RNA Velocity of Single Cells

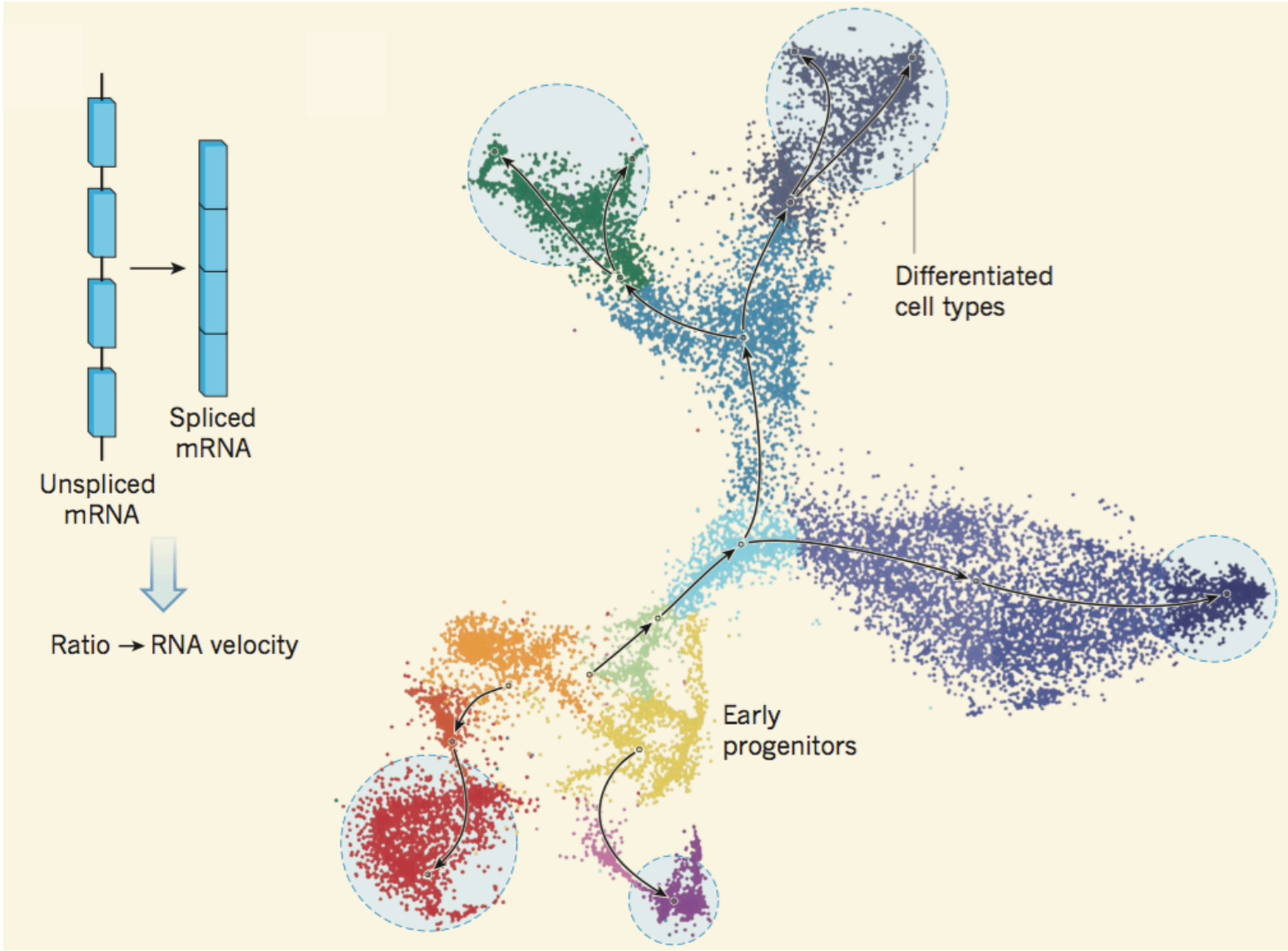
Min Dai

Dec 20 2018

daimin@zju.edu.cn

Outline

- Description
- Estimation
- Extrapolation
- Visualization
- Application



Cold Spring Harbor Laboratory

bioRxiv

THE PREPRINT SERVER FOR BIOLOGY

New Results

RNA velocity in single cells

Posted October 19, 2017

Gioele La Manno, Ruslan Soldatov, Hannah Hochgerner, Amit Zeisel, Viktor Petukhov, Maria Kastriti, Peter Lonnerberg, Alessandro Furlan, Jean Fan, Zehua Liu, David van Bruggen, Jimin Guo, Erik Sundstrom, Goncalo Castelo-Branco, Igor Adameyko, Sten Linnarsson, Peter Kharchenko

doi: https://doi.org/10.1101/206052

Now published in *Nature* doi: 10.1038/s41586-018-0414-6

LETTER

https://doi.org/10.1038/s41586-018-0414-6

Published August 23, 2018

RNA velocity of single cells

Gioele La Manno[1,2], Ruslan Soldatov[3], Amit Zeisel[1,2], Emelie Braun[1,2], Hannah Hochgerner[1,2], Viktor Petukhov[3,4], Katja Lidschreiber[5], Maria E. Kastriti[6], Peter Lönnerberg[1,2], Alessandro Furlan[1], Jean Fan[3], Lars E. Borm[1,2], Zehua Liu[3], David van Bruggen[1], Jimin Guo[3], Xiaoling He[7], Roger Barker[7], Erik Sundström[8], Gonçalo Castelo-Branco[1], Patrick Cramer[5,9], Igor Adameyko[6], Sten Linnarsson[1,2]* & Peter V. Kharchenko[3,10]*

Peter Kharchenko

Making sense of genomics data!

- Department of Biomedical Informatics
- Harvard Medical School

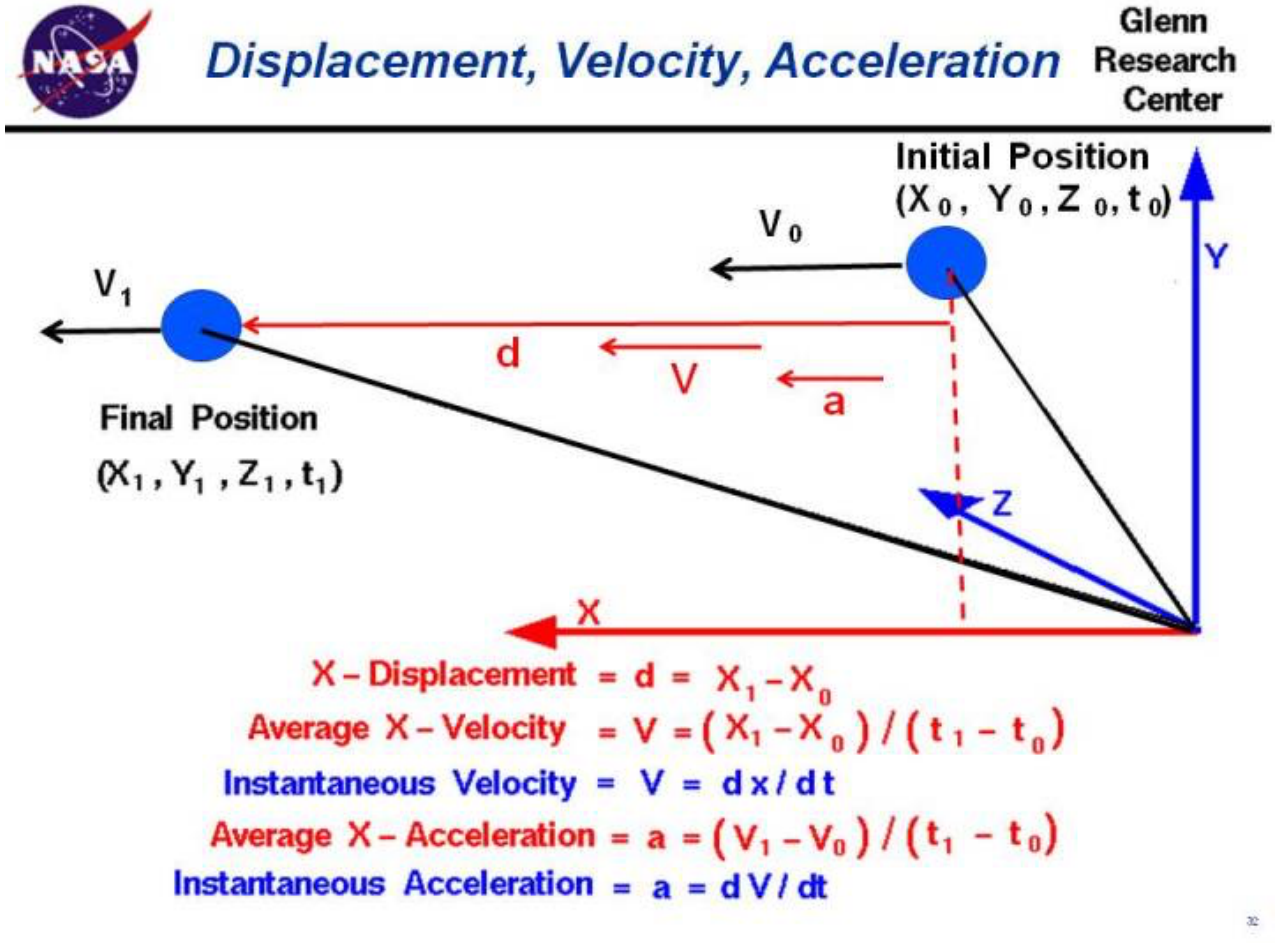
- PhD in Biophysics at Harvard University (studying gene regulation and metabolic networks under the advisement of George Church)

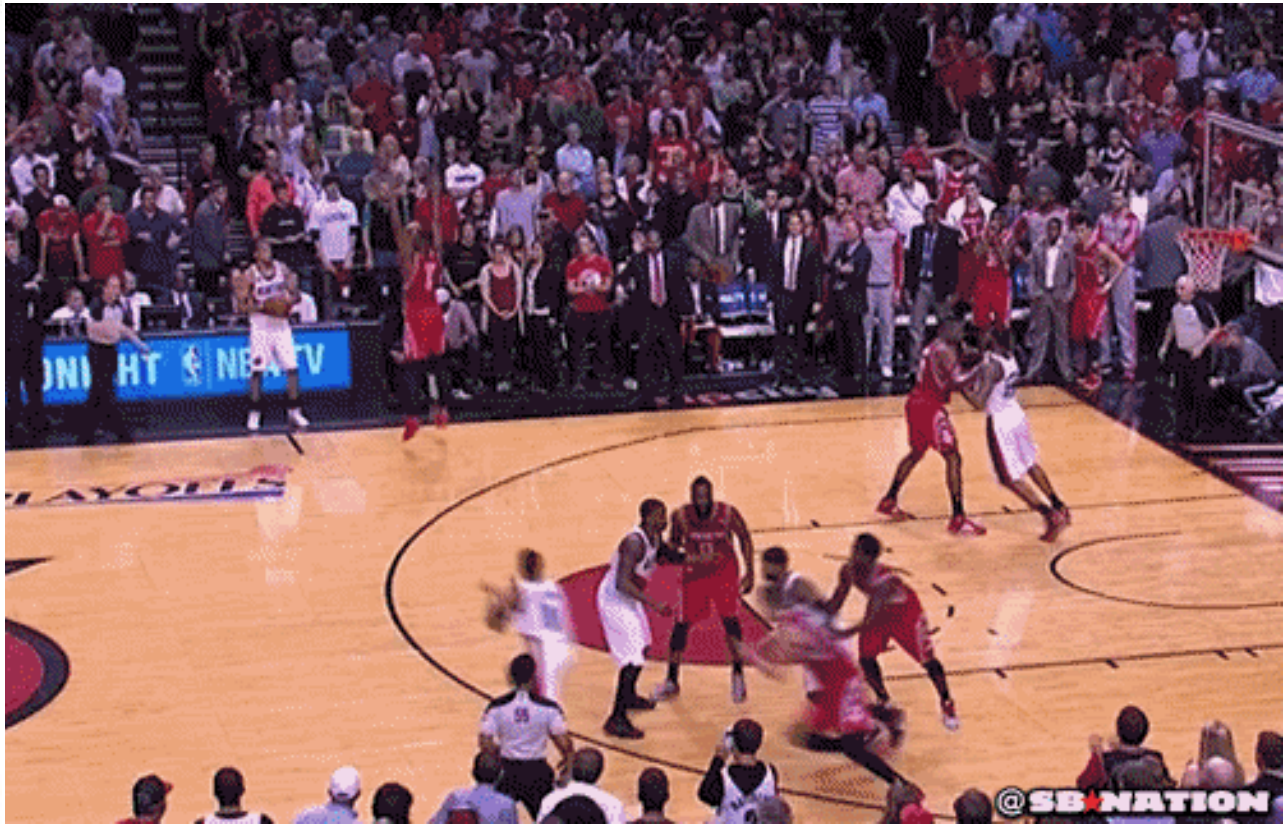
Research interest:

- Intratumoral heterogeneity in different cancer types
- Statistical modeling of the biological tissue from single-cell measurements
- Automating and scaling up key functional genomics assays using microfluidics

What is velocity?

