

Statistical modelling of cellular dynamics on differentiation and lineage

July 3, 2023 – Single Cell Plus (Banff & virtual)

Yuanhua Huang (黃淵華)

Email: yuanhua@hku.hk | Web: <https://statbiomed.github.io>

School of Biomedical Sciences
Dept Statistics & Actuarial Science



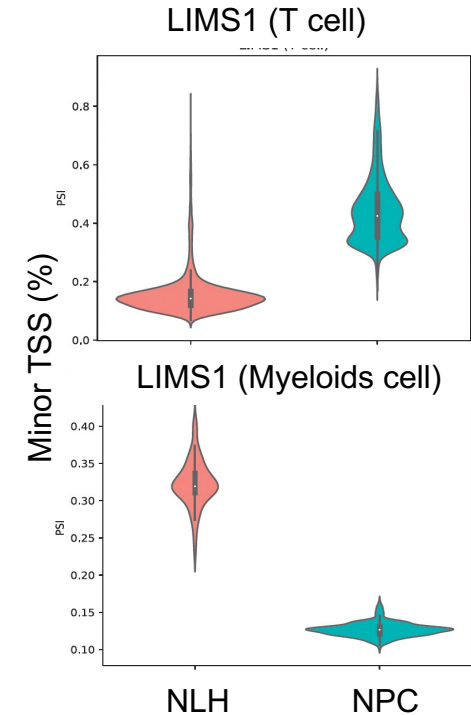
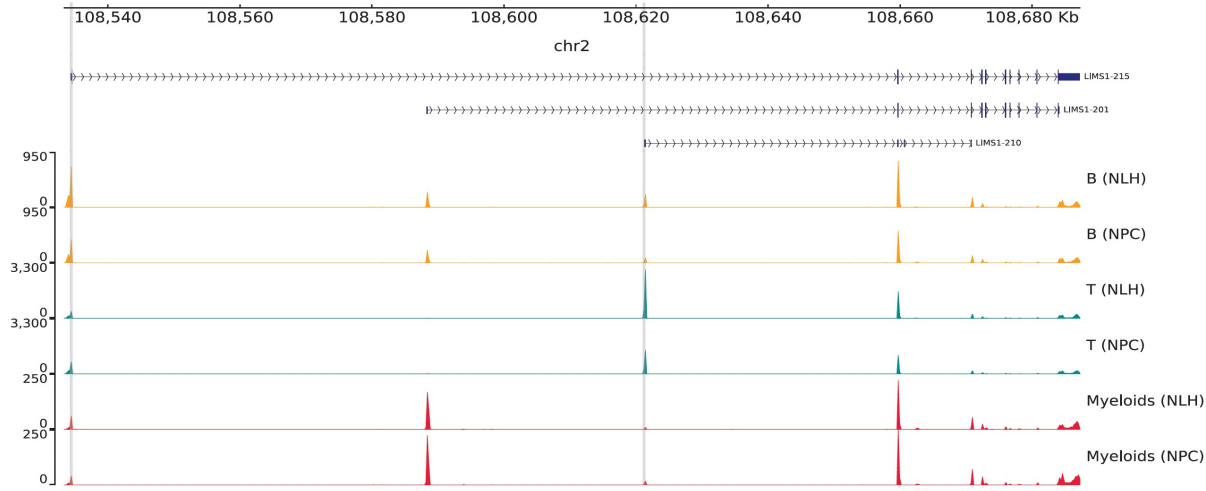
香港大學

THE UNIVERSITY OF HONG KONG



Centre for Translational
Stem Cell Biology
幹細胞轉化研究中心

5' scRNA-seq reveals regulatory patterns via TSS switch



Read 1 has cDNA (cap): precise TSS

Cell type specific: Major TSS & cancer effects

NPC: Nasopharyngeal Cancer

NLH: Nasopharyngeal Lymphatic Hyperplasia

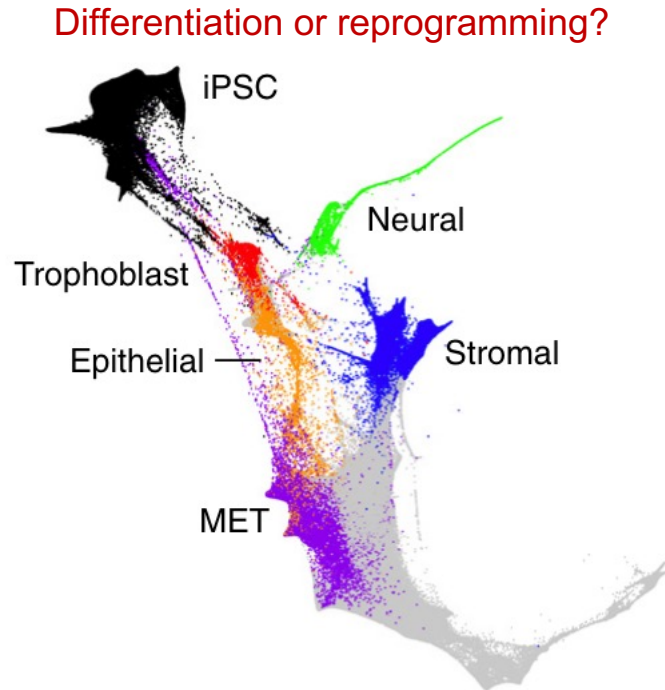
CamoTSS: *Cap* and *motif* based TSS modelling
R Hou, CC Hon & Y Huang, *BioRxiv* 2023



Outline – Cell differentiation and lineage

1. RNA velocity as a framework for directed trajectory inference
2. Lineage specific differentiation bias
3. Lineage markers with somatic mutations
 - Copy number variations
 - Mitochondrial SNV

How to identify the direction of cell transition processes?

