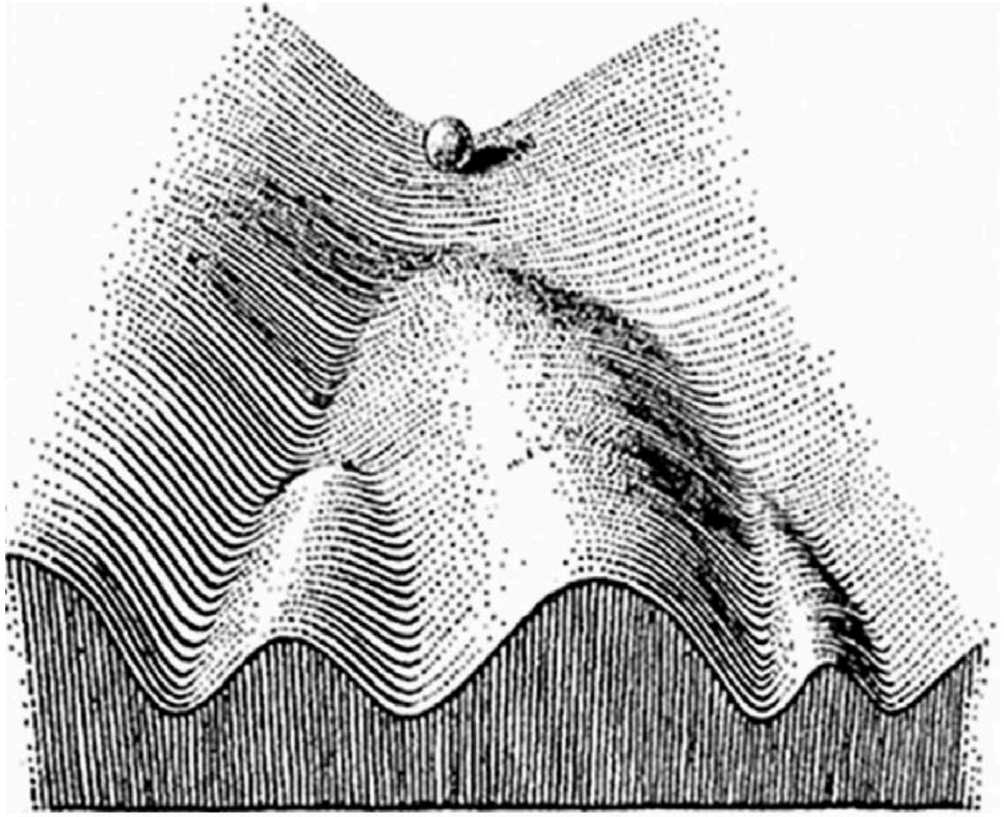
Coherent sparse optimization predicts early cell fate bias by integrating lineage and transcriptome information

Dr. Shou-Wen Wang
Damon Runyon Fellow

Allon Klein Laboratory
Department of Systems Biology
Harvard Medical School

Oct 19, 2022



About me

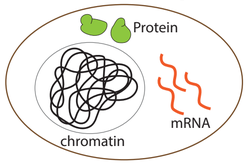
PhD in Statistical Physics, Tsinghua University, China
* Implications of hidden variables
* Physics of collective oscillations

Computational Biology Postdoc, Allon Klein lab @ Harvard
* Learning cell fate choice with lineage tracing

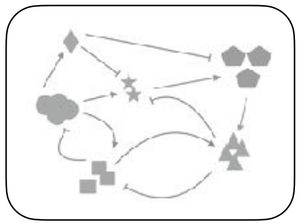
Assistant professor @ Westlake University in China since 2023

Cell fate choice is a complex process

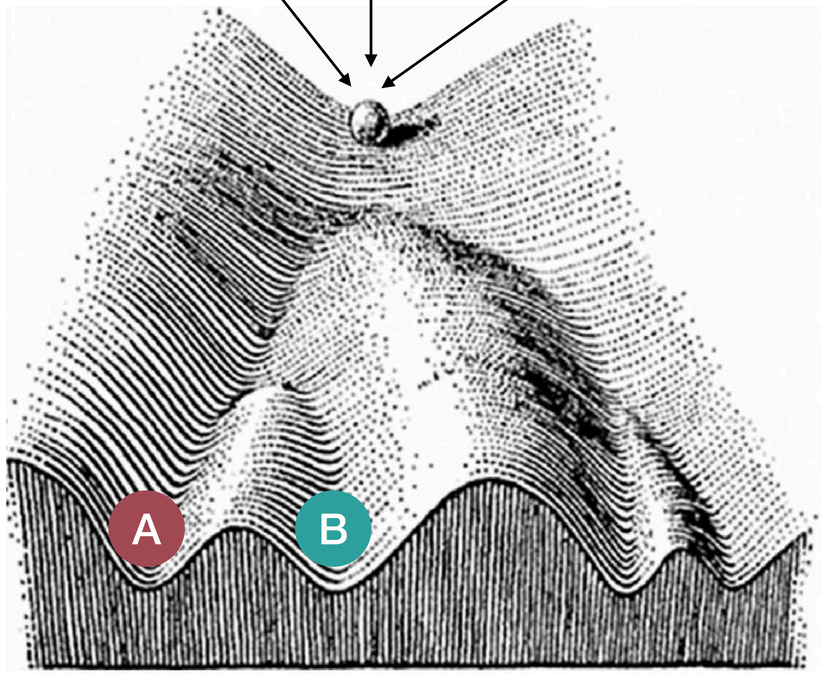
Internal input



Regulation



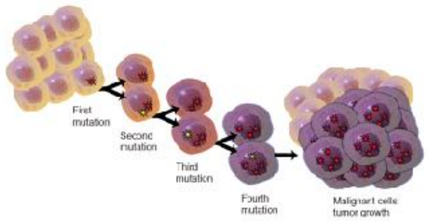
External input



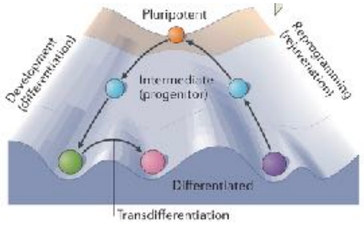
Waddington landscape

Development

Cancer



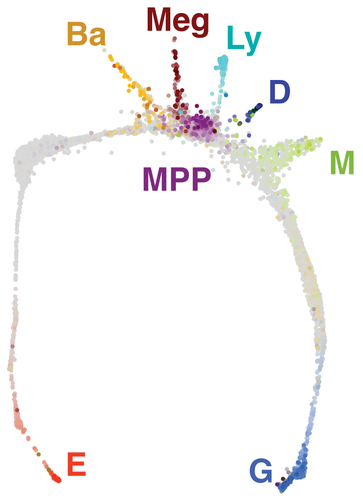
Cell engineering



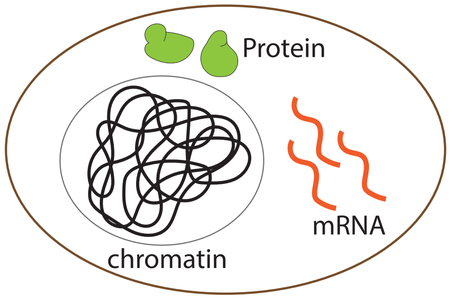
Takahashi & Yamanaka, Nat. Rev. Mol. Cell Biol. (2016)

Single-cell assays allow us to systematically measure all cell states of a system

High-throughput scRNA-seq



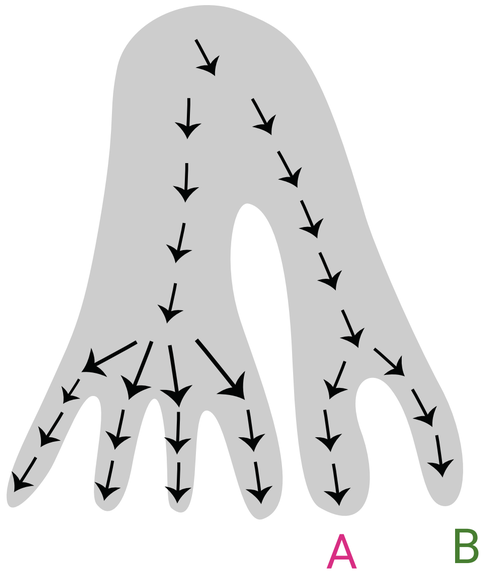
Multi-omics



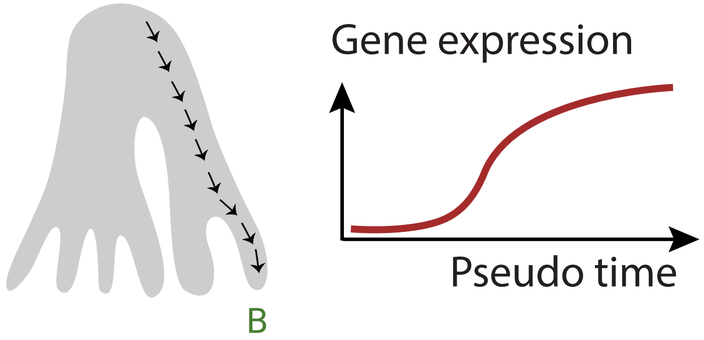
Klein *et.al.*, Cell (2015)
Macosko,…,McCarroll, Cell (2015)
Tusi/Wolock/Weinreb *et. al.,* Nature (2018)
Ma, …, Buenrostro,Cell 2020
Li,.., Tang, Cell Research (2020)
Rodriques, …, Macosko, Science (2019)

Single-cell genomics offers opportunities to learn differentiation dynamics

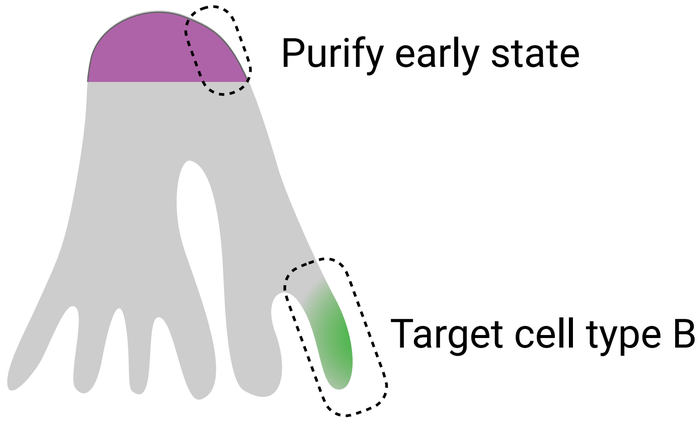
Differentiation dynamics



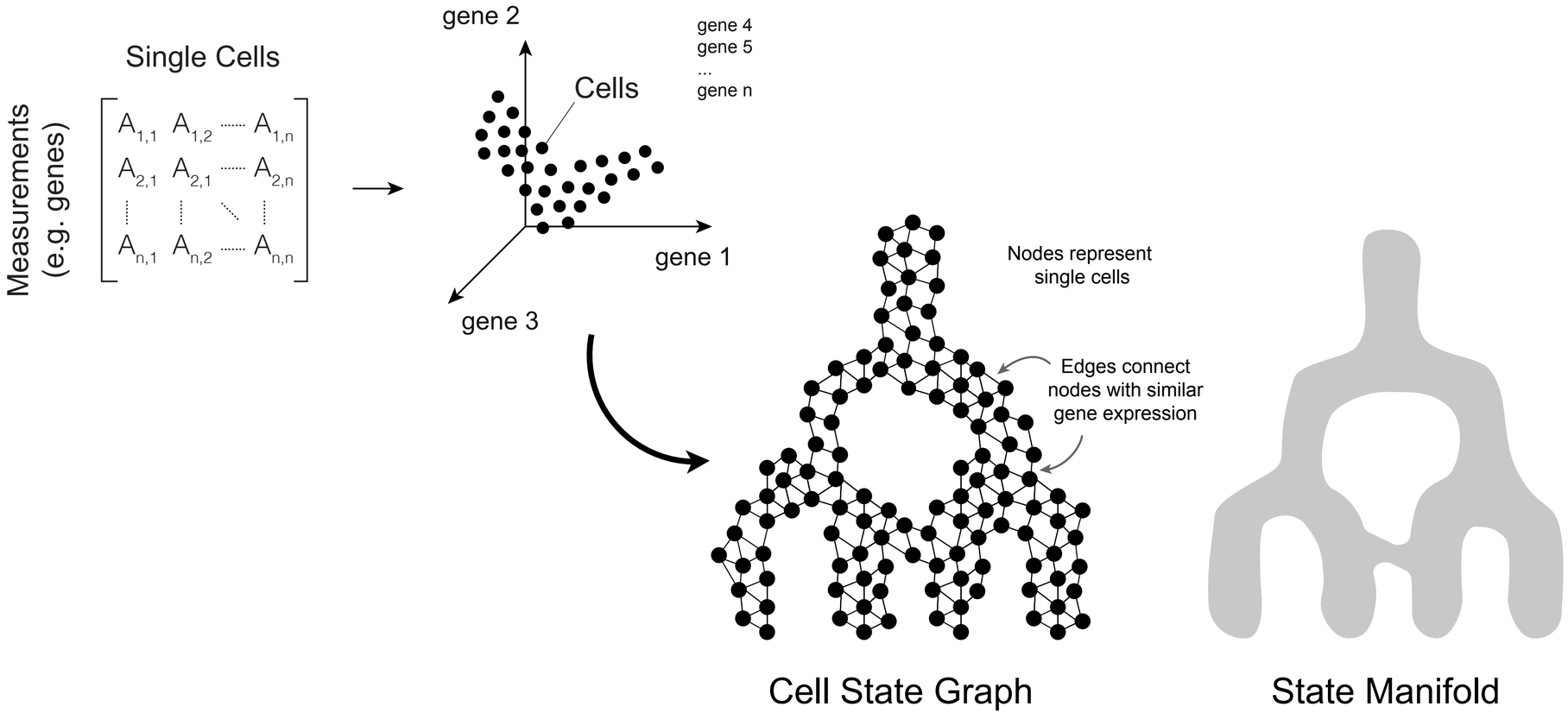
Understand gene regulation for fate choice