- Velocity: a vector quantity that designates how fast and in what direction a point is moving
- Can we predict future states of a point with estimated velocity?



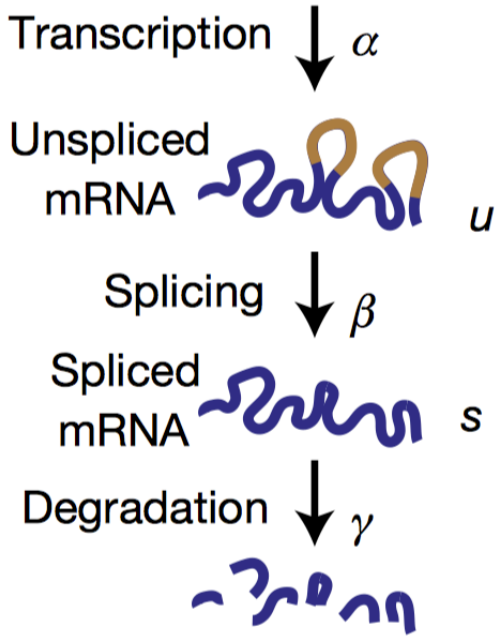


https://www.grc.nasa.gov/www/k-12/airplane/disvelac.html

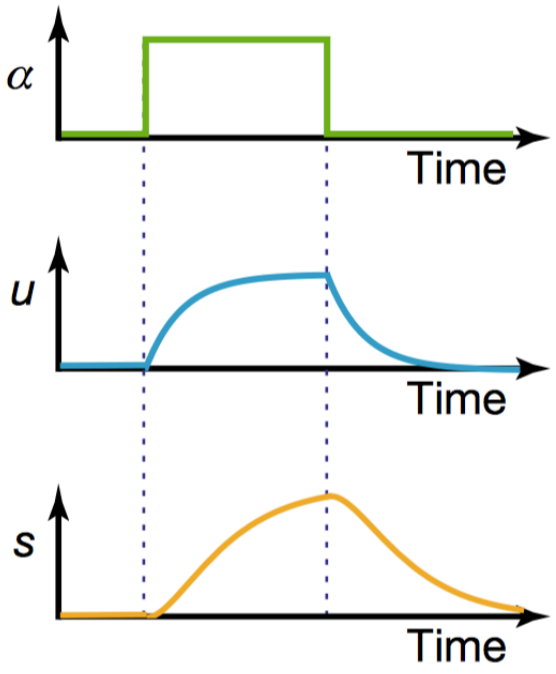
https://www.blazersedge.com

Theoretical description of RNA velocity

- The rate equations for a single gene describes how the expected number of unspliced mRNA molecules u, and spliced molecules s, evolve over time
- $\alpha(t)$ is the time-dependent rate of transcription, $\beta(t)$ is the rate of splicing, $\gamma(t)$ is the rate of degradation



The transcriptional dynamics



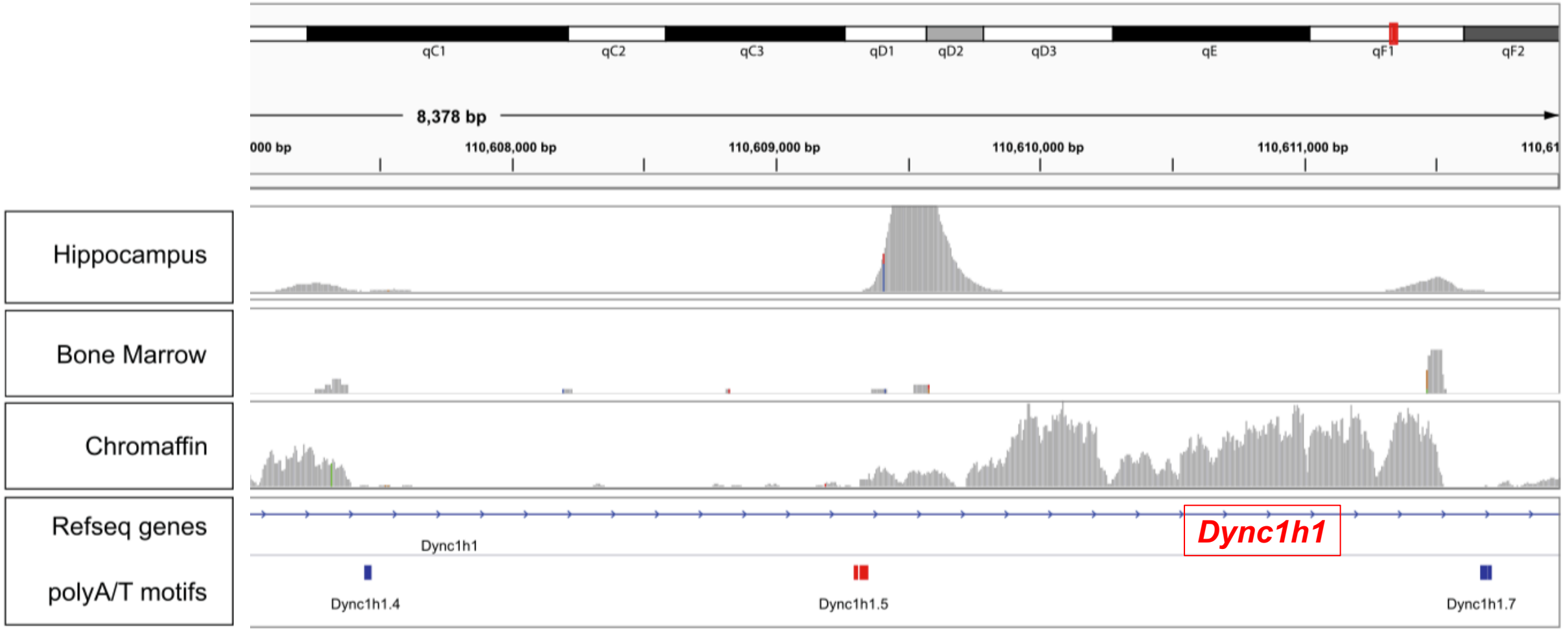
$$\frac{du}{dt} = \alpha(t) - \beta(t)u(t)$$

$$\frac{ds}{dt} = \beta(t)u(t) - \gamma(t)s(t)$$

The rate equations for a single gene

Most of the intronic reads arise due to internal priming from stable positions

- 15–25% of reads contained unspliced intronic sequences
- Track: 10x Chromium (top), inDrop (the second), Smart-seq2 (the third)



Simplify the model with assumptions

- Under an assumption of constant (time-independent) rates $\alpha(t) = \alpha$, $\gamma(t) = \gamma$, and setting $\beta(t) = 1$, the rate equations can be simplified
- The solution can be used to extrapolate mRNA abundance s to a future timepoint t_1

$$\frac{du}{dt} = \alpha(t) - \beta(t)u(t)$$

$$\frac{ds}{dt} = \beta(t)u(t) - \gamma(t)s(t)$$

$$\frac{du}{dt} = \alpha - u(t)$$

$$\frac{ds}{dt} = u(t) - \gamma s(t)$$

$$u(t) = \alpha(1 - e^{-t}) + u_0 e^{-t}$$

$$s(t) = \frac{e^{-t(1+\gamma)}\left[e^{t(1+\gamma)}\alpha(\gamma - 1) + e^{t\gamma}(u_0 - \alpha)\gamma + e^{t}(\alpha - \gamma(s_0 + u_0 + s_0\gamma))\right]}{\gamma(\gamma - 1)}$$

(where $u(0) = u_0$ and $s(0) = s_0$ are initial conditions)

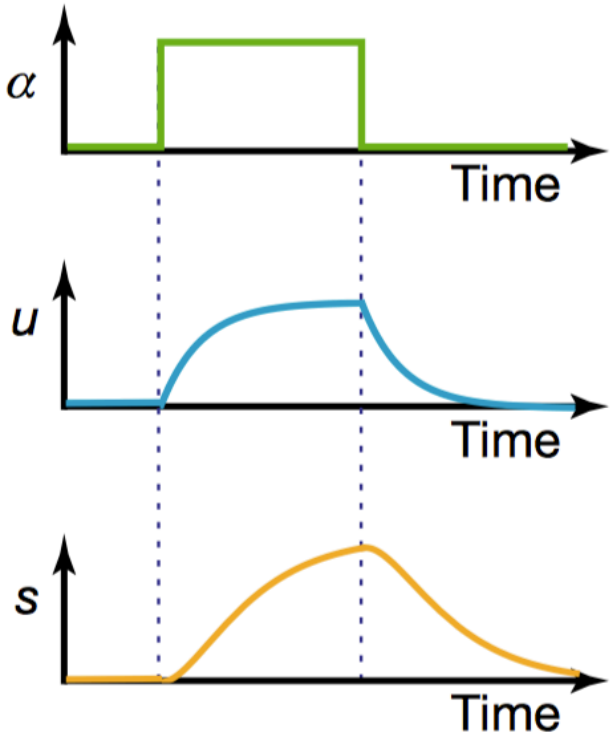
Estimate γ with the steady-state assumption

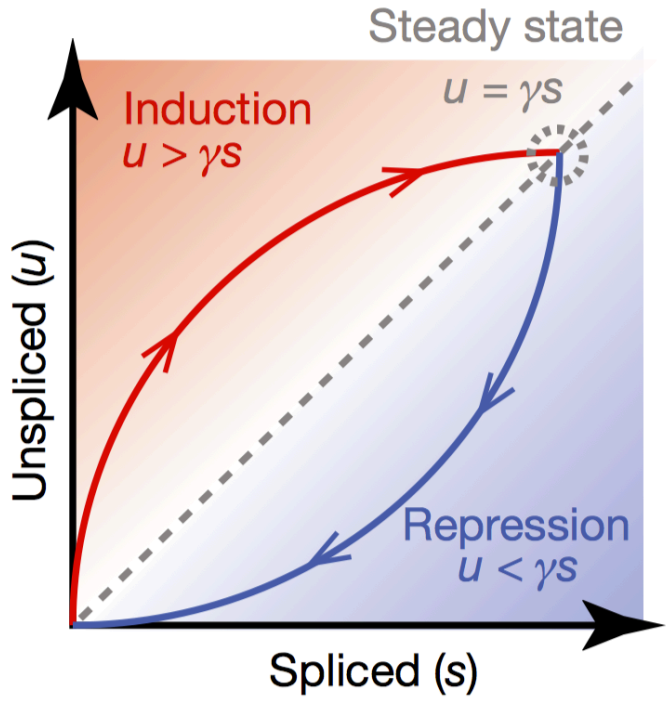
- The normalized degradation rate γ varies among genes and needs to be estimated in a gene-specifc manner
- In steady-state populations, where $ds/dt = 0$, γ of a given gene can be determined as the ratio of unspliced to spliced mRNA molecules
- The equilibrium slope γ combines degradation and splicing rates, capturing gene-specific regulatory properties, the ratio of intronic and exonic lengths, and the number of internal priming sites

$$\frac{du}{dt} = \alpha - u(t), \frac{ds}{dt} = u(t) - \gamma s(t)$$

$\frac{ds}{dt} = 0$ (steady state)