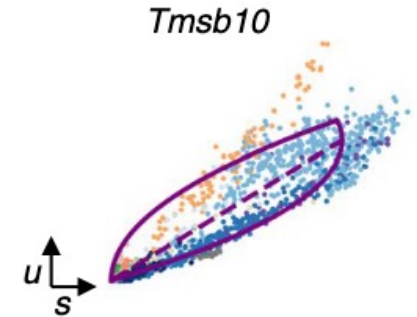
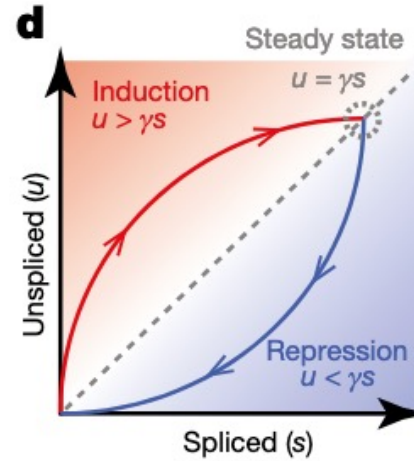
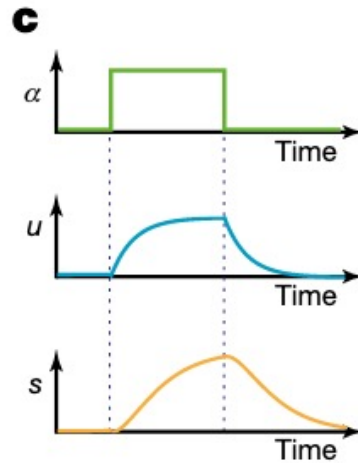
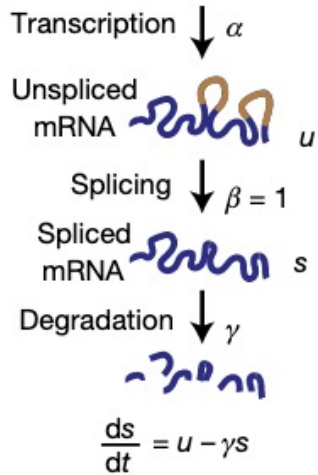


How do we know the direction?

- By experiment design & root cell
- By domain knowledge
- Stem cell have more expressed genes (or diverse expression profile)

Can we infer the direction from data?

Footprint on intrinsic RNA processing: RNA velocity



RNA velocity for cell transition probability matrix

Probability* of cell i transits to cell j

$$\pi_{ij} = \cos \angle(\delta_{ij}, \mathbf{v}_i) = \frac{\delta_{ij}^T \mathbf{v}_i}{\|\delta_{ij}\| \|\mathbf{v}_i\|}$$

$$\delta_{ij} = s_j - s_i$$

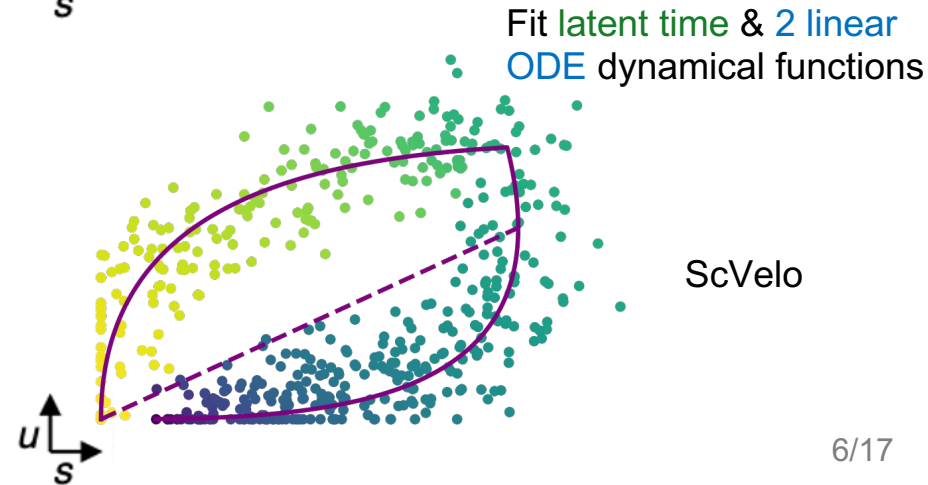
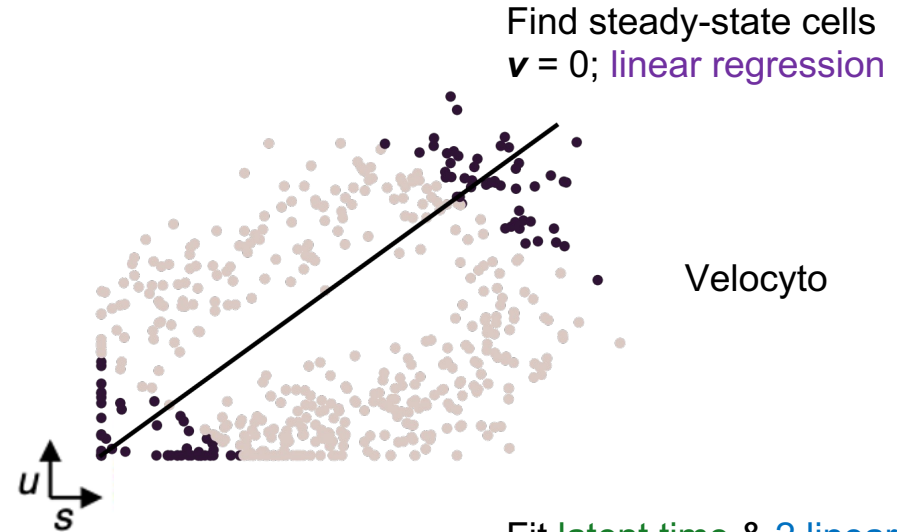
\mathbf{s} : vector of gene expression (spliced)

