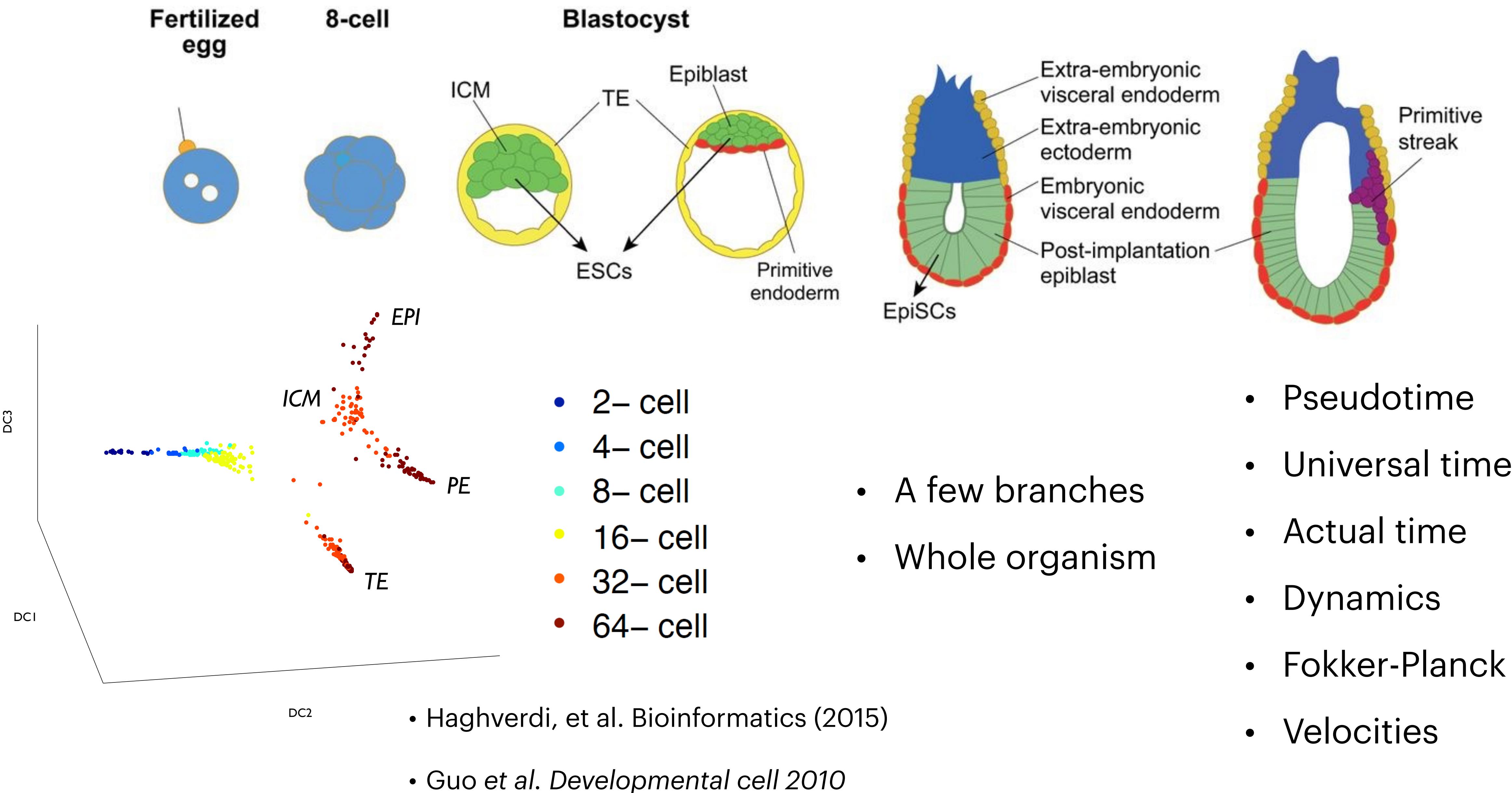


Gene regulatory networks

LALEH HAGHVERDI

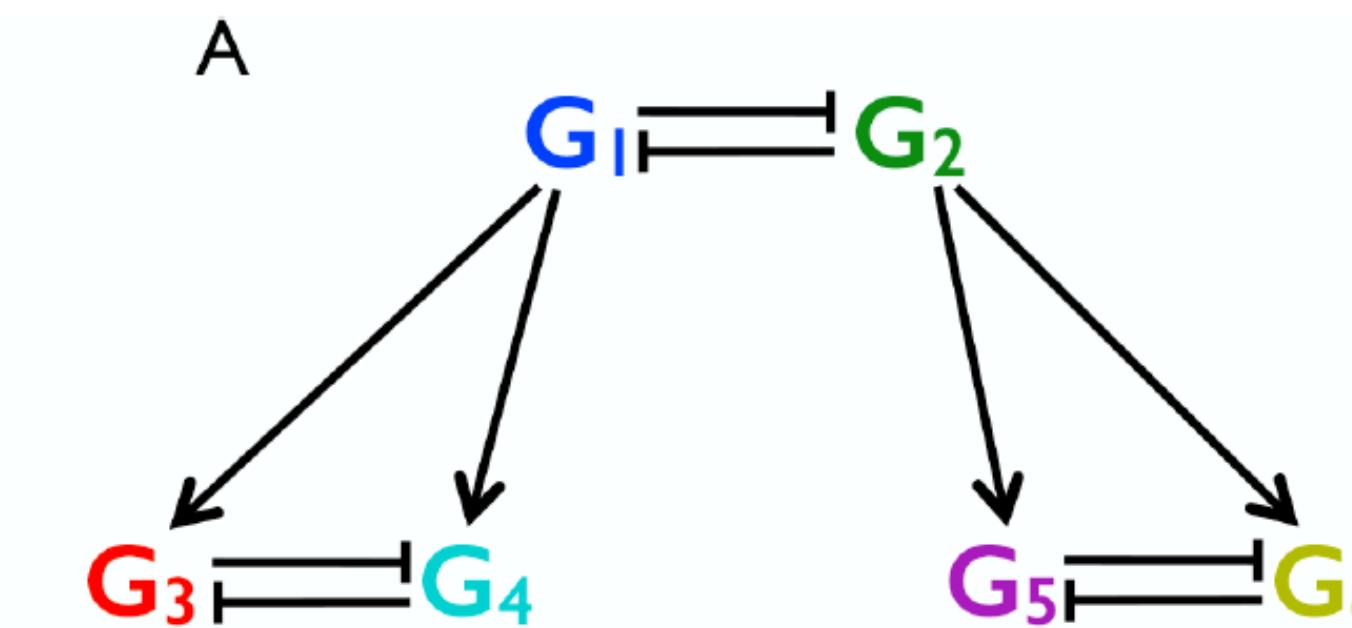
Cell differentiation data manifold and dynamics



Actual time, universal time, pseudotime

"Diffusion pseudotime robustly reconstructs lineage branching"
Haghverdi et al. Nature methods 2016 (Supplemental Note 1)

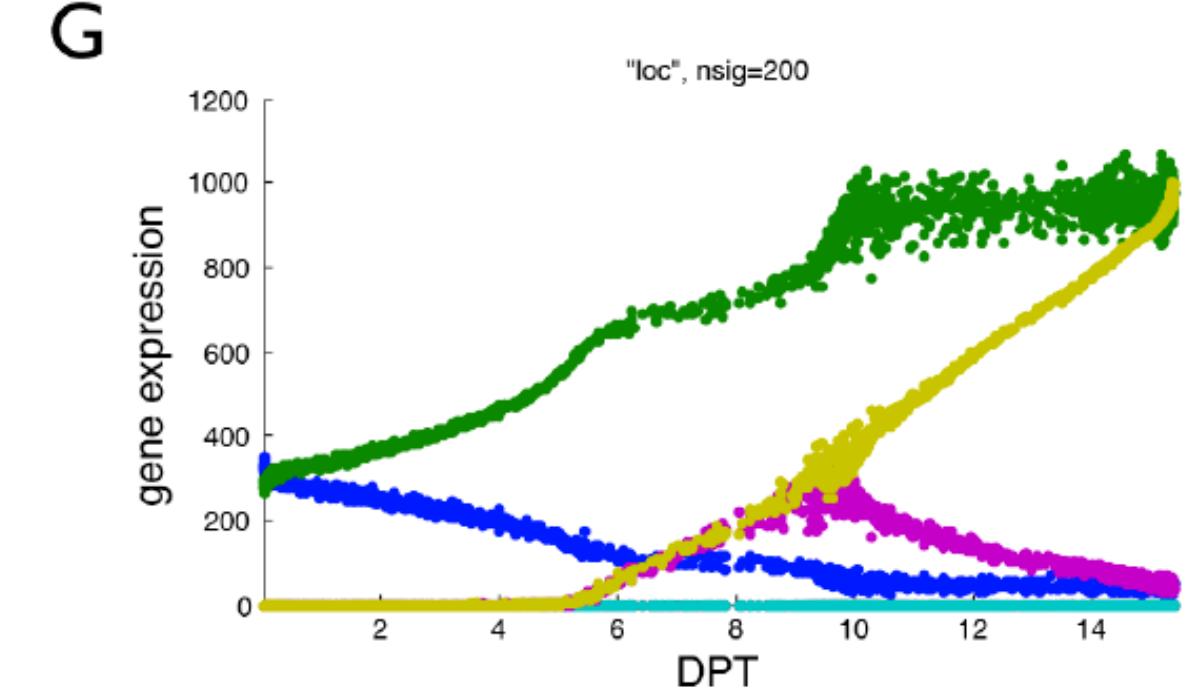
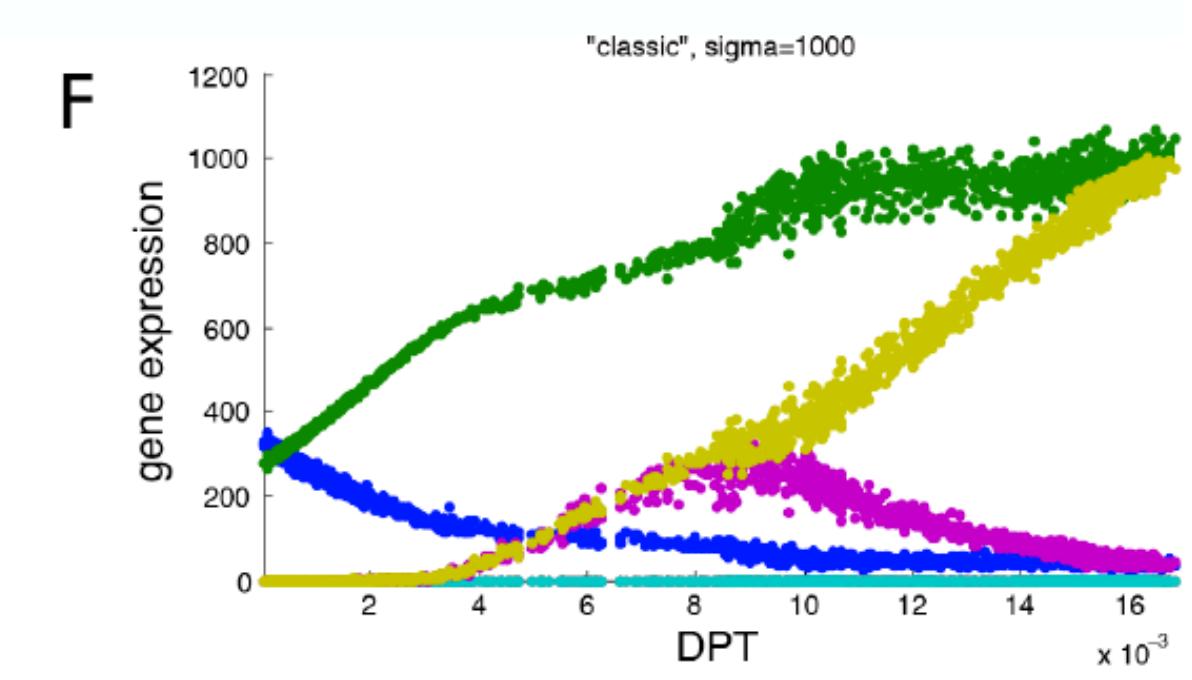
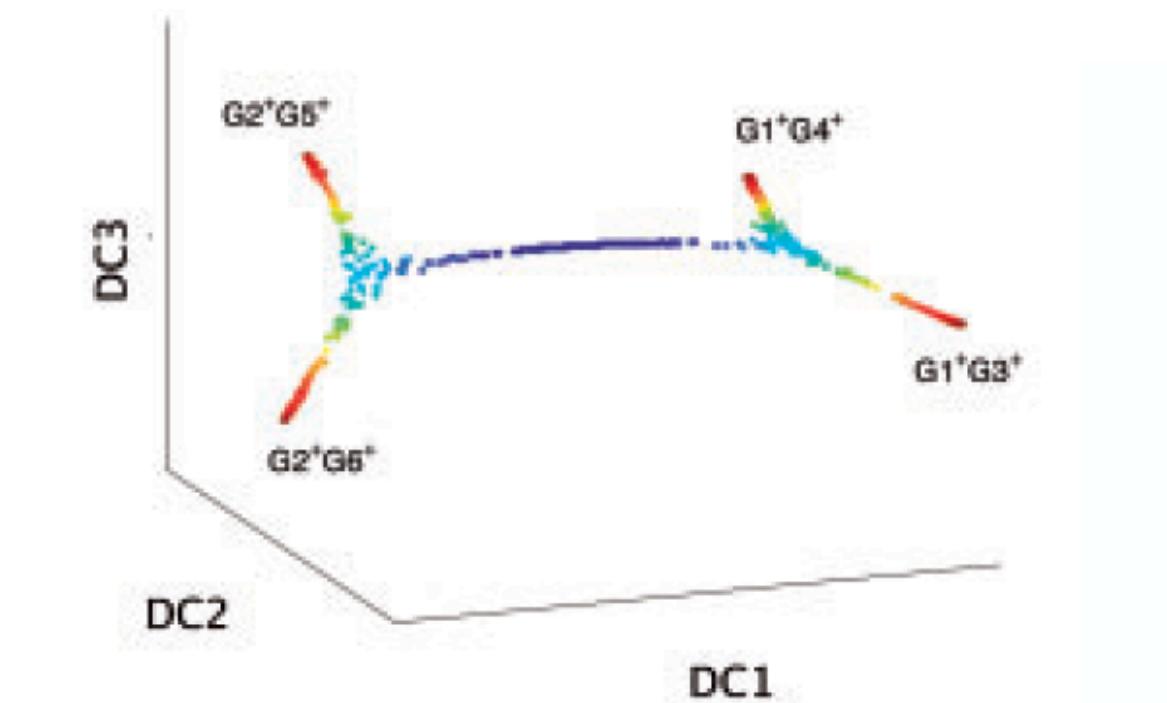
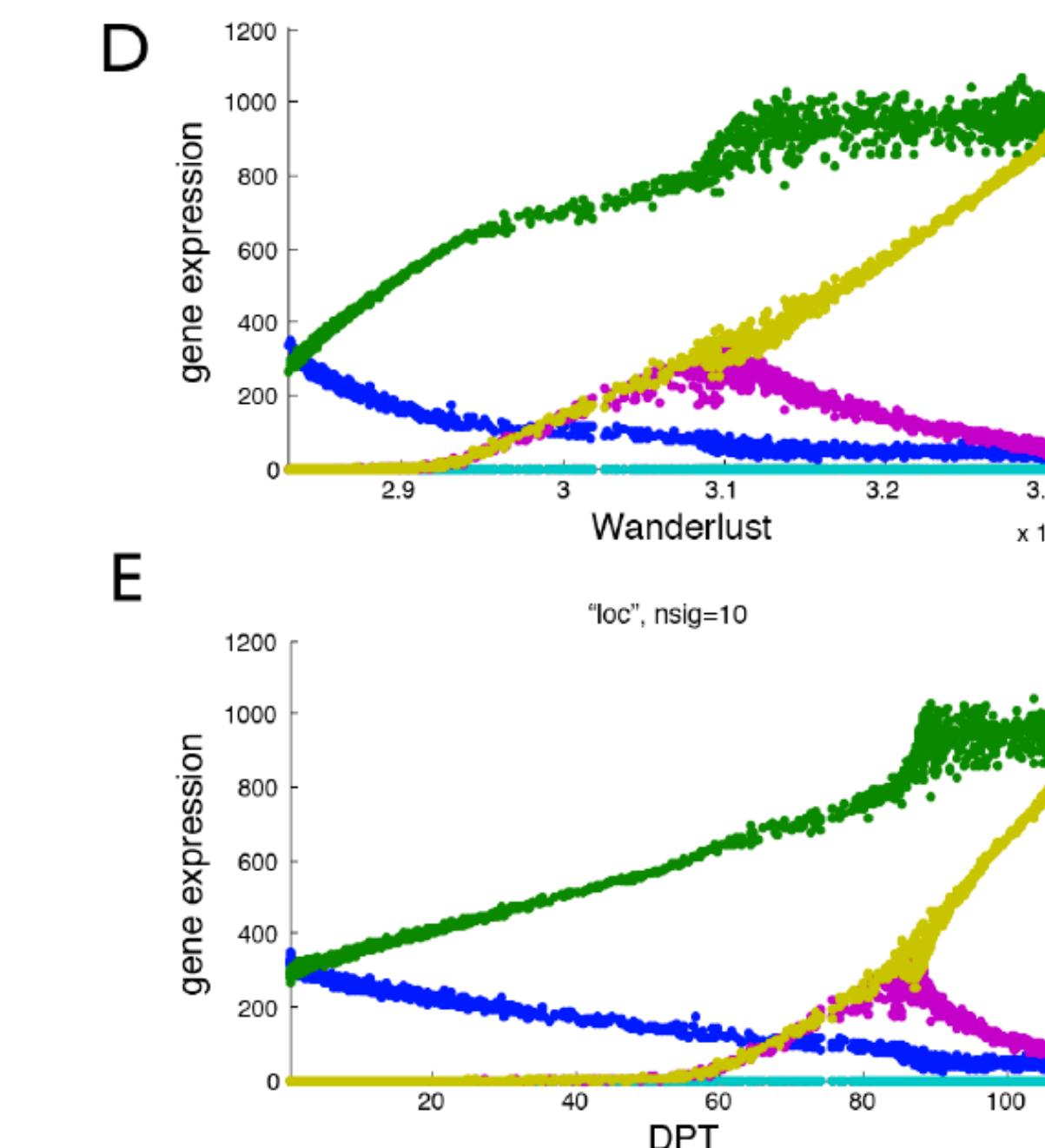
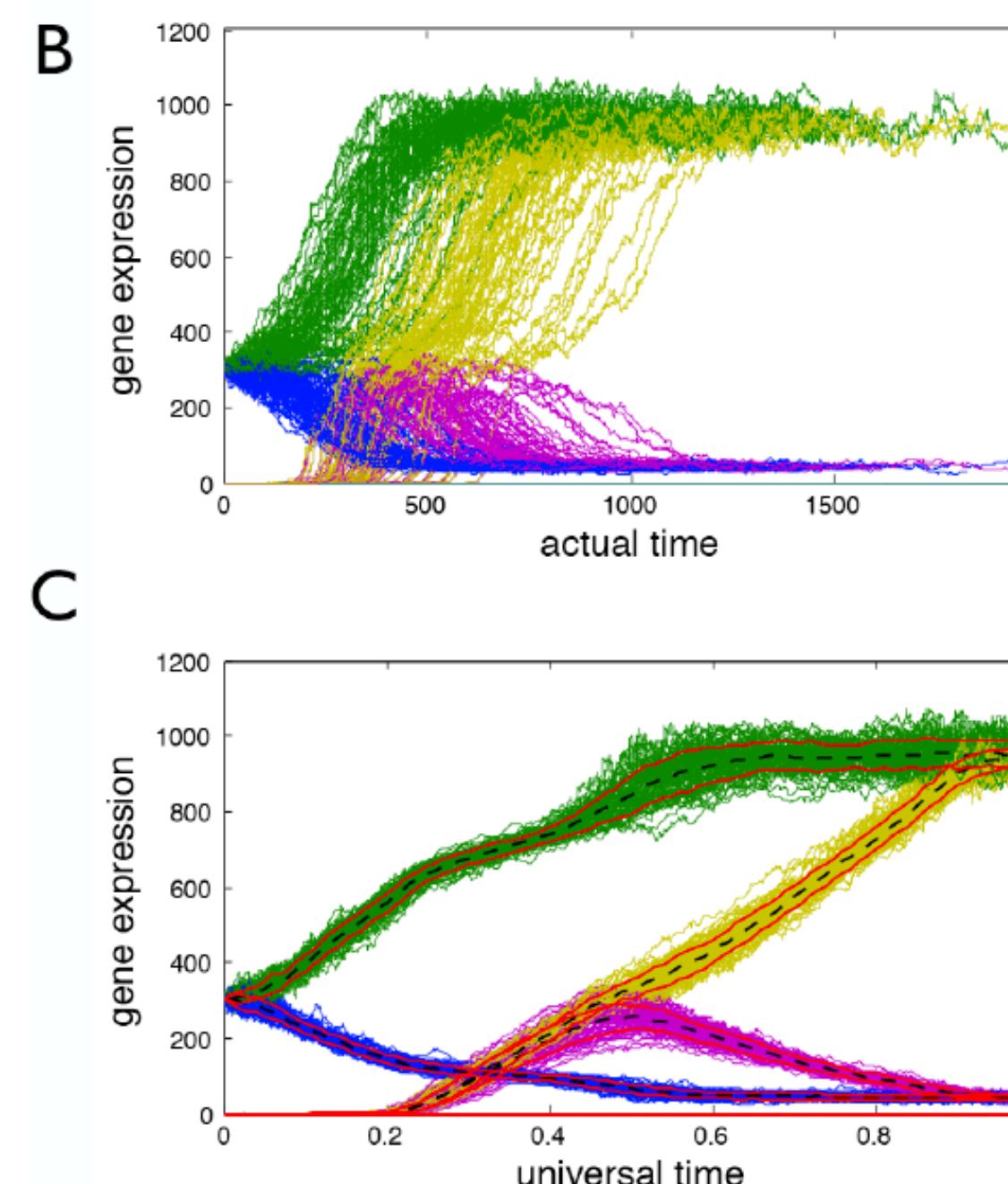
- Under assumption of equal time steps of sampling from an individual trajectory, as well as stationary distribution of cells sampled from multiple trajectories we have $1/\langle \text{density} \rangle \sim \langle \text{velocity} \rangle$