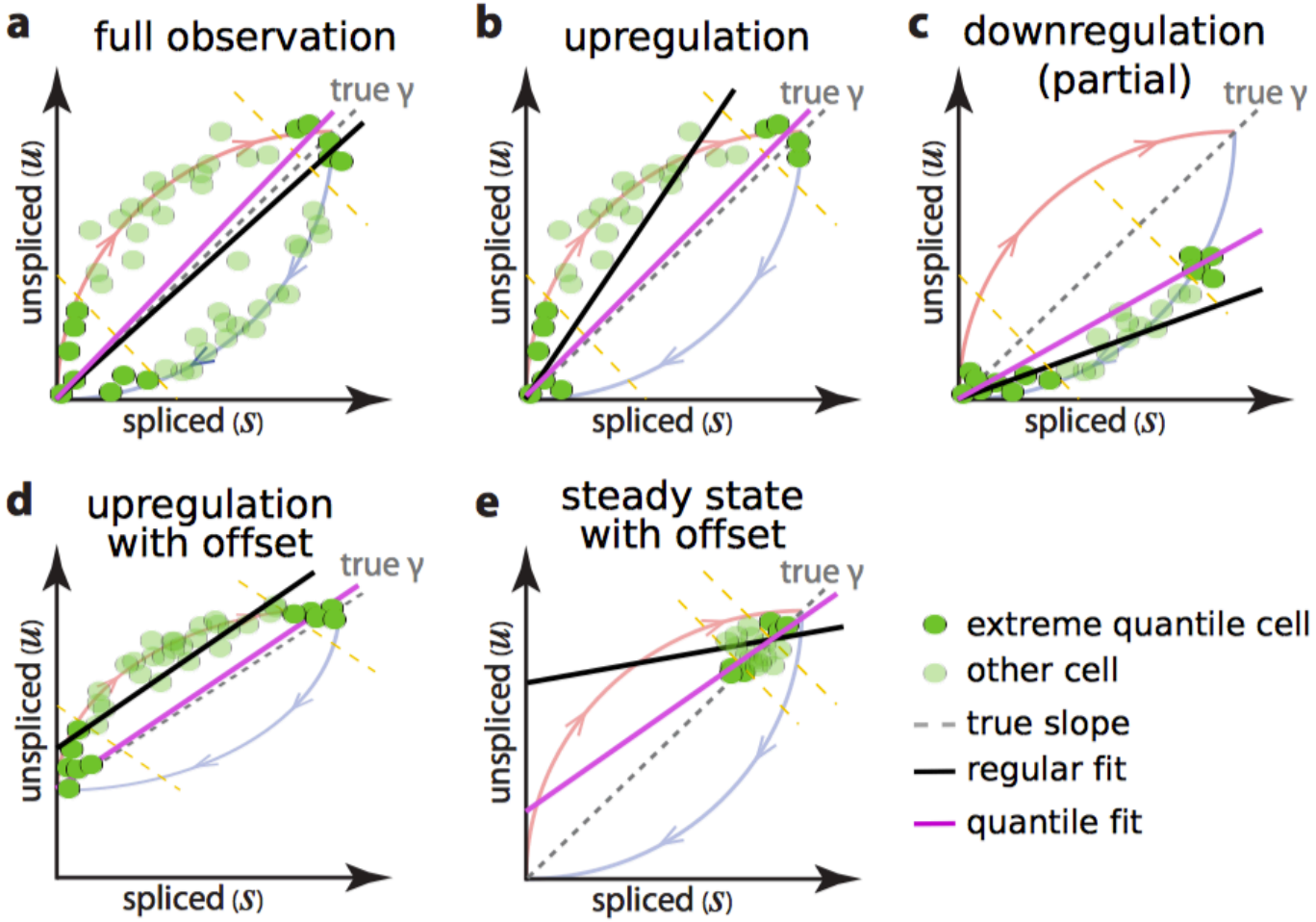
$$\gamma = \frac{u}{s}, \alpha = u$$

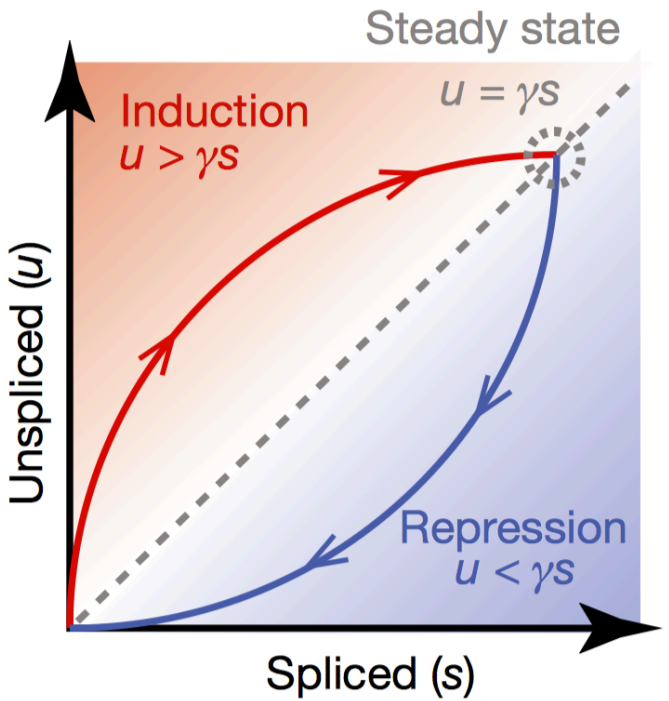




Estimate γ with the steady-state assumption

- Use a least squares fit of $u \sim \gamma * s$ to estimate γ for each gene
  - u and s are the size-normalized unspliced and spliced abundances observed for given gene across the cells
- An offset can optionally be included (baseline intronic counts that might be driven by unannotated transcripts)





Estimate $s(t)$ with two alternative assumptions

Model I. Constant velocity assumption

- The rate of change of the spliced molecules remains constant, i.e. that $\frac{ds}{dt} = v$ is constant

$$s(t) = s_0 + vt$$

- This assumption works well in practice as long as the time step is short

Model II. Constant unspliced molecules assumption

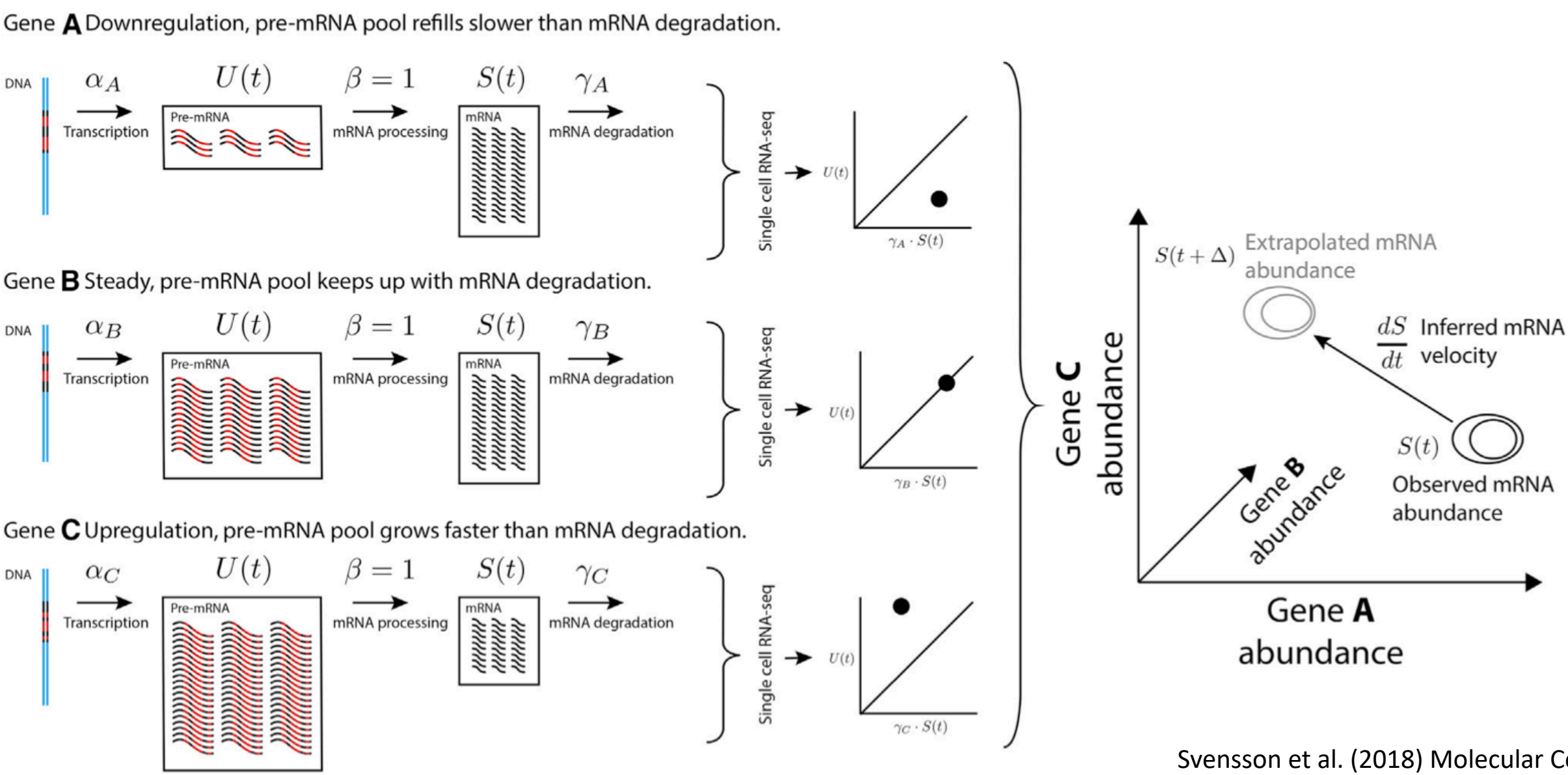
- The number of unspliced molecules stays constant, i.e. that $u(t) = u_0$. This reduces the problem to a single rate equation:

$$\frac{ds}{dt} = u_0 - \gamma s(t)$$

- The solution then becomes:

$$s(t) = s_0 e^{-\gamma t} + \frac{u_0}{\gamma}(1 - e^{-\gamma t})$$

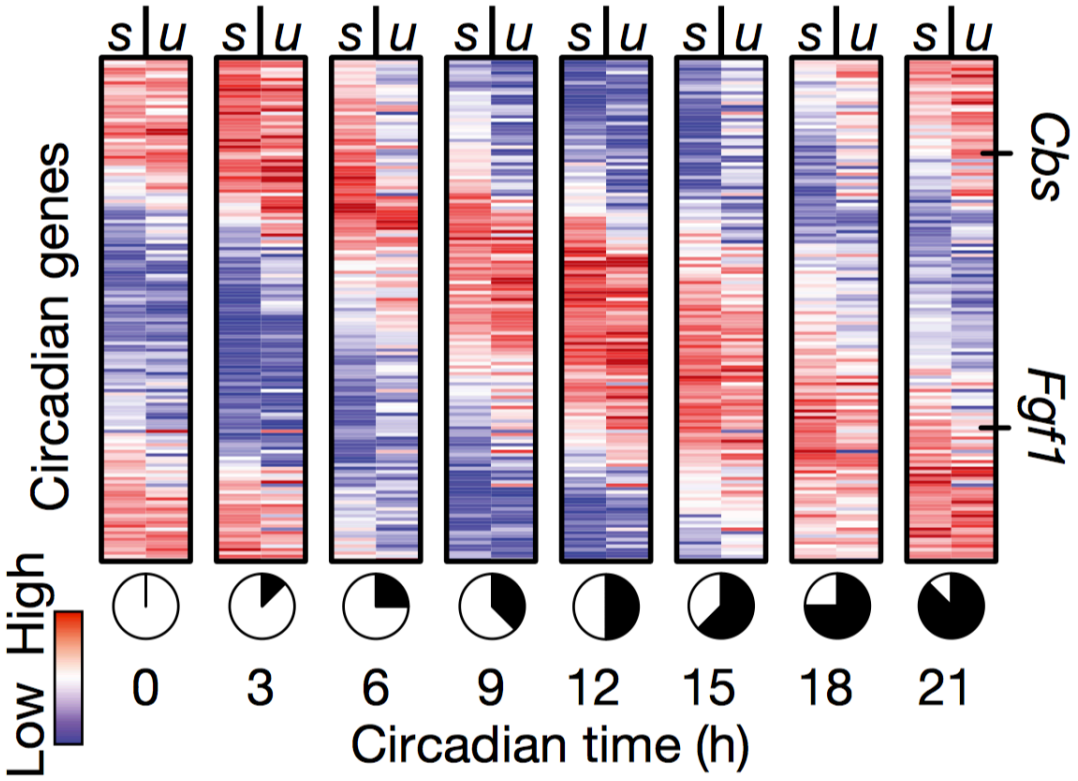
Illustration of the relation between pre-mRNA and mRNA in a single cell

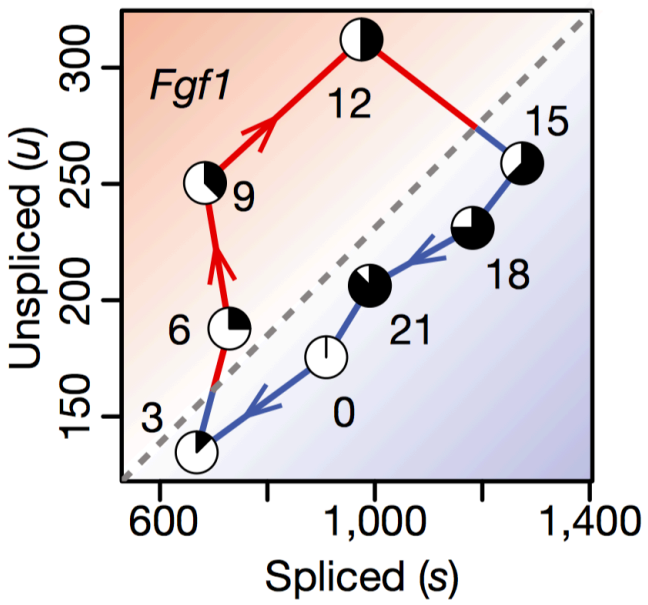


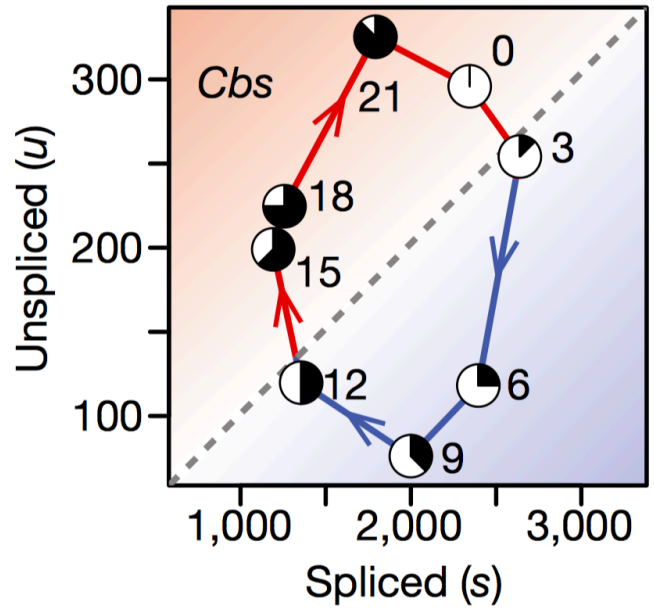
Svensson et al. (2018) Molecular Cell

Abundance of spliced (*s*) and unspliced (*u*) mRNAs for circadian-associated genes

