**Gerstein lab experience with *Interpretable Deep Learning Models to Characterize Cell-type-specific Variant Impact in Alzheimer’s Disease***

We have extensive experience in processing large-scale single cell data. Specifically, we are the lead analysis group in pioneering consortia, including ENCODE, PsychENCODE, and SCORCH, where we have uniformly processed population-scale single-cell transcriptomic data from over 300 postmortem brain regions. Our cell-type-specific gene expression signatures have been widely used in the PsychENCODE to prioritize risk factors in multiple brain disorders.

Our investigator team has been working together for years to build the official ENCODE and PsychENCODE annotation resources via bulk tissue and single-cell multi-omics integration. Specifically, we constructed gene-centric annotations by linking CREs to genes and constructing TRNs. In addition, we have extensive experience in using Hi-C to validate CRE-gene linkages1, and much previous experience with network analysis framework development2-4. We have processed cell-type-specific Hi-C in neural and microglial cells1 to validate our Direct-net CRE-gene linkage predictions.

We have extensive experience in quantifying TF and variant impacts in disease studies. For instance, we developed a variant-prioritization pipeline named FunSeq that included an adjustable data context 5,6. This tool has been widely used to identify disease-causing mutations for further in-depth analyses to understand the mechanisms underlying disease pathogenesis. FunSeq links each noncoding mutation to target genes, and prioritizes such variants based on functional annotation, sequence features, conservation, network connectivity, and mutation frequency in diseases. We also developed a generalized model named GRAM to predict cell-type-specific molecular effects of non-coding variants on their associated genes7. This tool has been used to predict the effects of fine-mapping causal variants from genome-wide association studies. Finally, we developed a variant-scoring framework named RADAR to pinpoint variants associated with RNA binding protein function dysregulation 8. In addition, we developed AlleleSeq, a tool for detecting candidate variants associated with allele-specific binding and allele-specific expression 9-11.

To fully characterize genotype-phenotype association, we harmonized bulk transcriptome and their genotype information for 1866 individuals with large-scale functional assays (e.g., ChIP-seq, Hi-C, and scRNA-seq). First, we created a comprehensive brain regulatory map, including enhancers and their targets, various quantitative-trait loci (e.g., expression, isoform, cell fraction, and chromatin QTLs), and gene regulatory networks. Then, we embedded such regulatory information into a deep-learning model to predict psychiatric phenotypes from genotype and transcriptome. Our DSPN model gives a ~6-fold improvement in prediction over additive polygenic risk scores (**Table 2**). Lastly, it highlights key genes and pathways associated with disorder prediction, including immunological, synaptic, and metabolic pathways, recapitulating *de novo* results from more targeted analyses.

1 Wang, D. *et al.* Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, eaat8464 (2018).

2 Lou, S. *et al.* TopicNet: a framework for measuring transcriptional regulatory network change. *Bioinformatics* **36**, i474-i481, doi:10.1093/bioinformatics/btaa403 (2020).

3 Yan, K. K., Wang, D., Xiong, K. & Gerstein, M. Comparing Technological Development and Biological Evolution from a Network Perspective. *Cell Syst* **10**, 219-222, doi:10.1016/j.cels.2020.02.004 (2020).

4 Gerstein, M. B. *et al.* Architecture of the human regulatory network derived from ENCODE data. *Nature* **489**, 91-100, doi:10.1038/nature11245 (2012).

5 Fu, Y. *et al.* FunSeq2: a framework for prioritizing noncoding regulatory variants in cancer. *Genome biology* **15**, 480, doi:10.1186/s13059-014-0480-5 (2014).

6 Khurana, E. *et al.* Integrative annotation of variants from 1092 humans: application to cancer genomics. *Science* **342**, 1235587, doi:10.1126/science.1235587 (2013).

7 Lou, S. *et al.* GRAM: A GeneRAlized Model to predict the molecular effect of a non-coding variant in a cell-type specific manner. *PLoS Genet* **15**, e1007860, doi:10.1371/journal.pgen.1007860 (2019).

8 Zhang, J. *et al.* RADAR: annotation and prioritization of variants in the post-transcriptional regulome of RNA-binding proteins. *Genome biology* **21**, 151, doi:10.1186/s13059-020-01979-4 (2020).

9 Onuchic, V. *et al.* Allele-specific epigenome maps reveal sequence-dependent stochastic switching at regulatory loci. *Science* **361**, doi:10.1126/science.aar3146 (2018).

10 Chen, J. *et al.* A uniform survey of allele-specific binding and expression over 1000-Genomes-Project individuals. *Nat Commun* **7**, 11101, doi:10.1038/ncomms11101 (2016).

11 Rozowsky, J. *et al.* AlleleSeq: analysis of allele-specific expression and binding in a network framework. *Mol Syst Biol* **7**, 522, doi:10.1038/msb.2011.54 (2011).