$$1/\rho(t) = v(t)$$

universal time

$$u(t) = \int_0^t \frac{1}{\rho(t')} dt$$

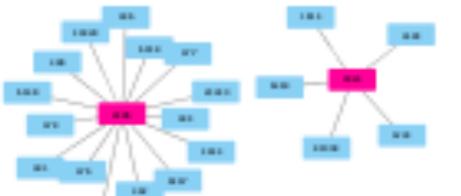
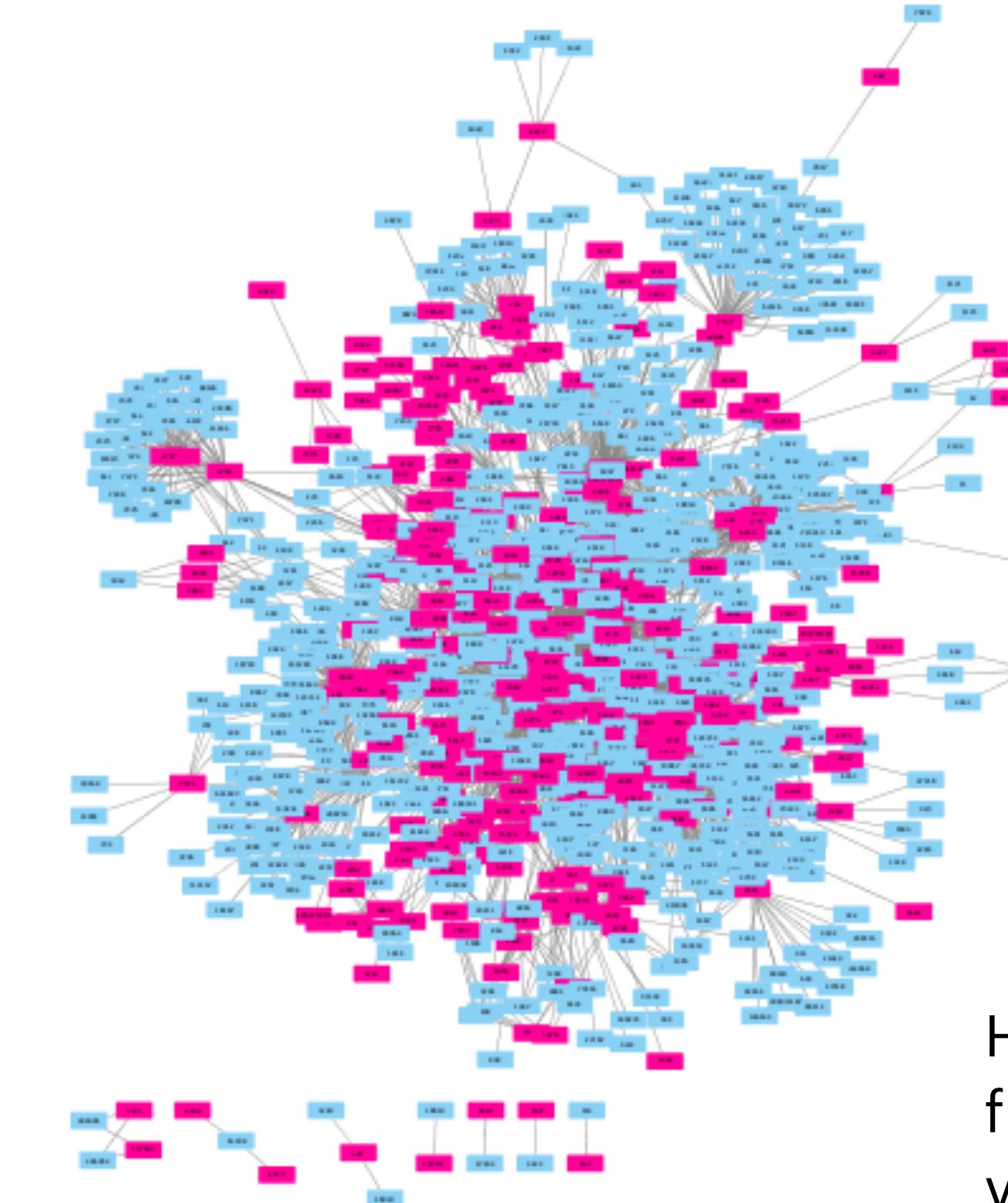
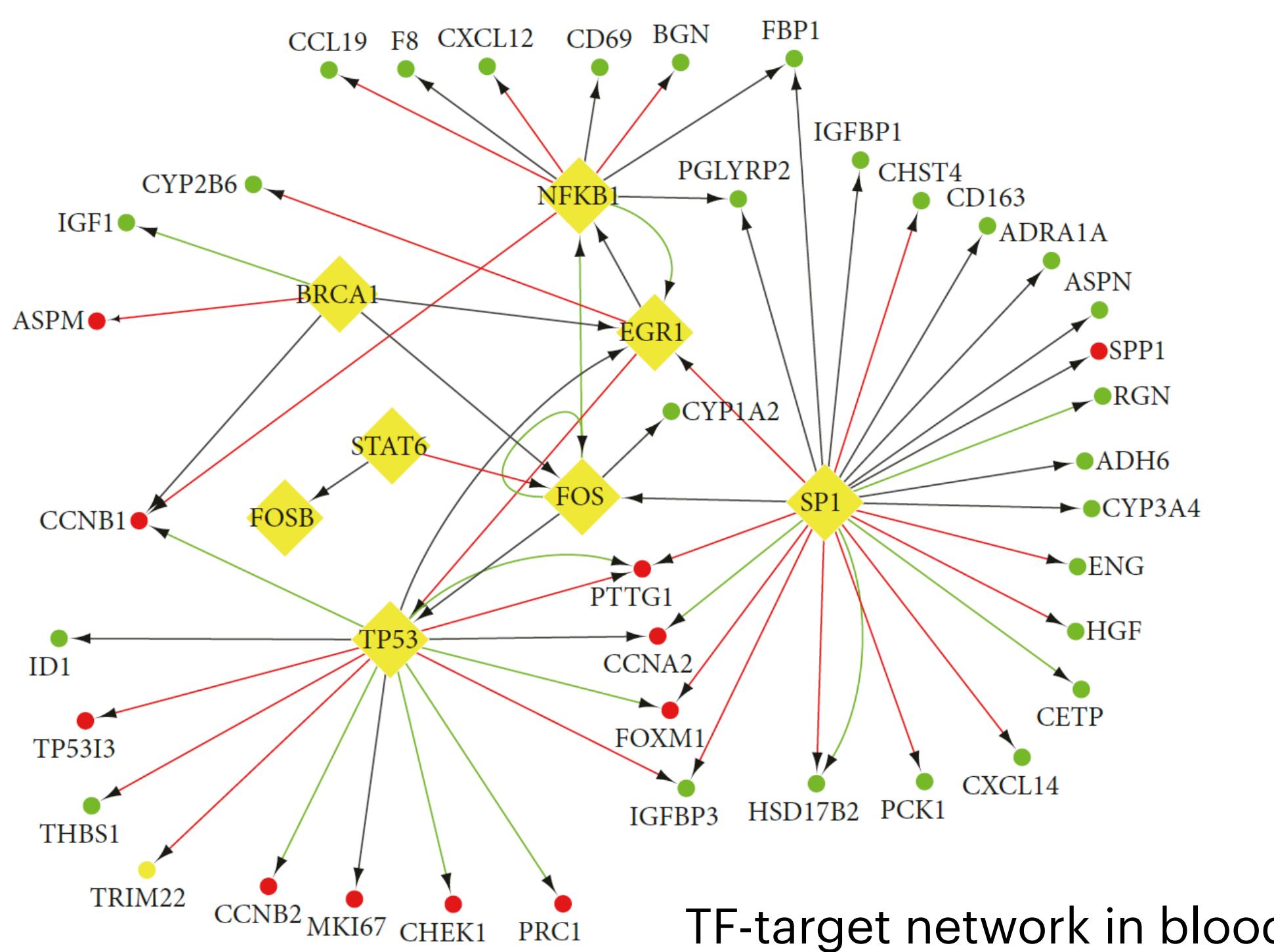


Exercises:

- a) Show that under assumption of equal time steps of sampling from an individual trajectory, as well as stationary distribution of cells sampled from multiple trajectories we have $1/\langle \text{density} \rangle \sim \langle \text{velocity} \rangle$ in each neighbourhood.

