**Comprehensive resource and integrative model for functional genomics of the adult brain**

Robust phenotype-genotype associations have been established for a number of neuropsychiatric disorders such as schizophrenia and bipolar disorder; however, the molecular mechanisms corresponding to these associations are unknown. Addressing this deficiency is a key aspect of the PsychENCODE consortium. To this end, the consortium has compiled comprehensive genotypic, transcriptomic, epigenetic, and single-cell sequencing data for thousands of individuals. To add to this, we have uniformly processed the overall PsychENCODE dataset and consistently merged it with other key large-scale genomic resources (e.g., Roadmap and GTEx) to develop a comprehensive data and analytic resource of the human adult brain, comprising over 2,000 individuals. This resource allows us to make a number of advances:
**\* We develop a definitive set of active brain enhancers, transcripts and regulatory networks.** In particular, we find ~88K brain enhancers that have high brain-activity signals across individuals.
**\* We place the gene expression and epigenetics of the brain in context of other tissues, in a consistent fashion.** Using a variety of spectral analyses, we have found that the brain has more distinct expression than most other tissues (Figure A) – in addition to a relatively large amount of non-coding transcription. However, the epigenetics differences are less.
\* **We develop the largest known expression quantitative trait loci (eQTL) set** for the brain. This comprises >1M variants and >11k eGenes (including non-coding genes, Figure B). In addition, we determined >5K associated chromatin QTLs. Collectively, these QTLs cover a larger fraction of brain-GWAS SNPs (e.g., 6% in schizophrenia, 10% in bipolar) than previously observed, providing clues to their molecular-level impact.
**\* Using the single-cell data, we can explain much (>80%) of the population-level variation in brain gene expression evident in eQTLs in terms of changing proportions of basic cell types, rather than changes in individual genes.** This facilitates segmentation of QTLs into distinct groups (eg those associated with overall cell-type selection vs. those clearly regulating an individual gene).
**\* We merge all the analytic results into an interpretable deep-learning model** relating genotypes, gene-expression levels, and epigenetics to the regulatory network and eQTLs. The model enables practical imputation of a subset of the transcriptome and epigenome, with an accuracy of ~70%.
**\* We use the model to improve prediction of biological variables and psychiatric diseases by the addition of transcriptomic data to genotype, as compared to genotype alone.** In particular, we show that we can predict biopolar disease and schizophrenia with much higher accuracy from the transcriptome than genotype alone i.ewith 8 Morever, we show clear predictive value of the imputed transcriptome over just using the genotype alone, demonstrating the usefulness of even a limited amount of functional genomics information for unraveling gene-disease relationships For instance, we can predict schizophrenia with an accuracy of 61% using our model and an imputed transcriptome, versus 56% with genotype alone.