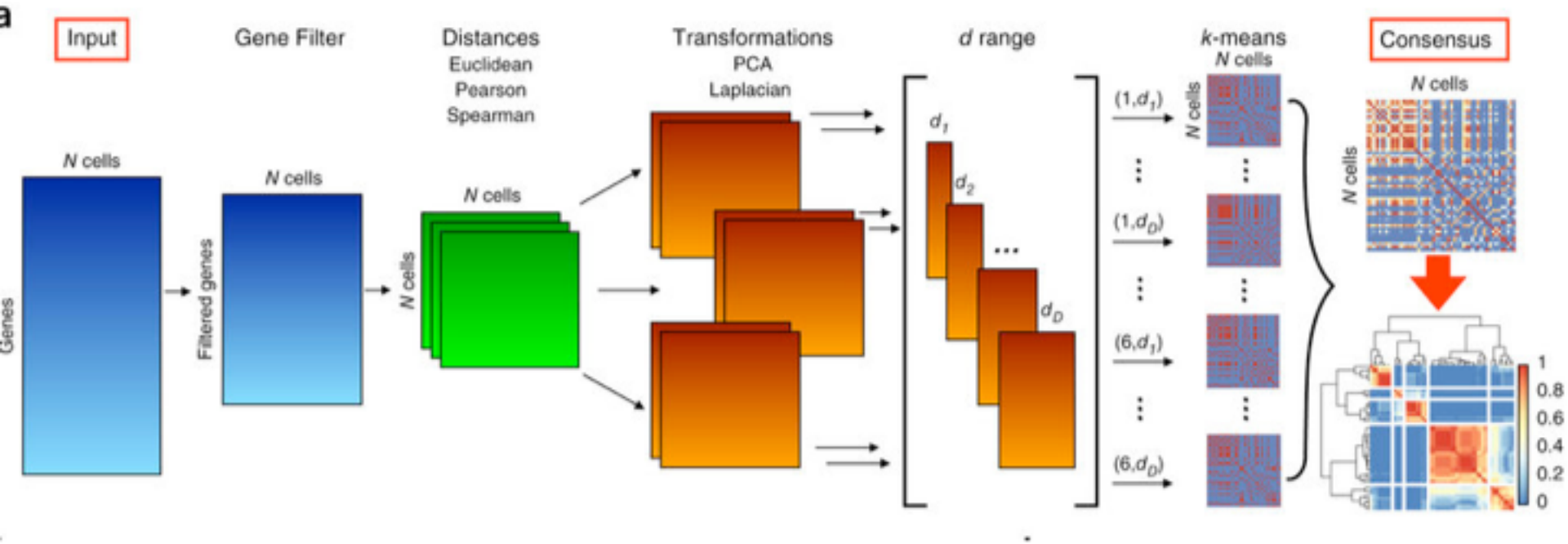
R toolkit for single-cell genomics



https://satijalab.org/seurat/

Useful tools – SC3

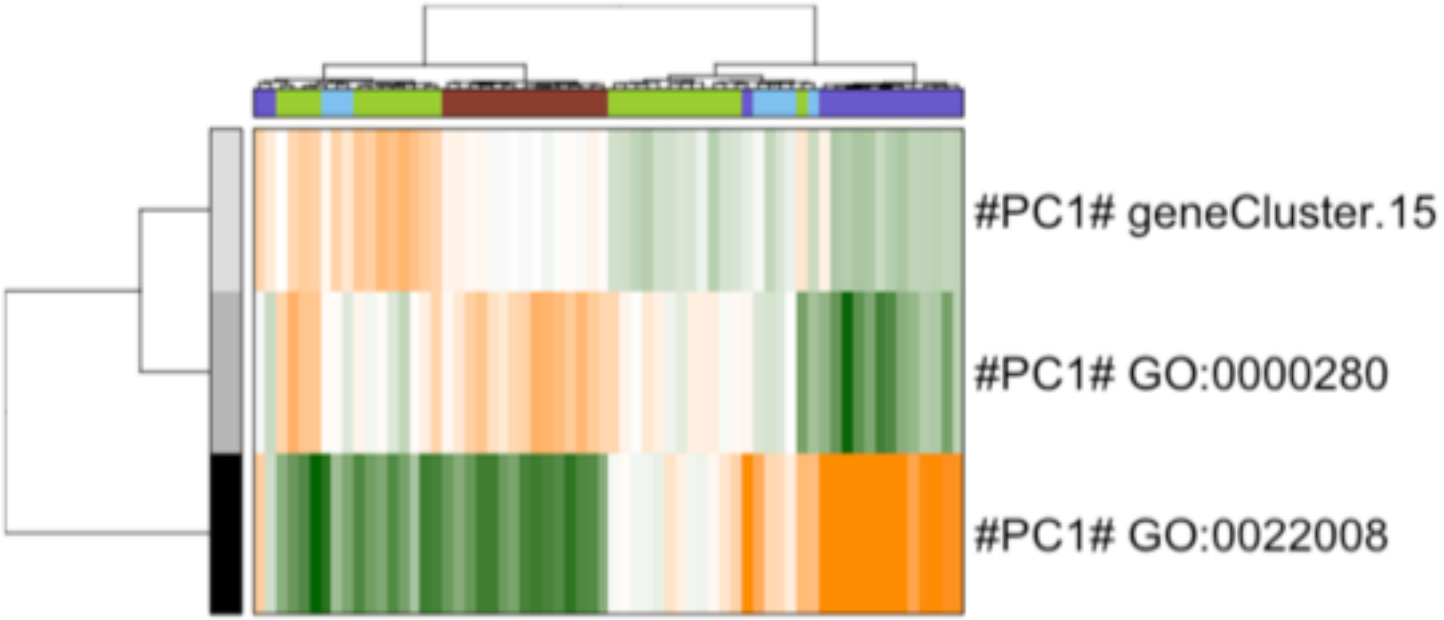
Single-cell Consensus Clustering



Kiselev et al., Nature Methods, 2017

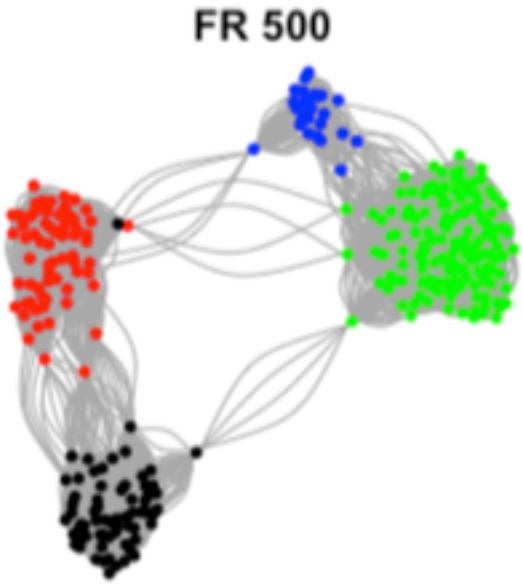
