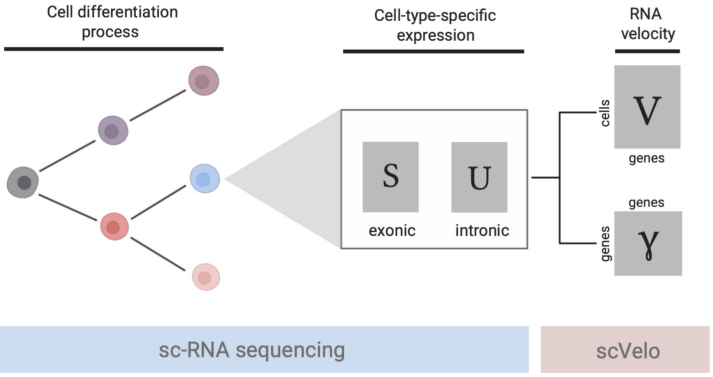
Neural Networks for Cellular Energy Landscapes from Single Cell Velocity Fields

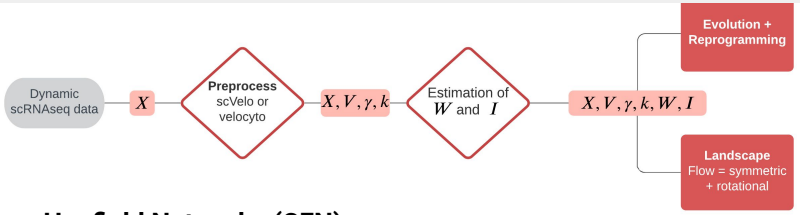
Juan P. Bernal-Tamayo[1,*],Larisa Morales-Soto[1],Robert Lehmann[1],Subash Balsamy[1],Xabier Martinez-de-Morentin[2],Amaia Vilas-Zornoza[3],Patxi San-Martin[3], Felipe Prosper[4],David Gomez-Cabrero[1,2,5],Narsis Kiani[6],Jesper Tegner[1,*]

Introduction

Recent development in the analysis of transcriptional regulation (Velocyto, scVelo) has made possible the estimation of RNA velocity and kinetic rates from scRNASeq data [2,3]. Yet, reconstructing the regulatory pathways underlying cellular processes remains a daunting challenge. Taking inspiration from Waddington, here we take an inverse approach and recover such networks from the cellular landscape they shape, using scRNASeq data.



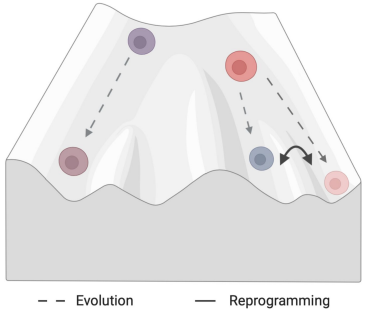
Methods



Continuous Hopfield Networks (CFN)

$$\dot{x}_i = \sum_{j=1}^{n} W_{ij}\varphi_j(x_j) - \gamma_i x_i + I_i$$

Here we model the dynamics of high-dimensional regulatory genomic circuits as a Continuous Hopfield Network [1].



Energy Function

$$E(X) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} s_i W_{ij} s_j + \sum_{i=1}^{n}\gamma_i \int_0^{s_i} \varphi_i^{-1}(\zeta)d\zeta - \sum_{i=1}^{n} s_i I_i$$

The CHN is endowed with an energy function for each state of the network, and, in the case W is a symmetric matrix, the stored states of the system correspond to local minima of this energy function.