GRNs and TF networks

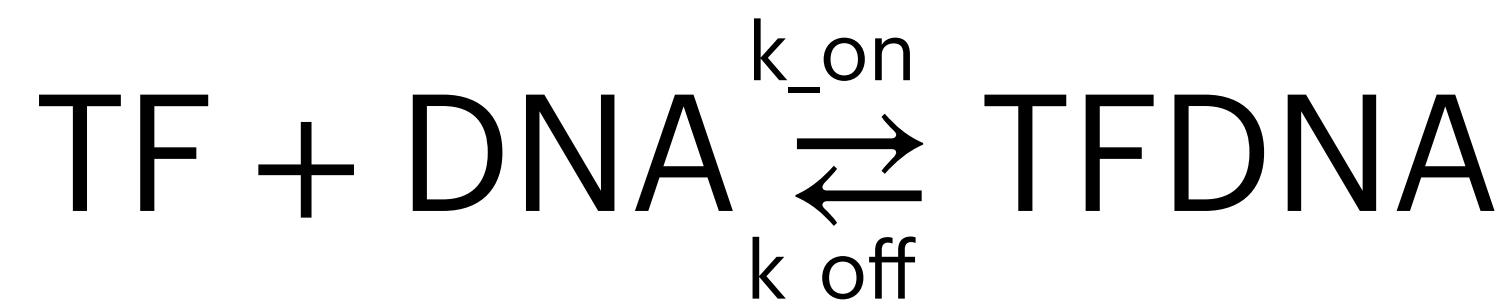
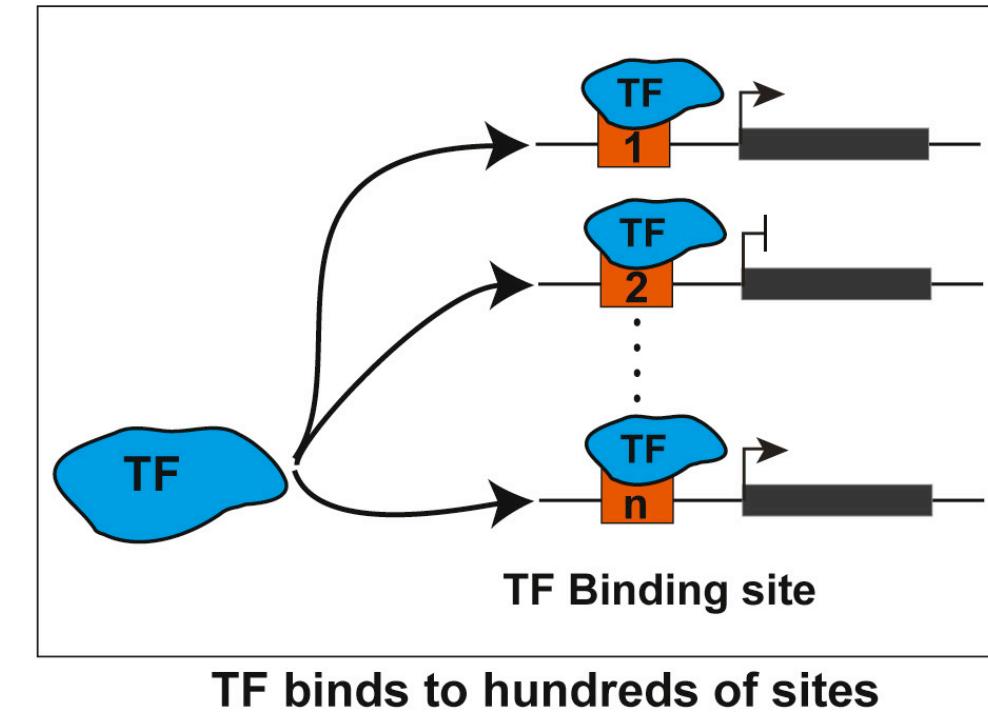
- Human and mouse: ~20k genes
- ~4000 TFs that directly bind to DNA
- Lowly expressed to regulate other genes



Human TFs and target genes from a database (STRING), visualisation by Cityscape

Single TF-promoter binding

Alon, U., 2019. An introduction to systems biology:
design principles of biological circuits.

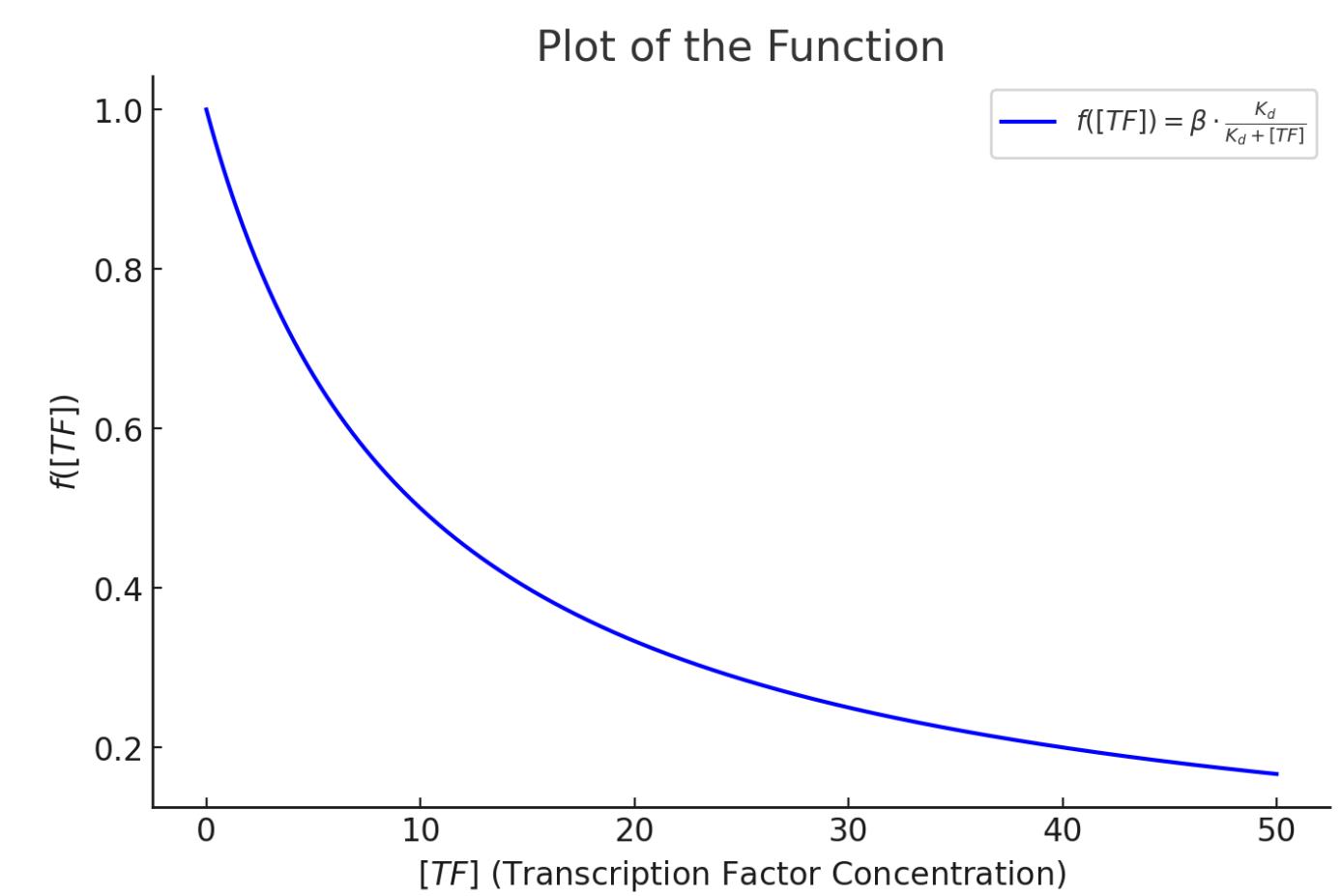


$$\frac{[TF][DNA]}{[TFDNA]} = \frac{k_{off}}{k_{on}} = K_d$$

Transcription rate \beta

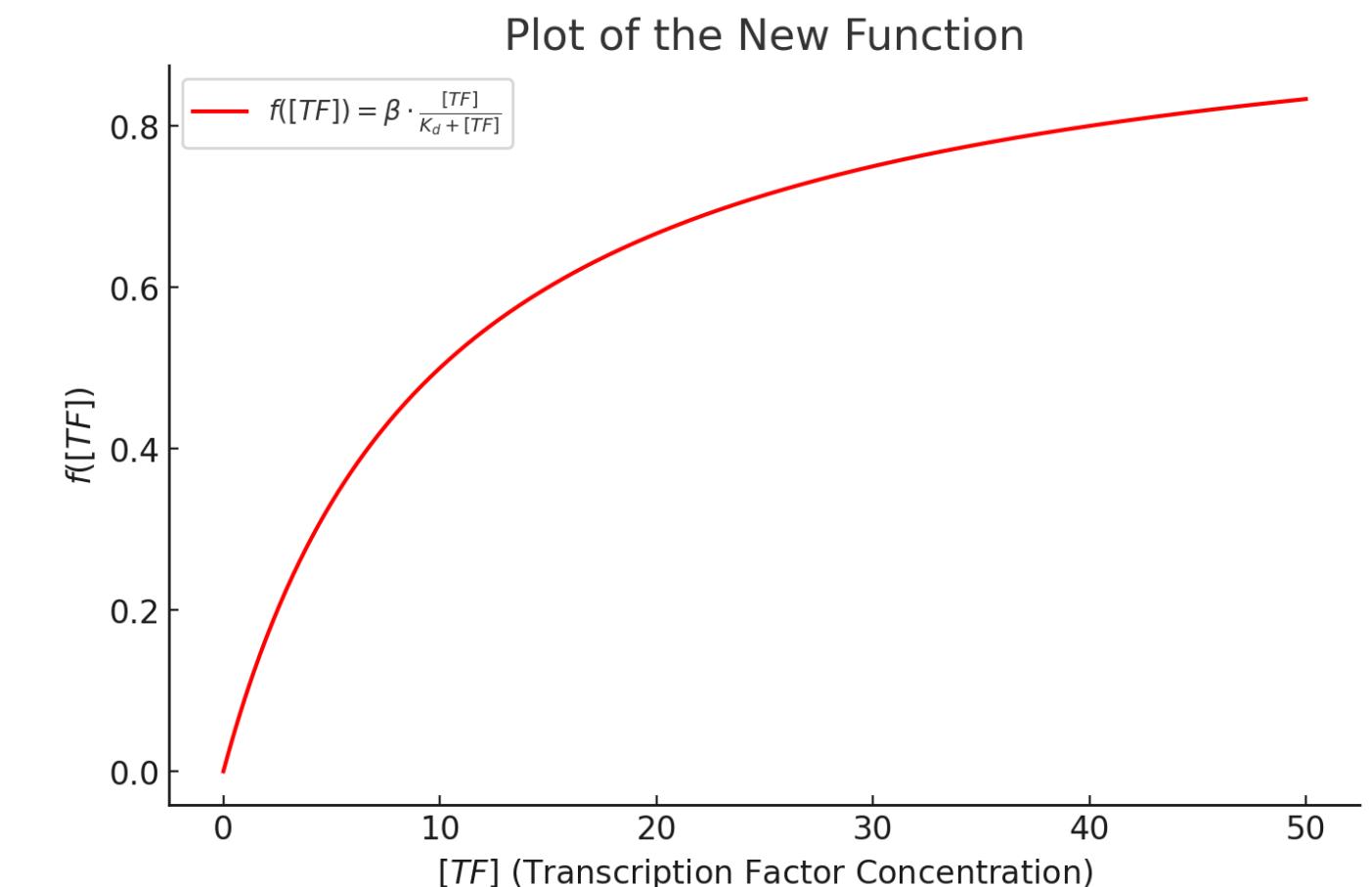
Inhibition

$$\text{Promoter activity} = \beta \cdot Pr(DNA) = \beta \cdot \frac{[DNA]}{[DNA] + [TFDNA]} = \beta \cdot \frac{K_d}{K_d + [TF]}$$

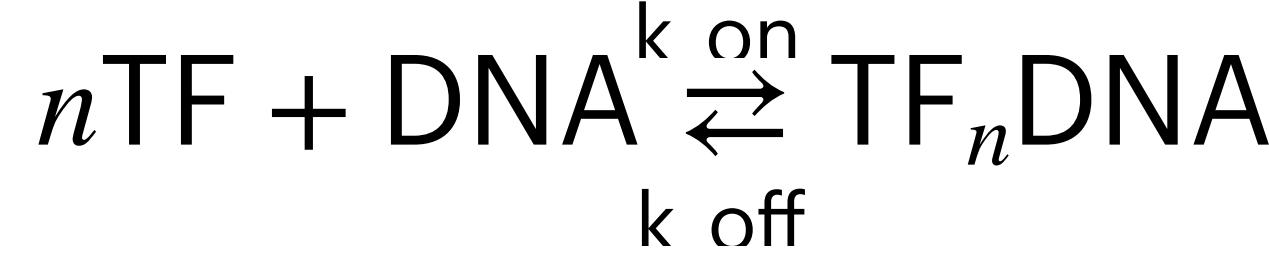


Activation

$$\text{Promoter activity} = \beta \cdot Pr(TFDNA) = \beta \cdot \frac{[TFDNA]}{[TFDNA] + [TFDNA]} = \beta \cdot \frac{[TF]}{K_d + [TF]}$$



Cooperative TF-promoter binding (activation)