Other tools

- Pagoda (*http://hms-dbmi.github.io/scde/pagoda.html*)



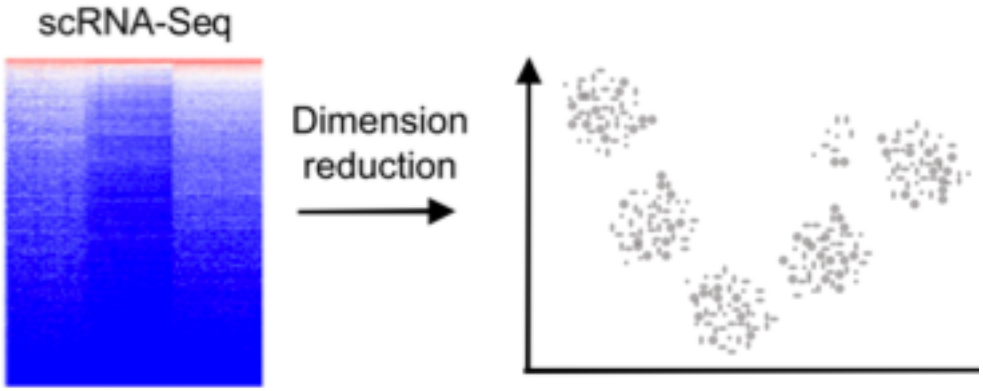
Other tools

- Pagoda (http://hms-dbmi.github.io/scde/pagoda.html)
- Graphs (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 step