Guide cell-type engineering

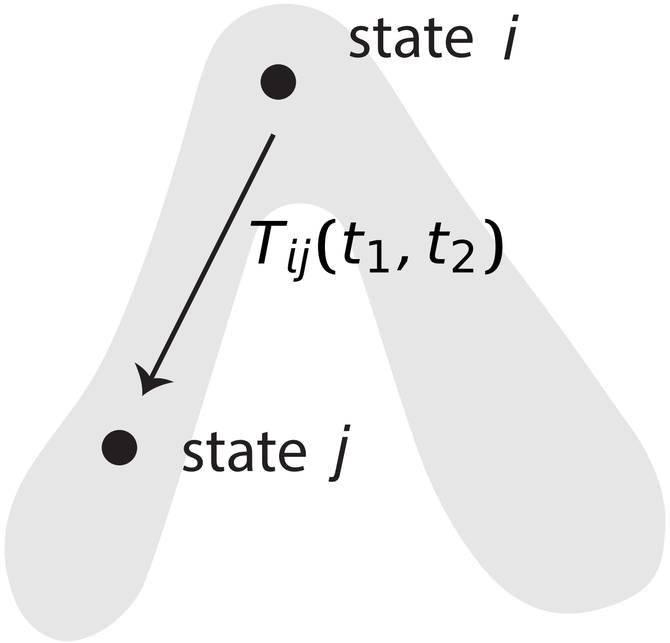


Embedding high-dimensional data to learn dynamics



Wagner and Klein, *Nat Rev Genetics* 2020

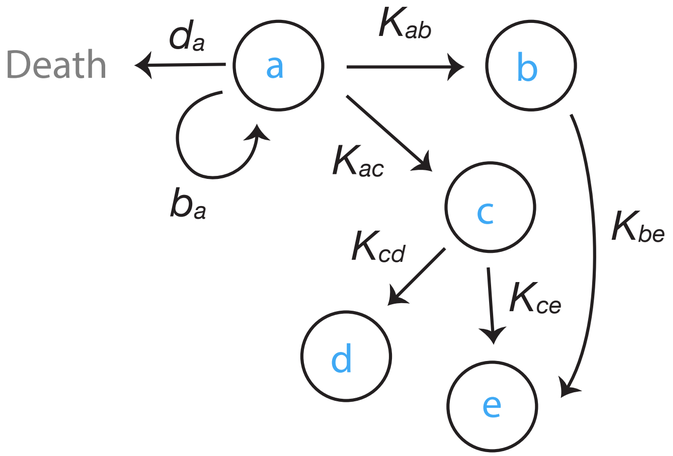
Dynamic inference as learning the transition probability matrix

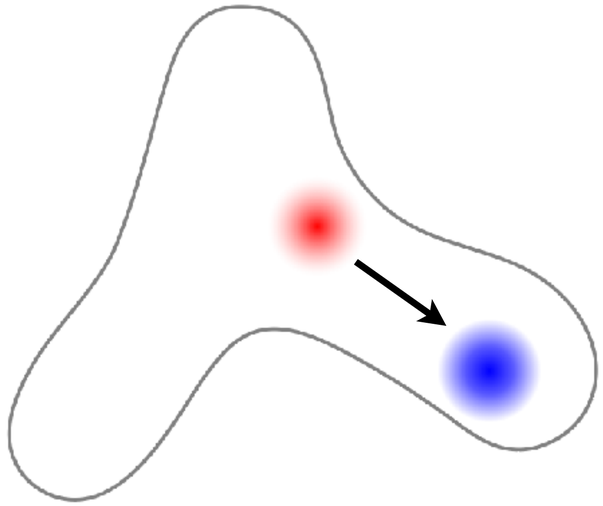


Schiebinger, Shu et.al., Cell, 2019

Linking transition map to stochastic processes of cell division, death, and differentiation

Instantaneous state dynamics





Cell density: $P_i(t) \equiv \frac{\bar{N}_i(t)}{\sum_i \bar{N}_i(t)}$

$$\vec{P}(t_2) = \vec{P}(t_1) T(t_1, t_2)$$

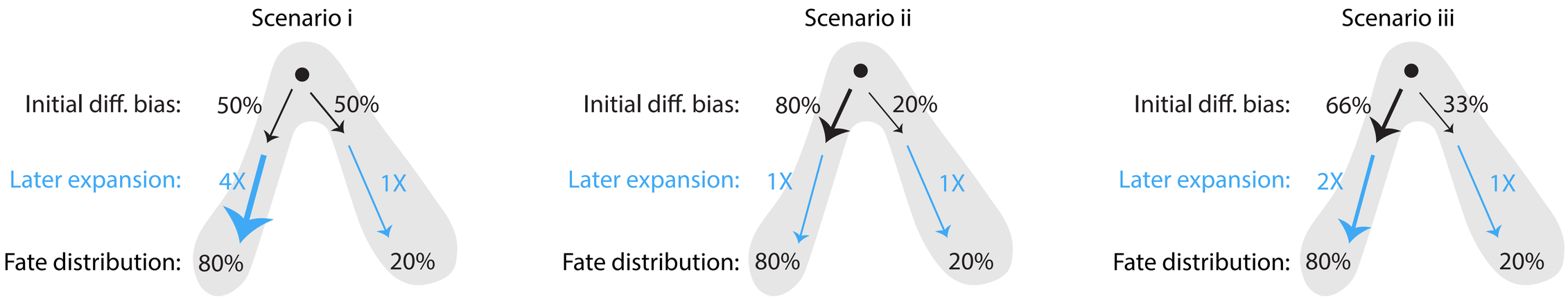
$$T(t_1, t_2) = \exp\left(\int_{t_1}^{t_2} \tilde{K} dt \right)$$

Cell number

$$\frac{d\bar{N}_i}{dt} = \sum_j \bar{N}_j K_{ji}$$

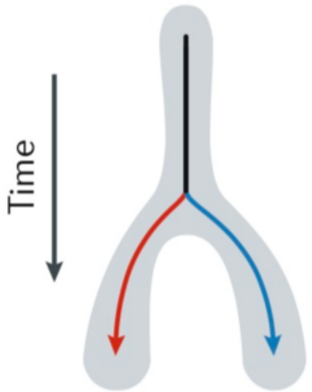
$$K_{ii} = b_i - d_i - \sum_{m \neq i} k_{im}$$

Transition probability does not distinguish differentiation bias and expansion bias

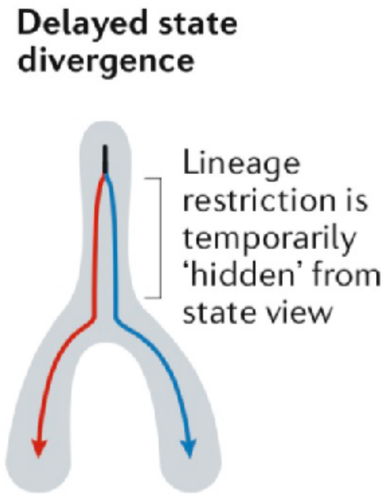


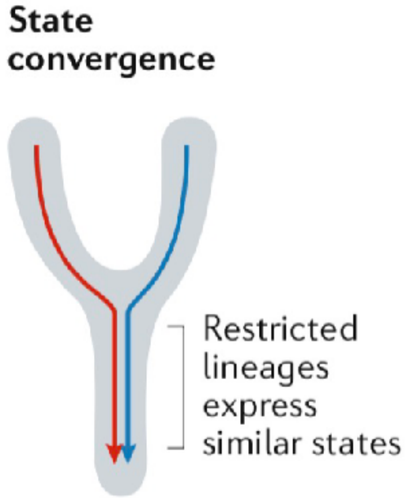
Challenges in dynamic inference from snapshot scRNAseq data

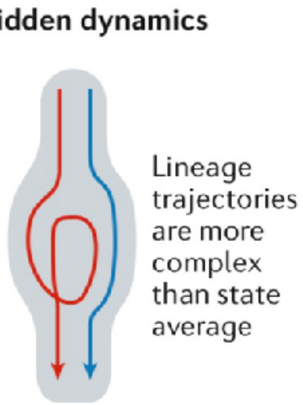
Lineage and state are consistent



Lineage and state diverges

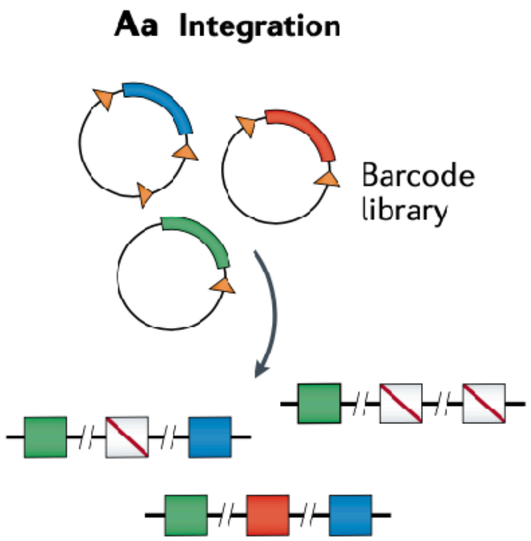


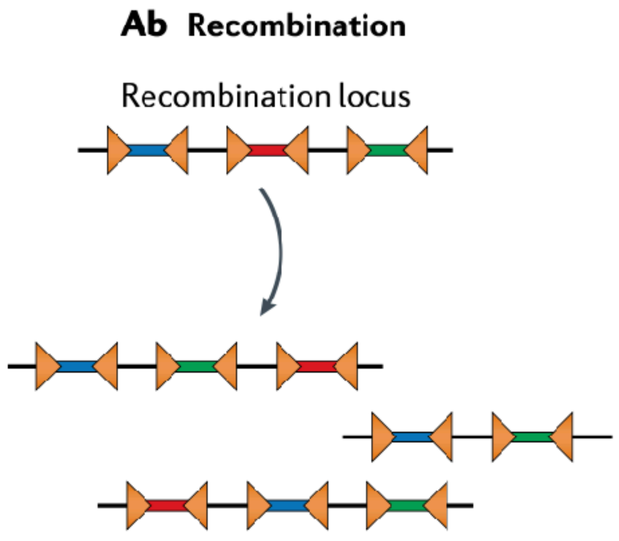


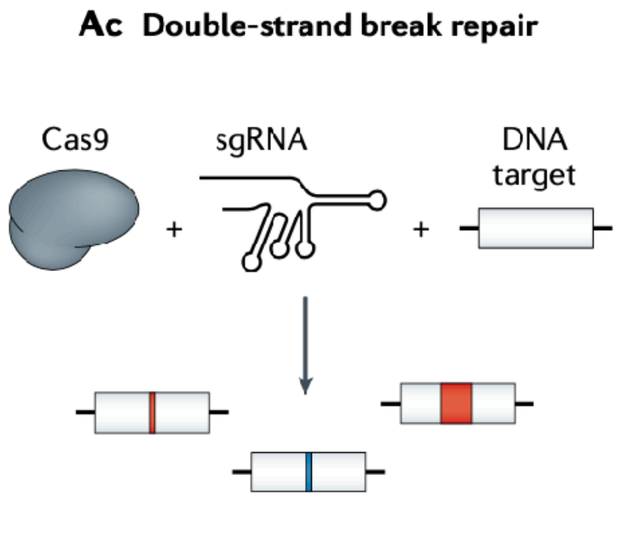


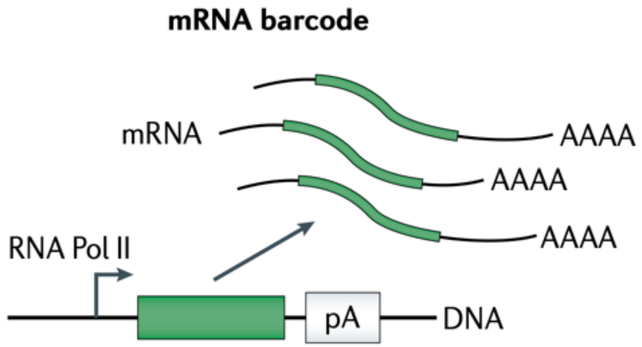
Weinreb et.al., PNAS, 2018
Wagner & Klein, Nat. Gen., 2020