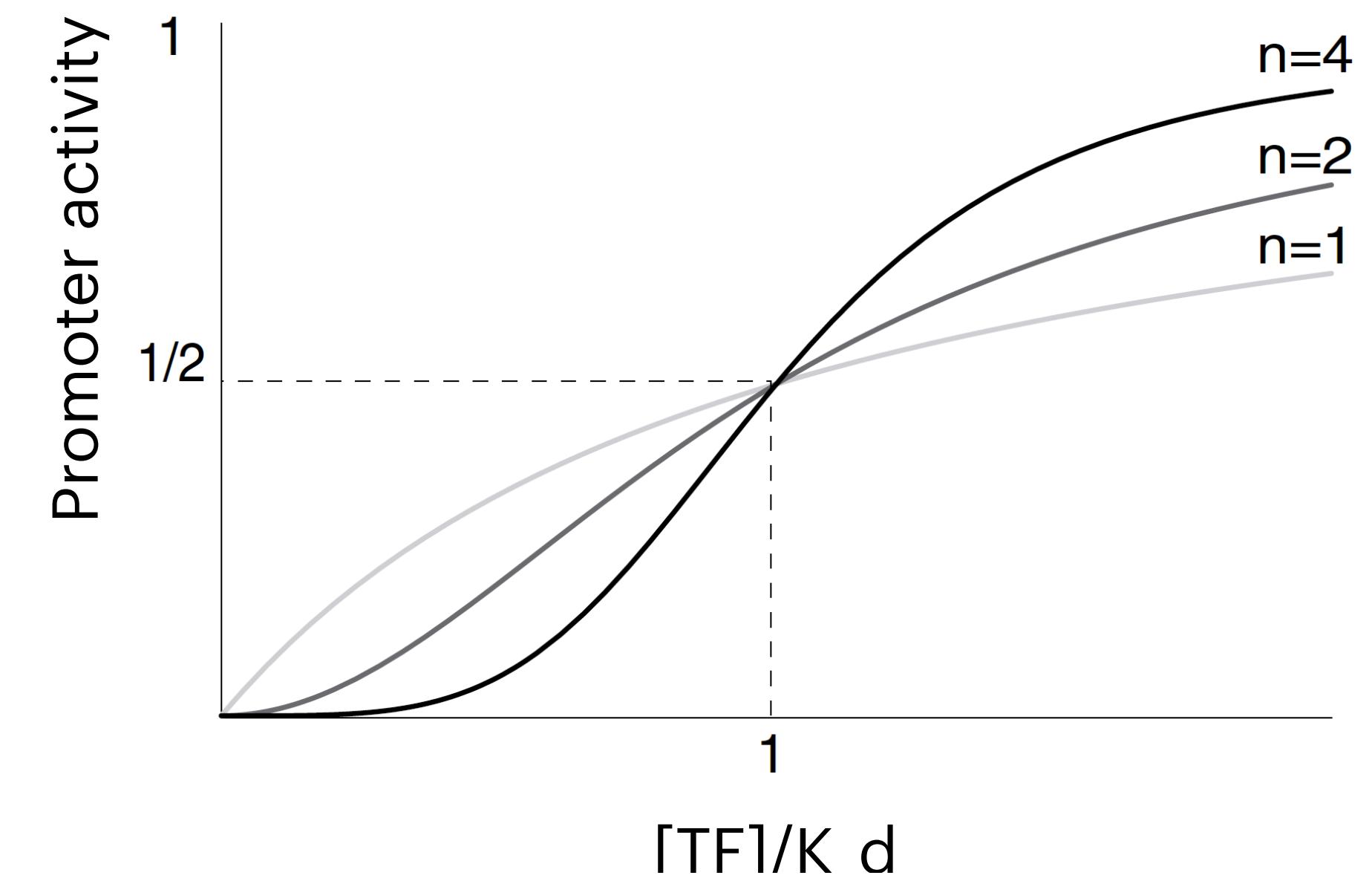
$$\frac{[TF]^n[DNA]}{[TF_nDNA]} = \frac{k_{off}}{k_{on}} = K_d^n$$

Hill function

Activation

Promoter activity = $dX/dt =$

$$\beta \cdot Pr(TF_nDNA) = \beta \cdot \frac{[TF_nDNA]}{[TF_nDNA] + [TF_nDNA]} = \beta \cdot \frac{[TF]^n}{K_d^n + [TF]^n}$$



- $N \geq 2 \rightarrow$ Sigmoid

Exercises:

- a) Write the promoter activity function for cooperative inhibition when two binding sites of the same TF on the promoter of a target gene are needed for its inhibition.
- b) Write the promoter activity function for cooperative inhibition by two distinct TFs, TF1 and TF2 and make the 3D plots for target gene activity (z axis) versus TF1 and TF2 concentration (x and y axes).

Bistability with Hill function production

- $\frac{dx}{dt} = 0$
 - (One or) two stable
 - One unstable

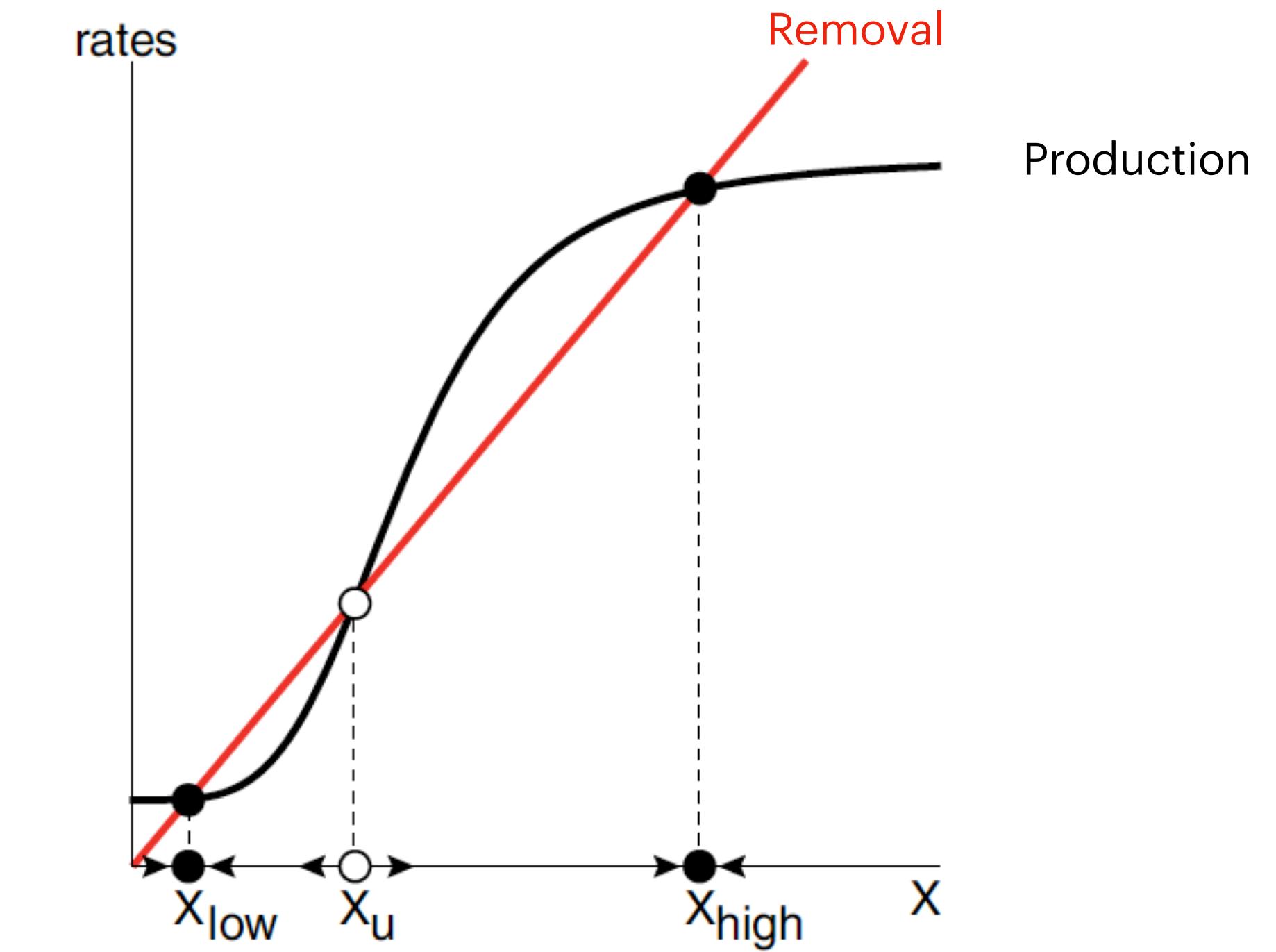
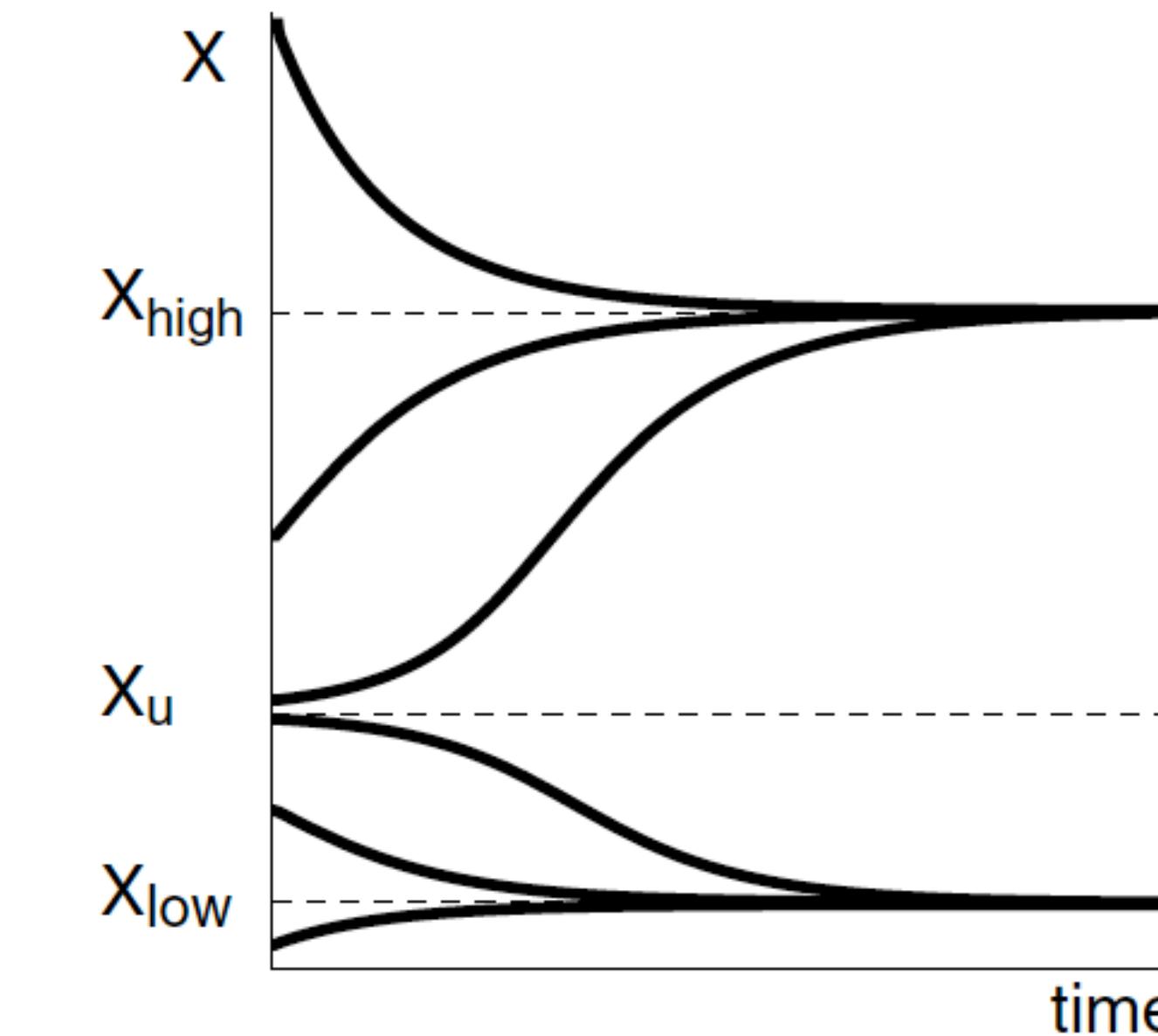


