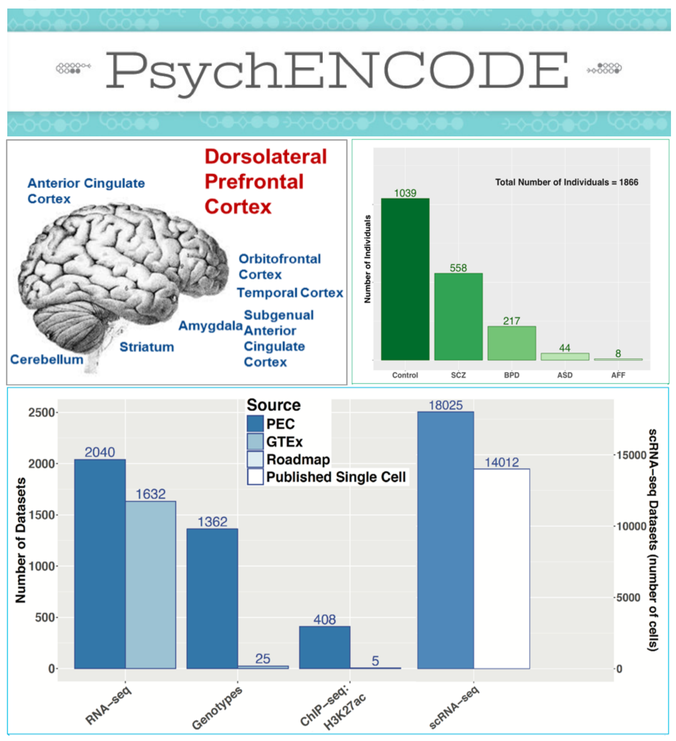
Single cell deconvolution reveals that the cell fractions vary across brain disorders and human aging

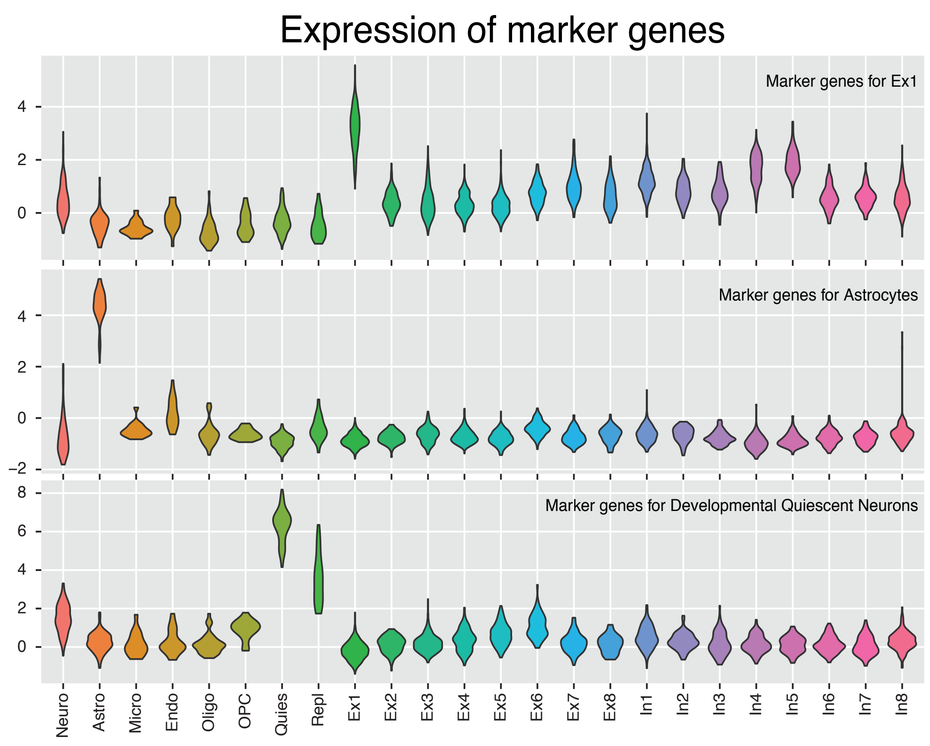
Xu Shi[1,2], Daifeng Wang[1,2,3], Declan Clarke[1,2], Mark Gerstein[1,2,4,5], PsychENCODE Consortium

