bioRxiv preprint first posted online Oct. 19, 2017; doi: http://dx.doi.org/10.1101/206052. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. All rights reserved. No reuse allowed without permission.

RNA velocity in single cells

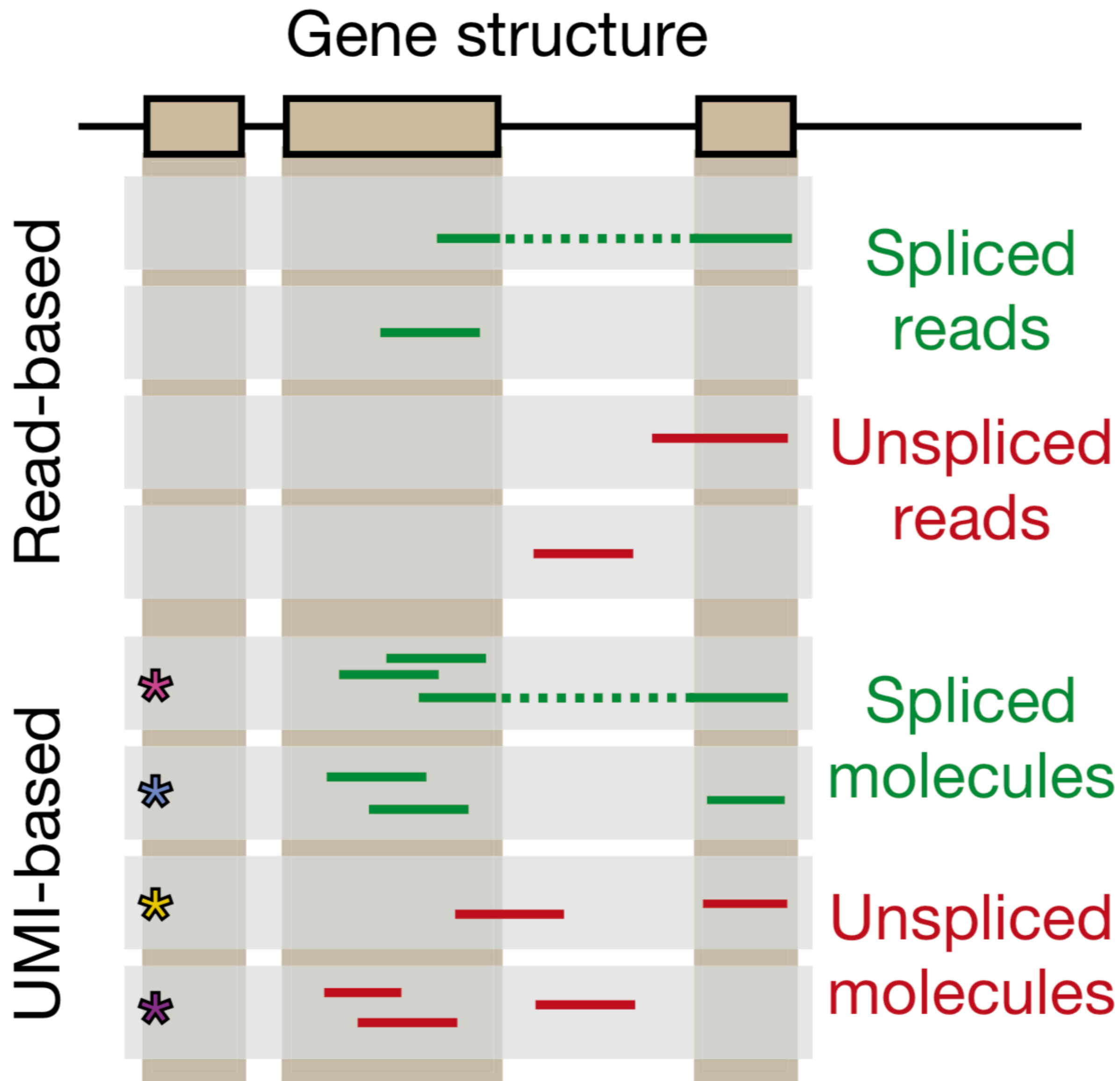
Authors: Gioele La Manno[1,2], Ruslan Soldatov[3], Hannah Hochgerner[1,2], Amit Zeisel[1,2], Viktor Petukhov[3,4], Maria E. Kastriti[5], Peter Lönnerberg[1,2], Alessandro Furlan[1], Jean Fan[3], Zehua Liu[3], David van Bruggen[1,2], Jimin Guo[3], Erik Sundström[6], Gonçalo Castelo-Branco[1,2], Igor Adameyko[5], Sten Linnarsson[1,2]† and Peter V. Kharchenko[3,7]†

LETTER

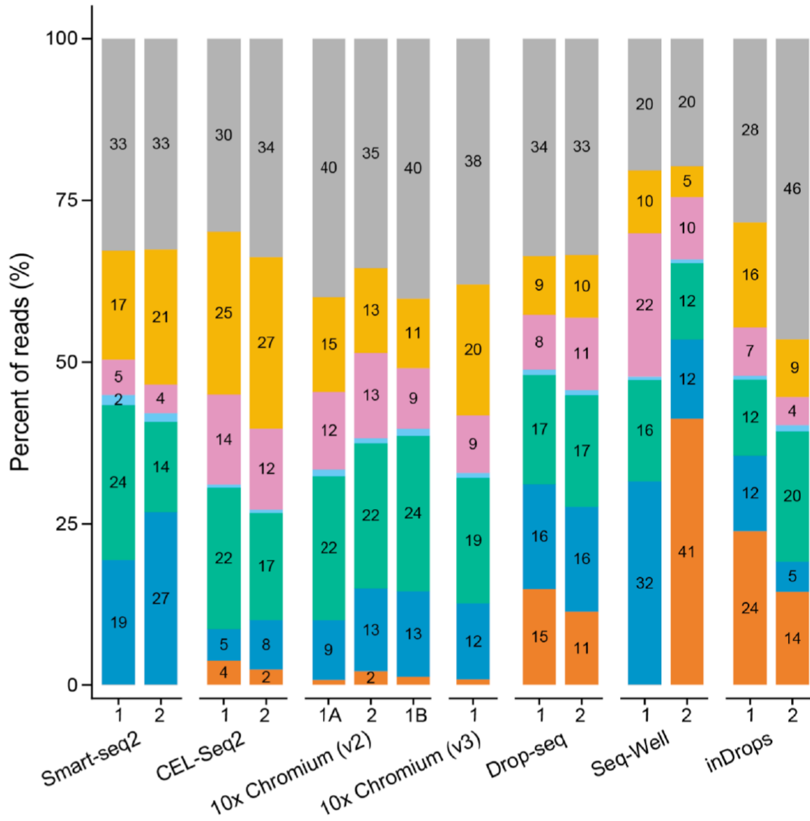
https://doi.org/10.1038/s41586-018-0414-6