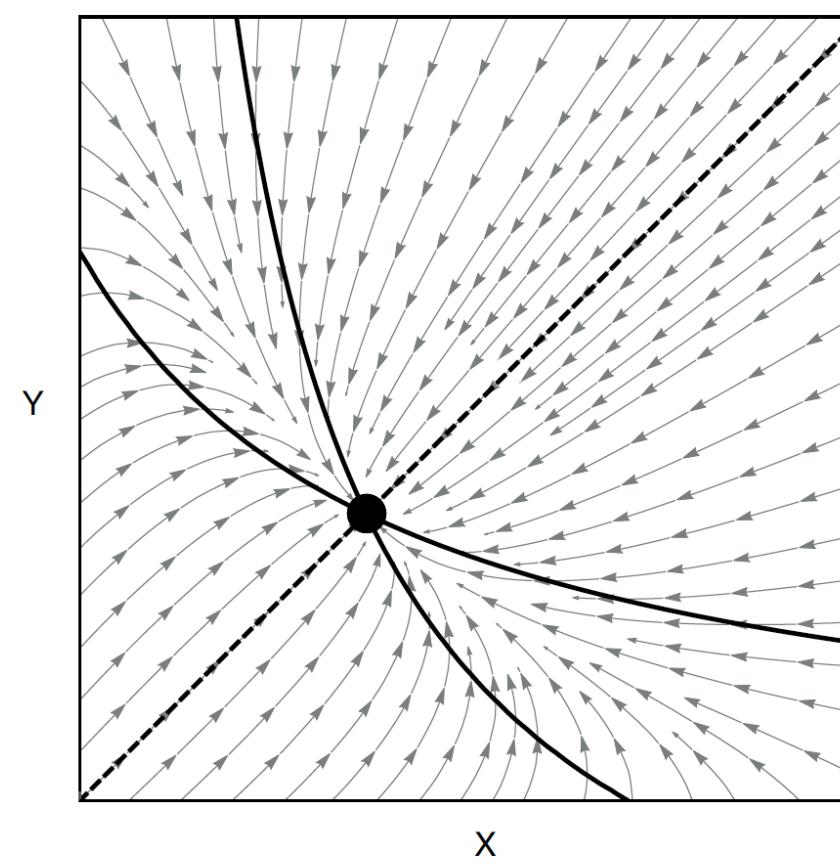
FIGURE 5.4



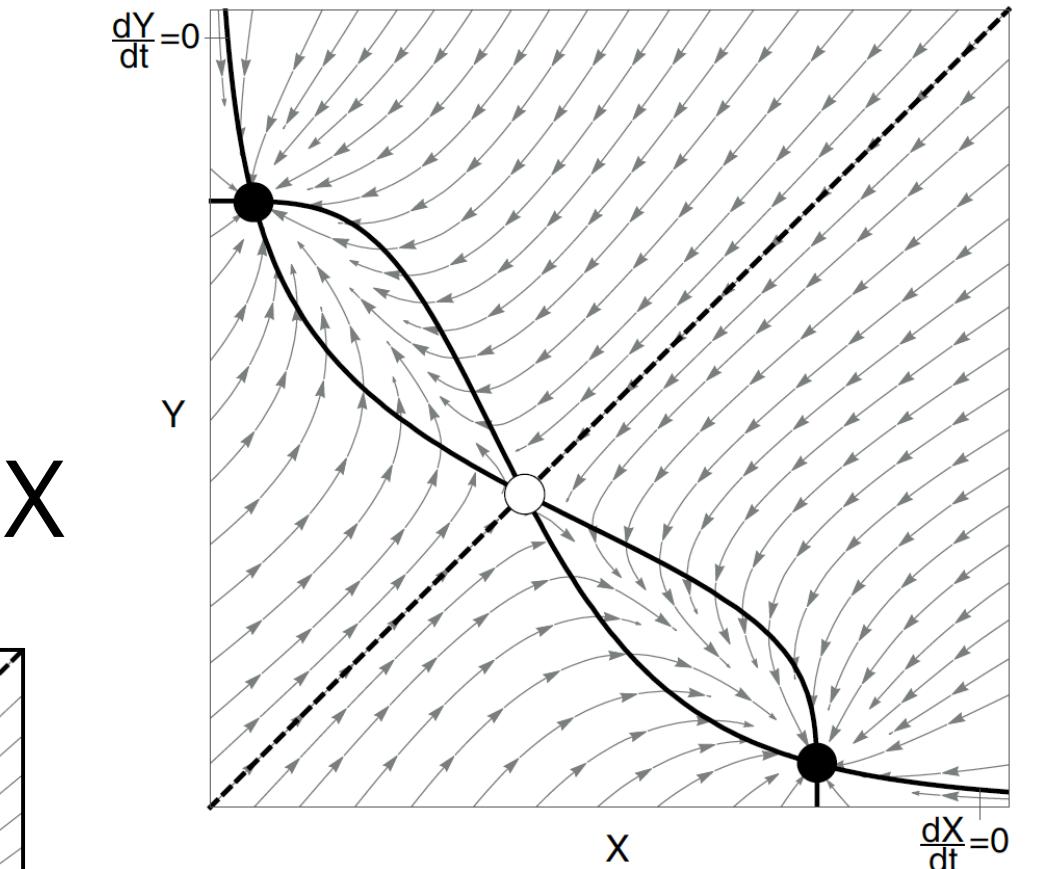
Toggle switch



- Toggle switch with Sigmoid (cooperative) activation functions for both X and Y → Bistable



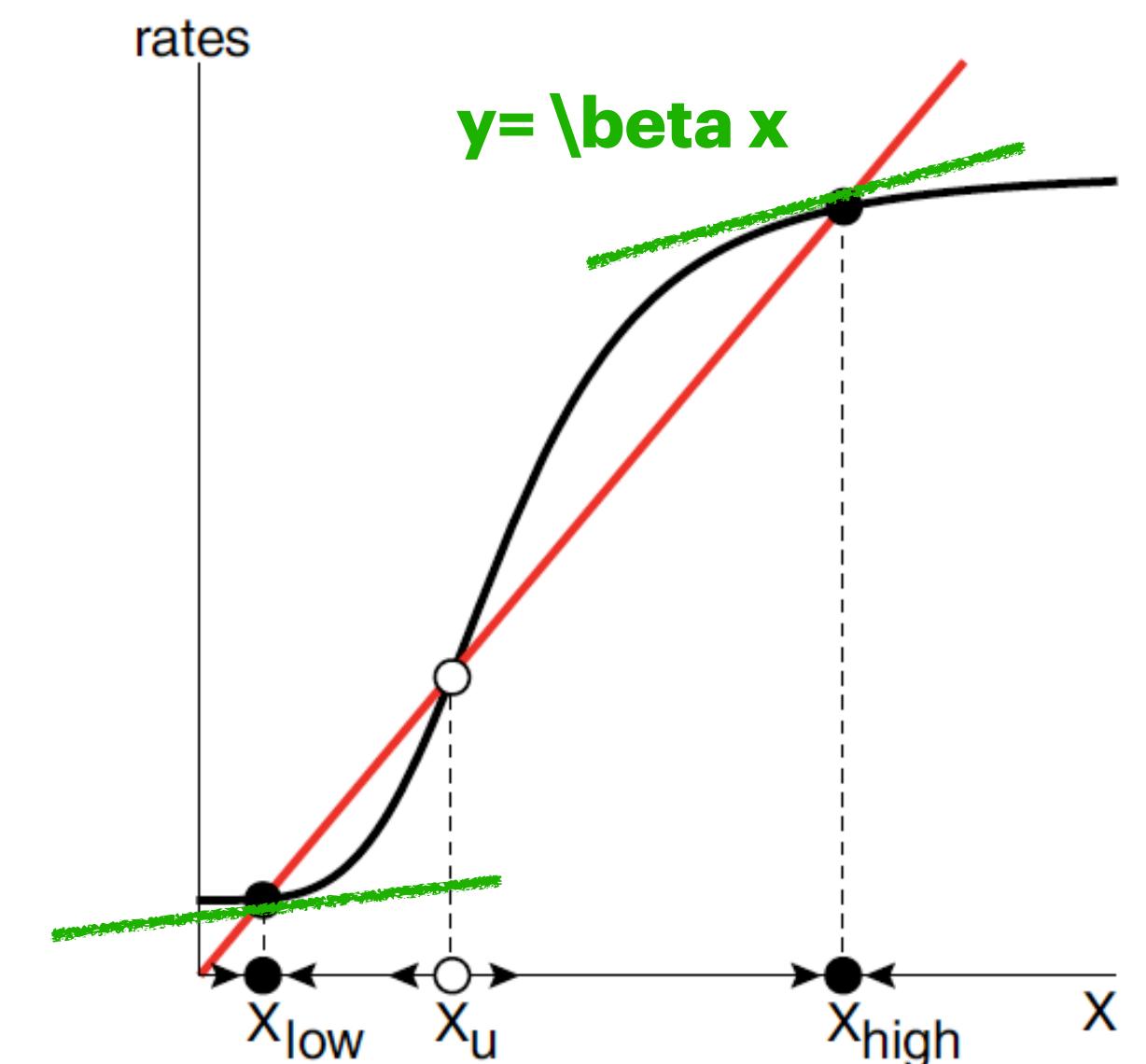
- Toggle switch with only one Sigmoid activation functions for → monostable



- Linear approximation around each fixed point

$$\frac{dX}{dt} = f(Y) - \alpha X$$

$$\frac{dY}{dt} = f(X) - \alpha Y$$



Oscillatory behaviour near the attractor

$$\frac{dX}{dt} = f(Y) - \alpha_1 X$$

$$\frac{dY}{dt} = g(X) - \alpha_2 Y$$

$$\frac{d}{dt} \begin{pmatrix} x \\ y \end{pmatrix} = J \begin{pmatrix} x \\ y \end{pmatrix}, \quad J = \begin{bmatrix} -\alpha_1 & \beta_1 \\ \beta_2 & -\alpha_2 \end{bmatrix}$$

$$2\lambda = -(\alpha_1 + \alpha_2) \pm \sqrt{(\alpha_1 - \alpha_2)^2 + 4\beta_1\beta_2}$$

- Real part positive or negative?
- Real part negative and Imaginary eigenvalues? → Damped oscillation
- Noise can induce oscillations in systems with theoretical Damped oscillation

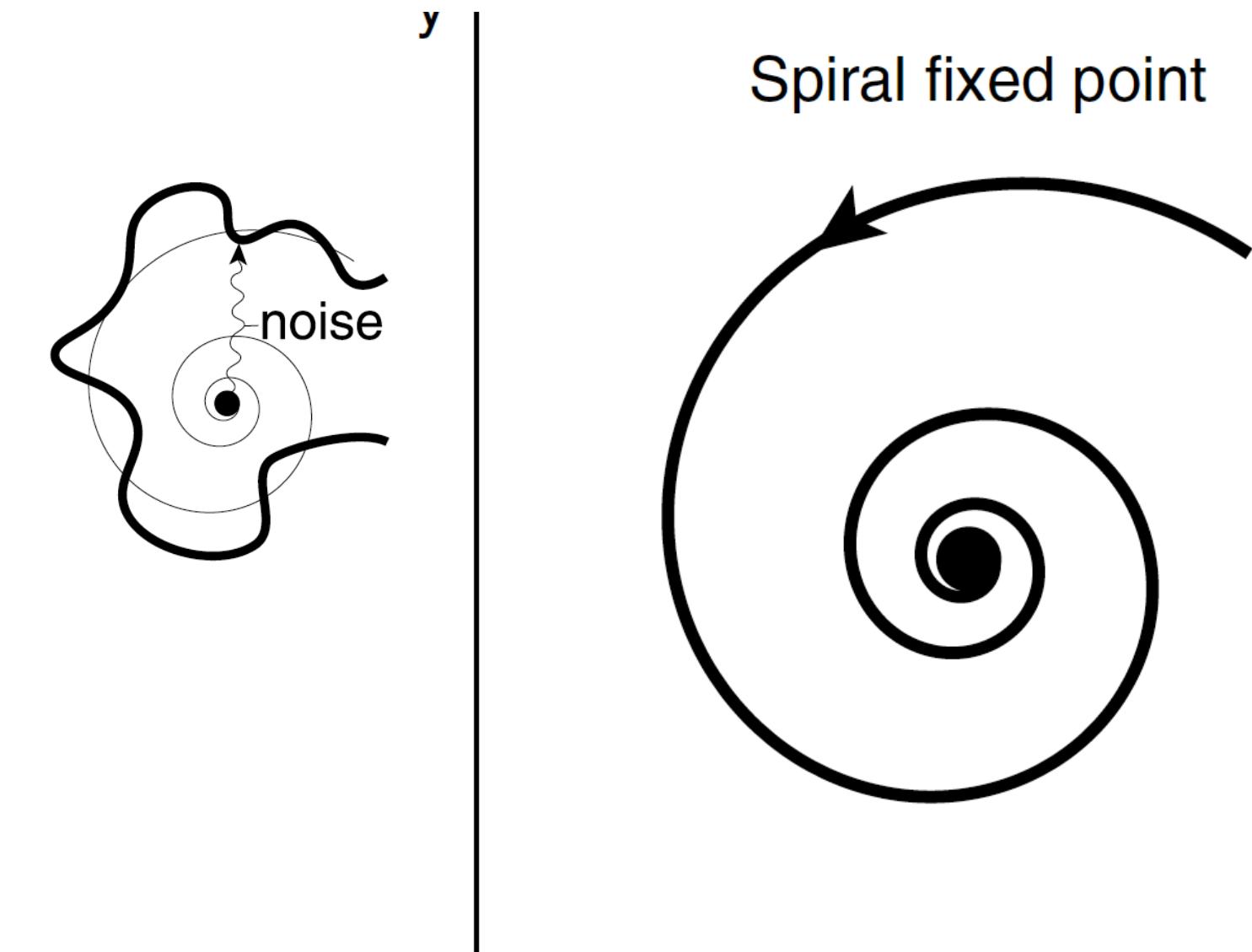
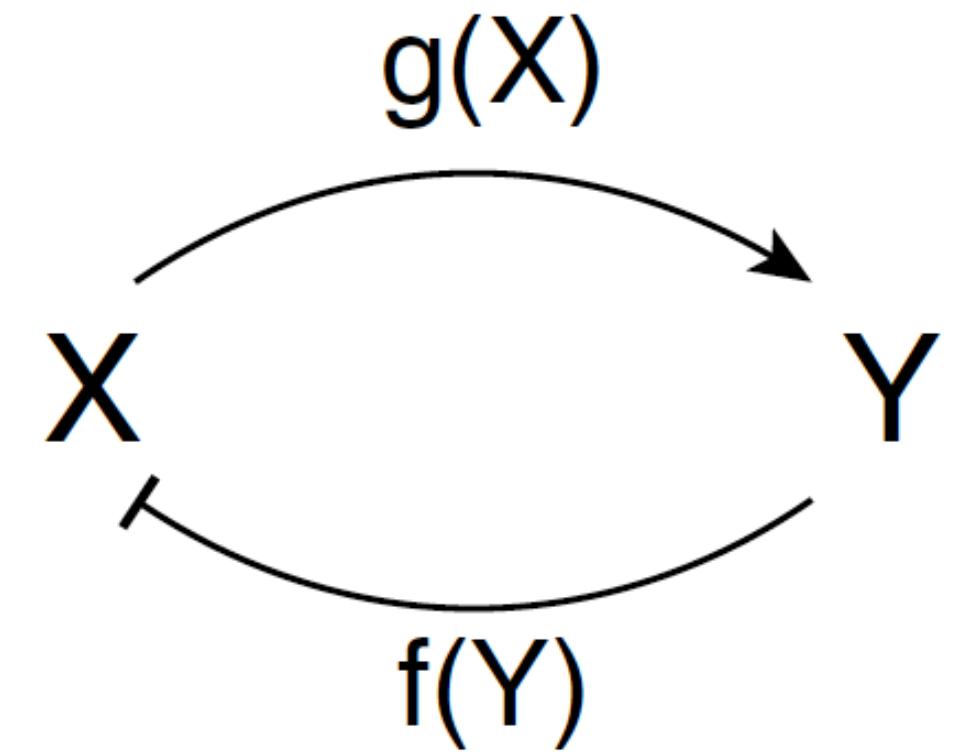
Linear expansion around an attractor

$$\frac{dx}{dt} = \beta_1 y - \alpha_1 x$$

$$\frac{dy}{dt} = \beta_2 x - \alpha_2 y$$

$$x = \operatorname{Re}(c_1 e^{\lambda_1 t} + c_2 e^{\lambda_2 t})$$

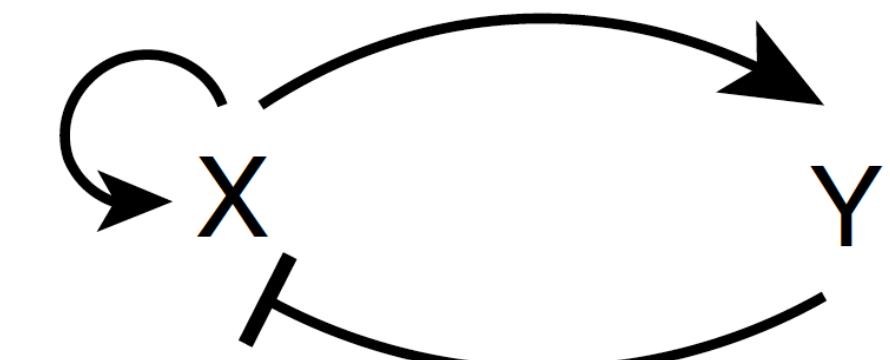
$$y = \operatorname{Im}(c_1 e^{\lambda_1 t} + c_2 e^{\lambda_2 t})$$