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

} able to detect nonlinear relationships between cells

- Graph-based techniques
  cells = nodes in a graph
  edges =connect transcriptionally similar cells
  It retains the most important edges in the graph → 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

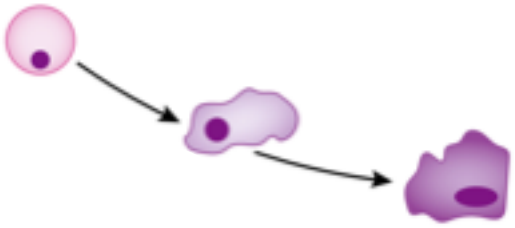
The basics

Discrete classification of cells is not appropriate

Summary of the continuity of cell states in the data
→ Trajectory Inference (TI) (or pseudotemporal ordering)

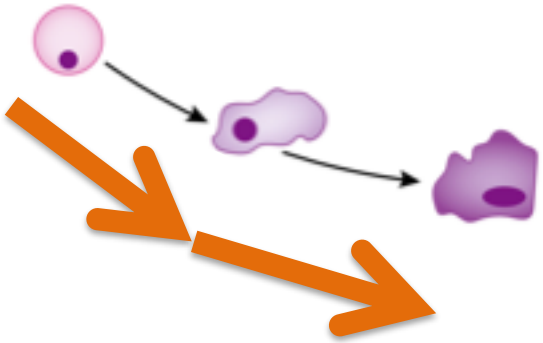
What is a trajectory?

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

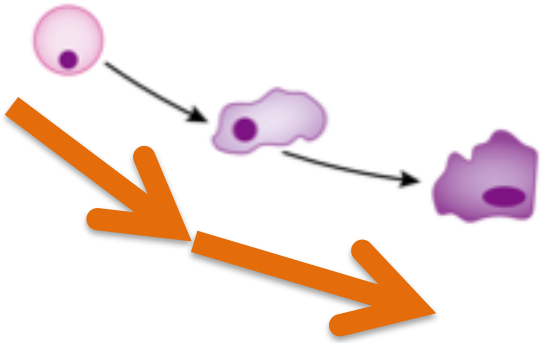


Track changes in gene expression:

- 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



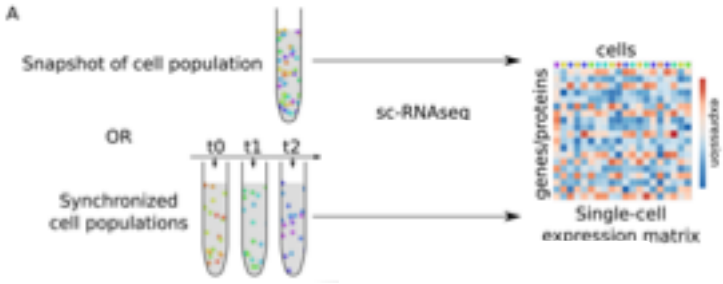
Track changes in gene expression:

- function of time
- function of progress along the trajectory

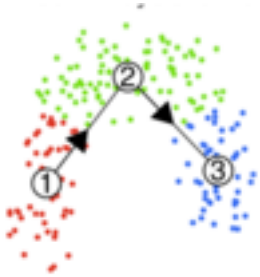
Pseudotime → abstract unit of progress:
distance between a cell and the start of the trajectory

How do TI tools work?

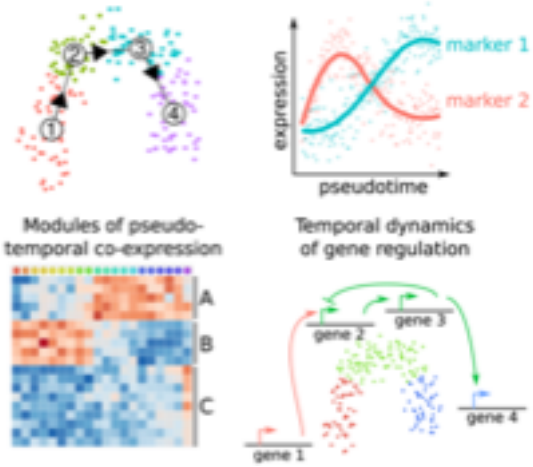
1. Population of single cells → different stages



