Pseudo time & Q&A/wrap up

Thomas D. Otto

BIOL5177

Content

1. What is pseudo time
2. Examples
3. Dive into one methods – Monocle
4. Is there a best method?

5. If you have any questions about the exercises/lecture – please add questions to menti
6. After this lecture, an example of how the COVID comparison could be done.

pseudotime analysis

Adapted from Lisa Hopcroft

Pseudotime in MVD

Snapshot in time

Pseudotime in MVD

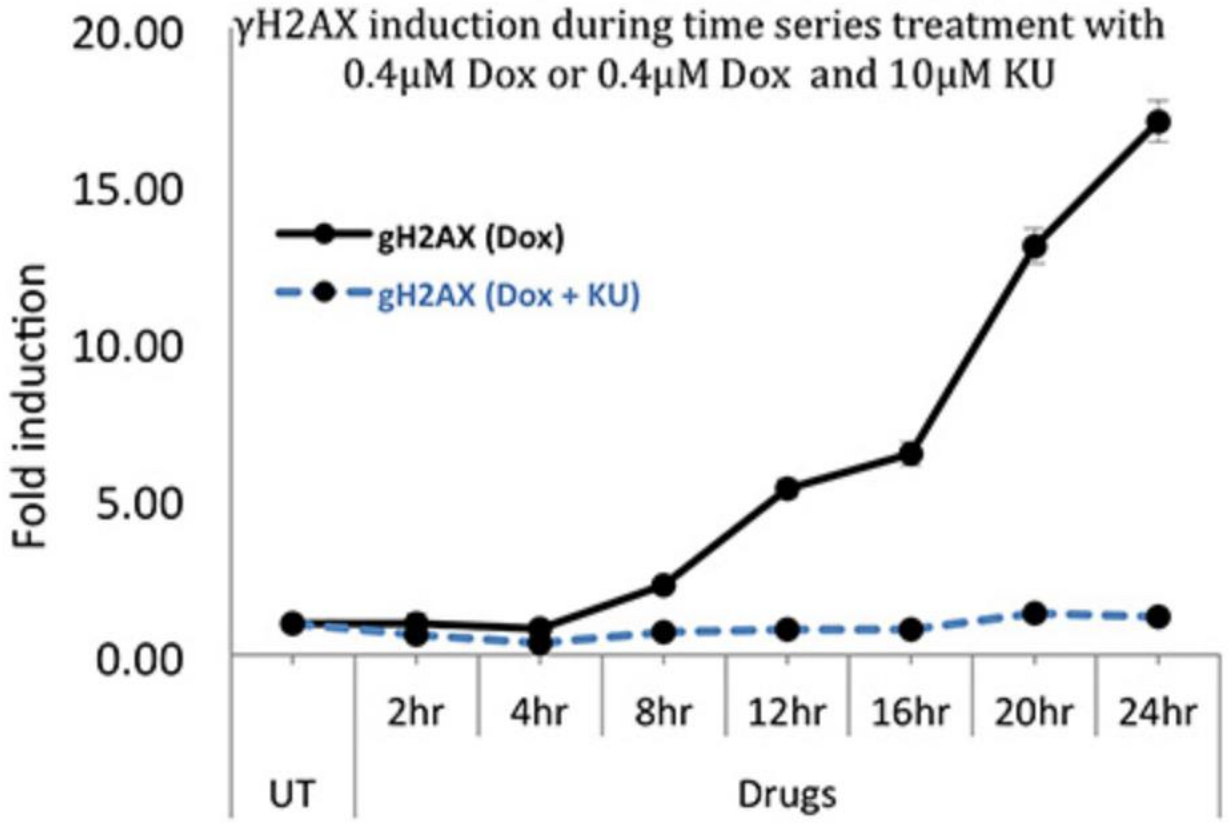
Snapshot in time

Pseudotime in MVD

Snapshot in time

Pseudotime in MVD

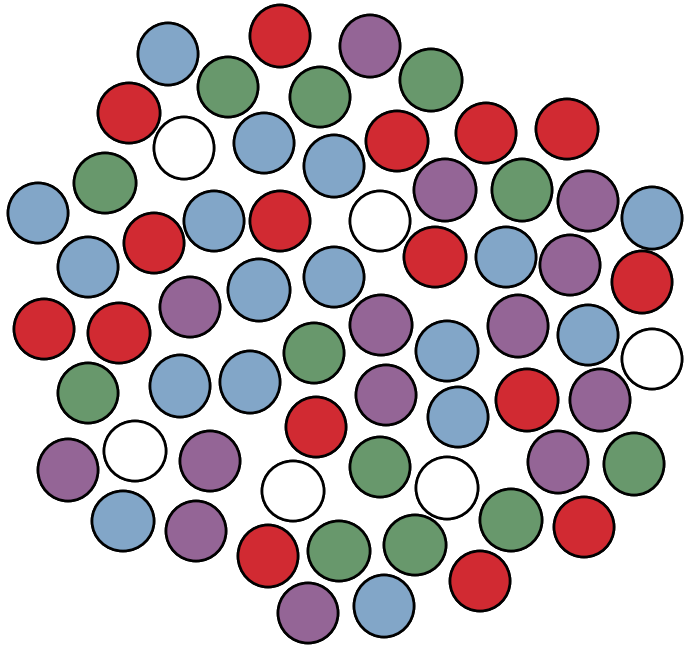
Time series data/analysis



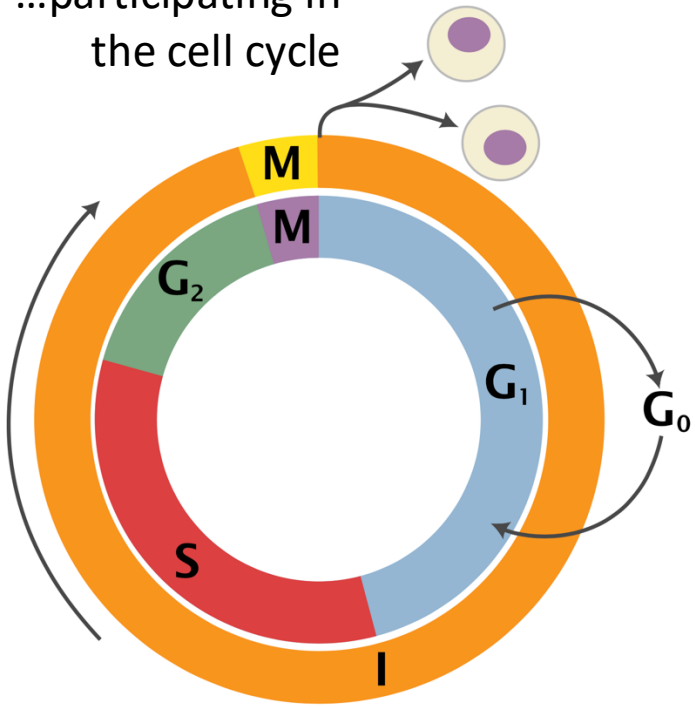
Taken from: Figure 9b | Idowu et al., 2013. Biodiscovery; 9:4.

“Homogeneous” population of cells

e.g., enriched by cell surface marker

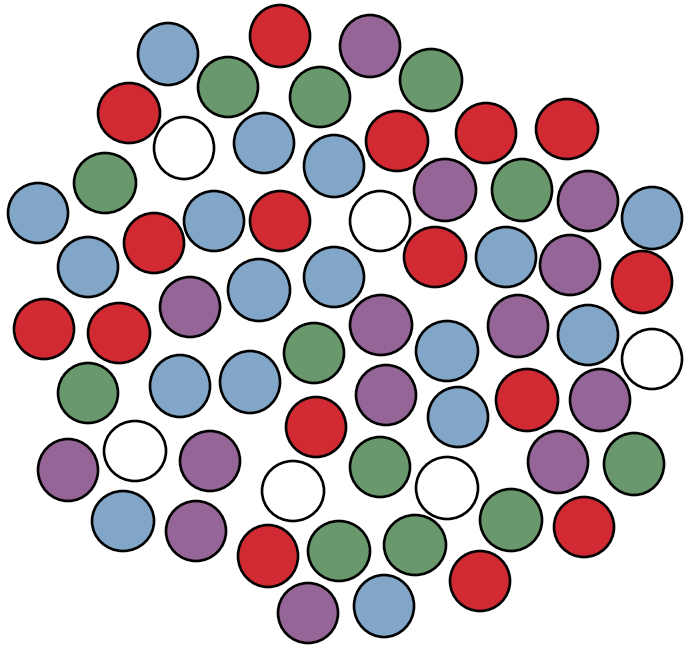


…participating in the cell cycle

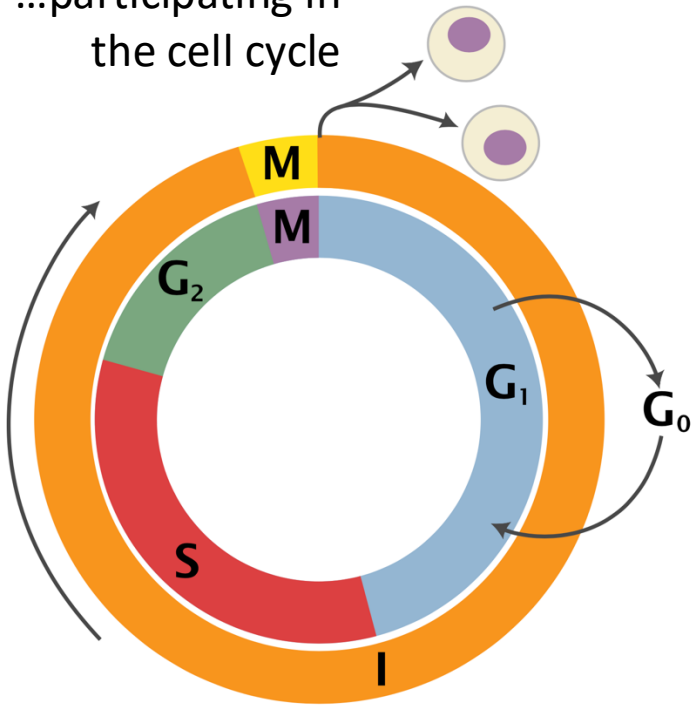


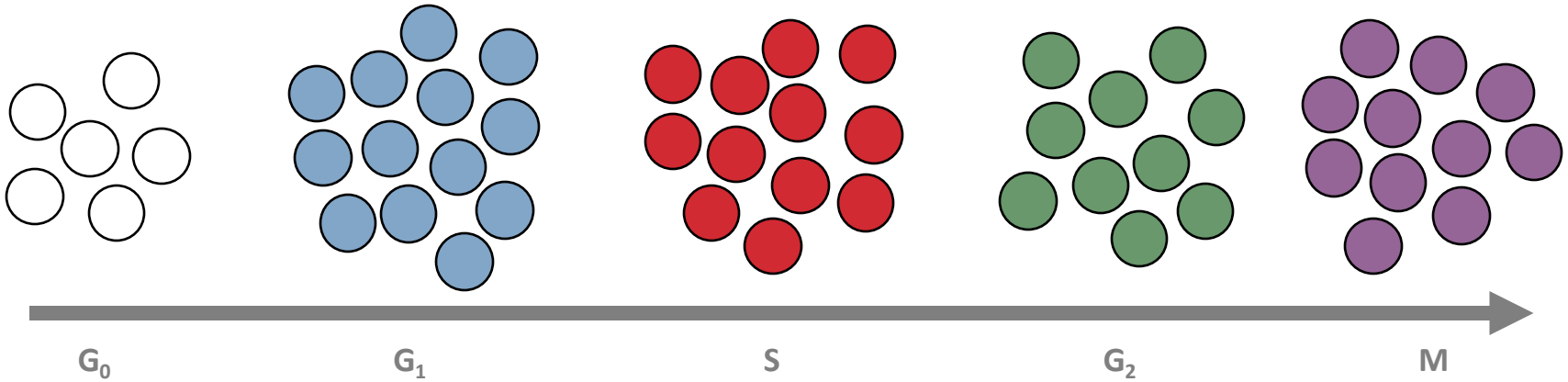
“Homogeneous” population of cells

e.g., enriched by cell surface marker



…participating in the cell cycle





Pseudotime in MVD

Studying mechanics of dynamic processes

Molecular mechanisms of dynamic processes – how do we identify and study them?

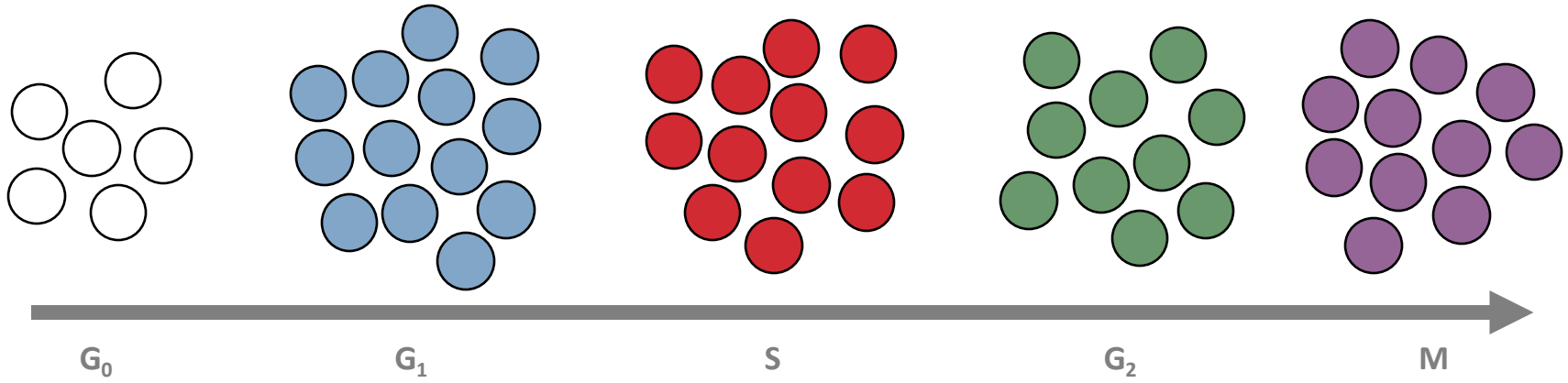
Sort cells?

Synchronize cells?

Is this possible/practical?

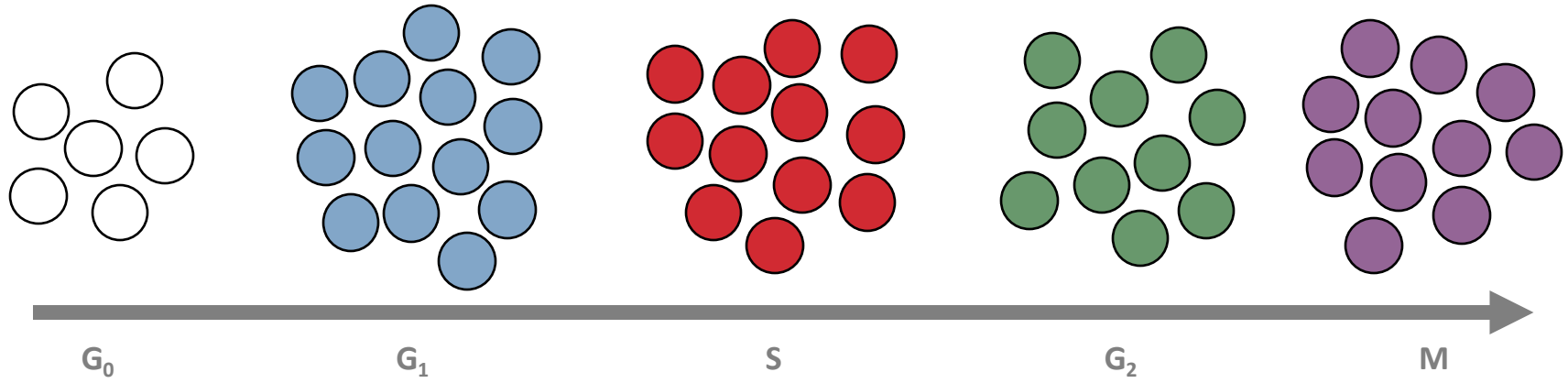
Are you going to cause unwanted changes?

Inferring a time series: “pseudotime”



- transcriptome dynamics
- inferring the order of cells along a trajectory
  assuming that transcription captures the relevant variability
- trajectory is through “transcriptional space”

Inferring a time series: “pseudotime”



curve reconstruction problem

1. identify the endpoints of the process
2. learn the shape of the path between them
3. order cells along this path

Inferring a time series: “pseudotime”

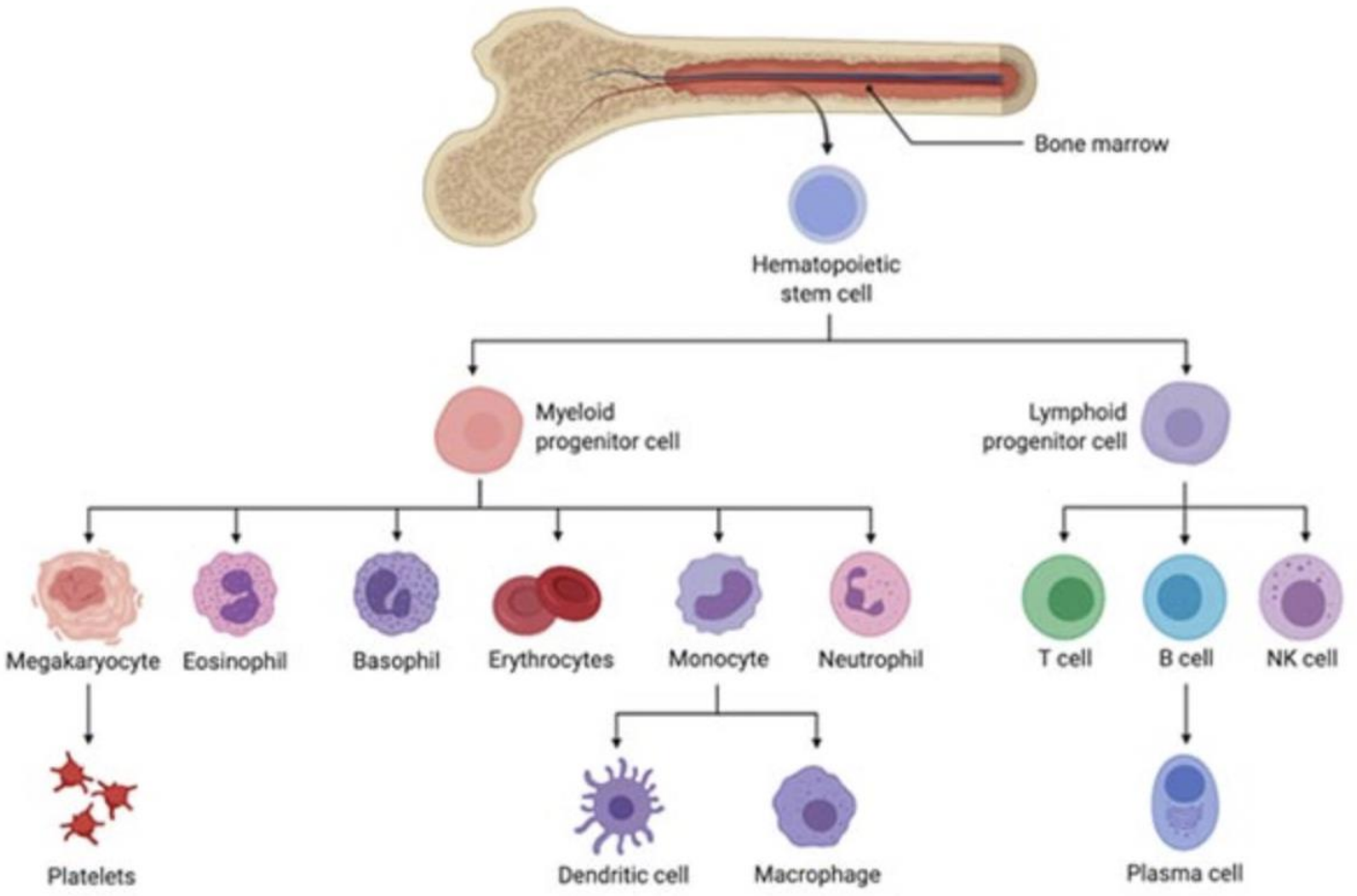
- When is this relevant?