[1] Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology, Thuwal, **Saudi Arabia**. [2] Mucosal and Salivary Biology Division, King's College London Dental Institute, London, **United Kingdom**. [3] Navarrabiomed, Complejo Hospitalario de Navarra, Universidad Pública de Navarra, Navarra, **Spain**. [4] Centro de Investigacion Medica Aplicada, Navarra, **Spain**. [5] Clínica Universidad de Navarra, Navarra, **Spain**. [6] Department of Oncology and pathology, Karolinska Institute, Stockholm, **Sweden**.
* Corresponding authors: jesper.tegner@kaust.edu.sa, juan.bernaltamayo@kaust.edu.sa

Explanation of the model

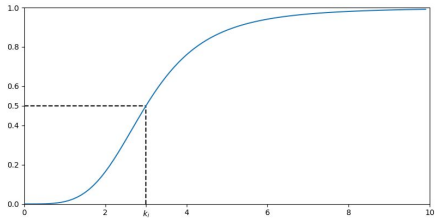
Hill functions

$$\varphi_i(x_i) = \frac{x_i^n}{x_i^n + k_i^n}$$

The gene dynamics is represented as a CHN, interacting via the *W* matrix, and using "biological" sigmoid Hill equations as transfer functions, modeling the switching between ON and OFF states. Where the dissociation rates k_i are inferred from the data as the steady state limit.

$$k_i = x_i^{\infty}/2$$

The exponent, *n*, controls the steepness of the change in the sigmoid function. Throughout this study a fixed exponent, *n=4*, is used.



Degradation rate and Steady-state limit

The degradation rates, γ_i , and steady state limit for each gene are obtained from the preprocessing of the data with scVelo.

Inference of W

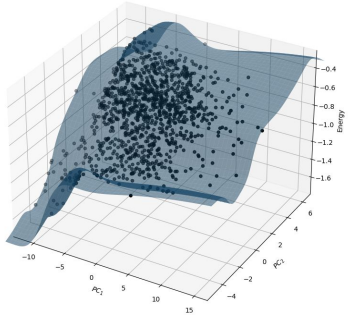
$$\dot{X} = \varphi(X)W - X\mathrm{diag}(\gamma) + I$$

$$\dot{X} + X\mathrm{diag}(\gamma) = [1|\varphi(X)]\begin{bmatrix} I \\ W \end{bmatrix}$$

$$\begin{bmatrix} I \\ W \end{bmatrix} = [1|\varphi(X)]^{+}\left(\dot{X} + X\mathrm{diag}(\gamma)\right)$$

The CHN model for the system can be written in a vectorial form, from here, the matrix, *W*, and the bias vector, *I*, can be approximated from the data taking the pseudoinverse of the augmented matrix of the ON/OFF gene states.

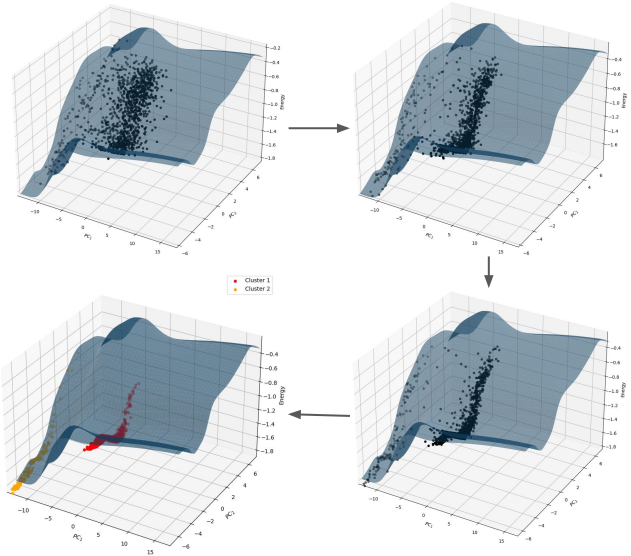
The following plots are generated using the Megakaryocyte-Erythroid Progenitor (MEP) cell cluster from an in-house generated hematopoietic differentiation data set.



Evolution

Asynchronous evolution

The system is evolved as a Neural Network in which the states are updated stepwise and asynchronously, *i.e.*, at each time-step only one gene is updated. In this scenario, at each time step the expression of every other gene remains constant, thus, the system, consisting of only the differential equation for the evolving gene, is solved numerically every time step.

