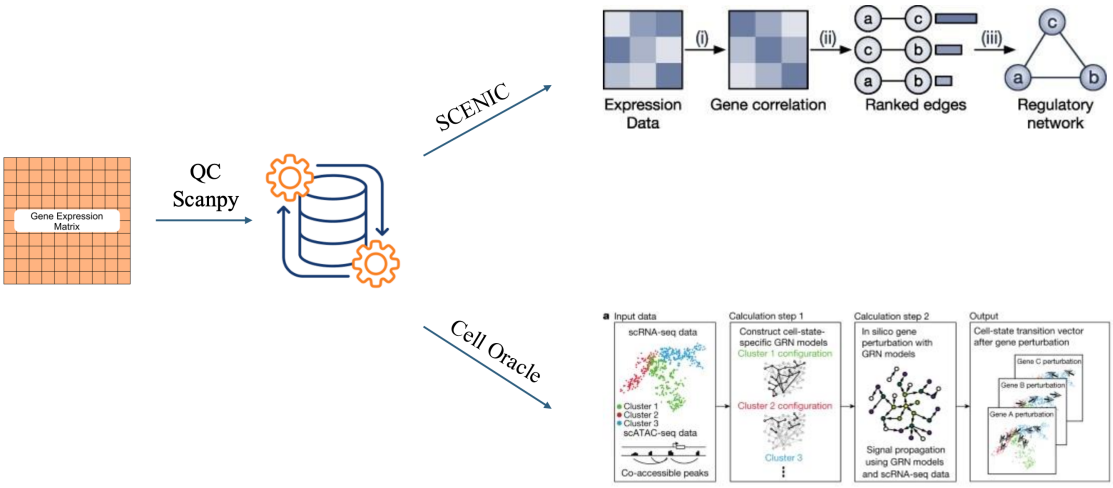
Inferring Gene Regulatory Networks of Single Cell Data using SCENIC and CellOracle

Rika Chan

December 2024

Abstract

Characterizing progenitors of Brown Adipose Tissue (BAT) is crucial to studying BAT development. SCENIC and CellOracle were used to inferred gene regulatory networks (GRNs), and identify possible BAT progenitor cell types in single cell data of mouse tissue taken at E12.5. Cxcl14 was identified as a marker for a potential BAT progenitor cell population.



Nguyen et al., 2020
Kamimoto et al., 2023

1 Introduction

Brown Adipose Tissue (BAT) is a thermogenesis, producing heat and increasing metabolic rate (Marlatt et al., 2017). BAT originates from the dermomyotome compartment of somites, whose tripotency also allows differentiation into skeletal muscles and dermal mesenchymes. Being a source of metabolic-related functions, BAT as a therapeutic treatment for obesity and other metabolic disorders is an active area of research. There are three major BAT depots in mice: interscapular (iBAT),

scapular (sBAT), and cervical (cBAT).

Identifying key genes and signaling pathways involved in lineage commitment and differentiation into brown adipocytes are key to understanding embryonic development of BAT and its potential in clinical therapy for metabolic-related disorders. However, BAT progenitors and its lineage commitment is still poorly characterized. Here, I ran SCENIC to determine relationship between transcription factors and target genes (regulons) on single cell expression data (scRNA-seq) from mouse tissue E12.5. SCENIC used a tree-based algorithm to infer connectivity between genes, allowing the model to capture interactions not just between two but across multiple genes. I also ran CellOracle, a regression-based network inference to determine cluster-specific gene regulatory networks and perform network analyses. From the two methods, I identified marker gene Cxcl14 as a marker for a potential brown adipocytes cell population.

2 Methods

Dataset and Data Pre-processing. Single cell expression data (scRNA-seq) was obtained from the lab of Professor Jennifer Mansfield of Barnard College. Quality controls were done using the python package Scanpy in order to remove genes with low counts, mitochondrial genes, and doublet cells. The filtered dataset contained 6492 genes and 4779 cells.