1. When the process **represents the biology** that you are interested in
2. When the process is **obscuring the biology** that you are interested in

2. Examples

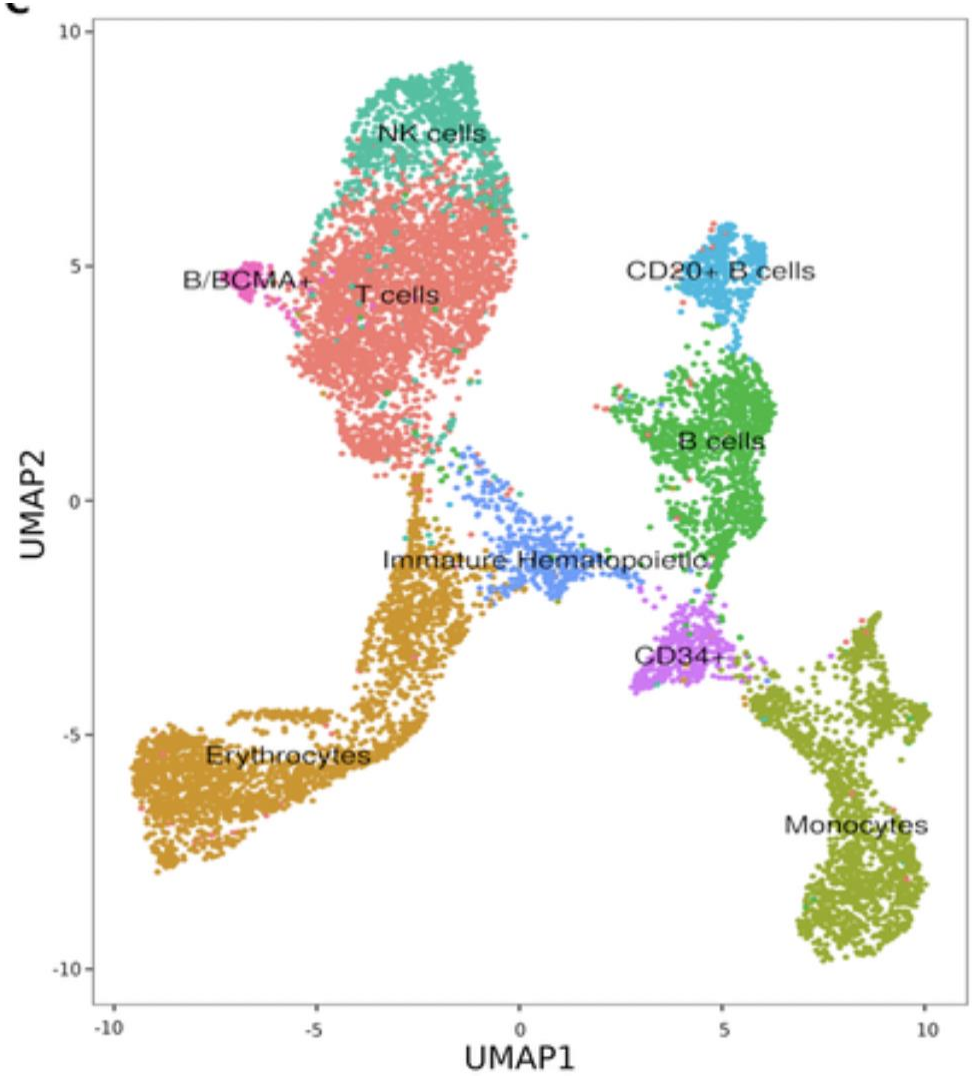
- cell cycle progression
- developmental/differentiation processes
  - Hematopoiesis (blood cell development)
  - oncogenic transformation
  - reprogramming
  - aging
  - cancer metastasis
- Parasite life cycle

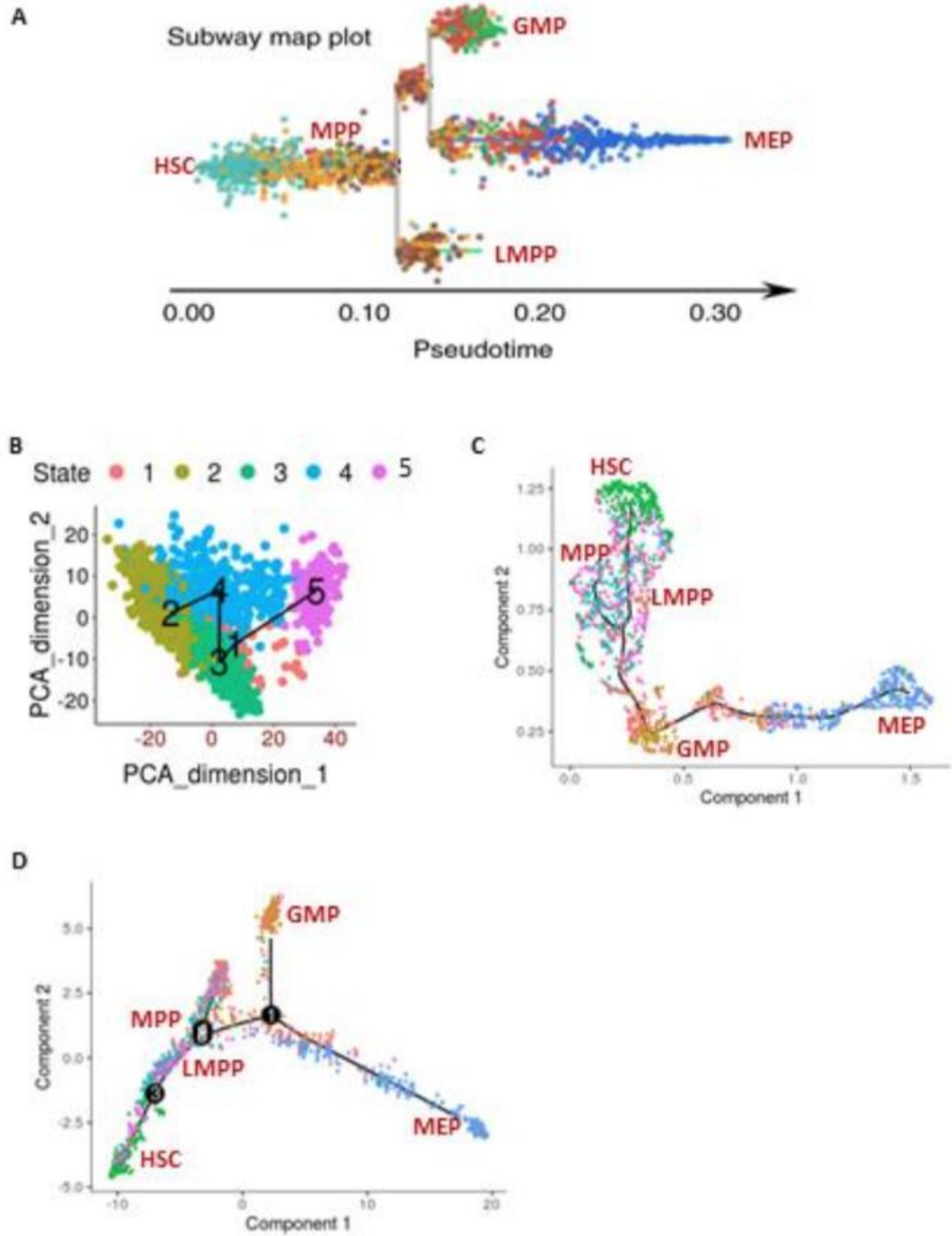
Hematopoiesis



https://icahn.mssm.edu/research/cardiovascular/focus/hematopoiesis

Different methods presenting Hematopoiesis dataset



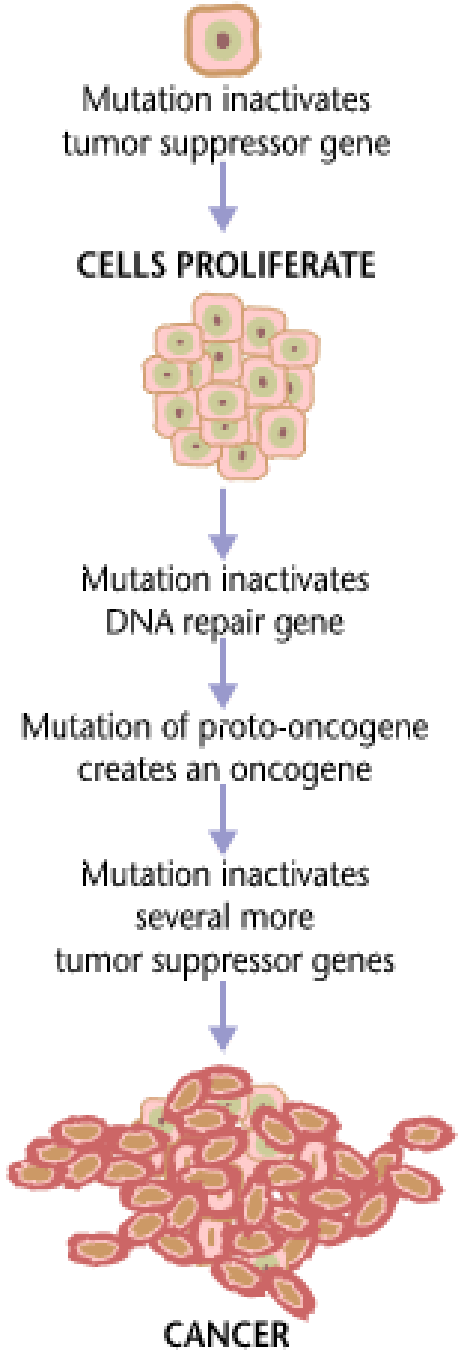


Pseudotime in MVD

Draft figure Rhona Millar

https://www.researchgate.net/publication/341407296_Single-cell_analysis_of_childhood_leukemia_reveals_a_link_between_developmental_states_and_ribosomal_protein_expression_as_a_source_of_intra-individual_heterogeneity/figures?lo=1&utm_source=google&utm_medium=organ

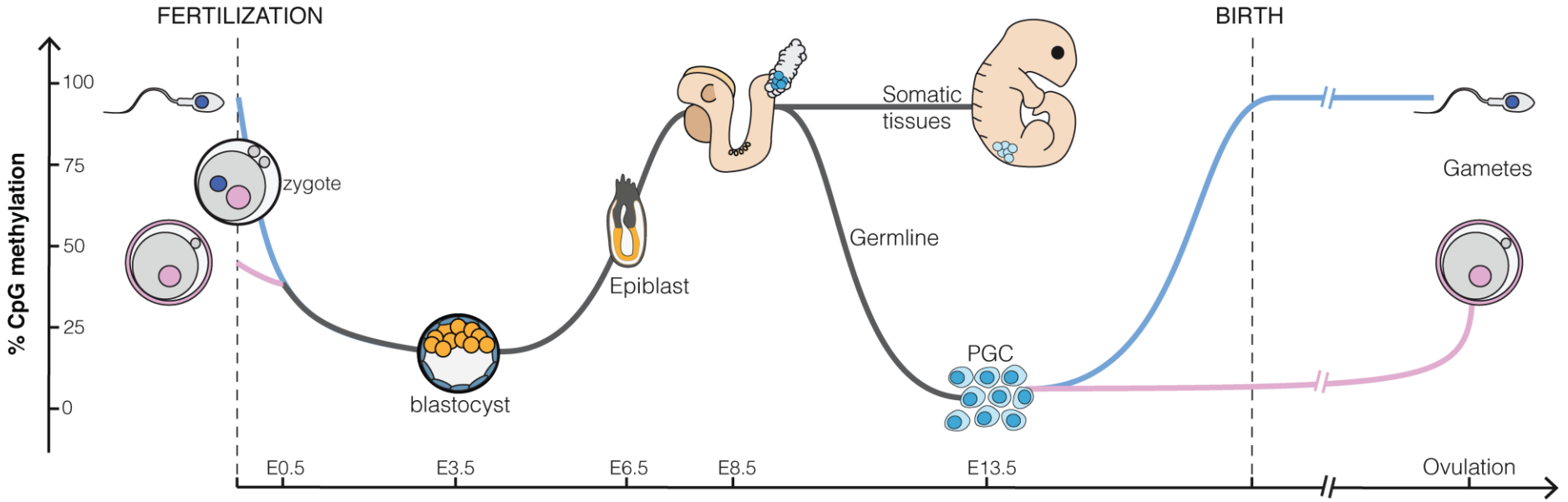
Oncogenic transformation



Wikipedia

Reprogramming

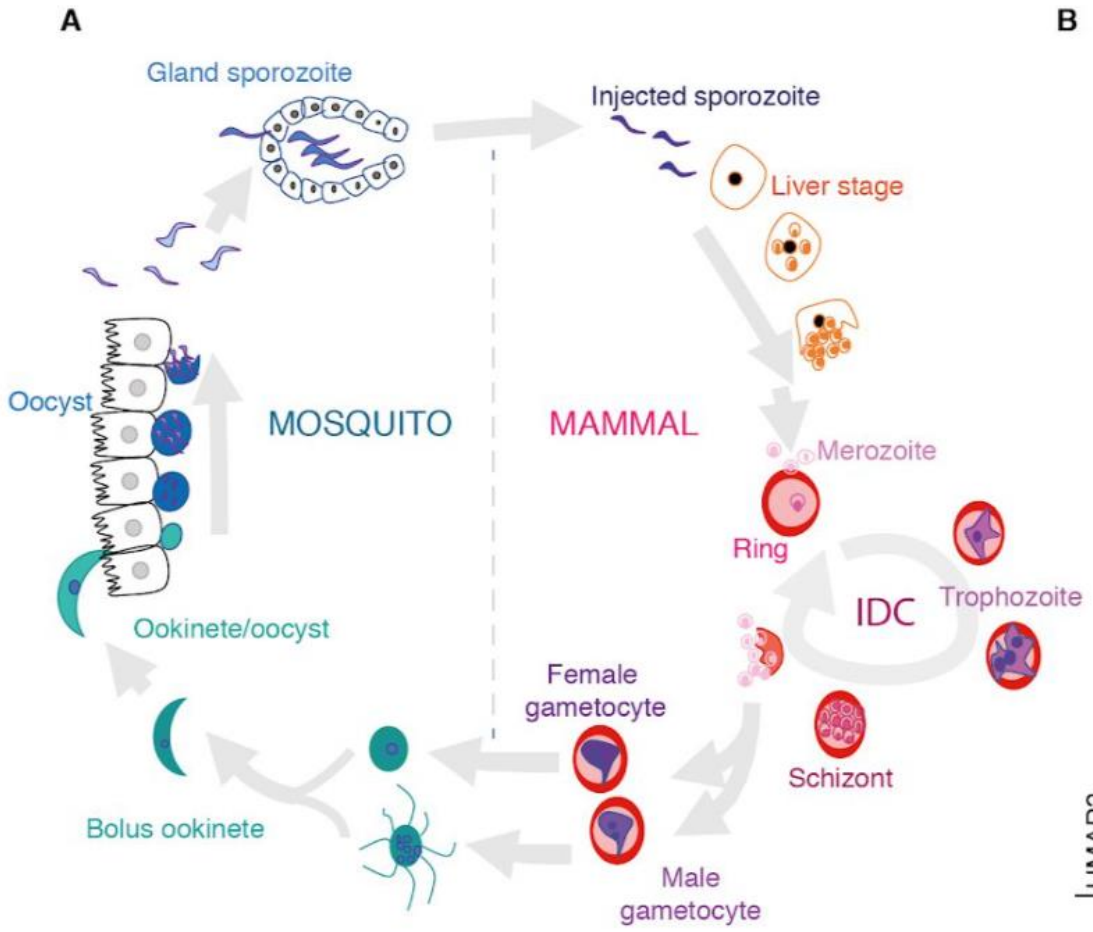
In biology, **reprogramming** refers to erasure and remodeling of epigenetic marks, such as DNA methylation, during mammalian development or in cell culture.[1] Such control is also often associated with alternative covalent modifications of histones

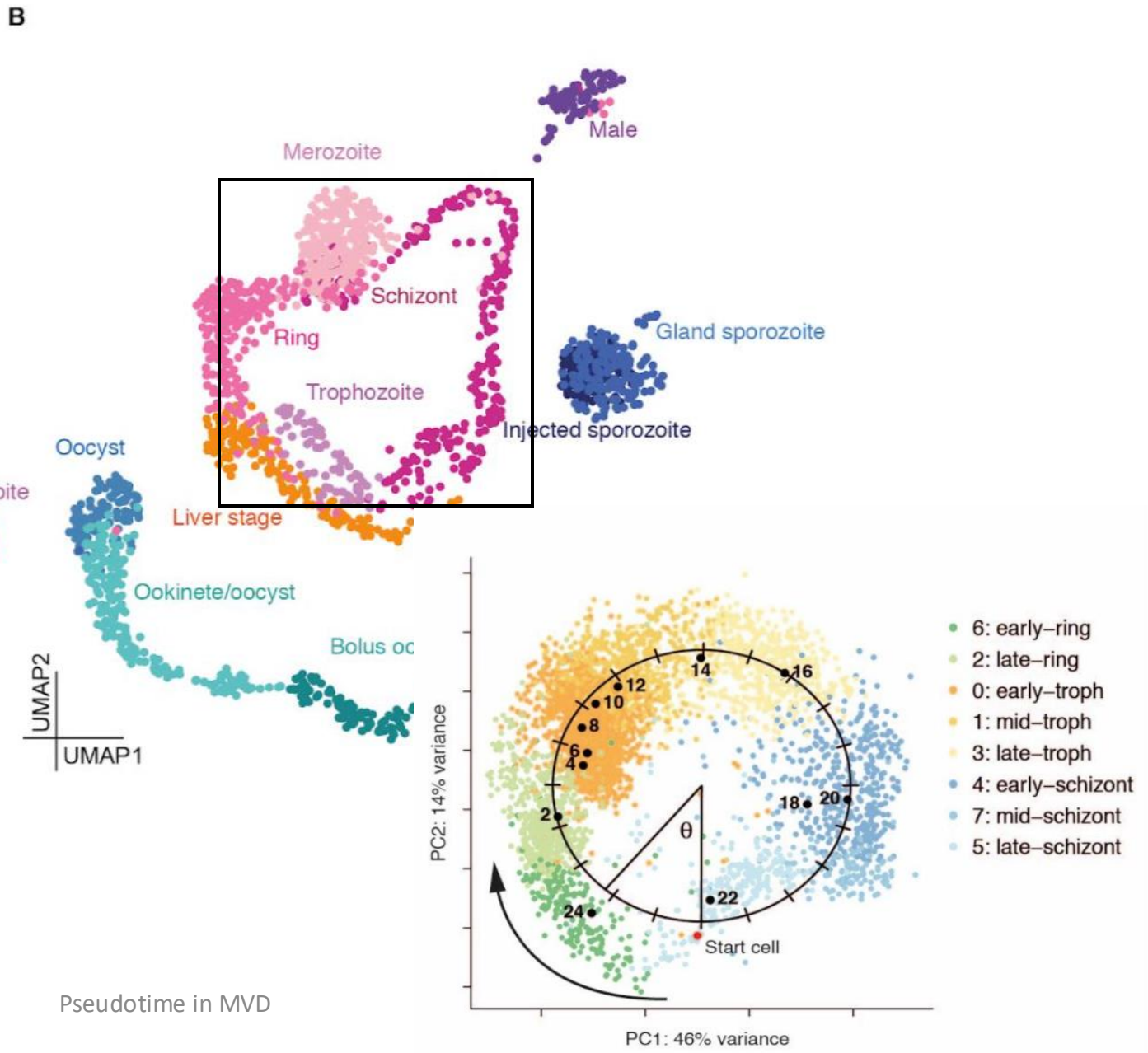


Pseudotime in MVD

Wikipedia

Full life cycle of malaria.