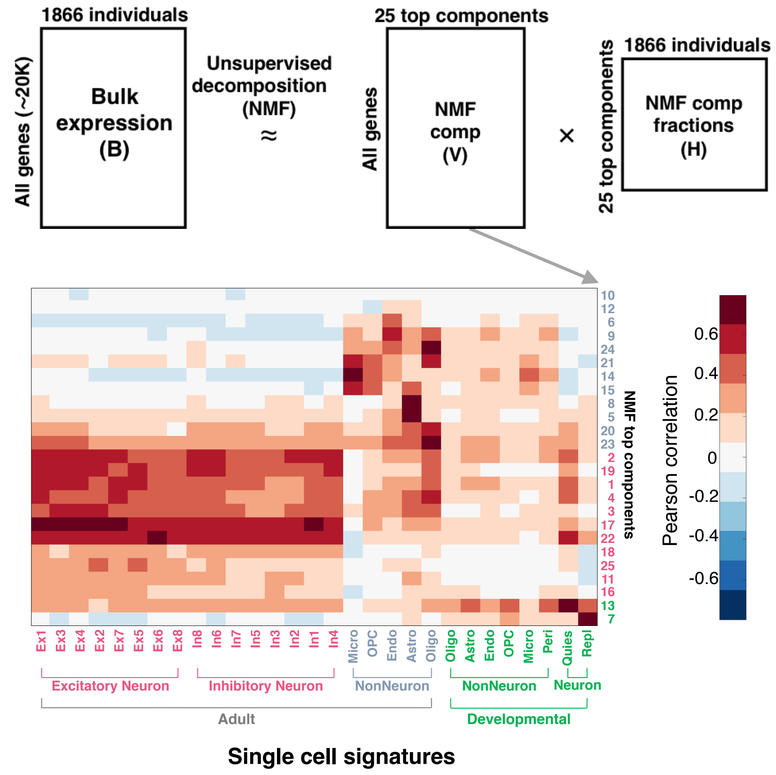
[1] Program in Computational Biology and Bioinformatics [2] Department of Molecular Biophysics and Biochemistry [4] Department of Computer Science [5] Department of Statistics and Data Science, Yale University, New Haven, CT 06520, USA

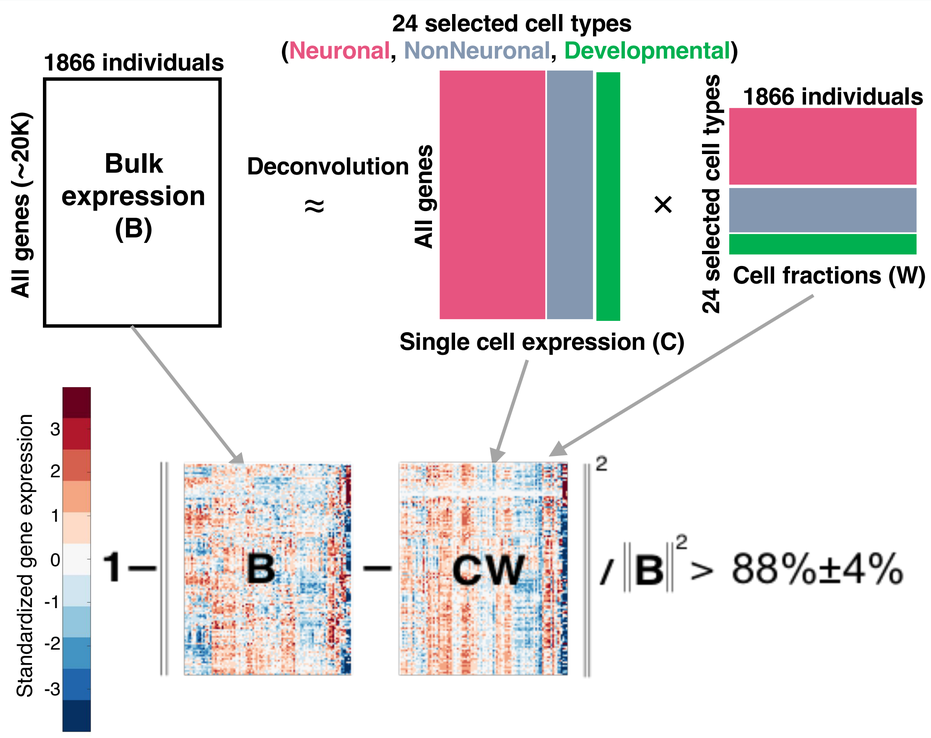
[3] Department of Biomedical Informatics, Stony Brook University, Stony Brook, NY 11794, USA