Gene Regulatory Network (GRN) Algorithm. SCENIC was used to implement a gene correlation algorithm. The program was chosen with the guidance of a survey of current GRN methods used for scRNA-seq (Nguyen et al., 2021). The command was run in its suggested default parameters with GRNboost2 method and 10 numworkers (Van et al., 2020). Outputs were compiled into a loom file, and results were analyzed in R (version 3.20) using SCopeLoomR. The same filtered data was used with CellOracle to determine cluster-specific GRNs, and ran using the default parameters (Kamimoto et al., 2023).

3 Results

3.1 GRNs construction with SCENIC

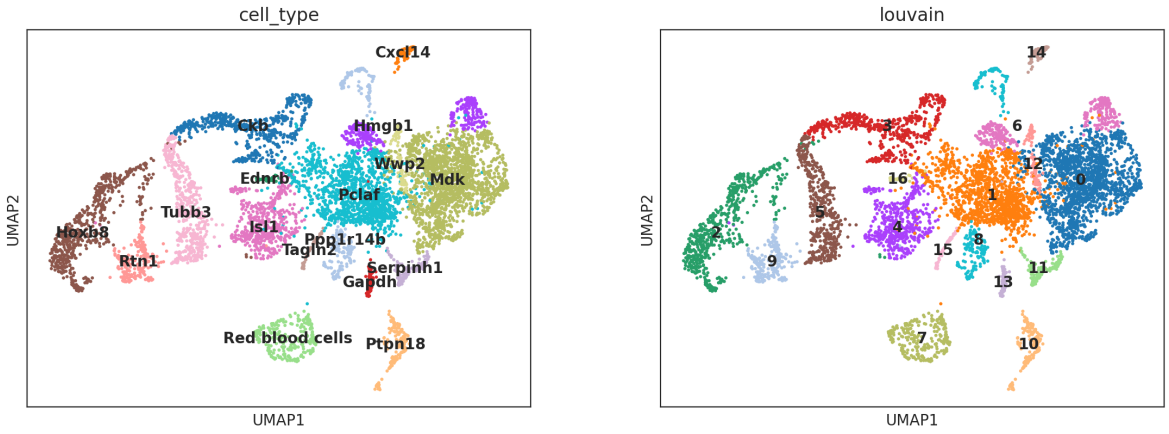


Figure 1: **UMAP.** UMAP clustering using louvain resolution of 0.35. (A) UMAP with marker genes labeled. (B) UMAP with cluster numbers.

16 clusters were identified using louvain with 0.35 resolution. Clusters were annotated with the top differentially expressed gene in that cluster due to largely unknown functions of some of the marker genes. Although cell type identification was limited, cluster 7 was able to be discerned as red blood cells due to high expression of *Hba-a1*, *Hba-a2*, and *Hbb-y* (data not shown). Cluster 14, marked by *Cxcl14*, was chosen to explore further due to *Cxcl14*'s function as a BAT progenitor (Maniyadath et al., 2023).

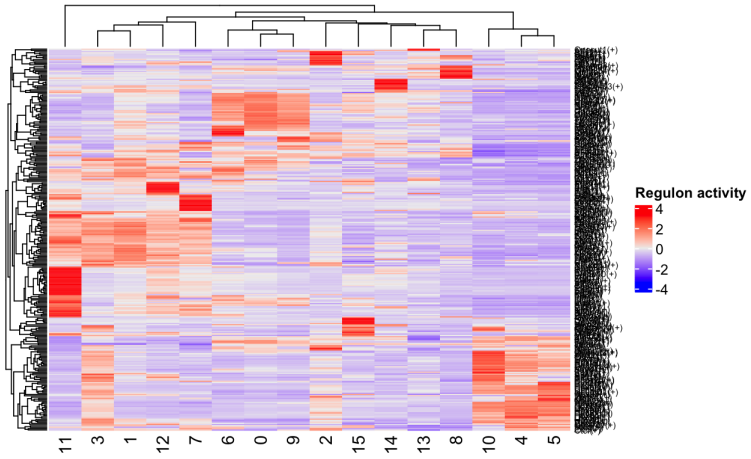


Figure 2: **Regulon Activity.** Heatmap showing activity of regulons across clusters from SCENIC. AUC score was used as measurement of activity of cells. AUC scores indicate relative expression of the genes in regulons in each cells.

