Reverse-Engineering the Sea Urchin: Uncovering Network Mechanisms of Differentiation Through Single-Cell Transcriptomics

Shakson Isaac, Mahadevan Subramanian, and Mamiko Yajima

Department of Molecular Biology, Cell Biology & Biochemistry, Division of Biology and Medicine, Brown University, Providence, RI, USA

ABSTRACT

- **Aim 1**: Infer cell lineage specific networks detailing important drivers of differentiation.
- **Aim 2:** Use scRNA-seq to infer both temporal and spatial gene expression patterns through trajectory analysis.
- **Aim 3**: Develop new hypotheses that describe the differentiation process of each cell lineage and detail future experiments to process.

BACKGROUND

- **Network** modeling is useful for extracting complexities within expression data.
- Regulation in developmental biology is dependent on **temporal** and **spatial** gene expression patterns.
- Constructing **cell lineage specific** gene regulatory networks (**GRNs**) can further molecular understanding of **differentiation** and gain new functional insights into underlying **circuits.**

METHODOLOGY

- **Single cell RNA sequencing** (scRNA-seq) data for the sea urchin embryo was obtained from Foster et al., 2020 [1].
- scRNA-seq data included **eight time points** covering 8-cell to late gastrula stage.
- **Semi-supervised clustering** (scSorter)[2] was used to obtain cell lineages from marker gene expression.
- **Trajectory inference** (Dynverse)[3] was performed on each cell-lineage and important **bifurcation** points.
- **Differentially expressed genes** (DEGs) were characterized by:
  - Genes having differences in expression between **cell lineage clusters** (Seurat)[4].
  - Genes whose expression significantly changing across **pseudotime** (GAM)[5].
  - Genes with differences in expression patterns over pseudotime between lineages (TradeSeq)[6].
  - Genes with difference in expression between lineages at early decision points (TradeSeq)[6].
- Set of all DEGs were used for **network inference** (BEELINE)[7].
  - Network methods consisted of using **normalized expression data** and potential uses of **pseudotime**.

RESULTS

Case 1: Primary Mesenchyme Cell Lineage (PMC)

