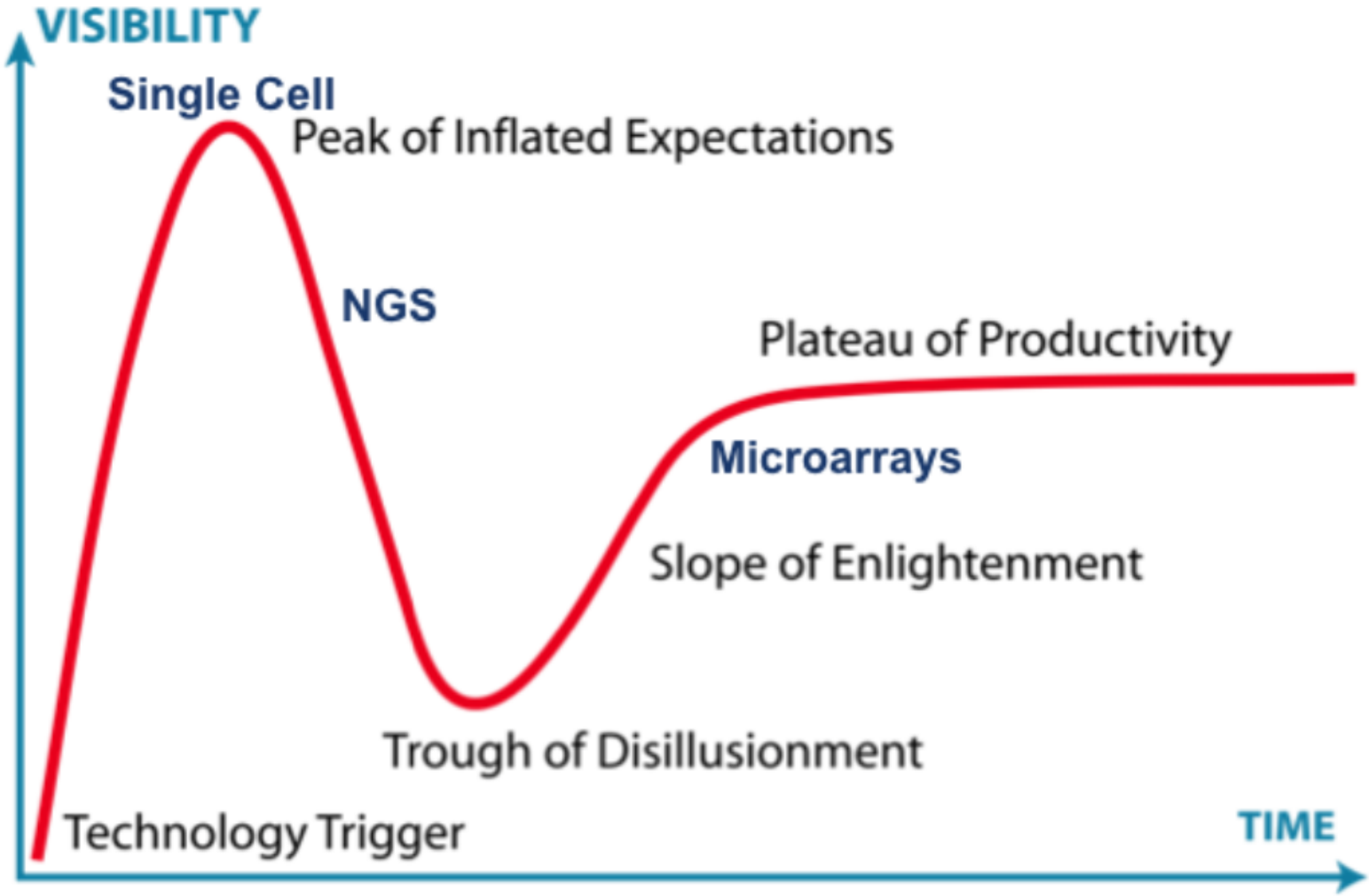
Single-cell transcriptomics overview

Alejandro Reyes
Huber group
CSAMA 2015



Rafael A Irizarry @rafalab · Feb 15

@Y_Gilad @joe_pickrell @davisjmcc @leonidkruglyak Are you not familiar with the "hype cycle" ?

Yoav Gilad @Y_Gilad · Feb 13

I guess single-cell data are so exciting that we all momentarily forgot everything we knew about study design, modeling, and multiple tests?

39 45

NATURE BIOTECHNOLOGY | OPINION AND COMMENT | CORRESPONDENCE

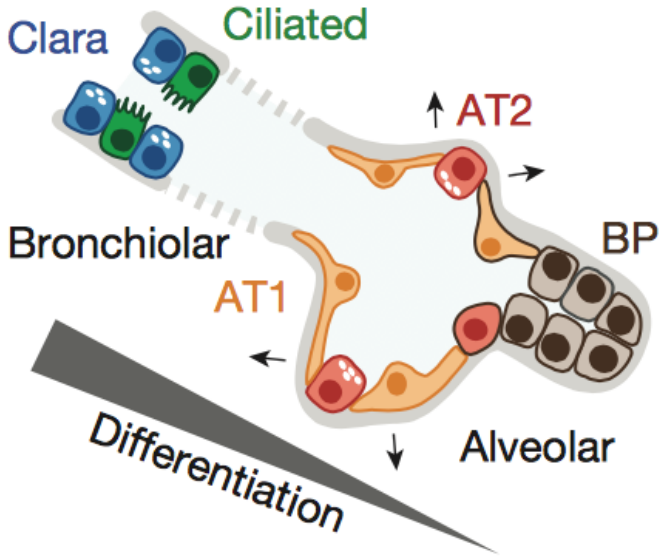
Sequencing technology does not eliminate biological variability

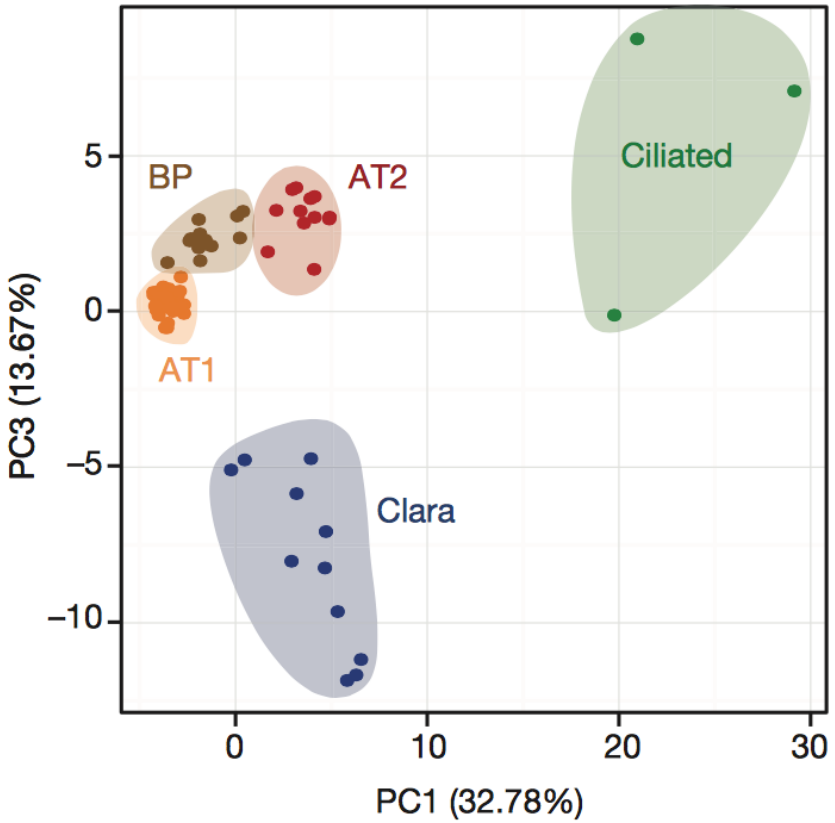
Kasper D Hansen, Zhijin Wu, Rafael A Irizarry & Jeffrey T Leek

Why?
(optimistic)

Research questions that can be answered with single-cell transcriptomics

1) Identification of cell-types from heterogeneous samples

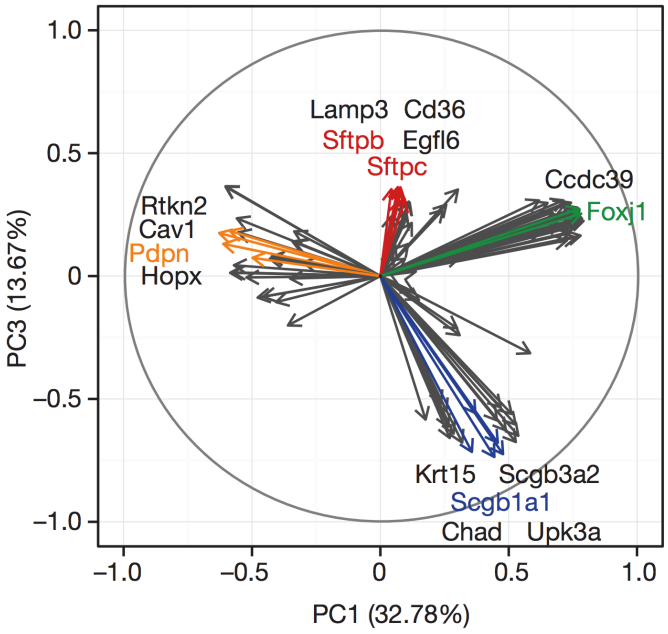


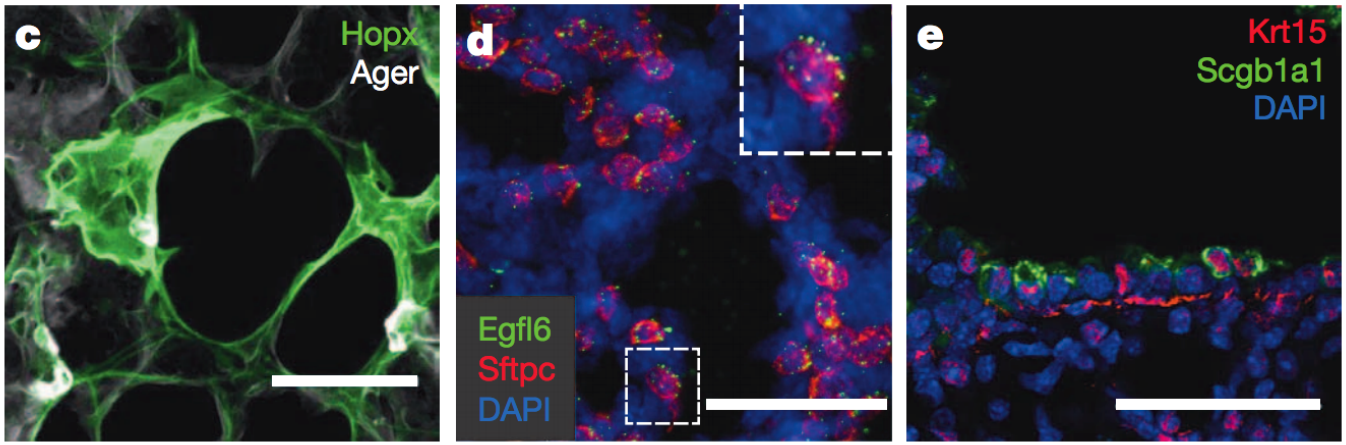


Treutlein et al, 2014

Research questions that can be answered with single-cell transcriptomics

2) Identification of cell-type specific markers (genes, isoforms)

