Context-Specific EMT Regulatory Networks from Temporal Single Cell RNA-Seq Data

Daniel A. Ramirez,[1,3] Vivek Kohar[2], and Mingyang Lu[2,3]

***Short Abstract* — We adopt a combined bioinformatics and mathematical modeling approach to construct context-specific EMT GRCs directly from time series transcriptomics data for four cancer cell lines treated with three EMT-inducing signals. We observe distinct paths during the forward and backward transitions, as is evident from the dynamics of major regulators such as NF-KB and AP-1. For each experimental condition, we systematically sample a large set of network models and identify the optimal GRC capturing context-specific EMT states, uncovering the role of common bioinformatics parameters and properties of network structures in determining the quality of GRCs.**

***Keywords* — Gene regulatory network, single cell transcriptomics, systems biology**

I. Introduction

EPITHELIAL-MESENCHYMAL transition (EMT) has been implicated in a number of biological phenomena including embryonic development, wound healing, and cancer metastasis [1]. Although a core decision-making circuit for EMT has been uncovered [2], the specific gene regulatory interactions that occur in an EMT could vary for different cell types, signaling states, and disease states [3]. Some attempts have been made to construct EMT regulatory circuits from evidence in literature [4-5], but this approach may fall short in new or understudied contexts.

Here, by using time-series single-cell transcriptomic data [3] and mathematical modeling [5], we investigated EMT in 12 experimental contexts by constructing numerous candidate networks and comparing their simulated behavior with observed gene expression. This allowed us to identify gene regulatory circuits or each condition containing both conserved and context-specific elements.

II. Methods

We developed a computational platform to construct and model gene networks from scRNA-seq data. Using processed scRNA-seq data from [3], transcription factor activities were first inferred using SCENIC. For each of twelve conditions – four cell lines and three signals – differential transcription factor (TF) activity was measured between timepoints. TFs with high differential activity across a large segment of the experimental conditions were used as the basis of a conserved core circuit. From this core circuit, we identified putative TF-target relationships for each condition using RcisTarget and the mutual information (MI) between TF and target activities. Varying the mutual information cutoff results in many putative networks. For each network, we simulated the gene expression using a mathematical modeling method named *Ra*ndom *Ci*rcuit *Pe*rturbation [5]. The simulated data were then quantitatively compared to the experimental gene expression profiles to obtain optimized networks.

III. Results

We identified a great amount of heterogeneity in the EMT signal response, with typical master regulators like SNAIL and ZEB not being consistently identified in the motif analysis, suggesting a partial EMT. Among the identified differentially active TFs, we observed more similarity in the EMT response of the same cell line treated with different signals, than with the same signal across different cell lines. However, the roles of many highly conserved differentially active TFs are context dependent.

From the network optimization, we can evaluate the effect of network features on their accuracy. We observed that moderate-sized networks provide optimal accuracy over very small or large ones. We further performed dynamical simulations on the optimized networks, from which we identified distinct forward and reverse trajectories of EMT, consistent with the observations from the single cell data.

IV. Conclusion

Our bioinformatics and modeling approach allowed us to uncover distinct features of EMT in different contexts, as well as the forward and reverse transition trajectories. We expect the integration of top-down bioinformatics and bottom-up systems biology modeling to be a powerful and generally applicable approach to elucidate gene regulatory mechanisms of cellular state transitions.

References

[1] Thiery JP, et al. (2009). Epithelial-Mesenchymal Transitions in Development and Disease. *Cell*, **139**, 871–890.

[2] Lu M, et al. (2013) MicroRNA-based regulation of epithelial–hybrid–mesenchymal fate determination. *Proc. Natl. Acad. Sci. U.S.A*, **110**, 18144–18149.

[3] Cook DP, Vanderhyden BC (2020) Context specificity of the EMT transcriptional response. *Nat. Commun.,* **11**, 2142.

[4] Katebi A, Ramirez D, Lu M (2021) Computational systems-biology approaches for modeling gene networks driving epithelial–mesenchymal transitions. *Comput Syst Oncol.*, **1**.

[5] Huang B, et al. (2017) Interrogating the topological robustness of gene regulatory circuits by randomization. *PLoS Comput. Biol.* **13**, e1005456.

Acknowledgements: This work was funded by NIH grant R35GM128717.

[1]College of Health Solutions, Arizona State University, Tempe, AZ. E-mail: darami14@asu.edu

[2]The Jackson Laboratory for Mammalian Genetics, Bar Harbor, ME, United States.

[3]Department of Bioengineering, Northeastern University, Boston, MA. E-mail: m.lu@northeastern.edu

Nothing should be here on page 2! Please limit your abstract to a single page, and create a one-page .pdf file for submission.