Bulk data from tissues can reliably be used to quantify average gene expression. However, the expression differences between two tissues may be due to different proportions of constituent cell types, or similar proportions of cell types but with varying gene expressions. To parse these disparate contributions to expression differences, we propose to apply two recently developed complementary strategies examining expression across individuals. First, we will generate single-cell gene expression profiles by uniformly processing single-cell RNA-seq data, in conjunction with other published single-cell studies on the brain, using standard pipelines. The gene expression profiles from multiple studies will be integrated into a coherent expression profile of brain cell types, including both excitatory and inhibitory neurons, and major non-neuronal types (e.g., microglia and astrocyte). Second, we will decompose the bulk expression in an unsupervised manner to identify the primary components of bulk expression variation. We will apply non-negative matrix factorization (NMF) on bulk expression data, and determine whether the top components are consistently associated with the single-cell signatures. This analysis has previously demonstrated that an unsupervised analysis derived solely from bulk data roughly recapitulates the single-cell signatures, partially corroborating them. We will then examine how variation in bulk expression can be explained by the proportions of basic cell types. To this end, we will estimate the relative proportions of various cell types for each tissue sample (i.e., "cell fractions"). In particular, we will deconvolve the bulk expression (**B**) using the uniformly generated single-cell signatures (**C**) to estimate cell fractions across individuals (**W**). This analysis will determine the degree to which population-level expression variation can be explained by a weighted combination of single-cell signatures. To further link the single-cell analysis with the mutations, we will associate SNPs with changes in the relative fractions of specific cell types across individuals. We term such relationships “cell-fraction QTLs” (fQTLs). The fQTL SNPs would be very important for cell fractions differences across individuals. Furthermore, we will also attempt to re-evaluate the observed gene-expression variation, accounting for cell fraction differences between individuals. In particular, our deconvolution analysis will generate a matrix showing the bulk expression after accounting for cell fractions changes. Specifically, it is the component of the bulk tissue expression variation that cannot be explained by the changing cell fractions alone: **Δ = B - CW**. We can subsequently use this quantity to determine “residual QTLs” by directly correlating it with genotype.

This method has been successfully applied to the bulk and single cell data from the PsychENCODE consortium. The unsupervised decomposition showed that the principal components of bulk expression highly correlated with single-cell signatures. Also, the deconvolution analysis demonstrated that weighted combinations of single-cell signatures could account for most of the population-level expression variation, with an accuracy of ~89%. The cell-fraction changes showed significant association with different traits such as gender, age and psychiatric disorders. For fQTLs, we identified 1672 distinct SNPs constituting 4199 fQTLs. Of these, two excitatory neurons subtypes were associated with the most fQTLs. We also identified 202,940 SNPs involved in residual eQTLs. These results support the use of a combination of unsupervised cell-signature analysis and supervised deconvolution to determine the underlying contributors to bulk gene-expression variation.