A range of methods allow lineage-tracing with single cell genomics





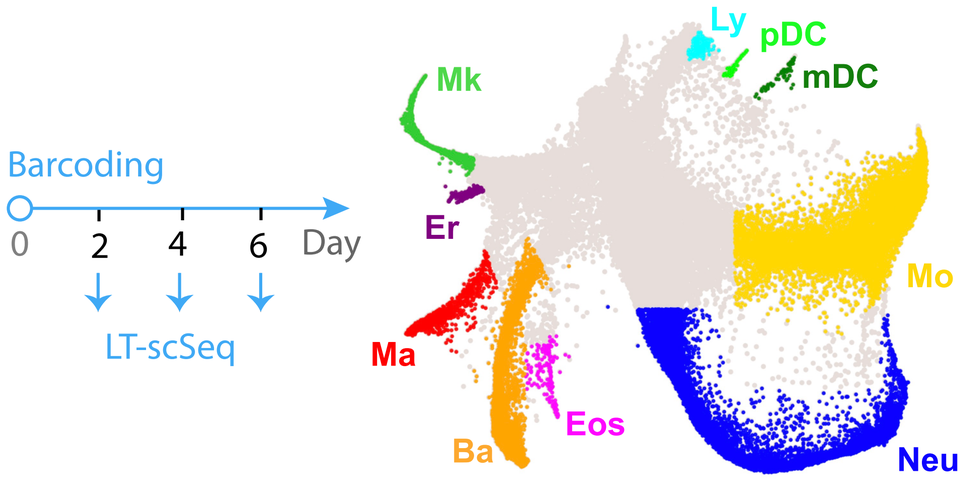


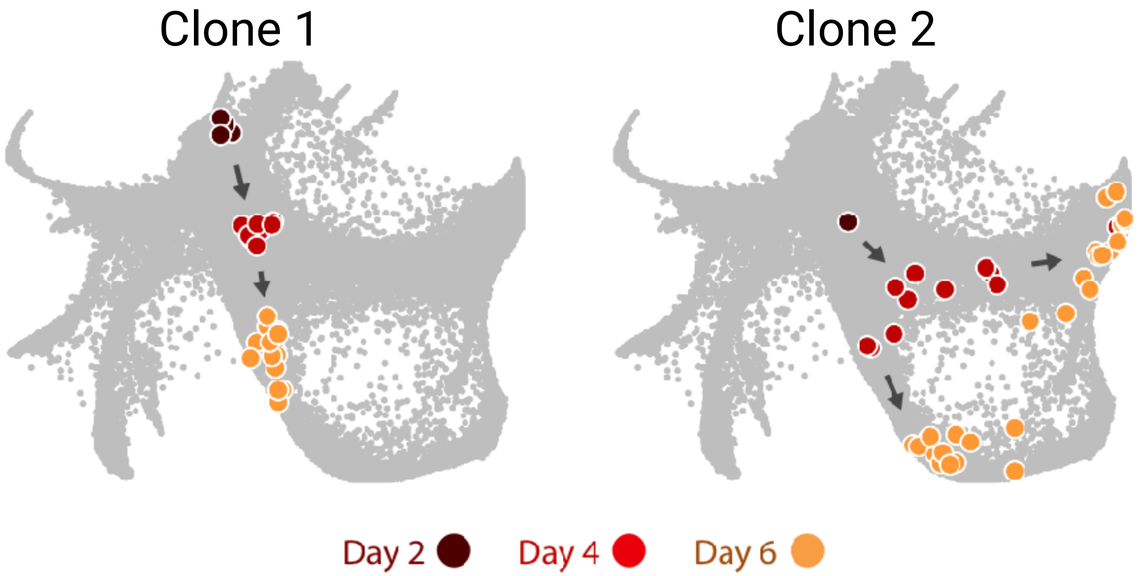


Lineage-Tracing scRNAseq (LT-scSeq)

Wagner and Klein, *Nat Rev Genetics* (2020)

Example: Lineage tracing to study *in vitro* hematopoiesis

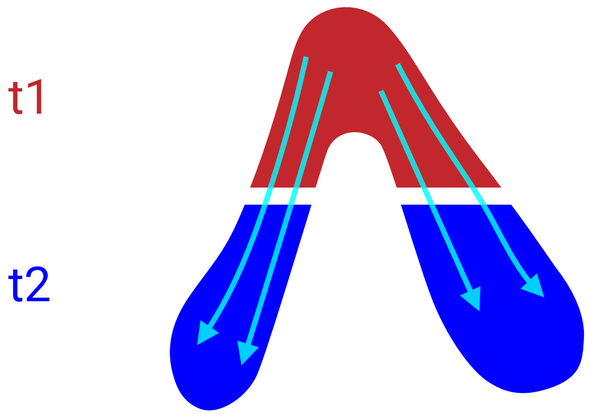




C. Weinreb, A. Rodriguez-Fraticelli et.al., Science 2020

How to constrain dynamic inference with lineage tracing?

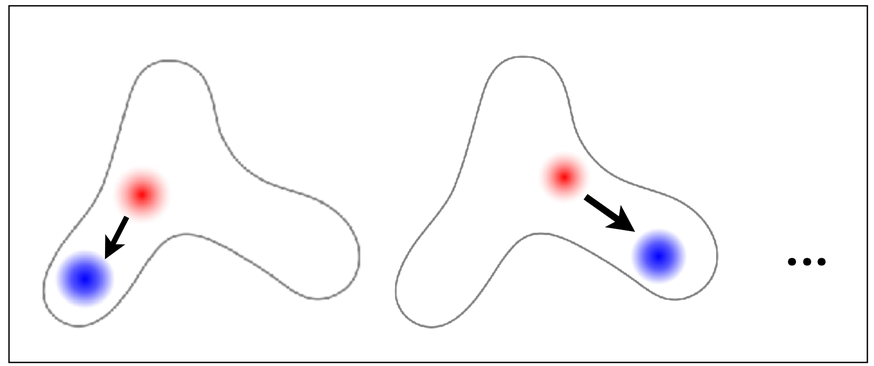
Without lineage tracing
(e.g., using Optimal Transport)



Schiebinger, Shu et.al., Cell, 2019
Nitzan et.al., Nature, 2019

With lineage tracing

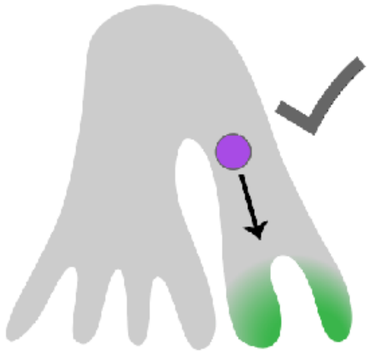
SuperOT: Prasad, Yang & Uhler, arXiv. 2020
LineageOT: Forrow & Schiebinger, Nat. Comm. 2021



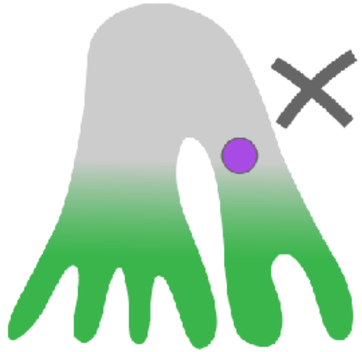
$$\sum_m ||\vec{P}(t_2; m) - \vec{P}(t_1; m)T(t_1, t_2)||_2 < \epsilon$$

(Under-determined)

The transition map is sparse



Sparse
(likely)

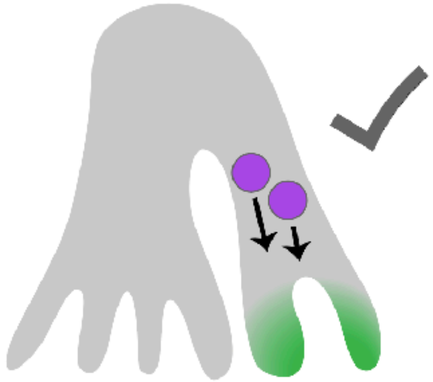


Non-sparse
(less likely)

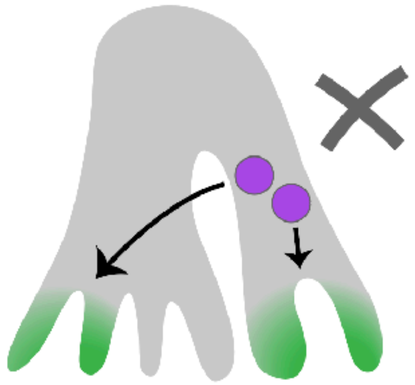
$$\min_{T} \underbrace{\|T\|_1}_{\text{Sparsity}}$$

L1-norm of a matrix: $||T||_1 = \sum_{ij} |T_{ij}|$

The transition map is coherent



Coherent
(likely)



Incoherent
(less likely)

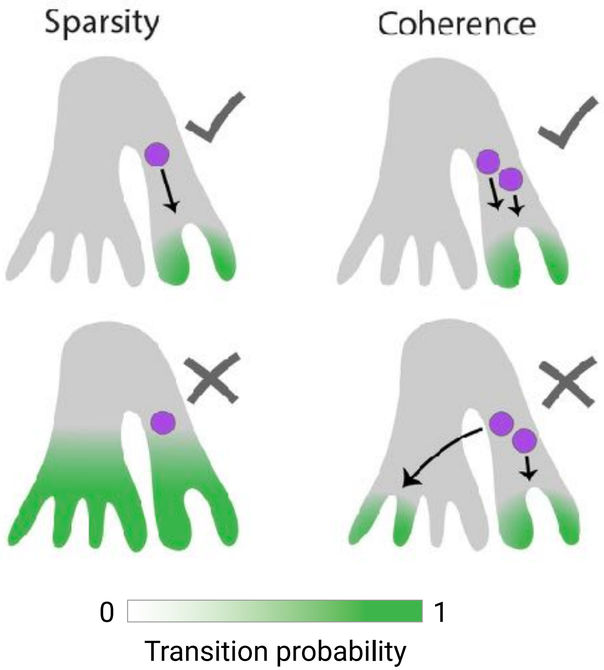
$$\min_T \underbrace{\alpha \|LT\|_2}_{\text{Coherence}}$$

L: graph laplacian

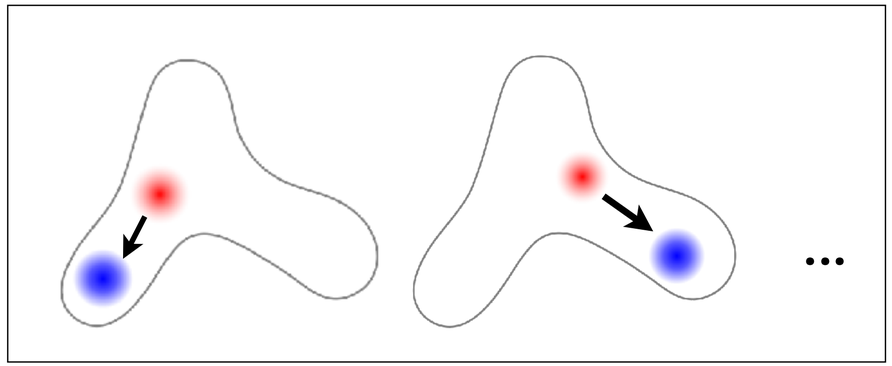
L2-norm (*Frobenius norm*) for a matrix

$$||T||_2 = (\sum_{ij} |T_{ij}|^2)^{1/2}$$

CoSpar: Coherent Sparse optimization

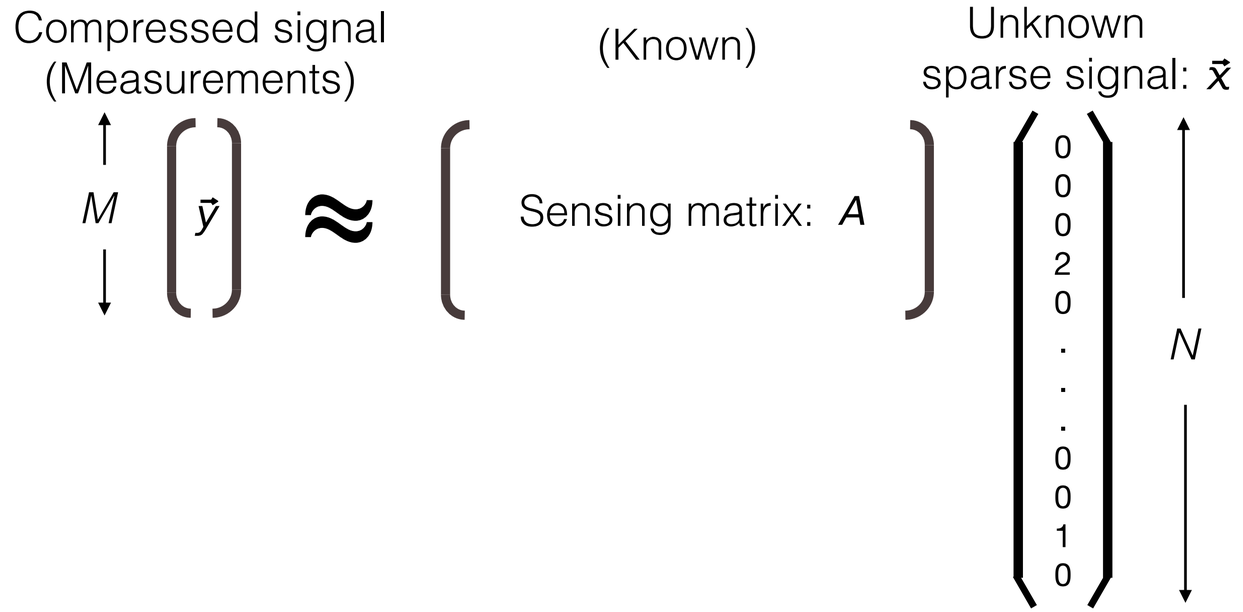


Clonal constraints



$$\min_{T} \underbrace{\|T\|_1}_{\text{Sparsity}} + \underbrace{\alpha\|LT\|_2}_{\text{Coherence}}, \qquad s.t. \underbrace{\|\mathbf{P}(t_2) - \mathbf{P}(t_1)T(t_1,t_2)\|_2 \leq \epsilon}_{\text{Clonal constraint}} ; \; T \geq 0 ; \text{Normalization}$$