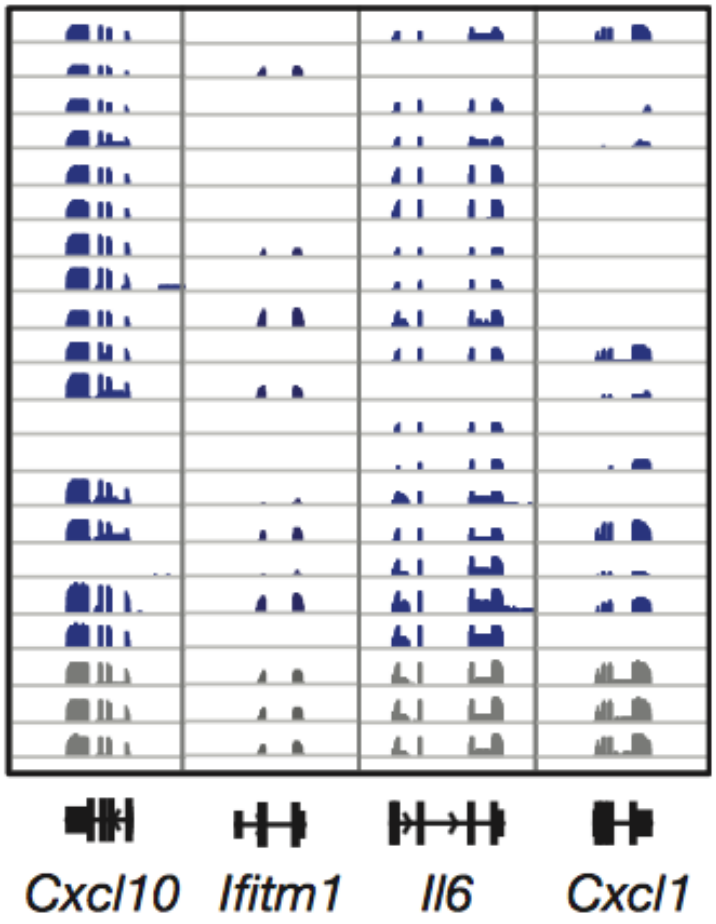


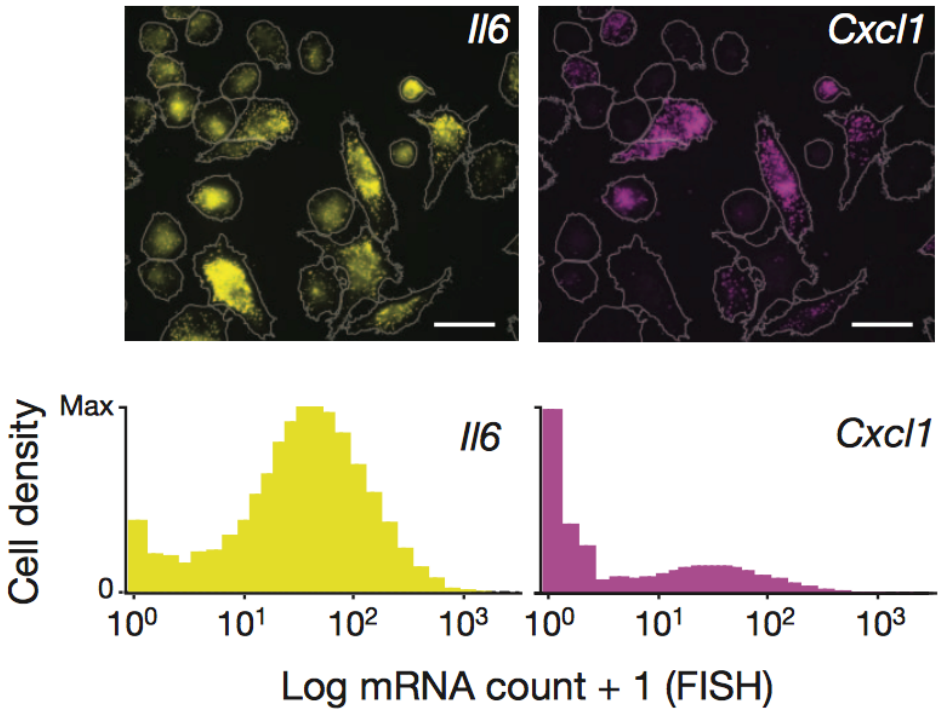


Treutlein et al, 2014

Research questions that can be answered with single-cell transcriptomics

3) Identifying highly varying genes across cells

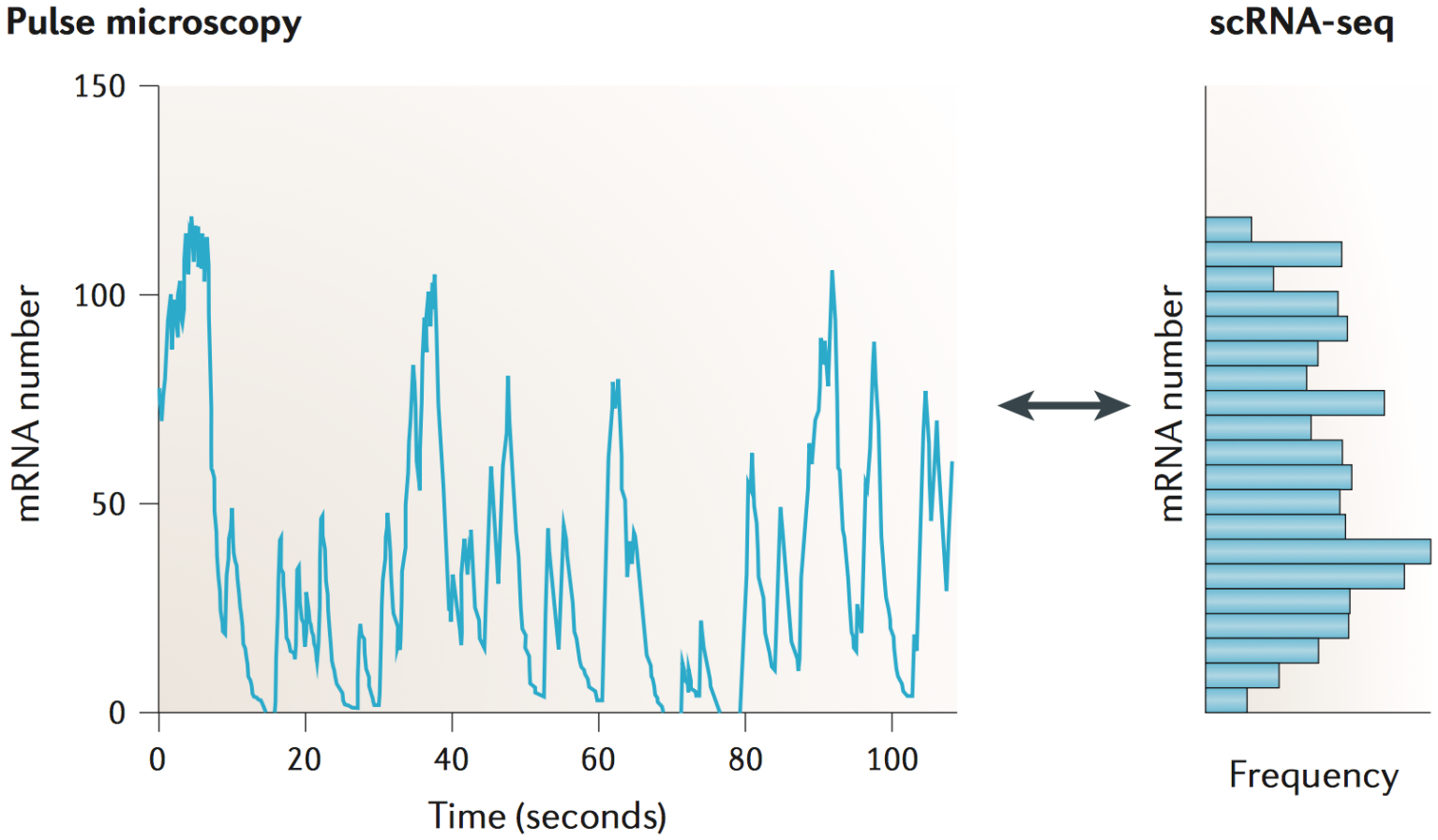




Shalek et al, 2013

Research questions that can be answered with single-cell transcriptomics

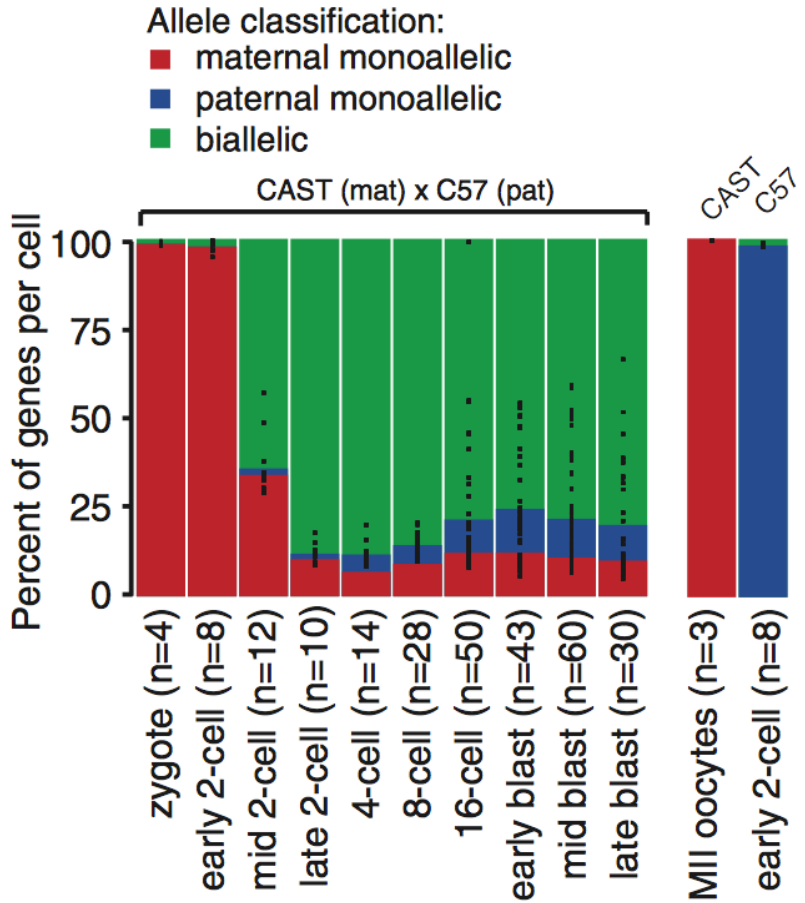
4) Study kinetics of transcription



Stegle et al, 2013

Research questions that can be answered with single-cell transcriptomics

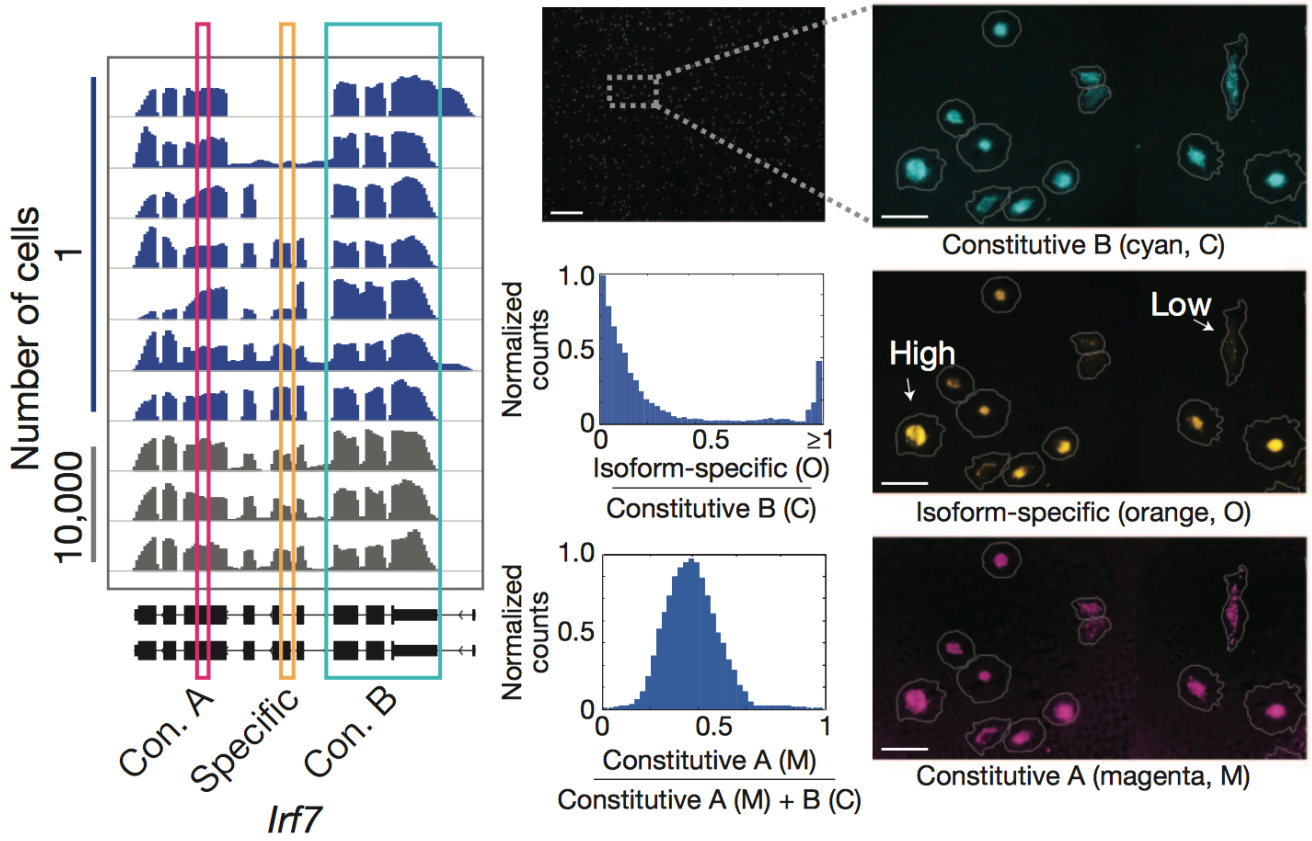
5) Allelic expression heterogeneity



Deng et al, 2014

Research questions that can be answered with single-cell transcriptomics

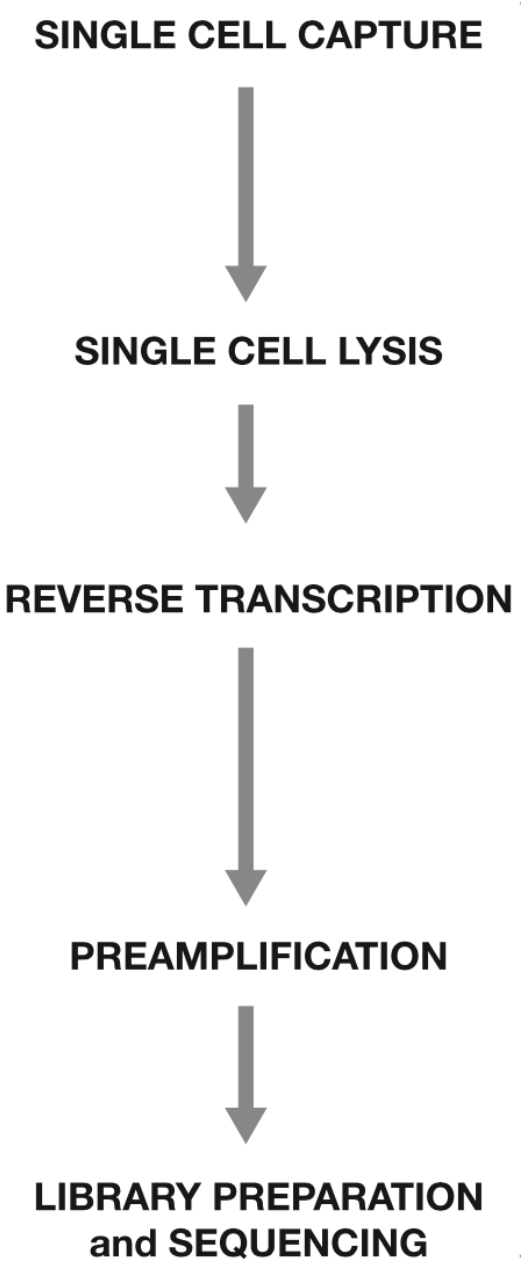
5) Transcript isoform expression heterogeneity