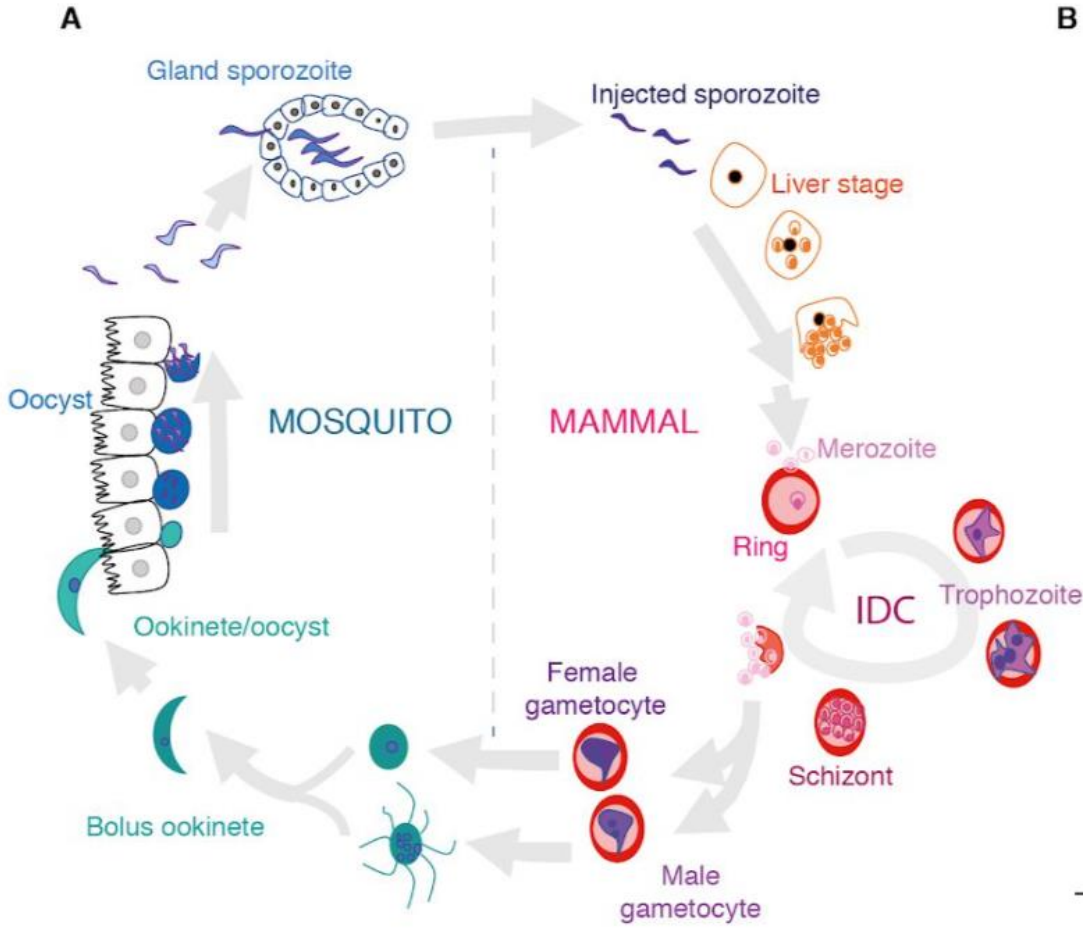


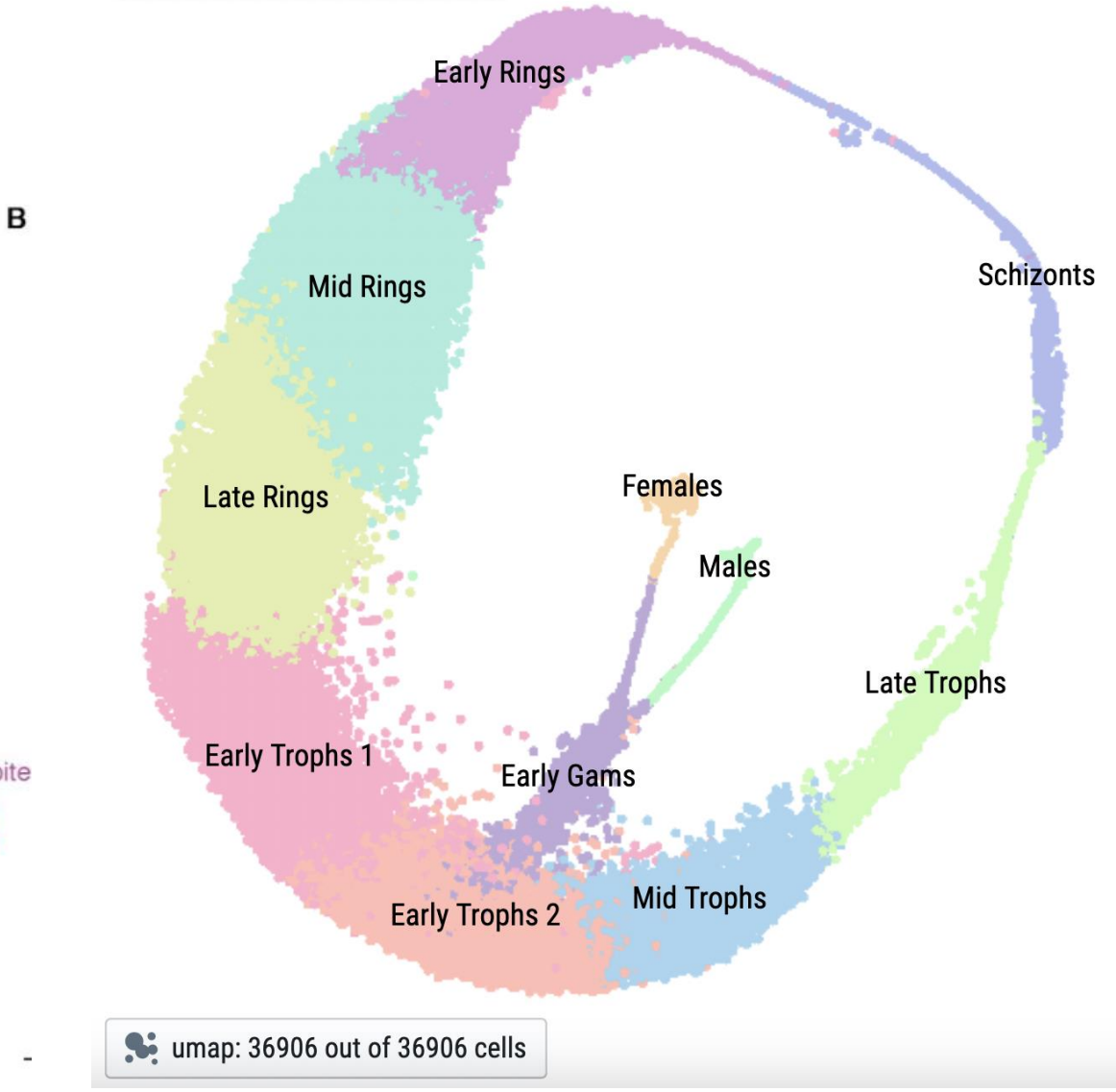


Pseudotime in MVD

Interesting for malaria is the formation of gametocytes

Relevant for 4th practical

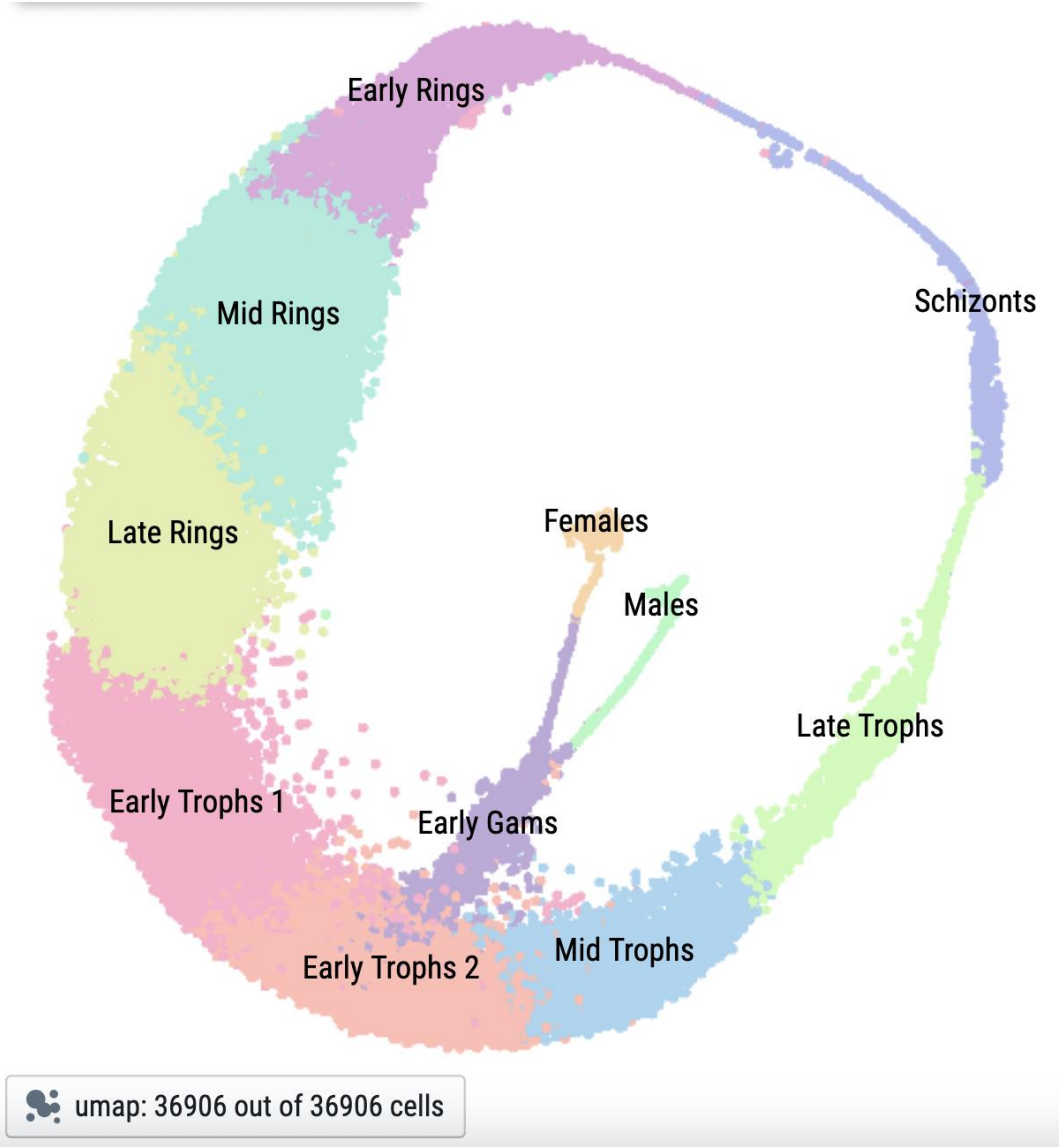


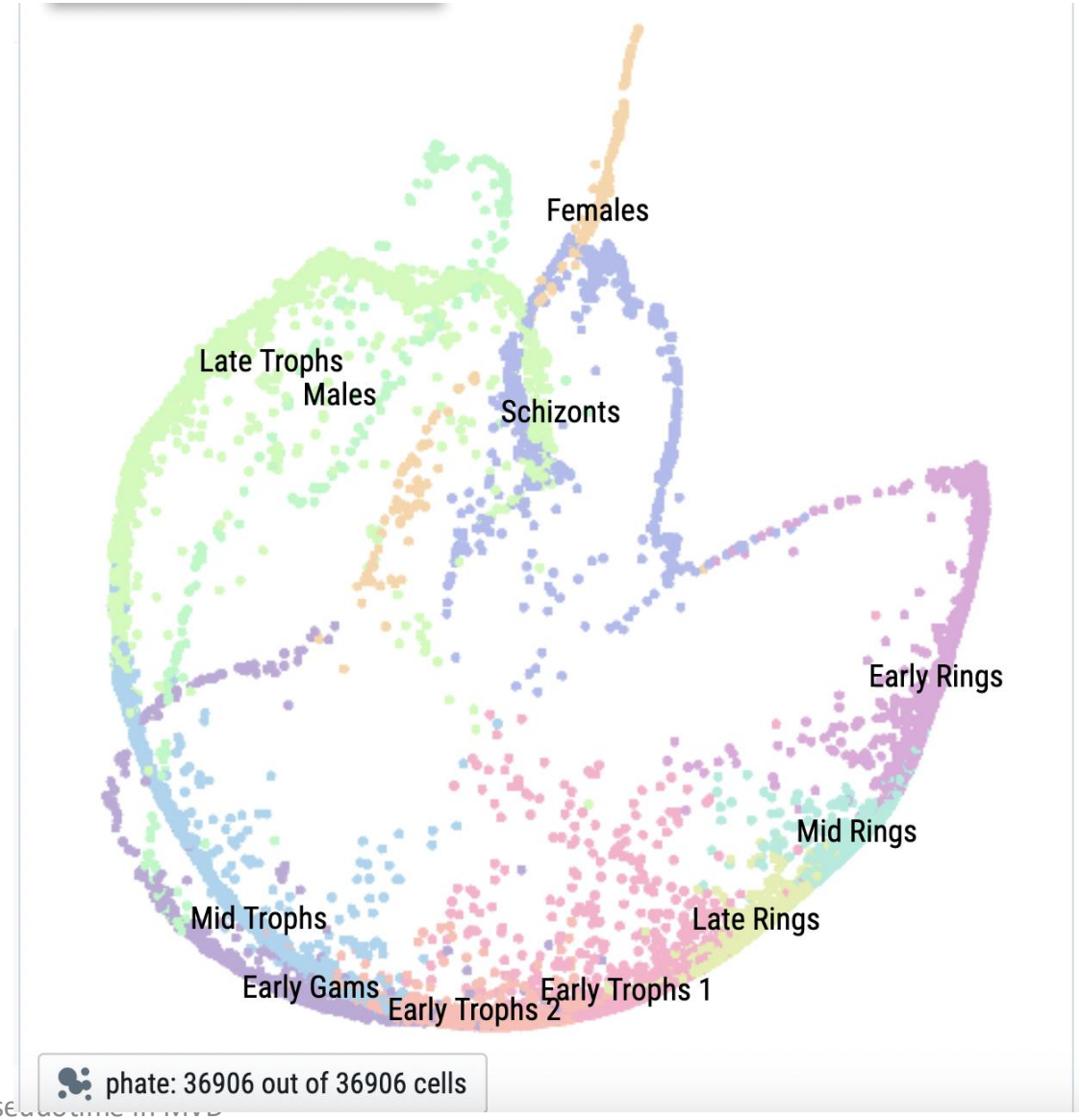


Pseudotime in MVD

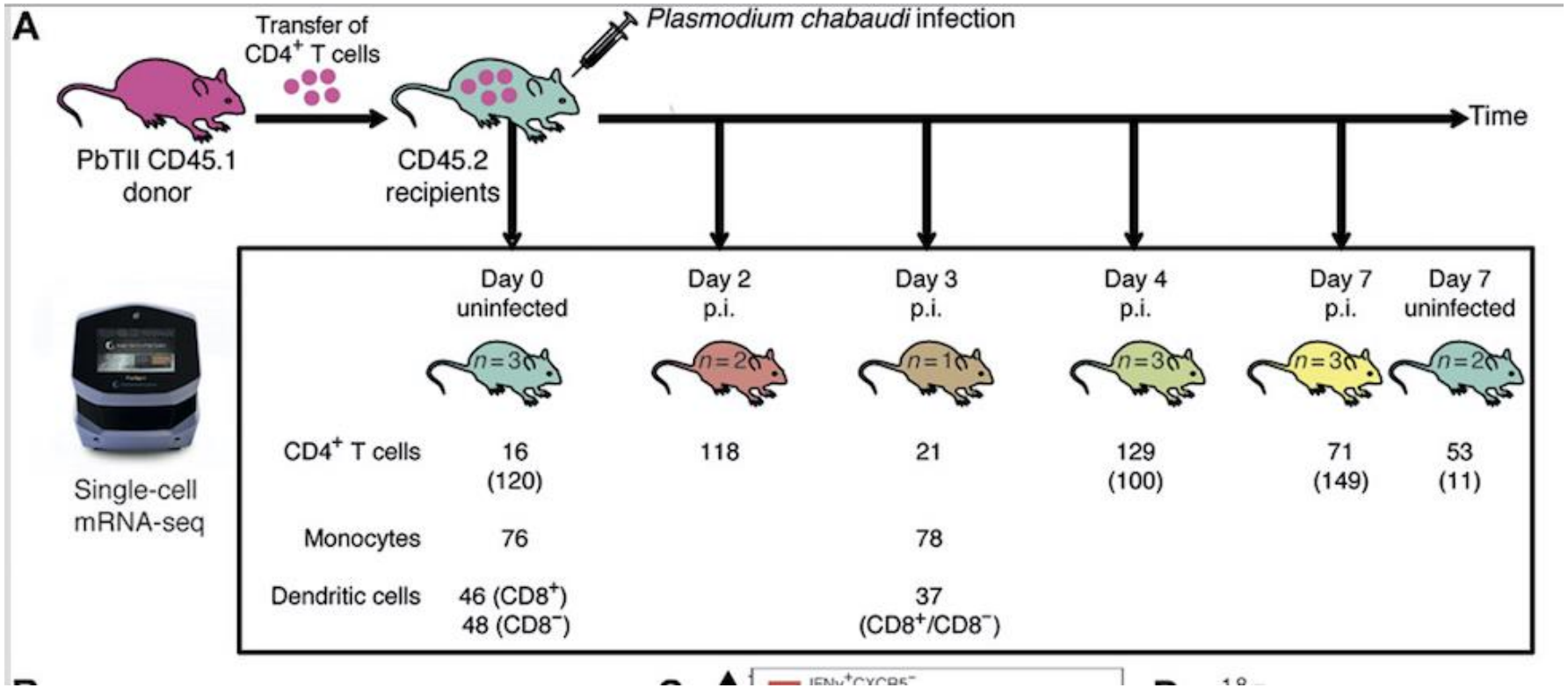
https://cellatlas-cxg.mvls.gla.ac.uk/view/Pb.Combined.h5ad/

UMAP is not the best representation for pseudo time



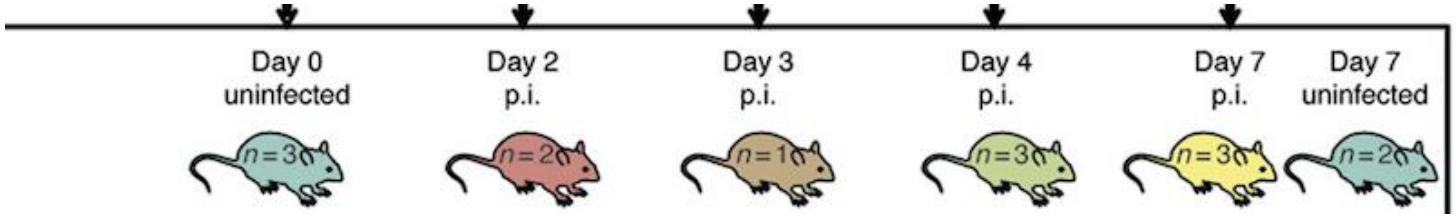


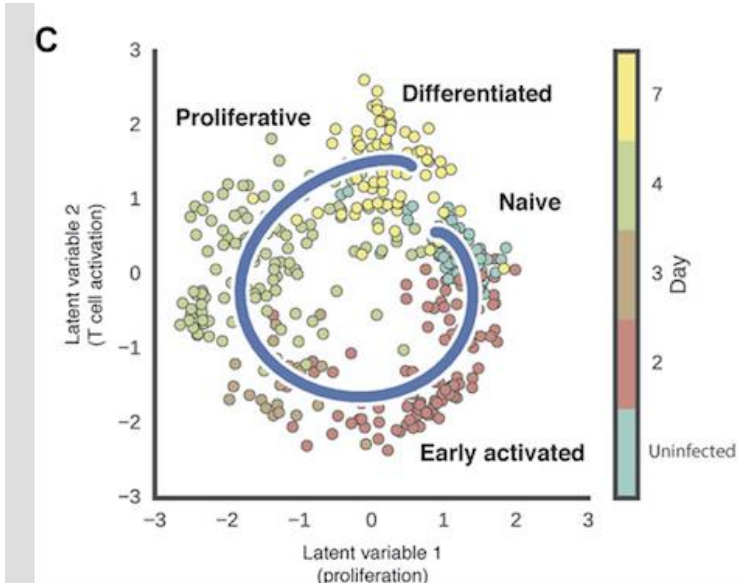
Understanding the Th cell activation and differentiation -



Lönnberg et al, Science Innumology 2017

Understanding the Th cell activation and differentiation -





Lönnberg et al, Science Immunology 2017

3. Exampling one method

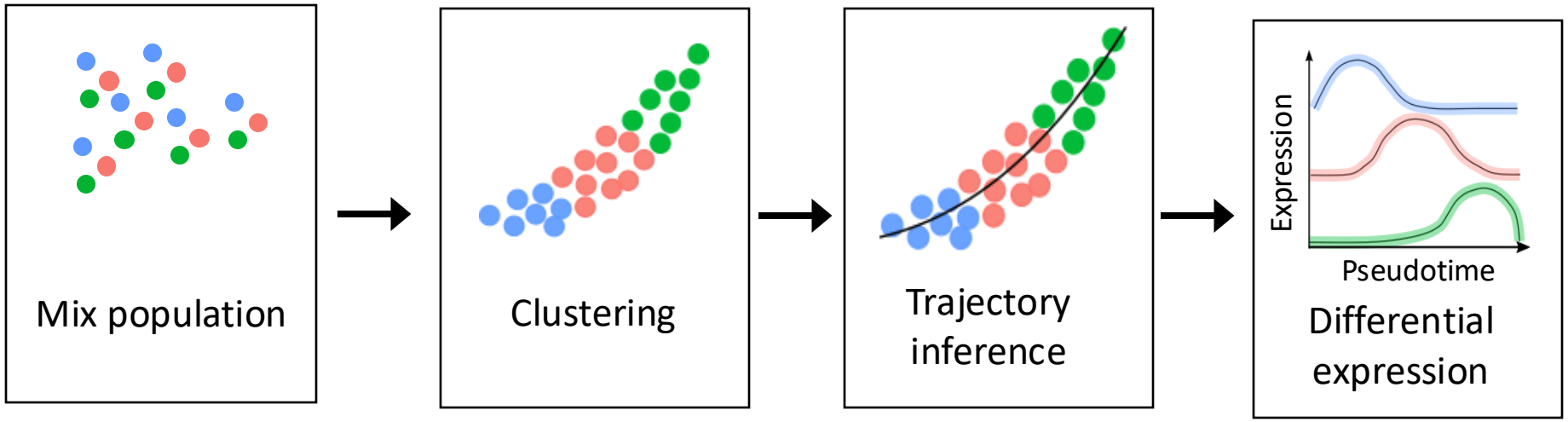
- Before need to explain:

Inferring a time series: “pseudotime”

pseudotime: a quantitative measure of progress through a biological process, as represented by gene expression

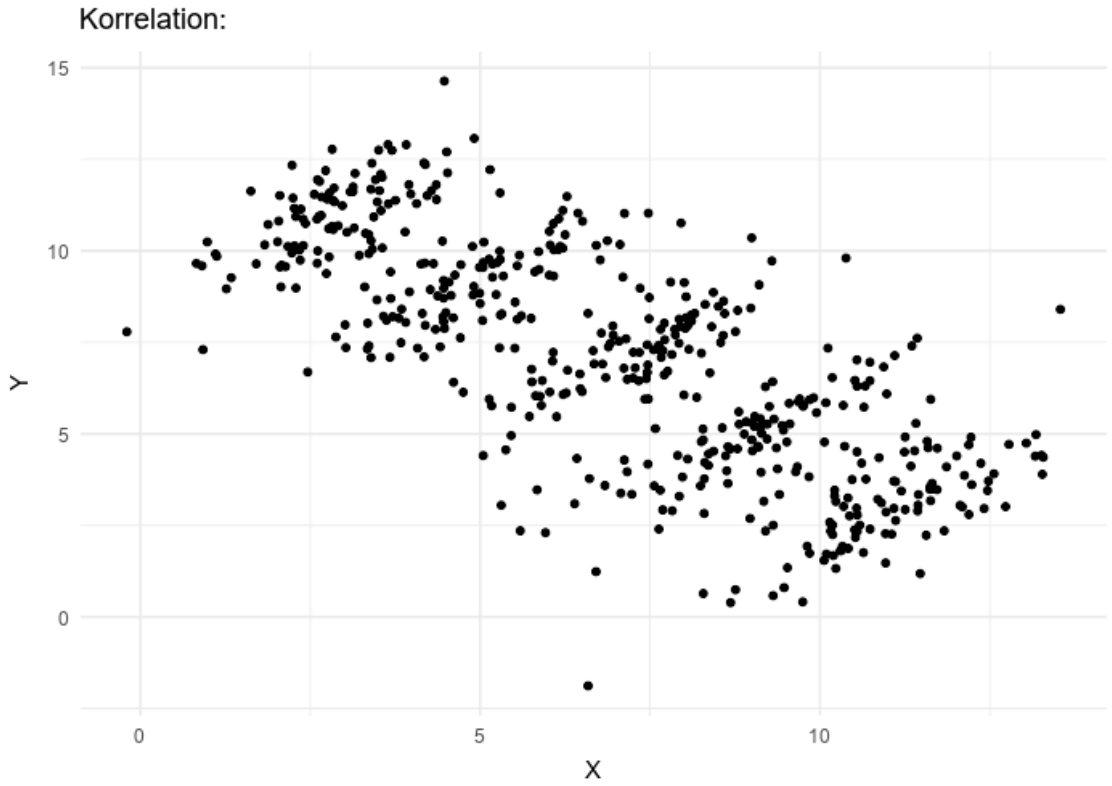
Discuss one methodology (Monocle) as an example

Word of caution (relevant to all *in vitro* biology)
In real life, any movement through transcriptional space is also a function of intercellular interactions and other stimuli – no cell operates in isolation



Simpson Paradox

Simpson's paradox (or Simpson's **reversal**, Yule–Simpson effect, amalgamation paradox, or **reversal** paradox), is a phenomenon in probability and statistics, in which a trend appears in several different groups of data but disappears or reverses when these groups are **combined**.



wikipedia

Some methodologies first (get an idea)

Travelling salesman problem

Minimum spanning tree (MSP)

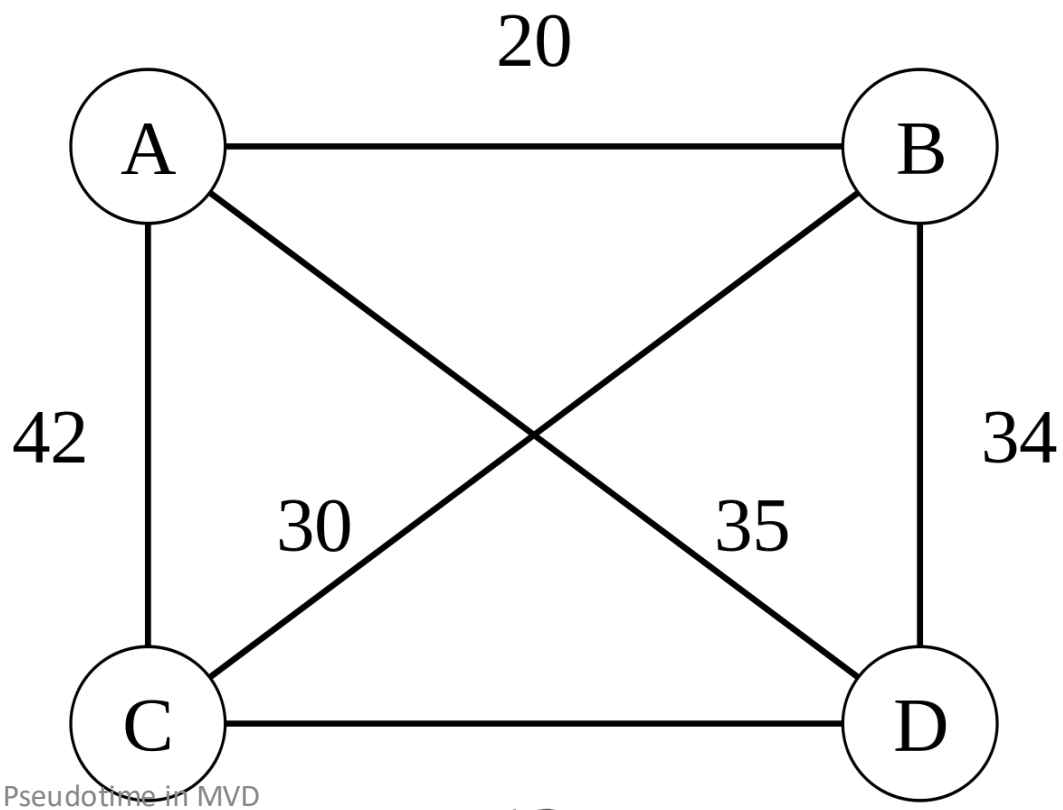
PG Tree

Independent component analysis

travelling salesman problem (TSP)

TSP: "Given a list of cities and the distances between each pair of cities, what is the shortest possible route that visits each city and returns to the origin city?"