- A time course of bulk RNA-seq measurements of the circadian cycle in the mouse liver
- The unspliced mRNAs are predictive of spliced mRNA at the next time point
- The dashed diagonal line shows the steady-state relationship, as predicted by γ fit

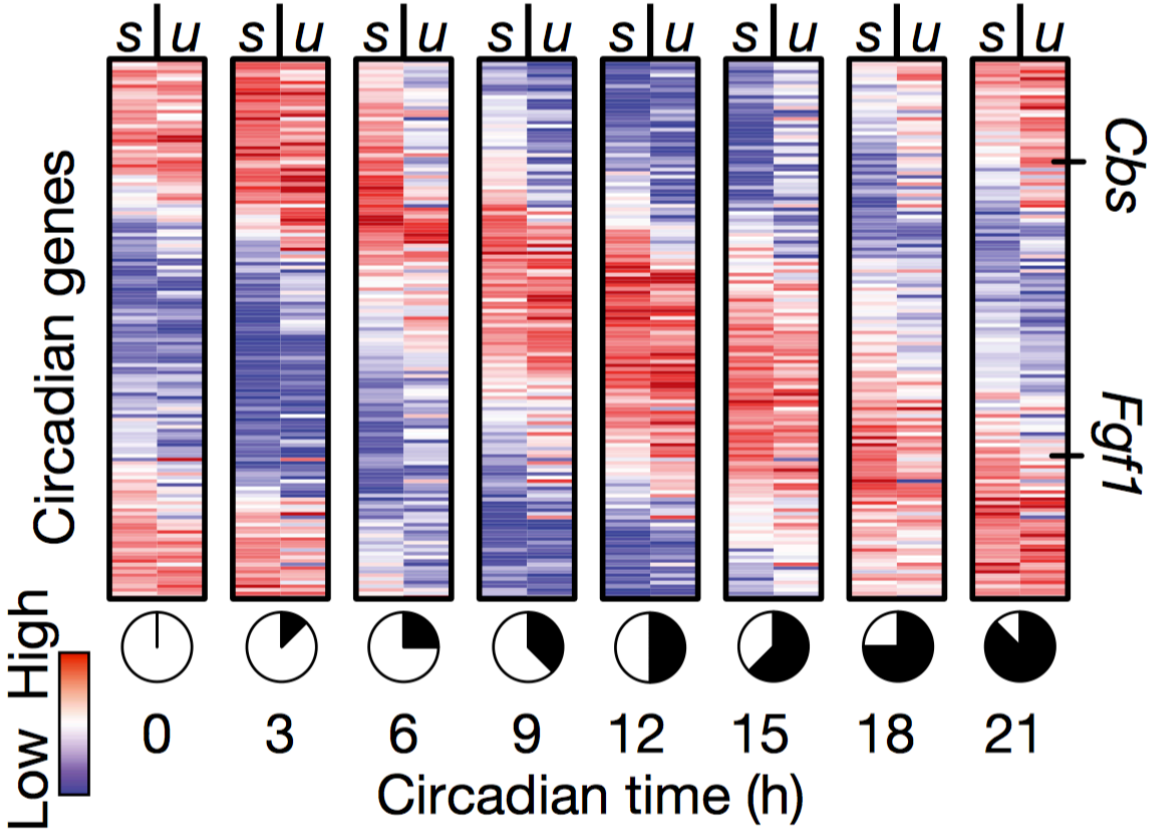


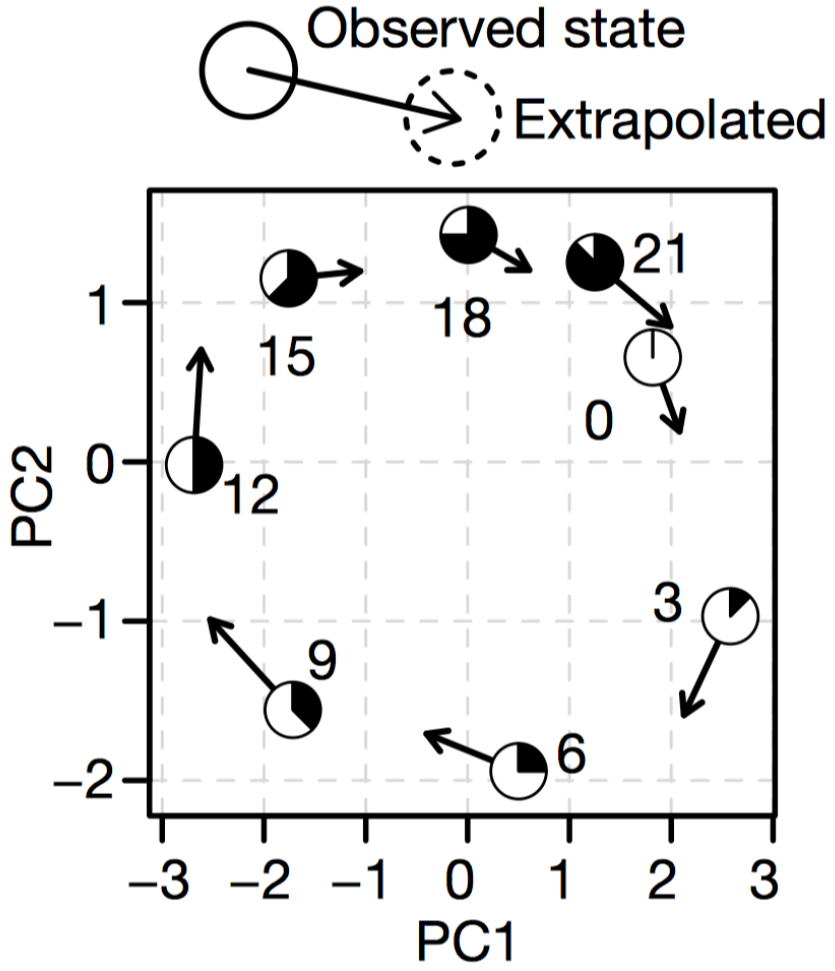




Balance between unspliced and spliced mRNAs is predictive of cellular state progression

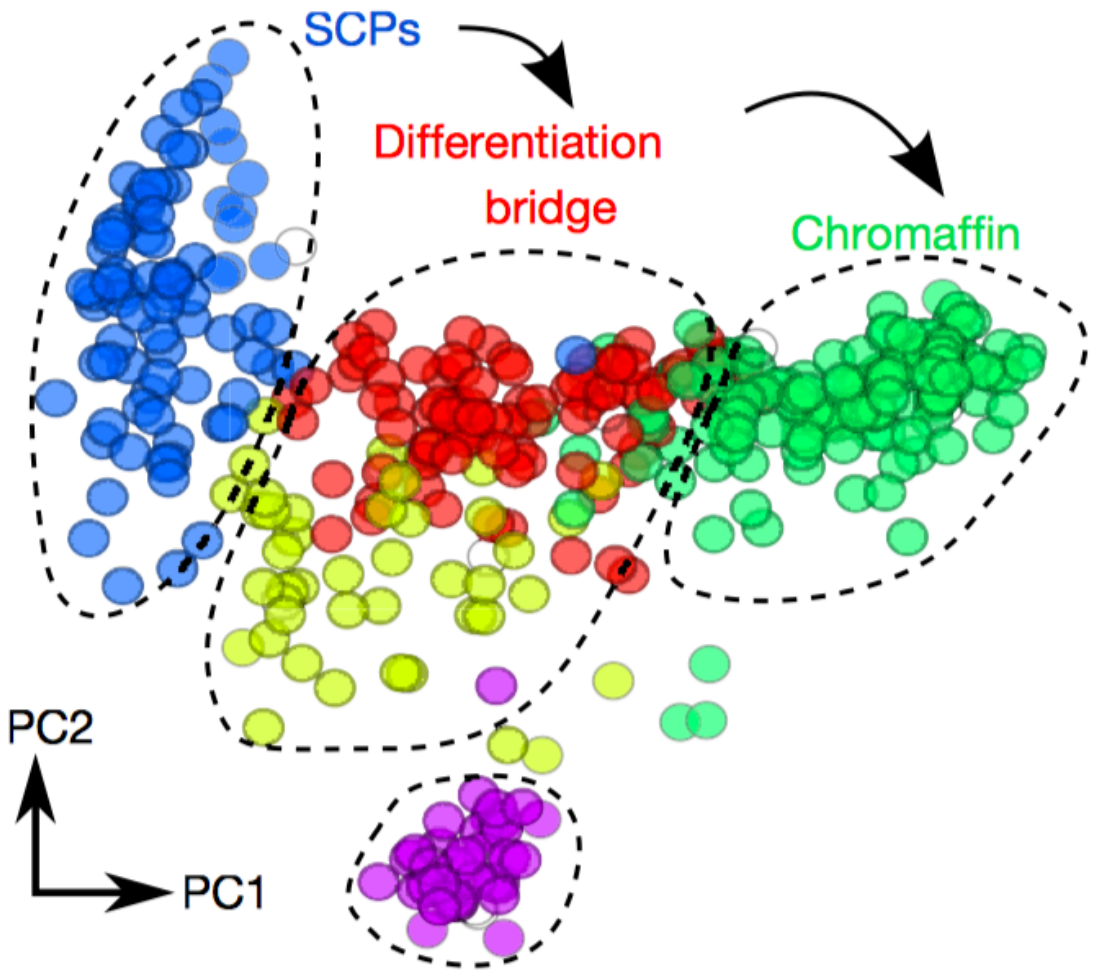
- Assuming gene independence, the overall RNA velocity of the cell is a multidimensional vector comprised of the individual gene velocities

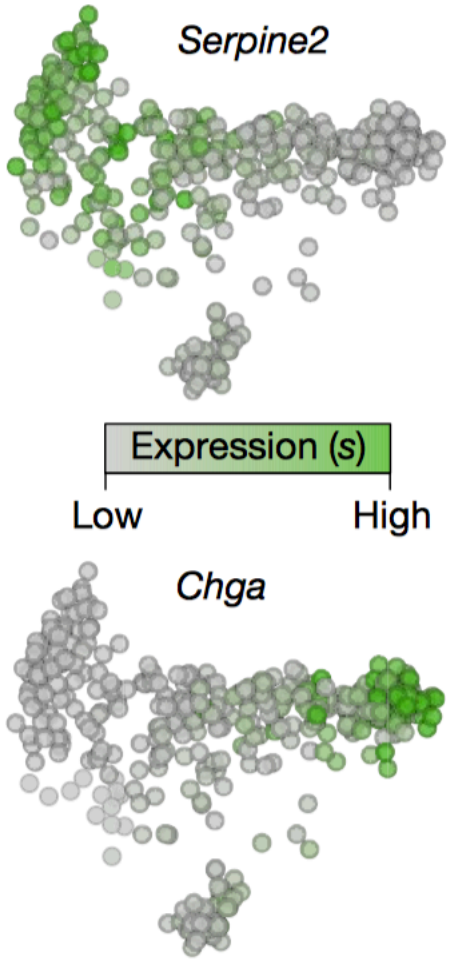


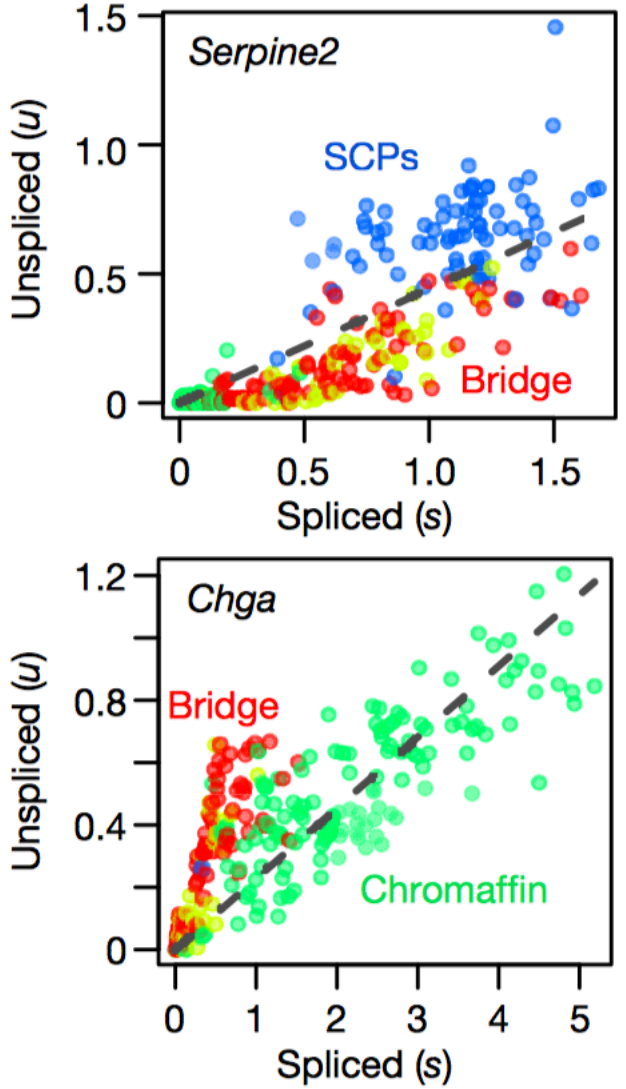


Major subpopulations of Schwann cell precursors (SCPs) differentiate into chromaffin cells