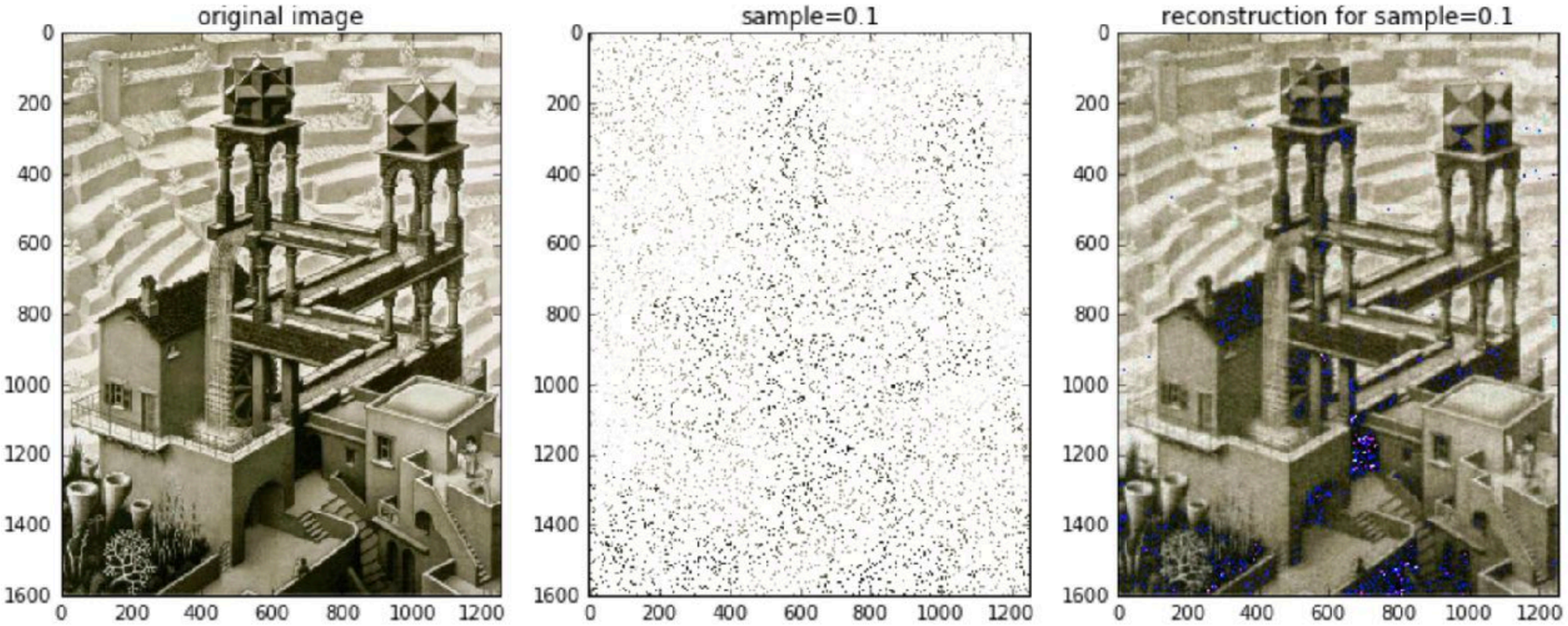
Detour: compressed sensing for sparse signal recovery from a few measurements



LASSO: $\min_{\vec{x}} ||\vec{x}||_1, \quad \text{subject to } |\vec{y} - A\vec{x}| < \epsilon$

E. J. Cand`es and M. B. Wakin, IEEE signal processing magazine 25, 21 (2008).

Detour: Compressed sensing reconstructs an image using only 10% randomly sampled pixels



Source: http://www.pyrunner.com/weblog/2016/05/26/compressed-sensing-python/

Applications of compressed sensing in biology

Infer transcriptome from random composite measurements

B. Cleary et.al., Cell, 2017
B. Cleary et.al., Nat. Biotech. 2021

Infer regulatory network

M. Nitzan et.al., Sci. Adv., 2017

Comparing LASSO, fused LASSO, and CoSpar

LASSO: $\min_{\vec{x}} ||\vec{x}||_1, \quad \text{subject to } |\vec{y} - A\vec{x}| < \epsilon$

Tibshirani, 1996

Fused LASSO: $\min_{\vec{x}} ||\vec{y} - A\vec{x}||_2, \quad \text{subject to } ||\vec{x}||_1 \leq \epsilon_1 \text{ and } \sum_j |x_j - x_{j-1}| \leq \epsilon_2$

Tibshirani et.al., 2005

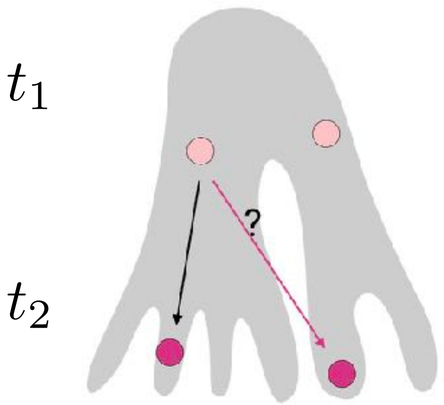
Locally smooth

CoSpar: $\min_T \underbrace{\|T\|_1}_{\text{Sparsity}} + \underbrace{\alpha \|LT\|_2}_{\text{Coherence}}, \quad s.t. \underbrace{\|\mathbf{P}(t_2) - \mathbf{P}(t_1)T(t_1, t_2)\|_2 \leq \epsilon}_{\text{Clonal constraint}} ; T \geq 0 ; \text{Normalization}$

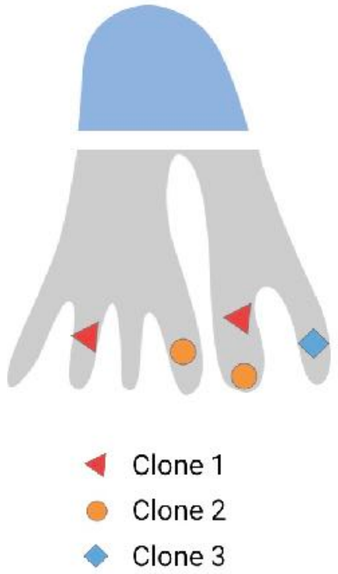
CoSpar extends LASSO and fused-LASSO by learning a **coherent** sparse **matrix** on an arbitrary cell-state **graph**

CoSpar helps to overcome key challenges for analyzing LT-scSeq datasets

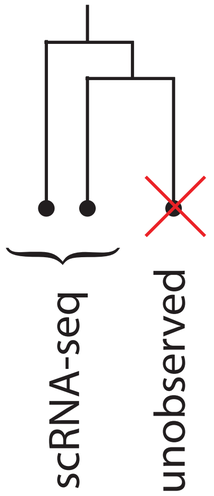
Heterogeneous early states



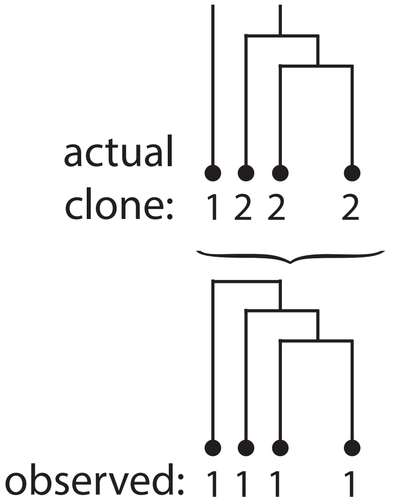
Missing early clonal states



Dropout



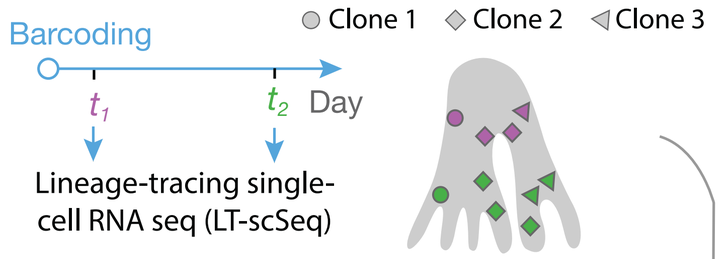
Barcode homoplasy

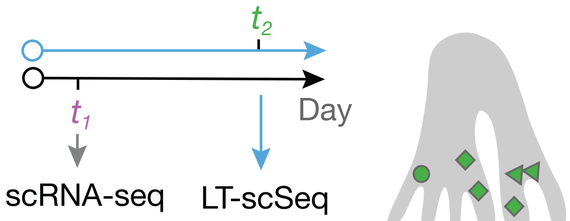


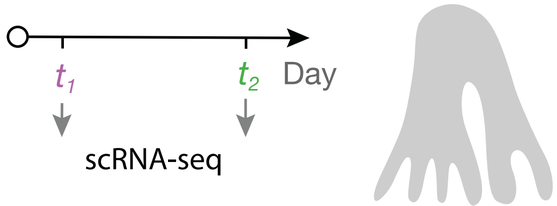
Also requires less input clonal data

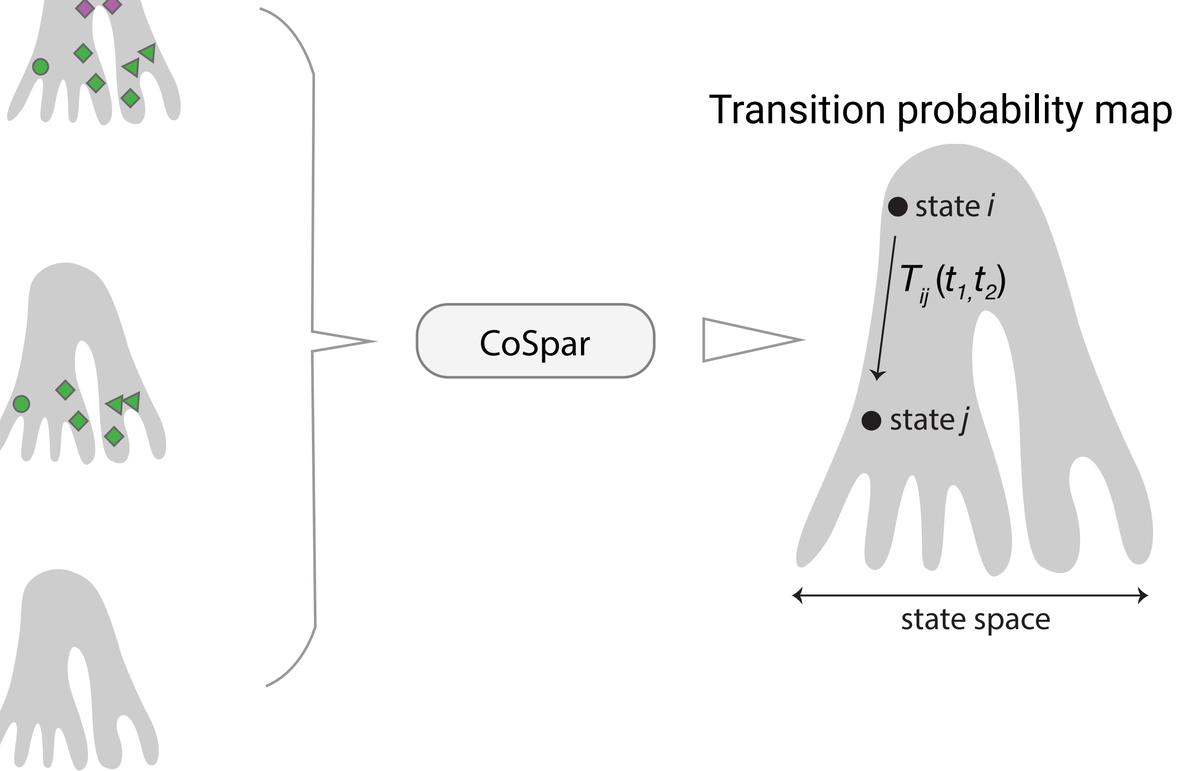
S.-W. Wang *et.al.,* Nat. Biotech. (2022)

CoSpar can work with 3 different experimental designs

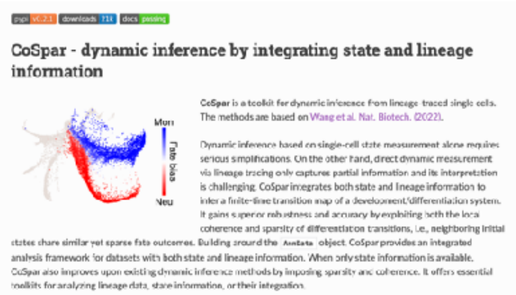








https://cospar.readthedocs.io/



S.-W. Wang *et.al.,* Nat. Biotech. (2022)

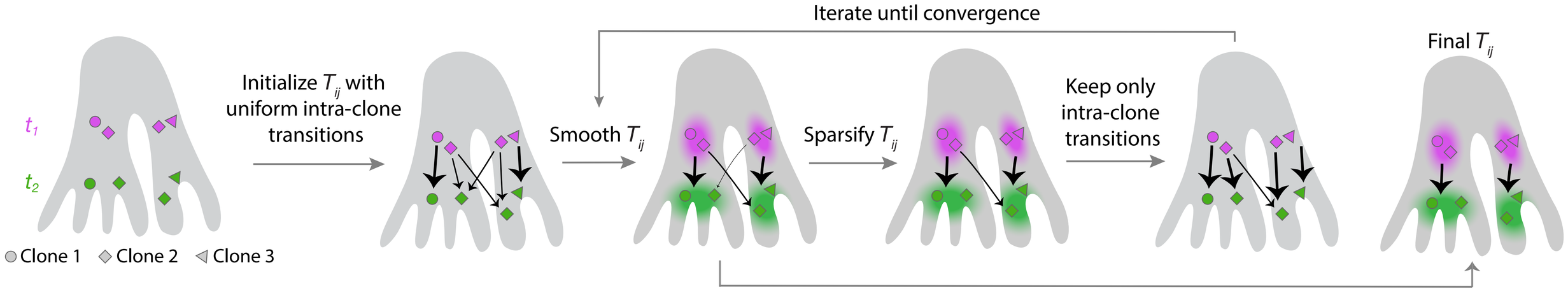
Part I: Transition map inference from re-sampled clones

Part II: Transition map inference from just one clonal time point

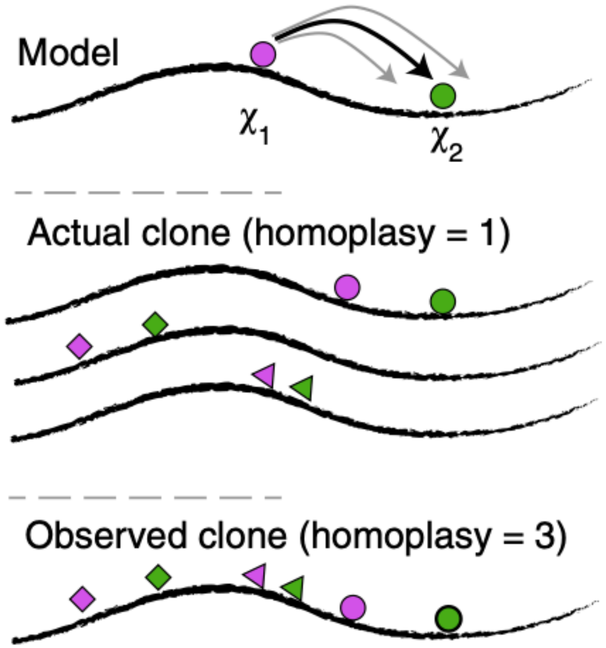
Part III: Applications to reprogramming and lung differentiation

