Bioinformatic Approaches to RNA Velocity

Basil Khuder
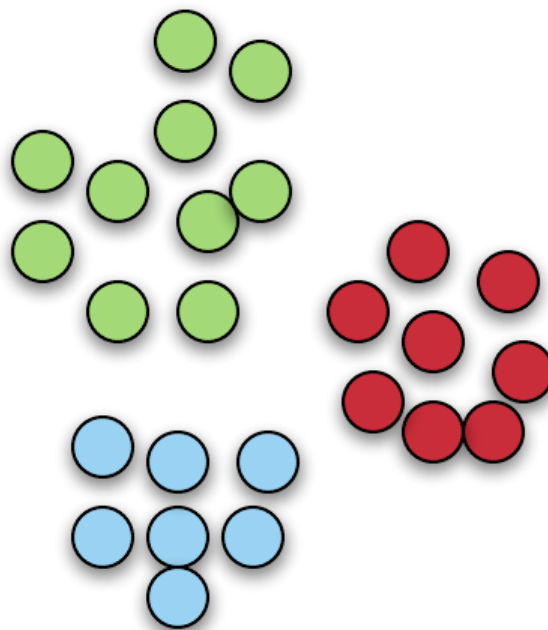
Northwestern Bioinformatics Meeting

2/21/20

Outline

- Mechanisms of RNA Velocity

- Conventional Bioinformatic Approaches

  - Limitations

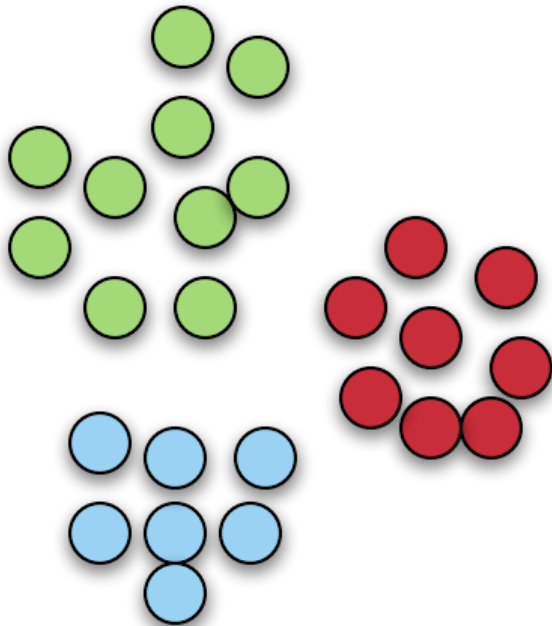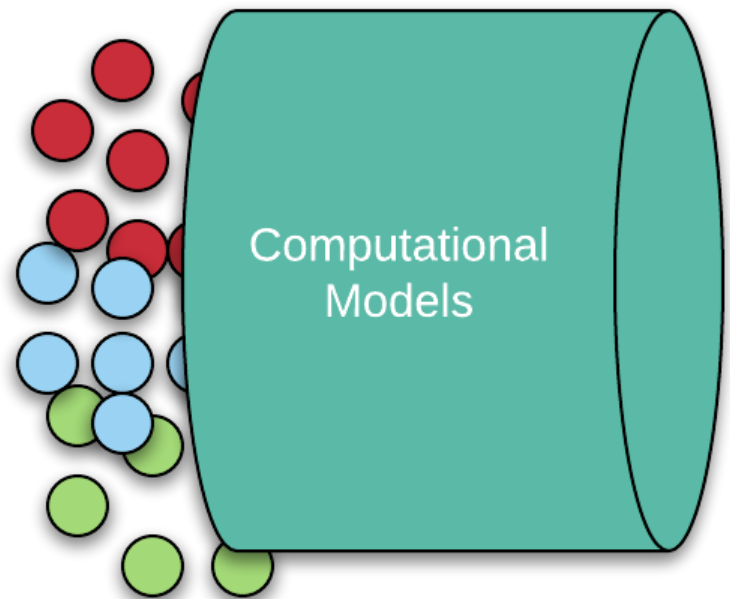- Latest Approaches