- Single-cell mRNA-seq data of mouse chromaffin cells (SMART-seq2), E12.5 mice (385 cells)
- The direction of differentiation can be validated by lineage tracing

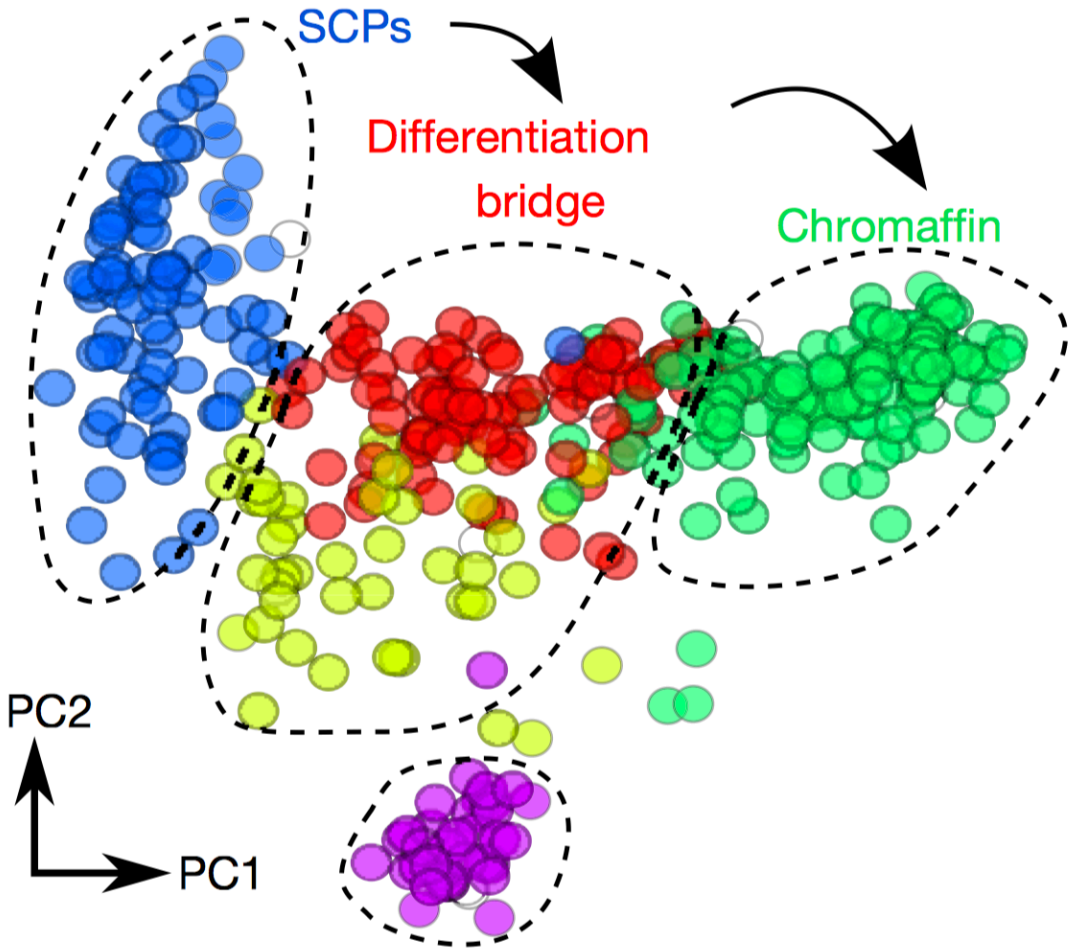


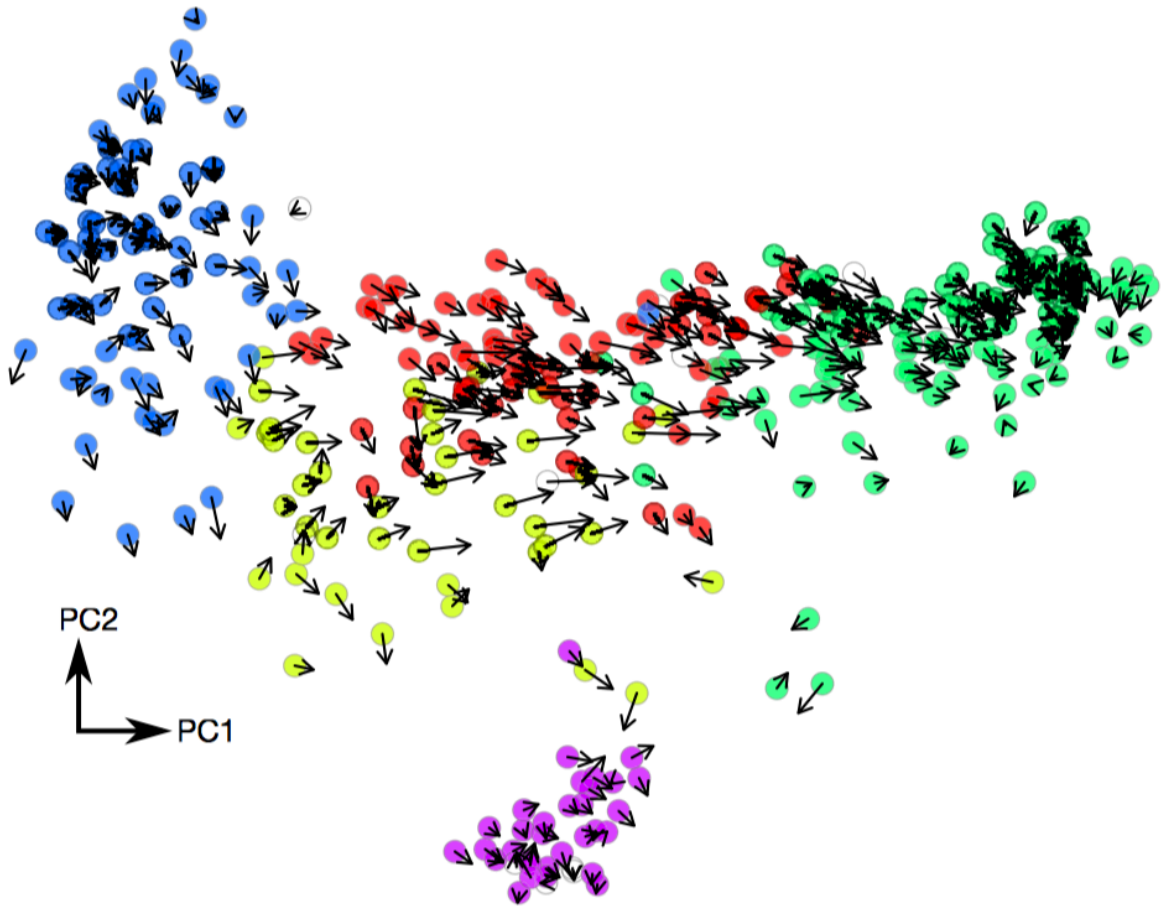




RNA velocity recapitulates dynamics of chromaffin cell differentiation

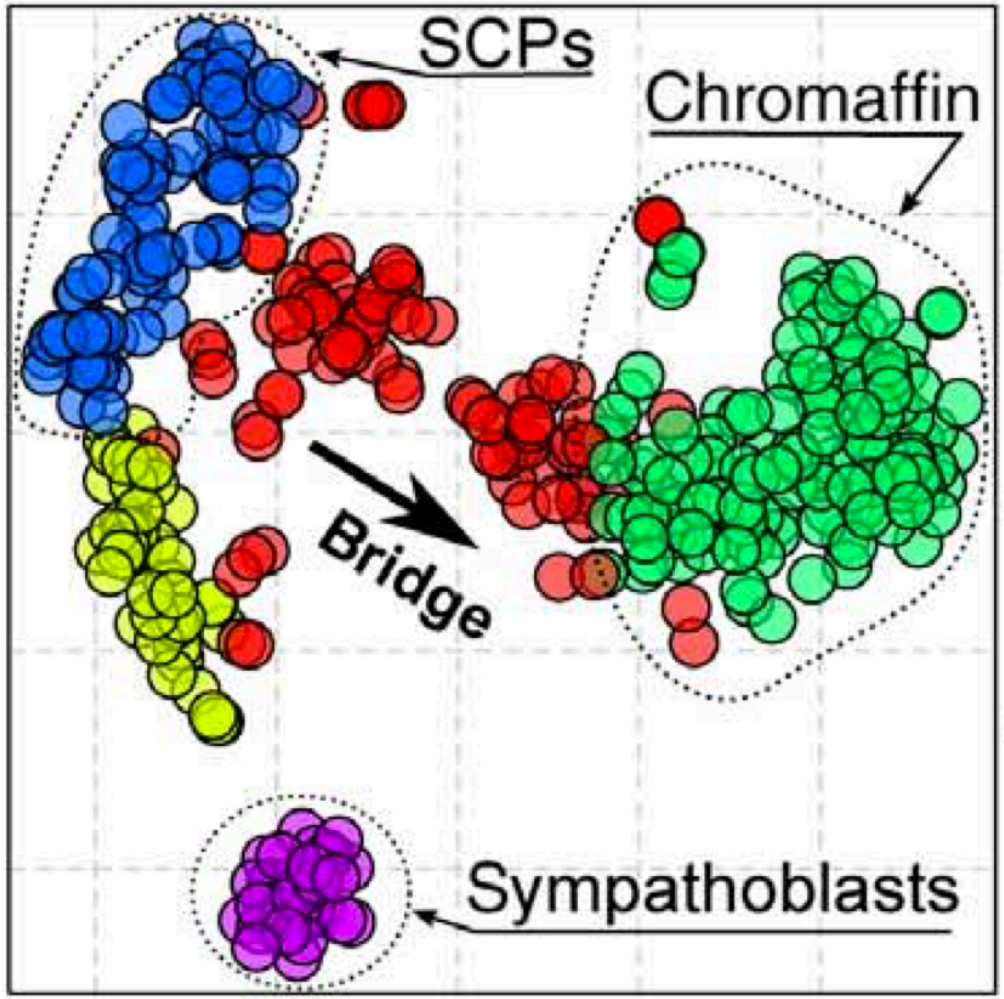
- The observed and extrapolated cell states can be jointly embedded in a common low-dimensional space

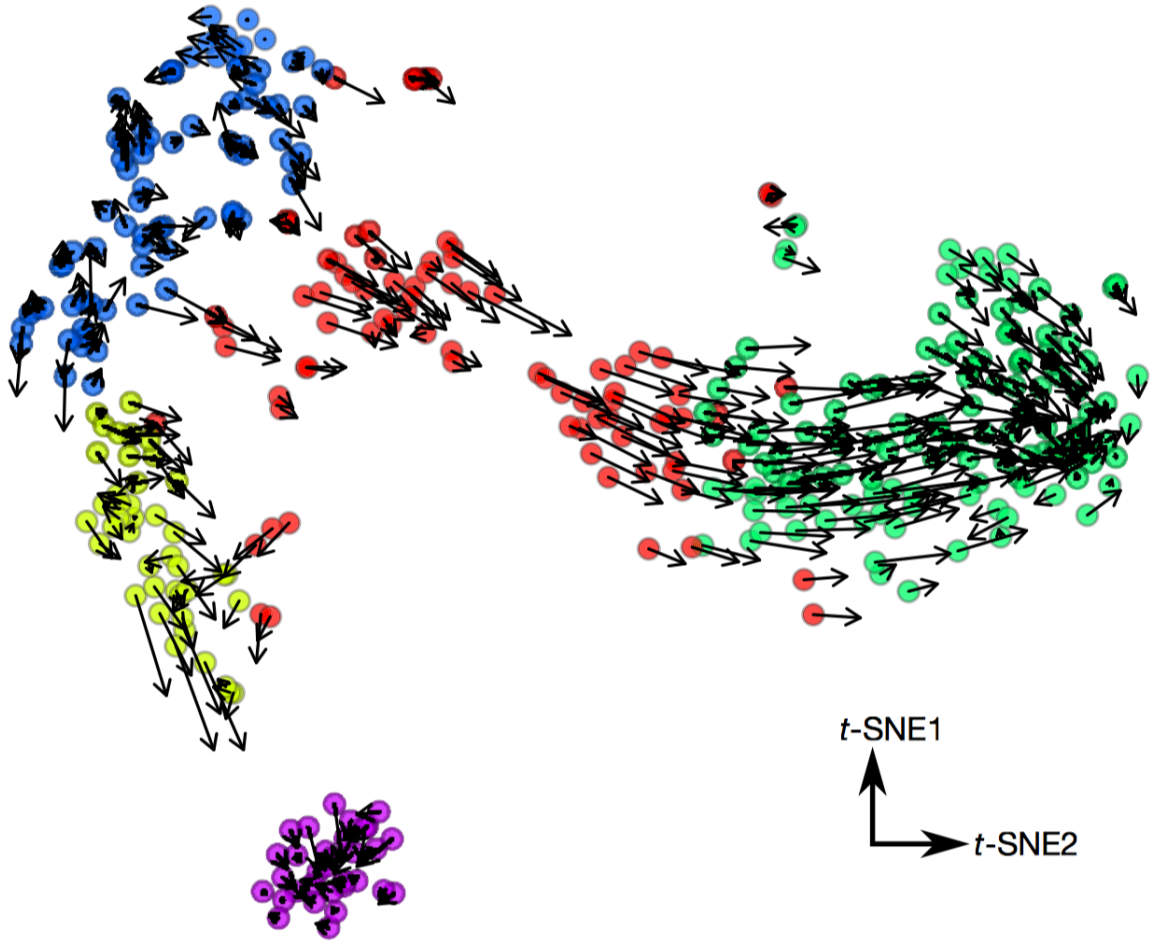




Velocities can be projected onto existing low-dimensional representations (e.g. t-SNE)