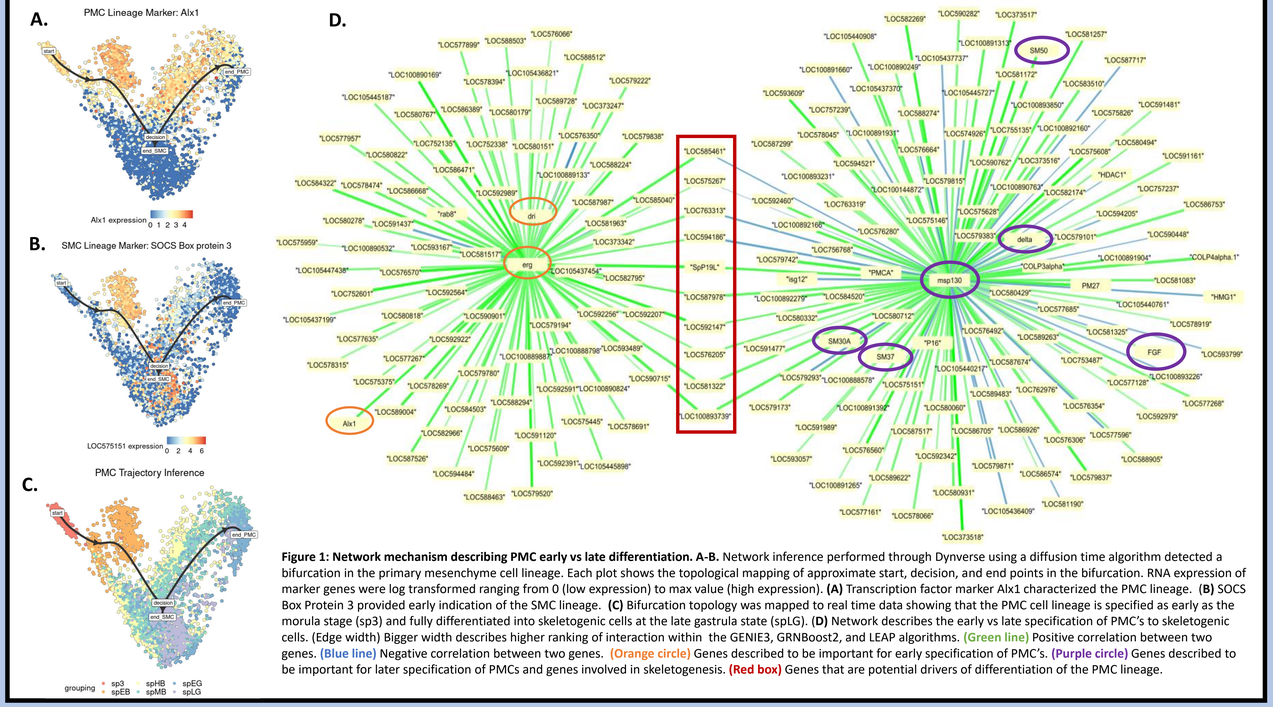

Figure 1: Network mechanism describing PMC early vs late differentiation. A-B. Network inference performed through Dynverse using a diffusion time algorithm detected a bifurcation in the primary mesenchyme cell lineage. Each plot shows the topological mapping of approximate start, decision, and end points in the bifurcation. RNA expression of marker genes were log transformed ranging from 0 (low expression) to max value (high expression). **(A)** Transcription factor marker Alx1 characterized the PMC lineage. **(B)** SOCS Box Protein 3 provided early indication of the SMC lineage. **(C)** Bifurcation topology was mapped to real time data showing that the PMC cell lineage is specified as early as the morula stage (sp3) and fully differentiated into skeletogenic cells at the late gastrula state (spLG). **(D)** Network describes the early vs late specification of PMC's to skeletogenic cells. (Edge width) Bigger width describes higher ranking of interaction within the GENIE3, GRNBoost2, and LEAP algorithms. (Green line) Positive correlation between two genes. (Blue line) Negative correlation between two genes. (Orange circle) Genes described to be important for early specification of PMC's. (Purple circle) Genes described to be important for later specification of PMCs and genes involved in skeletogenesis. (Red box) Genes that are potential drivers of differentiation of the PMC lineage.

Case 2: Secondary Mesenchyme Cell Lineage (SMC)