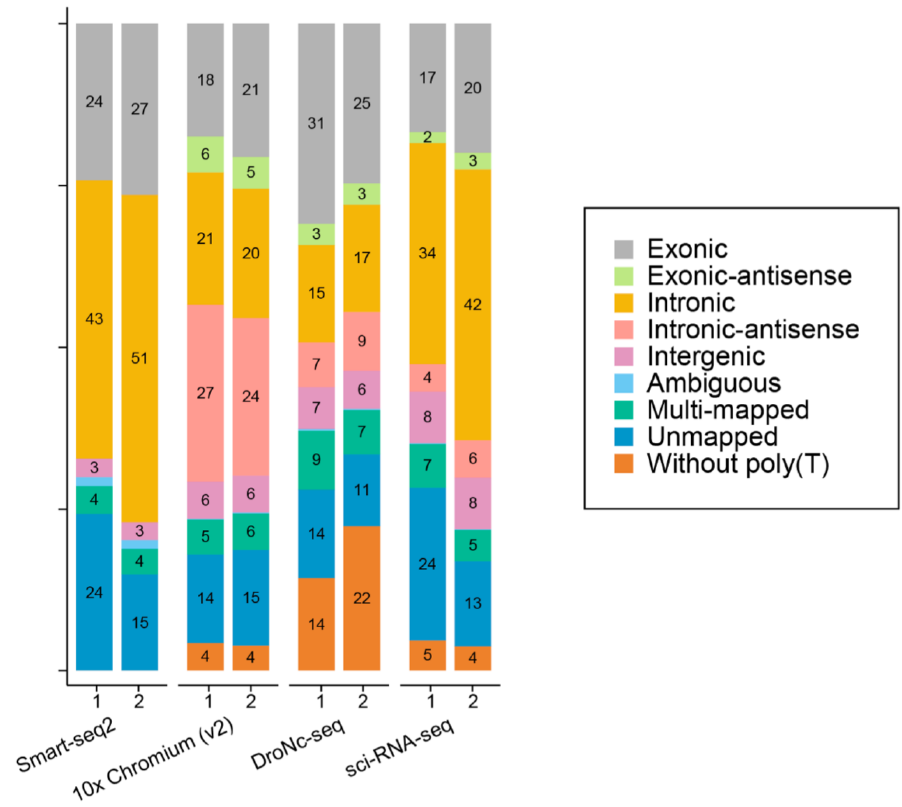
RNA velocity of single cells

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]*



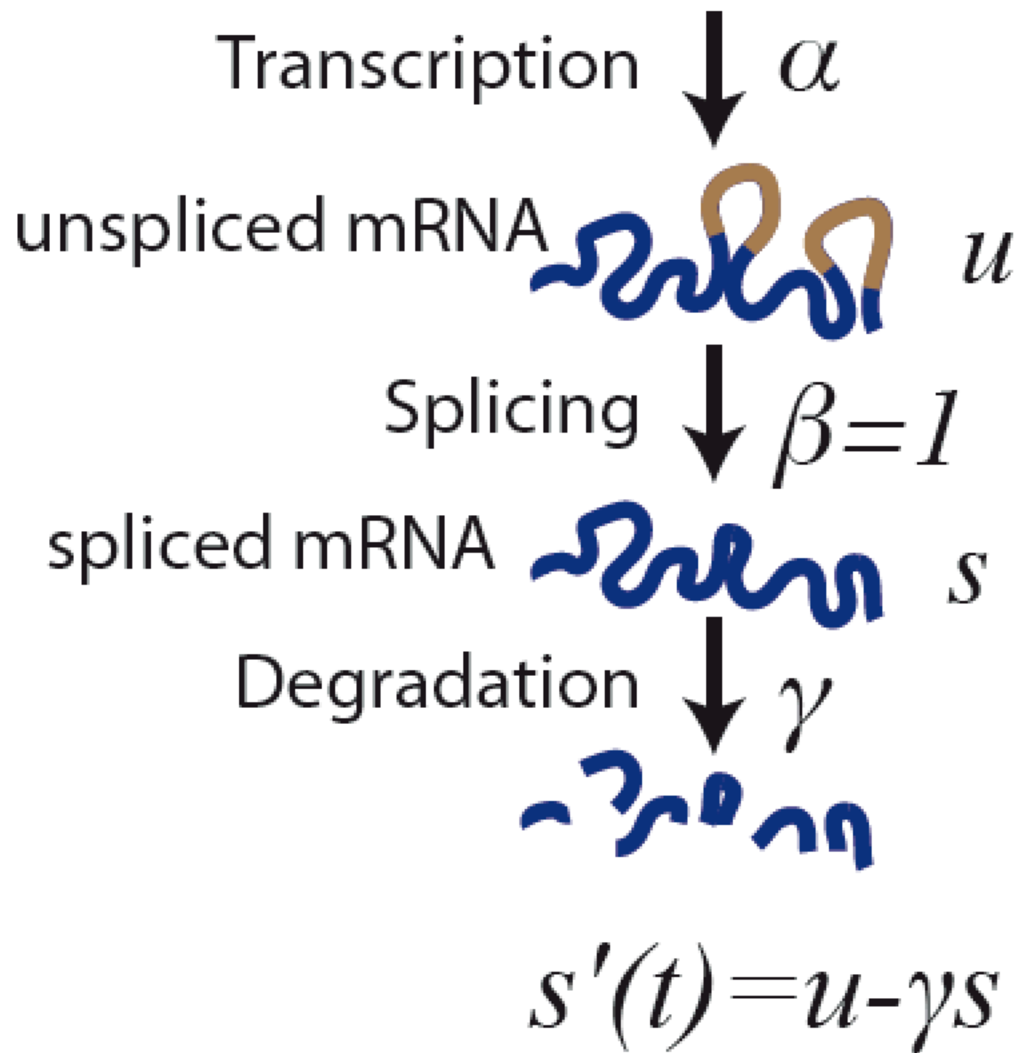
Statement 1: on peut trouver des pré-mRNAs dans les libraries RNAseq