- Based the similarity of the extrapolated state to other cells in the local neighbourhood

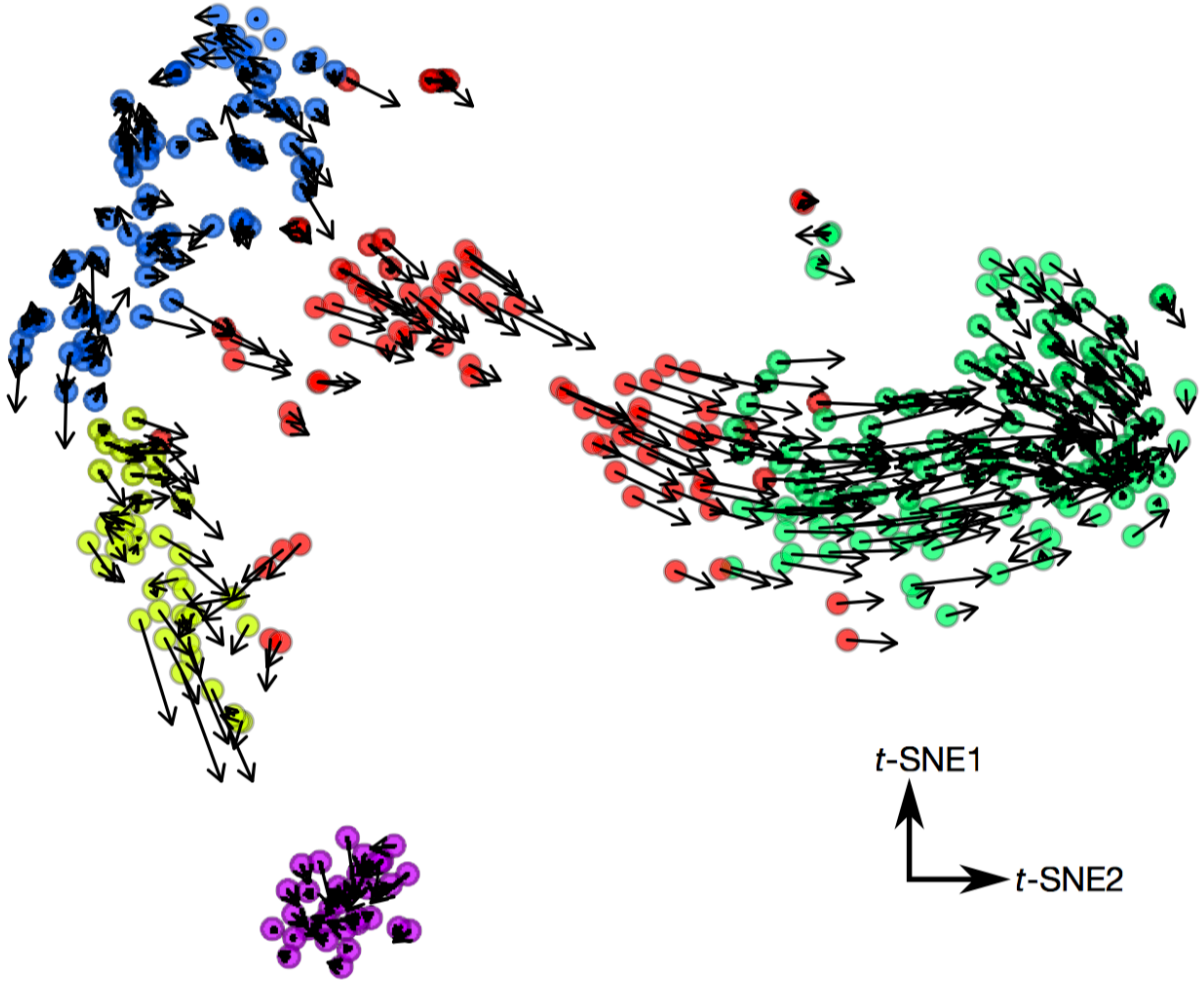


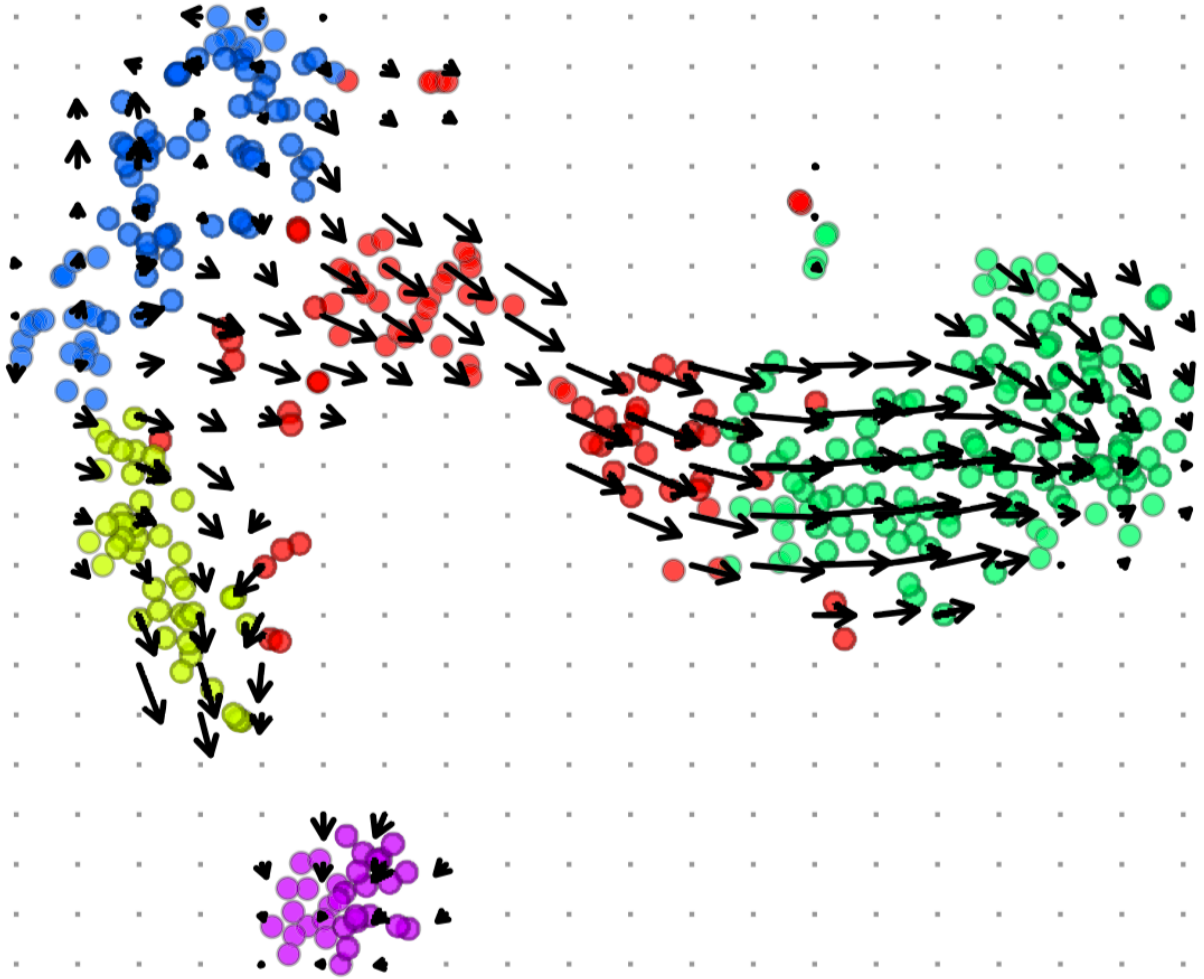


Furlan et al. (2017) Science

Visualize the prevalent pattern of cell velocities with locally averaged vector fields