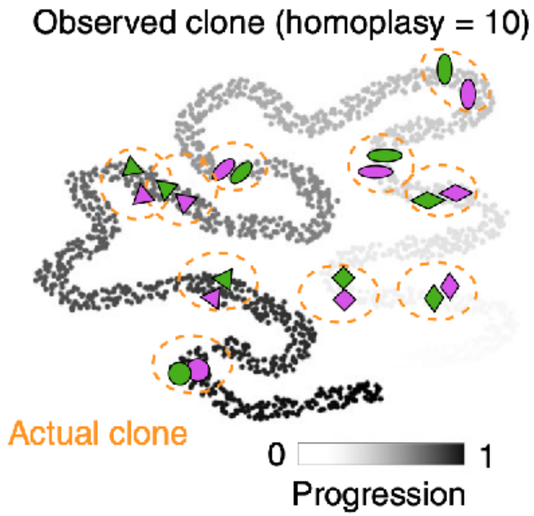
Heuristic implementation: sequentially apply coherence, sparsity, and clonal constraints until convergence

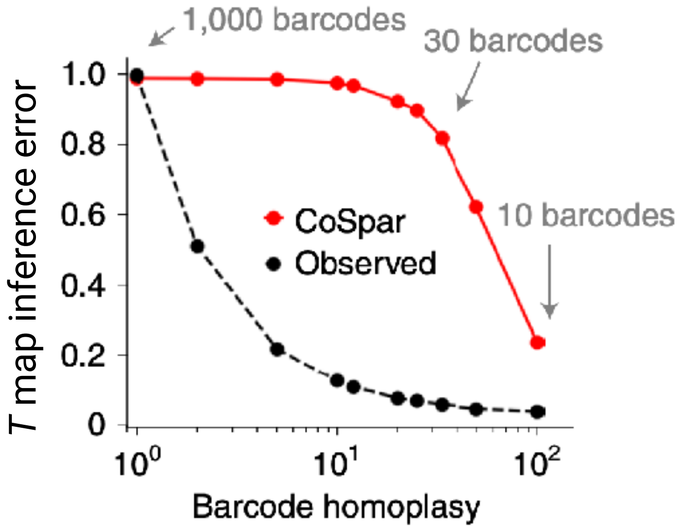
$$\min_{T} \underbrace{\|T\|_1}_{\text{Sparsity}} + \underbrace{\alpha\|LT\|_2}_{\text{Coherence}}, \qquad s.t. \underbrace{\|\mathbf{P}(t_2) - \mathbf{P}(t_1)T(t_1,t_2)\|_2 \leq \epsilon}_{\text{Clonal constraint}} ; T \geq 0 ; \text{Normalization}$$



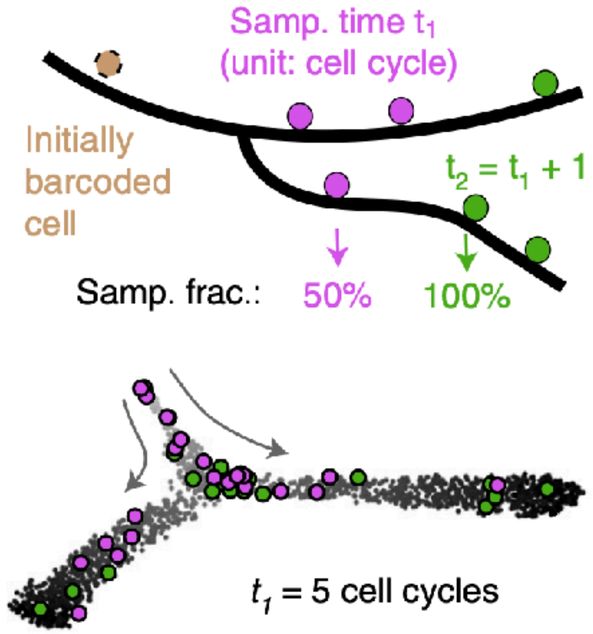
CoSpar correctly infers transition map from a simulated linear differentiation process with extensive barcode homoplasy

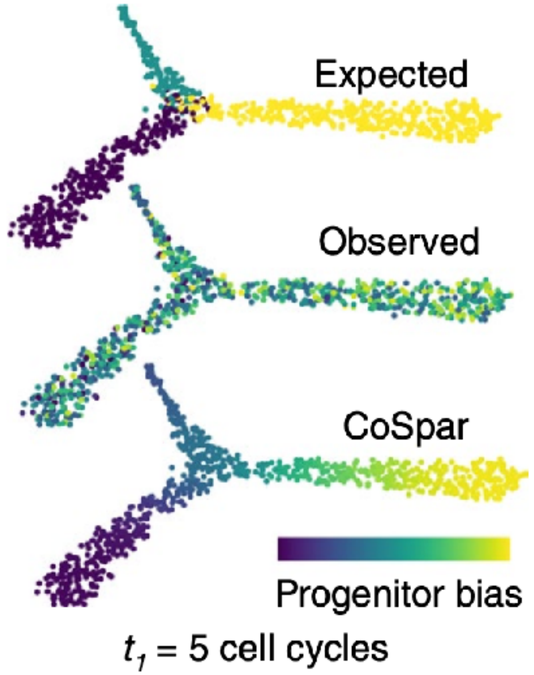


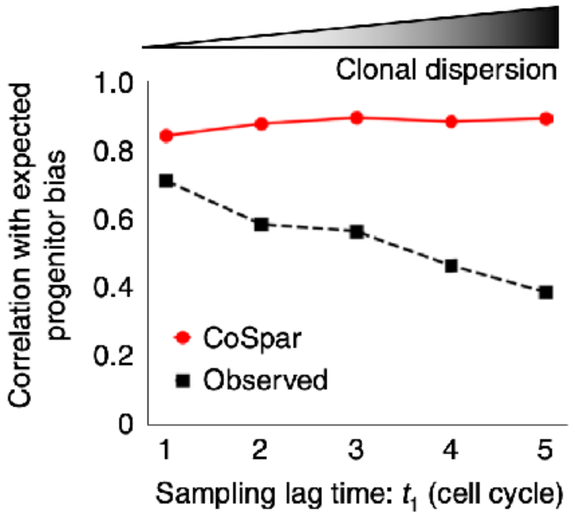




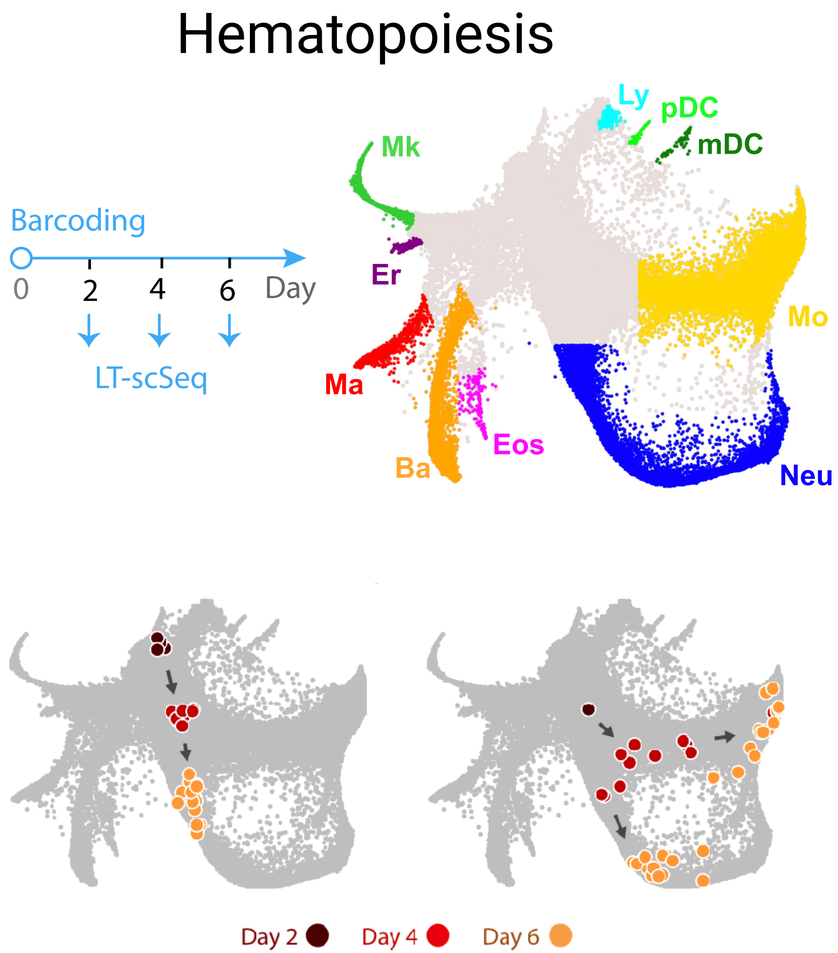
CoSpar correctly infers transition map from a simulated bifurcation process with high clonal dispersion

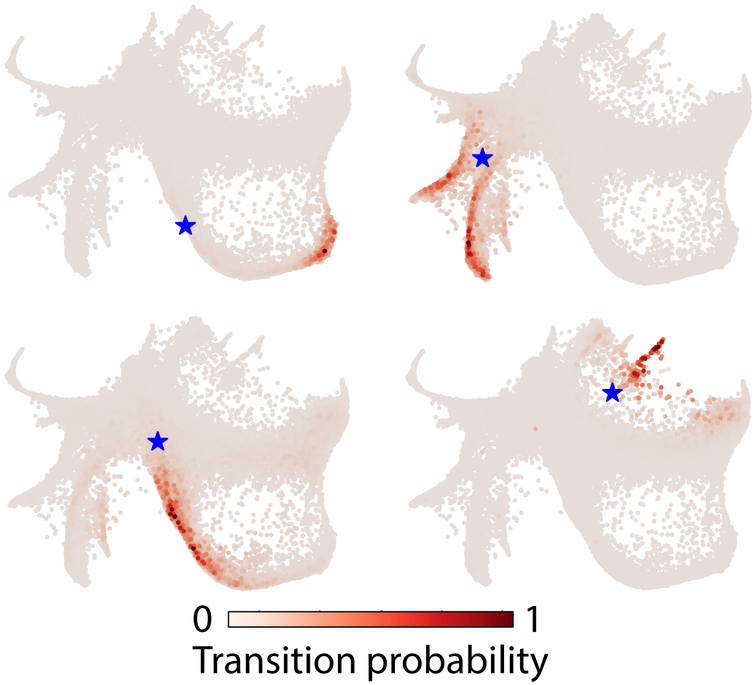






CoSpar learns single-cell transition map in hematopoiesis



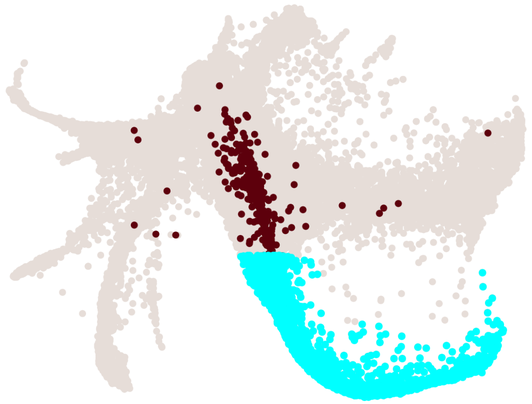


C. Weinreb, A. Rodriguez-Fraticelli *et.al.*, Science (2020)

CoSpar correctly predicts early progenitors of neutrophil using highly dispersed clones from day 4 and 6

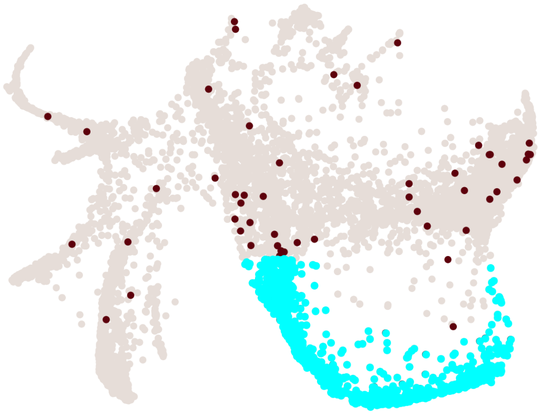
Observed progenitors
from all day 2-4 clones

Method in Weinreb et al., 2020

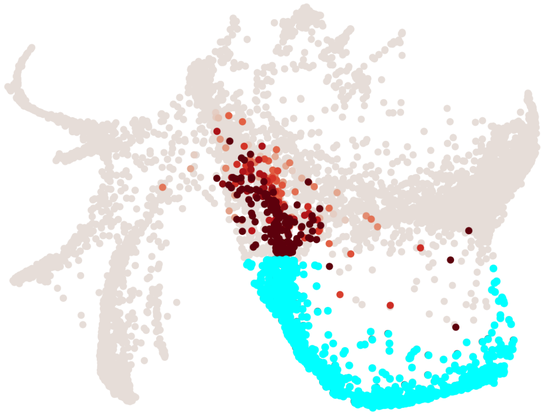


Using highly-dispersed day 4-6 clones
(dispersion>4.3, top 15%)