Shalek et al, 2013

Protocols and noise (pesimistic)

Single-cell transcriptomics protocols overview



Kolodziejczyk et al, 2013

Single-cell transcriptomics protocols overview

SINGLE CELL CAPTURE

↓

SINGLE CELL LYSIS

↓

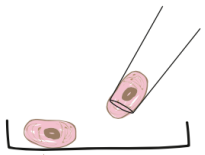
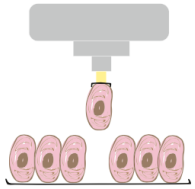
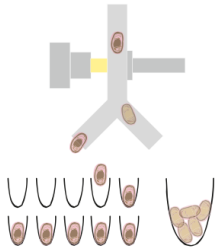
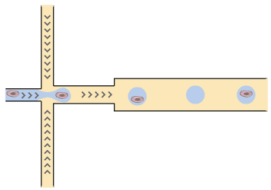
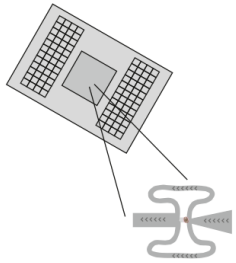
REVERSE TRANSCRIPTION

↓

PREAMPLIFICATION

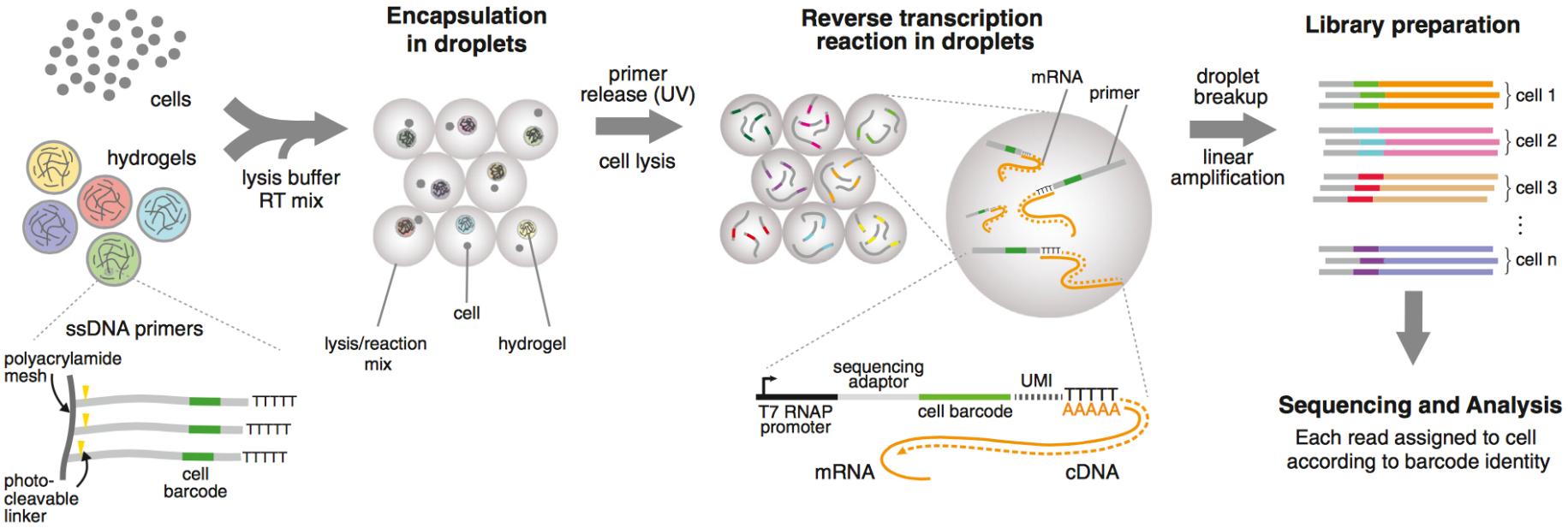
↓

LIBRARY PREPARATION and SEQUENCING

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

Kolodziejczyk et al, 2013

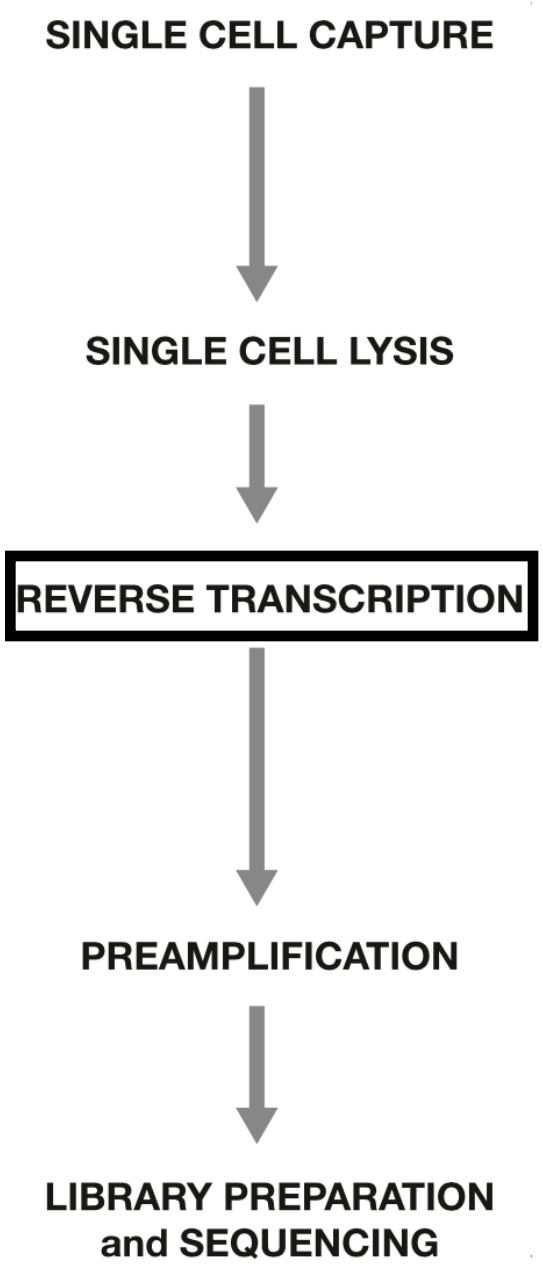
Single cell protocols



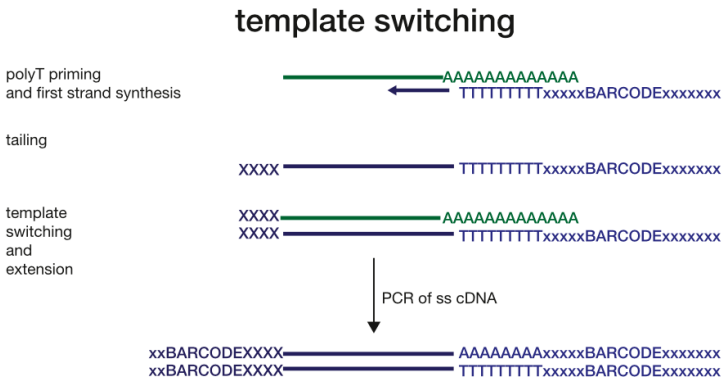
Thousands of cells!