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

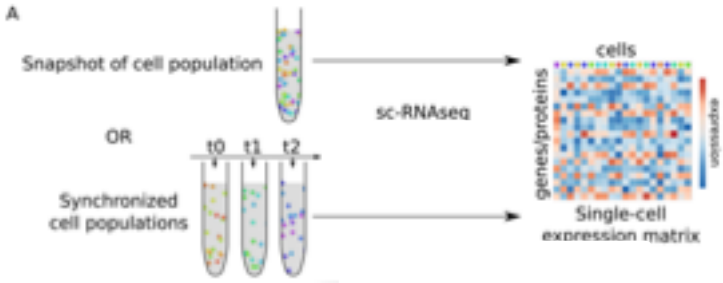


3. Identify the different stages in the dynamic process and their interrelationships

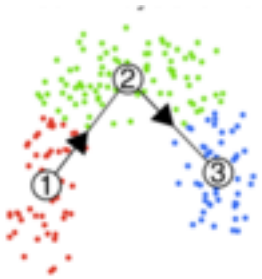


Cannot et al., Eur. J. Immun, 2016

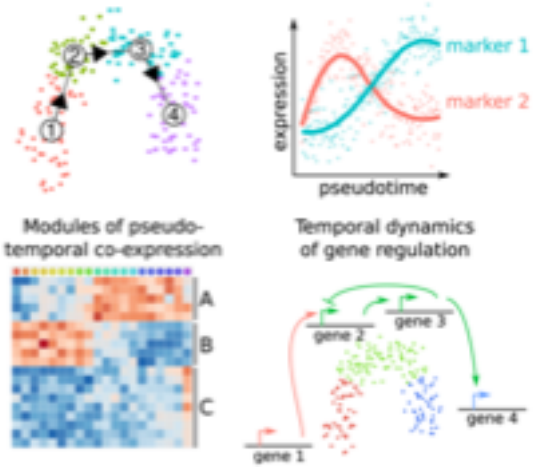
What TI offers



- Unbiased and transcriptome-wide understanding of a dynamic process



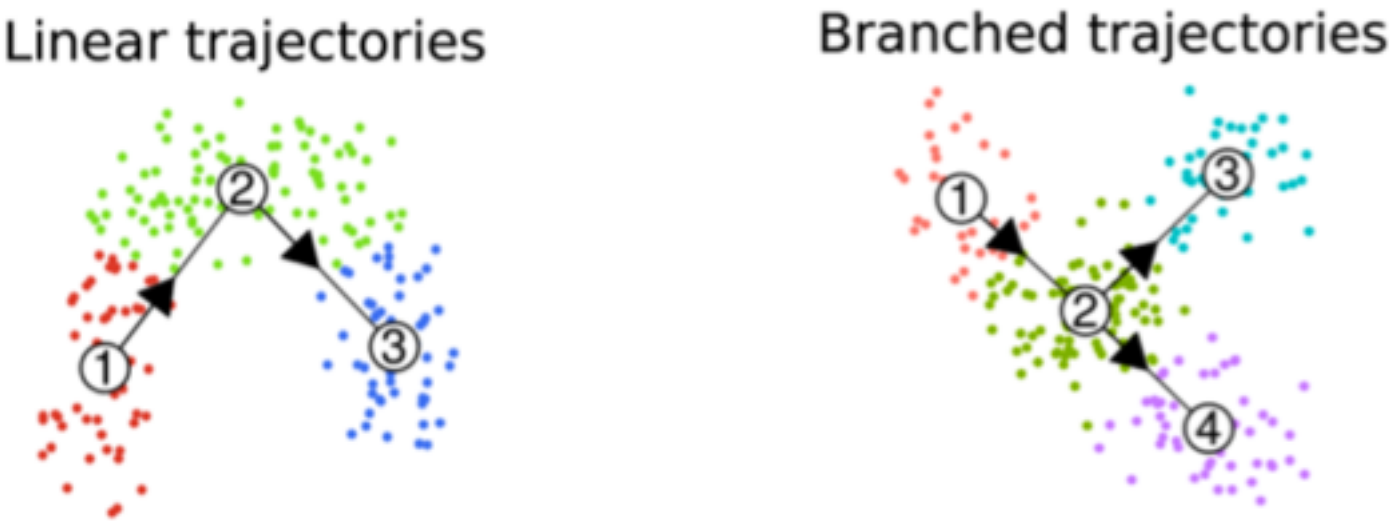
- They allow the objective identification of new subsets of cells



Cannot et al., Eur. J. Immun, 2016

Type of trajectories

Trajectory's total length: total amount of transcriptional change that a cell undergoes at 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.