How computational expensive is that?

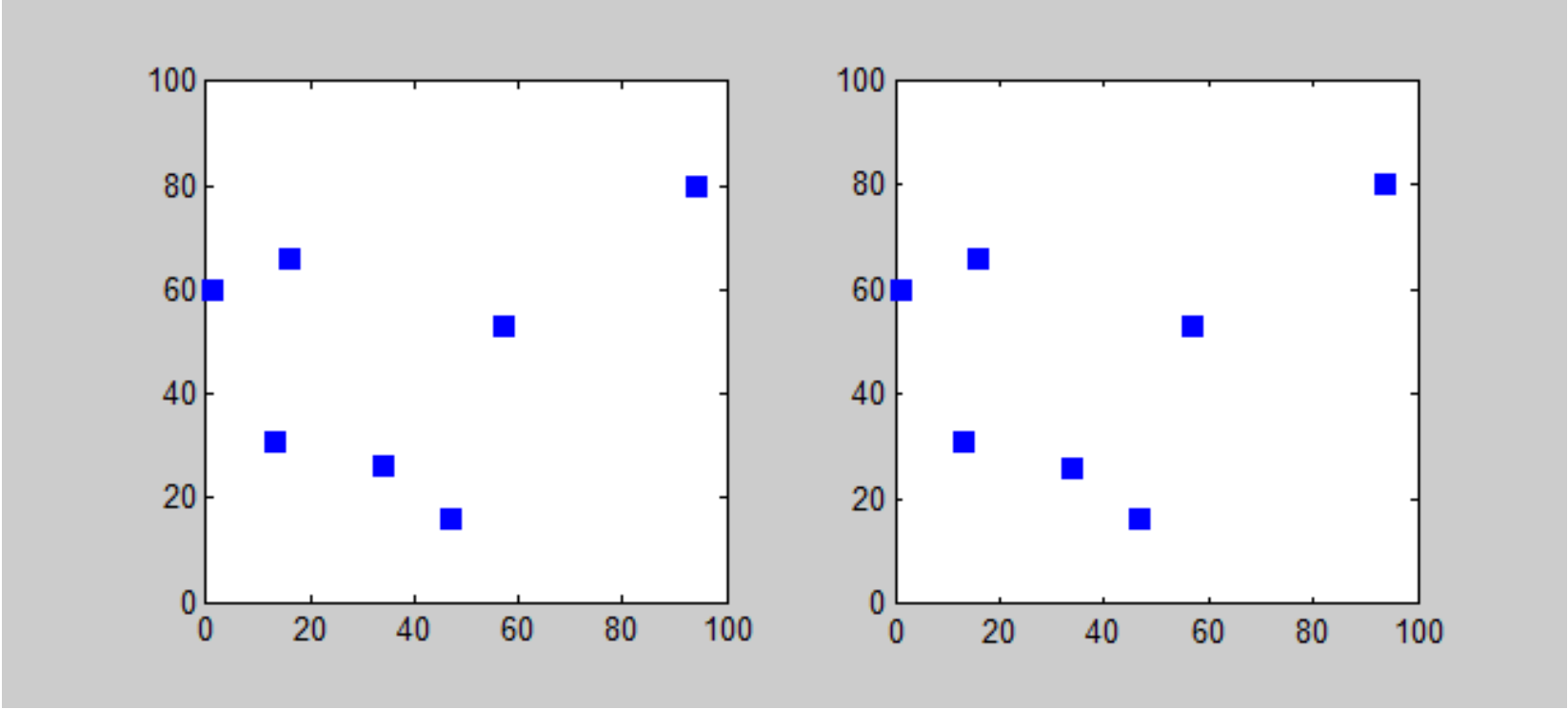


wikipedia

travelling salesman problem (TSP) 2/2

It is NP-hard!

Need heuristic for larger size graphs



Brute force example, looking for the best solution, 7 cities 360 steps!

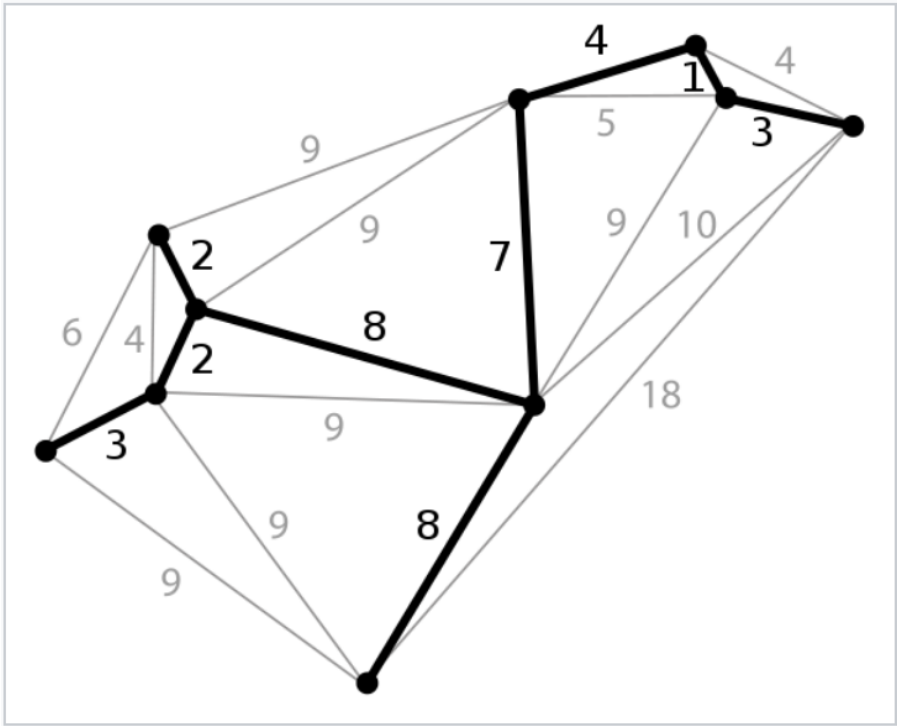
wikipedia

Minimum spanning tree (MSP)

It is NP-hard!

Without cycles;

connect all the edges with a minimal sum



A minimum spanning tree of a weighted planar graph. Finding a minimum spanning tree is a common problem involving combinatorial optimization.

wikipedia

PQ Tree – find the order

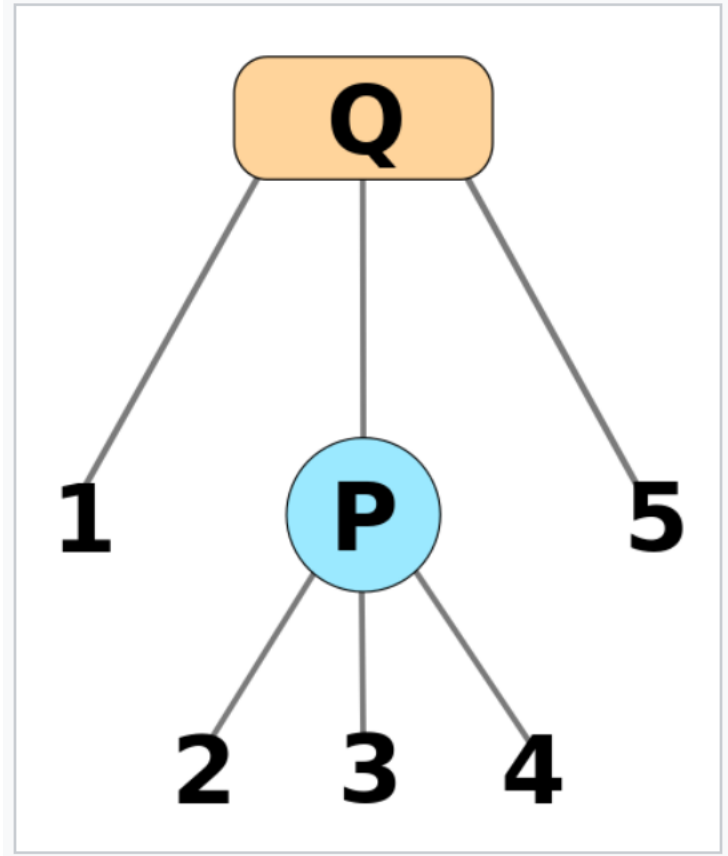
Data structure to store “order”

Tree with “fixed nodes Q” (just reverse order)

P can permutate the order

Example: [1 (2 3 4) 5]

12345, 12435, 13245, 13425, 14235, 14325, 52341, 52431, 53241, 53421, 54231, 54321



The PQ tree representing [1 (2 3 4) 5]

wikipedia

Let’s take a deep breath ☺

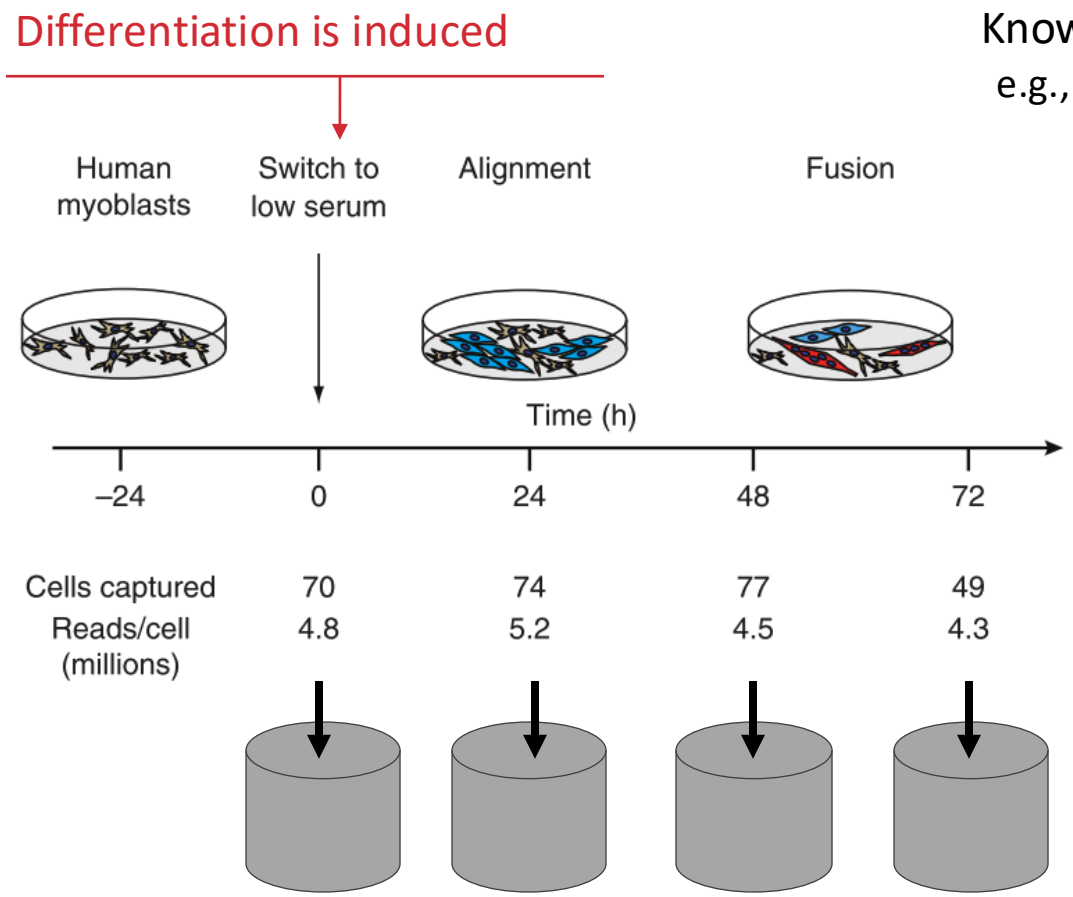
I think that a bioinformatician should have an idea of what is under the hood!

MONOCLE

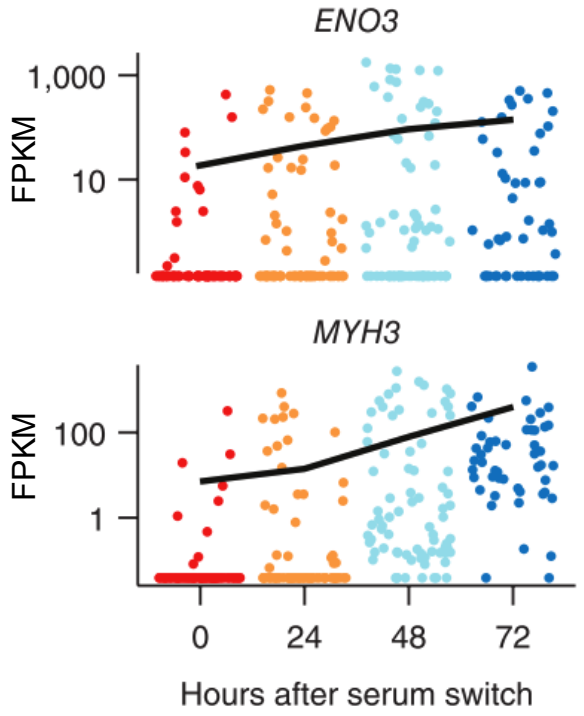
Trapnell et al., 2014. Nature Biotechnology, 32: 381–386.

Monocle: data and motivation

Differentiation of **H**uman **S**keletal **M**uscle **M**yoblasts



Known that several TFs are important
e.g., ENO3 / MYH3 are late stage markers



Pseudotime in MVD

Taken from: Figure 1a,c | Trapnell et al., 2014. Nature Biotechnology, 32: 381–386.

Monocle: algorithm overview

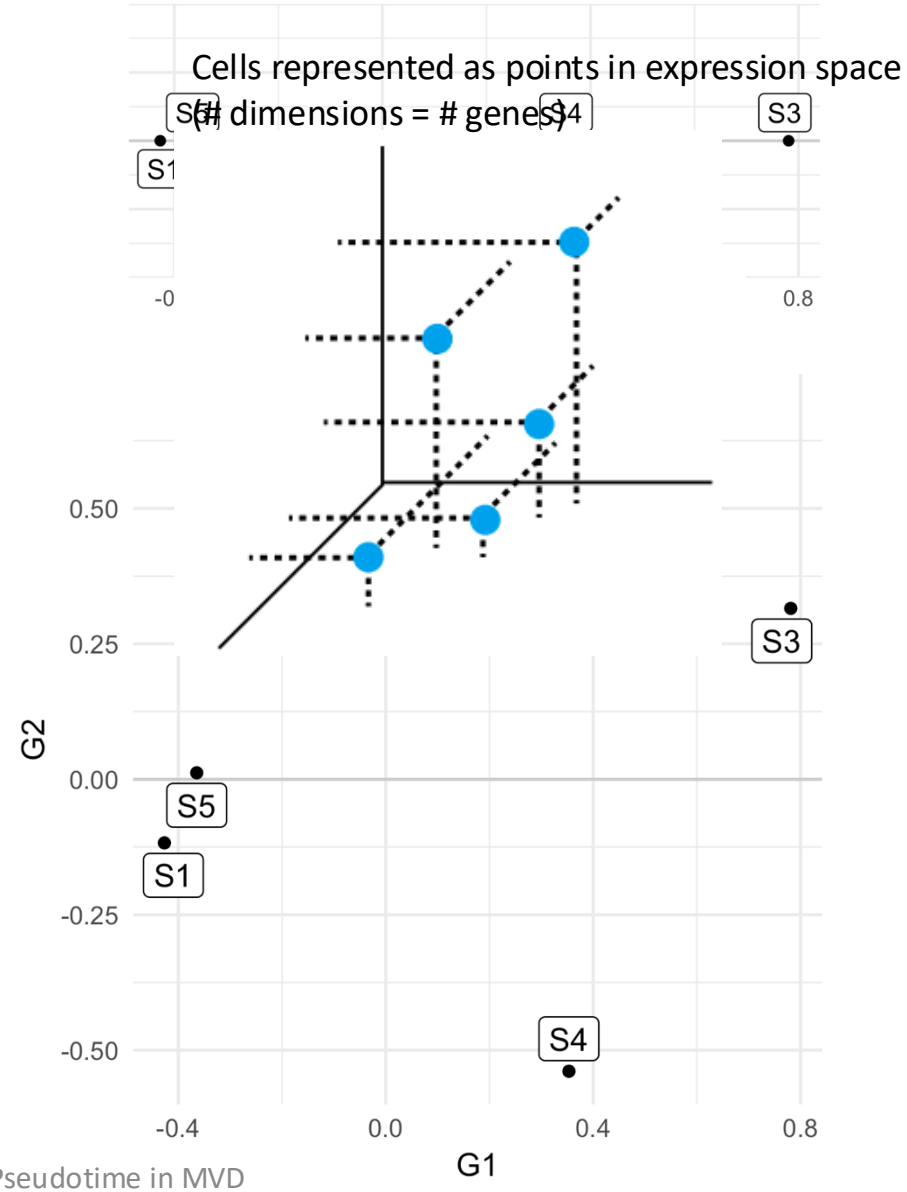
Read count data
genes x cells

Reduce dimensionality
***ICA**, PCA, tSNE, …

Identify trajectory
Minimal spanning tree

Order cells
…to define pseudotime axis

Overlay additional data
gene expression, cluster etc…



Pseudotime in MVD

Taken from: Figure 2a | Trapnell et al., 2014. Nature Biotechnology, 32: 381–386.

