Gerstein Lab experience in hierarchical deep learning model to predict variant impact related to aging and disease phenotype (case / control)

We develop a flexible hierarchical phenotype network to predict variant impact. Ithas several visible layers associated with the inferred components, including genotypes at scQTL and bulk eQTL sites, cell-type specific and bulk tissue-based gene regulatory network (GRNs), cell-type fractions, cell-to-cell communication networks, gene co-expression modules, and sample covariates. Moreover, bulk and cell-type gene expression levels can be imputed from the genotype (eQTL and scQTL variables), using a conditional energy-based model incorporating GRNs and cell-to-cell network connections as indicated. These cell-type-specific nodes with dense connectivity are then incorporated into a deep linear model to predict phenotypes in each sample and prioritize cell types and genes associated with each trait. Furthermore, the model can impute missing data for the discovery of cell-type specific molecular phenotypes important for neuropsychiatric disorders. Doing so allows us to link variants with intermediate cell-type functional genomic activities, such as cell-type-specific gene expression, pathway activity, and cell-cell communication.

Gerstein lab experience in linking cell-specific genotype with aging:

To explore the relationship between the transcriptome and aging, we constructed a model to predict an individual's age from the single-cell data. The model shows that the transcriptomes of six cell types (i.e., L2/3 IT, L4 IT, L5 IT, L6 IT, Oligodendrocytes, and OPC) have strong predictive value. It also shows that many genes contribute to the model, highlighting the broad transcriptome changes in aging.

Gerstein lab experience with cell-specific modeling to predict Alzheimer’s disease in the context of aging: We obtained the cell-type fraction for each bulk sample. Specifically, we used the bMIND software package1 to infer cell-type-specific gene expression, cell-type-specific methylation and cell fraction for each sample. We assessed cell fraction changes with aging and found that the fraction of SST neurons decreases significantly with age. We also selected differentially methylated regions to infer cell-type specific methylation patterns using methylation data of 740 individuals in ROSMAP. We assessed the contributions of gene expression and methylation features from each cell type to predicting Alzheimer’s disease status by the random forest model. Oligodendrocytes are found to be significantly associated with an increased cell fraction in the Alzheimer’s disease, while L2/3 IT and Pvalb and Sst are significantly associated with a decreased cell fraction in Alzheimer’s disease.

References:

1. Wang, J., Roeder, K. & Devlin, B. Bayesian estimation of cell type-specific gene expression with prior derived from single-cell data. *Genome Res* **31**, 1807-1818 (2021). <https://doi.org/10.1101/gr.268722.120>