- Future Directions

Static

RNA Velocity

Static

Dynamic

Computational Models

RNA Velocity

RNA velocity of single cells

Gioele La Manno, Ruslan Soldatov, Amit Zeisel, Emelie Braun, Hannah Hochgerner, Viktor Petukhov, Katja Lidschreiber, Maria E. Kastriti, Peter Lönnerberg, Alessandro Furlan, Jean Fan, Lars E. Borm, Zehua Liu, David van Bruggen, Jimin Guo, Xiaoling He, Roger Barker, Erik Sundström, Gonçalo Castelo-Branco, Patrick Cramer, Igor Adameyko, Sten Linnarsson ✉ & Peter V. Kharchenko ✉
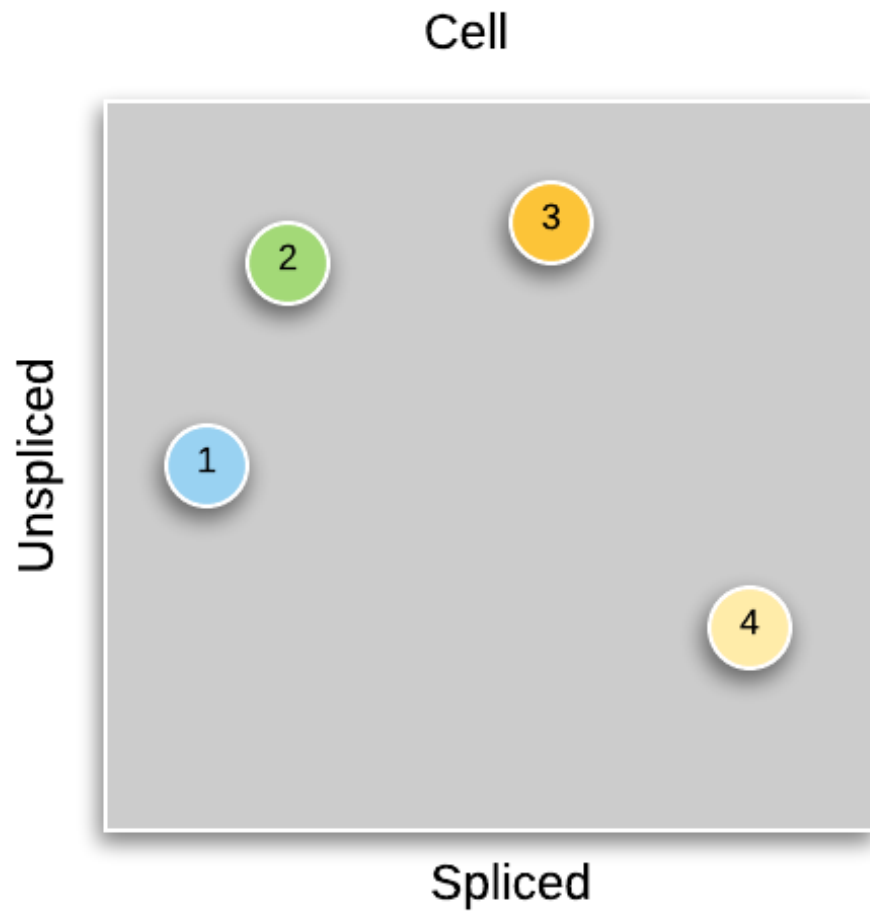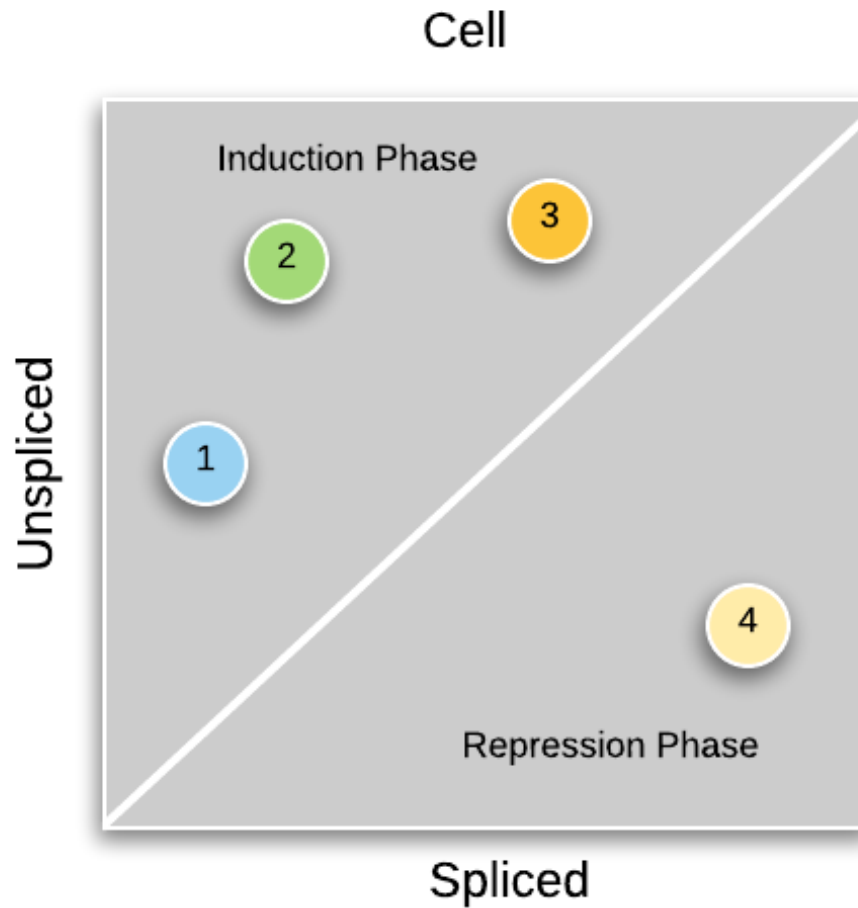
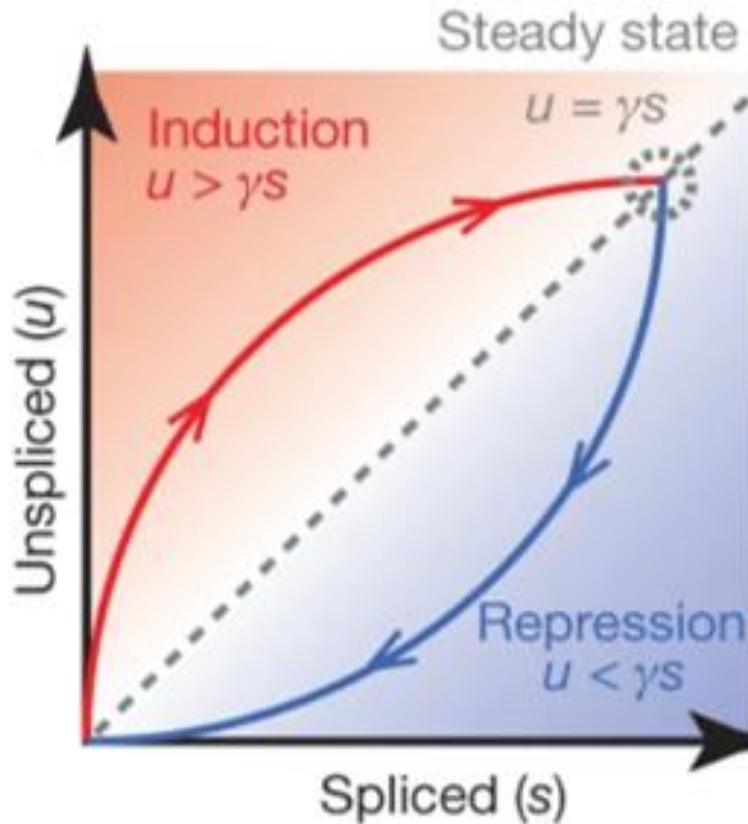# The First Derivative of Gene Expression