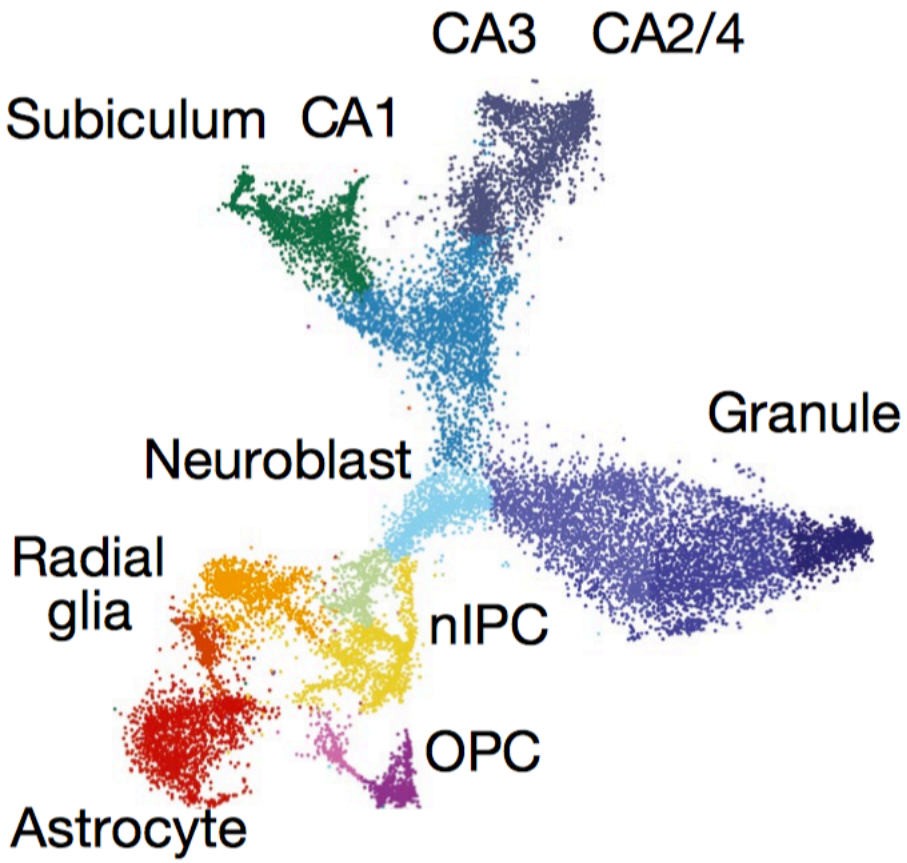
- Suitable for large datasets

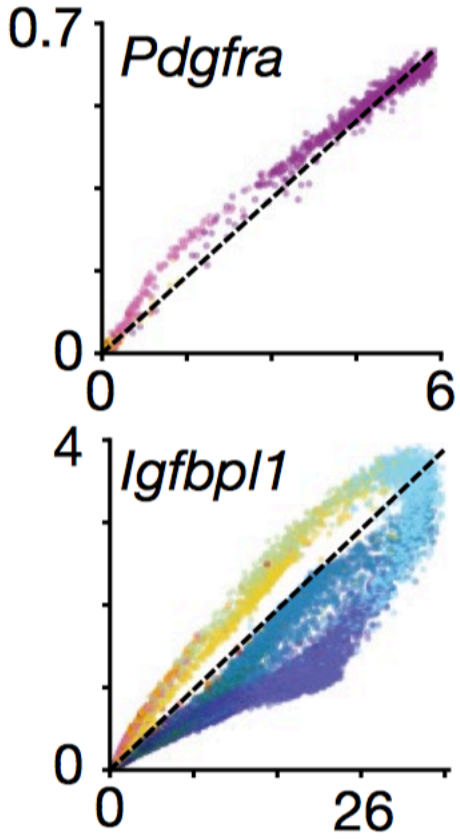


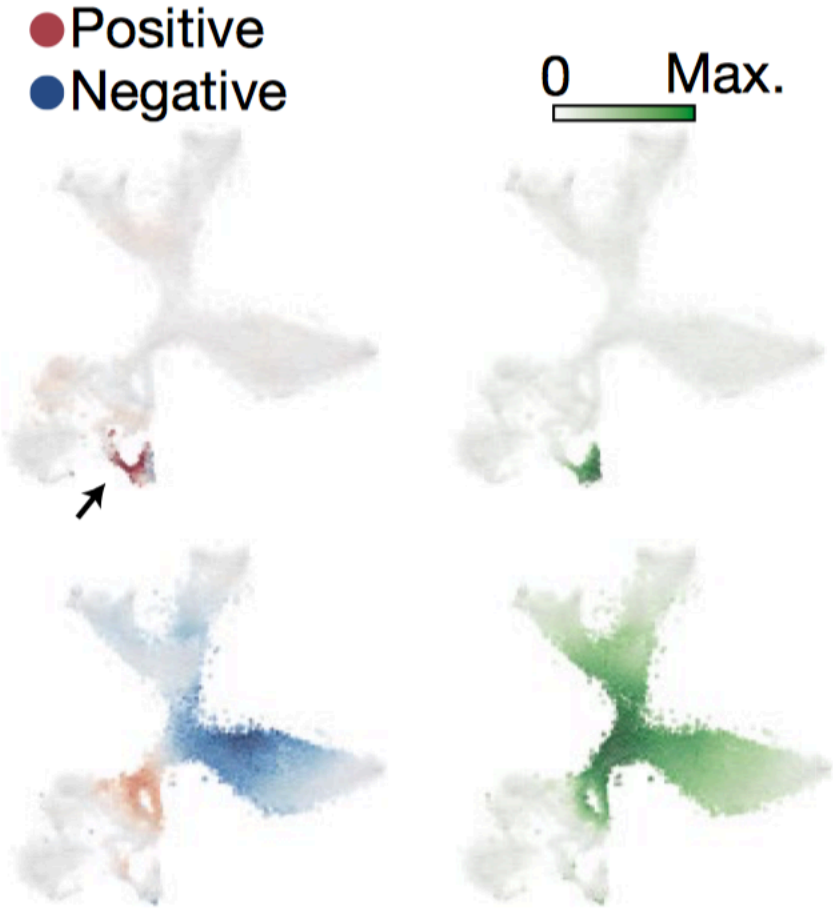


Apply RNA velocity to the branching lineage of the developing mouse hippocampus

- t-SNE plot reveals a complex manifold with multiple branches
- Phase portraits of individual genes showed specific induction and repression of gene expression along the manifold

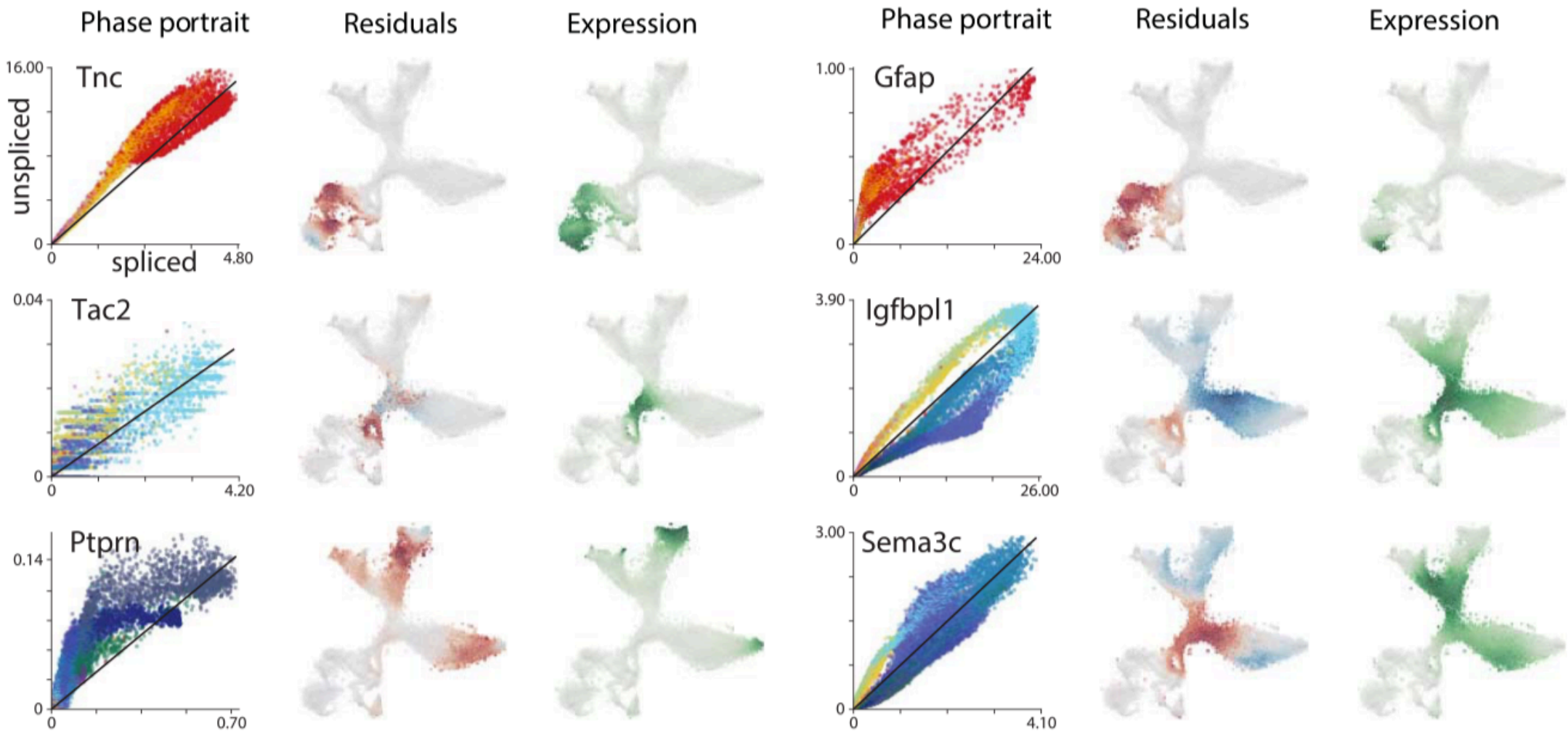






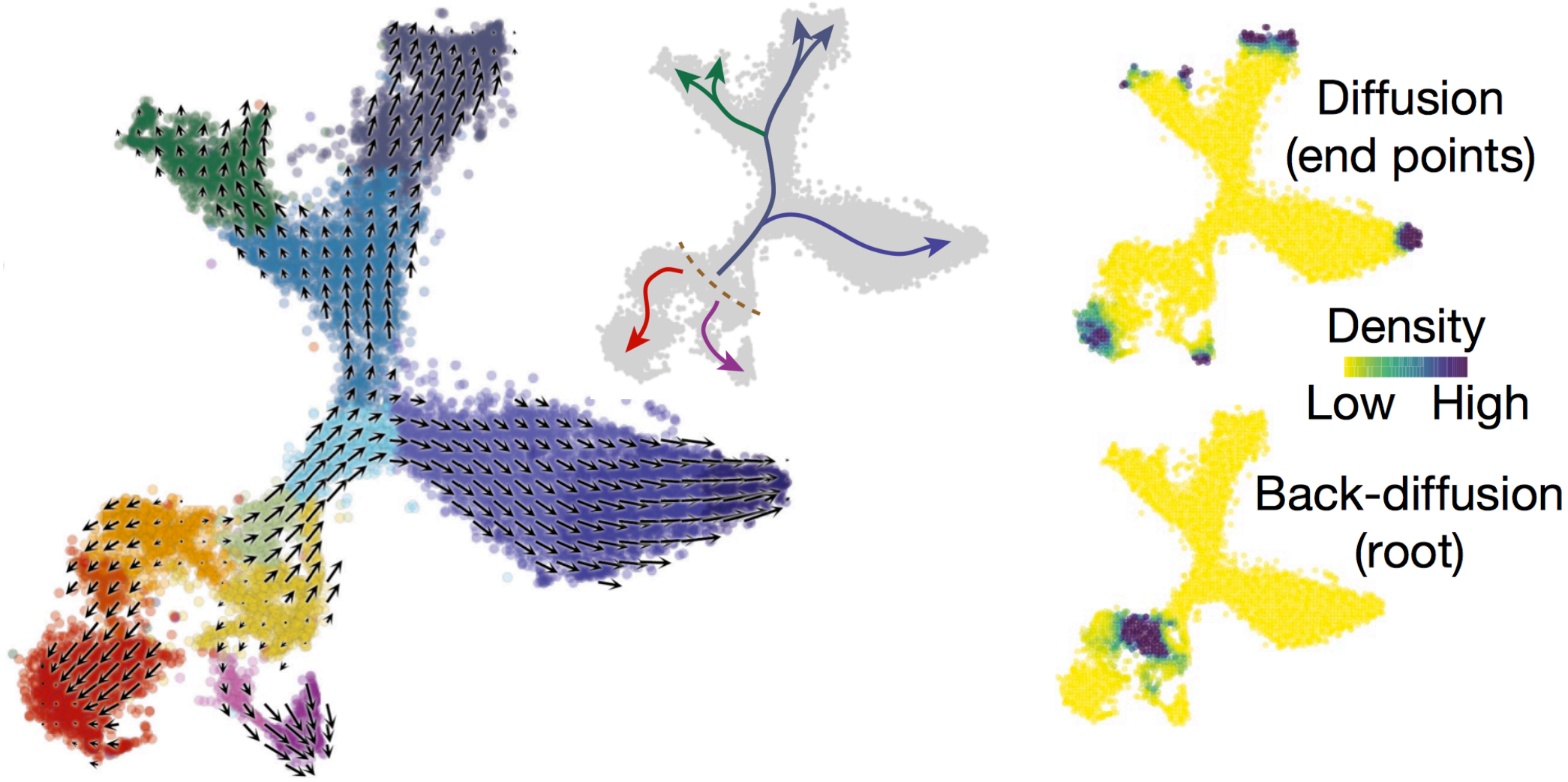
Selected phase portraits and fits of the equilibrium slope (γ) for the developing cells

- The residuals are the difference between observed and expected unspliced abundance, which closely tracks with velocity



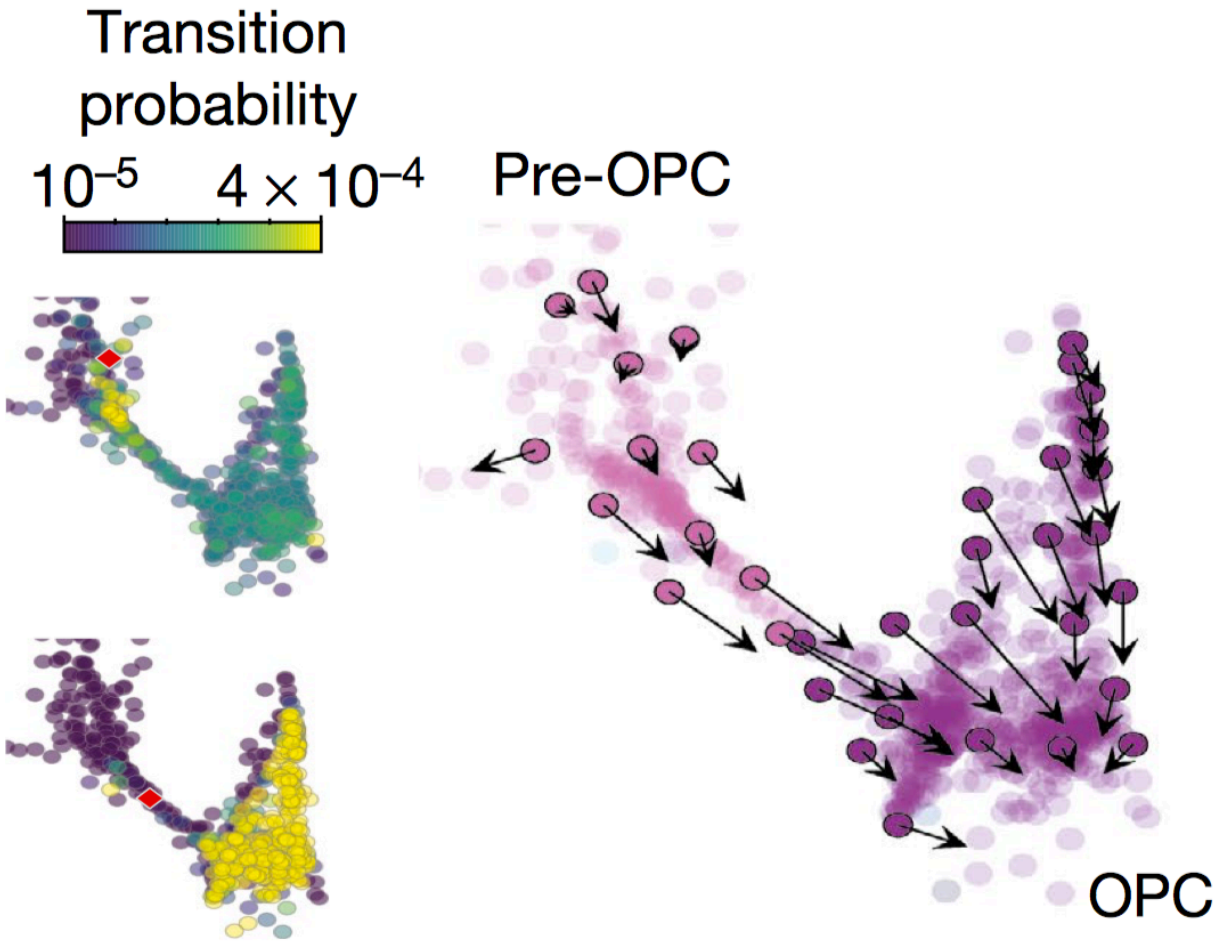
RNA velocity shows a strong directional flow towards each of the main branches

