Gerstein Lab's Expertise in Transcriptome, Epigenome and Proteome Analysis for Constructing Gene Regulatory Networks and Identifying Disease Variants

We have utilized deep neural networks (e.g., DSPN[1], LNCTP[2]) to accurately impute cell-type specific expression and phenotype from genotype in the context of neuropsychiatric disorders. 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 established from previous aims or existing methods. Our modeling approach aims to address how variants jointly contribute to brain diseases/disorders, and through which cell types these variants exert their influence on the disease. Furthermore, we leverage the capability of a large language model (LLM) from predicting protein 3D structure to predict phase transitions -- a key biophysical mechanism underlying Alzheimer's disease (AD) initiation and progression. Specifically, our findings indicate that among the two main proteins involved in AD, amyloid beta precursor protein (APP) is more prone to aggregation (by $A\beta 42$ formation) than tau, suggesting its predominant role in the early stages of the disease. We also demonstrated that the embedding vectors derived from the LLM effectively identify aggregation-prone proteins, aligning with the expected connection between structure and associated physical behaviors. Finally, we found that both APP and MAPT genes are downregulated in AD; however, APP showed a significant decrease while tau-related gene (MAPT) did not significantly change between AD and controls. We further investigated the regulatory mechanisms of these genes by deriving the gene regulatory networks for APP and tau. We found that APP is more tightly regulated, suggesting that higher aggregation propensity is naturally subjected to more regulation to minimize its impact on the disease[3]. In addition, we demonstrated the effectiveness of graph neural networks in predicting liquid-liquid phase separation (LLPS), which serves as a precursor to protein aggregation and is therefore crucial to identify as a potential drug target[4].

[1] Wang, Daifeng, et al. "Comprehensive functional genomic resource and integrative model for the human brain." *Science* (2018).
[2] Emani, Prashant S., et al. "Single-cell genomics and regulatory networks for 388 human brains." *Science* (2024).
[3] Frank, Mor, et al. "Leveraging a large language model to predict protein phase transition: A physical, multiscale, and interpretable approach." *Proceedings of the National Academy of Sciences* (2024).
[4] Wang, Gaoyuan, et al. "A Variational Graph Partitioning Approach to Modeling Protein Liquid-liquid Phase Separation. Cell Rep Phys Sci (2024).