RNA Velocity
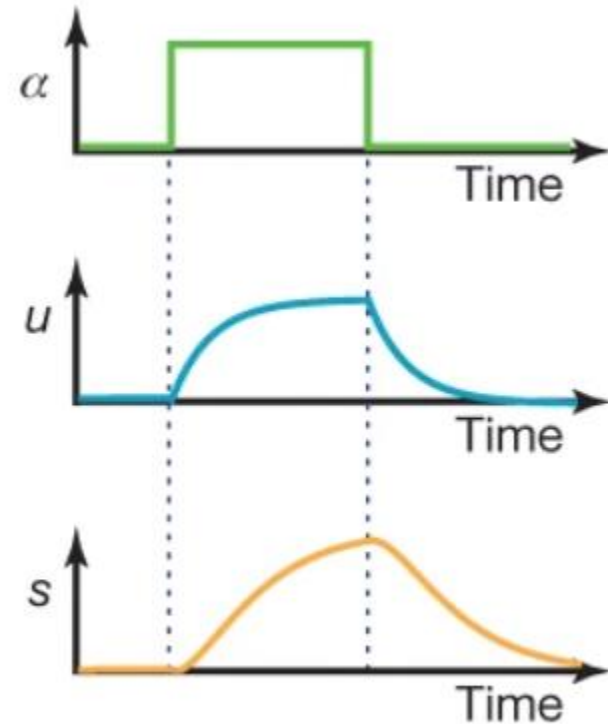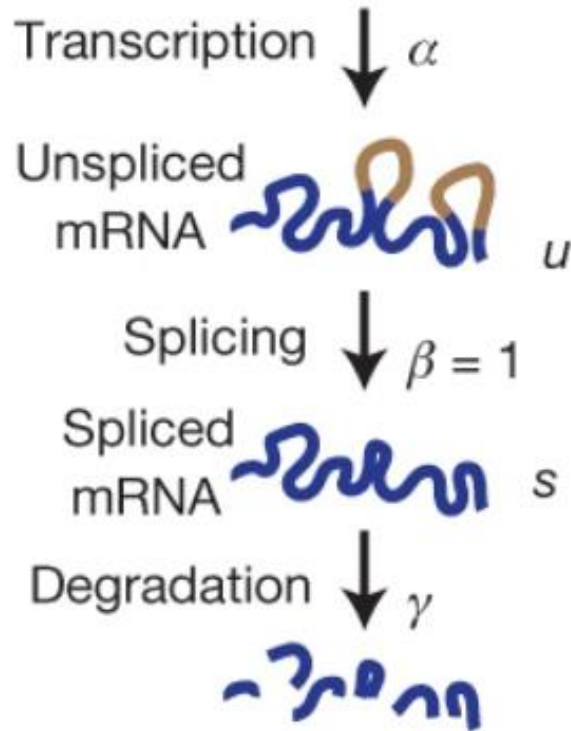
RNA Velocity

RNA Velocity

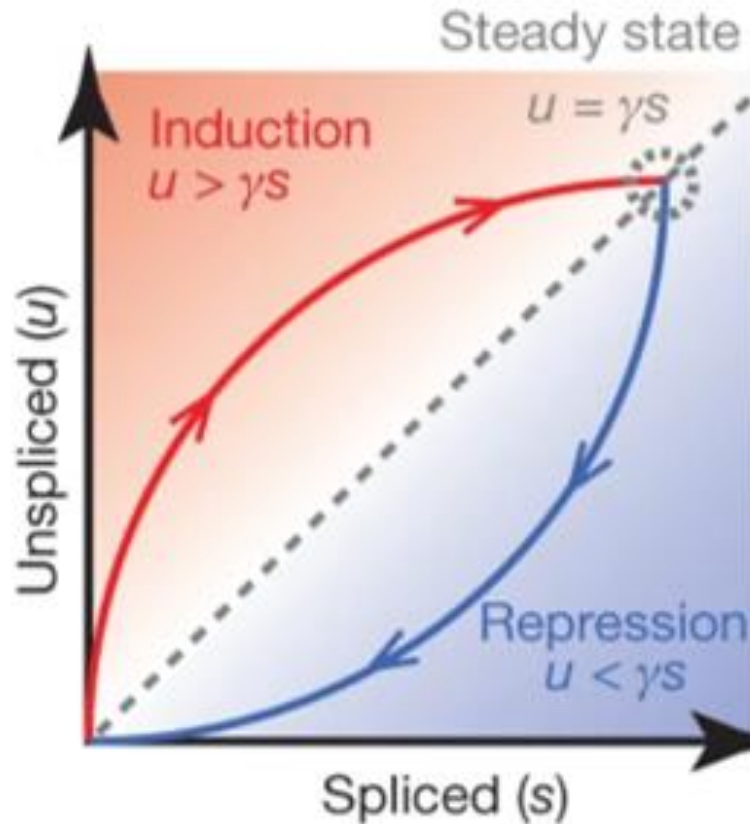# "Steady-State Model"

At the steady-state, the amount of unspliced (u) transcripts being produced is equal to the amount of spliced transcripts (s) multiplied by the degradation rate (γ)

Deviations away from this steady-state denote RNA Velocity

"Steady-State Model"
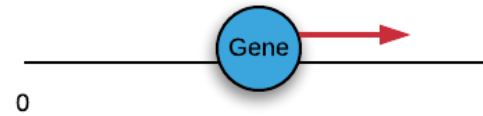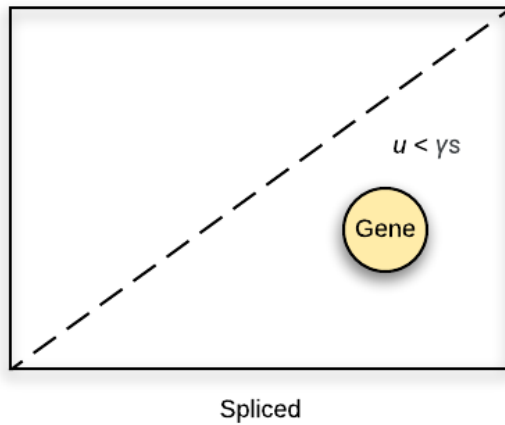

Assumptions:

All genes have the same splicing rate, and will reach a steady state.