Though unspecific to a particular gene or pathway, the heatmap generally shows the activity of

regulons across cell types. For instance, cluster 11 (*Serpinh1*) showed a high regulon activity. Cluster 14 (*Cxcl14*) showed relatively less regulon activity, but remain crucial to explore. Although data not shown, *Barx2*(+) was identified as the regulon with the highest RSS score in cluster 14 (a cluster-specific measurement of regulon activity proposed by Suo et al. for the Mouse Cell Atlas in 2018).

3.2 GRNs construnction with CellOracle

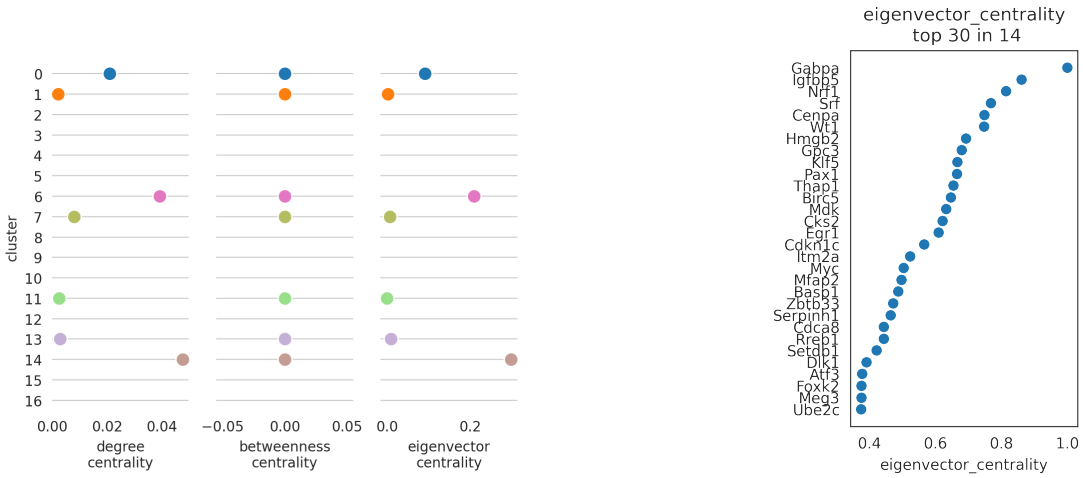


Figure 3: **Network scores of *Cxcl14* and cluster 14.** (A) Degree centrality is defined as the number of ties linked to a node. Betweenness centrality refers to the shortest path between all pairs of vertices. Eigenvector centrality refers to the influence a node has on the network. (B) Top 30 genes with the highest Eigenvector centrality score in cluster 14.

Here, *Cxcl14* showed high network score in cluster 14 as expected (fig. 4A); however, it did not show up as the top 30 genes with the highest eigenvector centrality score in cluster 14. Although data not shown, *Cxcl14* also did not show up in the top 30 genes with the highest degree centrality and betweenness centrality, despite being the marker gene for that cluster. Although this does not delegitimize *Cxcl14* role in cluster 14, it suggests that all three metrics are essential in assessing importance of a gene in a network as well as assessing the network from the gene and the cluster perspectives.

4 Discussion

SCENIC and CellOracle are powerful tools to construct GRNs and perform network analyses. The main limitation is that they are unable to infer whether it is an activation or repression regulation. However, the results provide pathways in which future experiments could be followed.

The lab is currently interested in characterizing mechanisms around BAT lineage commitment by looking at scRNA-seq to identify possible BAT progenitors. From this preliminary result, cluster 14,

marked by *Cxcl14* gene, is proposed to be a possible BAT progenitor, mainly because of *Cxcl14* being a known BAT progenitor and presence of other important genes involved in brown adipocytes differentation and adipogensis such as *Pax1* and *Gabpα* (figs. 3B). *Gabpα* along with phospho-*ERK* have been shown to regulate mitochondrial development in brown adipocytes (Kato et al., 2020). In addition, *Barx2*(+) regulon, which was identify as a regulon with the highest activity score in cluster 14, is known to inhibit Wnt/β-catenin pathway (Mi et al., 2016). This is relevant as Wnt/β-catenin was shown to inhibit brown adipogenesis (Chen et al. 2023).