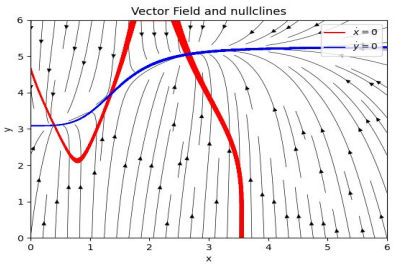


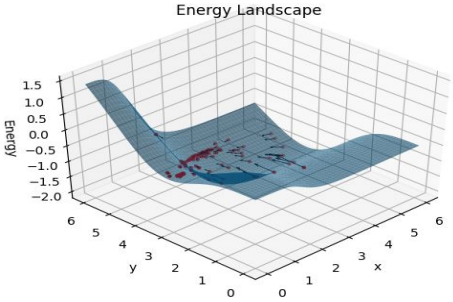
Decomposition of the Flux

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 & -1 \\ 6 & 1 \end{bmatrix} \begin{bmatrix} \varphi(x; k_x = 1.5) \\ \varphi(y; k_y = 4) \end{bmatrix} - \begin{bmatrix} x \\ 3y \end{bmatrix} + \begin{bmatrix} 0.65 \\ 9 \end{bmatrix}$$

Partial derivatives of the energy

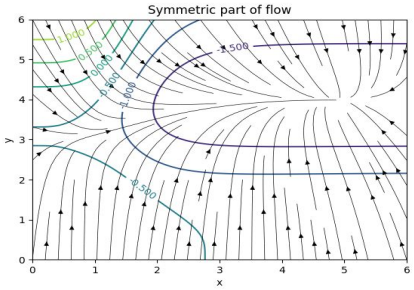
$$\frac{\partial E}{\partial x_i} = -\frac{d\varphi_i}{dx_i}\left(\sum_{j=1}^{n} W_{ij}^S s_j - \gamma_i x_i + I_i\right)$$

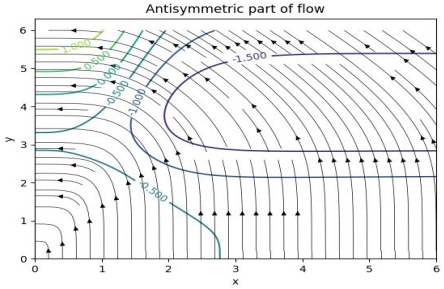




Symmetric-Antisymmetric decomposition

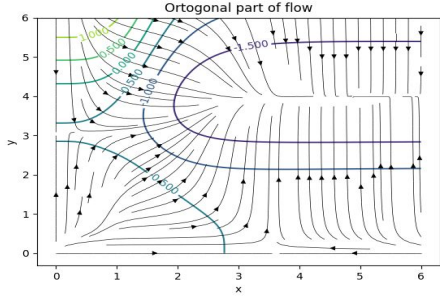
$$\dot{X} = F_S(X) + F_A(X) = \left(W^S\varphi(X) - \gamma X + I\right) + \left(W^A\varphi(X)\right)$$

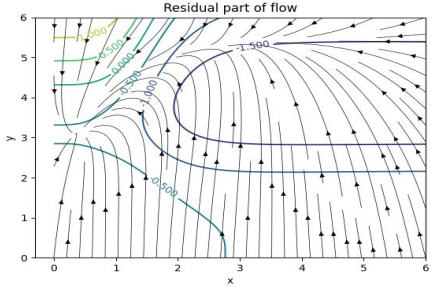




Orthogonal-Residual decomposition

$$\dot{X} = -\nabla E + F_R(X) = -\nabla E + \left(W^A\varphi(X) + (1 - \nabla\varphi(X))\left(W^S\varphi(X) - \gamma X + I\right)\right)$$





When the gradient of the energy function is taken, it can be noted that a factor of the partial derivatives corresponds to the symmetric part of the differential equation driving the system. When the matrix, *W*, is symmetric, the energy function assimilates a Lyapunov function for the system.

Inspired by this fact, two different decompositions of the flow driving the system are made. First, separating the system into the antisymmetric part of *W*, and the symmetric part. Second, separating the system into the gradient of the energy function, and the residual part.