"Steady-State Model"
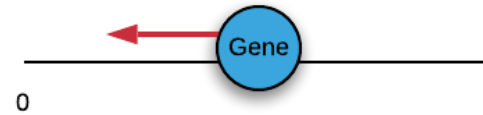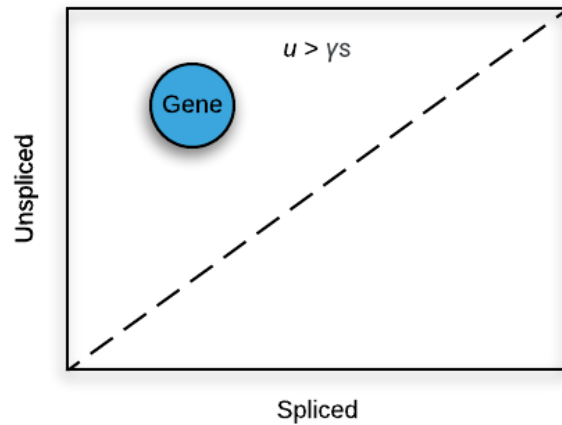
"Steady-State Model"
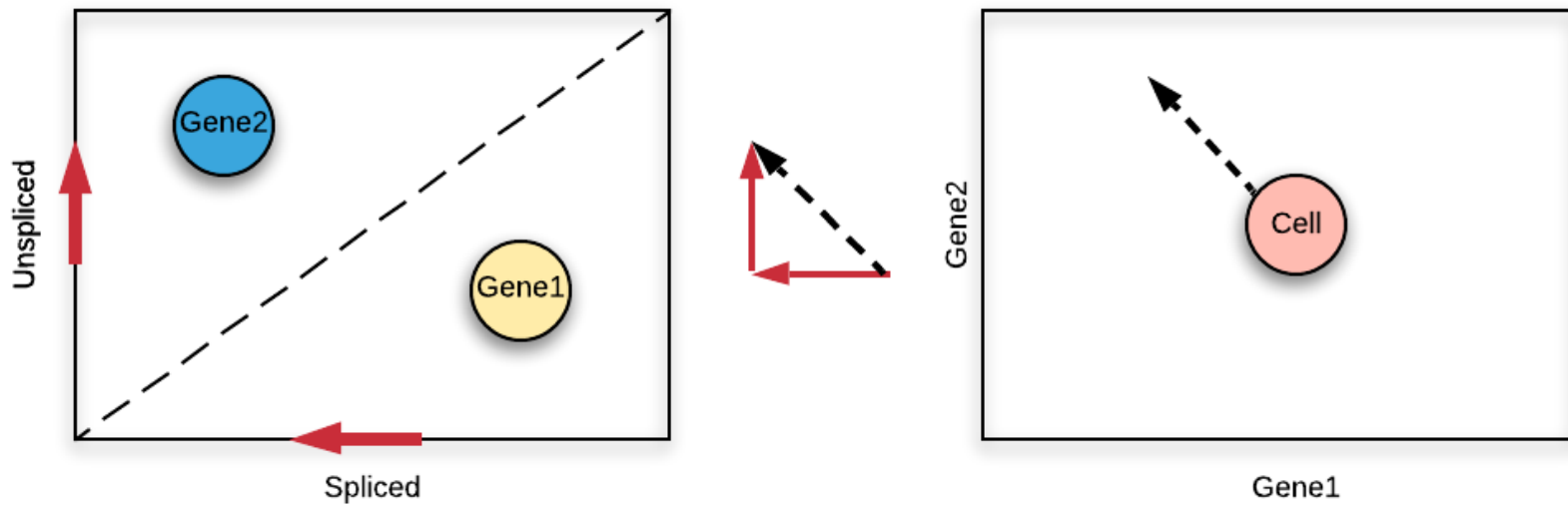
Schwann cells
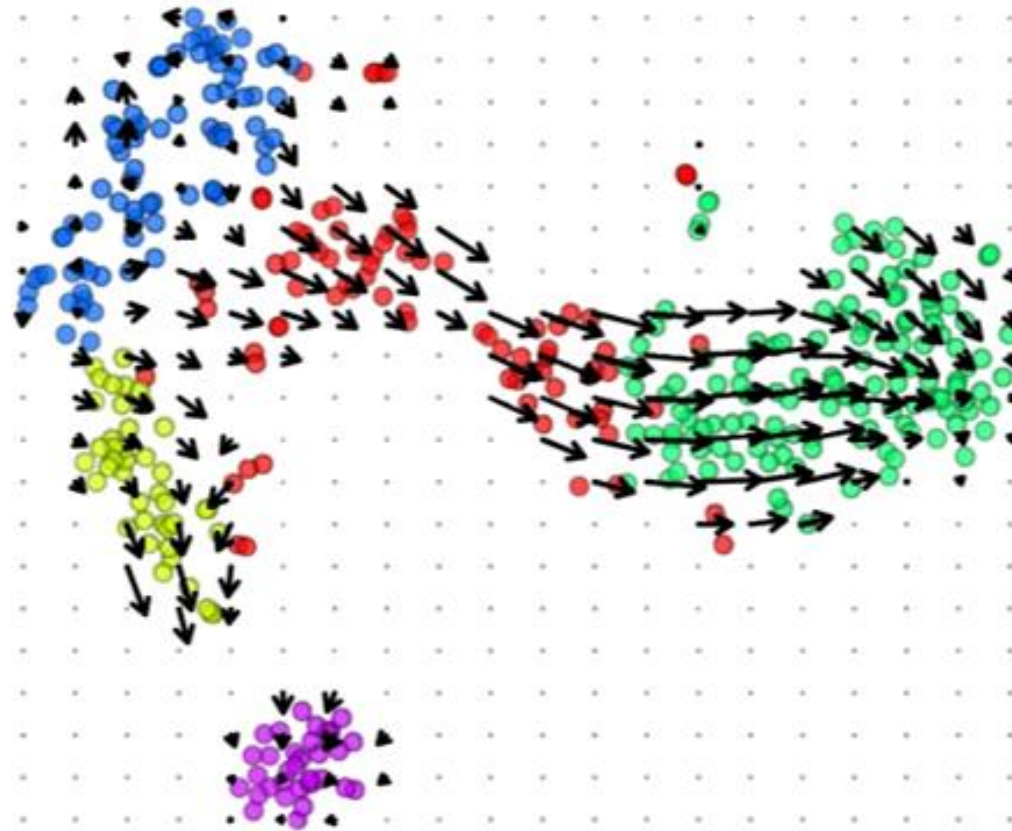
Chromaffin cells