\mathbf{v} : vector of gene expression velocity

How to predict \mathbf{v} (a big assumption):

$$\mathbf{v} = \frac{ds}{dt} = f(u, s) = \beta * u - \gamma * s$$

*: followed by KNN clipping & softmax



UniTVelo: dynamical function w/ spliced-oriented design

Generalised RNA velocity framework:

$$\hat{s}_{c,g} = f(t_{c,g}; \boldsymbol{\theta}_g),$$
$$\hat{u}_{c,g} = w_1 \cdot f(t_{c,g}) + w_2 \cdot f'(t_{c,g})$$

Loss:

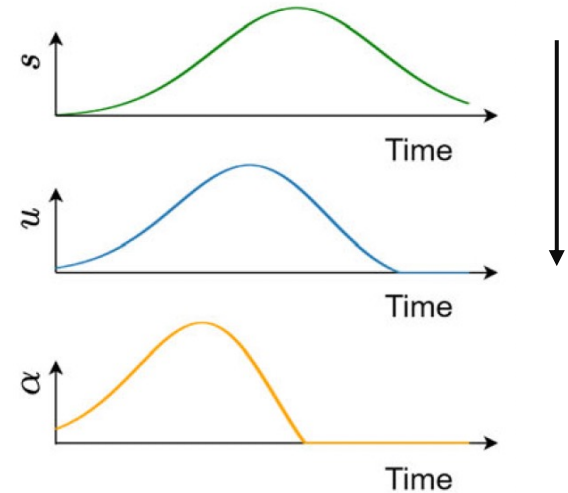
$$\text{Loss} = \text{MSE}(\mathbf{S}, \hat{\mathbf{S}}) + \text{MSE}(\mathbf{U}, \hat{\mathbf{U}})$$

Link function (differentiable; RBF as example):

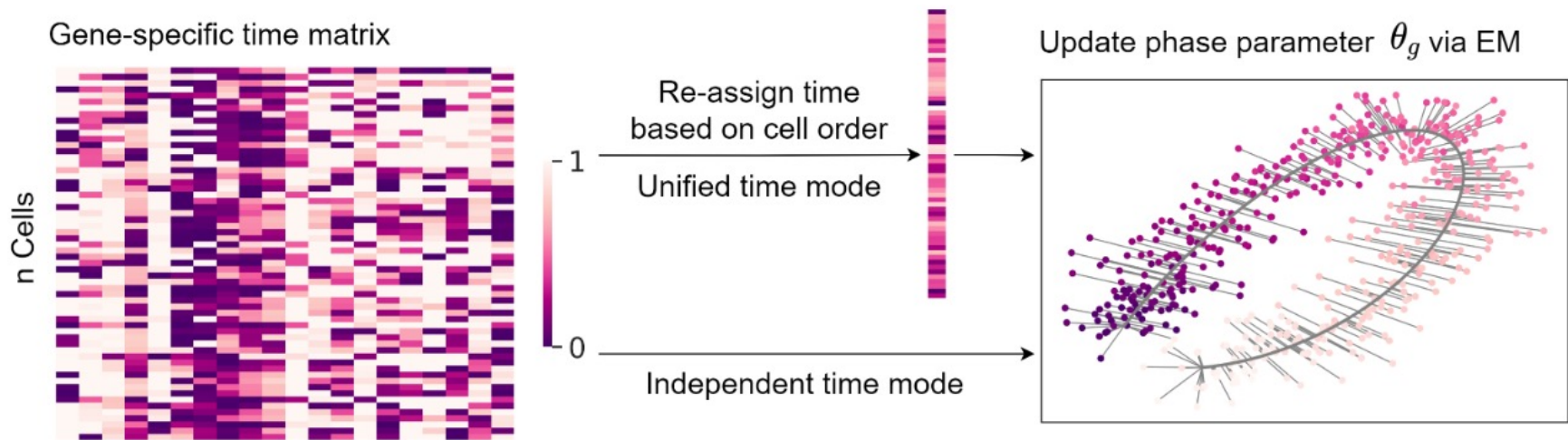
$$f(t; \boldsymbol{\theta}) := h \cdot \exp(-(t - \tau)^2 / l); \boldsymbol{\theta} = \{h, l, \tau\}$$

Latent variables (time or rank; fully constrained as example):