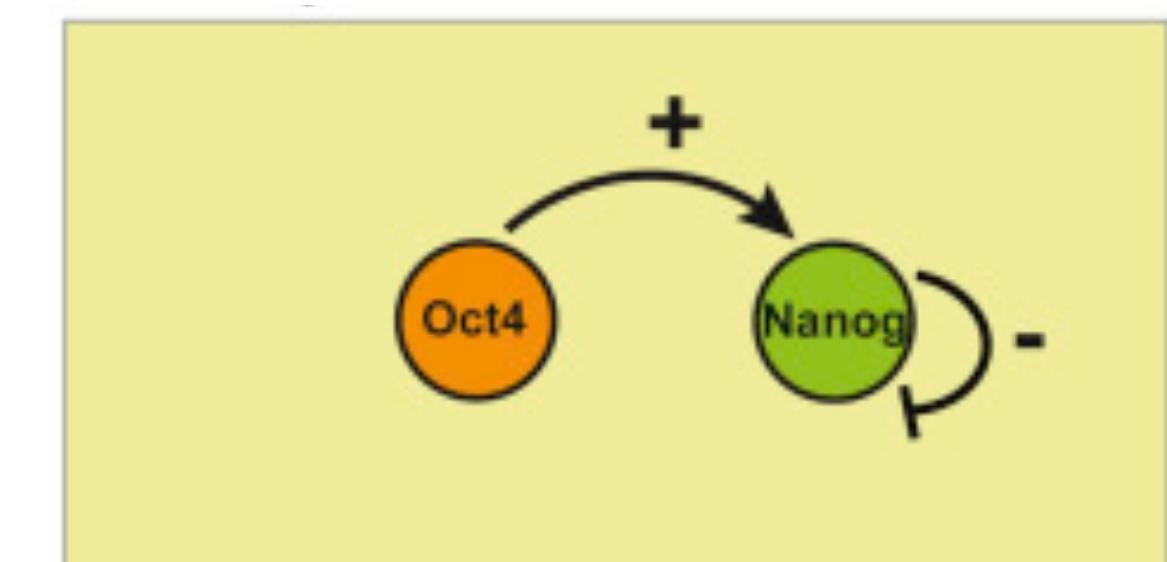
Exercises:

- a) Show that the following motif can result in oscillation around an attractor.

oscillator motif



- b) Show the conditions on alpha, beta that result in damped oscillation around the attractor
- c) What if Y has self activation instead of X?
- d) What if both X and Y have self activation?
- e) Which of these motifs can be assumed for cell type transitions (i.e., exiting one attractor and entering another)?
- f) Analyse the attractor states of the following motif of pluripotency (stemness). Under what parameter ranges is we get an stable attractor?



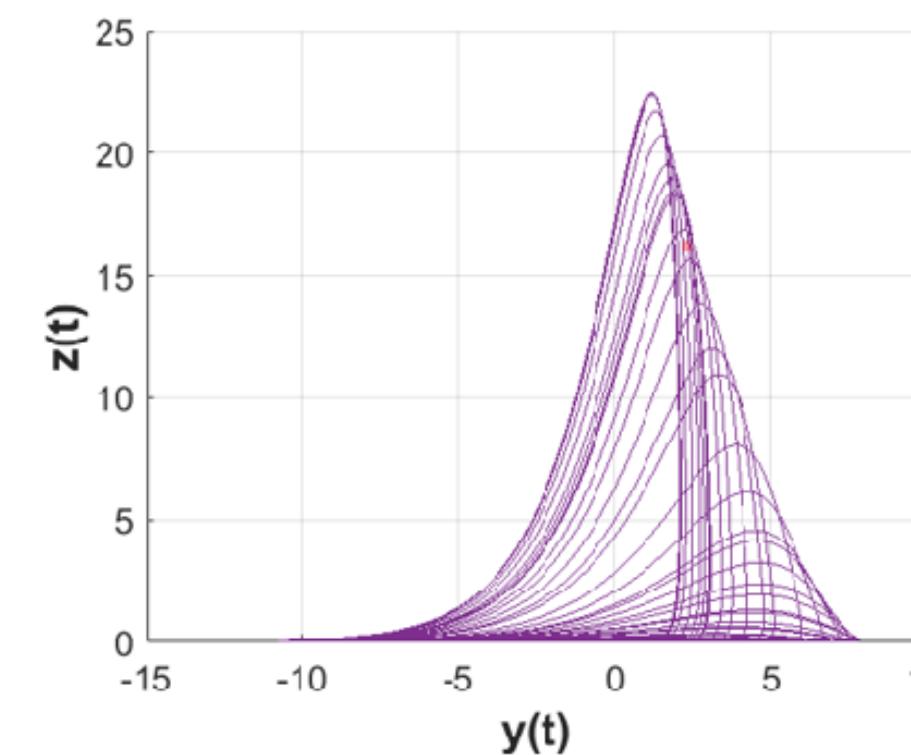
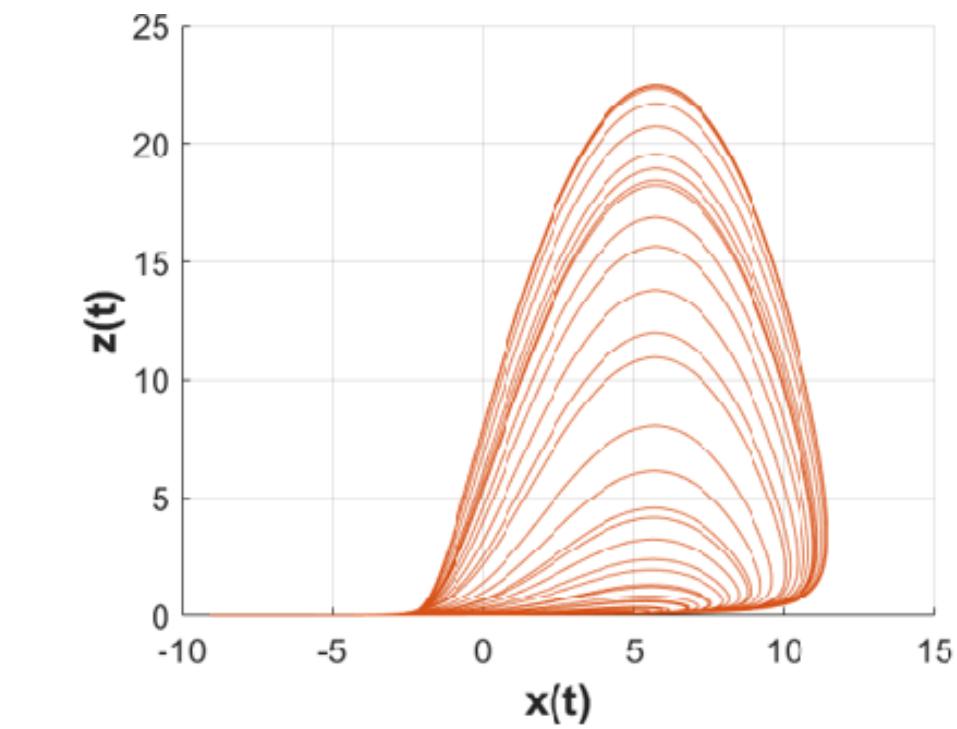
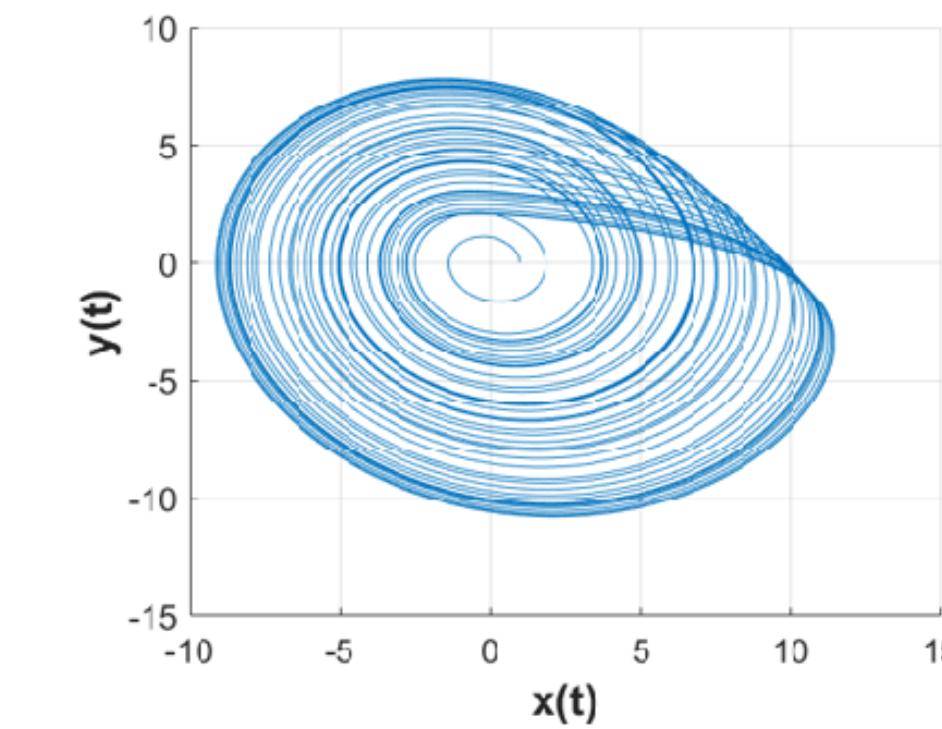
Direct disruption of core pluripotency transcription motif using CRISPRd

Chaotic attractor? (e.g. Roessler attractor)

- Close initial conditions → deviating in trajectories
- Long-Term versus Short-Term Haematopoietic Stem Cells
 - LT-HSCs → ST-HSCs → MPPs → CLPs (lymphoid) or CMPs (myeloid): life-long self renewal
 - ST-HSCs dominate during infections/bleeding (rapid myeloid output): a few weeks self renewal

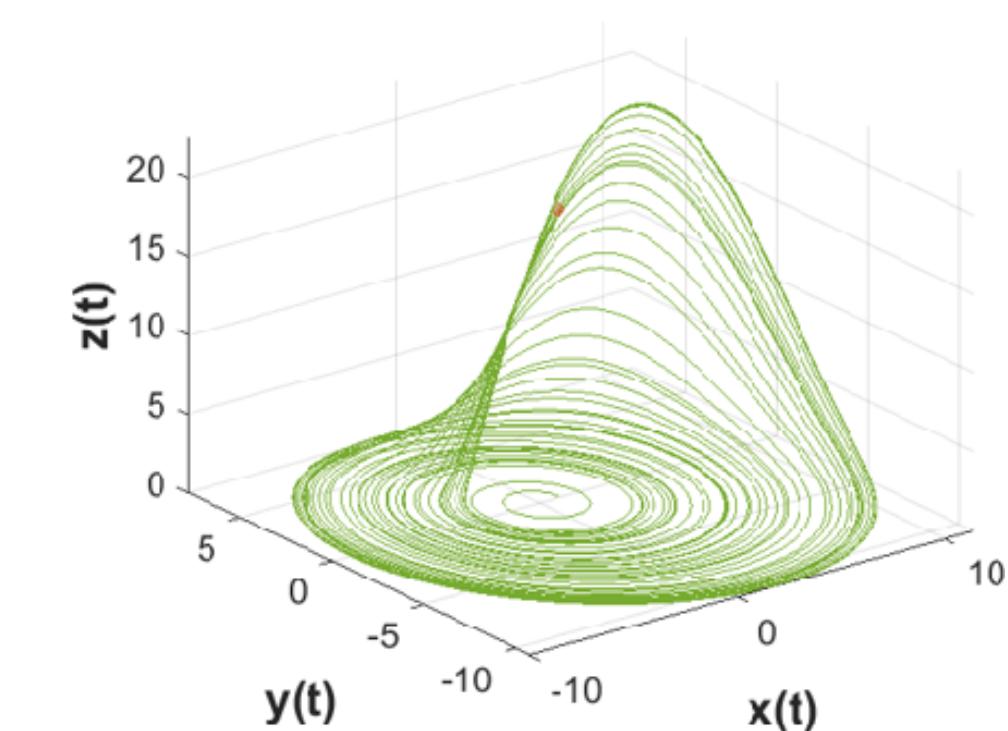
Nonlinear
ODEs

$$\begin{cases} \frac{dx}{dt} = -y - z \\ \frac{dy}{dt} = x + ay \\ \frac{dz}{dt} = b + z(x - c) \end{cases}$$



For $z=0$ becomes linear →

$$\begin{cases} \frac{dx}{dt} = -y \\ \frac{dy}{dt} = x + ay \end{cases}$$



Exercises:

We saw how a toggle GRN with two hill functions can give rise to two attractors

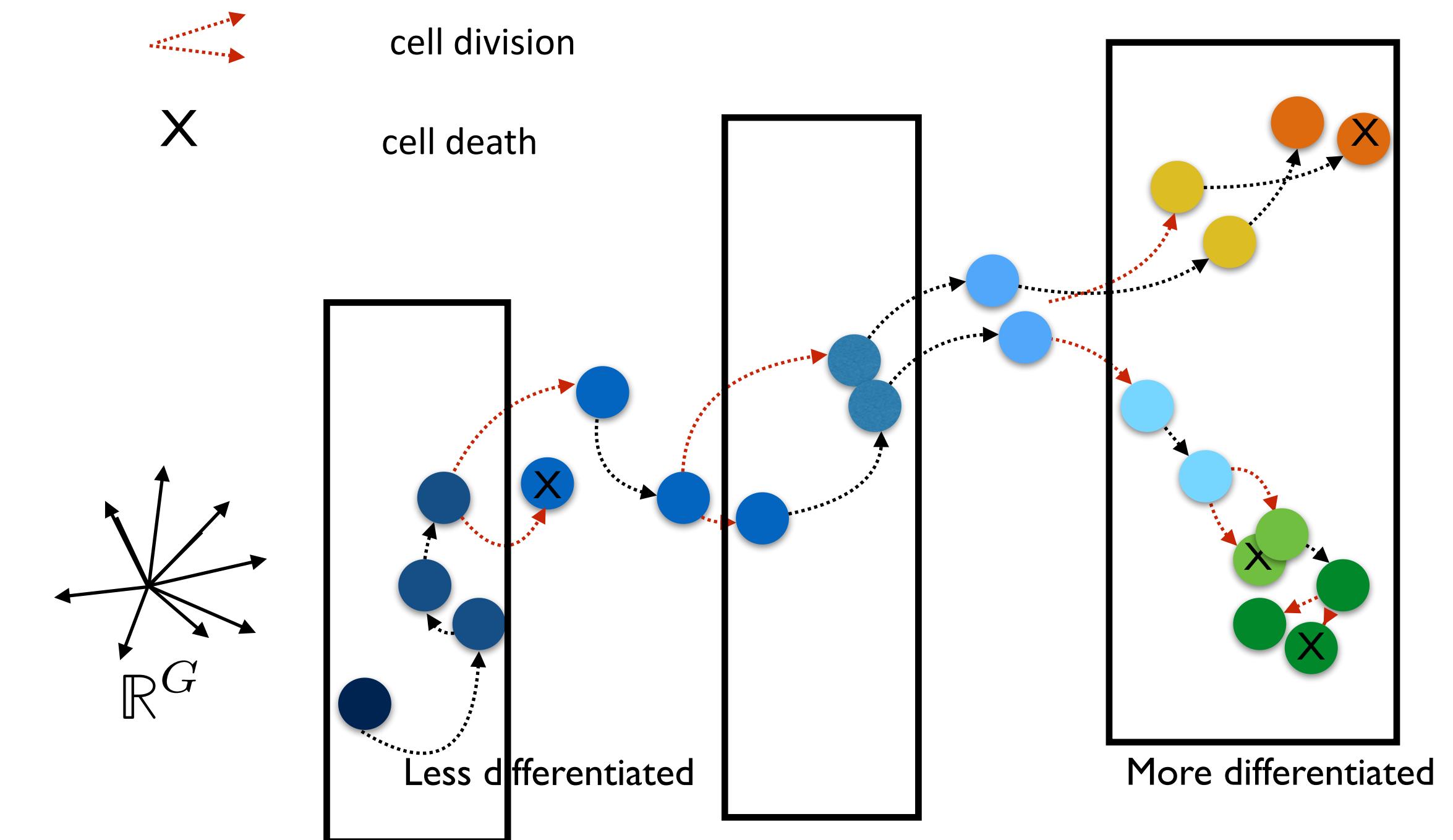
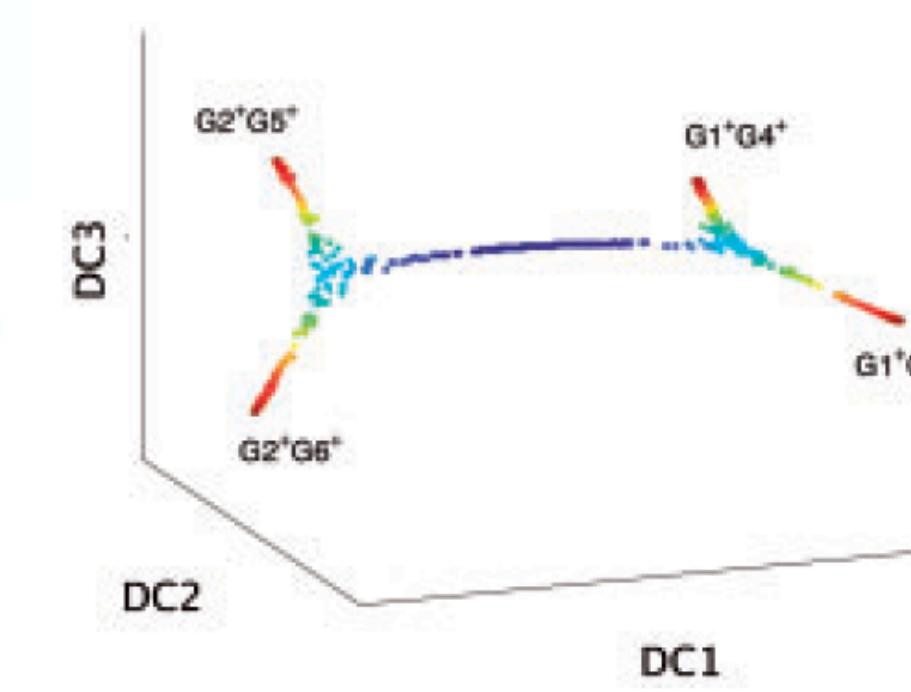
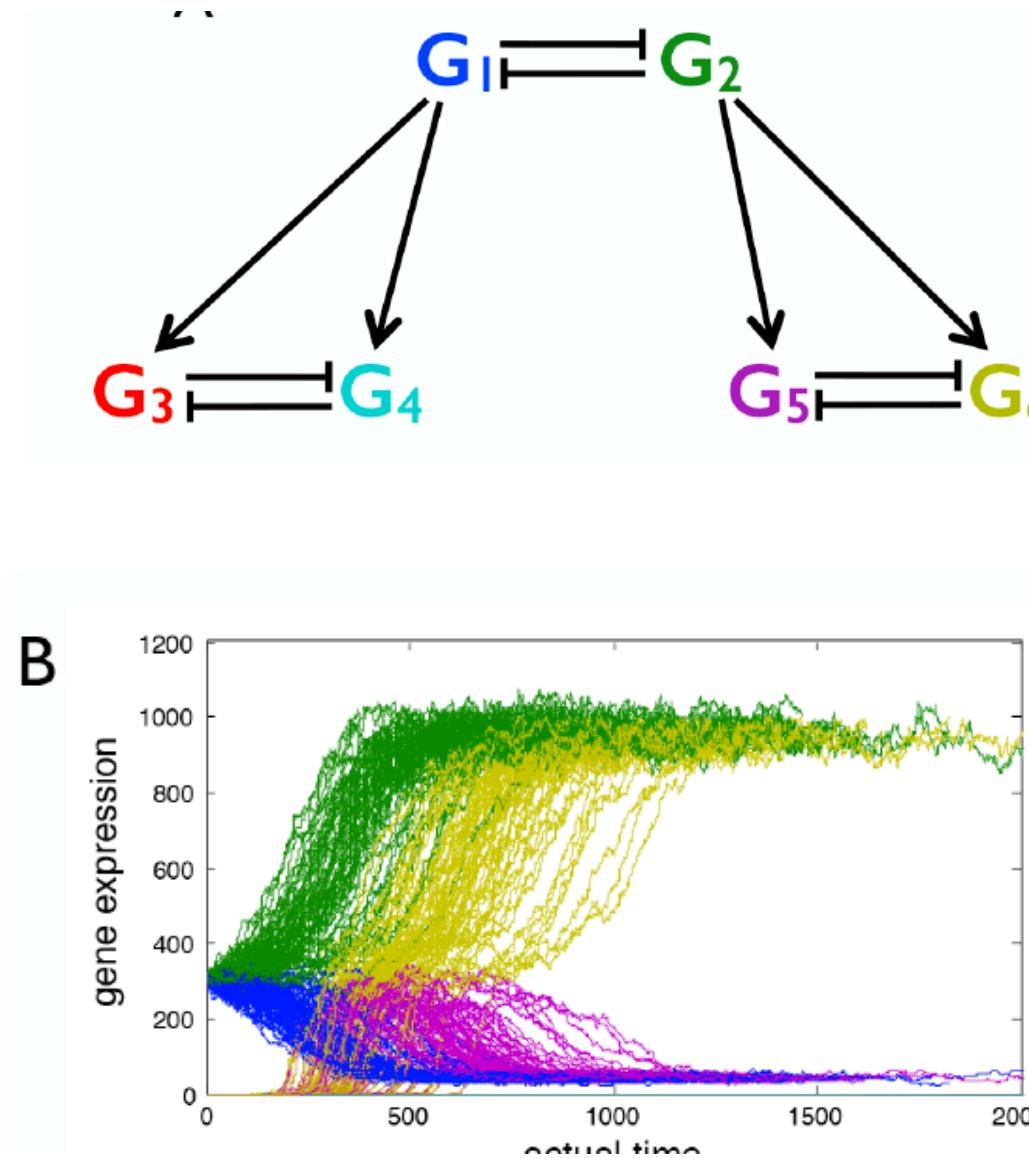
We also saw oscillator and damped oscillator GRNs

- a) Can you make a GRN with chaotic dynamics? E.g. based on the Roessler attractor?
- b) Is it likely or unlikely that such a dynamics appears in natural GRNs?
- c) How will the dynamics of a chaotic system (e.g. Roessler) be with added noise?
- d) How can we characterise the dynamics around an attractor of a D-dimensional GRN? (So far we have considered 2 genes only)
- e) Can you by considering D genes, each with a Sigmoid activation function create a 3D chaos system (Roessler, Lorenz, etc.)?

Exercises:

In one of the exercises from the Dynamical Inference session, we show from a snapshot data if we have the coordinate and velocity of each cell, we can determine the diffusion/drift magnitude, assuming a Fokker-Planck time evolution dynamics.

- a) Can we from a snapshot data tell if the dynamics of a system is chaotic?
- b) How about from two or more time windows?
- c) Does clonal structure (a genomics measure which does not change over time) help to identify if the system is chaotic or not?
- d) Can you demonstrate your arguments with a simulation of cells sampled from multiple trajectories of a chaotic, versus a Wiener process or a GRN circuit?



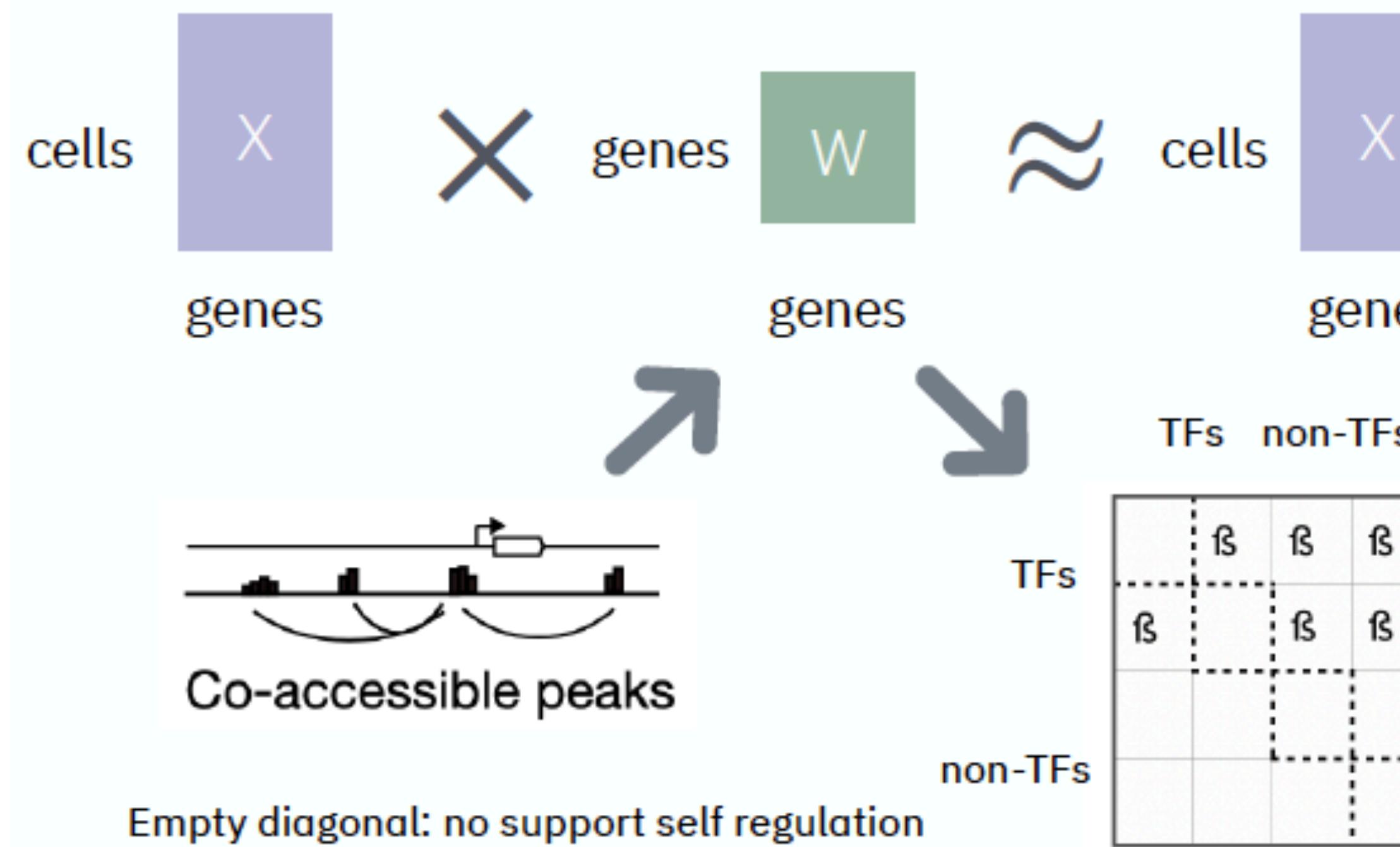
GRNs inference from single-cell omics

- (In a few datasets that we checked) on average ~30 TFs binding sites at the promoter of a target gene
- Forming TF complexes
- Enhancer binding in addition to promoter
- ~4000 TFs encoded in human and mouse genome
- More cooperating signalling and external factors
- We discussed creation of dynamics from a GRN but GRN inference is an inverse problem