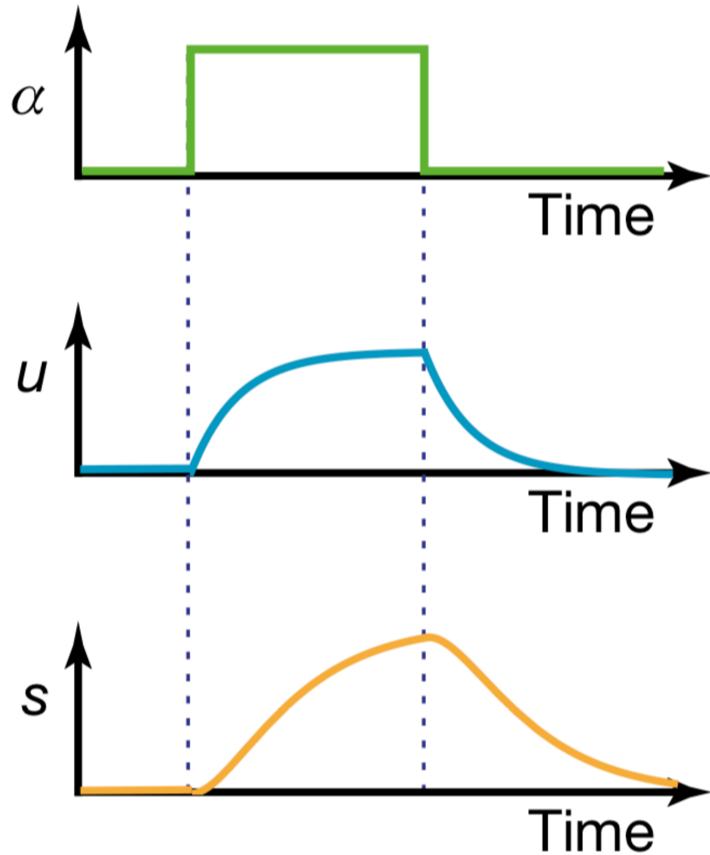


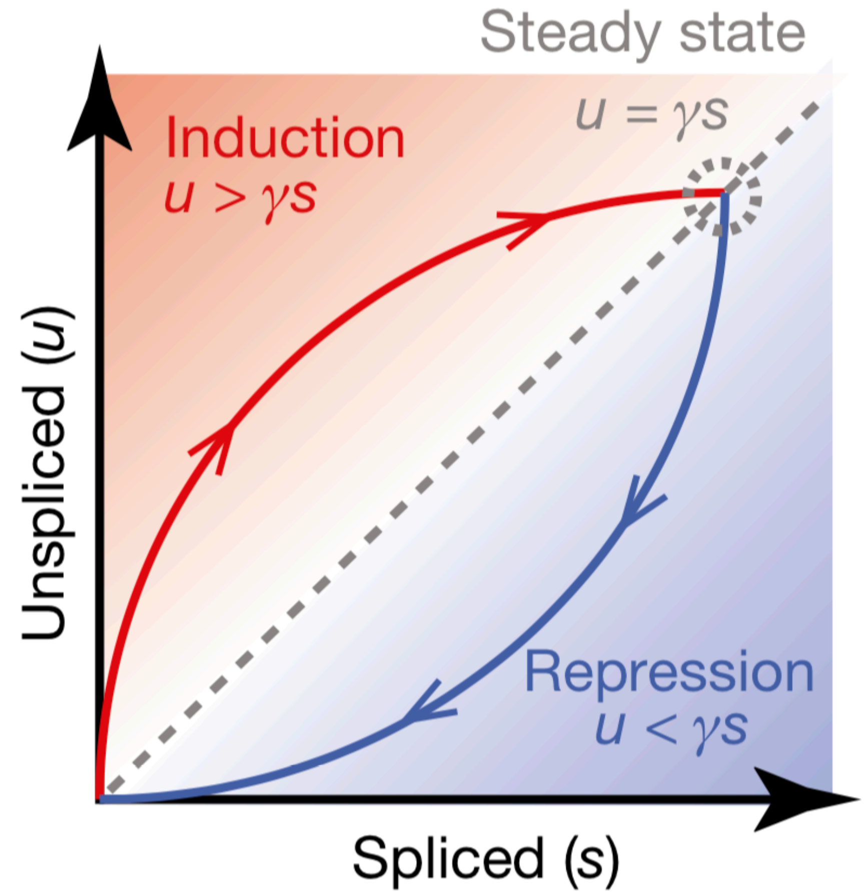
doi: http://dx.doi.org/10.1101/632216.

Statement 1: on peut trouver des pré-mRNAs dans les libraries RNAseq



Statement 2: on peut écrire un modèle dynamique simple





Si on fait varier α (le taux de transcription), on voit le déplacement dans l'espace s/u

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

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

Here, $\alpha(t)$ is the time-dependent rate of transcription, $\beta(t)$ is the rate of splicing, $\gamma(t)$ is the rate of degradation. Under an assumption of constant (time-independent) rates $\alpha(t) = \alpha$, $\gamma(t) = \gamma$, and setting $\beta(t) = 1$ (i.e. measuring all rates in units of the splicing rate), the rate equations simplify to:

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

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

The complete solution to the rate equations is given by:

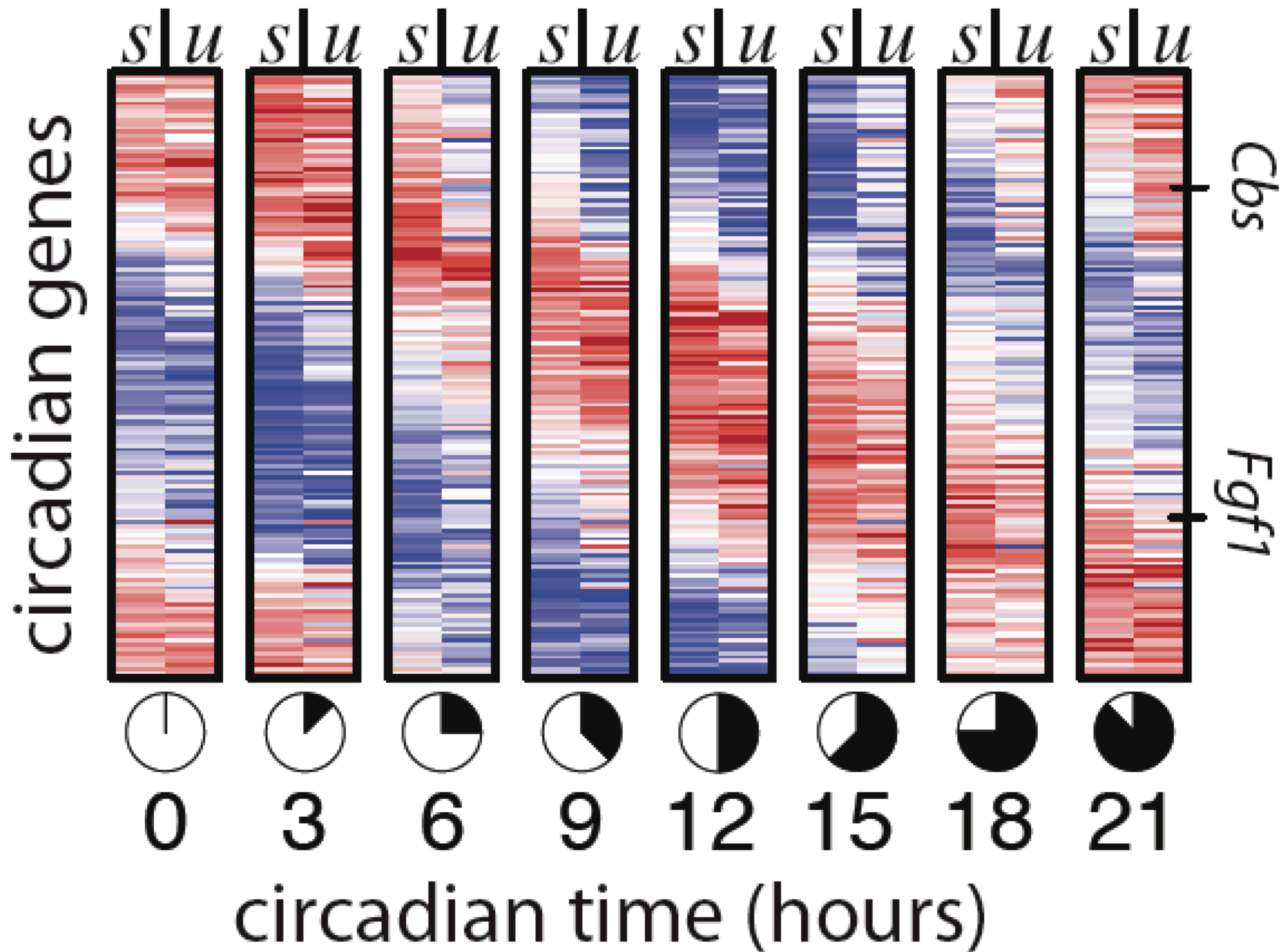
$$u(t) = \alpha(1 - e^{-t}) + u_0 e^{-t}$$

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

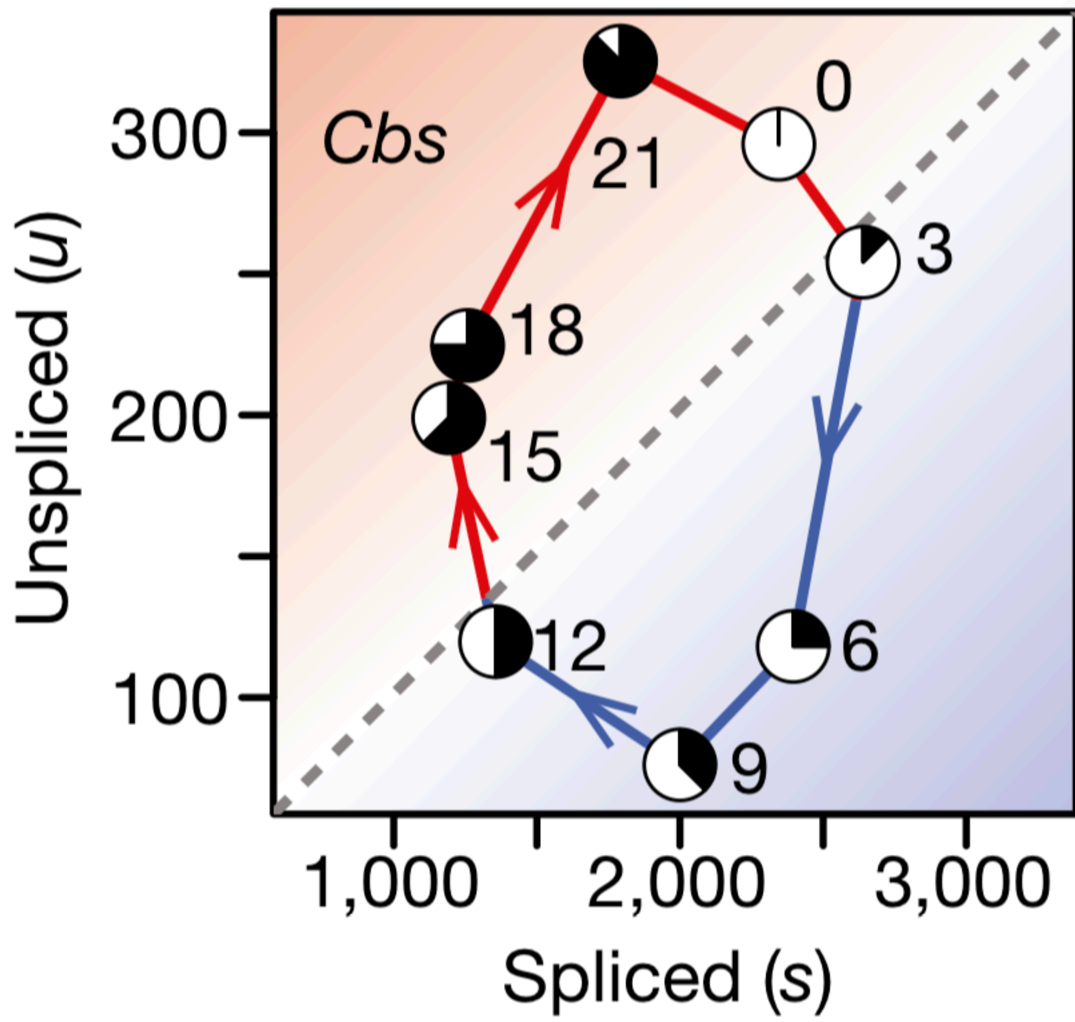
with the initial conditions $u(0) = u_0$ and $s(0) = s_0$. This solution can be used to extrapolate mRNA abundance s to a future timepoint t_1, under the assumption stated above, by entering the current state of the cell as u_0 and s_0, and then computing $s(t_1)$.