- Using a Markov random-walk model on the velocity field, the terminal and root states could be automatically identified

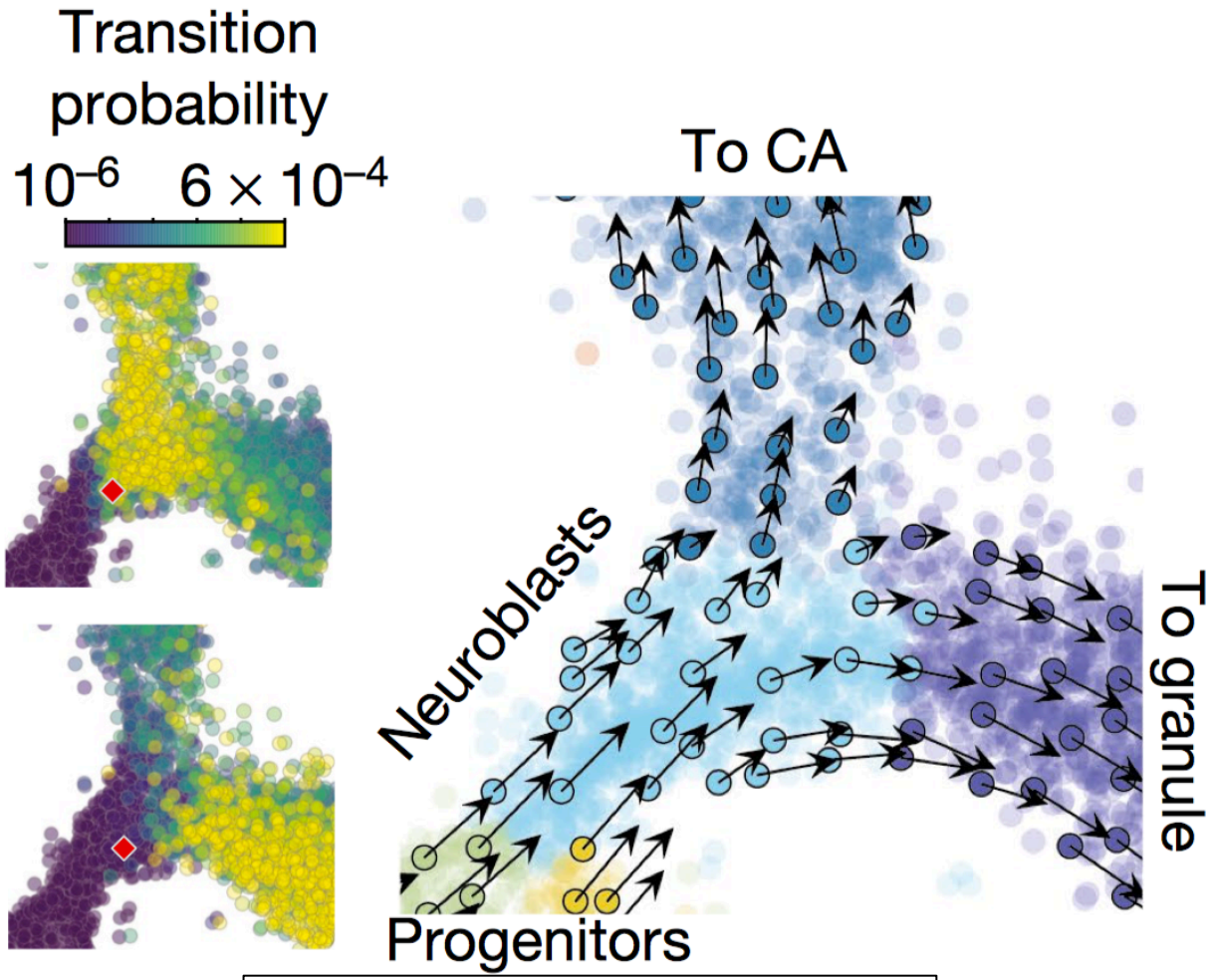


Visualization of single-step transition probabilities from two starting cells (red) to neighbouring cells

- The detailed, single-cell view of a branching lineage allowed us to interrogate fate choice



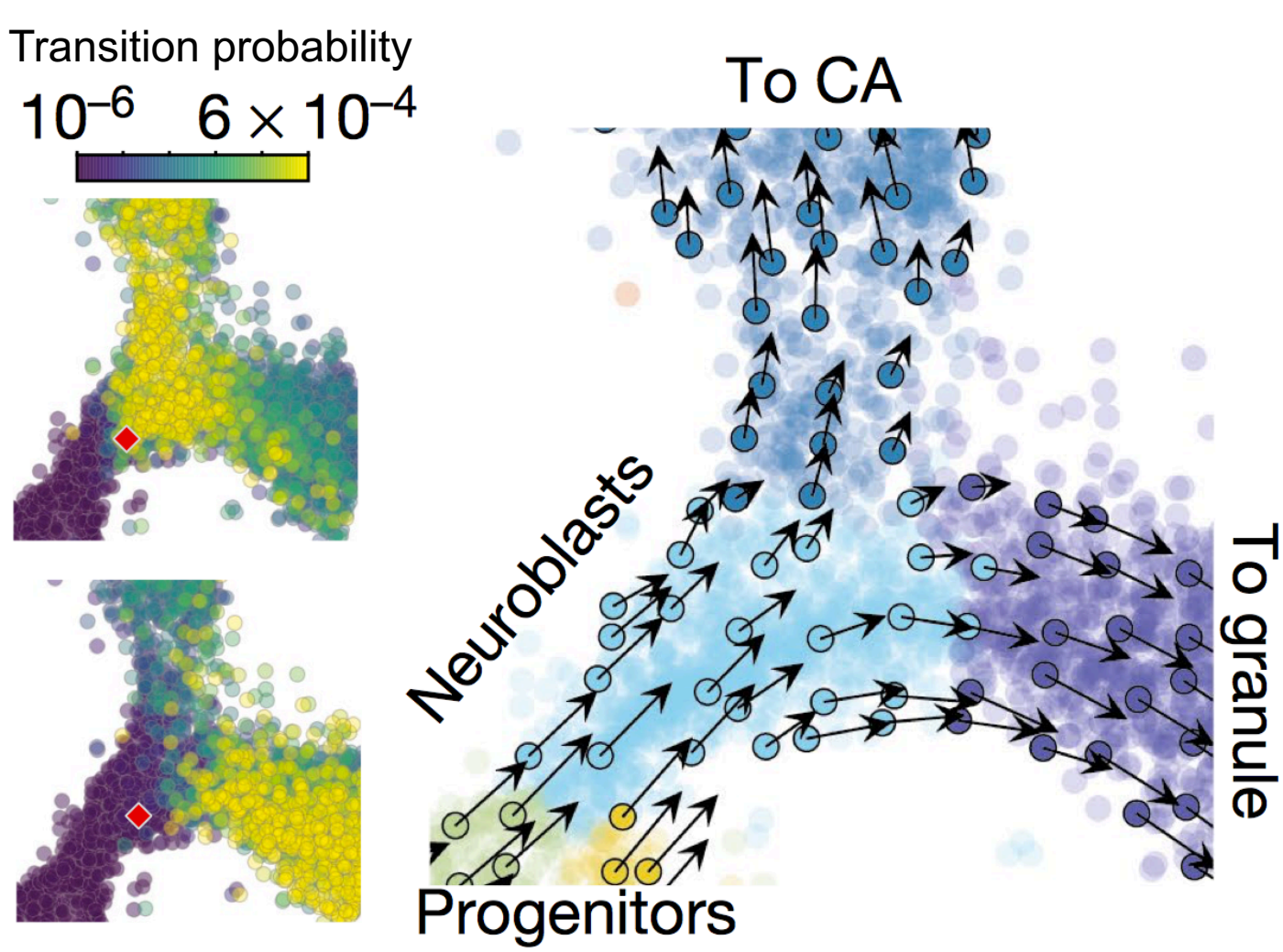
Commitment to oligodendrocyte fate

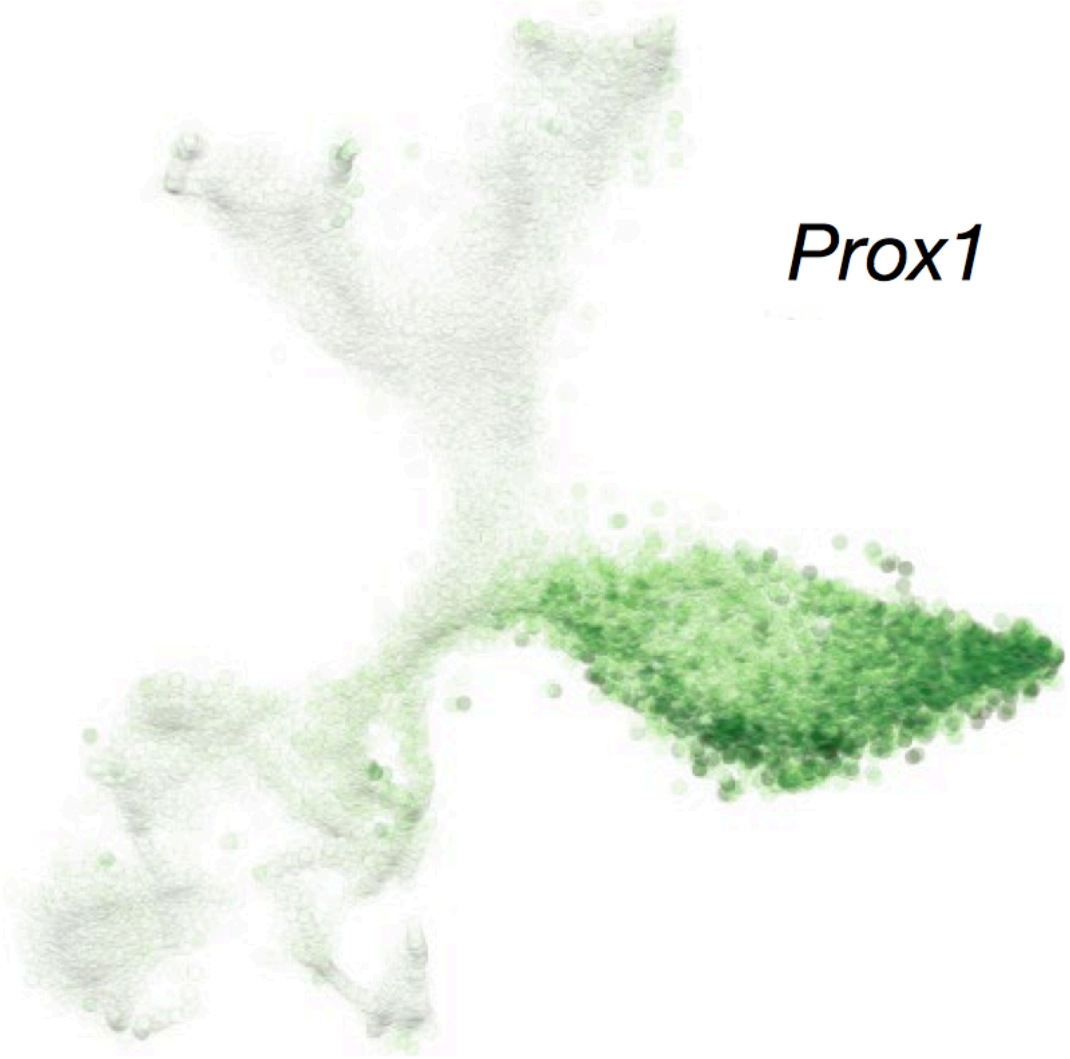


Fate decision of neuroblasts

Two neighbouring cells with different fates are distinguished by activation of *Prox1*

- *Prox1* is required for the formation of granule neurons and that, when *Prox1* is deleted, neuroblasts instead adopt a pyramidal neuron fate





Summary

- Across diverse scRNA-seq pipelines, 15–25% of reads originated from priming in the intronic regions of unspliced molecules

- RNA velocity leverages nascent unspliced RNA to deduce the future transcriptional states of cells, thus providing insights into developmental trajectories

- RNA velocity showed the expected developmental trajectory and helped deduce potentially key developmental transition genes that are responsible for driving cell fate choices

Thank you!

Discussion