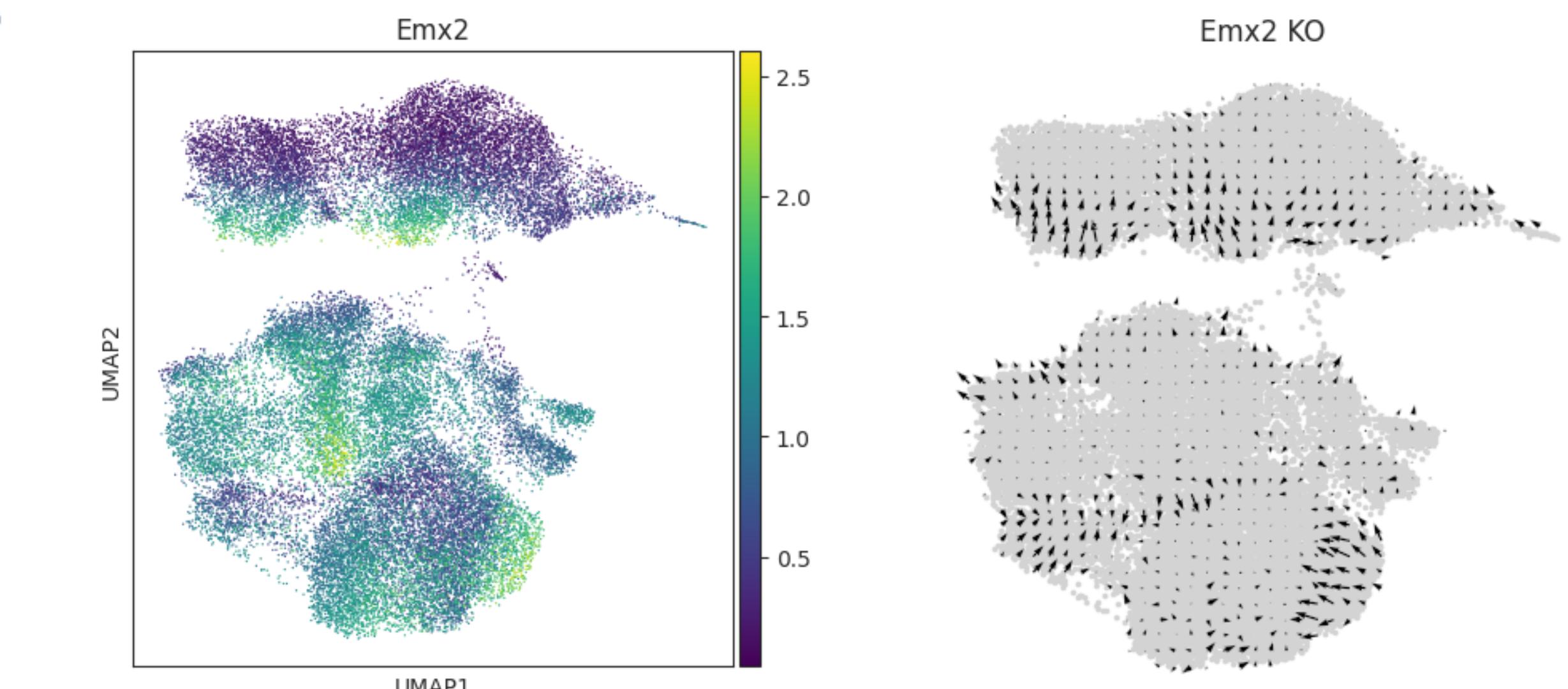
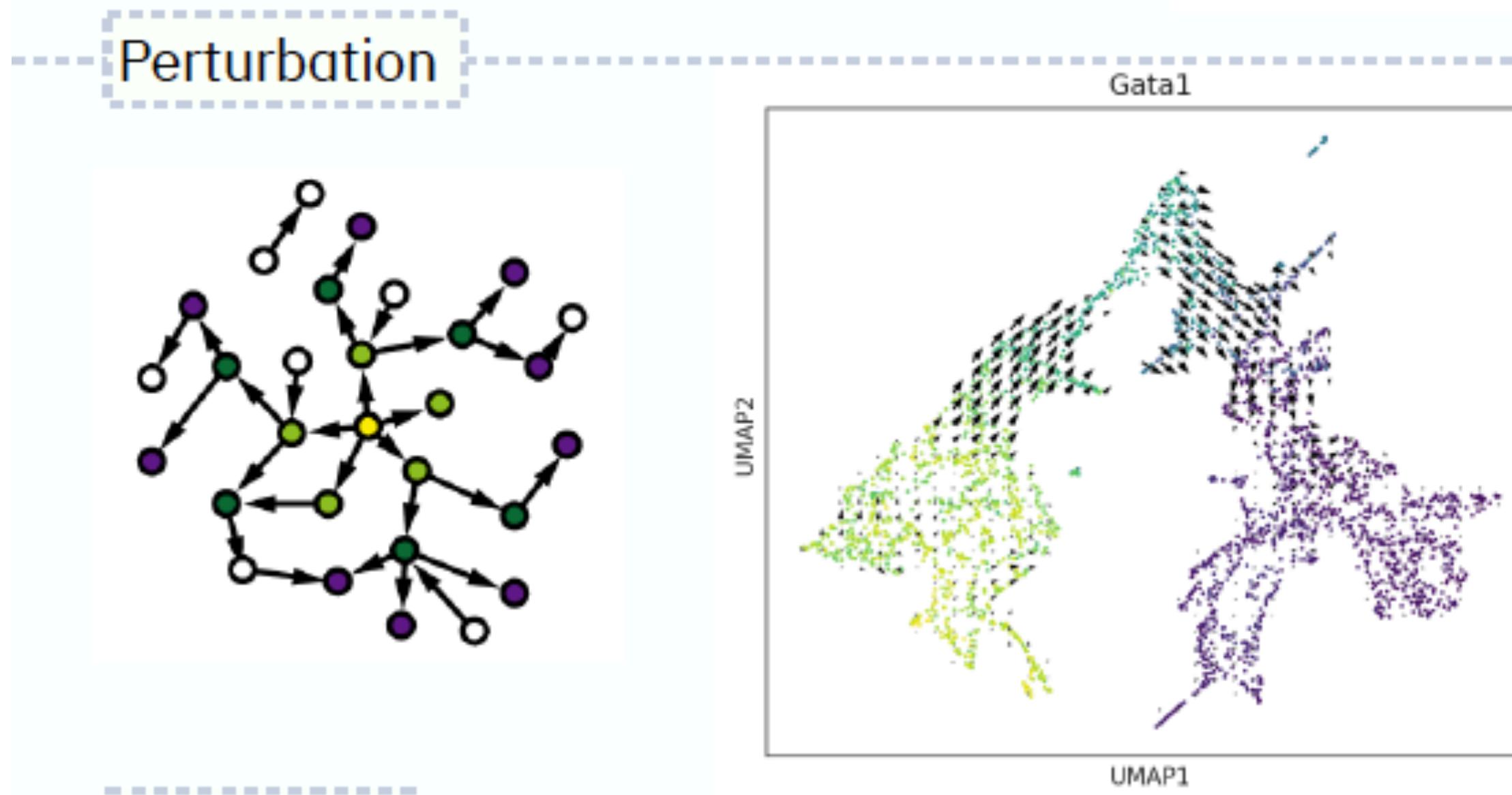
Simplifying assumptions:

- Context specific instead of universal rules
- Linearisation around attractors (cell types)
- Focus on prediction (e.g. in response to perturbation) rather than the true GRN

CellOracle (Kamimoto et al. Nature 2023)



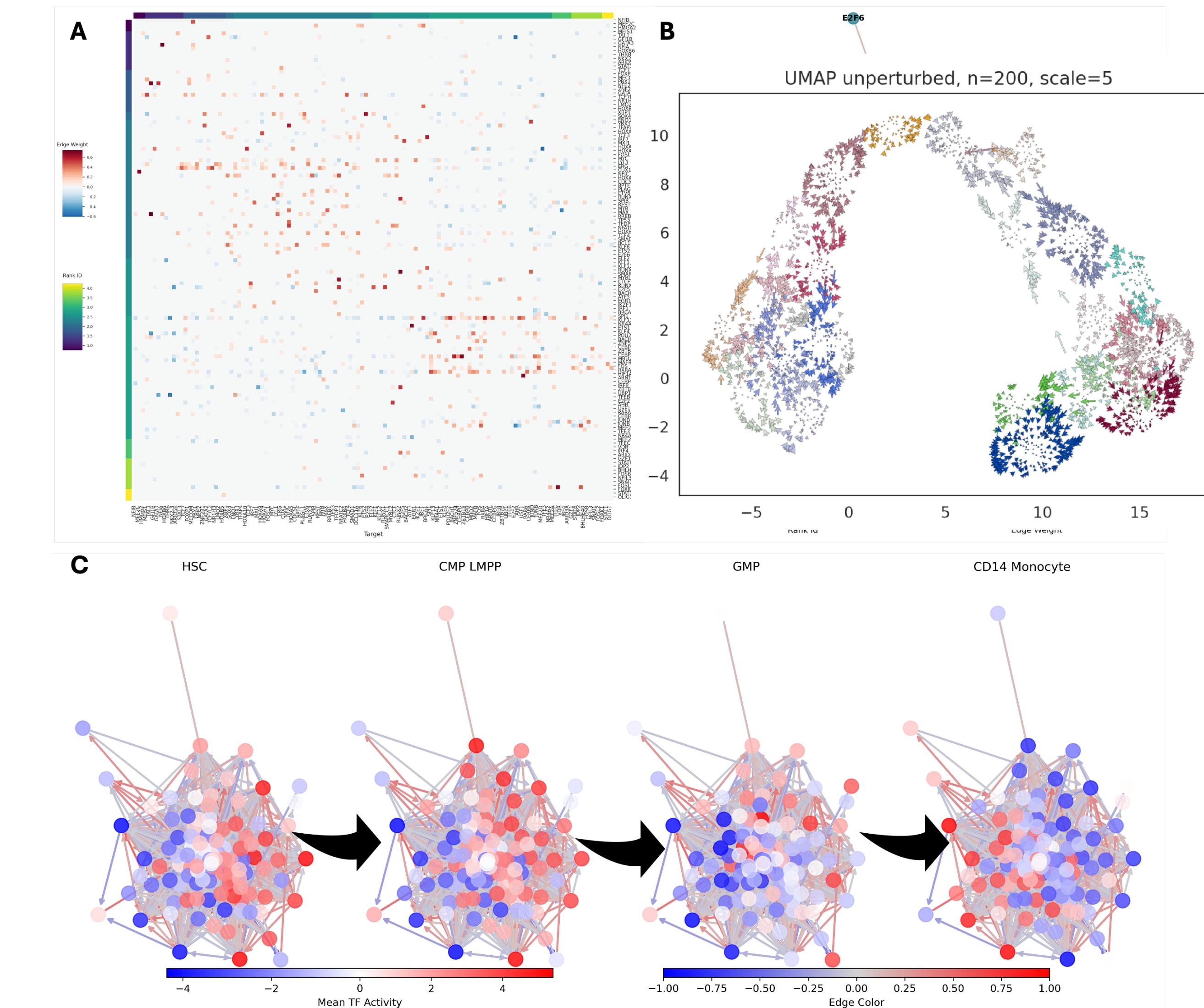
- Use chromatin accessibility + motif binding site information to identify potential regulators
- Focus on prediction
- But yield trivial predictions: flow in the direction of gradient of expression



Blood myeloid branch development GRN

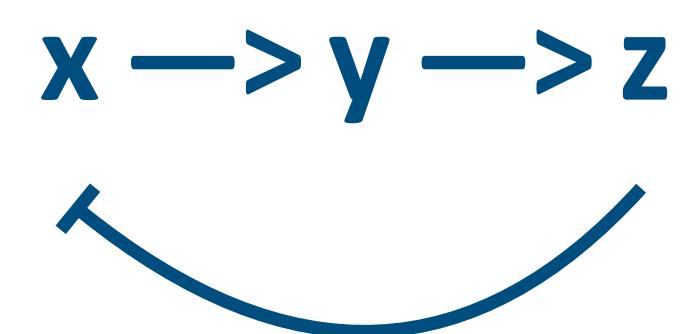
- Modified Cell Oracle (Kamimoto et al. 2023) pipeline
- Combine locally linear GRNs
- Focus on prediction (e.g. in response to perturbation) rather than the true GRN

$x \rightarrow y \rightarrow z$



Exercises:

- a) Analyse and simulate (using Beeline and BoolODE) the dynamics of the following GRN and show it can result in cell state transition.
- b) What attractor states does this GRN motif have?



GRNs inference from pseudo-time ordered cells (by Leap)

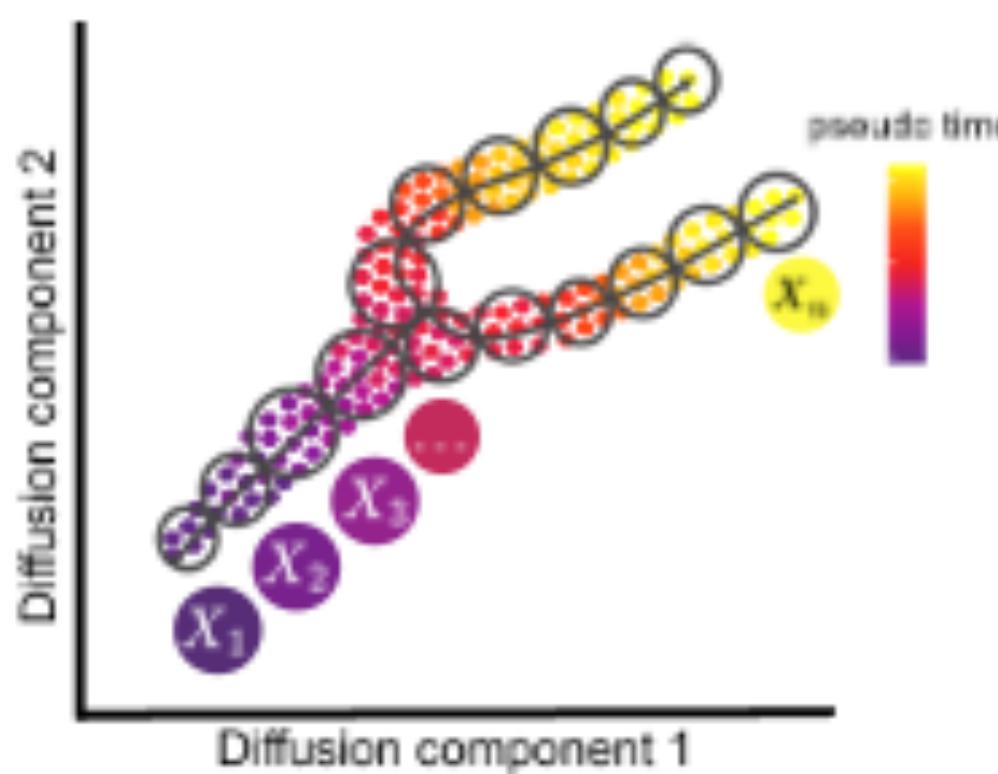
Dai, Hao, et al. "Reconstructing gene regulatory networks in single-cell transcriptomic data analysis." *Zoological research* 2017

- Considers correlations over different time lags
- Not only correlation but causal
- Cannot say anything about auto-regulation
- Rarely misses a regulator but too many false positives

Exercises:

- a) Why is the solution to the below pseudo-temporal model not unique?
- b) How does non-uniqueness (unidentifiability) affect our access to the true GRN?
- c) How does it affect our power of prediction?
- d) Linearisation around an attractors is theoretically justified. But how good/bad is assumption of a linear model over a trajectory?
- e) Can you solve for a W such that $X_t W^n = X_{t+n}$? Is this solution unique? Is the predictions from this model as trivial as CellOracle's or more interesting?

A linear regression model to find the $[G,G]$ regulator matrix (W) which brings every pseudo-cell to the next pseudo-cell over the pseudo-time order



$$\begin{matrix} \text{pseudo cell} \\ \text{transcription factor} \end{matrix} \times \begin{matrix} \text{regulatory transcription factor} \\ \text{target transcription factor} \end{matrix} = \begin{matrix} \text{pseudo cell} \\ \text{transcription factor} \end{matrix}$$

The diagram illustrates a linear regression model for gene expression dynamics. It shows a sequence of pseudo-cells $X_1, X_2, X_3, \dots, X_n$ plotted against diffusion components 1 and 2. A color scale indicates the progression of pseudo-time. The model is represented by the equation $X \times W = \tilde{X}$, where X is the matrix of transcription factors at each pseudo-cell, W is the regulatory transcription factor matrix, and \tilde{X} is the target transcription factor matrix.

Exercises:

- a) Can you suggest another simple model with either improved ground truth capture? Or improved prediction power?
- b) Any idea for a nonlinear but still simple model?

Some references:

- Strogatz, S.H., 2018. Nonlinear dynamics and chaos with student solutions manual: With applications to physics, biology, chemistry, and engineering. CRC press.
- Alon, U., 2019. An introduction to systems biology: design principles of biological circuits. Chapman and Hall/CRC.
- CellOracle: Kamimoto, Kenji, et al. "Dissecting cell identity via network inference and in silico gene perturbation." *Nature* 614.7949 (2023): 742-751.

**Thank you for your
attention!**