On peut donc prédire l’état futur du système au temps t (s(t)) connaissant son état actuel (s0 et u0)

Preuve du concept: cycle cellulaire



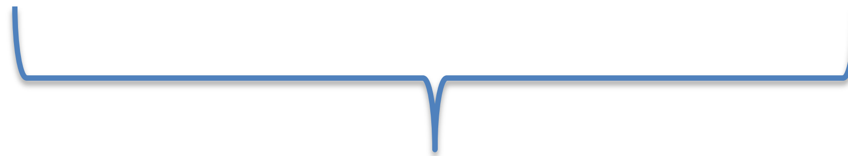
On voit en effet que les formes pré-mRNA (u) précèdent les formes épissées (s).



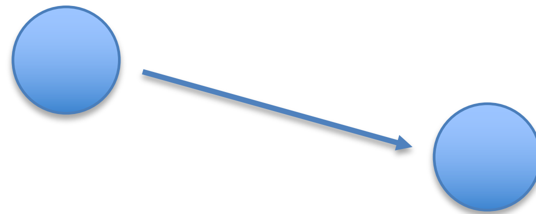
On peut verifier l'évolution d'un gène

Puis prédire l'état futur du système pour une combinaison de tous les gènes, et le projeter en 2D

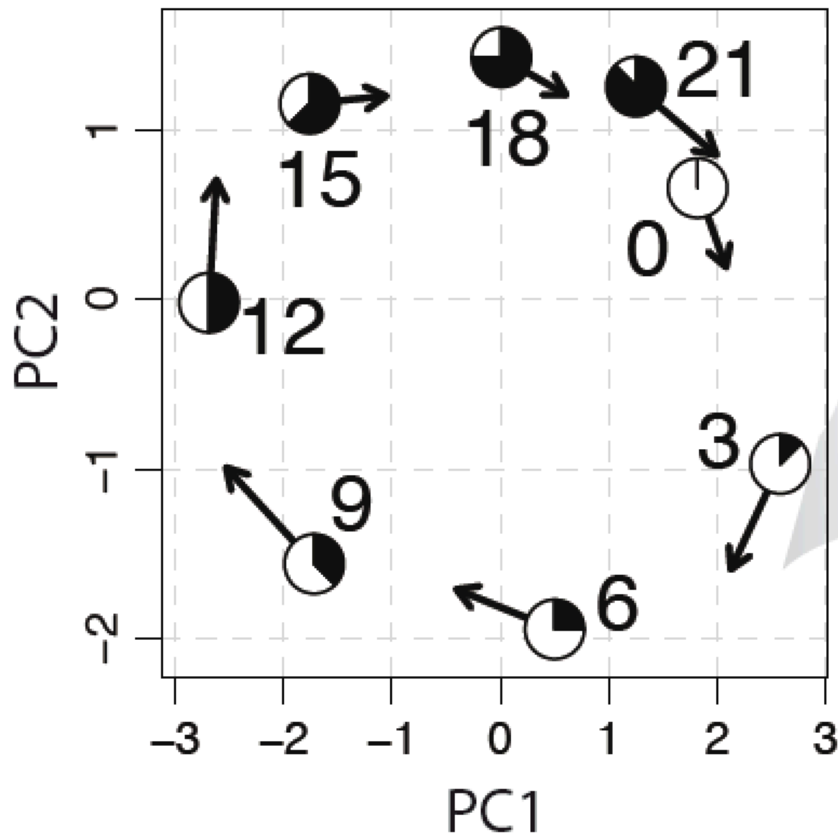
	Cellule à l'instant t	Cellule à l'instant t+1
G1	U_0, S_0	S_1
G2	U_0, S_0	S_1
.		
.		
.		
Gn	U_0, S_0	S_1



Projection dans un espace 2D (PCA, UMAP, tSNE)



(On ne représente pas en général la cellule d'arrivée)



observed state:
s_0

extrapolated state:
$s_1 = s_0 e^{-\gamma t} + u_0/\gamma(1-e^{-\gamma t})$

Donne de très jolies figures
(differentiation of Schwann cell precursors (SCPs) into chromaffin cells in E12.5 mouse)

Les limitations de cette approche (the steady-state model) :

1. on the gene level, assumes that the full splicing dynamics with transcriptional induction, repression and steady-state mRNA levels are captured

2. on the cellular level, assumes that all genes share a common splicing rate (β =1).