Quantification of RNA-Velocity
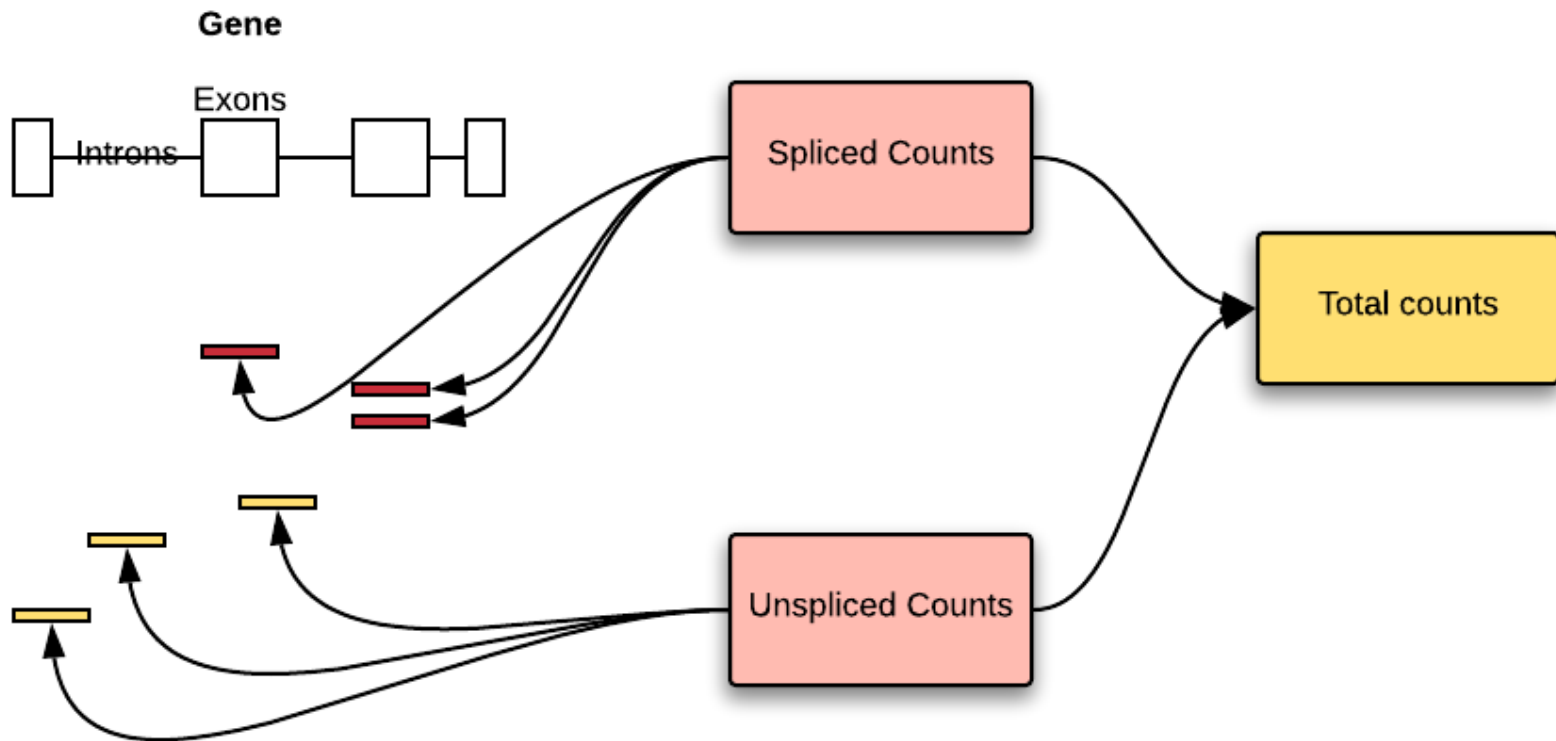
Quantification of RNA-Velocity
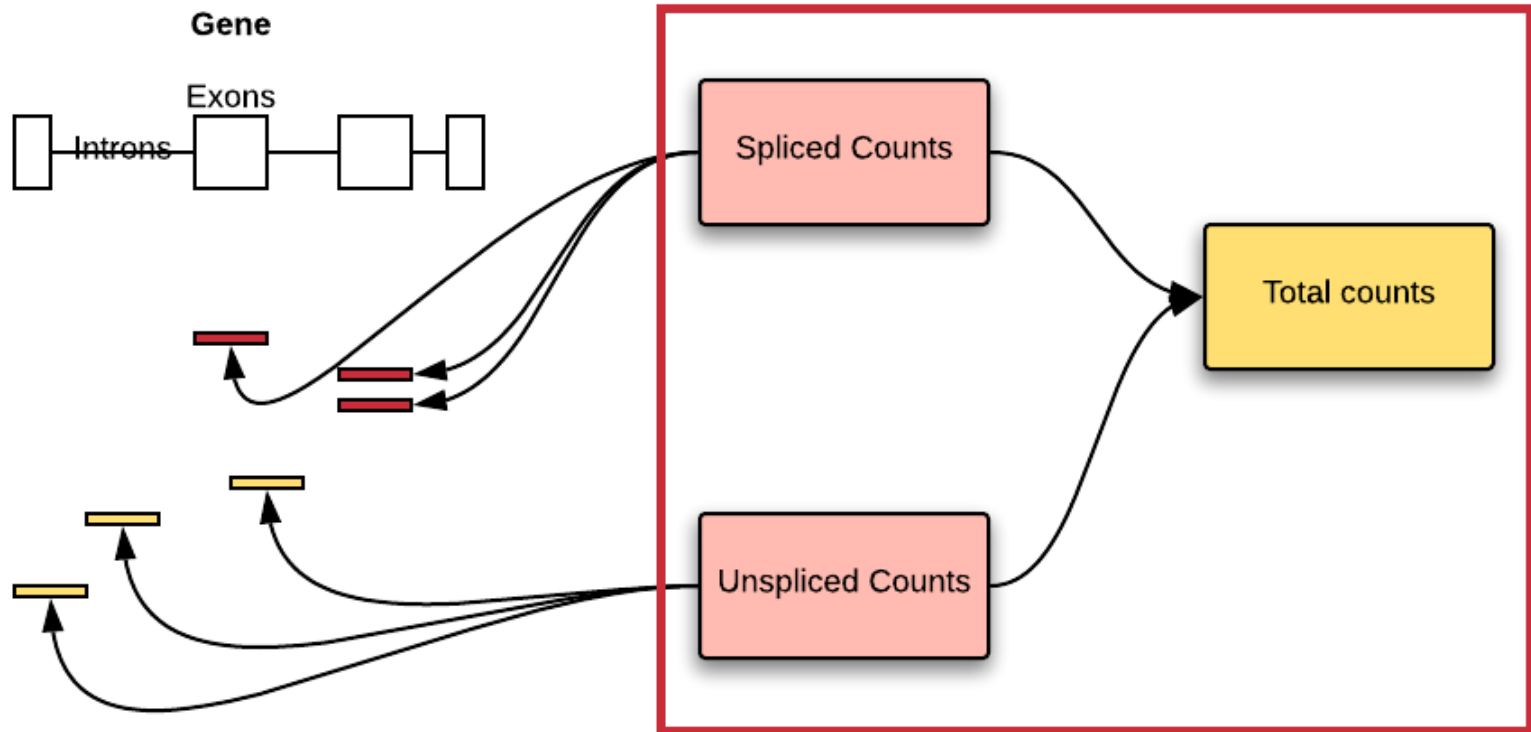
Loom Format and Velocyto
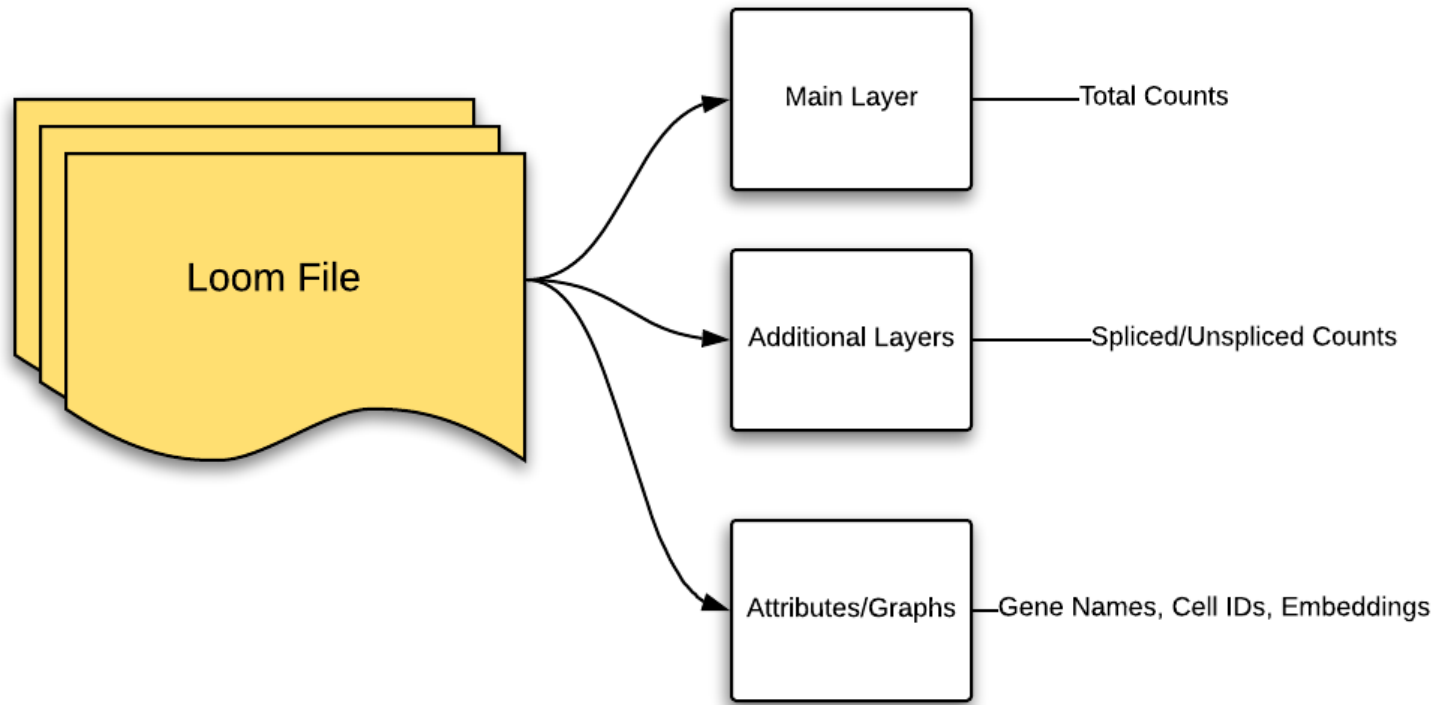
The main matrix in a Loom file is in a chunked format that is automatically compressed and decompressed.