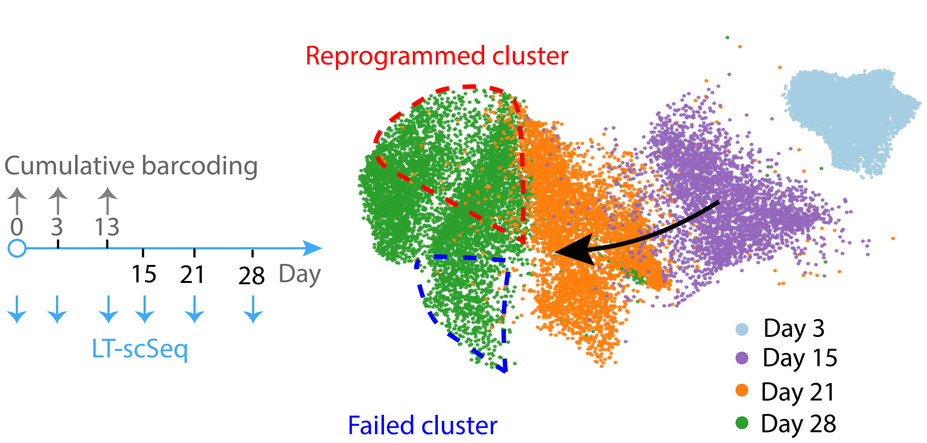
Weinreb

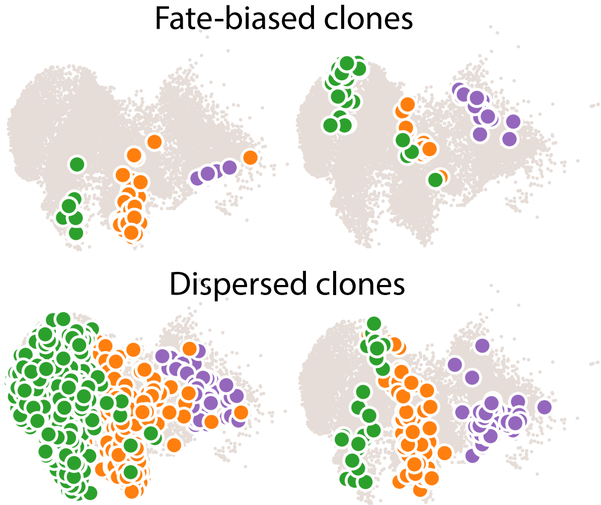


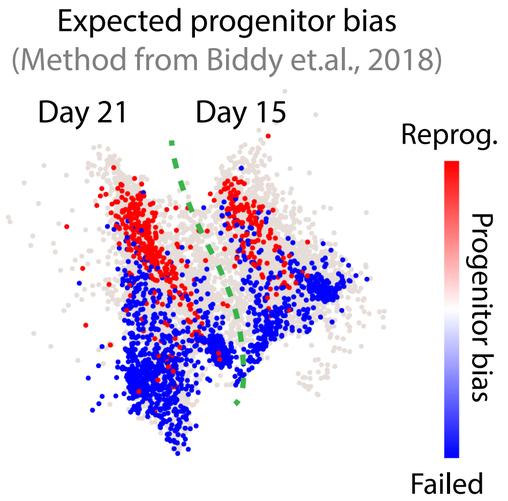
CoSpar



Clonal re-sampling during reprogramming reveals two differentiation trajectories

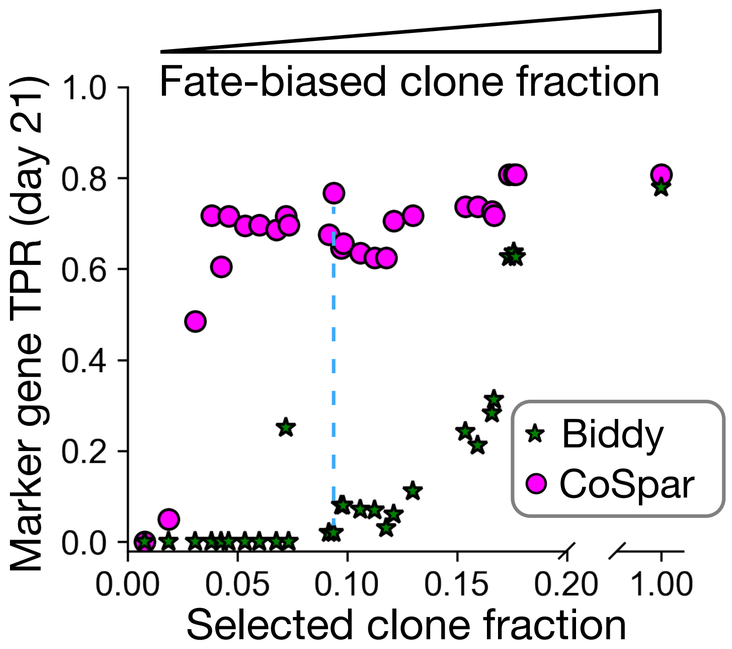


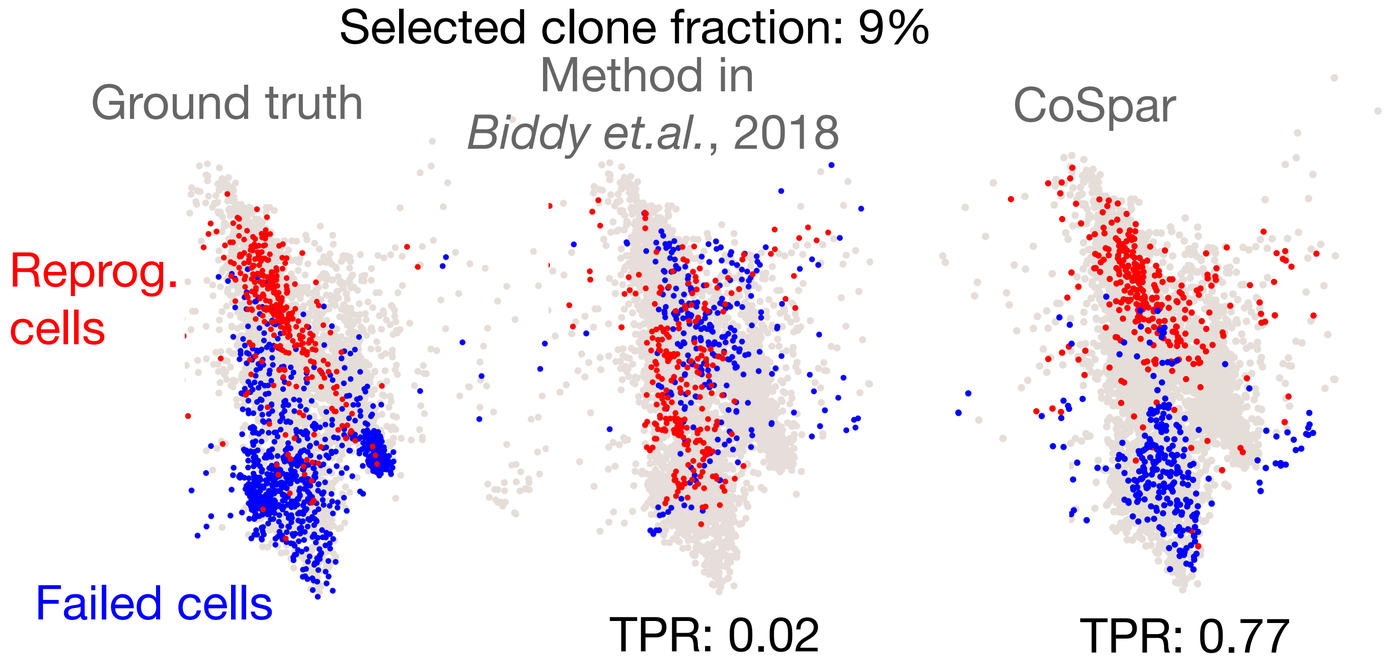




B. A. Biddy et.al., Nature (2018)

CoSpar predicts cell fate choice from as low as 5% of clones in reprogramming





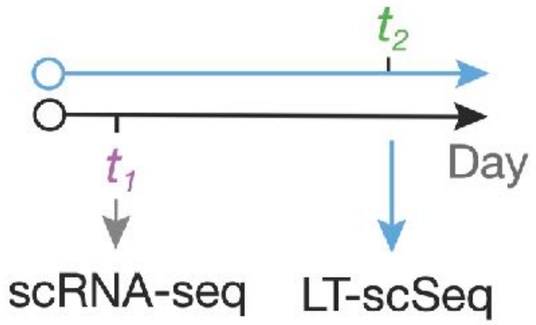
B. A. Biddy et.al., Nature (2018)

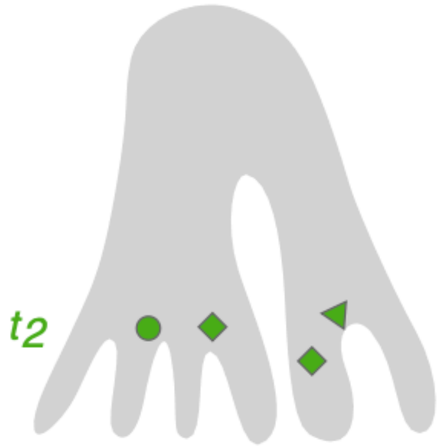
Part I: Transition map inference from re-sampled clones

Part II: Transition map inference from just one clonal time point

Part III: Applications to reprogramming and lung differentiation

Learning transition map from a single clonal time point is an under-determined problem

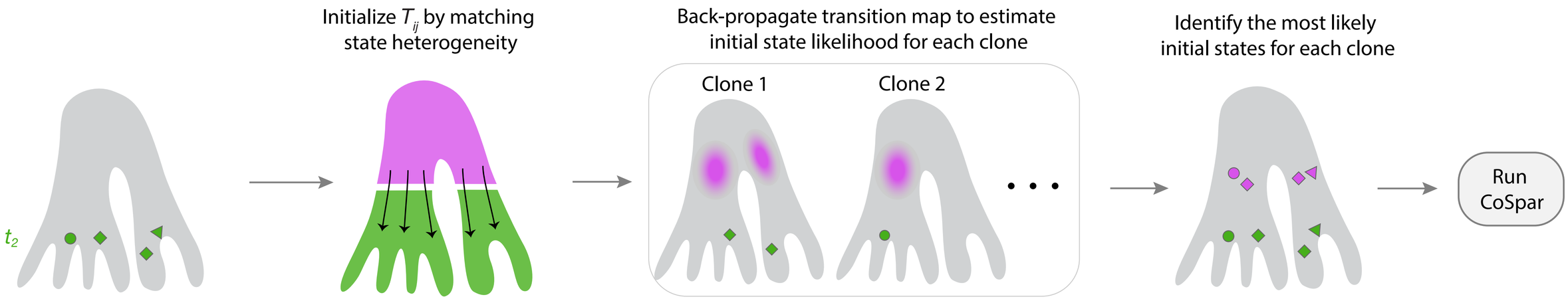




Ideal property of the transition map

* Coherence and sparsity
* Maximize transition prob. from putative progenitors to observed clonally-related cells
* Minimize transport cost between early and late populations (e.g. using OT)

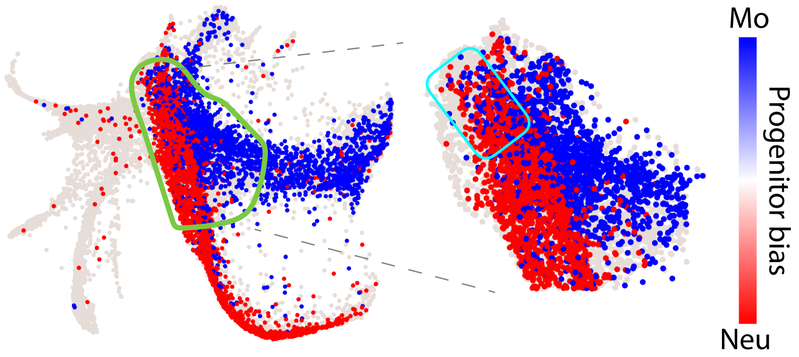
A simple approximation: sequentially apply each constraints



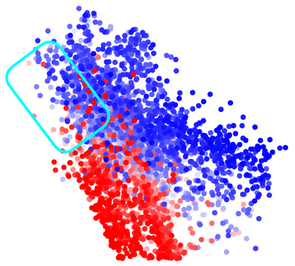
We further extends CoSpar to time-series data without clonal observations

CoSpar predicts early fate bias in hematopoiesis, outperforming existing fate-prediction methods based on transcriptome

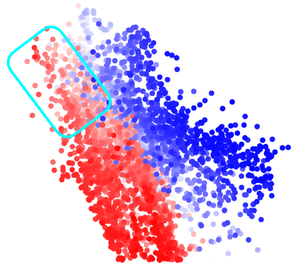
Ground truth: Method from Weinreb et al., 2020