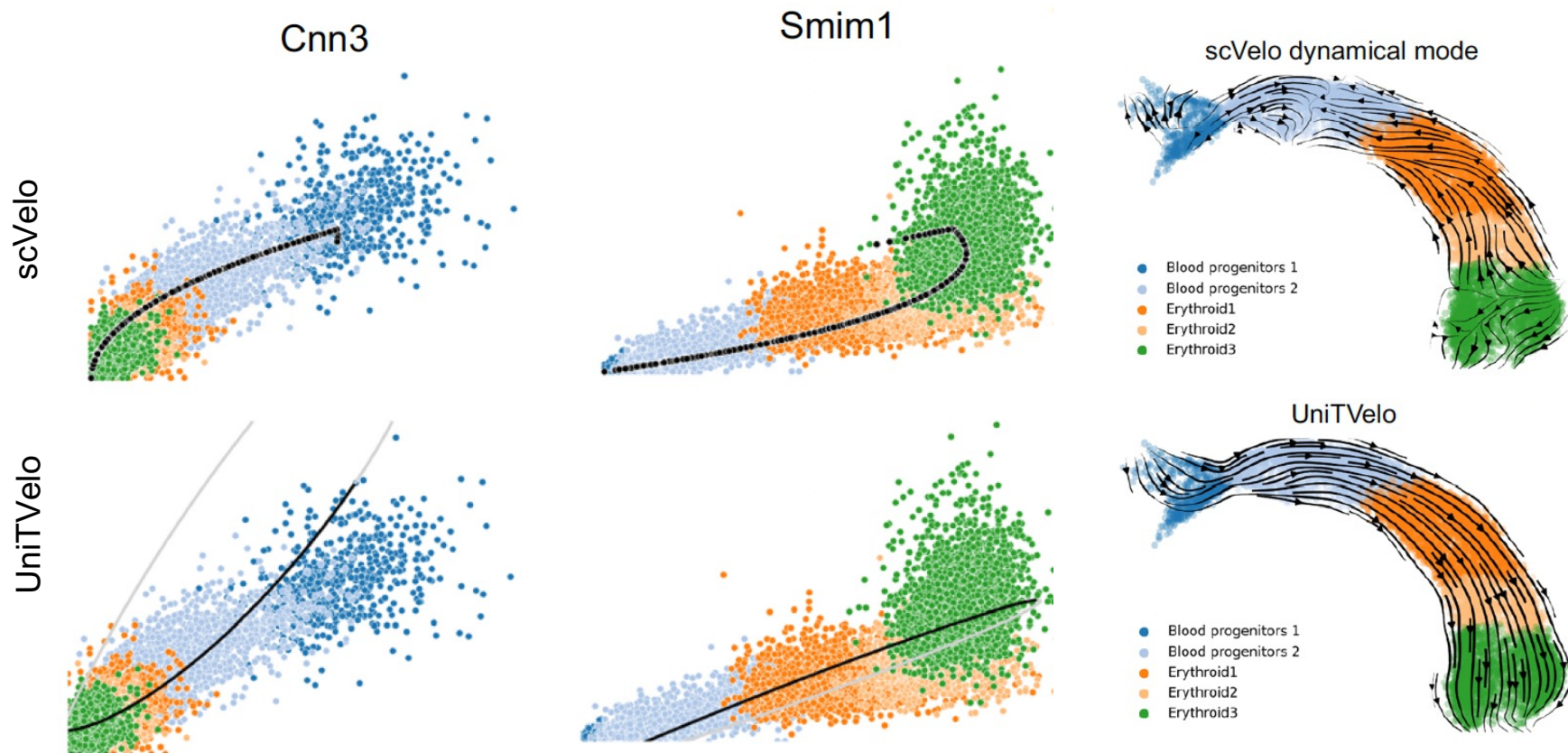
$$t_{c,g} := t_c, \text{ for any } g$$



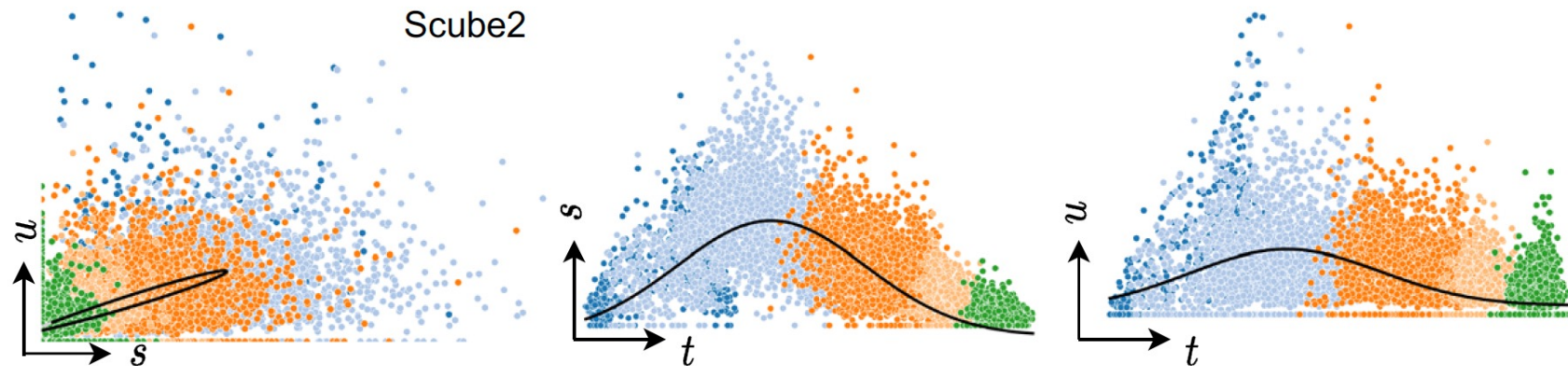
UniTVelo: unified time-based RNA velocity



UniTVelo for analyzing mouse Erythroid



UniTVelo for analyzing mouse Erythroid (example)