Inferring gene regulatory networks allow for the lab to continue exploring avenues in which BAT progenitors are expressed. From this analysis, genes with high network scores can be explored further experimentally to confirm its importance in BAT lineage commitment. Clusters with a high regulon activity could also be explored to determine the differentiation trajectory in that particular embryonic stage (fig. 2). A notable limitation to this report is the lack of pseudotime analysis. In silico perturbation of *Cxcl14* is possible with CellOracle; however, without pseudotime analysis the alternative differentiation trajectory with *Cxcl14* knockout does not yield meaningful comparison.

5 References

1. Chen, N., Yuan, M., Zhang, N., Chen, M., Liu, R., Wang, J., Lu, P. (2023). Ctnnb1/-catenin inactivation in ucp1-positive adipocytes augments the browning of white adipose tissue. iScience, 26(5), 106552. https://doi.org/10.1016/j.isci.2023.106552

2. Kamimoto, K., Stringa, B., Hoffmann, C.M. et al. Dissecting cell identity via network inference and in silico gene perturbation. Nature 614, 742–751 (2023). https://doi.org/10.1038/s41586-022-05688-9

3. Kato, H., Okabe, K., Miyake, M., Hattori, K., Fukaya, T., Tanimoto, K., Beini, S., Mizuguchi, M., Torii, S., Arakawa, S., Ono, M., Saito, Y., Sugiyama, T., Funatsu, T., Sato, K., Shimizu, S., Oyadomari, S., Ichijo, H., Kadowaki, H., Nishitoh, H. (2020). ER-resident sensor PERK is essential for mitochondrial thermogenesis in brown adipose tissue. Life Science Alliance, 3(3), e201900576. https://doi.org/10.26508/lsa.201900576

4. Maniyadath, B., Zhang, Q., Gupta, R. K., Mandrup, S. (2023). Adipose tissue at single-cell resolution. Cell Metabolism, 35(3), 386–413. https://doi.org/10.1016/j.cmet.2023.02.002

5. Marlatt, K. L., Ravussin, E. (2017). Brown Adipose Tissue: an Update on Recent Findings. Current obesity reports, 6(4), 389–396. https://doi.org/10.1007/s13679-017-0283-6

6. Mi Y., Zhao S., Zhou C., Weng J., Li J., Wang Z., Sun H., Tang H., Zhang X., Sun X., Peng Z., Wen Y. Downregulation of homeobox gene Barx2 increases gastric cancer proliferation and metastasis and predicts poor patient outcomes. Oncotarget. 2016; 7: 60593-60608. Retrieved from https://www.oncotarget.com/article/11260/text/

7. Nguyen H, Tran D, Tran B, Pehlivan B, Nguyen T. A comprehensive survey of regulatory network inference methods using single cell RNA sequencing data. Brief Bioinform. 2021 May 20;22(3):bbaa190. doi: 10.1093/bib/bbaa190. PMID: 34020546; PMCID: PMC8138892.

8. Rao J, Djeffal Y, Chal J, Marchianò F, Wang CH, Al Tanoury Z, Gapon S, Mayeuf-Louchart A, Glass I, Sefton EM, Habermann B, Kardon G, Watt FM, Tseng YH, Pourquié O. Reconstructing human brown fat developmental trajectory in vitro. Dev Cell. 2023 Nov 6;58(21):2359-2375.e8. doi: 10.1016/j.devcel.2023.08.001. Epub 2023 Aug 29. PMID: 37647896; PMCID: PMC10873093.

9. Suo S, Zhu Q, Saadatpour A, Fei L, Guo G, Yuan GC. Revealing the Critical Regulators of Cell Identity in the Mouse Cell Atlas. Cell Rep. 2018 Nov 6;25(6):1436-1445.e3. doi: 10.1016/j.celrep.2018.10.045. PMID: 30404000; PMCID: PMC6281296.

10. Van de Sande, B., Flerin, C., Davie, K. et al. A scalable SCENIC workflow for single-cell gene regulatory network analysis. Nat Protoc 15, 2247–2276 (2020). https://doi.org/10.1038/s41596-020-0336-2