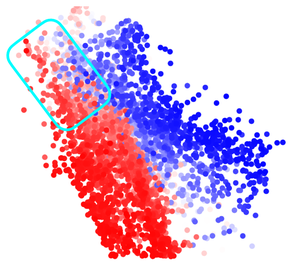
WOT: state only



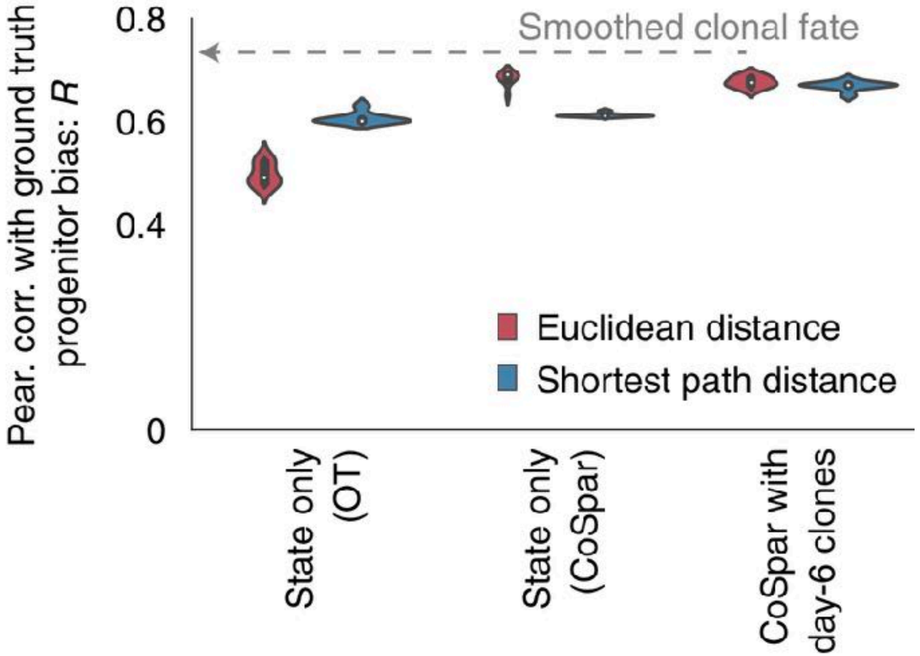
CoSpar: state only



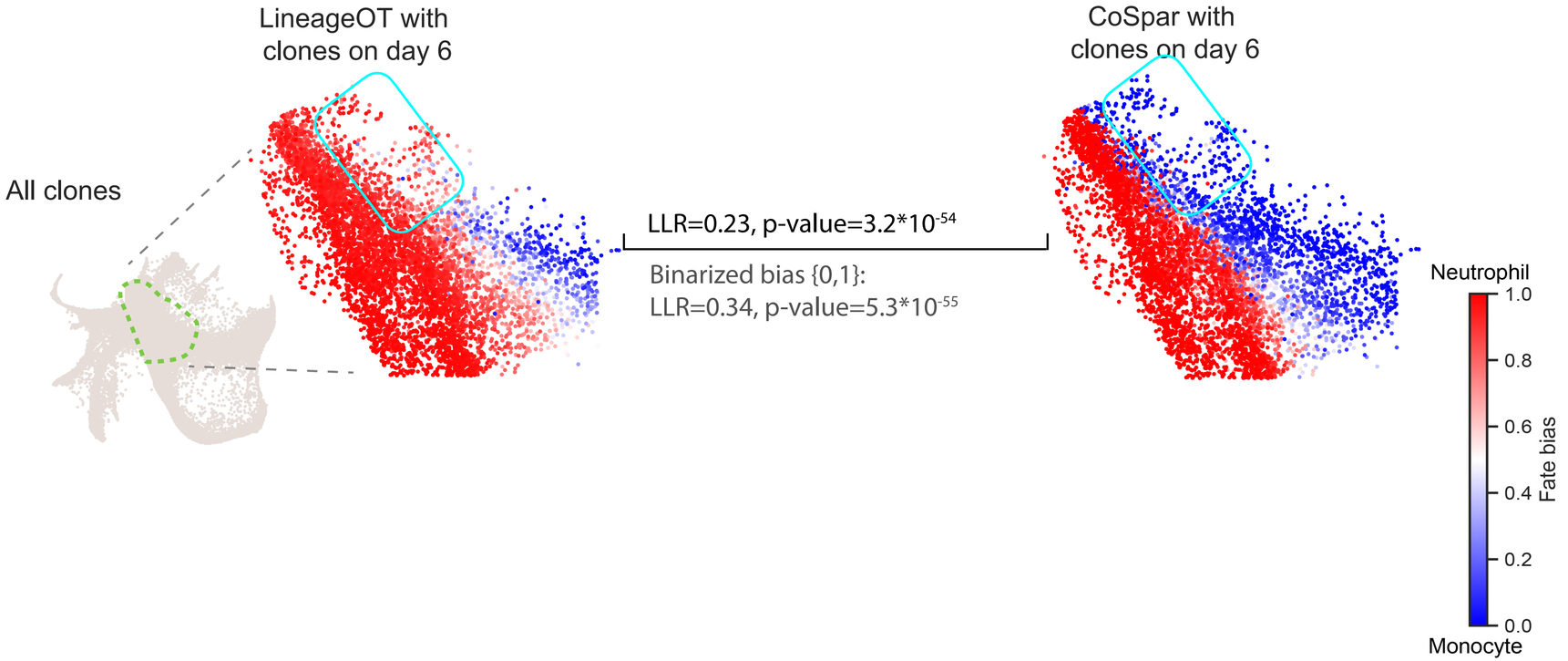
CoSpar: + day-6 clones



Metric used: Euclidean distance



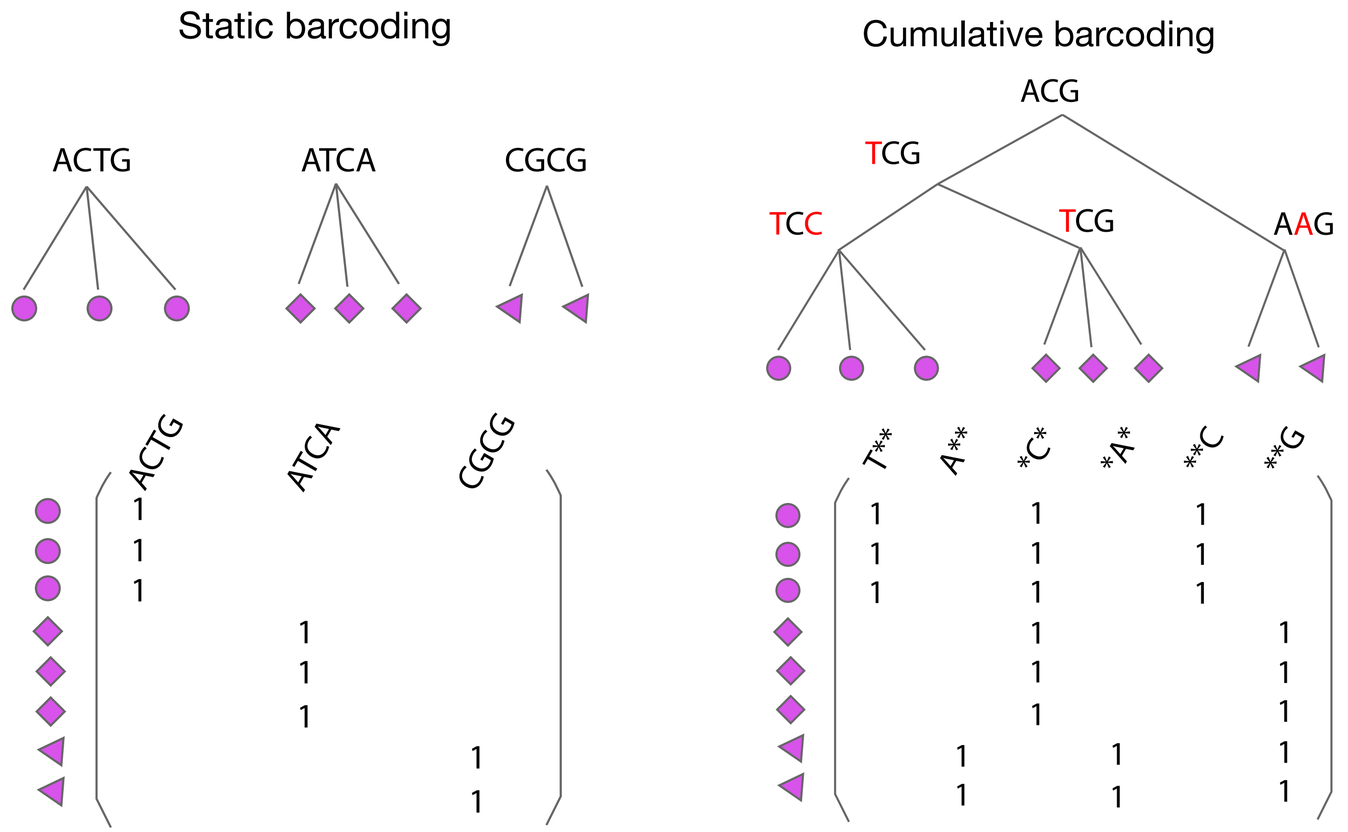
CoSpar outperforms LineageOT that also integrates lineage and transcriptome information



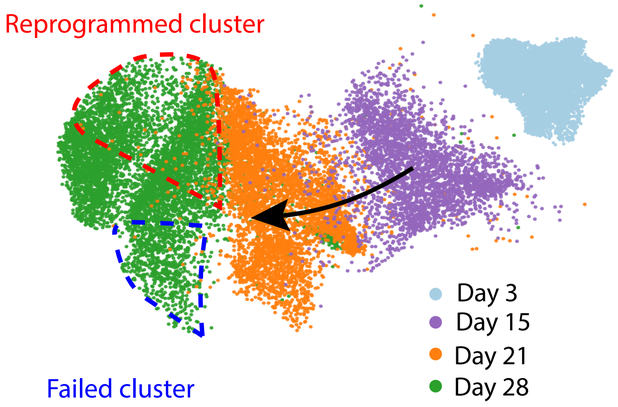
Mean log-likelihood ratio: $\mathrm{LLR} = \mathrm{Mean}_i \log\left(\frac{p_i(\mathrm{CoSpar}) + c_0}{p_i(\mathrm{LineageOT}) + c_0}\right)$

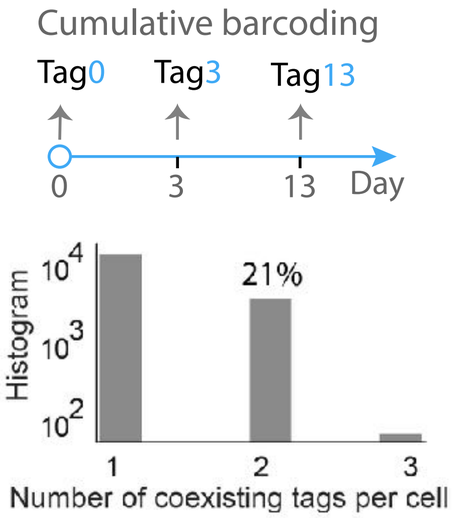
A. Forrow and G. Schiebinger, Nat. Commun. (2021).

Using a *cell-by-mutation* matrix as input, CoSpar works both for static and cumulative barcoding

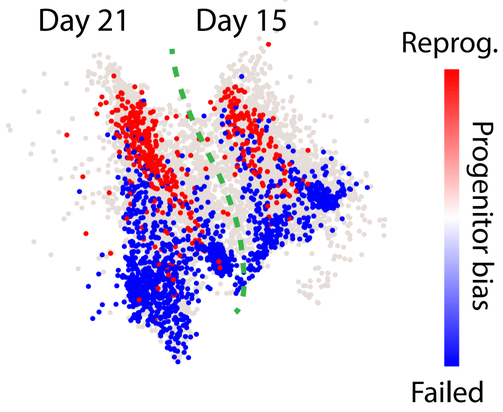


CoSpar works both for static and cumulative barcoding





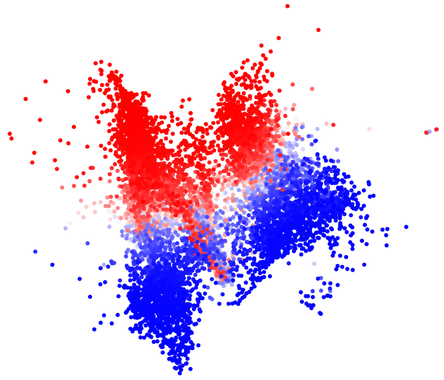
Expected progenitor bias
(Method from Biddy et.al., 2018)



Using only clones on day 28

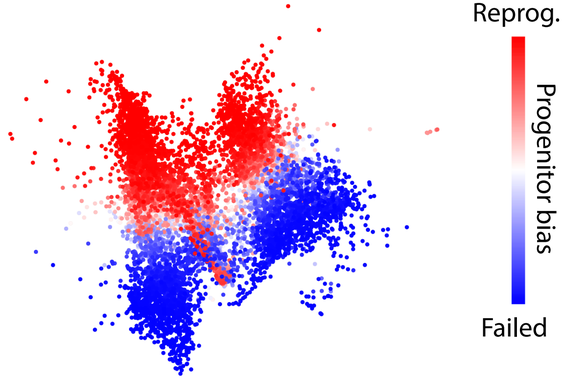
As cumulative BC

Isolate: {Tag0, Tag3, Tag13}



Convert to static BC

Concatenate BC: Tag0-Tag3-Tag13

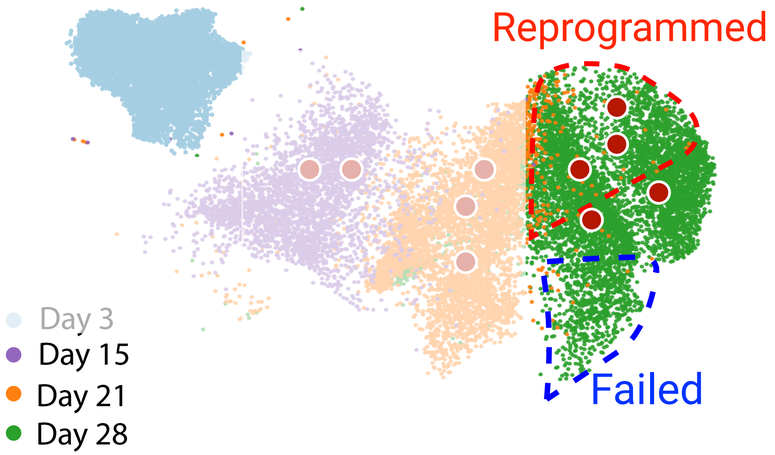


Part I: Transition map inference from re-sampled clones

Part II: Transition map inference from just one clonal time point

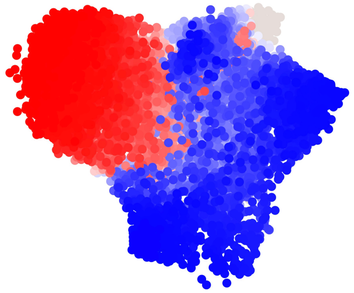
Part III: Applications to reprogramming and lung differentiation

CoSpar reveals fate choice as early as day 3 in reprogramming



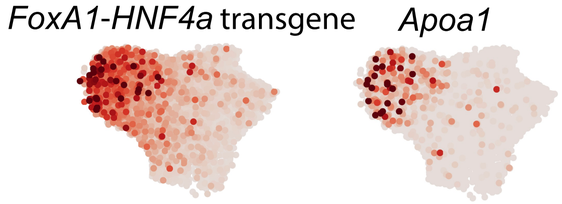
Biddy *et.al.*, Nature (2018)