Scube2 Expression

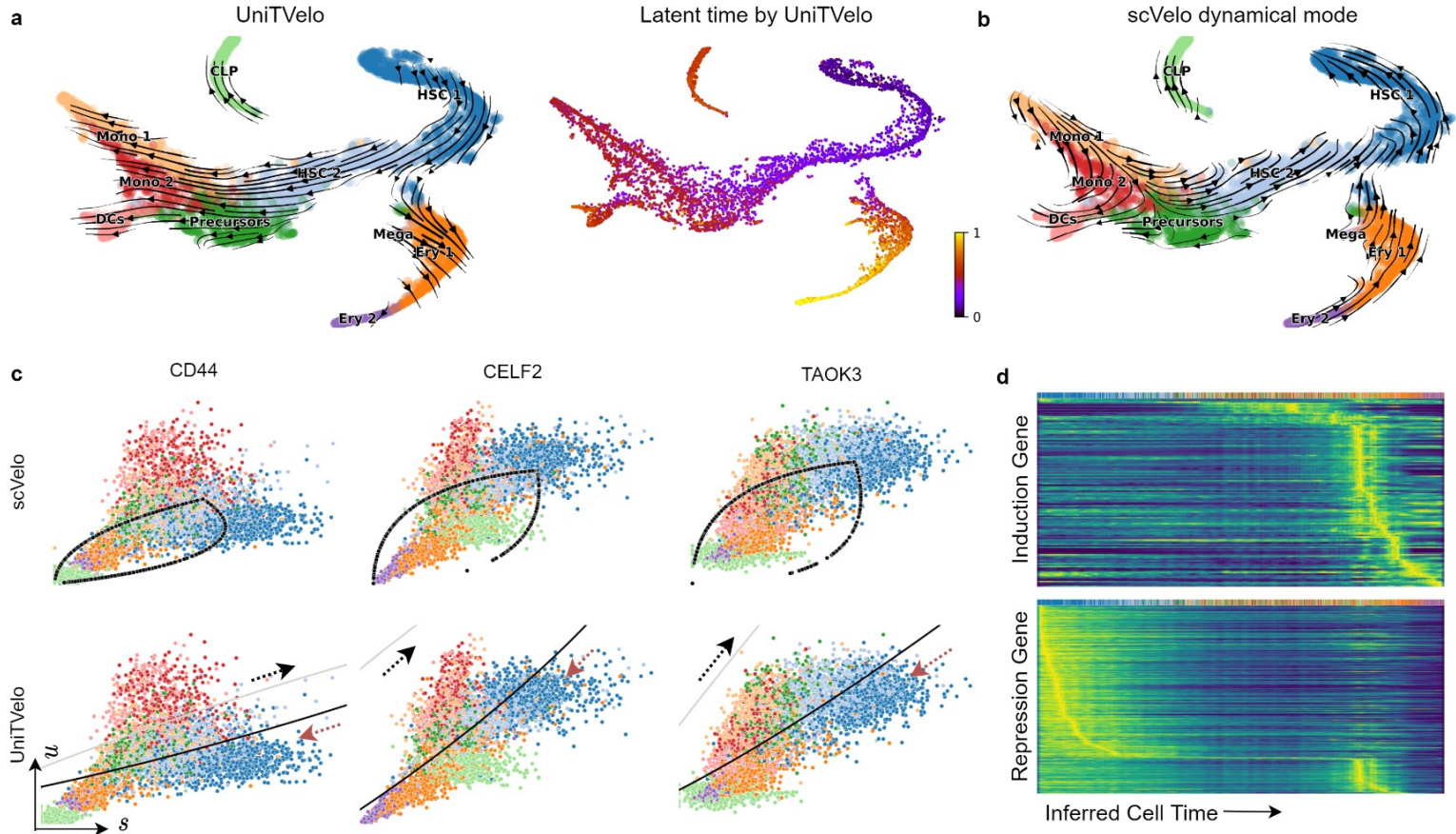


Scube2 Velocity



Transient expression:
Only detected by using
unified time

UniTVelo for analyzing human hematopoiesis



Open challenges

Velocity specific challenges

- **Select features** to view the temporal structure
- **Uncertainty** of the transition probability: parameters, variable selection
- Surrogate predictors of RNA derivative