Klein et al, 2015

Single-cell transcriptomics protocols overview



Tang protocol (Tang et al 2009)
CELseq/MARSseq (Hashimony et al. 2013, Jaitin et al. 2014)
QuartzSeq (Sasagawa et al. 2013)



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

Kolodziejczyk et al, 2013

Single-cell transcriptomics protocols overview

SINGLE CELL CAPTURE

SINGLE CELL LYSIS

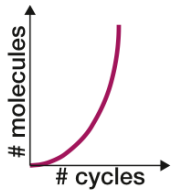
REVERSE TRANSCRIPTION

PREAMPLIFICATION

LIBRARY PREPARATION and SEQUENCING

PCR

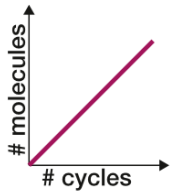
- exponential amplification
- PCR base specific biases



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)

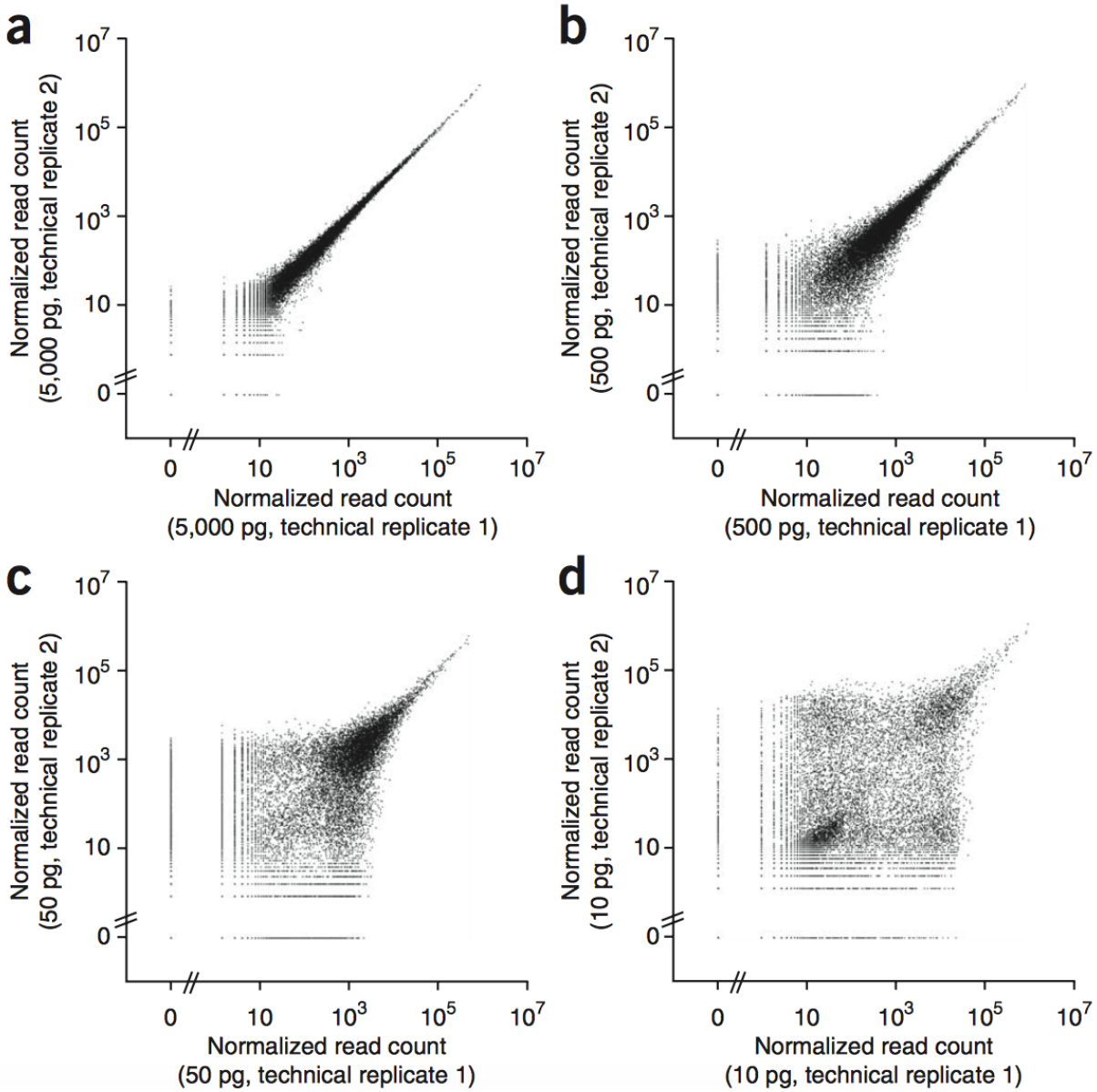
Kolodziejczyk et al, 2013

Analysis

Observed read counts are a combination of different factors

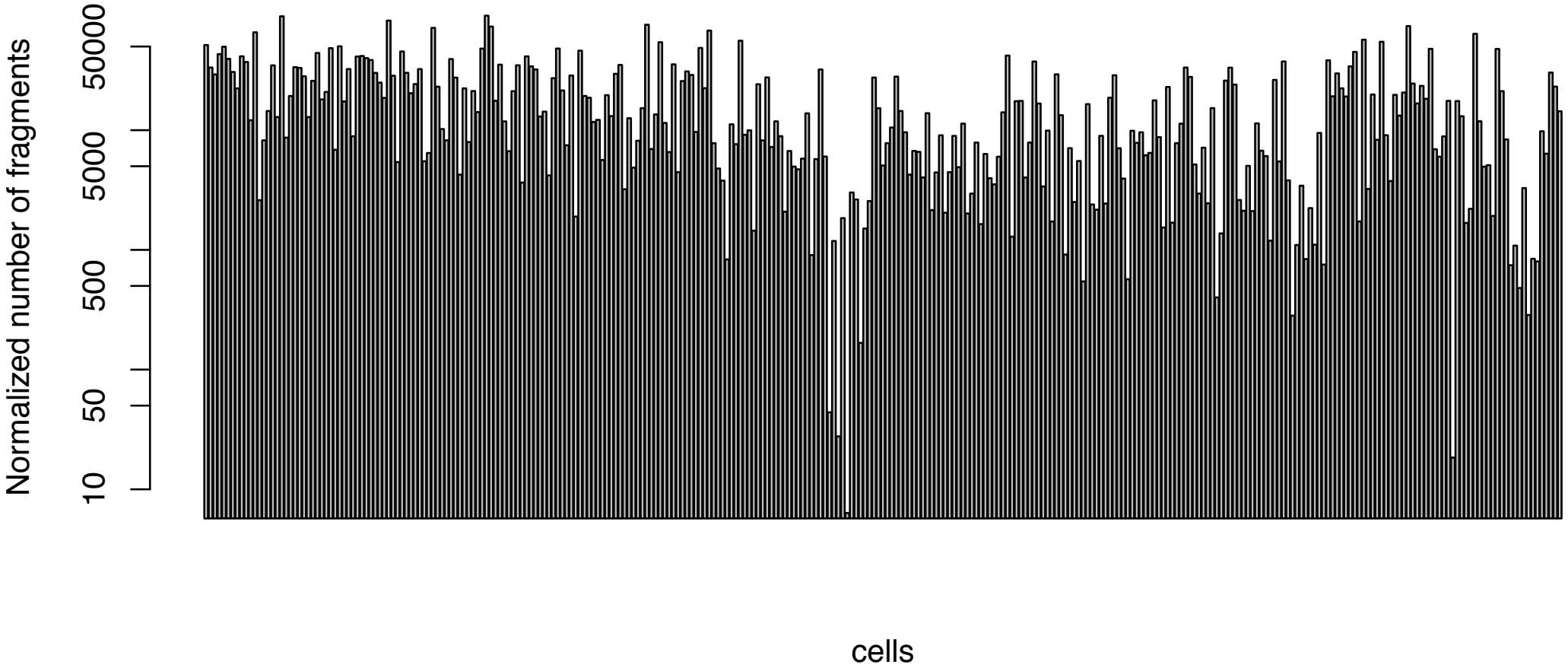
counts = cell state + cell cycle + cell size +
apoptosis + …+ technical noise