Input data – potential risks

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

Method	Date	Most complex trajectory type	Fixes topology	Prior required	Prior optional	Evaluated	Reference
Monocle ICA	01/04/2014	Tree	Parameter	# branches	None	Yes	[13]
Wanderlust	24/04/2014	Linear	Fixed	Start cell(s)	None	Yes	[14]
SCUBA	30/12/2014	Tree	Free	None	Time course, Marker genes	Yes	[15]
Sincell	27/01/2015	Tree	Free	None	None	Yes	[16]
NBOR	08/06/2015	Linear	TBD	TBD	TBD	No[2]	[6]
Waterfall	03/09/2015	Linear	Fixed	None	None	Yes	[17]
gpseudotime	15/09/2015	Linear	TBD	TBD	TBD	No[2]	[18]
Embeddr	18/09/2015	Linear	Fixed	None	None	Yes	[19]
ECLAIR	12/01/2016	Tree	TBD	TBD	TBD	No[2]	[20]
DPT	08/02/2016	Bifurcation	Fixed	None	Marker genes	Yes	[21]
Pseudogp	05/04/2016	Linear	Fixed	None	None	Yes	[22]
SLICER	09/04/2016	Graph	Free	Start cell(s)	End cell(s), Marker genes	Yes	[23]
SCell	19/04/2016	Linear	TBD	TBD	TBD	No[2]	[24]
Wishbone	02/05/2016	Bifurcation	Parameter	Start cell(s), # end states	Marker genes	Yes	[25]
TSCAN	13/05/2016	Tree	Free	None	None	Yes	[26]
SCOUP	08/06/2016	Multifurcation	Parameter	Start cell(s), Cell grouping, # end states	None	Yes	[27]
DeLorean	17/06/2016	Linear	TBD	TBD	TBD	No[2]	[28]
SlingshD	21/06/2016	Tree	Free	None	None	Yes	[29]
Ouija	23/06/2016	Linear	Fixed	Marker genes	None	Yes	[30]
Mpath	30/06/2016	Tree	Free	Cell grouping	None	Yes	[31]
cellTree	13/08/2016	Tree	Free	None	Cell grouping	Yes	[32]
WaveCrest	17/08/2016	Linear	TBD	Time course	None	No[2]	[33]
SCIMITAR	04/10/2016	Linear	Fixed	None	None	Yes	[34]
SCORPIUS	07/10/2016	Linear	Fixed	None	None	Yes	[35]
SCENT	30/10/2016	Linear	TBD	TBD	TBD	No[2]	[36]
k-branches	15/12/2016	Tree	TBD	TBD	TBD	No[2]	[37]
SLICE	19/12/2016	Tree	Free	None	Cell grouping, Marker genes	Yes	[38]
Topslam	13/02/2017	Linear	Fixed	Start cell(s)	None	Yes	[39]
Monocle DDRTree	21/02/2017	Tree	Free	None	# end states	Yes	[40]
Granatum	22/02/2017	Tree	TBD	TBD	TBD	No[2]	[41]
GPfates	03/03/2017	Multifurcation	Parameter	# end states	None	Yes	[42]
MFA	15/03/2017	Multifurcation	Parameter	# end states	None	Yes	[43]
PHATE	24/03/2017	Tree	TBD	TBD	TBD	No[2]	[44]
TASIC	04/04/2017	Tree	TBD	TBD	TBD	No[2,4]	[45]
SOMSC	05/04/2017	Tree	TBD	TBD	TBD	No[2]	[46]
Slingshot	19/04/2017	Tree	Free	None	Start cell(s), End cell(s)	Yes	[47]
scTDA	01/05/2017	Linear	TBD	TBD	TBD	No[2]	[48]
UNCURL	31/05/2017	Linear	TBD	TBD	TBD	No[2]	[49]
reCAT	19/06/2017	Cycle	Fixed	None	None	Yes	[50]
FORKS	20/06/2017	Tree	TBD	Start cell(s)	None	No[1]	[51]
MATCHER	24/06/2017	Linear	TBD	TBD	TBD	No[2]	[52]
PhenoPath	06/07/2017	Linear	Fixed	None	None	Yes	[53]
HopLand	12/07/2017	Linear	TBD	TBD	TBD	No[2]	[54]
SoptSC	26/07/2017	Linear	TBD	Start cell(s)	None	No[2]	[55]
PBA	30/07/2017	Multifurcation	TBD	TBD	TBD	No[2]	[56]
BGP	01/08/2017	Bifurcation	TBD	TBD	TBD	No[2]	[57]
scanpy	09/08/2017	Bifurcation	TBD	TBD	TBD	No[2]	[58]
B-RGPs	01/09/2017	Acyclic graph	TBD	TBD	TBD	No[2]	[59]
WADDINGTON-OT	27/09/2017	Graph	TBD	TBD	TBD	No[2,3]	[60]
AGA	27/10/2017	Disconnected graph	TBD	TBD	TBD	No[2]	[61]
GPseudoRank	30/10/2017	Linear	TBD	TBD	TBD	No[2]	[62]
p-Creode	15/11/2017	Tree	TBD	TBD	TBD	No[2]	[63]
iCyclic	30/11/2017	Linear	TBD	TBD	TBD	No[2]	[64]
GrandPrix	03/12/2017	Multifurcation	TBD	Time course	None	No[2]	[65]
Topographer	21/01/2018	Tree	TBD	None	Start cell(s)	No[2]	[66]
CALISTA	31/01/2018	Graph	TBD	None	None	No[2]	[67]
scEpath	05/02/2018	Tree	TBD	TBD	TBD	No[2]	[68]
MERLoT	08/02/2018	Tree	TBD	TBD	TBD	No[2]	[69]
ElPiGraph.R	04/03/2018	Graph	TBD	TBD	TBD	No[2]	

