Transiently increased intercommunity regulation characterizes concerted cell phenotypic transition

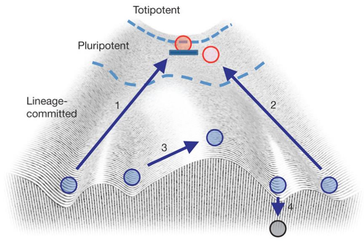
Weikang Wang1*, Dante Poe1,2, Ke Ni1,2, Jianhua Xing1,3,4,*
1 Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, PA 15232, USA.
2Joint CMU-Pitt Ph.D. Program in Computational Biology, University of Pittsburgh, Pittsburgh, PA, USA.
3 Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, PA 15232, USA.
4 UPMC-Hillman Cancer Center, University of Pittsburgh, Pittsburgh, PA, USA.

Abstract

Cell phenotype transition (CPT) takes place in many biological processes. A fundamental question is how cells coordinate switching of expressions of genes. In this work, we integrated dynamical systems theories and single cell RNA sequencing analyses. We found that the effective number of regulation between different communities in gene regulation network (GRN) increase first and then decrease, accompanying with similar change of the GRN frustration.

Introduction

Studying dynamics of cell phenotypic transition emerges as a new field. In recent years, hundreds of papers published in this field especially single cell RNA sequencing (scRNA-seq).

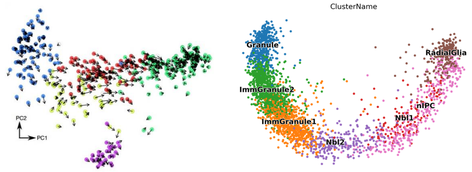


But scRNA-seq data analyses are typically statistics-based approaches with limited mechanistic insights on cellular dynamics. How do we study the dynamics equation of CPT from it?

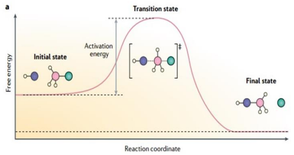
$$d\mathbf{x}/dt = \mathbf{F}(\mathbf{x}) + \eta(t)$$

RNA velocity analysis

To obtain the dynamics information, we adopt the method of RNA velocity. RNA velocity is a high-dimensional vector that predicts the future state of individual cells on a timescale of hours (Left figure). RNA velocity can be inferred from the quantity of spliced and unspliced RNA. With RNA velocity, we can obtain transition graph of CPT (development of the granule cell lineage in dentate gyrus)

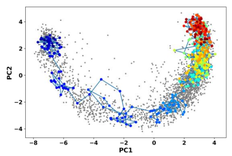


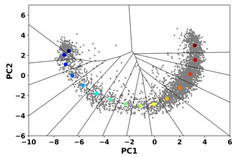
Reaction coordinates



Reaction coordinates (RC) is an abstract one-dimensional coordinate which represents progress along a reaction pathway.

Here we simulate the trajectories that originate from the initial state (Radial glia-like cells) to the final state (Granule cells) in this transition graph. We also calculate the reaction coordinate in the gene expression space.





GRN inference

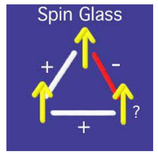
Now we obtained dx/dt and x from the single cell data. We want to explore the properties of F. From the biological perspective, F is the gene regulation network(GRN). We use partial linear square regression (PLSR) to infer F and study its properties along the RC.

Measured/derived from single cell data

$$\frac{d\mathbf{x}}{dt} = \mathbf{F} \cdot \mathbf{x} + \eta(t)$$

Constant matrix learned from **x** and dx/dt with partial linear square regression (PLSR) and local false discover rate (LFDR)

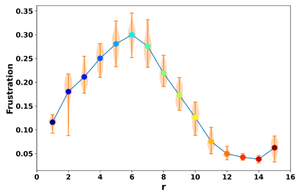
Frustration



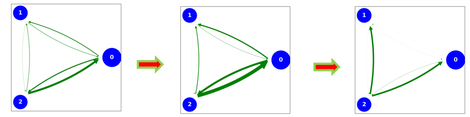
$$x_i \to s_i = \begin{cases} 0, & \text{gene off} \\ 1, & \text{gene on} \end{cases}$$

$$\mathrm{sgn}((2s_i - 1)F_{ij})s_j = \begin{cases} 1, & \text{unfrustrated} \\ 0, & \text{no regulation} \\ -1, & \text{frustrated} \end{cases}$$

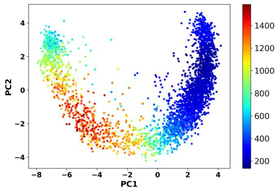
The frustration is that the spin value is against the interaction from the other. We binarized the gene expression and defined the frustration value of gene regulation. The uniqueness of gene regulation is that it is directional. If the regulation gene is off, its regulations on other genes are not frustrated or unfrustrated. The frustration score of GRN is the proportion of frustrated regulation to the total number of gene regulations. We found that during CPT, the frustration score increase first and then decrease.



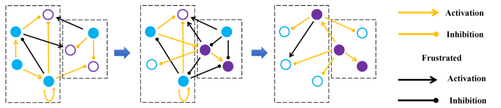
Inter-community interactions



We divide the GRN into several communities, represented by the blue nodes. We found that the number of active regulations (represented by the width of arrows) between the communities increase first and then decrease along RC.



In the scatter plot , we showed the number of inter-community edges in each cell .



The variation of GRN in CPT can be summarized with this plot. In the beginning and end of CPT, the GRN has high modularity, against global propagation of perturbations. In the transition state, the GRN has low modularity and high frustration, which reflect the coordinated reprogramming of gene expression profiles.

Acknowledgement

This work was partially supported by National Cancer Institute (R37 CA232209), and National Institute of Diabetes and Digestive and Kidney Diseases (R01DK119232) to JX.

* xing1@pitt.edu