Gerstein's lab experience in analyzing transcriptome and epigenome datasets to build gene regulatory networks and models for identifying disease variants

Gerstein's lab focuses on analyzing transcriptome and epigenome datasets to construct gene regulatory networks and predictive models related to brain disorders. Its expertise lies in leveraging large-scale genomic data to uncover molecular mechanisms underlying these conditions. Our team has conducted snRNA-seq and snATAC-seq on frozen postmortem tissue from multiple brain regions, yielding over 4 million single nuclei from 388 prefrontal cortex (PFC) samples and 55 central nucleus of the amygdala (CNA) samples. These datasets accurately reflect the diversity of cell types in the human brain, demonstrating high concordance across individuals with consistent rediscovery of the same cell populations, highlighting our capability in handling population-scale single-cell data.

In addition, our investigator team has been collaborating for years to build the official ENCODE[1] and PsychENCODE[2] annotation resources for single-cell multi-omics integration. By integrating multiple data modalities, including scQTLs, snATAC-seq, TF-binding sites, and gene coexpression, we constructed gene regulatory networks for PFC cell types. We linked TFs to potential target genes based on their coexpression relationships from snRNA-seq data and mapped scQTLs to connect promoters and enhancers. These networks are available in a variety of user-friendly formats. Finally, we utilize deep neural networks (e.g., DSPN[3], LNCTP[4]) to accurately impute cell-type-specific expression and phenotype from genotype. This effort identified over 250 risk genes and potential drug targets for brain-related disorders, along with associated cell types. Building on gene regulatory rules, we build interpretable deep-learning approaches to address how common variants jointly contribute to brain disorders (e.g., Alzheimer's disease), and through which cell types these variants exert their influence on the disease.

[1] Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan KK, Cheng C, Mu XJ, Khurana E, Rozowsky J, Alexander R, Min R. Architecture of the human regulatory network derived from ENCODE data. Nature. 2012 Sep 6;489(7414):91-100.

[2] Akbarian S, Liu C, Knowles JA, Vaccarino FM, Farnham PJ, Crawford GE, Jaffe AE, Pinto D, Dracheva S, Geschwind DH, Mill J. The psychencode project. Nature neuroscience. 2015 Dec;18(12):1707-12.

[3] Wang D, Liu S, Warrell J, Won H, Shi X, Navarro FC, Clarke D, Gu M, Emani P, Yang YT, Xu M. Comprehensive functional genomic resource and integrative model for the human brain. Science. 2018 Dec 14;362(6420):eaat8464.

[4] Emani PS, Liu JJ, Clarke D, Jensen M, Warrell J, Gupta C, Meng R, Lee CY, Xu S, Dursun C, Lou S. Single-cell genomics and regulatory networks for 388 human brains. Science. 2024 May 24;384(6698):eadi5199.