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

Monocle 2

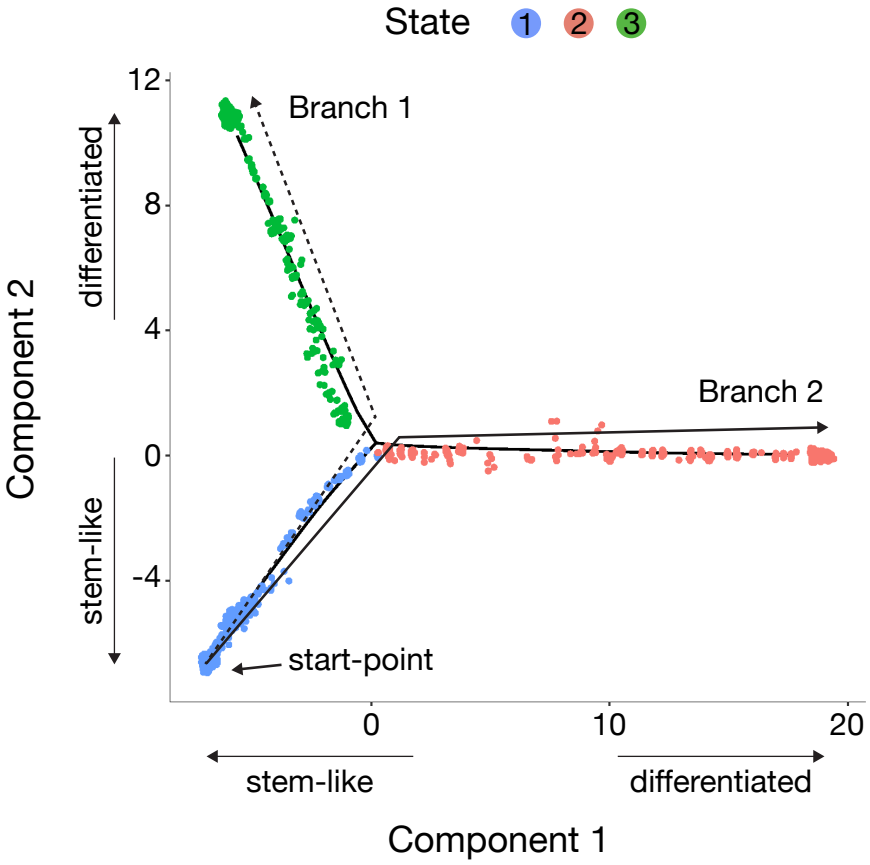
Trajectory inference workflow:

1. Choosing genes to order the data

2. Reducing dimensionality of the data

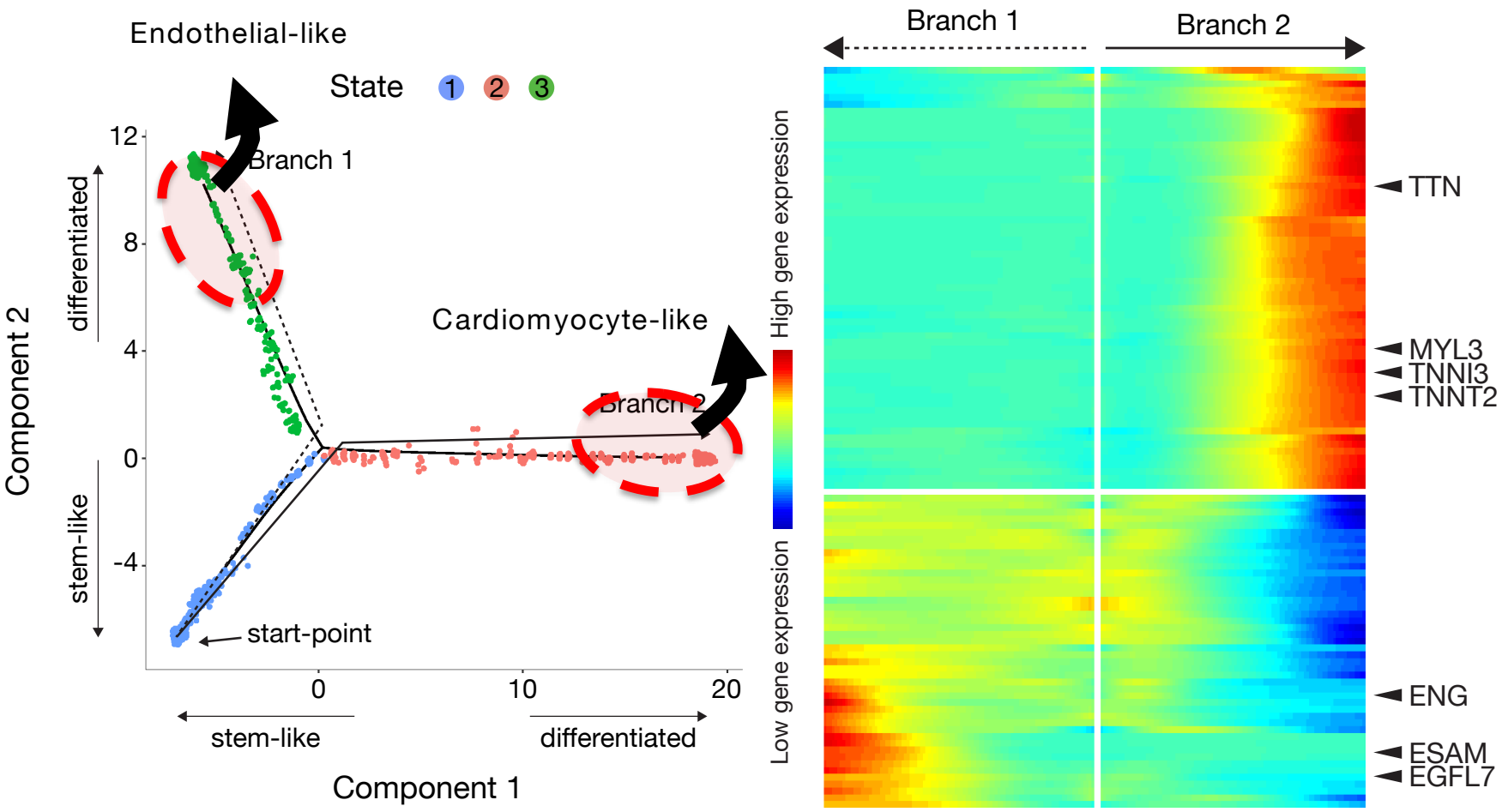
3. Ordering cells in pseudotime