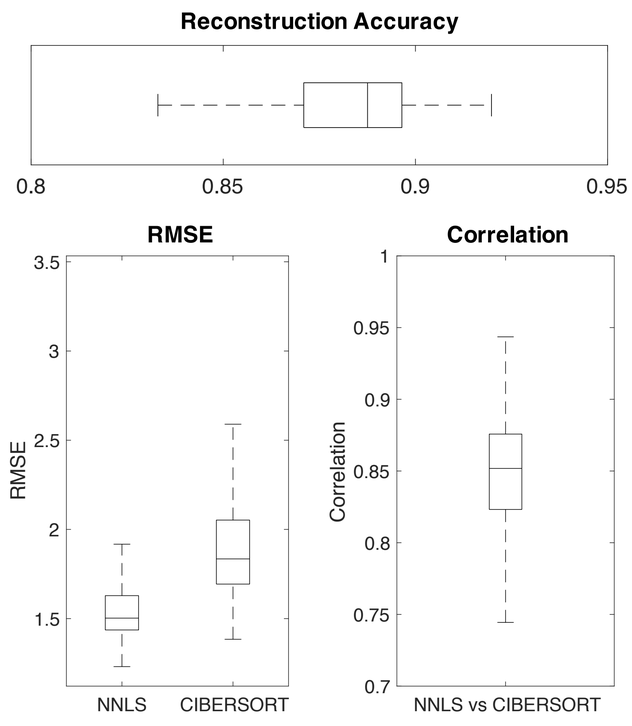
The PsychENCODE Consortium (PEC) has generated over 2,000 human brain samples for healthy controls and individuals afflicted by neuropsychiatric disorders including autism spectrum disorder (ASD), bipolar disorder (BPD), schizophrenia (SCZ) and affective disorder (AFF). These samples were measured in different assays such as full genotyping, RNA-seq, scRNA-seq, etc. To build a comprehensive expression dataset, we uniformly process the RNA-seq data from PEC, ENCODE, CommonMind, GTEx and Roadmap, which includes 1866 individuals. For the scRNA-seq data, we used standard pipelines to uniformly process single-cell RNA-seq data from PEC, in conjunction with other single-cell studies on the brain. Then we assembled profiles of brain cell types, including major neuron and non-neuronal, and additional developmental cell types. In total, we have built the expression profiles of 32,037 cells. We used an unsupervised analysis to identify the primary components of bulk expression variation. We decomposed the bulk gene-expression matrix using non-negative matrix factorization, and determined whether the top components, capturing the majority of covariance, were consistently associated with the single cell signatures. This demonstrates that an unsupervised analysis derived solely from bulk data roughly recapitulates the single-cell signatures, partially corroborating them.

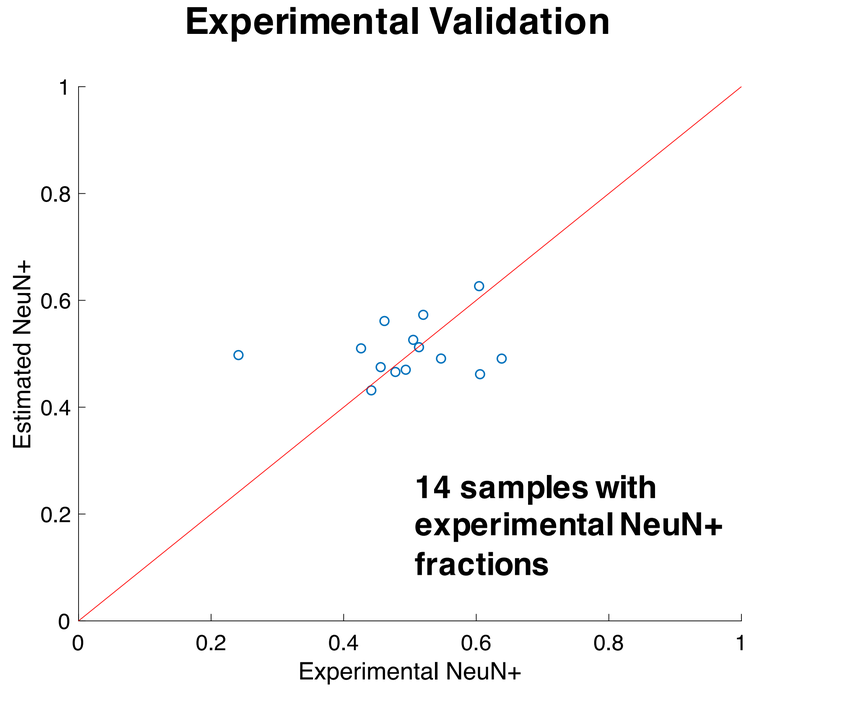












We identified cell-fraction changes associated with different traits. For example, there are different fractions of particular types of excitatory and inhibitory neurons in samples with disorders. ASD samples tend to have higher fractions of astrocytes and microglia cells compared to healthy control samples. We also observed an association with age. These changes may be associated with differential expression of specific genes. We further associated SNPs with changes in the relative fractions of specific cell types to identify "cell-fraction QTLs" (fQTLs).