Small amounts of starting material impact on technical noise

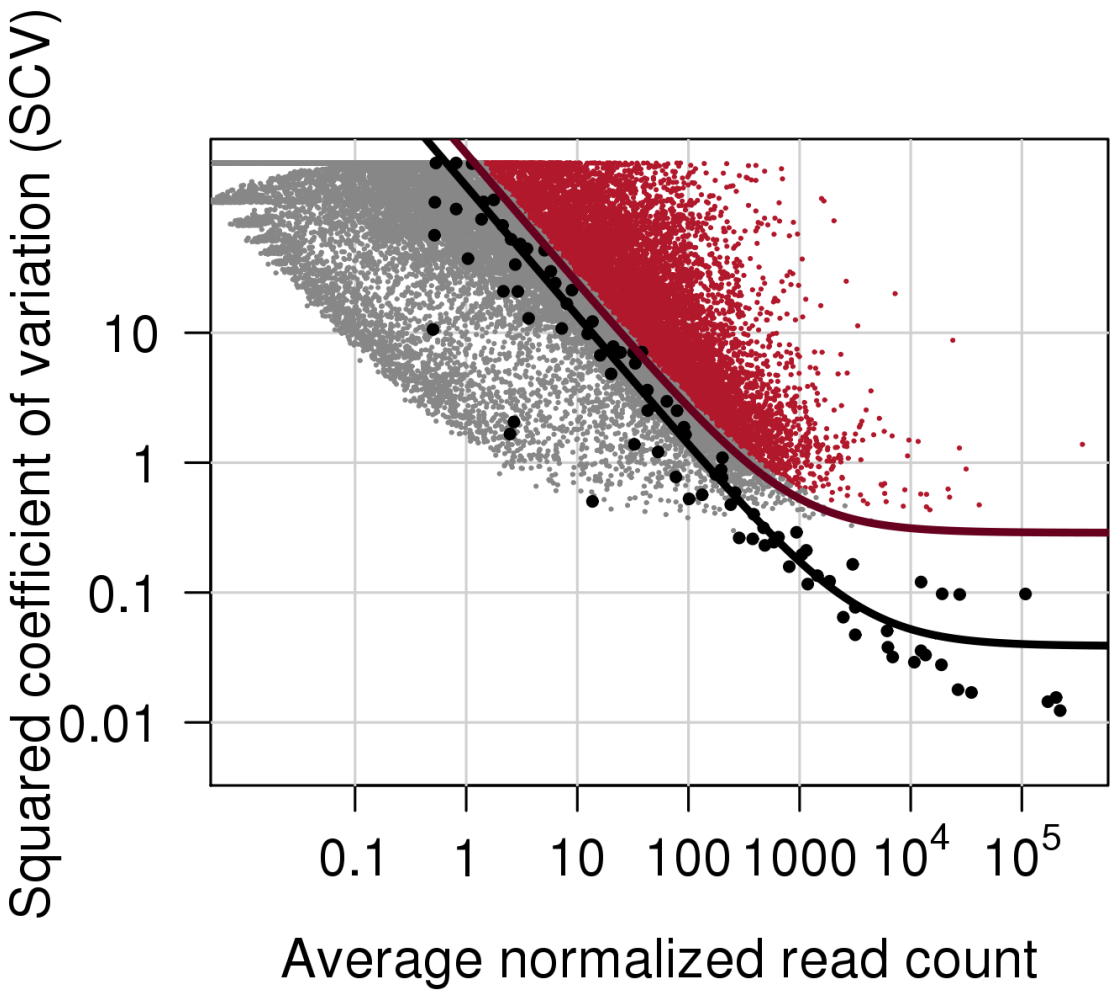


Brennecke, Anders et al, 2013

Detection problems

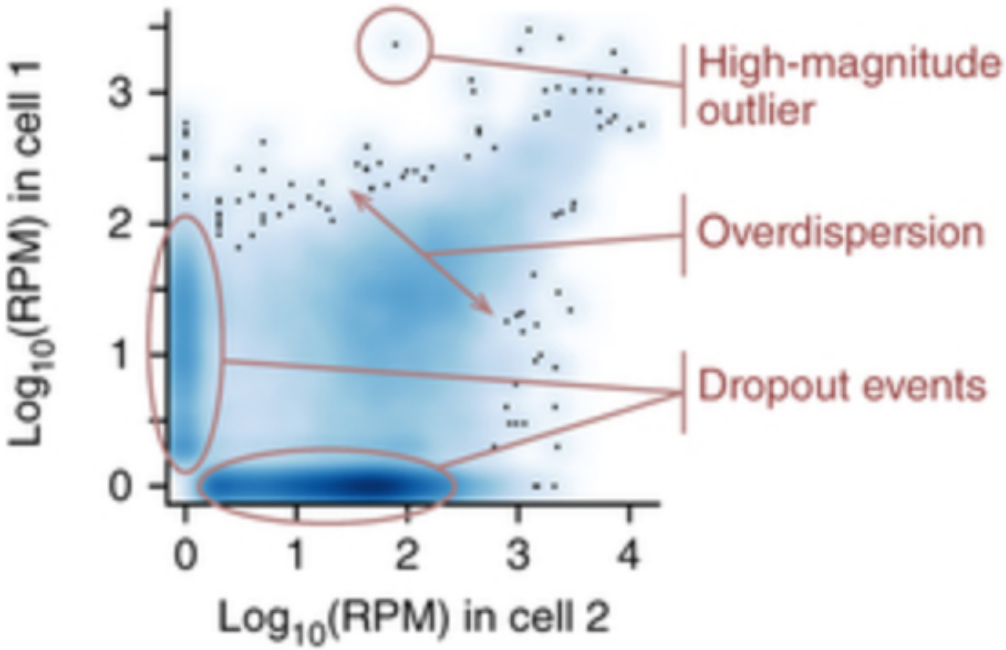


Accounting for technical noise using spike-in sequences



Method by Brennecke, Anders et al, 2013
Data from Brennecke, Reyes et al, 2015

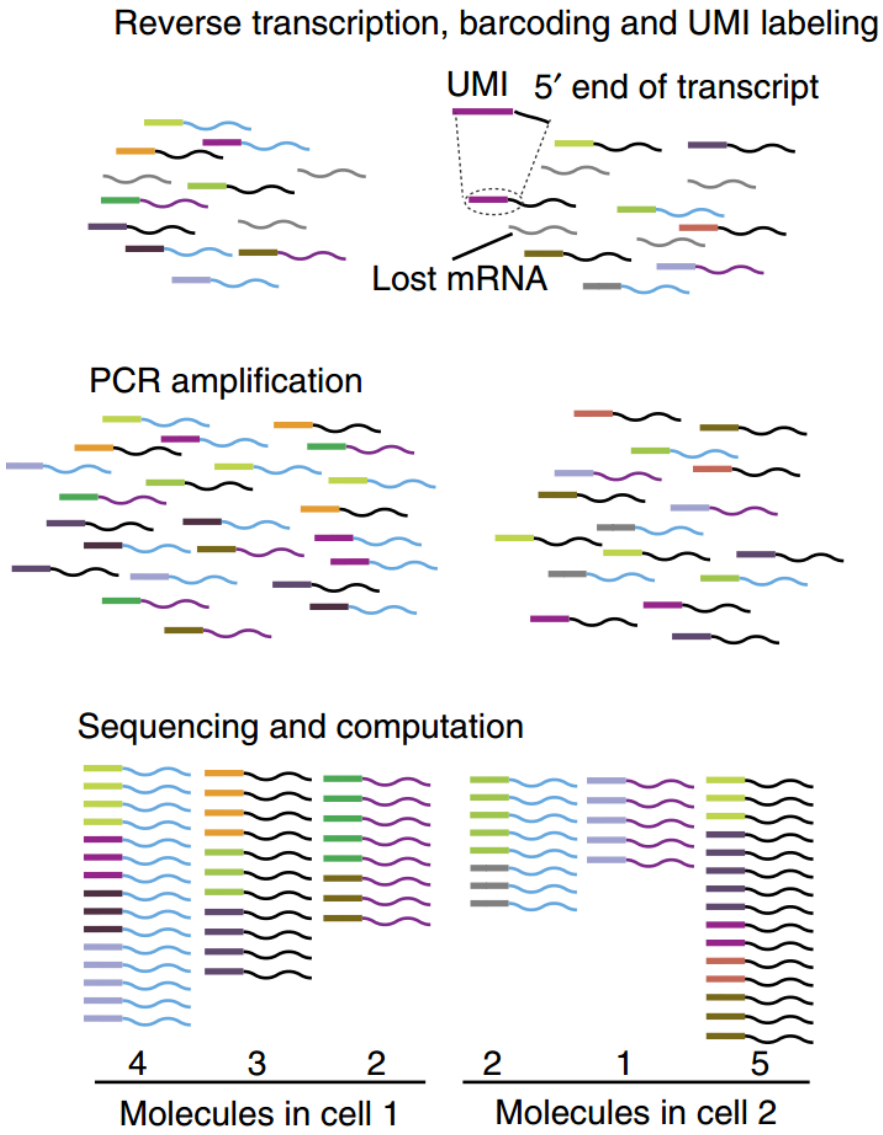
Accounting for technical noise by considering "dropout" events



$$\begin{cases} r_1 \approx Poisson(\lambda_0) & \text{Dropout in } c_1 \\ \begin{cases} r_1 \approx NB(r_2) \\ r_2 \approx NB(r_1) \end{cases} & \text{Amplified} \\ r_2 \approx Poisson(\lambda_0) & \text{Dropout in } c_2 \end{cases}$$

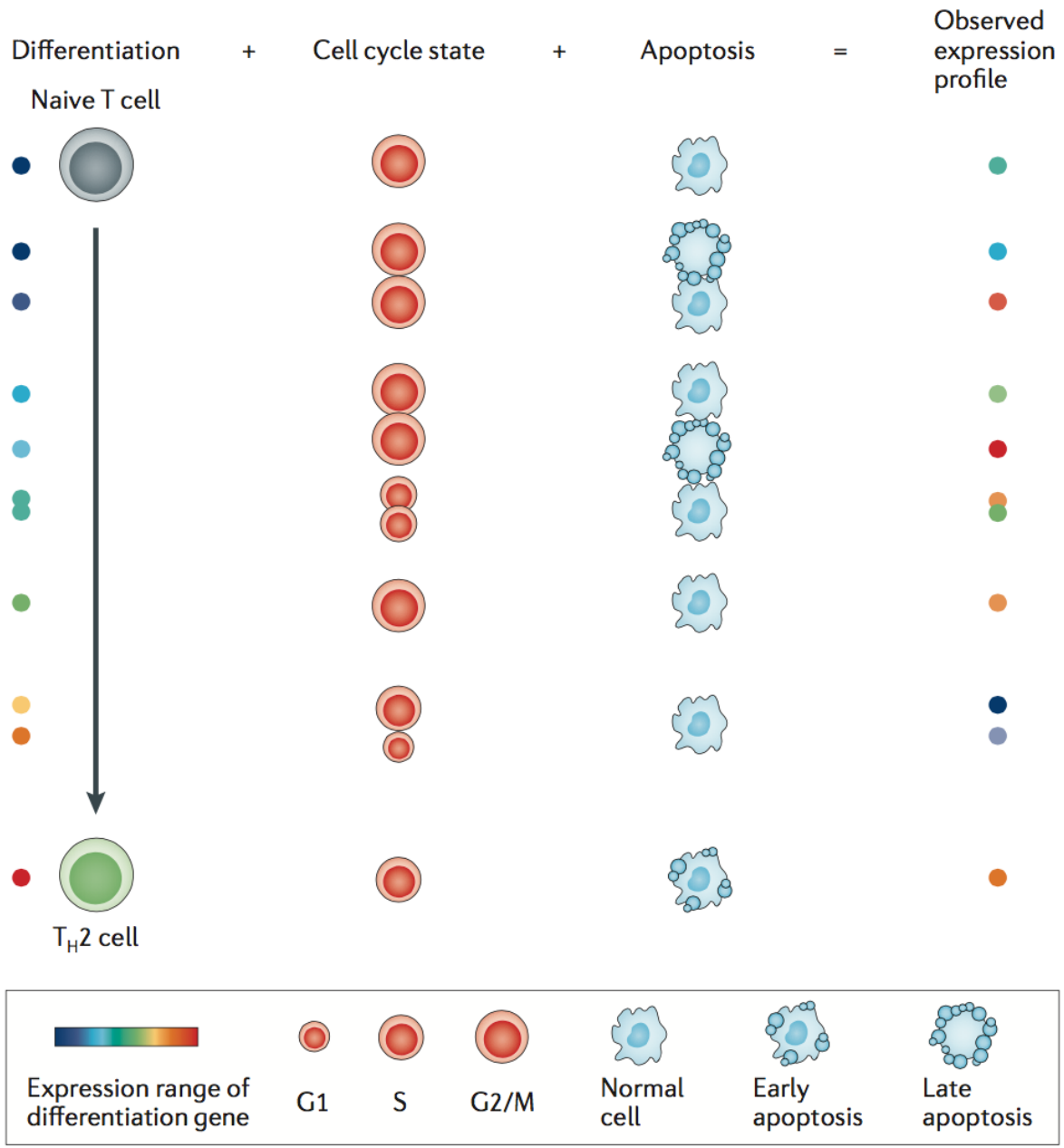
Kharchenko et al, 2013

Accounting for technical noise using unique molecular identifiers



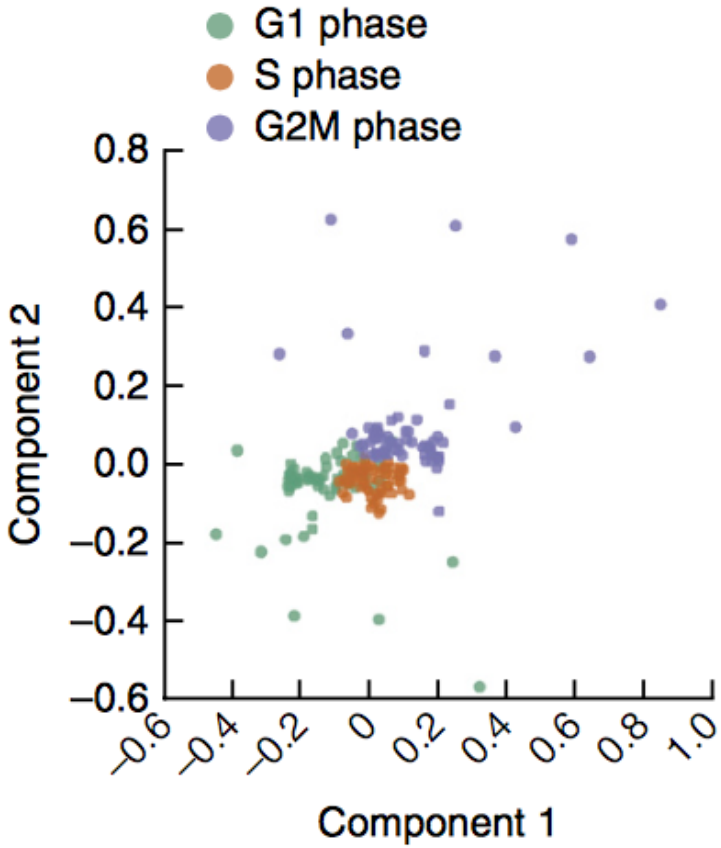
Islam et al, 2014

Accounting for “biological” confounders



Stegle et al, 2015

scLVM is useful to regress out variation explained by latent variables



Buettner et al, 2014

scLVM is useful to regress out variation explained by latent variables

Given H hidden factors,

$$y_g \sim \mathcal{N}\left(\mu_g \mathbf{1}, \sum_{h=1}^{H} \sigma_{gh}^2 \Sigma_h + \nu_g^2 \mathbf{I} + \delta_g^2 \mathbf{I}\right)$$