Proposition: scVelo
(Bergen et al., doi: https://doi.org/10.1101/820936)

It generalizes RNA velocity estimation to transient systems and systems with heterogeneous subpopulation kinetics.

bioRxiv preprint doi: https://doi.org/10.1101/820936. this version posted October 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

October 28, 2019

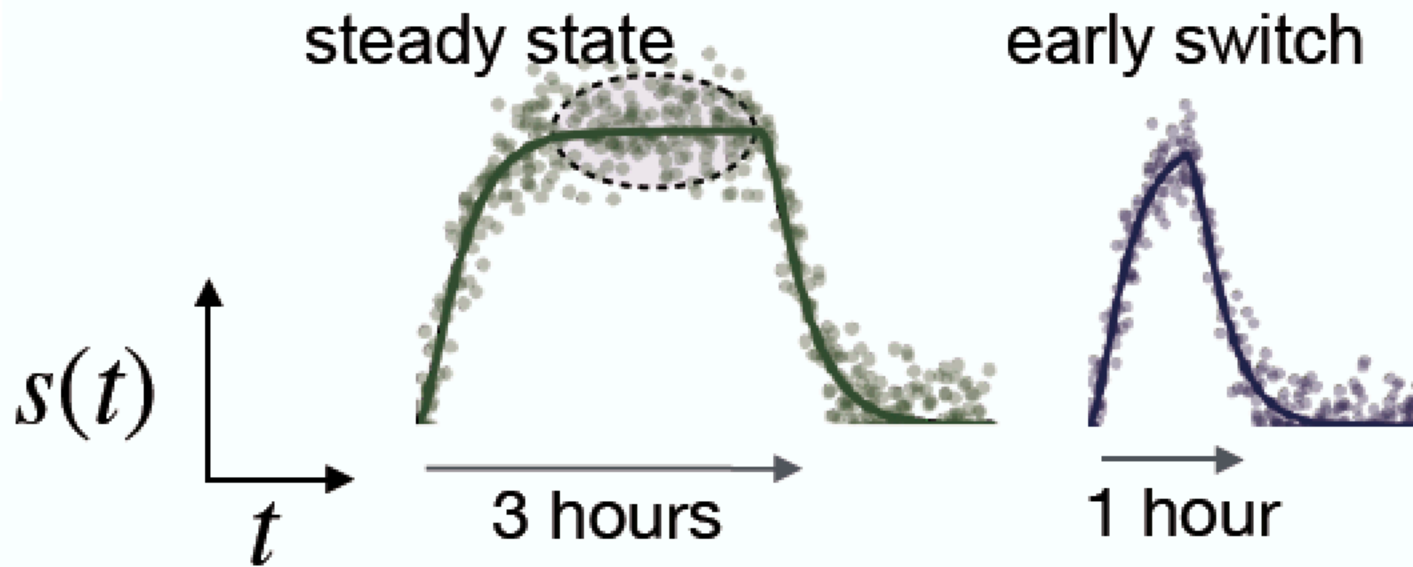
Generalizing RNA velocity to transient cell states through dynamical modeling

Volker Bergen[1,2], Marius Lange[1,2], Stefan Peidli[2], F. Alexander Wolf[1*], Fabian J. Theis[1,2*]

1 Institute of Computational Biology, Helmholtz Center Munich, Germany.
2 Department of Mathematics, TU Munich, Germany.

*Corresponding authors: alex.wolf@helmholtz-muenchen.de, fabian.theis@helmholtz-muenchen.de

the steady-state model   scVelo



Pour chaque cellule, on va estimer les valeurs de deux jeux de paramètres:

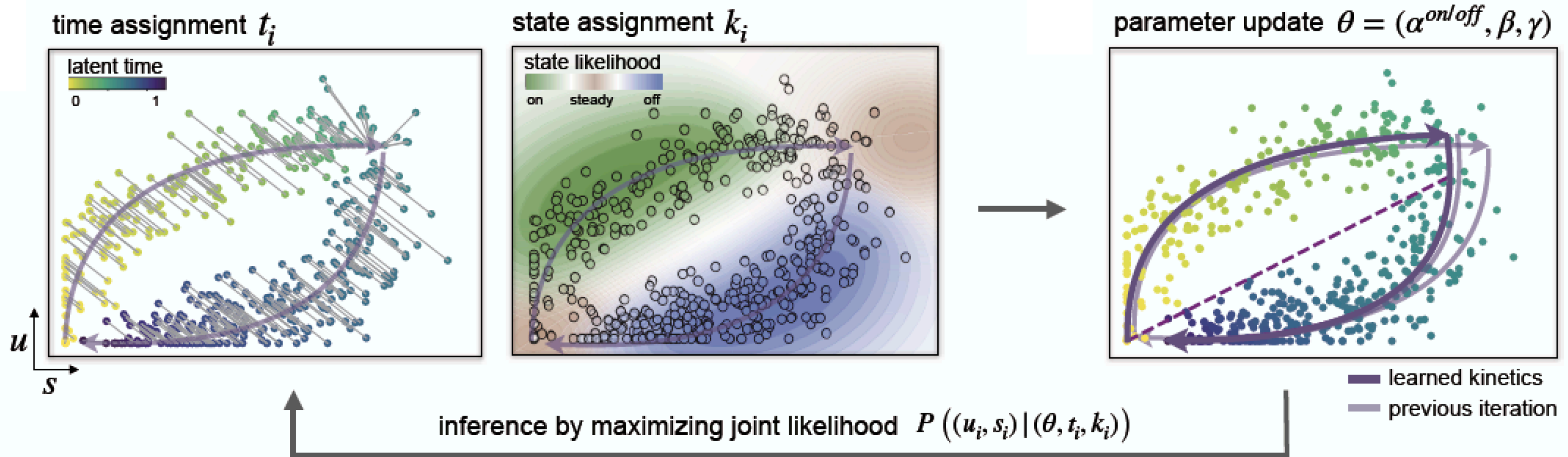
1. Les taux de réaction par gène: (i) transcription: $\alpha_k(t)$; (ii) épissage: β; et (iii) dégradation: γ