Fates of human fetal heart cells



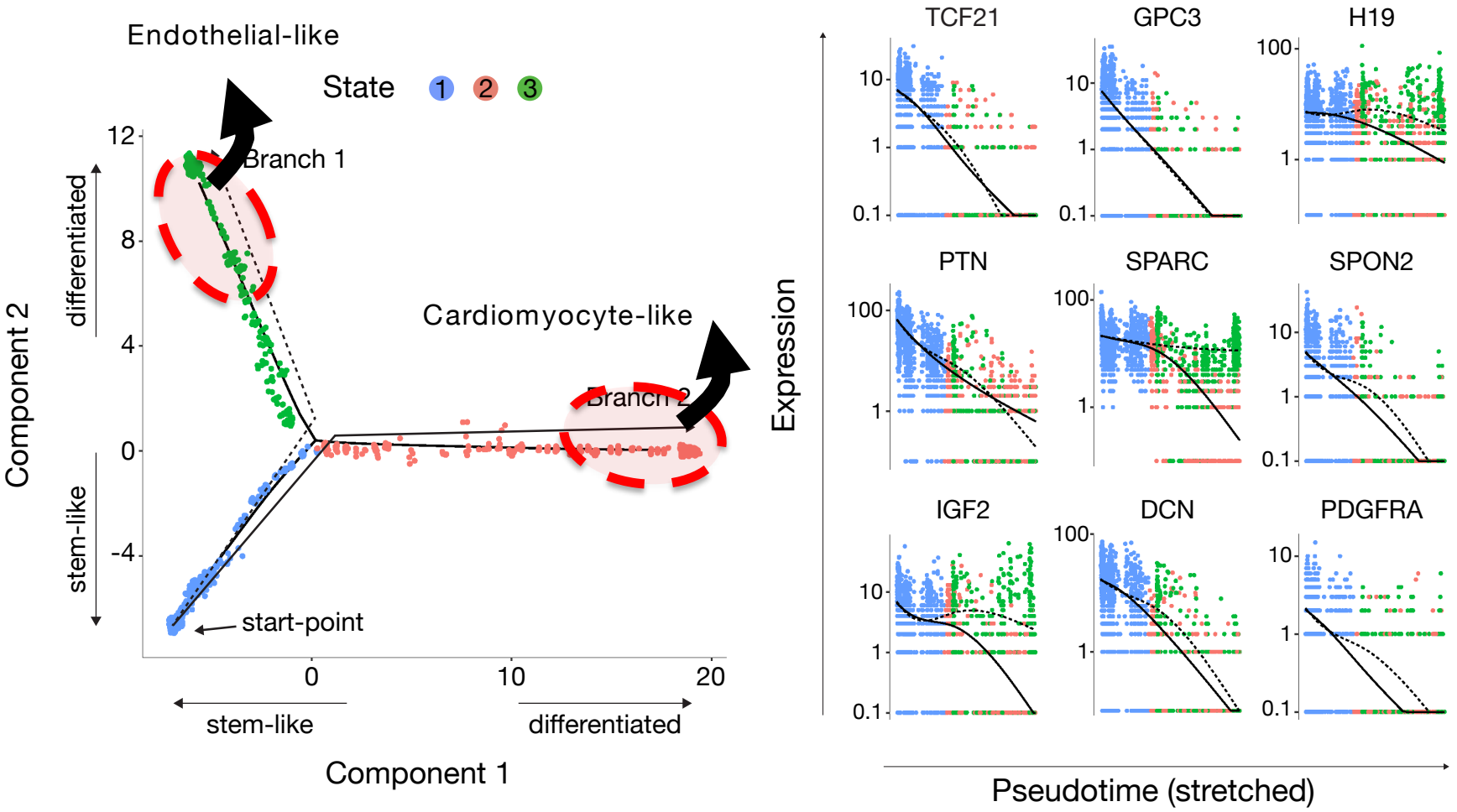
SciLifeLab

Fates of human fetal heart cells



SciLifeLab

Fates of human fetal heart cells



SciLifeLab