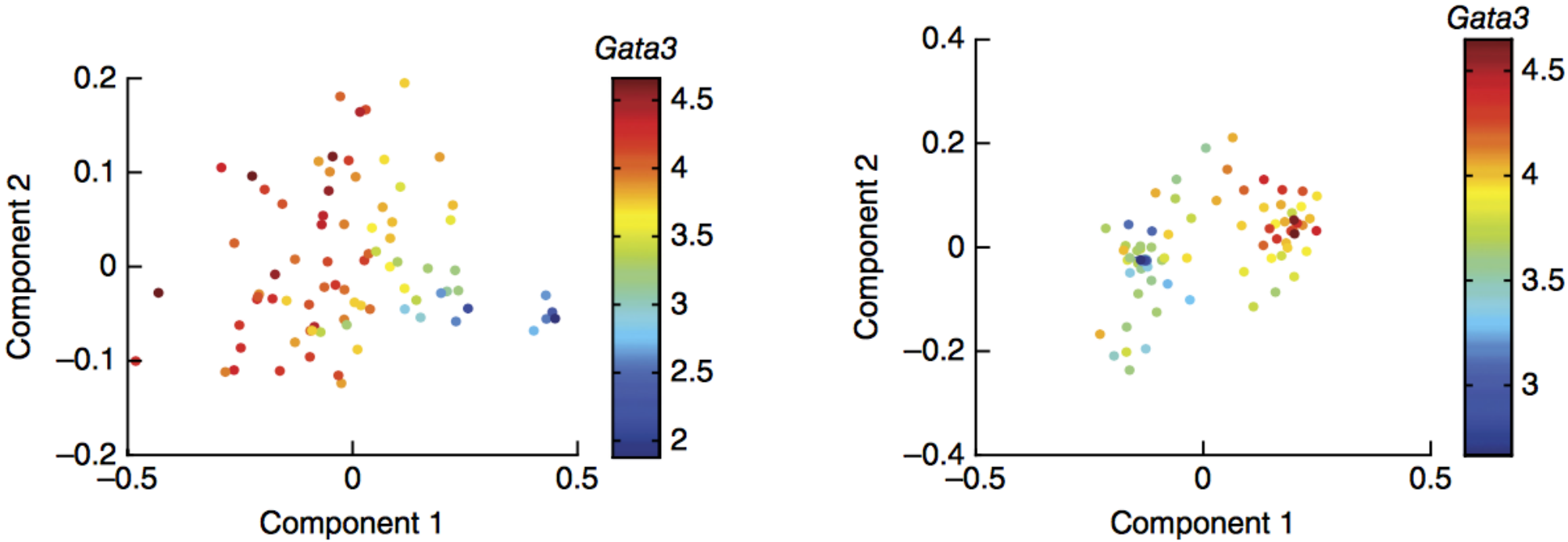
Observed expression for gene g

Variance attributed to hidden factors

Residual "biological" variance

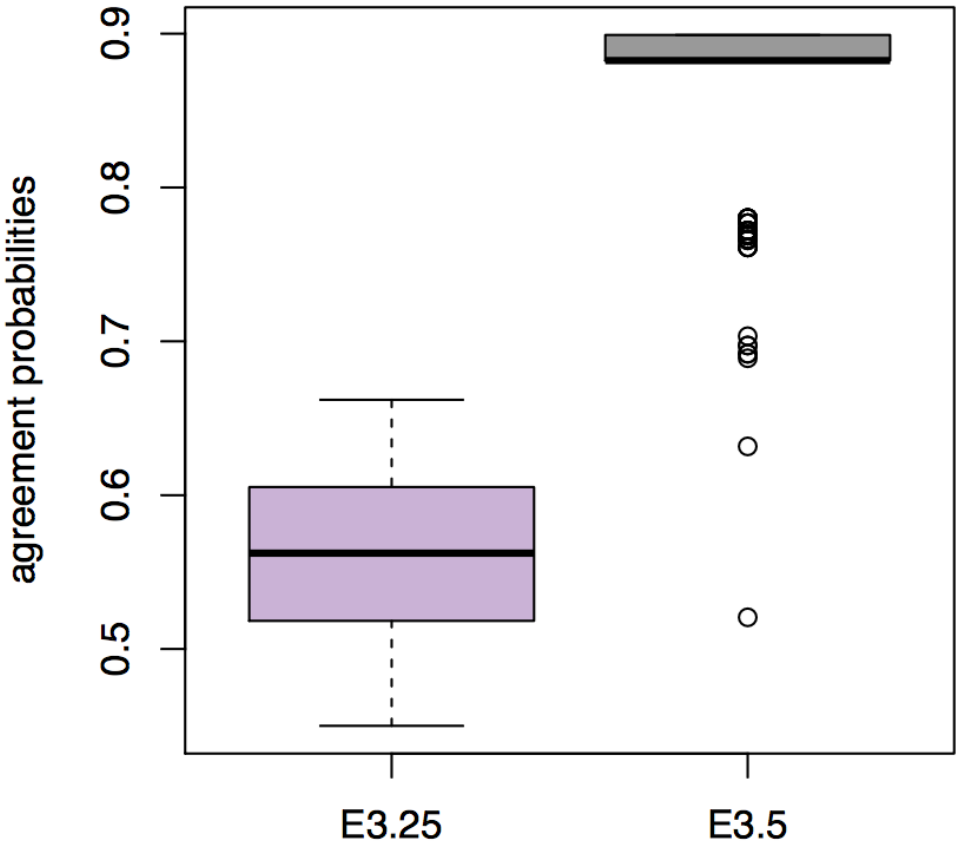
Technical variance

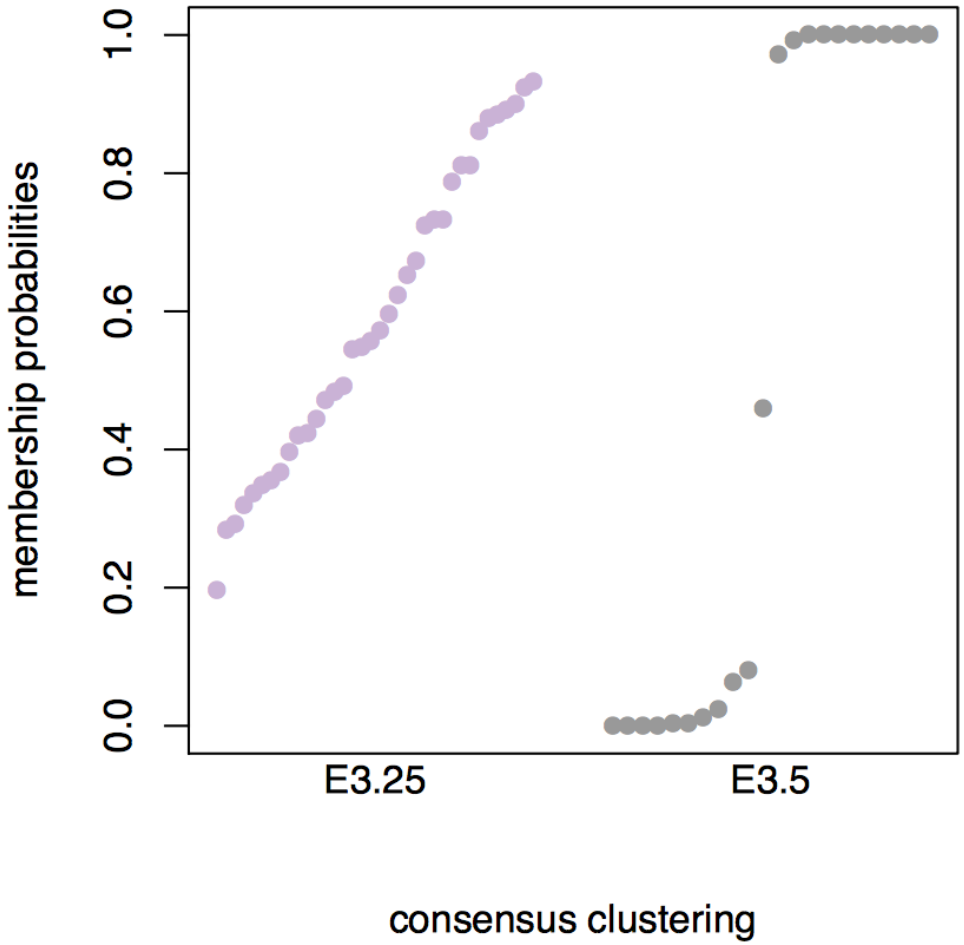
scLVM is useful to regress out variation explained by latent variables