2. Deux variables latentes spécifiques de la cellule: k_i et t_i

k_i prend 4 valeurs: induction, repression, active and inactive steady states

t_i est un temps continu: latent time

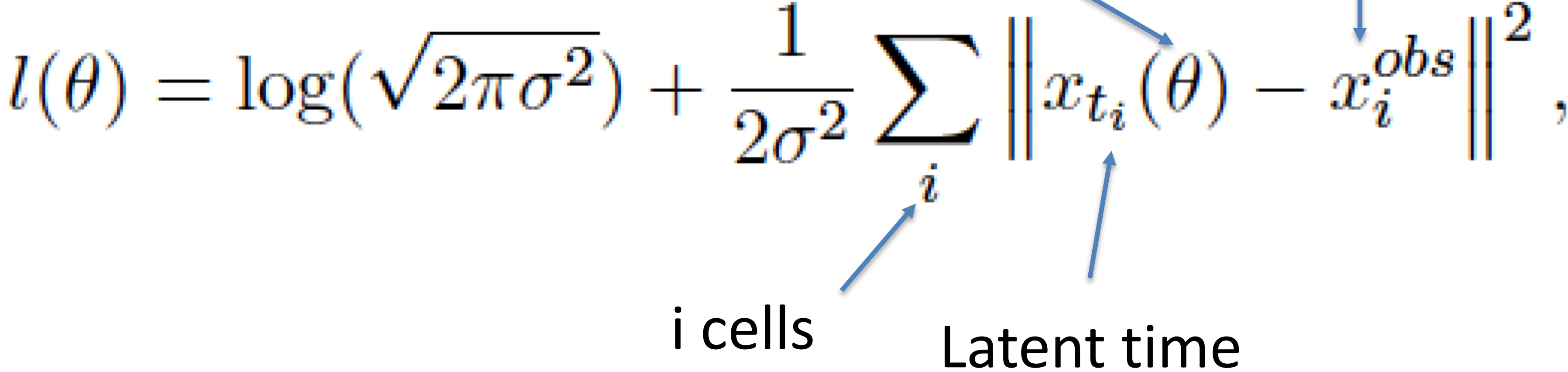
Sachant un jeu, on peut déduire l'autre: L'algo va alterner: optimiser 1, déduire 2, optimiser 2, déduire 1, etc..

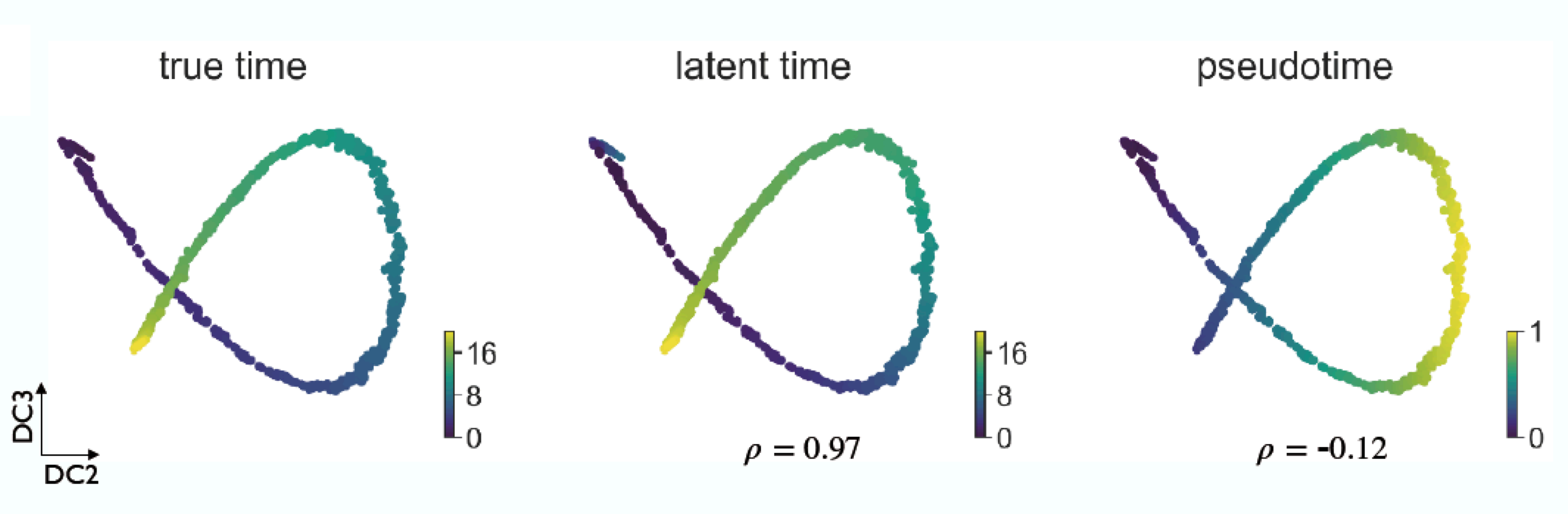


Let the model estimate be $\hat{x}(t) = (\hat{u}(t), \hat{s}(t))$, and let the observations be $x_i^{obs} \sim N(\hat{x}(t), \sigma^2)$. With the assumption of the gene-specific σ to be constant across cells within one transcriptional state, and the observations to be i.i.d., the likelihood-based framework is derived in the following.

The negative log-likelihood to be minimized is given by

$$l(\theta) = \log(\sqrt{2\pi\sigma^2}) + \frac{1}{2\sigma^2} \sum_i \left\| x_{t_i}(\theta) - x_i^{obs} \right\|^2, \tag{7}$$





On simulated splicing kinetics, latent time is able to reconstruct the underlying real time at **near perfect correlation** and correct scale, clearly outperforming pseudotime. In contrast to pseudotime methods, our latent time is **grounded on transcriptional dynamics** and internally accounts for speed and direction of motion. Hence, scVelo's latent time yields faithful gene expression time-courses to delineate dynamical processes, and to **extract gene cascades**.

