Gerstein Lab Experience in Gene Regulatory Networks, Cell-to-Cell Communication, and Shared Molecular Mechanisms

Our lab has extensive experience in RNA-seq, ChIP-seq, and other genomic data processing. We have implemented standardized pipelines such as the ENCODE/GTEx pipelines for uniform processing across bulk and single cell datasets. This allowed consistent comparison across tissues and integration of data from different sources. Specifically, to identify brain transcriptional signatures, we applied standardized and established ENCODE pipelines to uniformly process RNA-seq data across PsychENCODE, GTEx, and Roadmap Epigenomics datasets [1]. This consistency ensured our expression data and downstream analyses were comparable to previous work. Using these data, we identified noncoding regions of transcription and sets of differentially expressed and coexpressed genes.

The work from Prof. Mark Gerstein's lab sought to elucidate how genetic diversity influences gene expression across different cell types in the human brain, particularly focusing on neuropsychiatric disorders [2]. We processed single-nuclei multi-omics datasets into a significant resource, containing data from over 2.8 million nuclei derived from the prefrontal cortex. This resource spans 28 distinct cell types, allowing detailed analysis of gene expression and chromatin states across a diverse set of individuals, including those with neuropsychiatric disorders. We utilized snATAC-seq and snRNA-seq data from healthy individuals to predict regulatory links between transcription factors (TF) and their target genes, identifying both proximal and distal regulatory interactions. Using a combination of SCENIC and scGRNom pipelines, we successfully constructed cell-type-specific GRNs that captured key regulatory elements across the brain cell types. These networks elucidate the interactions between transcription factors, enhancers, and target genes, offering a dynamic view of gene regulation across different cell types and conditions. Figure 3 presents the construction of cell-type-specific gene regulatory networks through the integration of snRNA-seq and snATAC-seq datasets from the brain tissues of healthy individuals [3].

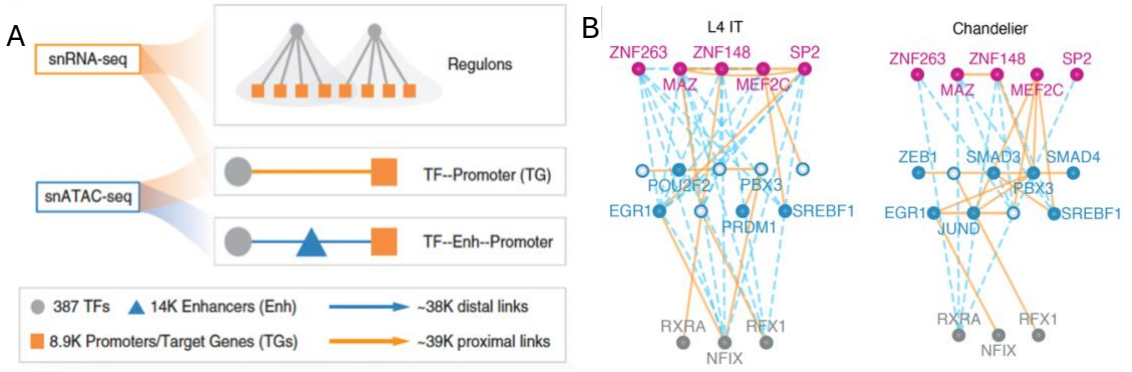

Figure 3. (A) Overview of cell-type-specific GRN construction based on snRNA-seq and snATAC-seq datasets. (B) Network graphs illustrating selected excitatory (L4 IT) and inhibitory (Chandelier) gene regulatory networks. Transcription factor nodes are colored pink, blue, or gray to denote out-hubs, bottlenecks, and in-hubs, respectively. TFs without blue fill are not functioning as bottlenecks in the corresponding cell type. Proximal connections are depicted with solid orange lines, while distal connections are represented by dashed blue lines.

Leveraging the regulatory and expression data from published resources, we developed networks depicting cell-to-cell communication within the brain by integrating publicly accessible ligand-receptor pair data with snRNA-seq. We identified distinct broad communication patterns among excitatory, inhibitory, and glial cell types, with significant alterations in these patterns observed in patients with neuropsychiatric disorders such as schizophrenia and bipolar disorder. Specifically, we noted that excitatory neurons showed reduced incoming signaling in schizophrenia, whereas inhibitory neurons exhibited increased signaling. These networks reveal how cells interact molecularly, shifting the understanding of cellular interactions in both healthy and diseased brain states. Figure 4 illustrates the construction of intercellular communication networks in our recent publication through systematic mapping of co-expressed ligand-receptor pairs, with a circos plot visualizing the signaling interactions between distinct brain cell populations [3].

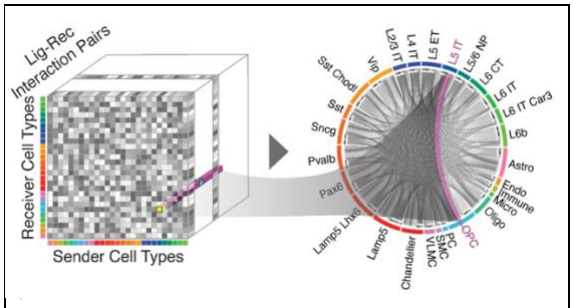

Figure 4. Intercellular communication networks derived from ligand-receptor co-expression analysis, with circos visualization highlighting L5 IT to OPC communication.

Our lab has integrated data from various sources to develop a unified view of variant effects across multiple disorders [4]. This effort underscored the need for unified frameworks to address common challenges in investigating and interpreting the impact of genomic variants. This idea can also be applied to study disparate diseases that share epidemiological relationships. Our lab has extensive experience in studying common, shared, or interconnected elements. We have used Fisher's exact test and Jaccard index to classify enhancer co-regulation modes by analyzing regulation modules, each consisting of two enhancers and a TSS [5]. For each module, we represented enhancer activity across samples as binary vectors and assessed their relationships using Fisher's exact test and the Jaccard index. Modules were categorized as simultaneous co-regulation or as mutually exclusive co-regulation based on the Fisher's exact test and the Jaccard index. This framework provided insights into enhancer interactions in samples of human primary cells, tissues and cell lines. We also have experience in gene set enrichment analysis (GSEA). We have ever utilized normalized expression data processed through DESeq2 and normalized microarray gene expression values, applying GSEA to curated gene sets from the Molecular Signatures Database (MSigDB) Hallmark and Gene Ontology collections of functional gene sets [6]. Standard GSEA protocols were followed, including phenotype permutations and collapsing gene identifiers to gene symbols, ensuring robust and reproducible pathway enrichment analysis.

References

1. Wang, D., et al., *Comprehensive functional genomic resource and integrative model for the human brain.* Science, 2018. **362**(6420).
2. Emani, P.S., et al., *Single-cell genomics and regulatory networks for 388 human brains.* Science, 2024. **384**(6698): p. eadi5199.
3. Emani, P.S., et al., *Single-cell genomics and regulatory networks for 388 human brains.* bioRxiv, 2024.
4. Kumar, S. and M. Gerstein, *Unified views on variant impact across many diseases.* Trends in Genetics, 2023. **39**(6): p. 442-450.

5. Cao, Q., et al., *Reconstruction of enhancer–target networks in 935 samples of human primary cells, tissues and cell lines.* Nature Genetics, 2017. **49**(10): p. 1428-1436.
6. Wang, G., et al., *Cross-platform transcriptomic profiling of the response to recombinant human erythropoietin.* Scientific Reports, 2021. **11**(1): p. 21705.