Buettner et al, 2015

Cluster stability analysis





Ohnishi, Huber, et al, 2013

Dimensionality reduction

Multidimensional scaling*
Isomap*
t-SNE*
Diffusion maps

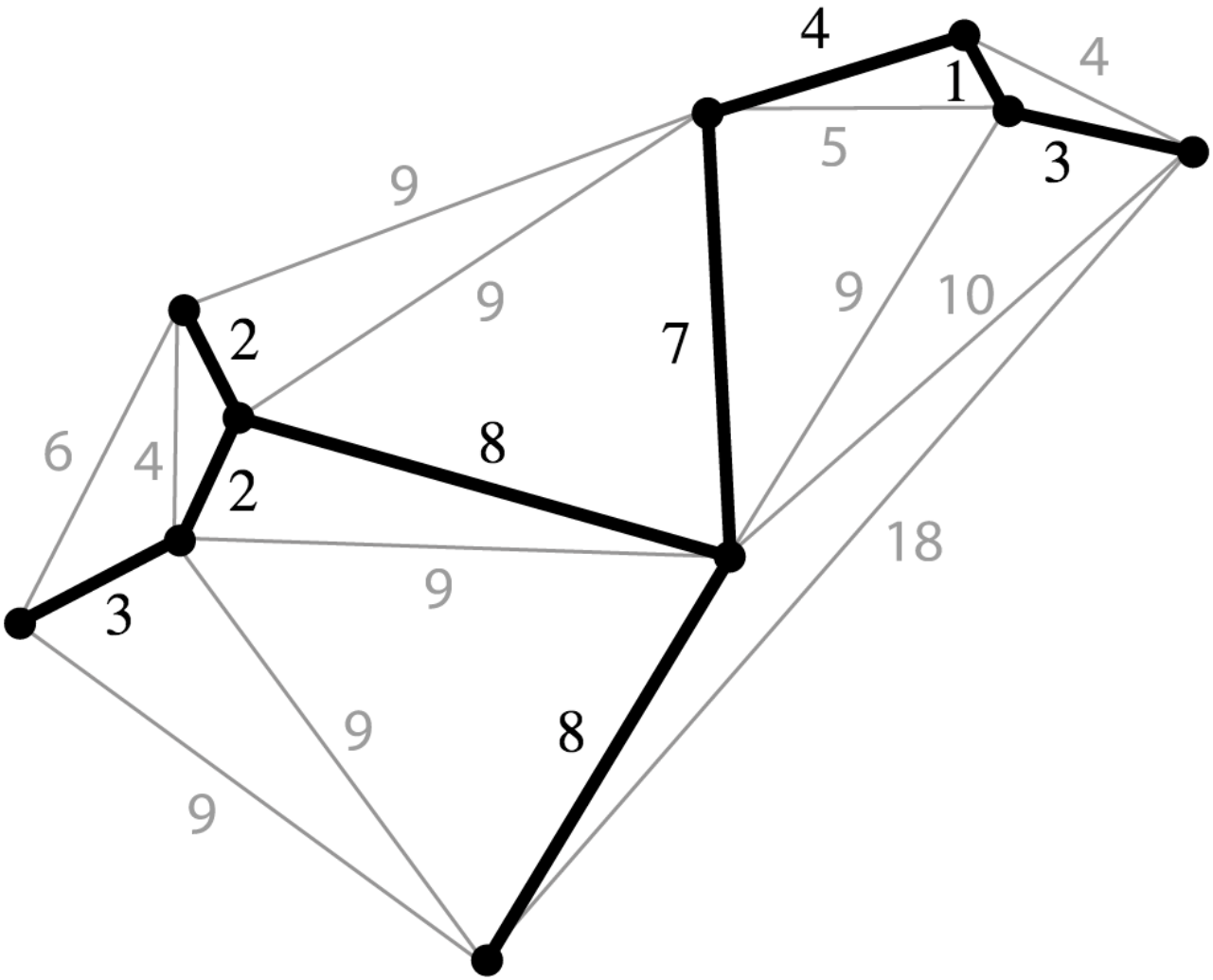
*Bioc Package: sincell

Cell hierarchy reconstruction

Minimum Spanning Tree (MST)*
Maximum Similarity Spanning Tree (SST)*
Iterative Mutual Clustering Graph (IMC)*
Wanderlust

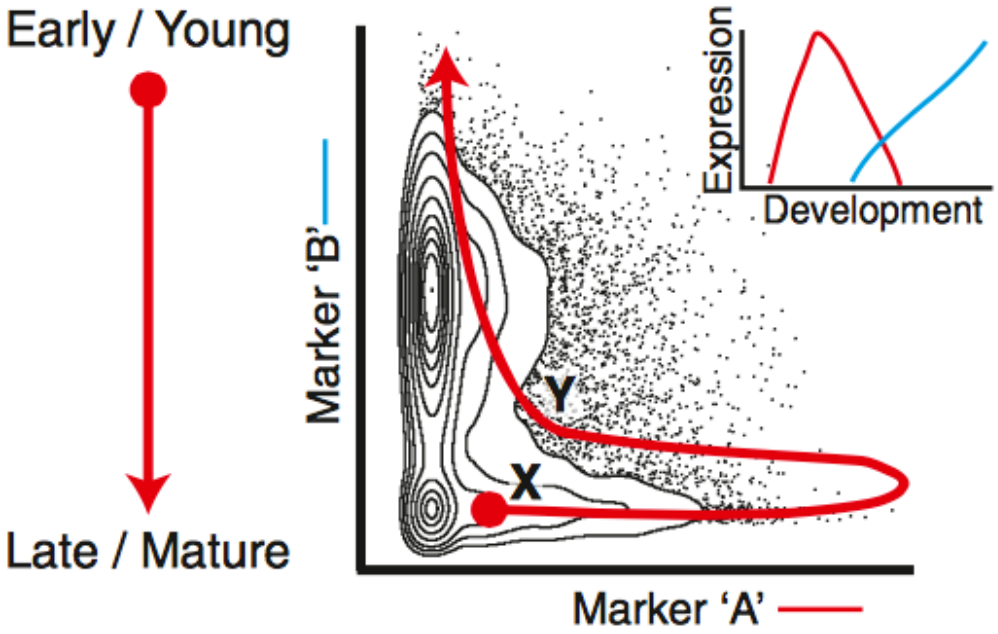
*Bioc Package: sincell

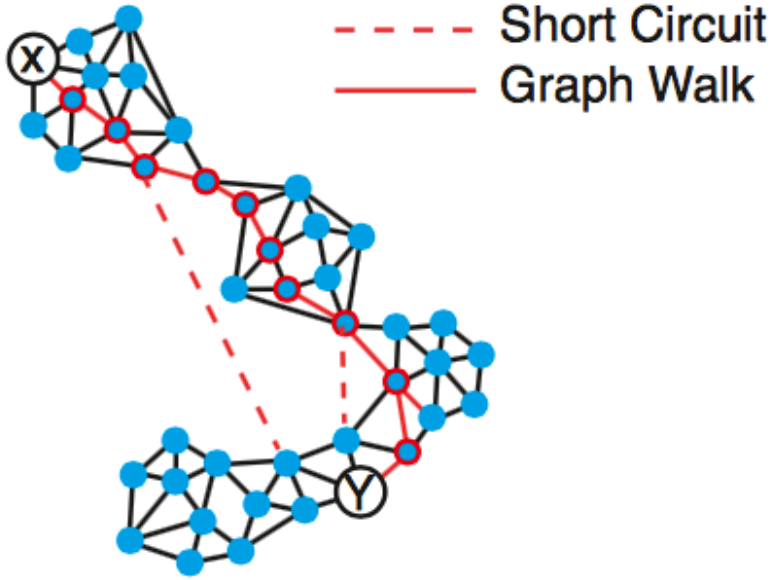
Construction of cell state hierarchies



Minimum Spanning Tree Concept (wikipedia)