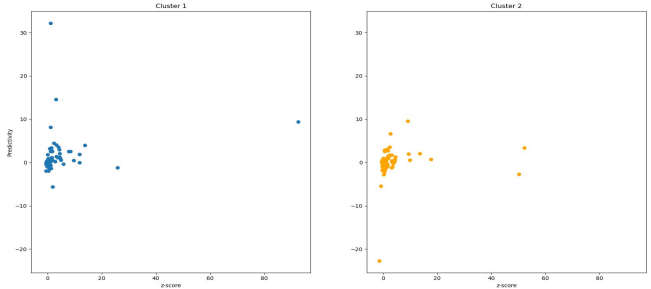
The symmetric/orthogonal part drives the system to the local minima of the energy, while the antisymmetric/residual part drives the system out of them.

Predictivity and Reprogramming

$$\eta_i^\mu = \sum_{\nu=1}^{m} (A^{-1})^{\mu\nu} X_i^\nu$$

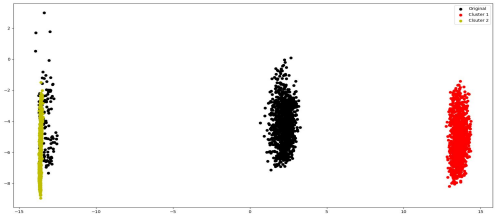
$$A = \mathrm{cov}(C_1, C_2)$$

Following the method by [4] to calculate the predictivity of a given cell fate, first the covariance matrix of the fates is calculated, and then used to calculate the corresponding predictivities. η_i^μ, corresponds to the predictivity of gene *i* over cell fate μ.



Then, among the genes with a high Z-score we picked the top 10 with the highest predictivities for each fate. As shown in the equation below the biases for the selected genes are modified and the system is evolved from the beginning with the corresponding modified biases.

$$I_i^\mu = \left| \gamma_i X_i^\mu - \sum_{j=1}^{n} W_{ij} \varphi_j(X_j^\mu) \right|$$



The projection over the first two PCs of the cells evolved with the modified biases can be observed. In the table the top 10 most predictive genes for each cell fate are shown.

Conclusions

A generalized framework for the construction of the energy field of a high dimensional biological system without having governing equations of the system has been formulated in this paper. Moreover, our results contribute to a deeper understanding of non-gradient cellular dynamical energy fields of biological systems and provide a way for steering a biological system towards the desired direction.

References

1. Hopfield, John J. “Neural Networks and Physical Systems with Emergent Collective Computational Abilities.” *Feynman and Computation*, pp. 7–19., doi:10.1201/9780429500459-2
2. Volker Bergen et al. “Generalizing RNA velocity to transient cell states through dynamical modeling”. Unpublished Work. 2019. URL: https://dx.doi.org/10.1101/820936.
3. Gioele La Manno et al. “RNA velocity of single cells”. In: Nature 560.7719 (2018), pp. 494–498. ISSN: 0028-0836. DOI: 10.1038/s41586-018-0414-6. URL: https://dx.doi.org/10.1038/s41586-018-0414-6.
4. Alex H. Lang et al. “Epigenetic Landscapes Explain Partially Reprogrammed Cells and Identify Key Reprogramming Genes”. In: PLoS Computational Biology 10.8 (2014), e1003734. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1003734. URL: https://dx.doi.org/10.1371/journal.pcbi. 1003734.
5. Sai Teja Pusuluri et al. “Cellular reprogramming dynamics follow a simple 1D reaction coordinate”. In: Physical Biology 15.1 (2017), p. 016001. ISSN: 1478-3975. DOI: 10.1088/1478-3975/aa90e0. URL: https://dx.doi.org/10.1088/1478-3975/aa90e0.

Cluster 1	**HSP90AB1**	**TOP2B**	**PLAUR**	**PSME2**	LY86	**USP1**	**EZH2**	**RORA**	**IKZF2**	**LST1**
Cluster 2	HSP90AB1	**LY86**	**TOP2B**	**RNF152**	**OLFM3**	**SLBP**	**MED12L**	PLAUR	**LIF**	**HBEGF**