Variant Effect Prediction Using Deep Neural Networks for Alzheimer's Disease

Lan, Alexander (School: The Harker School)

Alzheimer’s disease (AD), affecting >50 million individuals worldwide, is characterized by a progressive loss of cognitive function for which no effective therapies currently exist. To identify new genes and pathways involved in the disease, genome-wide association studies (GWAS) have been performed across thousands of cases and controls, identifying genetic variants that are statistically associated with AD. However, >90% of these variants reside within the noncoding genome, exerting their effects without altering gene sequences. Instead, these noncoding variants affect how genes are expressed by altering the genetic sequences of enhancers and promoters which are activated by sequence-specific transcription factors (TFs). Unlike amino-acid altering variants within genes, the complex grammar of TF binding and gene regulation makes it challenging to predict which variants are truly functional and which genes and cell types they affect. To pinpoint causal variants in AD, I developed complex convolutional neural networks which learn the grammar and syntax of the noncoding genome, enabling cell type-specific prediction of TF binding from the sequence alone. Highly predictive models (AUC ~0.9, Pearson r ~0.8) were trained using single-cell chromatin accessibility data from seven human brain cell types, and the importance of AD-related noncoding variants was ranked using pseudo-back-propagation. These models both verified known variants such as rs636317 and rs13025717 and nominated new candidate causal variants. To make these models broadly accessible, I developed a production-ready Variant Effect Prediction Platform, which gives researchers a scalable, accurate computational method for evaluating and prioritizing their variants of interest in all genetic brain diseases.

Awards Won:
Third Award of $1,000