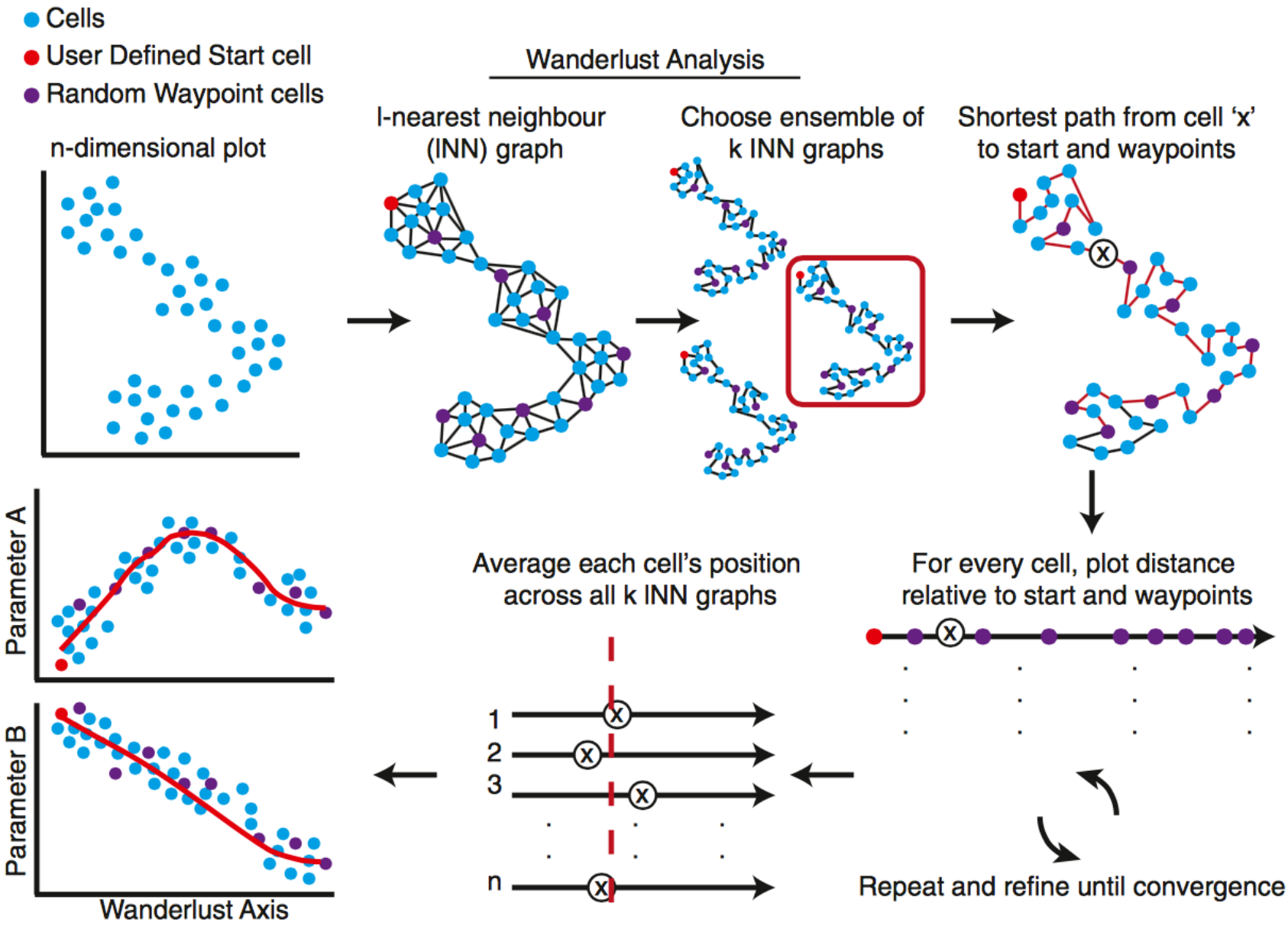
Construction of cell state hierarchies





Bendall et al, 2014

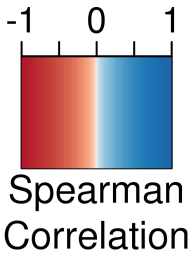
Wanderlust algorithm

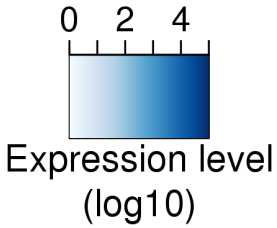


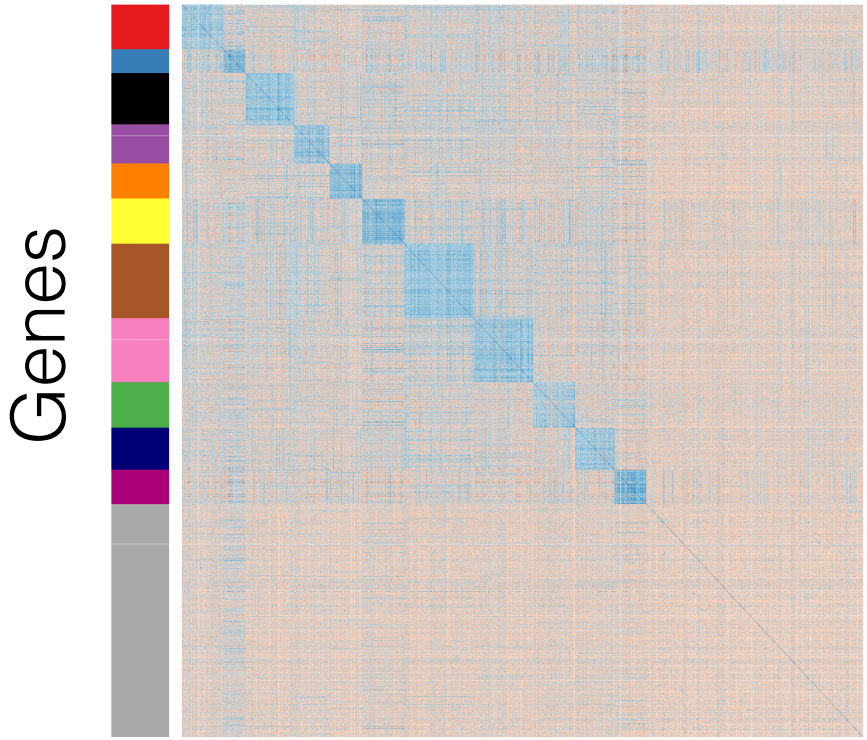
Bendall et al, 2014

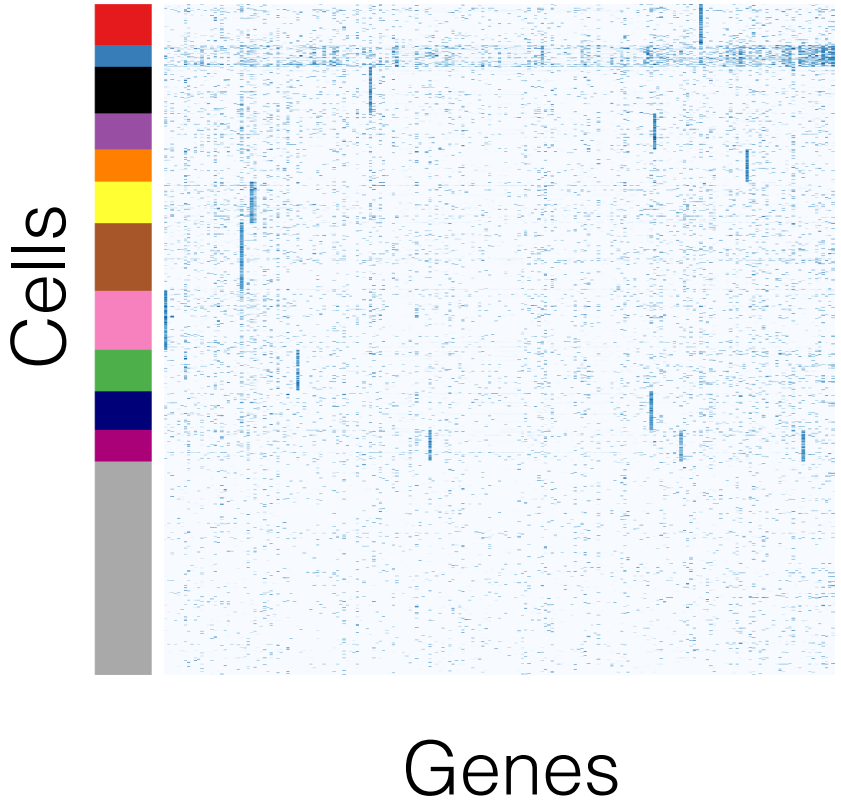
Validate!

K-medoids clustering suggested non-random gene expression patterns

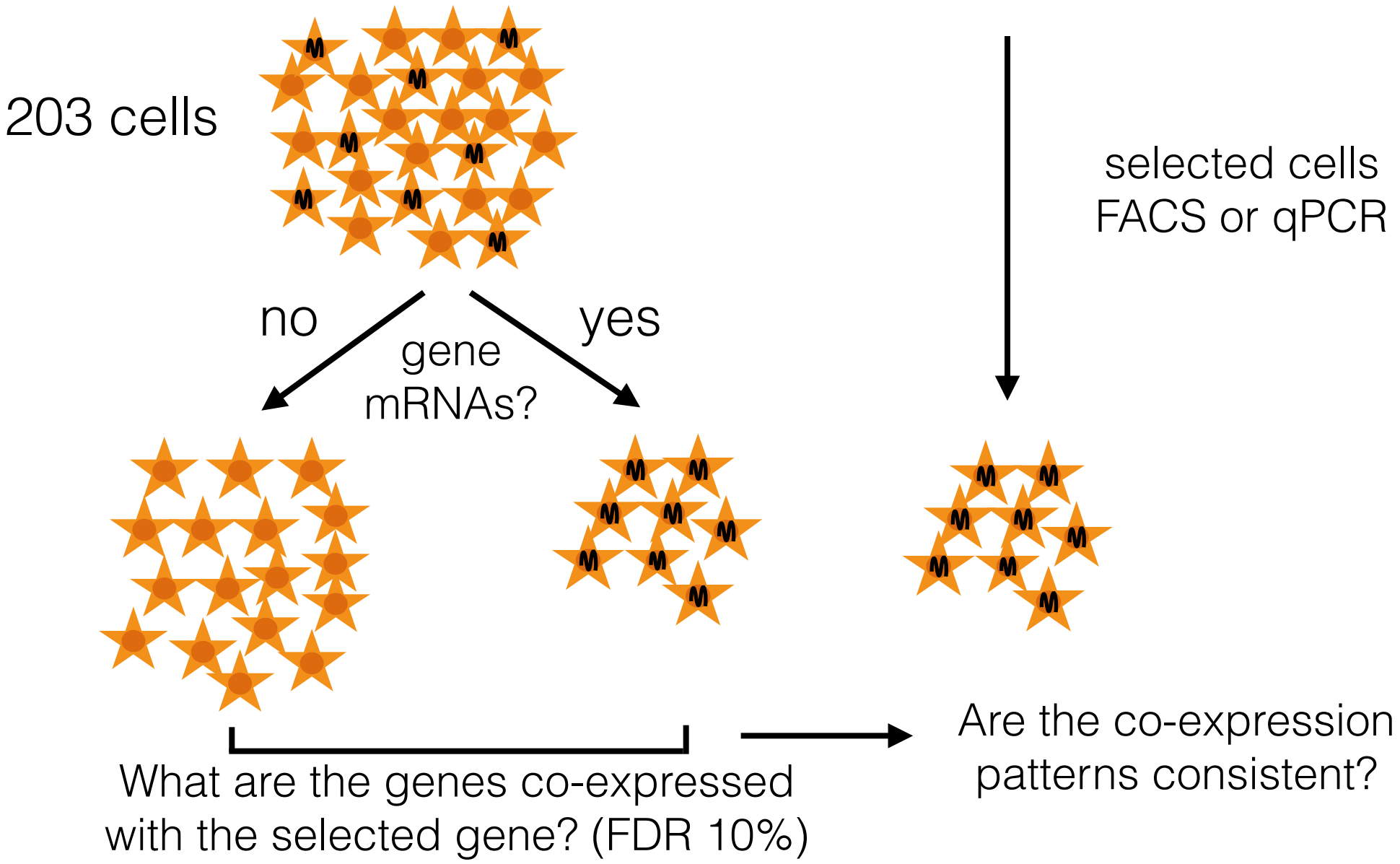




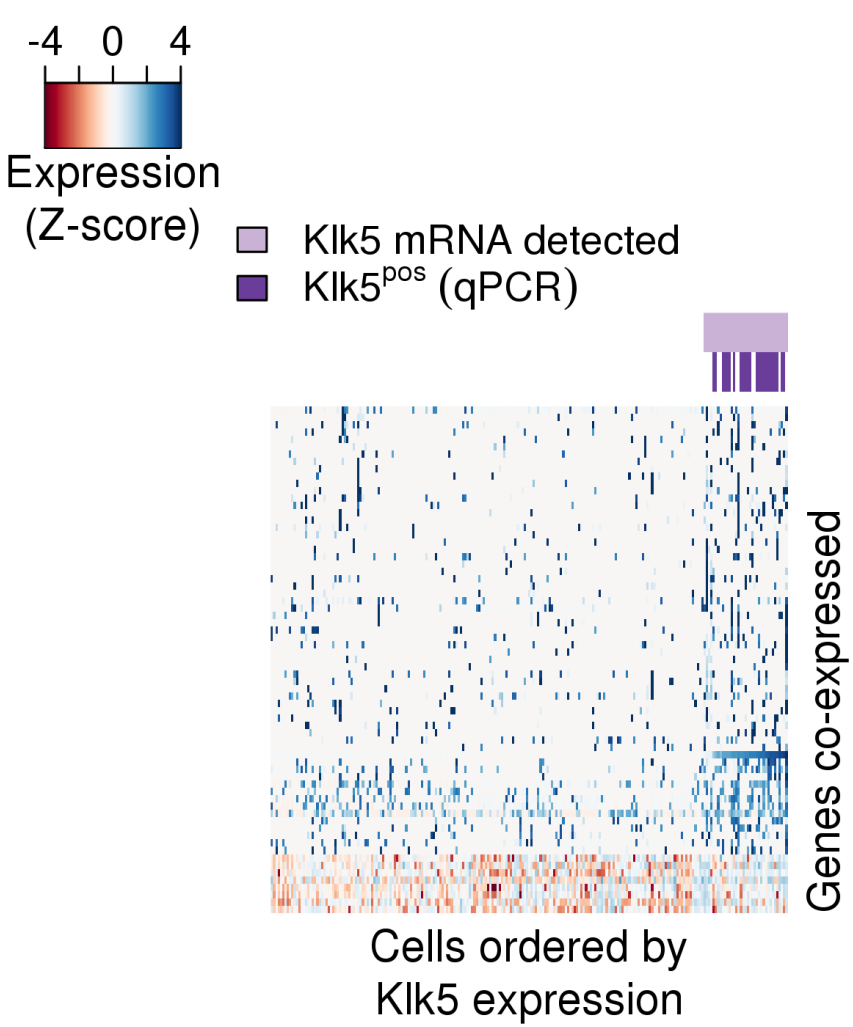


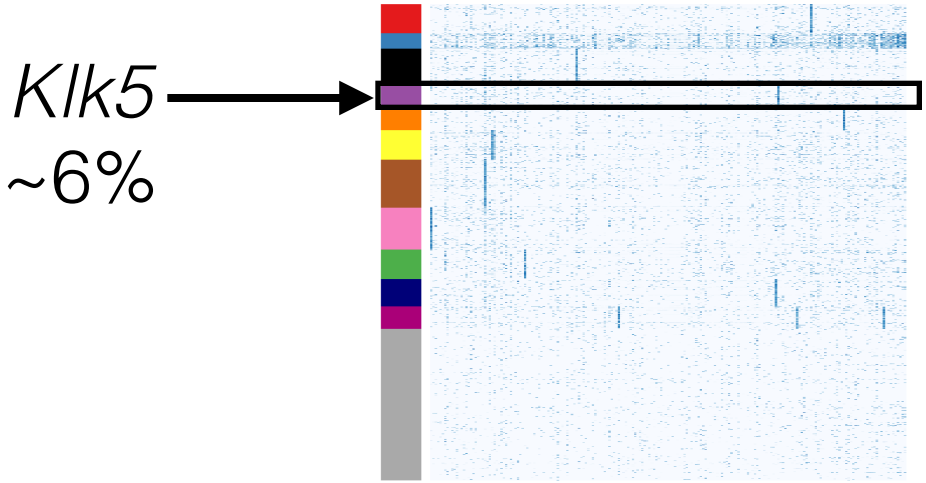


Co-expression of genes was confirmed using independent analytical and experimental validations



Co-expression of genes was confirmed using independent analytical and experimental validations





Thanks!