Map: day-3 to day-28



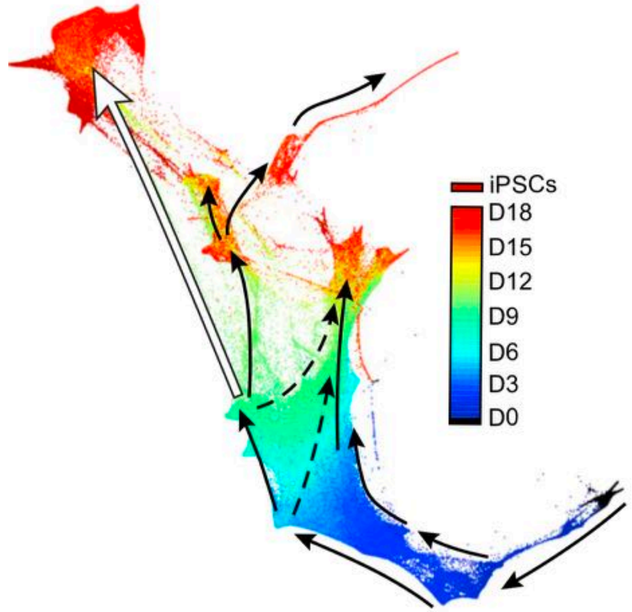
Progenitor bias

Reprog. Failed



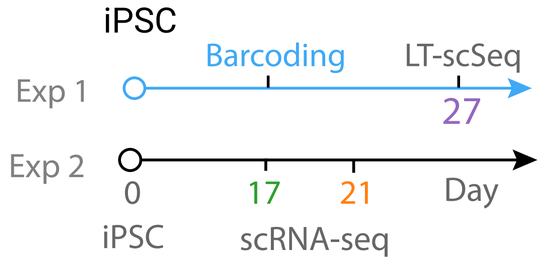
Normalized gene expression

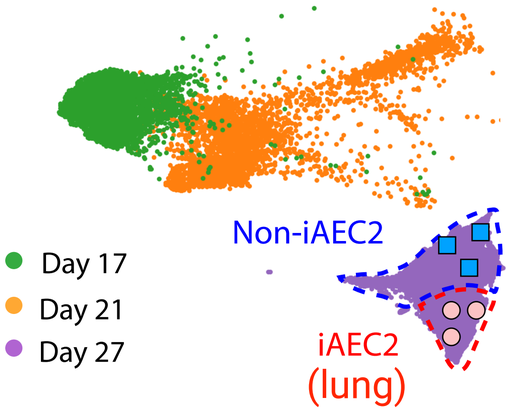
WOT: sampling every 12h over reprogramming



Schiebinger *et.al.*, Cell (2019)

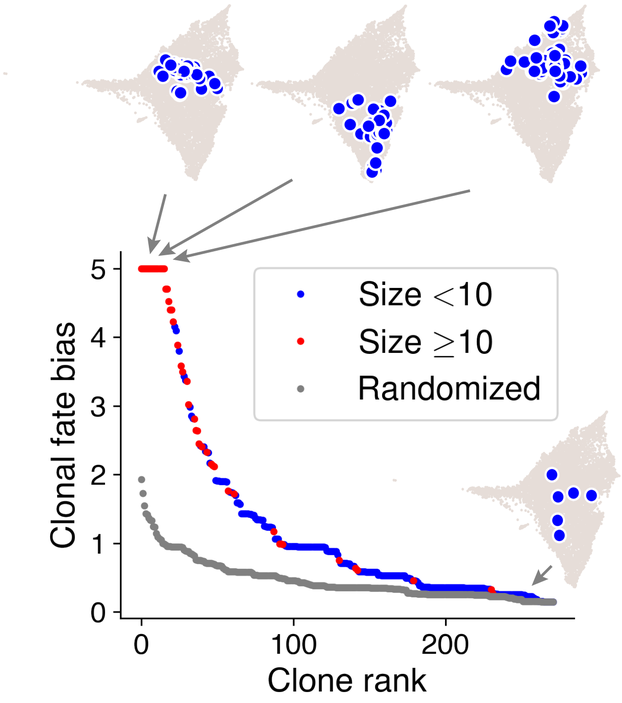
CoSpar reveals fate choice as early as day 3 in reprogramming

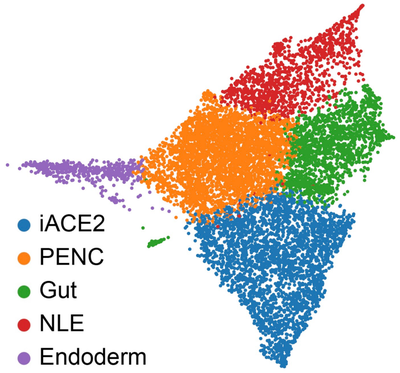


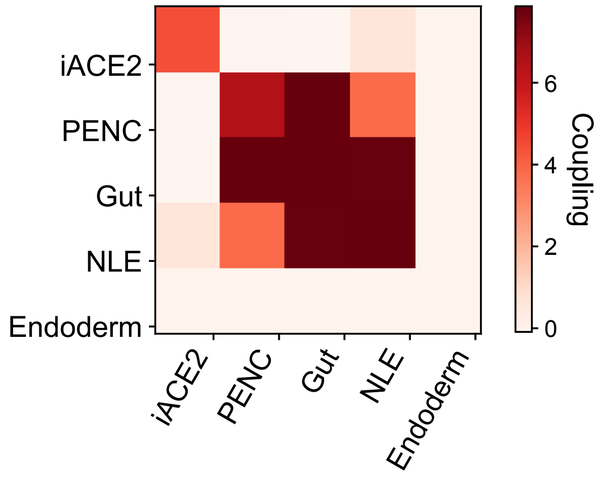


K. Hurley *et.al.*, Cell Stem Cell (2020)

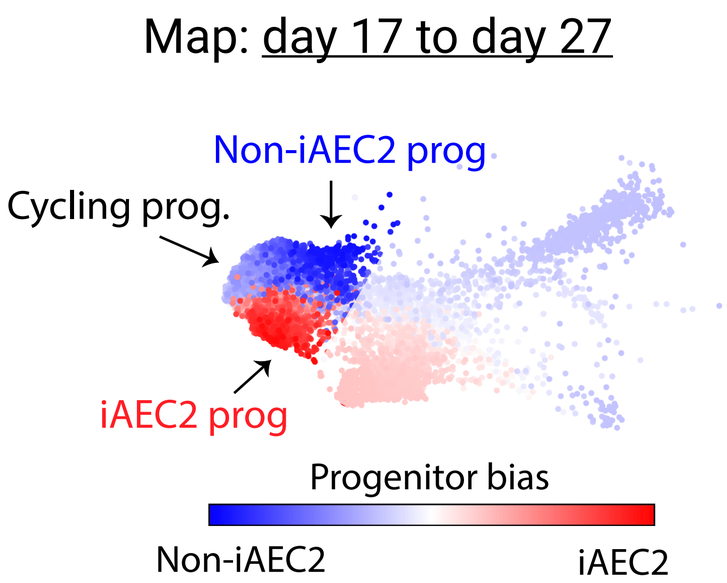
77% fate-biased clones
(≥ 10 cells)

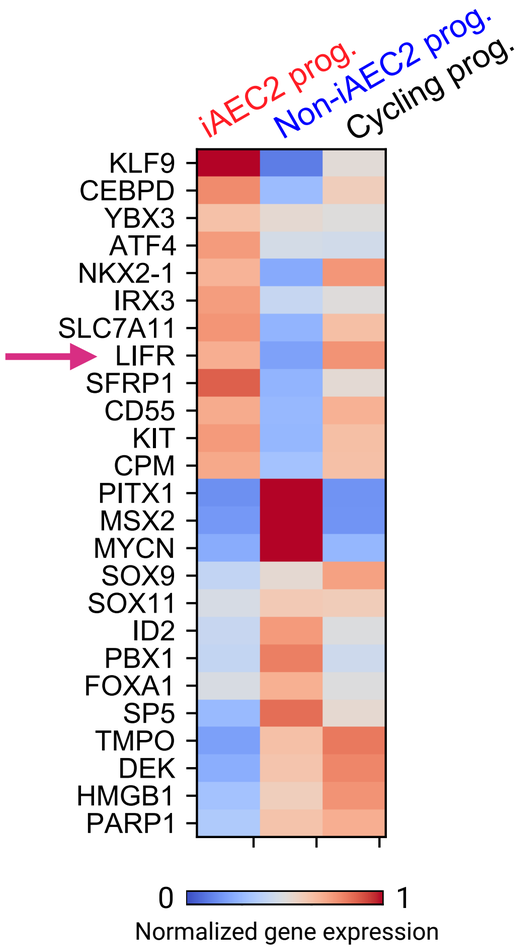


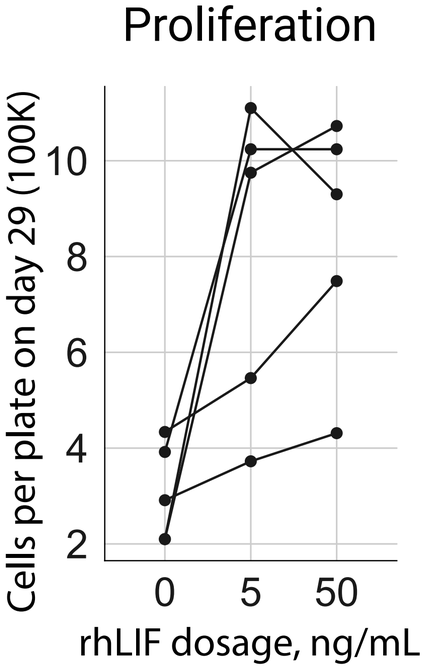




CoSpar predicts early fate boundary in lung differentiation



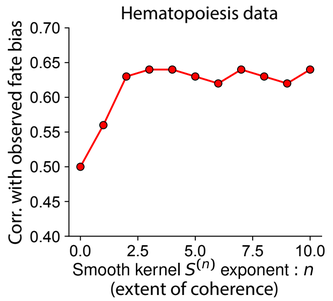




Limitations of CoSpar

CoSpar only learns the “average” fate bias of observed states

CoSpar is sensitive to the degree of smoothing or how cell-cell similarity is measured



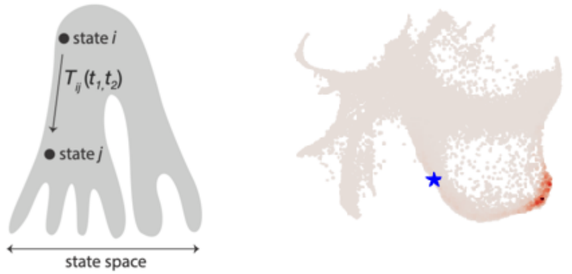
CoSpar would not work if early transcriptome heterogeneity relevant to fate choice is not measured or not preserved after data preprocessing

With a single clonal time point, CoSpar leans more on state information to infer transition map

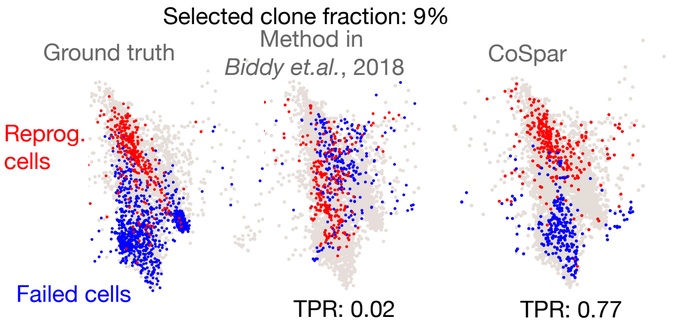
Conclusion

CoSpar extends the framework of compressed sensing by learning a coherent and sparse transition map from re-sampled clonal data

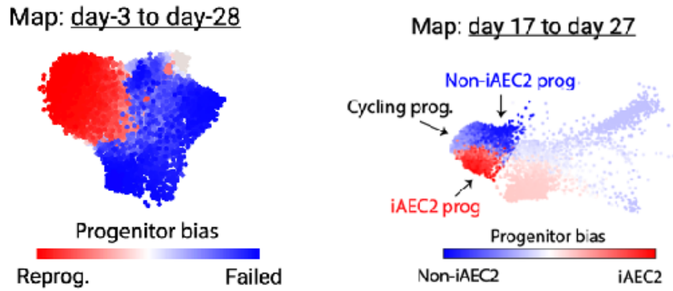
$$\min_{T} \underbrace{\|T\|_1}_{\text{Sparsity}} + \underbrace{\alpha\|LT\|_2}_{\text{Coherence}}, \qquad s.t. \underbrace{\|\mathbf{P}(t_2) - \mathbf{P}(t_1)T(t_1,t_2)\|_2 \le \epsilon}_{\text{Clonal constraint}} ;\ T \ge 0 ;\ \text{Normalization}$$



CoSpar overcomes challenges from clonal dispersion and barcode homoplasy, and requires much fewer clones



CoSpar successfully learns transition map from a single clonal time point to predict fate outcome several weeks later in reprogramming and direct lung differentiation



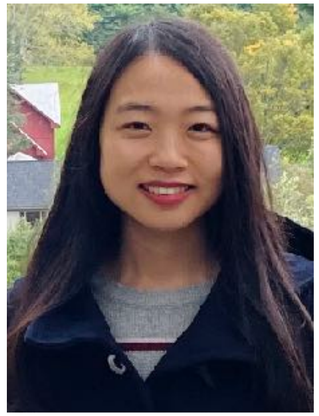
https://cospar.readthedocs.io/

S.-W. Wang *et.al.,* Nat. Biotech. (2022)

Acknowledgement

Allon Klein, HMS

Li Li, BCH



<u>Klein lab</u>
Charlotte Strandkvist
Laura Bagamery
Sean Mcgeary
Hailey Cambra
Kalki Kukreja
Qiu Wu
Tal Deborah Scully
Nicolas Gort Freitas
Ignas Mazelis

<u>Funding:</u>

Damon Runyon Fellow
DAMON RUNYON CANCER RESEARCH FOUNDATION

WESTLAKE UNIVERSITY

<u>Other collaborators</u>
Darrell Kotton
Michael J. Herriges
Kilian Hurley

MIA: Alex Bloemendal & Lois Doolittle