Loom Format and Velocyto (continued...)

- Velocyto (Python) and Velocyto.R
- Both have options to generate unspliced/spliced count matrices.
- Built-in QC and normalization.
- Projection of Velocity onto a manifold.
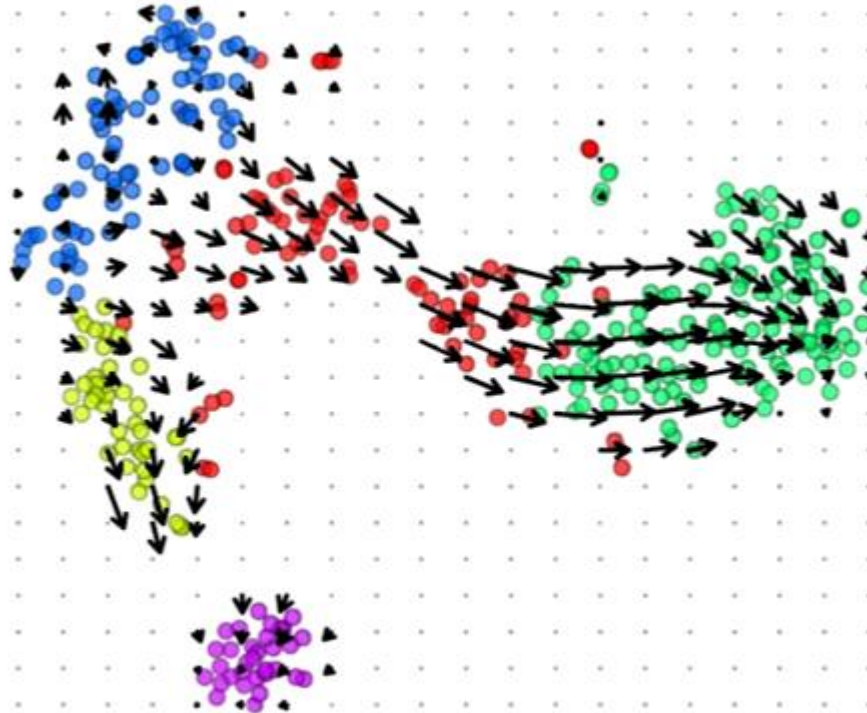
Limitations

- Steady-State Model Assumption

- Computational Resources

- Conflicting Programming Languages/Poor Integration *(Velocyto.R)*

Limitations

- ~~Steady-State Model Assumption Violations~~

- **Dynamical Model**

- ~~Computational Resources~~

- **Kallisto BusTools/scVelo**

- ~~Conflicting Programming Languages/Poor Integration~~

- **R Jupyter Notebooks/R-Reticulate/Seurat Wrappers**

Software Packages

Matrix/Counts Generation

- Velocyto Run
- Kallisto-Bustools (KB)