Bonuses:

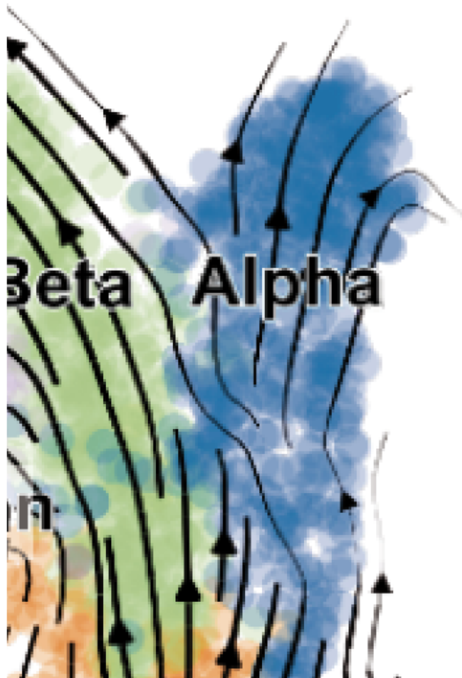
1. We can obtain relevant parameter values that can be directely re-injected into WASABI, like the mRNAs degradation rates

2. We can identify putative driver genes (so called « dynamical genes » as genes characterized by high likelihoods.

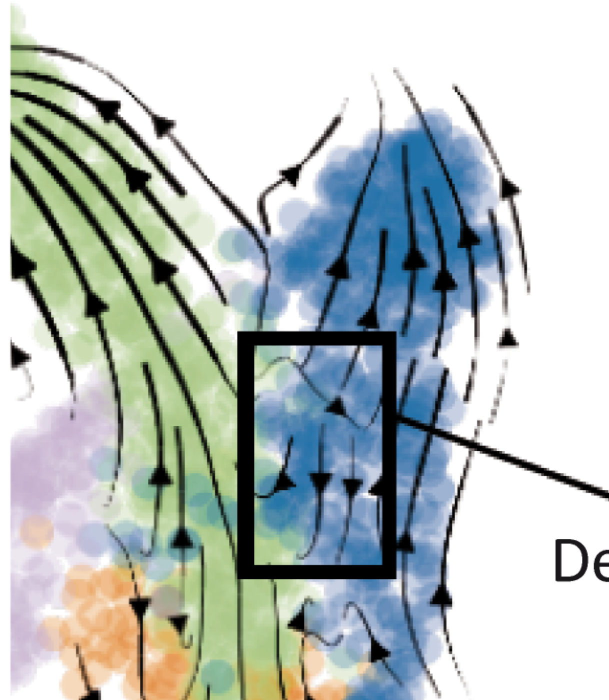
3. We can identify transcriptional switches (when a gene switches its regime, identified by k) -> ranking of genes in WASABI

4. Toward a stochastic dynamical model (left to future work) -> identify bursting frequencies??

Dynamical model

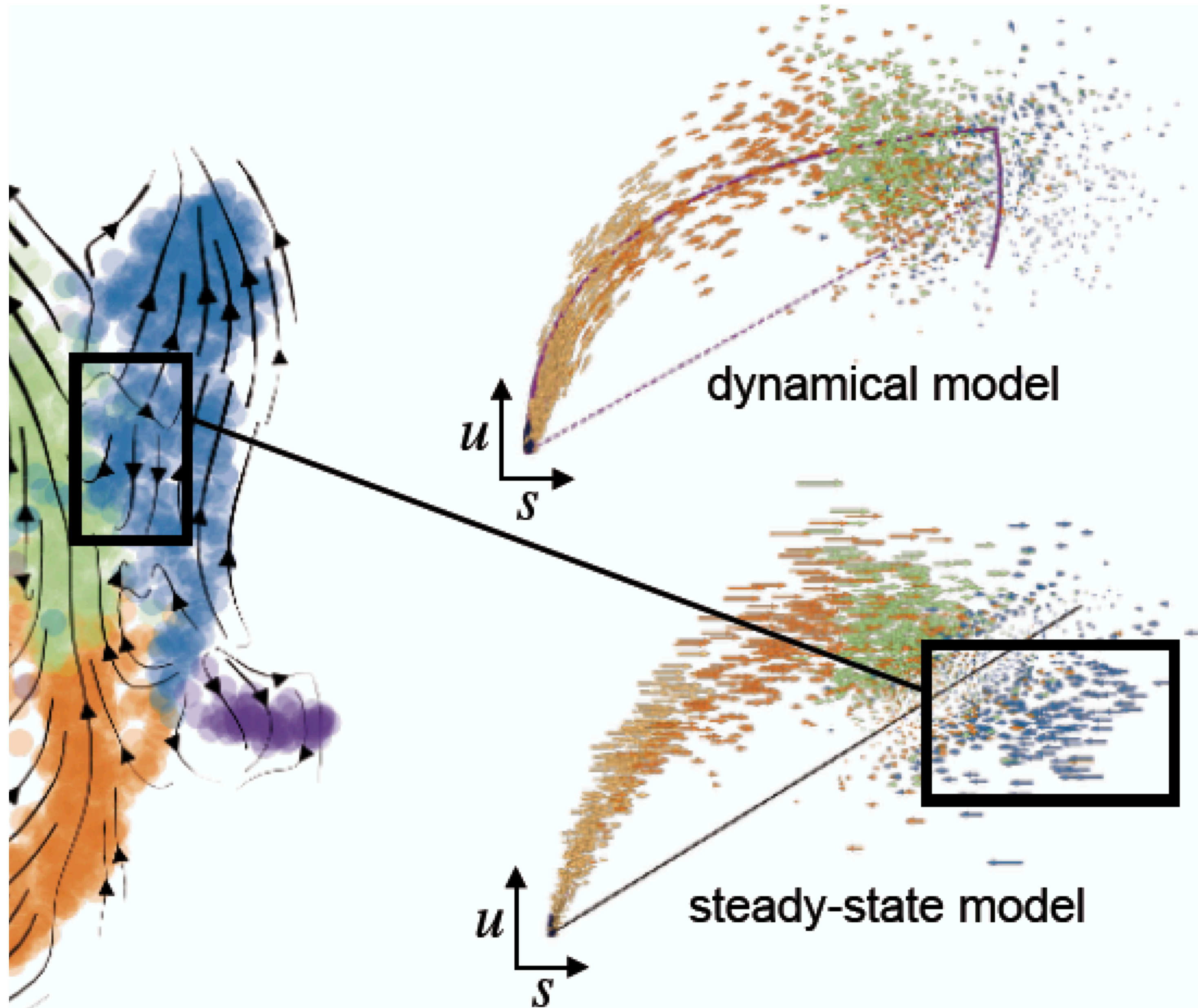


Steady-state model



Dedifferentiating cells?

pancreatic endocrinogenesis



Cpe gene-velocities
dynamical model
u
s
steady-state model
u
s