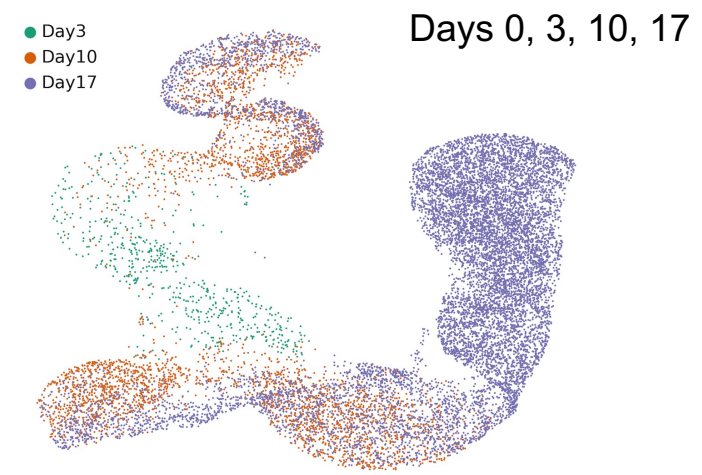
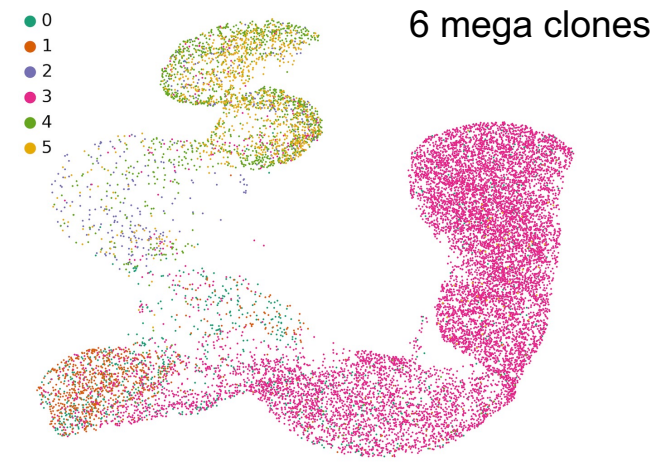
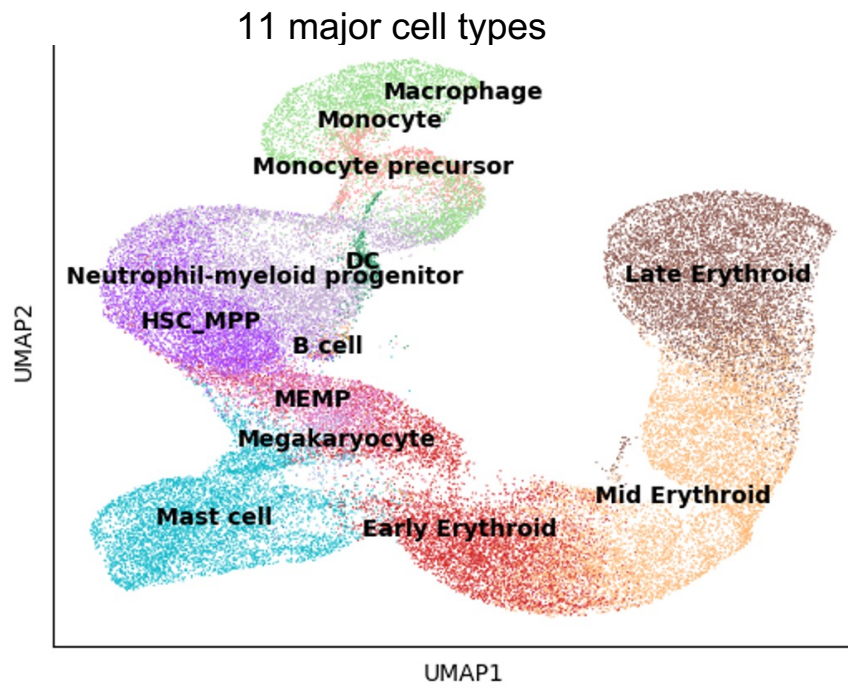
General challenges (as pattern discovery models)

- Balance between model **flexibility** and **overfitting**
- Local optima: **lowest loss** VS **heuristic diagnosis**

Biotechnology challenges

- Enhance the intronic reads (e.g., metabolic labelling)
- Primers targeting a set of genes & introns for universal velocity space

Modelling clone specific differentiation (Hematopoiesis)