Velocity

- scVelo
- Velocyto

Data Manipulation

- **Anndata (h5ad -> hdf5)**
- **Bustools (Bus)**
- Loompy (Loom)

scVelo
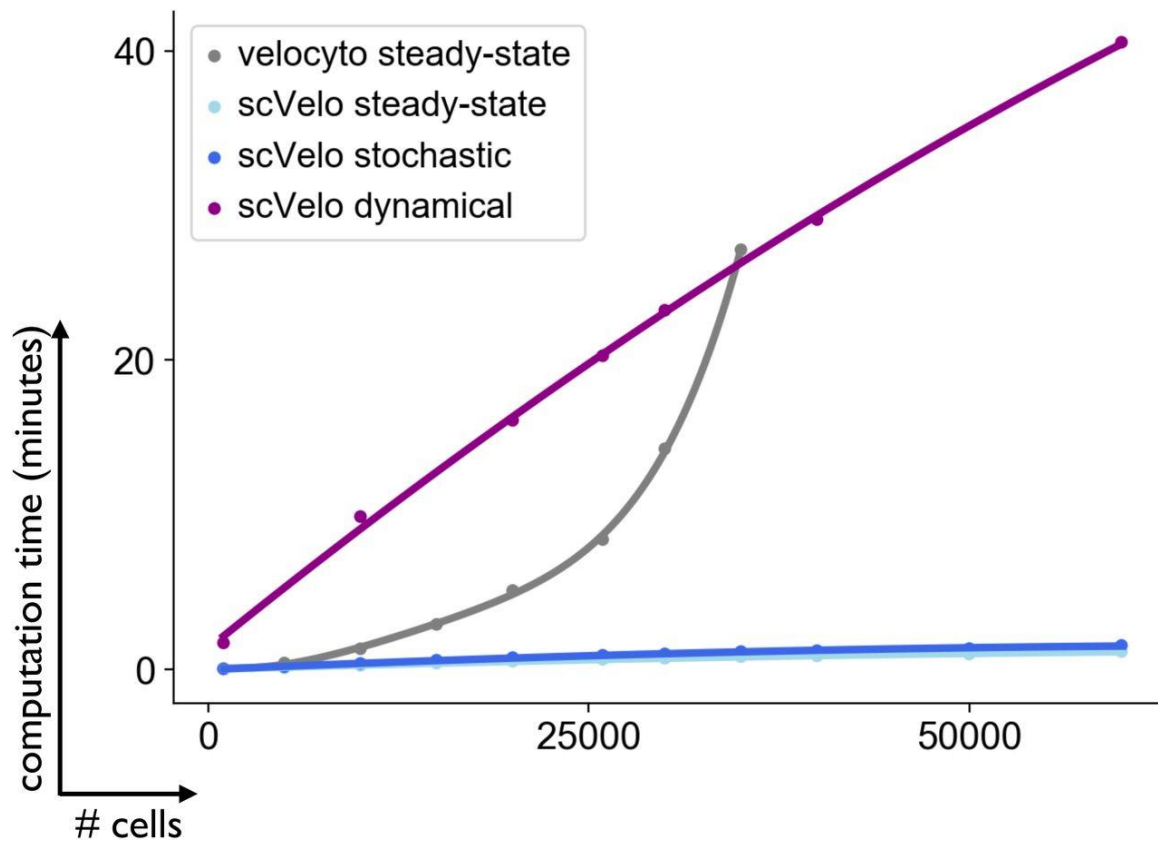
scVelo (continued..)

- 10-fold speed up versus original Velocyto implementation (linear runtime vs quadratic.)
- Includes not only dynamical model, but steady-state (Le manno) as well.

a transcription $u(t)$ splicing $s(t)$ degradation
$\alpha^{on/off}$ β γ

b steady state   early switch
$s(t)$
t
3 hours   1 hour

c time assignment t_i state assignment k_i
latent time
0   1
state likelihood
on   steady   off
u
s
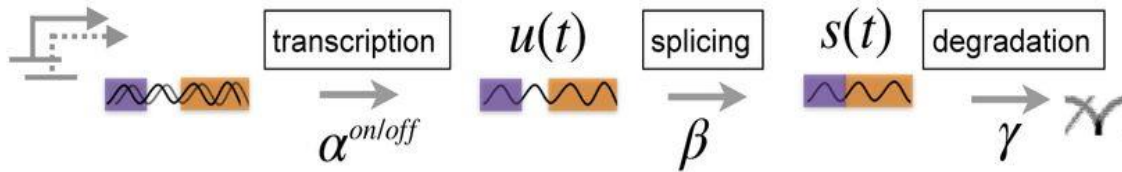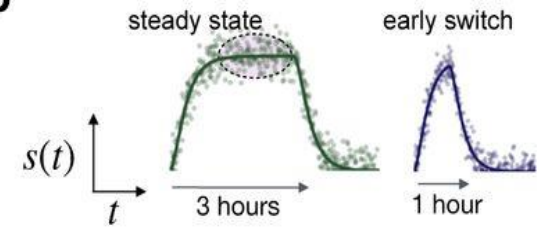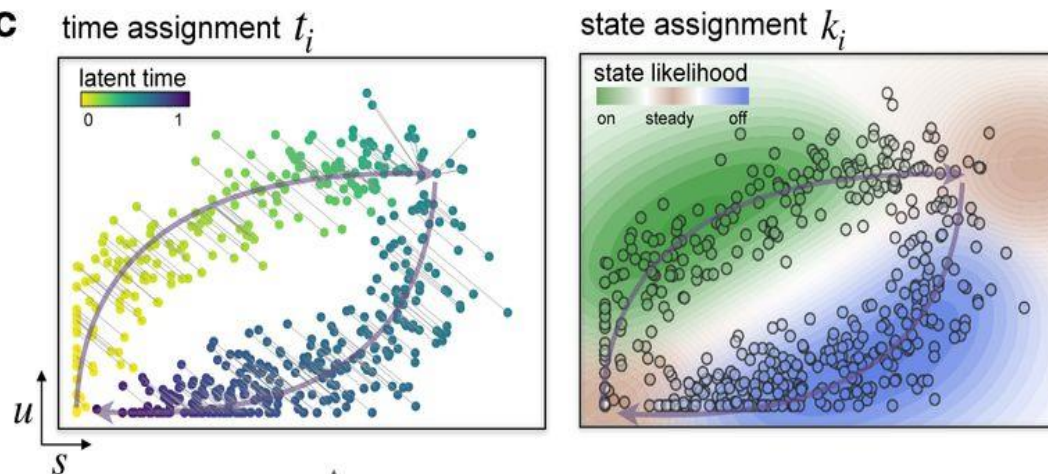
d parameter update $\theta = (\alpha^{on/off}, \beta, \gamma)$
learned kinetics
previous iteration
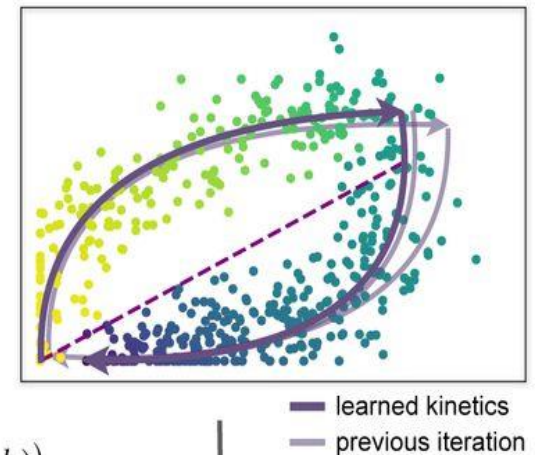
inference by maximizing joint likelihood $P\left((u_i, s_i) \mid (\theta, t_i, k_i)\right)$