Monocle: algorithm overview

Read count data
genes x cells

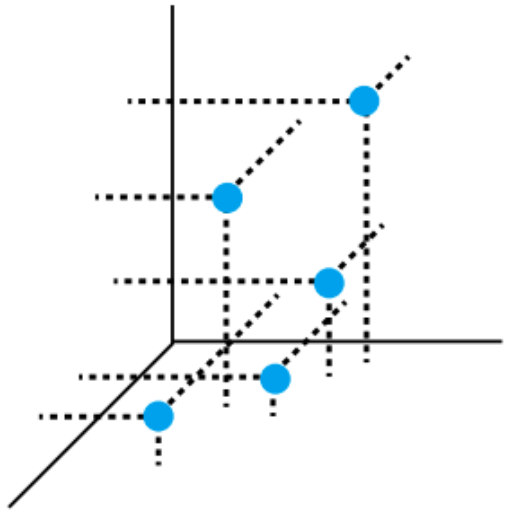
Reduce dimensionality
***ICA**, PCA, tSNE, …

Identify trajectory
Minimal spanning tree

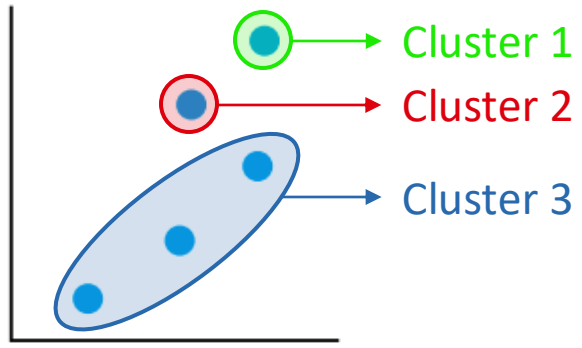
Order cells
…to define pseudotime axis

Overlay additional data
gene expression, cluster etc…

Cells represented as points in expression space
(# dimensions = # genes)

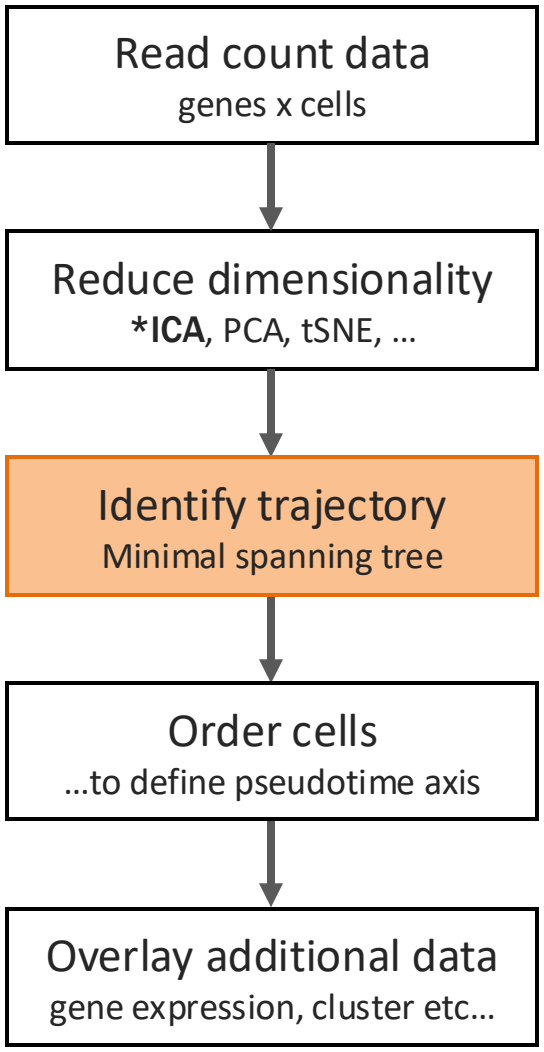


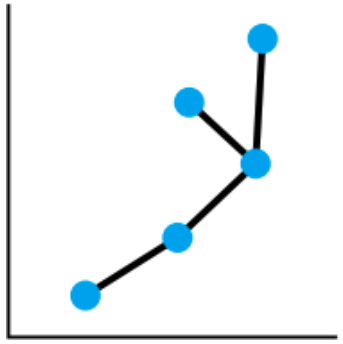
Reduce dimensionality



Taken from: Figure 2a | Trapnell et al., 2014. Nature Biotechnology, 32: 381–386.

Monocle: algorithm overview





Reduce dimensionality
Build MST on cells

Taken from: Figure 2a | Trapnell et al., 2014. Nature Biotechnology, 32: 381–386.

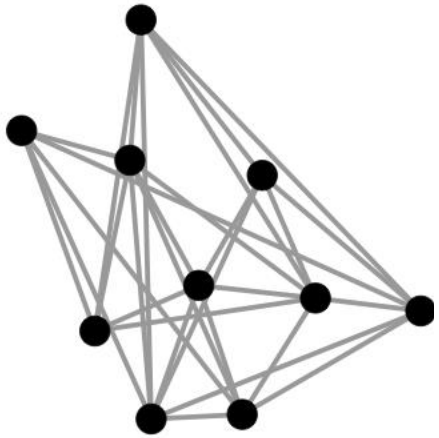
Monocle: minimum spanning tree

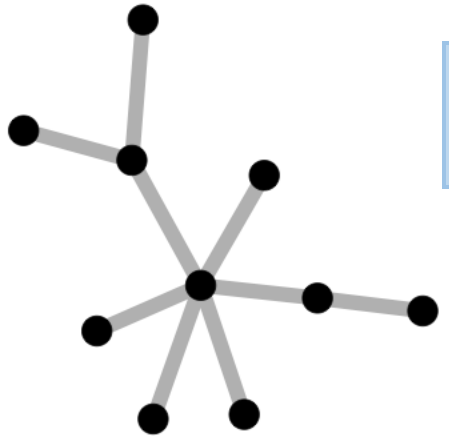
Shortest path that incorporates all points

- (similar to traveling salesman problem)

Shortest = least cost

Edges are weighted by Euclidean distance between cells





curve reconstruction problem

1. identify the endpoints
2. learn the path between them
3. order cells along this path

But how do we order the cells when the MST has branches?

PQ Tree – find the order

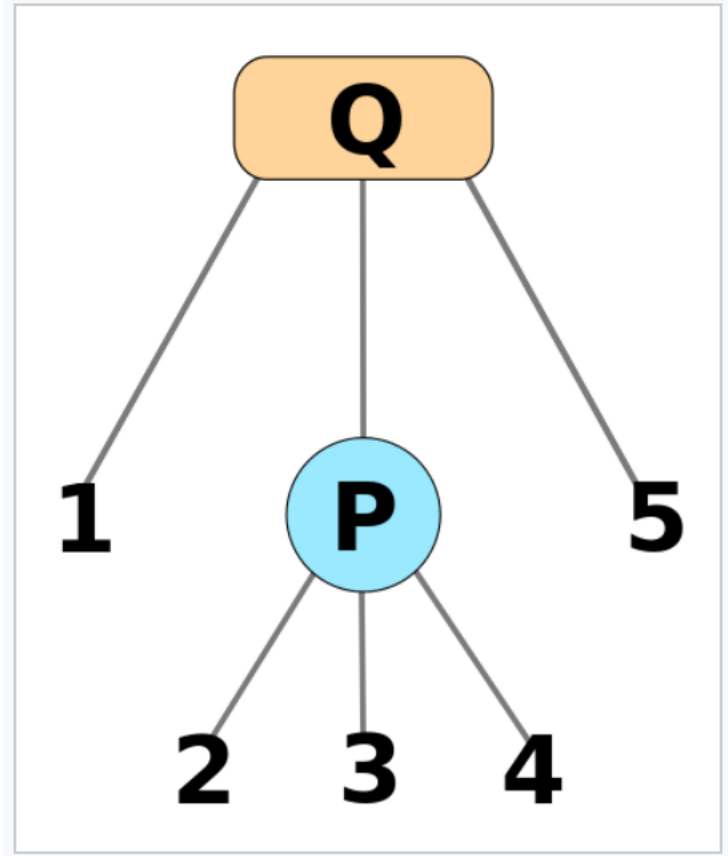
Data structure to store “order”

Tree with “fixed nodes Q” (just reverse order)

P can permutate the order

Example: [1 (2 3 4) 5]

12345, 12435, 13245, 13425, 14235, 14325, 52341, 52431, 53241, 53421, 54231, 54321



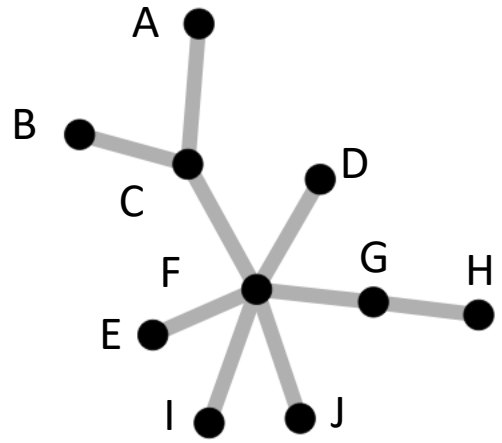
The PQ tree representing [1 (2 3 4) 5]

wikipedia

Monocle: resolving ordering in MST

Construct a PQ tree to represent legal orderings

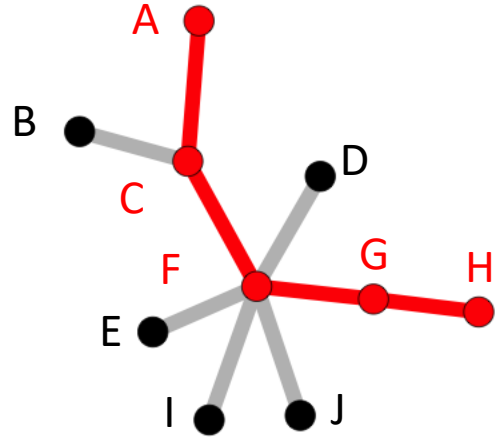
"Q" nodes have ordered children (order can be reversed) - rectangle
"P" nodes have a set of unordered children - circle



Monocle: resolving ordering in MST

Construct a PQ tree to represent legal orderings

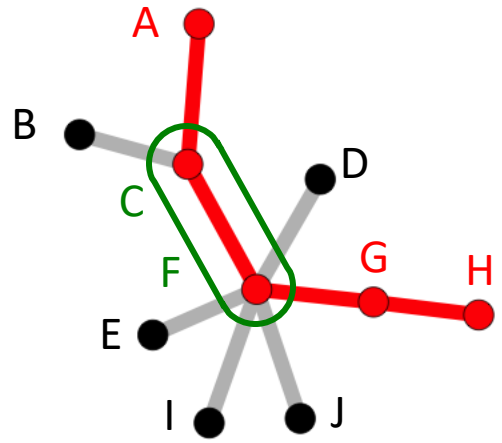
"Q" nodes have ordered children (order can be reversed) - rectangle
"P" nodes have a set of unordered children - circle

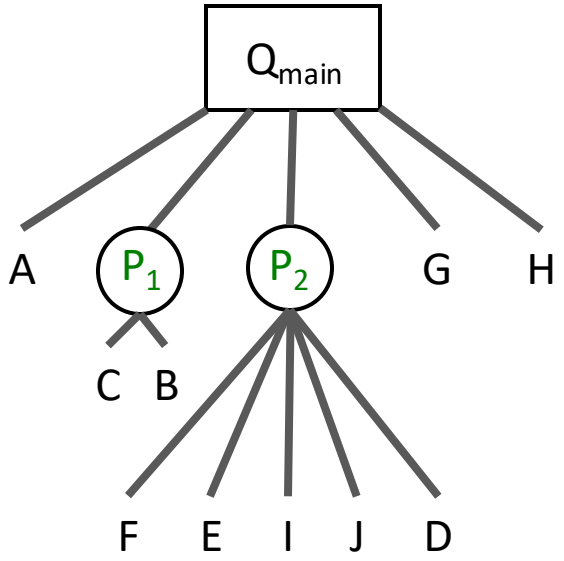


Monocle: resolving ordering in MST

Construct a PQ tree to represent legal orderings

"Q" nodes have ordered children (order can be reversed) - rectangle
"P" nodes have a set of unordered children - circle





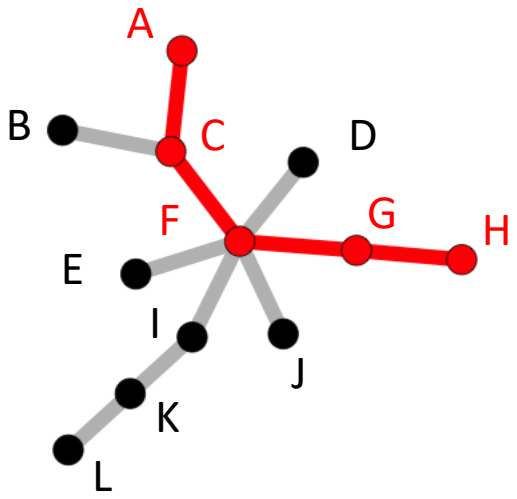
ACBFEIJDGH
ABCFEIJDGH
ACBFIEDJGH
ABCFEIJDGH
HGFEIJDCBA
HGFEIJDBCA
HGJEIFDCBA
etc…

1. Calculate MST
2. Find graph diameter (i.e., the longest path)
3. Initiate PQ tree with Q_{main} node
4. Work across diameter adding P/Q nodes to encode a family of orderings
5. Monocle searches all orderings to obey constraints and minimise distance

Monocle: pseudotime in branches

Backbone P nodes represent random variation

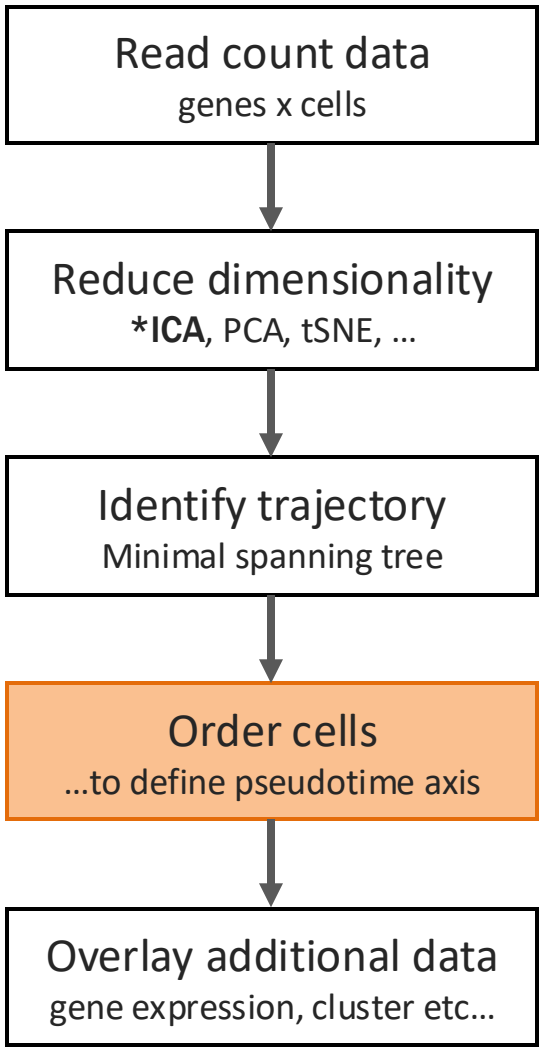
Backbone Q nodes indicate *branching points*
i.e., where trajectory is certain

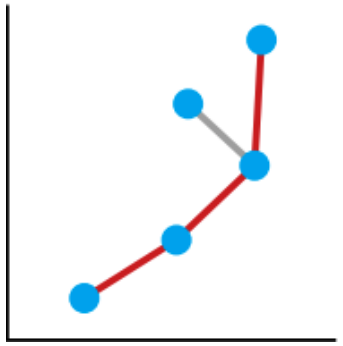


k (user defined) = # branches permitted

Branch dependent pseudotime

Monocle: algorithm overview

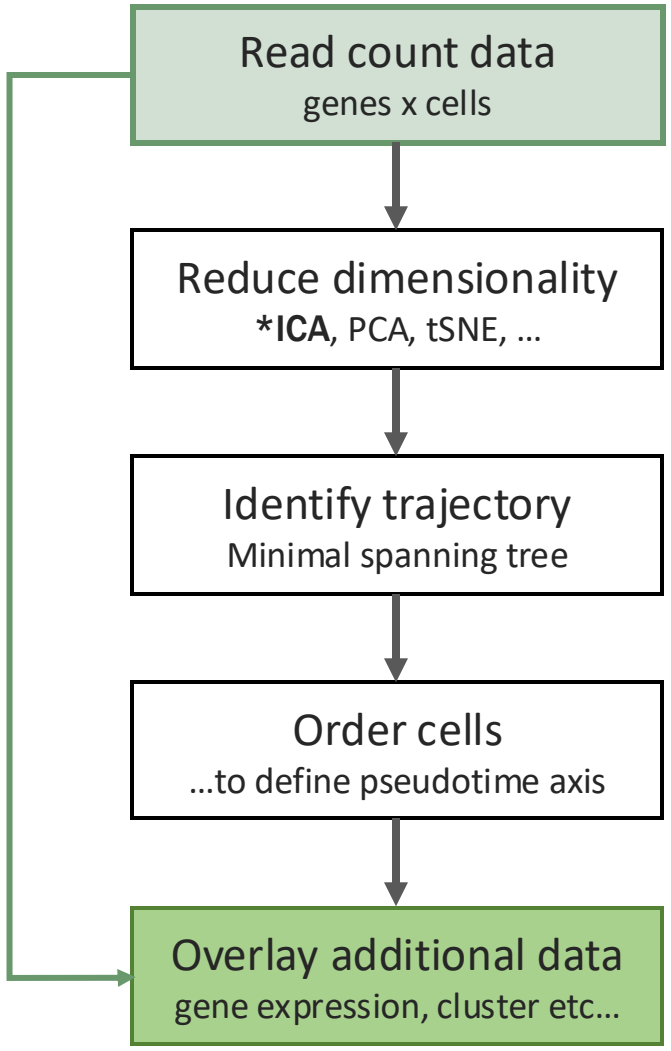




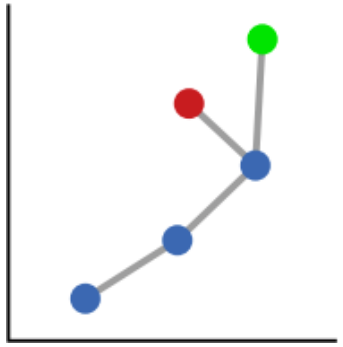
Reduce dimensionality
Build MST on cells
Order cells in pseudotime

Taken from: Figure 2a | Trapnell et al., 2014. Nature Biotechnology, 32: 381–386.

Monocle: algorithm overview



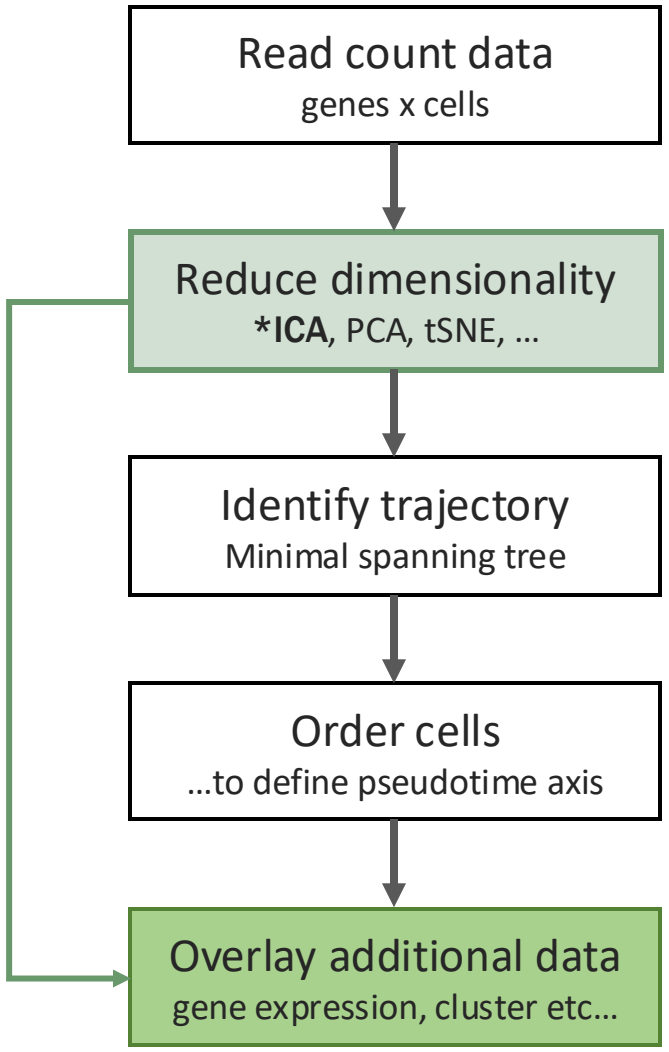
Investigate relationship between cell markers and pseudotime



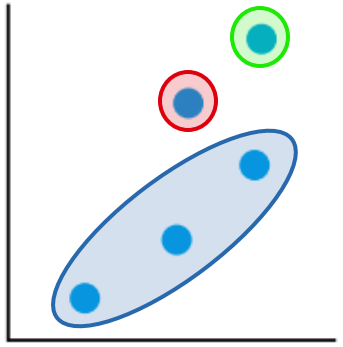
Reduce dimensionality
Build MST on cells
Order cells in pseudotime
Overlay marker expression

Taken from: Figure 2a | Trapnell et al., 2014. Nature Biotechnology, 32: 381–386.