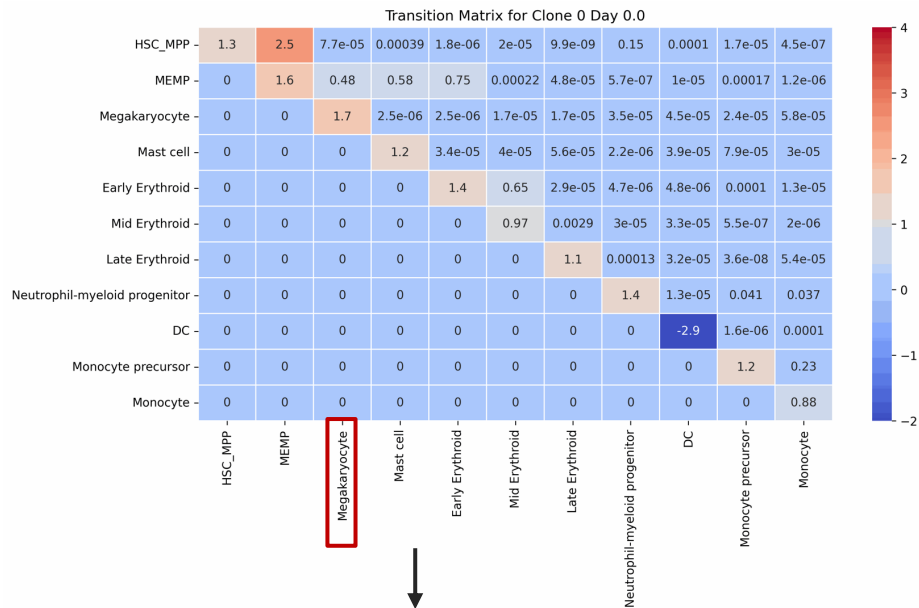


LARRY-barcoding system



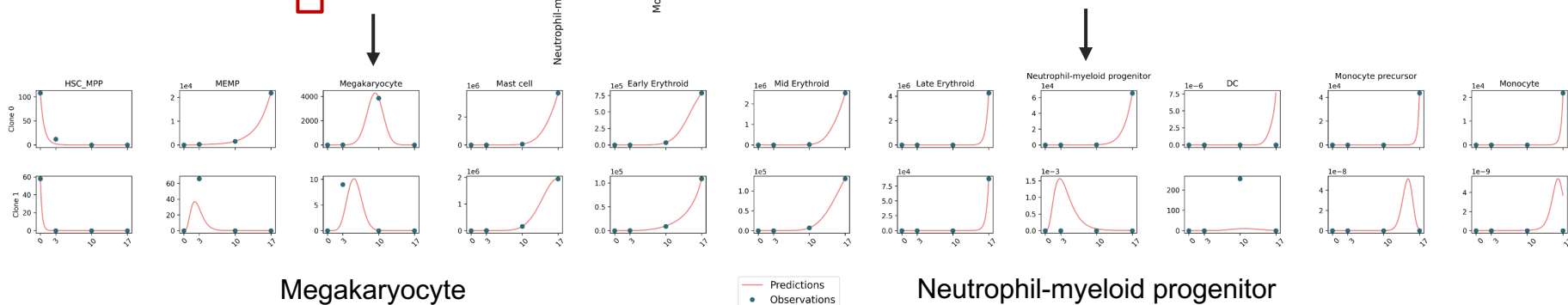
From Bertie Göttgens Lab:
Shirom Chabra
Melania Barile

Clone & time specific transition rate estimation

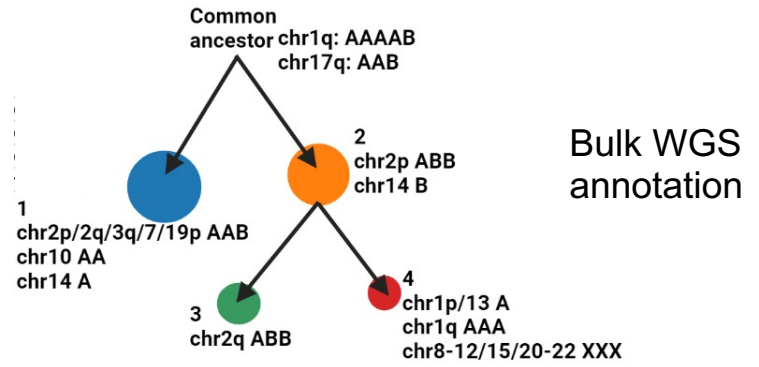
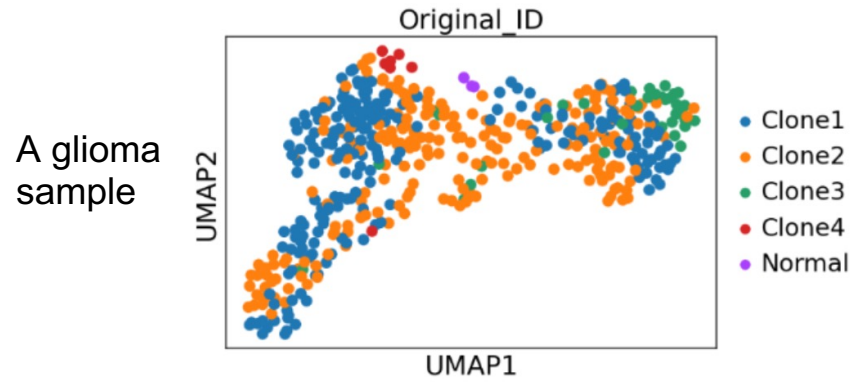


Cell counts by Neural ODE (per clone)

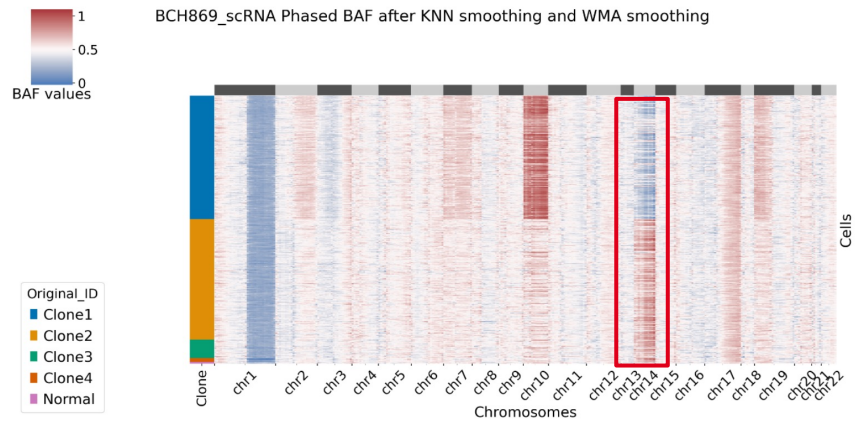
$$\begin{aligned}
 \mathbf{y}'(t) &= f(\mathbf{y}(t)) \\
 &= \mathbf{y}(t) \cdot \mathbf{K}_1 + \mathbf{y}(t) \cdot \text{MLP}(\mathbf{y}(t)) \\
 \hat{\mathbf{y}}(t_i) &= \text{ODEsolver}(\mathbf{y}(t_0), t_0, t_i, \boldsymbol{\theta})
 \end{aligned}$$



