Gerstein Lab's Experience in Extending Omics-Based Genotype-to-Phenotype Prediction Models

Extension of an Omics-based Genotype-to-Phenotype prediction Model: We have created a predictive model that quantifies phenotypic risk for certain brain-related conditions conditional on genotypes. The model, termed the Linear Network of Cell Type Phenotypes (LNCTP, which is derived from the proposed model). For Aim 1, we have extended this model by incorporating public omics datasets, such as the Accelerating Medicines Partnership for Alzheimer's Disease (AMP-AD) datasets. These datasets were uniformly processed and integrated into our model, enhancing the richness of data for training and validation. Additionally, we are currently harmonizing tissue-level and single-cell genomics data from various public sources including NeMO, GTEx, BrainSpan, PsychENCODE, and the Allen Human Brain Atlas (AHBA), ensuring compatibility and consistency across different datasets.

Gerstein Lab's Experience in Enhancing Model Interpretability with Graphical LASSO

Graphical LASSO Implementation: To improve model interpretability and robustness, we integrated the LNCTP with a Graphical LASSO approach, which imposes a sparse structure on the precision matrix to select the most relevant features. This method enhances the model's ability to highlight key regulatory interactions and pathways, thereby improving interpretability. By inducing sparsity, Graphical LASSO mitigates overfitting, especially in high-dimensional data typical of genomics studies, and aids in feature selection and precision matrix estimation for gene regulatory networks. The Graphical LASSO was implemented through unary and GMRF (Gaussian Markov Random Field) training steps, with the precision matrices being optimized to reflect sparse regulatory connections. Incorporating Graphical LASSO into the LNCTP model has led to improved predictive accuracy for cognitive and psychiatric traits. The enhanced model uses both tissue-specific and cell-type-specific regulatory networks and achieves better performance.

Gerstein Lab's Experience in Integrative Data Analysis and Model Expansion

Data Integration and Model Expansion: We integrated single-cell RNA-seq data from multiple databases capturing cell-type-specific regulatory mechanisms. Furthermore, the model was used to simulate the effects of gene perturbations on cellular phenotypes, providing insights into potential therapeutic targets. By simulation, we predicted the downstream effects on gene regulatory networks and phenotypic outcomes.