Monocle: algorithm overview



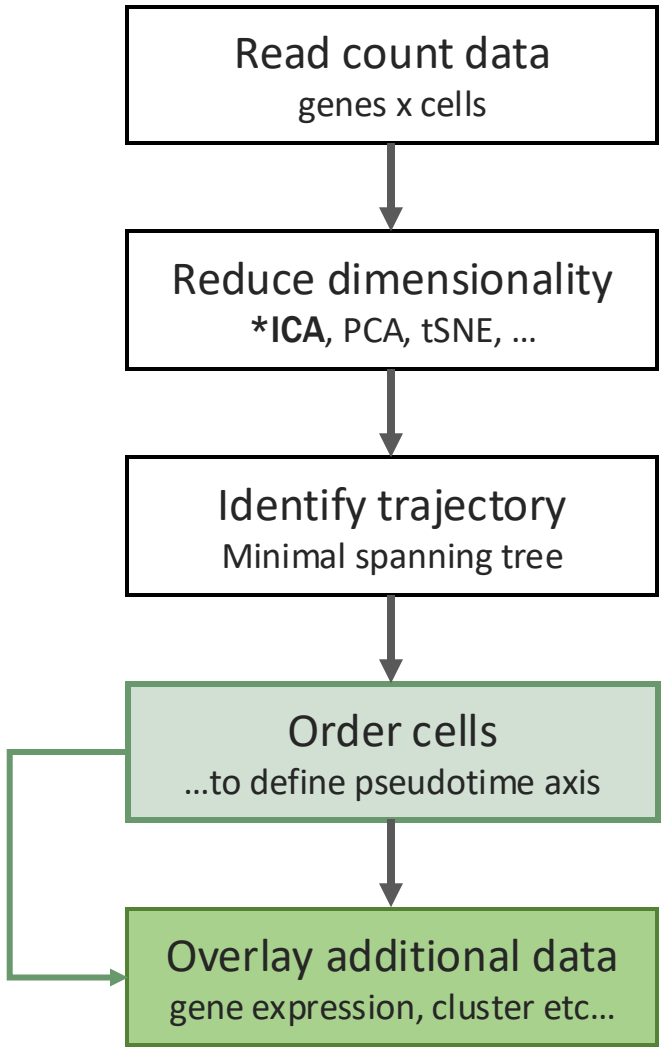
Investigate relationship between cell clusters and pseudotime



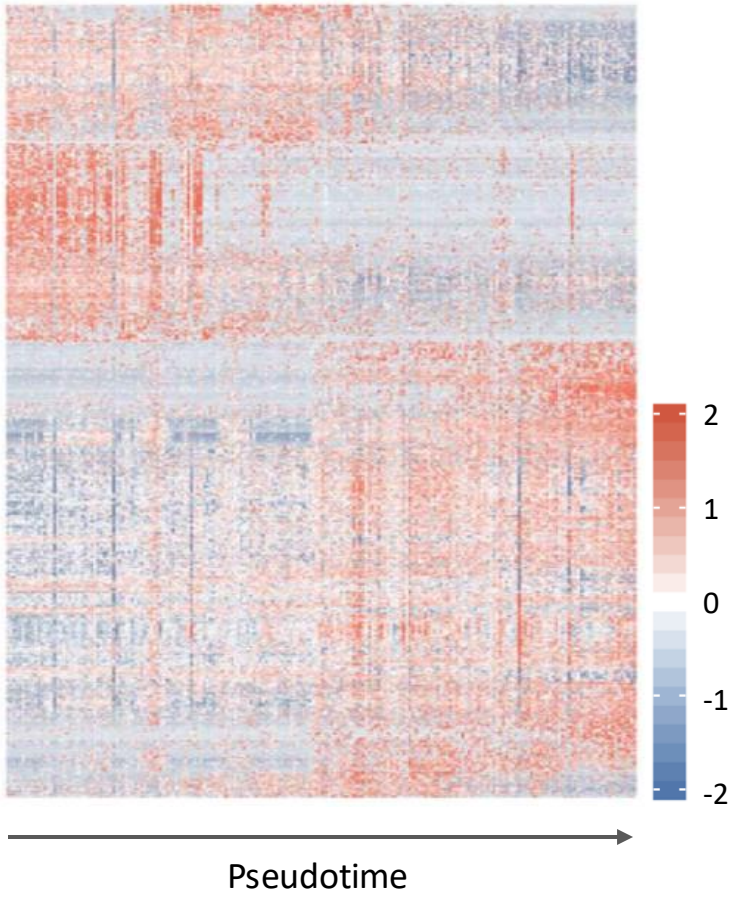
Reduce dimensionality
Build MST on cells
Order cells in pseudotime
Label cells by type

Taken from: Figure 2a | Trapnell et al., 2014. Nature Biotechnology, 32: 381–386.

Monocle: algorithm overview

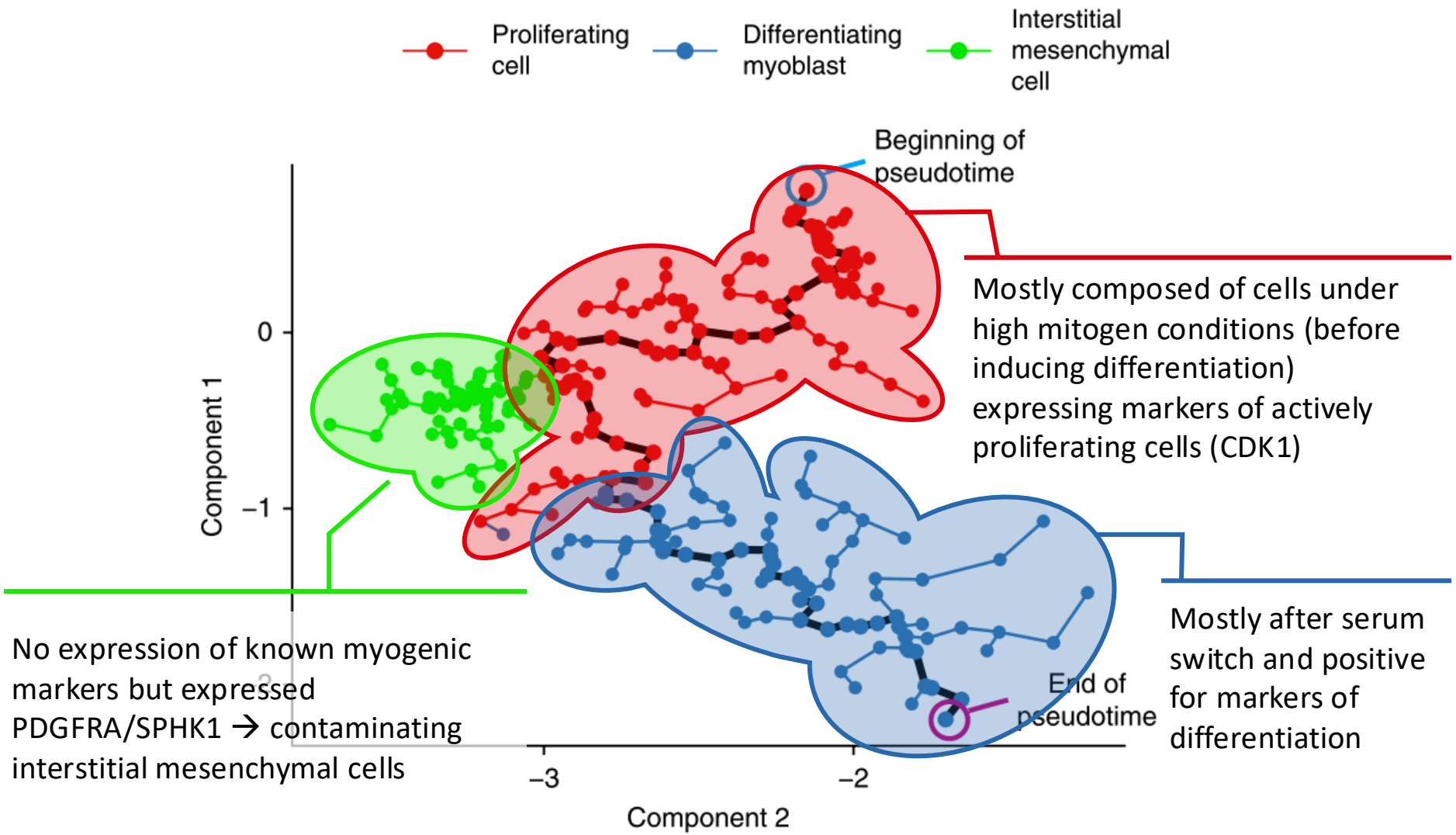


Model gene expression as a function of pseudotime



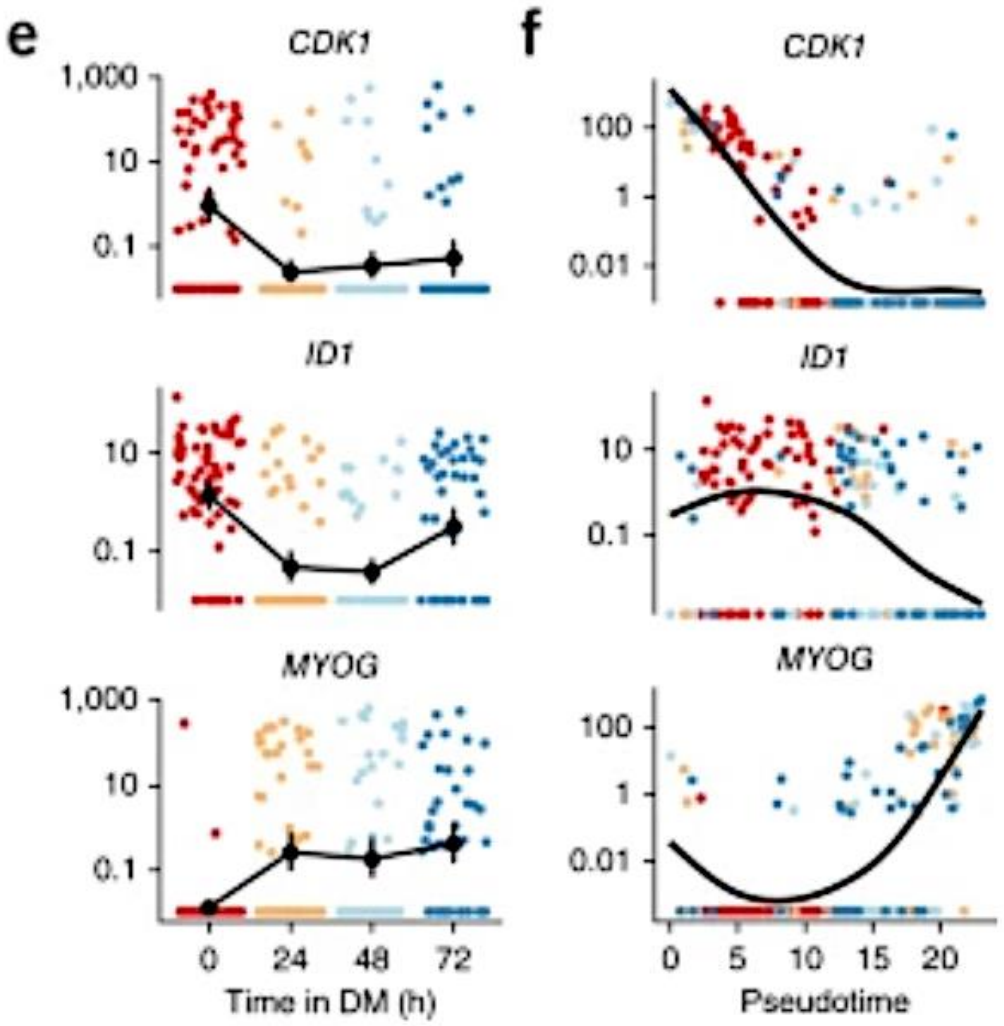
Taken from: Figure 2a | Trapnell et al., 2014. Nature Biotechnology, 32: 381–386.

Monocle: results and interpretation



Monocle able to construct this trajectory despite the presence of these contaminating cells (Not possible in bulk RNA-seq data)

Monocle: results and interpretation



Pseudotime in MVD

Monocle: results and interpretation

Interpretation

- overlay what you know
  function/pathway membership, other data
- overlay aspects of the data structure
  clustering, variability,
- identify genes potentially driving transition
  gene expression that correlates with pseudotime
  model gene expression as a function of pseudotime

Remember: *in silico* results are only **hypotheses**

Must validate *in vitro* (KD, RNAi, drug perturbation + functional studies)/*in vivo* (KO, transgenic models + functional studies)

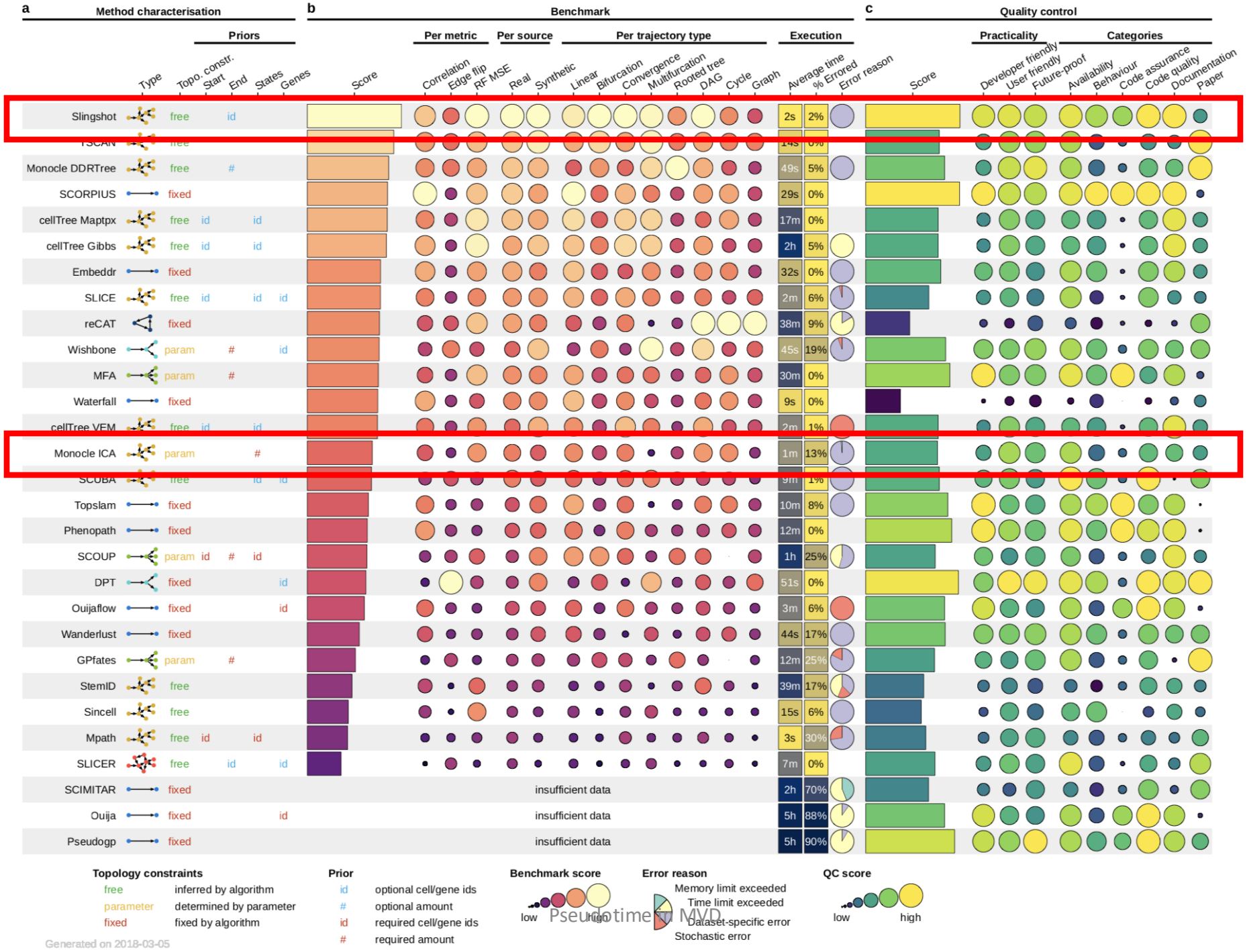
Other pseudotime methodologies

Benchmarking (Monocle DDRTree among the best)

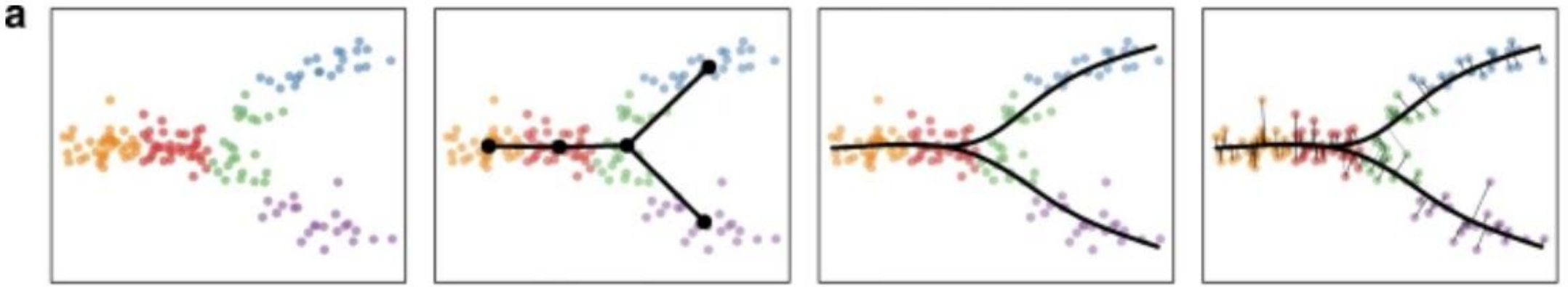
https://www.nature.com/articles/s41587-019-0071-9

(*Nature Biotechnology)*

29 methodologies assessed with real and simulated data


a
Method characterisation
Priors
Type
Topo. constr.
Start
End
States
Genes
b
Benchmark
Per metric
Per source
Per trajectory type
Execution
Score
Correlation
Edge flip
RF MSE
Real
Synthetic
Linear
Bifurcation
Convergence
Multifurcation
Rooted tree
DAG
Cycle
Graph
Average time
% Errored
Error reason
c
Quality control
Practicality
Categories
Score
Developer friendly
User friendly
Future-proof
Availability
Behaviour
Code assurance
Code quality
Documentation
Paper
Slingshot free id 2s 2% 14s 0% Monocle DDRTree free # 49s 5%
SCORPIUS fixed 29s 0%
cellTree Maptpx free id id 17m 0%
cellTree Gibbs free id id 2h 5%
Embeddr fixed 32s 0%
SLICE free id id id 2m 6%
reCAT fixed 38m 9%
Wishbone param # id 45s 19%
MFA param # 30m 0%
Waterfall fixed 9s 0%
cellTree VEM free id id 2m 1%
Monocle ICA param # 1m 13%
SCUBA free id id 9m 1%
Topslam fixed 10m 8%
Phenopath fixed 12m 0%
SCOUP param id # id 1h 25%
DPT fixed id 51s 0%
Ouijaflow fixed id 3m 6%
Wanderlust fixed 44s 17%
GPfates param # 12m 25%
StemID free 39m 17%
Sincell free 15s 6%
Mpath free id id 3s 30%
SLICER free id id 7m 0%
SCIMITAR fixed insufficient data 2h 70%
Ouija fixed insufficient data 5h 88%
Pseudogp fixed insufficient data 5h 90%
Topology constraints
free inferred by algorithm
parameter determined by parameter
fixed fixed by algorithm
Prior
id optional cell/gene ids
optional amount
id required cell/gene ids
required amount
Benchmark score
low high
Error reason
Memory limit exceeded
Time limit exceeded
Dataset-specific error
Stochastic error
QC score
low high
Generated on 2018-03-05
Pseudotime in MYD

Slingshot – same same?



Step 1: A minimum spanning tree is constructed on the clusters to determine the number and rough shape of lineages.

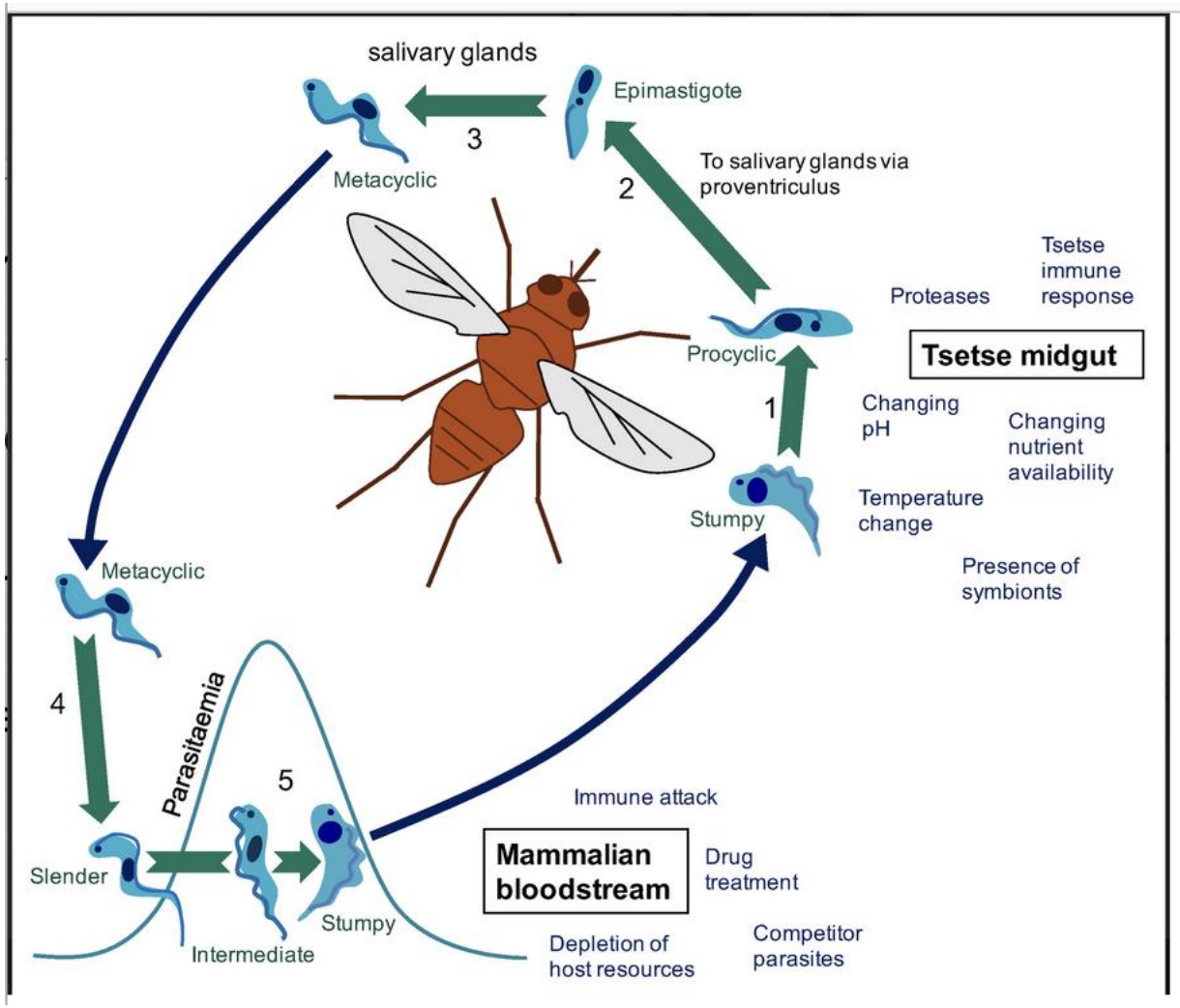
Step 2: Simultaneous principal curves are used to obtain smooth representations of each lineage.