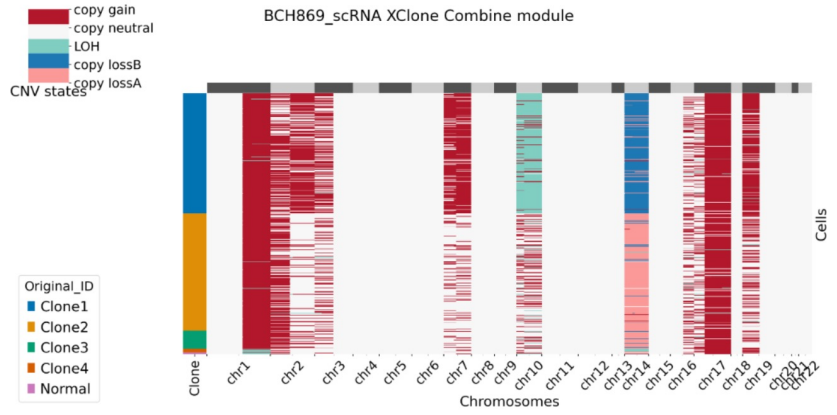
XClone: detect allelic copy number variation from scRNA



XClone Signal (BAF)



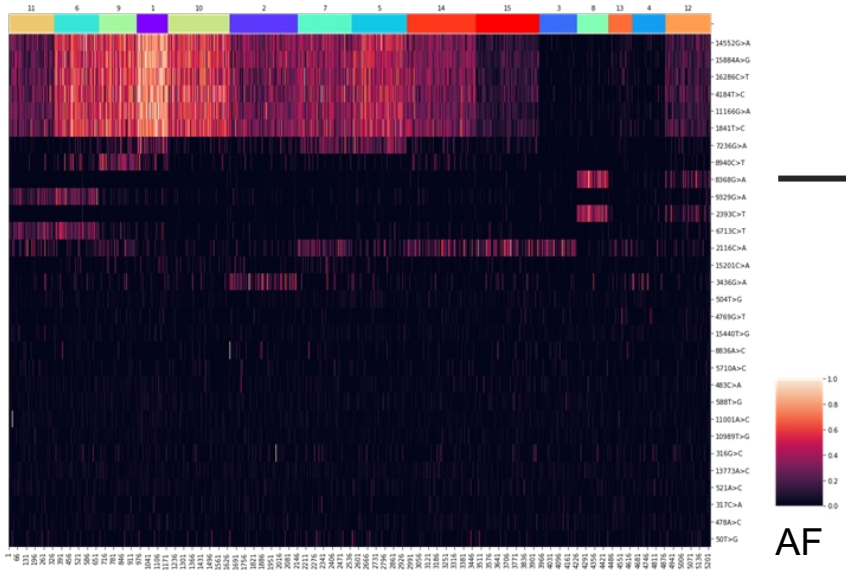
XClone Detection



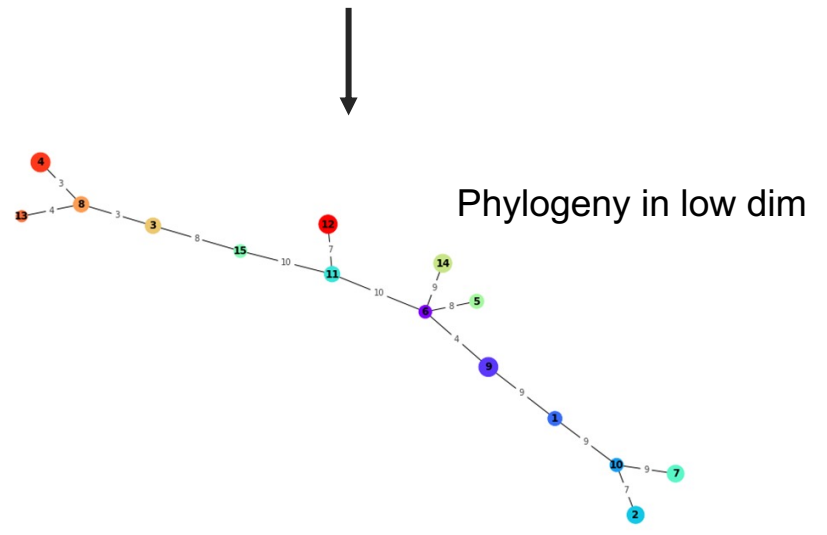
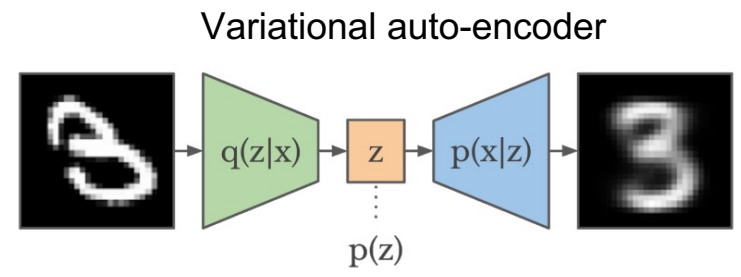
Three-step phasing & Two HMM models

XClone: R Huang, X Huang, et al. *BioRxiv* 2023

Representation learning for phylogenetic inference



Mitochondrial variants & clones (MKN45, gastric cancer)



Acknowledgements

Lab members

Single-cell transcriptomics

Mingze Gao

Ruiyan Hou

Fangxin Cai

Xinyi Lin (w/ Joshua)

Single-cell genetics

Rongting Huang

Xianjie Huang

Yiming Chao (w/ Rio)

Chen Cheng

Kevin Chung

Spatial transcriptomics

Zhuoxuan Li (w/ Pentao)

Chrissy Ma (w/ Joshua)

Biomedical AI

Weizhong Zheng (w/ Joshua)

Yu Lu

Yijun Liu

Zhihan Xu



Collaborators

Prof. Pentao Liu

Dr. Joshua Ho

Dr. Rio Sugimura

Prof. SY Leung

Prof. Bertie Göttgens

Dr. Melania Barile

Shirom Chabra

