Submitter Name: Nathan Nakatsuka Submitter Email: 08nanaka@gmail.com
PI Name: Eric Nestler PI Email: eric.nestler@mssm.edu

A Meta-Analysis Method for Finding Reproducible Differentially Expressed Genes in Single-Cell Transcriptomic Case-Control Studies

Nathan Nakatsuka[1,2,3,4], Longda Jiang[1,2], Austin Hartman[1,2], Eric Nestler[4], Rahul Satija[1,2]

[1] New York Genome Center; New York, NY 10013
[2] Center for Genomics and Systems Biology, New York University; New York, NY 10003
[3] Department of Psychiatry, New York University Grossman School of Medicine; New York, NY 10016
[4] Nash Family Department of Neuroscience and Friedman Brain Institute, Icahn School of Medicine at Mount Sinai; New York, NY 10029

Our goal is to develop improved analytical methods to enable the identification of dysregulated genes in neuropsychiatric disorders, with a main interest in drug addiction. As a first step toward developing such approaches, we compiled data from 16 previously published Alzheimer's Disease (AD) and 16 COVID-19 single cell RNA sequencing (scRNA-seq) studies to systematically examine the reproducibility of differentially expressed genes (DEGs) across studies. We found that while COVID-19 DEGs had moderate reproducibility, AD DEGs had relatively poor reproducibility, though larger datasets were better than smaller ones. We developed a non-parametric meta-analysis method that prioritizes reproducibility by summing relative differential expression ranks across datasets and obtained empirical p-values by permutations. We found 521 genes across 7 cell types in AD and 1,638 genes across 8 cell types in COVID-19 reaching statistical significance (FDR<0.05). The meta-analysis genes had good prediction for AD and COVID-19 case-control status (AUC of 0.81 and 0.91, respectively) in left-out datasets. Specificity and sensitivity of these genes were substantially higher than those discovered by dataset merging and inverse variance weighted p-value aggregation methods. The DEGs were enriched for genes found in GWAS and exome-sequencing studies and revealed known and novel biological pathways based on gene ontology. We then adapted this method and discovered several genes with sex-biased expression. We are now applying this method to whole brain scRNA-seq data we are generating on rodent cocaine intravenous self-administration models and previously published studies to promote the rigorous discovery of reproducible DEGs in the heterogeneous syndrome of addiction.