Step 3: Pseudotime values are obtained by orthogonal projection onto the curves

https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-018-4772-0

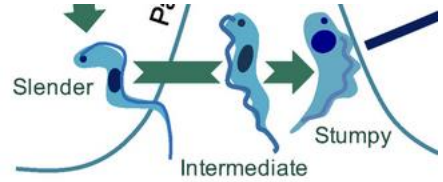
Example – *Trypanosoma brucei*

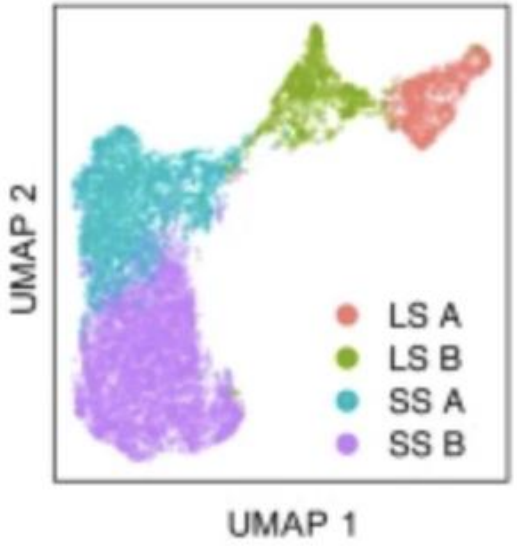
- Understand **slender to stumpy differentiation**
- Parasite Knock out
- Projected cell into PHATE space – a different manifold
- Trajectory build with slingshot
- TradeSeq to find genes expressed along the trajectory

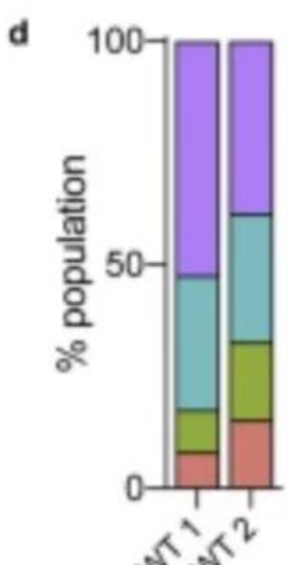


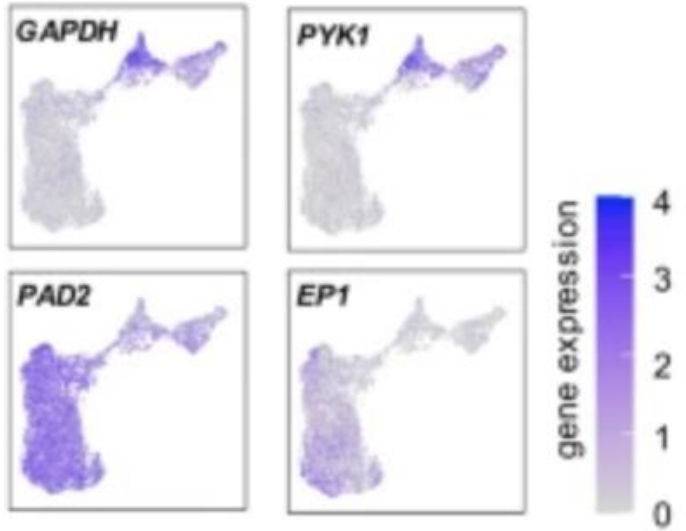
https://www.nature.com/articles/s41467-021-25607-2

Analysis of wild type

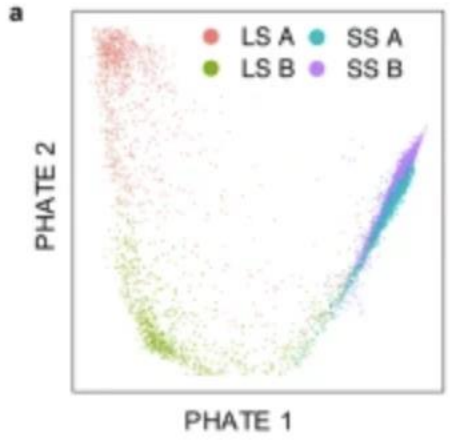


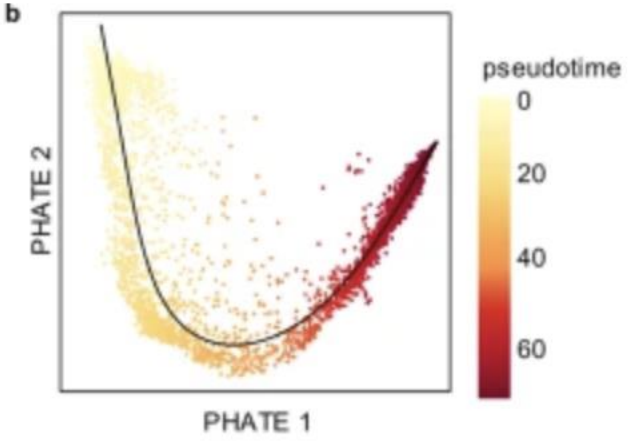


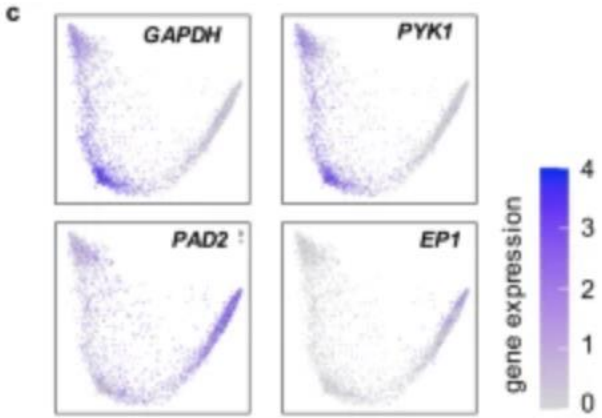




Slingshot works better in PHATE space





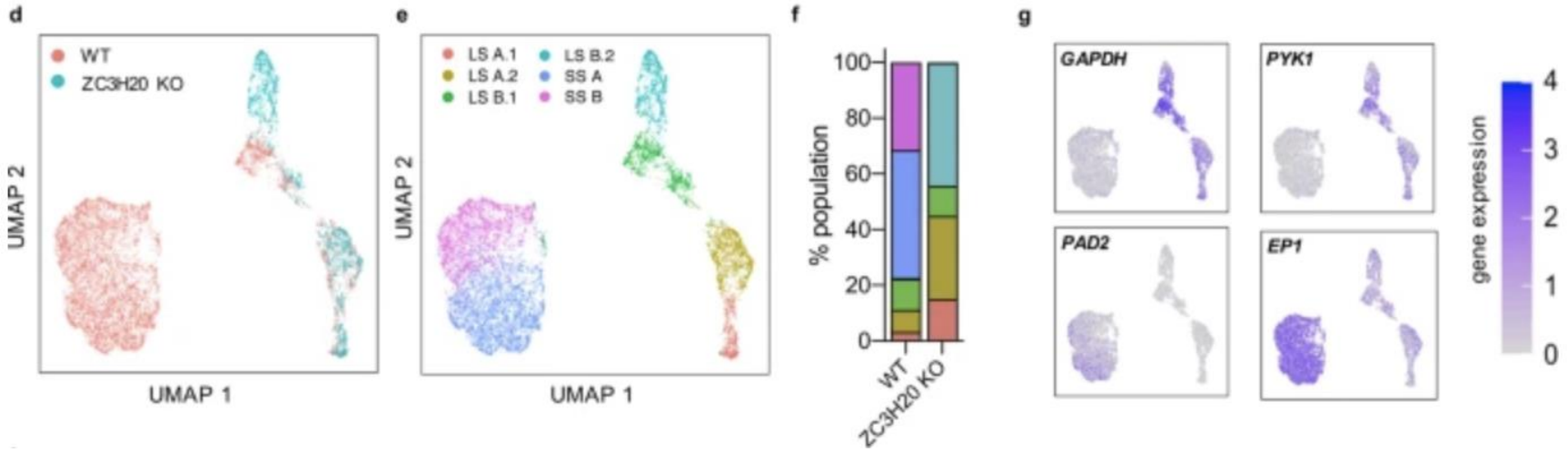


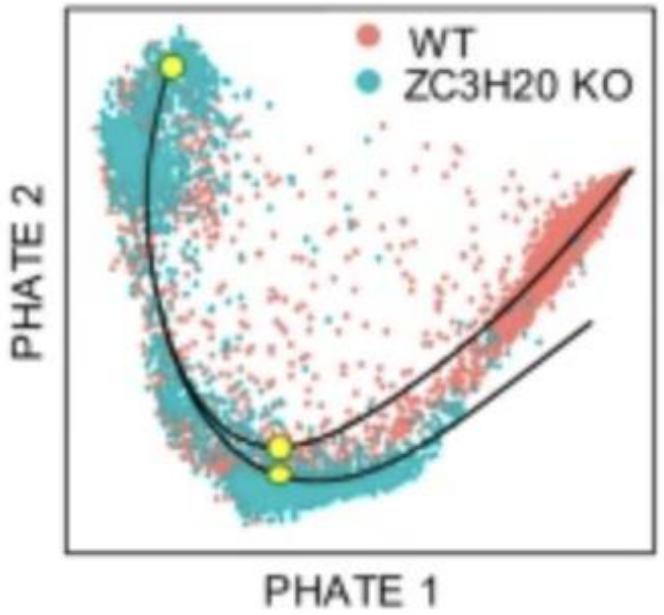
Why PHATE over UMAP for pseudotime

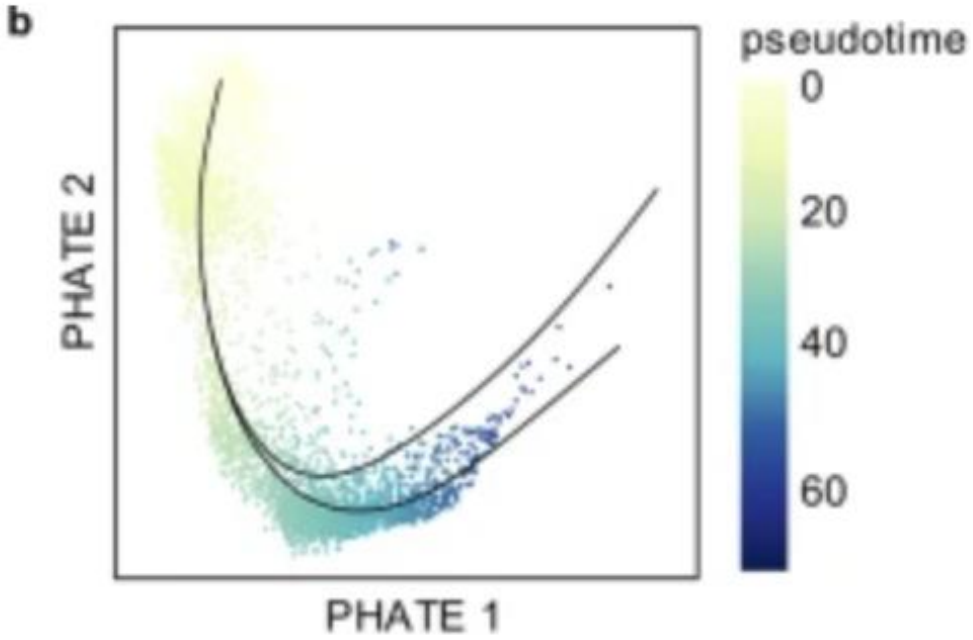
- Again:

PHATE captures better global distances, therefore it is better for processes

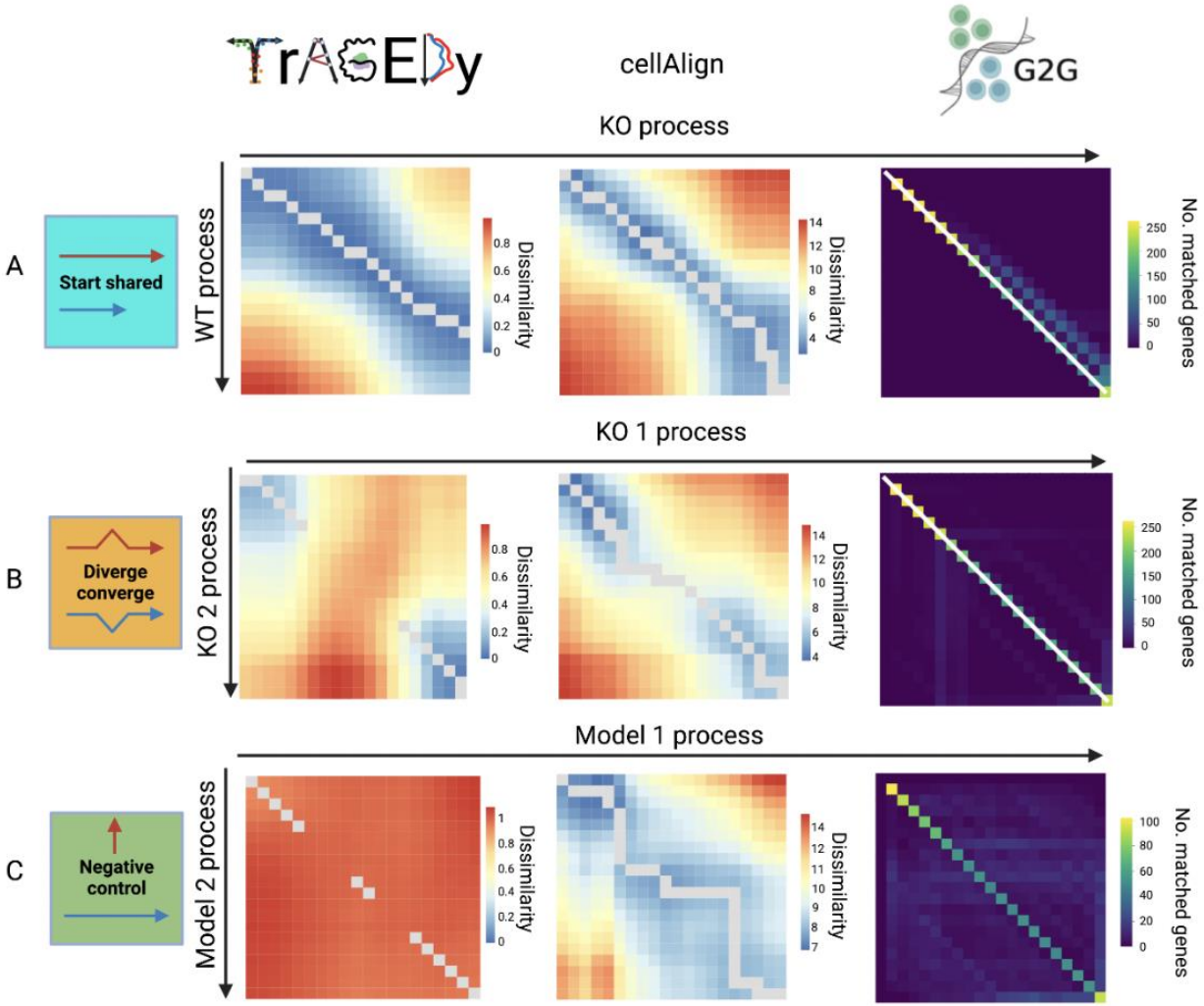
ZC3H20 KO

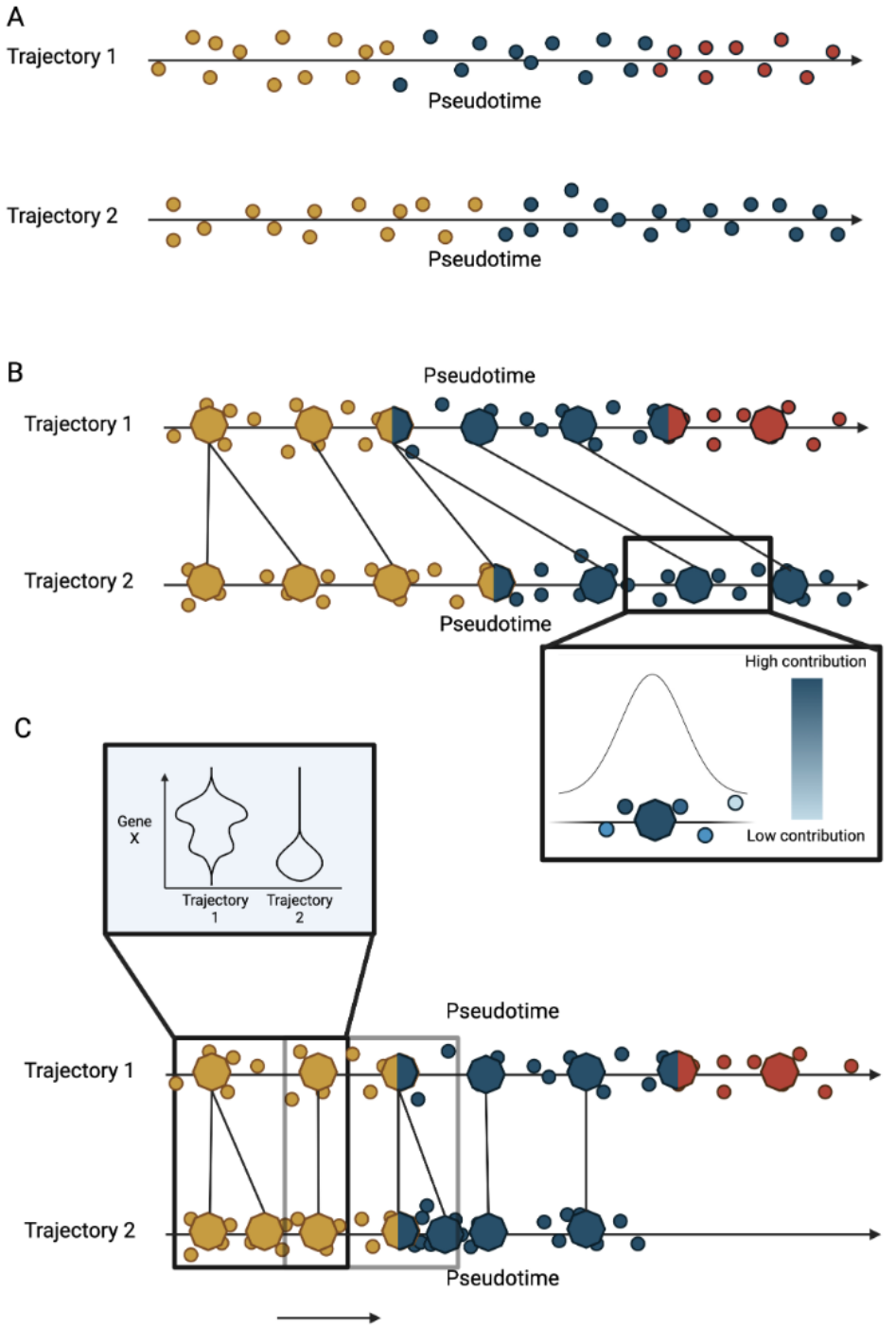

d
WT
ZC3H20 KO
UMAP 2
UMAP 1
e
LS A.1
LS A.2
LS B.1
LS B.2
SS A
SS B
UMAP 2
UMAP 1
f
% population
100
80
60
40
20
0
WT
ZC3H20 KO
g
GAPDH
PYK1
PAD2
EP1
gene expression
4
3
2
1
0


