PsychENCODE consortium

* database for studying effects of functional genomic elements in patients with psychiatric disorders
* Consists of a number of research groups

Studies included:

* CMC, UCLA-ASD, SZBDMulti-seq, MultiomeBrain, DevBrain, IsoHuB, PTSDBrainomics, LIBD
* Includes samples of PFC with control, schizophrenia, BPD, ASD, alzheimer’s, PTSD

snRNAseq processing

* Count matrix generated w cellranger, each sample run independently w no aggregation, but pooled if one sample run through multiple lanes
* Demultiplexing:
  + Quantify per-cell hashtag oligos (used cli CITE-seq-count package embedded in python)
* Ambient RNA cleanup: cellbender remove-background program
* Per-fastq set/samples processed w pegasus, applied to cellbender output individually
  + Removed mitochondrial genes etc
  + Singlets retained
  + Doublets IDd and removed w scrublet doubletdetection
  + Dempultiplexed samples aggregated again for gene ID, highly variable gene selection, PCA, Harmony batch correction, nearest-neighbor detection, Leiden clustering, and UMAP dimensionality reduction

DEA with t test, compare expression of genes in clusters against all others

* For disease traits:
  + Filtering: CPM normalization to filter out lowly expressed genes (>0.5 in <30% of samples) and indivs with <50 cells detected. After filtering, cell types with <16 samples also removed
  + DEA performed on raw counts using Deseq2 likelihood ratio test standard pipeline
    - Covariates: age, gender, genotype ancestry, PMI, average UMI per cell, and disease status
    - Contrasts made between disease and healthy
    - Multiple testing corrections performed
    - Adjusted p value of <0.05 defined as differentially expressed
* For aging:
  + Control split into young (25-70) and old (70-90) with balanced numbers
    - Filtering: CPM normalization to filter out lowly expressed genes (>0.5 in <30% of samples) and indivs with <50 cells detected. After filtering, cell types with <16 samples also removed
    - DEA performed on raw counts using Deseq2 likelihood ratio test standard pipeline
      * Covariates: age, gender, genotype ancestry, PMI, average cell UMI count, and disease status
      * Contrasts made between old and young
      * Multiple testing corrections performed
      * Adjusted p value of <0.05 defined as differentially expressed
  + Schizophrenia split into young (25-70) and old (70-90) with balanced numbers, samples only from SZBD and CMC
  + Same DEA as control
* Celltype inference based on marker gene collection
* Raw count matrices for each sample used in following steps