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

Contact PI(s): M Gerstein, J Knowles & D Geschwind

Robust phenotype-genotype associations have been established for a number of neuropsychiatric diseases such as schizophrenia and bipolar disorder; however, the molecular mechanisms underlying these associations are unknown. Addressing this deficiency is a key aspect of the PsychENCODE Consortium. To this end, the consortium has generated genotypic, transcriptomic, epigenetic, Hi-C and single-cell sequencing data from thousands of individuals and processed them uniformly to ensure consistency. In addition, we have used the same analytic approaches to re-process data from other large-scale genomic projects (e.g., Epigenomics Roadmap and GTEx) and merge them with the PsychENCODE corpus to develop the largest analytic resource for the adult human brain, comprising 1,945 individuals (freely available online via psychencode.org). This resource allows us to make a number of advances:

* We compare transcriptomic and epigenomic data between the brain and other tissues in a consistent fashion, and develop the largest **reference set** of functional genomic elements for the adult brain, including active enhancers, transcripts, enhancer-gene linkages (supported by a full Hi-C data from adult brain), and regulatory networks. For instance, we identified ~88,800 enhancers and ~79,000 transcripts active in pre-frontal cortex and similar lists for a number of other brain regions. Moreover, using a variety of spectral analyses, we find that the brain has more distinct expression patterns compared to most other tissues (Fig. A), including a greater amount of non-coding transcription. However, the differences in epigenetics are relatively smaller.

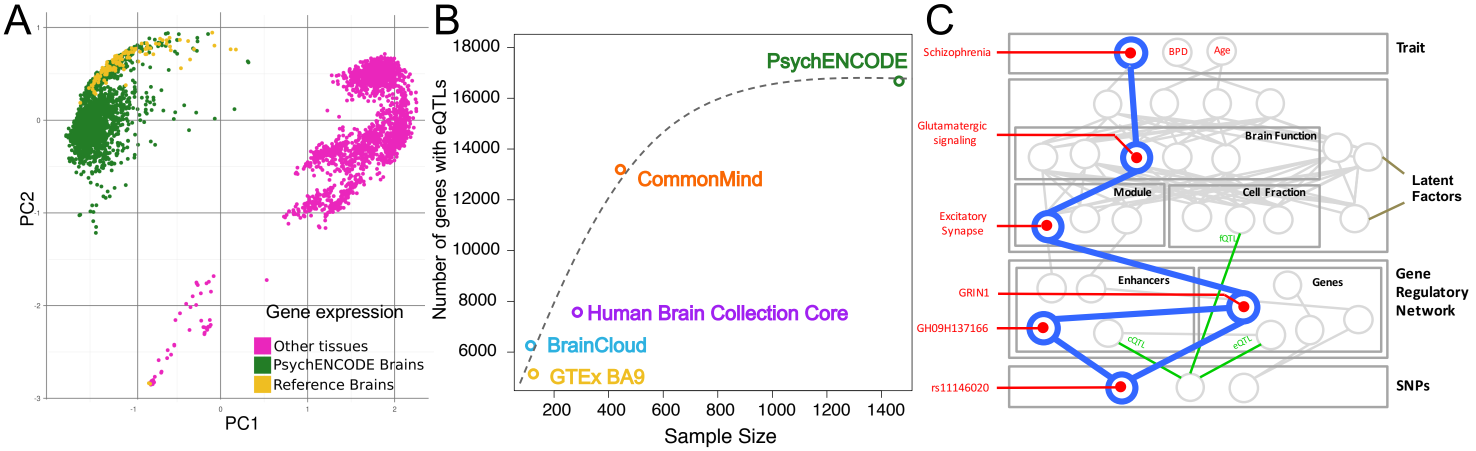
* Using **single-cell data**, we deconvolve the gene expression data from bulk brain tissue. We find that >80% of the inter-individual variation in tissue expression can be accounted for by alterations in the proportions of basic cell types, rather than by changes in individual genes. Moreover, cell fractions vary across brain phenotypes and disease. For instance, we see evidence of more excitatory neurons in autism spectrum disorder.

* We develop the largest set of brain quantitative trait loci (**QTLs**), including those for expression (eQTLs), chromatin (cQTLs), alternative splicing (sQTLs) and even cell fractions (fQTLs, from the single-cell analysis). For example, we have found >1M significant eQTLs, controlling >11K coding and non-coding genes (considerably more than previous studies, Fig. B). We also observed >5K chromatin cQTLs. Collectively, these QTLs annotate a larger fraction of GWAS SNPs involving the brain (e.g., 6% in schizophrenia, 10% in bipolar) than previously observed, providing leads on which genes are affected in disease.

* We merge the analytic results into a generative, **deep-learning model**, where we enforce interpretable constraints on the connectivity to mirror the regulatory architecture at multiple levels (Fig. C). This allows us to relate genotypes, gene-expression levels and epigenetics to the regulatory network and QTLs. This model enables practical imputation of a subset of the transcriptome and epigenome with an accuracy of >70%. Furthermore, we can use this integrated model to improve prediction of psychiatric diseases and other biological variables by the addition of functional genomics data to genotype. In particular, we show that we can predict bipolar disease and schizophrenia with much higher accuracy from the transcriptome, than from genotype alone (i.e., for schizophrenia, a three times accuracy improvement over a random baseline of 50%, +18% vs +6%).

Why Science?

This paper is one of the major "capstones" of the consortium, presenting the largest genomic resource to date on adult brain. It is the first to create a "next-generation resource," merging human population variation, epigenetics and single-cell data with interpretable, deep-learning models for predicting psychiatric disorders.


A
PC2
PC1
Gene expression
Other tissues
PsychENCODE Brains
Reference Brains
B
Number of genes with eQTLs
Sample Size
PsychENCODE
CommonMind
Human Brain Collection Core
BrainCloud
GTEx BA9
C
Trait
Schizophrenia
BPD
Age
Glutamatergic signaling
Brain Function
Latent Factors
Module
Cell Fraction
Excitatory Synapse
fQTL
Enhancers
Genes
Gene Regulatory Network
GRIN1
GH09H137166
cQTL
eQTL
rs11146020
SNPs