WT
ZC3H20 KO
PHATE 2
PHATE 1

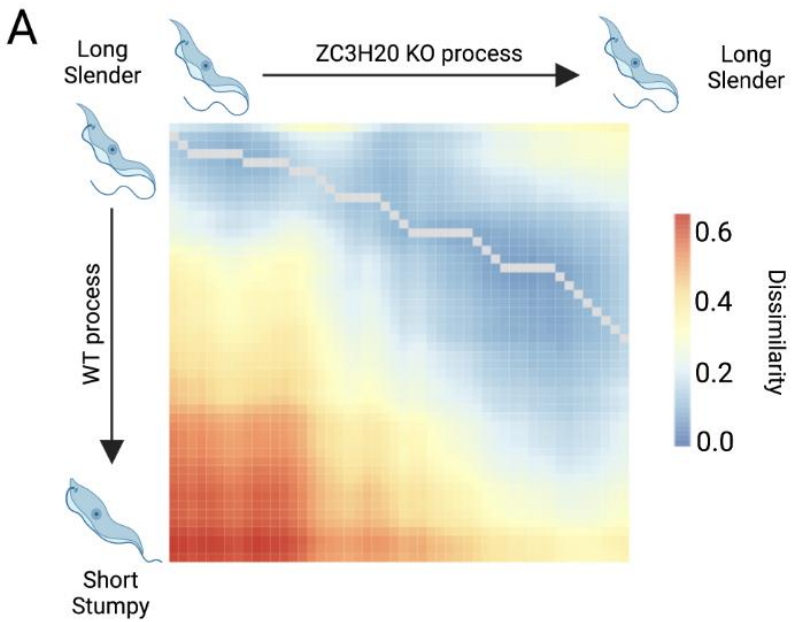

b
PHATE 2
PHATE 1
pseudotime
0
20
40
60

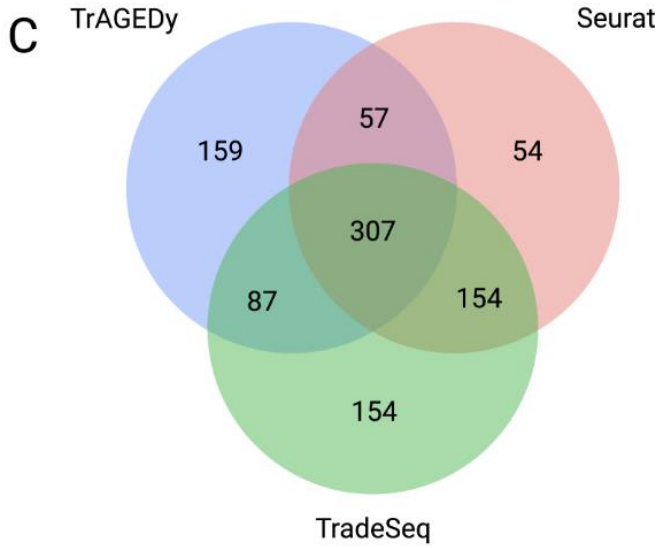
Tragedy – aligning trajectories

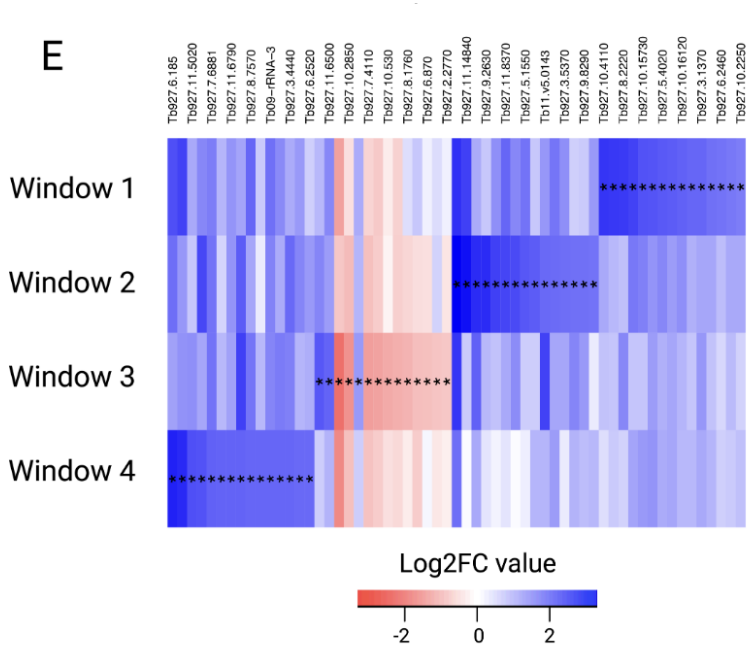


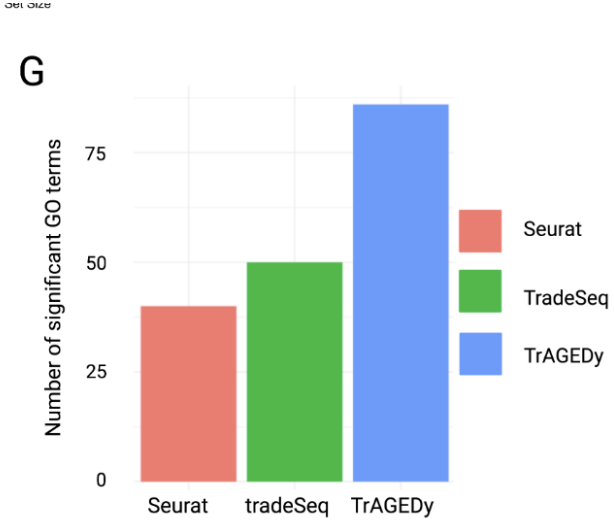


Results









JOURNAL ARTICLE ACCEPTED MANUSCRIPT

TrAGEDy—Trajectory Alignment of Gene Expression Dynamics

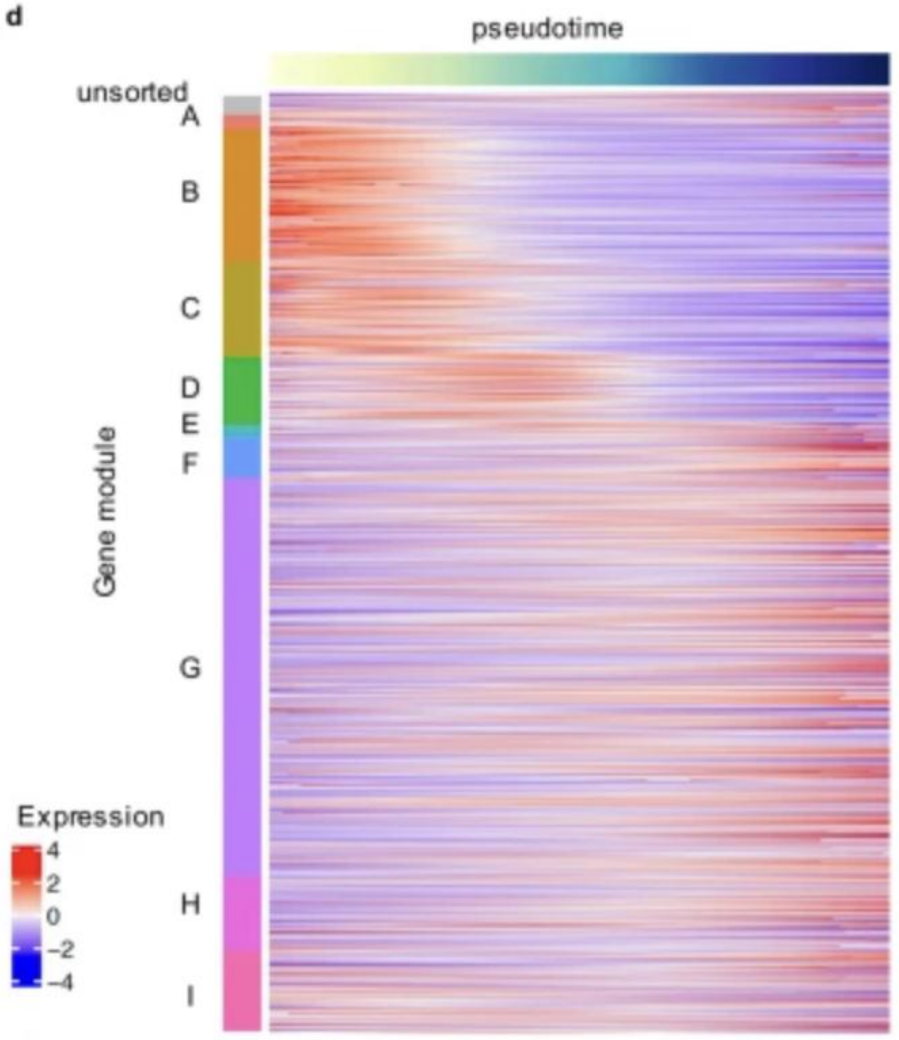
Ross F Laidlaw, Emma M Briggs, Keith R Matthews, Amir Madany Mamlouk, Richard McCulloch, Thomas D Otto

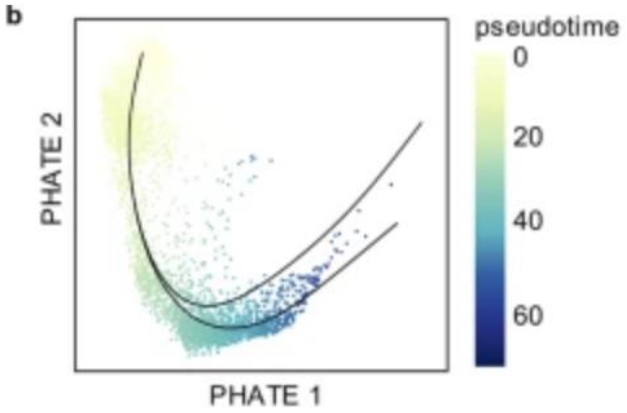
Bioinformatics, btaf073, https://doi.org/10.1093/bioinformatics/btaf073

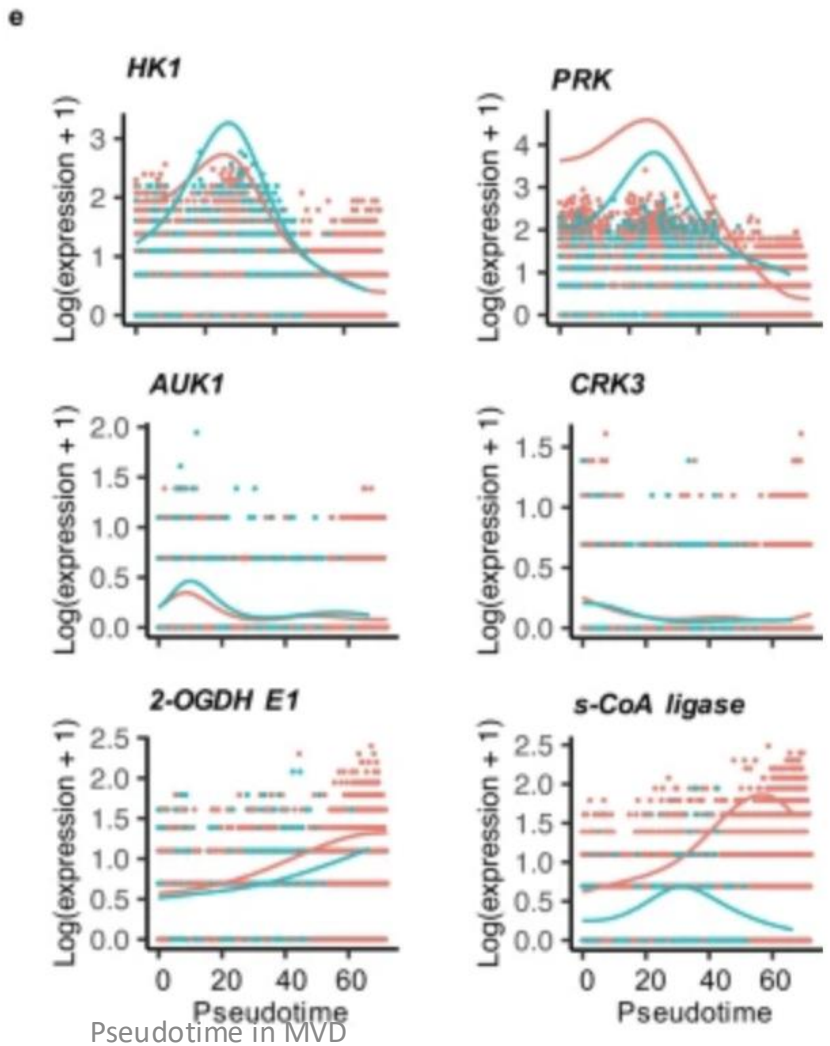
Published: 11 March 2025 **Article history** ▾

Pseudotime in MVD

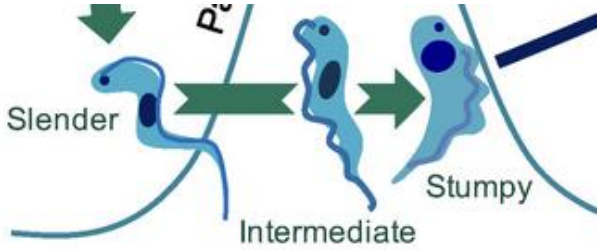
Expression along the trajectory







Conclusions paper



- Slingshot was able to detect two trajectories
- TradeSeq visualised genes expressed between the two trajectories

- **Trajectories project into PHATE space – is the better reduction than UMAP**

- KO did keep parasites in Slender form!

nature

Letter | Published: 08 August 2018

RNA velocity of single cells

Gioele La Manno, Ruslan Soldatov, Amit Zeisel, Emelie Braun, Hannah Hochgerner,

- RNA velocity—the time derivative of the gene expression state
- Distinguish nascent (unspliced) and mature (spliced) mRNA in common single-cell RNA sequencing protocols
- Expect RNA velocity to greatly aid the analysis of developmental lineages and cellular dynamic

RNA velocity: spliced versus unspliced RNA

Fig. 1: Balance between unspliced and spliced mRNAs is predictive of cellular state progression.

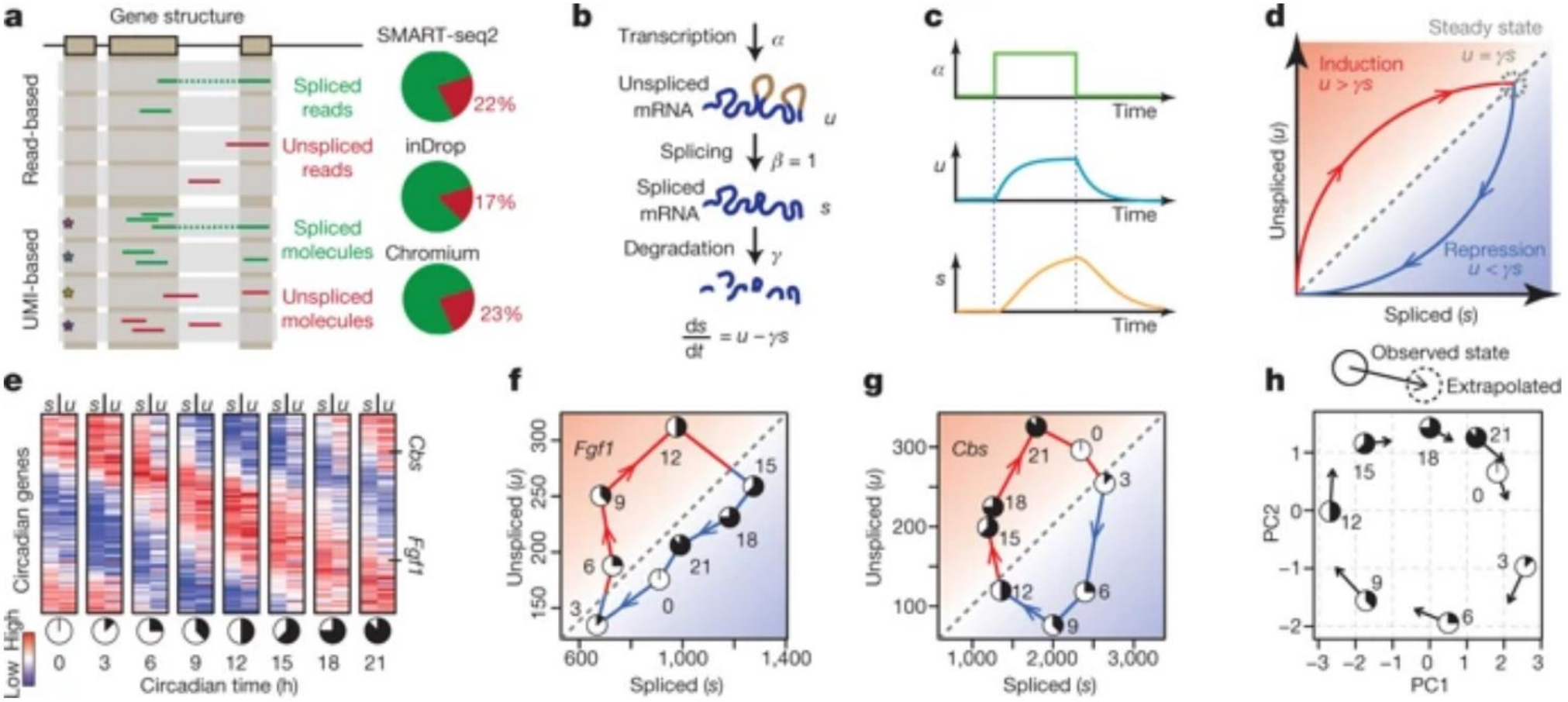
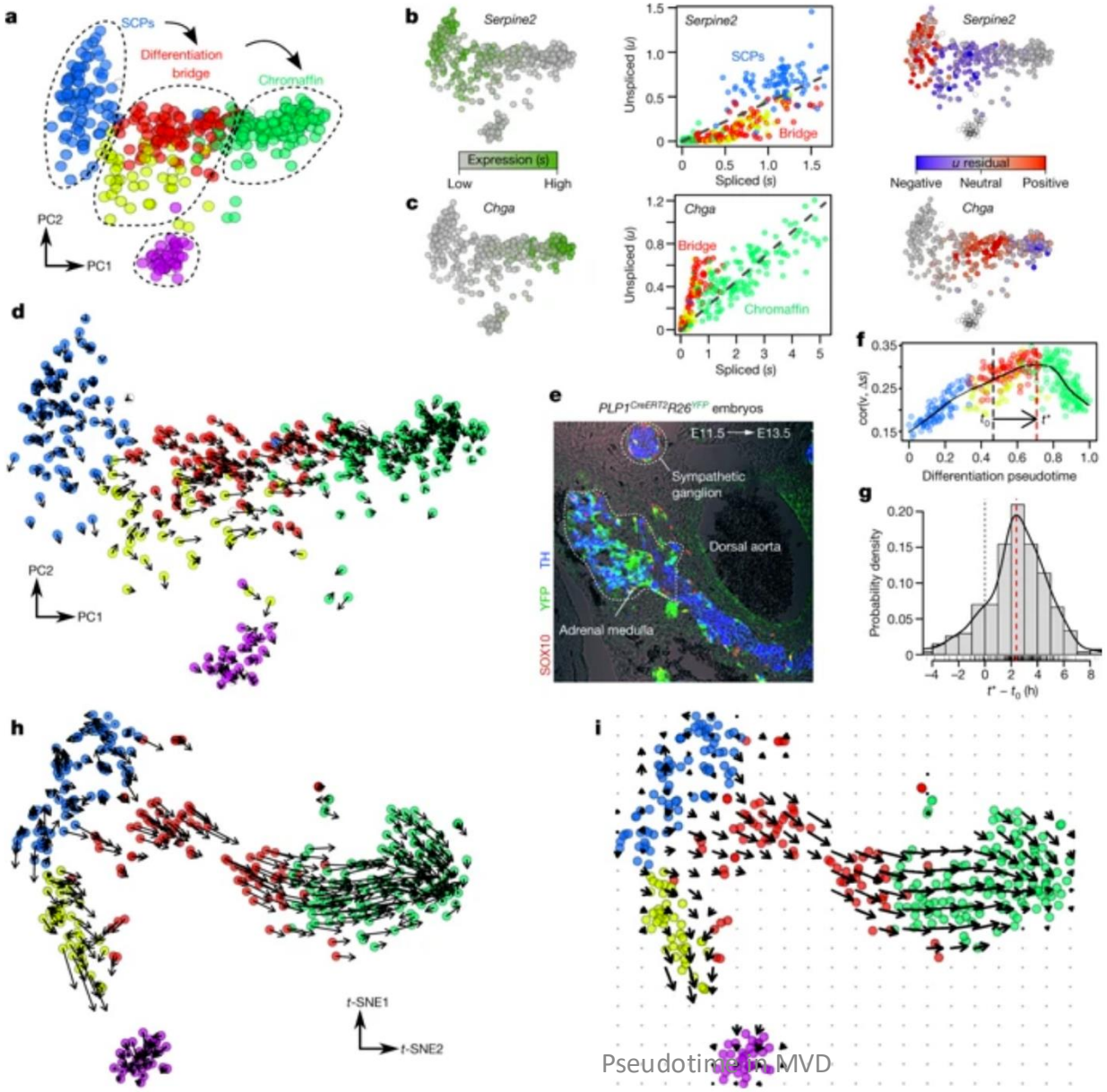


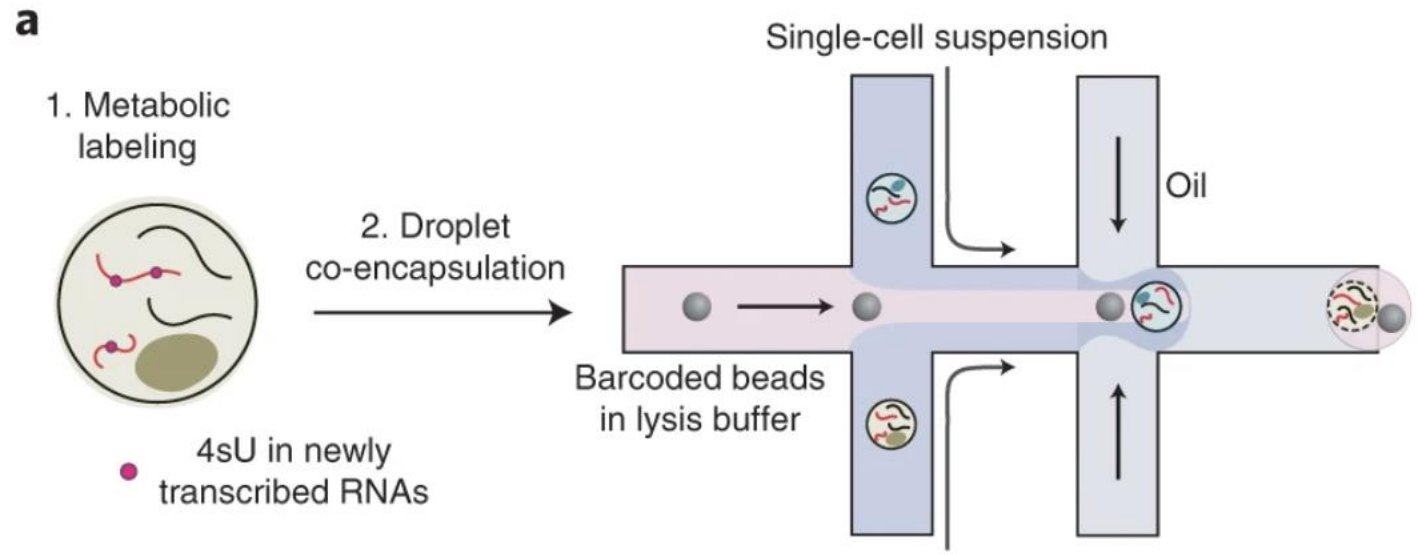
Fig. 2: RNA velocity recapitulates dynamics of chromaffin cell differentiation.



How to capture time (dynamic) better - Metabolic labelling

Single-cell RNA sequencing offers snapshots of whole transcriptomes but obscures the temporal RNA dynamics

marking newly transcribed mRNAs with T-to-C substitutions



https://www.nature.com/articles/s41592-020-0935-4

https://www.youtube.com/watch?v=ckjN1Nud3oM

IMPORTANT

- Don’t just apply algorithms! They always give you something back!

- Are clusters ”connected”?
- Can you show that there is a process?
- Are these same cell types?

- **PHATE space – is the better reduction than UMAP**

Conclusions

single cell RNA-Seq is noisy

Need to relate the finding to “biology”

Many methods exists

Still open problems: How good is integration? DE!

Vehicle to understand your biological question

Pseudo time is powerful to find dynamics

Several exiting new tools

Q&A

- Please ask questions, so far
  - Lecture
  - Practicals

Learning aims

- Introduce more technical concepts
- Explore how to process scRNA-Seq data
- Learn new methods, normalization, integration and DE in scRNA-Seq
- Critical evaluation of scRNA-Seq
- Overview of developments in scRNA-Seq


University of Glasgow
Questions?
#UofGWorldChangers
@UofGlasgow