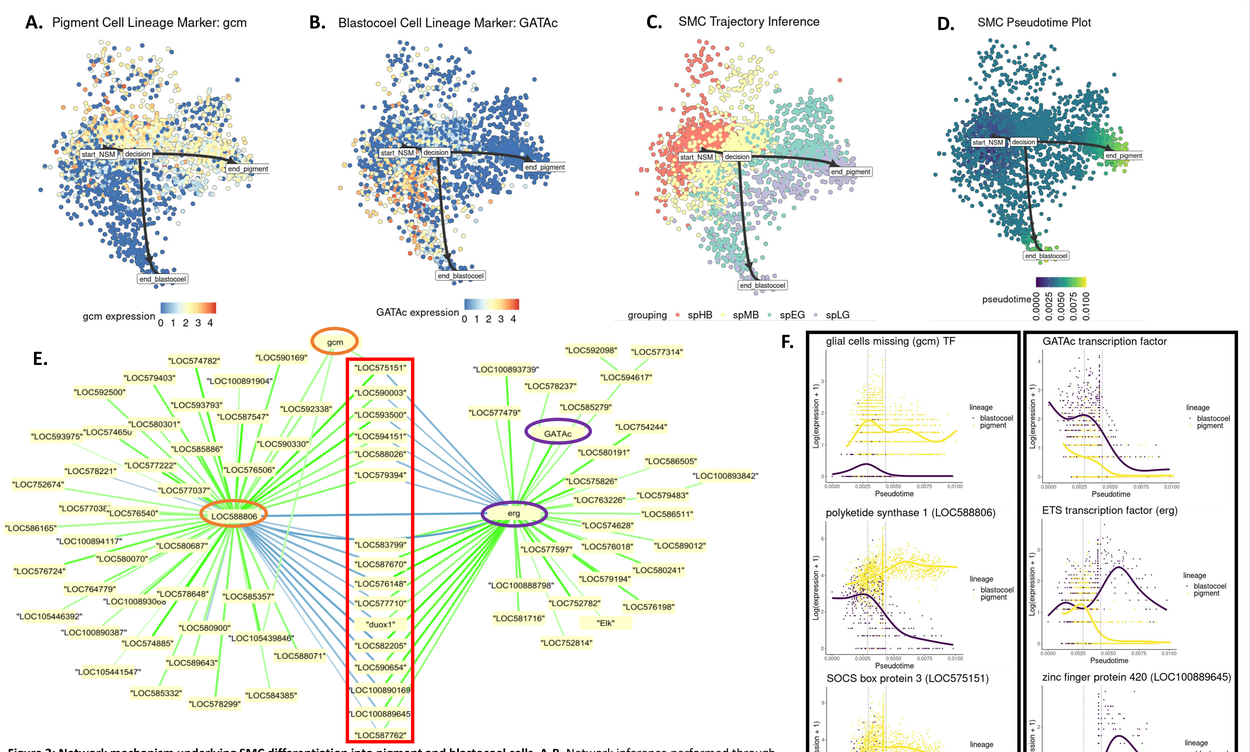

Figure 2: Network mechanism underlying SMC differentiation into pigment and blastocoel cells. A-B. Network inference performed through Dynverse using a diffusion time algorithm detected a bifurcation in the secondary mesenchyme cell lineage. Each plot shows the topological mapping of approximate start, decision, and end points in the bifurcation. RNA expression of marker genes were log transformed ranging from 0 (low expression) to max value (high expression). **(A)** Transcription factor marker glial cell missing characterized the pigment cell branch. **(B)** Transcription factor marker GATAc characterized the blastocoel cell branch. **C-D.** Bifurcation topology was mapped to temporal indications of both real-time and pseudotime. Real time values showed that pseudotime accurately reflects the real-time course from the hatched blastula (HB) to late gastrula (LG) stage. Additionally, the decision point of the bifurcation is within the mesenchyme blastula stage aligning with other records of the differentiation of pigment and blastocoel lineages. **(E)** Network describes the differentiation of pigment vs blastocoel cells. (Edge width) Bigger width describes higher ranking of interaction within the GENIE3, GRNBoost2, and LEAP algorithms. (Green line) Positive correlation between two genes. (Blue line) Negative correlation between two genes. (Orange circle) Genes described to be important for pigment cells in literature. (Purple circle) Genes described to be important for blastocoel cells in literature. (Red box) Genes that are potential drivers of differentiation that are regulated by transcription factors erg (blastocoel cell) and gcm (pigment cell). **(F)** Dynamic expression patterns of genes involved in pigment and blastocoel cells. In between the vertical dashed lines indicate decision regions. (Left box) Gcm, pks1, and SOCS box protein 3 have high expression in the pigment cell lineage and are marked to be positively correlated to one another. (Right box) GATAc, erg, and ZNF420 have high expression in the blastocoel cell lineage and have positive correlation with one another. Between both cell lineage networks the two example genes, SOCS box protein 3 and ZNF420 are potential genes that may regulate the differentiation circuit in SMC's.

DISCUSSION

- Analyzing networks of differentiation in the SMC and PMC cell lineage through scRNA-seq data and methods involving trajectory and network inference are possible.
  - The usage of these methods invoked new insights into **mechanisms of differentiation** and contained known interactions described in previous literature.
- **Genes not described in literature** and regulated by known transcription factors can be experimentally verified to provide major **insight into regulatory circuits**.
  - Such genes are highlighted in Figure 1 and 2.
- The GRN's created reflect correlation between genes based on temporal and spatial expression data. These networks contain a mixture of **direct and indirect interactions**. However, usage of **prior knowledge** (Davidson GRN) can highlight these differences.

FUTURE DIRECTIONS

- Perform trajectory and network analysis for **additional cell lineages** found in the scRNA-seq dataset.
- **Experimentally verify new interactions** through whole-mount in-situ hybridization, qPCR, and knockdown/inhibitor experiments.
- Expand networks to include **proteomics and metabolomics** data to describe systems level information of the differentiation process.
- Understand differentiation of cell fate in **other echinoderms** at the **single cell level**.

ACKNOWLEDGEMENTS

Thank you to the Brown University Karen T. Romer UTRA and Brown University RAB grant for supporting this project, Professor Mamiko Yajima and Yajima Lab members for continued support and insight, Professor Gary Wessel and Stephany Foster for providing initial access to the scRNA-seq data, Dr. Ashok Ragavendran (CBC) for their additional insights and feedback on computational resources and analysis, and CCV for their support in processing data through the OSCAR supercomputer.

REFERENCES

1. Foster, S., Oulhen, N., & Wessel, G. (2020). A single cell RNA sequencing resource for early sea urchin development. *Development (Cambridge, England)*, *147*(17), dev191528. https://doi.org/10.1242/dev.191528
2. Guo, H., & Li, J. (2021). scSorter: assigning cells to known cell types according to marker genes. *Genome biology*, *22*(1), 69. https://doi.org/10.1186/s13059-021-02281-7
3. Saelens, W., Cannoodt, R., Todorov, H., & Saeys, Y. (2019). A comparison of single-cell trajectory inference methods. *Nature biotechnology*, *37*(5), 547–554. https://doi.org/10.1038/s41587-019-0071-9
4. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., 3rd, Zheng, S., Butler, A., Lee, M. J., Wilk, A. J., Darby, C., Zager, M., Hoffman, P., Stoeckius, M., Papalexi, E., Mimitou, E. P., Jain, J., Srivastava, A., Stuart, T., Fleming, L. M., Yeung, B., Rogers, A. J., ... Satija, R. (2021). Integrated analysis of multimodal single-cell data. *Cell*, *184*(13), 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048
5. Trevor Hastie (2020). gam: Generalized Additive Models. R package version 1.20. https://CRAN.R-project.org/package=gam
6. Van den Berge, K., Roux de Bézieux, H., Street, K., Saelens, W., Cannoodt, R., Saeys, Y., Dudoit, S., & Clement, L. (2020). Trajectory-based differential expression analysis for single-cell sequencing data. *Nature communications*, *11*(1), 1201. https://doi.org/10.1038/s41467-020-14766-3
7. Pratapa, A., Jalihal, A. P., Law, J. N., Bharadwaj, A., & Murali, T. M. (2020). Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nature methods*, *17*(2), 147–154. https://doi.org/10.1038/s41592-019-0690-6