Kallisto Bus Tools

- Kallisto "revolutionized" the concept of pseudoalignment which allows for quick alignment of reads by direct comparison of raw reads to transcript sequences.

- Kallisto Bustools allows for pseudoalignment on single-cell data.

- KB-Python is a wrapper that allows for rapid quanitifcation, including for RNA-Velocity matrices.

Kallisto Bus Tools

Lior Pachter
@lpachter

We (w/@sinabooeshaghi @VeigaBeltrame) computed the carbon footprint of running Cell Ranger vs. kallisto bustools for scRNA-seq. Turns out for one dataset it's the difference between driving a car from LA to Mexico vs. driving a few blocks in Pasadena.
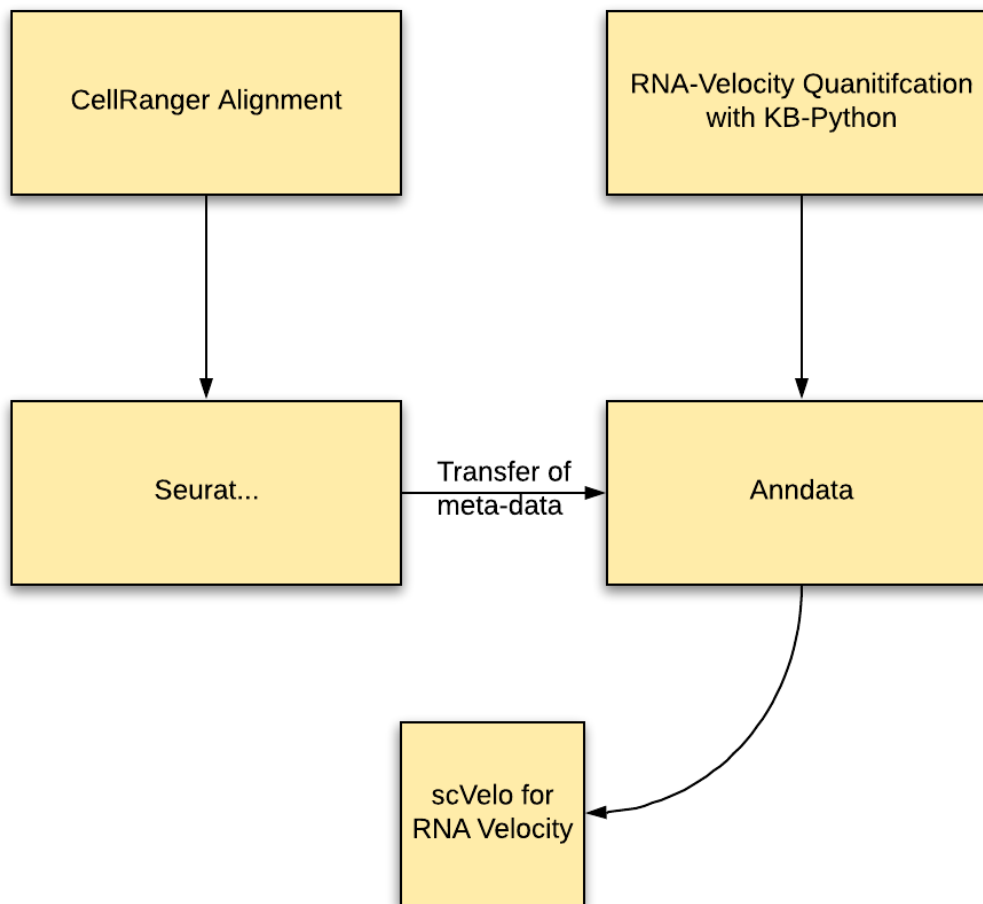
Workflow

Future Directions

"The second derivative of mRNA abundance…"

Protein velocity and acceleration from single-cell multiomics experiments

Gennady Gorin, Valentine Svensson ✉ & Lior Pachter ✉

Genome Biology **21**, Article number: 39 (2020) | Cite this article

49 Altmetric | Metrics

Abstract

The simultaneous quantification of protein and RNA makes possible the inference of past, present, and future cell states from single experimental snapshots. To enable such temporal analysis from multimodal single-cell experiments, we introduce an extension of the RNA velocity method that leverages estimates of unprocessed transcript and protein abundances to extrapolate cell states. We apply the model to six datasets and demonstrate consistency among cell landscapes and phase portraits. The analysis software is available as the *protaccel* Python package.

Protein velocity and acceleration from single-cell multiomics experiments

Gennady Gorin, Valentine Svensson ✉ & Lior Pachter ✉

Abstract

The simultaneous quantification of protein and RNA makes possible the inference of past, present, and future cell states from single experimental snapshots. To enable such temporal analysis from multimodal single-cell experiments, we introduce an extension of the RNA velocity method that leverages estimates of unprocessed transcript and protein abundances to extrapolate cell states. We apply the model to six datasets and demonstrate consistency among cell landscapes and phase portraits. The analysis software is available as the *protaccel* Python package.