**Excitability Datasets**

[Schaukowitch 2017](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5524384/) contains bulk RNA-seq data of mouse primary cortical and hippocampal neurons with excitability perturbations. They treated neurons with TTX or bicuculline for 2h or 6h at days *in vitro* 12. I need to read an excel file into a dataframe in python (the gene list in my *Analysis* folder) and compare it to primary data sources to see if autistic individuals give evidence of a change in homeostatic plasticity—through synaptic scaling in the case of TTX and bicuculline perturbations. Gene list available.

[Boulting, *et al.* 2020](https://ncbi.nlm.nih.gov/pmc/articles/PMC7933108/) contains bulk RNA-seq and scRNA-seq of human iPSC-derived GABAergic neurons subjected to excitability perturbations *in vitro*. These neurons were sequenced 15 minutes, and 1/2/4h post-KCl exposure. They also measured H3K27Ac modifications to the chromatin architecture of under similar conditions. They also predicted enhancer regions and their gene targets. Gene list available.

* Isogenic human iPSC GABAergic neurons show expression most similar to ventral telencephalon, specifically post-conception week 16 embryonic ganglionic eminence and septum-derived cell types, which are endogenous progenitors to interneurons and spiny projection neurons in the striatum and basal ganglia. Differentiation yielded 93% GABAergic neurons *in vitro.*
* Single cell RNA sequencing of 37,101 isogenic cells over three batches of unperturbed neurons, basically an atlas effort to characterize what cell types these hGNs are most similar to.
* Total RNA sequencing of 1G hGNs at 15 min (n = 3), 1 h (n = 3), 2 h (n = 6) and 4 h (n = 3) after membrane depolarization via 55 mM KCl **or current injection?**, compared to expression in unstimulated cultures (n = 6). Membrane depolarization was measured by whole-cell patch-clamp electrophysiology (**voltage clamp?**). Unique differentially expressed genes detected from 1G = 655, 4G = 854, 432 genes in common, z score = 84.59, FDR = 0.05, fold-change threshold = 1.5. Repeated experiments in 4 other isogenic lines, denoted as 4G here.
* In the data file “Boulting\_2020\_Differential\_RNA,” the 8 tabs “9-1G **[#]**h ASDgenes”, “9-4G **[#]**h ASDgenes” are the SFARI genes induced 15 min, 1/2/4h post-KCl from total RNA sequencing of the 5 isogenic lines. The equivalent dataset for single cell RNA sequencing is the single tab, “9-scRNAseq inducible ASD.” The total inducible genes that overlap from the 1G scRNAseq and 1G total RNAseq dataset is “8-SingleCellseq\_1GtotalRNAseq.” A deeper look into the scRNAseq clusters is “7-scRNAseq\_privately-ind.” The intersection of the differentially expressed genes between the 1G and 4G genotypes is “3-inducible in both 1G and 4G.” There are the individual differential expression analyses using EdgeR, which are the first 8 tabs which are like the ‘ASDgenes’ tabs except they are more inclusive to genes not annotated in the SFARI database. ***The first analyses*** ***will be done*** on the genes at the intersection of the methods, “8-SingleCellseq\_1GtotalRNAseq” (452 genes) and the different genetic backgrounds, “3-inducible in both 1G and 4G” (432 genes). First, I’ll check how many genes appear in both tabs and take the intersection as one gene set, but I’ll also treat each tab as its own gene set for a total of 3 comparisons to the primary post-mortem data.

[Valakh, *et al.* 2023](https://elifesciences.org/articles/74899) knocked out PARbZIP family transcription factors (*Hlf, Tef, Dbp,* and *Nfil3*) that modulate expression of genes ontologically associated with excitability in mice and electrophysiologically and transcriptionally characterized cortical slices. These murine cortical slices from the neocortex were collected at P7, then treated with TTX for equivalent postnatal days 12-17 (5d) before bulk RNA-seq. They FACS sorted deep layer 5/6 neurons and parvalbumin-positive interneurons for RNA-seq. These same cortical slices were TTX treated from EP12-14 before electrophysiological characterization. The raw sequencing reads are available for download on the NCBI SRA Run Selector, an interface that allows you to deliver public datasets to your preferred cloud computing environment. You need the SRA accession number for a specific dataset, or the NCBI BioSample ID for the related datasets to a paper, which are provided in the paper link.

[Genc, *et al.* 2020](https://elifesciences.org/articles/55775#data) created *Drosophila* with heterozygous null mutations of x5 ASD-associated genes (*RIMS1, CHD8, CHD2, WDFY3* and *ASH1L*) and administered philanthotoxin-433 at the neuromuscular junction as a perturbation to postsynaptic homeostatic plasticity (excitability?). PhTx is a postsynaptic glutamate antagonist, which decreases the amplitude of postsynaptic mEPSPs, dampening synaptic transmission. They also created double knockouts of these same 5 genes (second site non-complementation) and found suppression of PHP. Then, they did a forward genetic screen using a small library of known karyotypical abnormalities tiling the 3rd chromosome, containing 5-50 genes per mutation and covering 6000 genes total. The more genes per mutation had no correlation with the magnitude of mEPSP amplitude disruption. Gene list available.

**Primary Datasets**

[Velmeshev, et al. 2019](https://www.science.org/doi/10.1126/science.aav8130) performed single nucleus RNA sequencing on 41 samples from post-mortem human prefrontal cortex and anterior cingulate cortex from 15 ASD patients yield 52,003 nuclei and 16 controls yielded 52,556 nuclei. There was an additional 22,000 nuclei sequenced from 8 PFC samples of sporadic epilepsy patients and 7 PFC matched control samples.

* Ages 4-22 years old and matched for age, sex, RNA integrity number, and post-mortem interval. None of the ASD patients had intellectual disability, but ~8 had seizures alongside ASD. They detected 17 cell types including subtypes of excitatory neurons, interneurons, and astrocytes with enrichment of layer 4 excitatory neurons and protoplasmic astrocytes in ASD patients.
* Using a linear mixed model, they identified 692 differential expression events in 513 unique differentially expressed genes at a *q* value < 0.05; expression level change ≥ 10%. 407 out of 513 were differentially expressed in a single cell type (Data S4). Only 17% of DEGs were cell-type specific genes, suggesting cell type–selective dysregulation of ubiquitously expressed genes. The top neuronal-specific DEGs were dysregulated in L2/3 excitatory neurons and VIP-expressing interneurons, while the top DEGs in non-neuronal cell types were dysregulated in protoplasmic astrocytes and microglia. 75 of the 513 DEGs are found in the SFARI database (Fig. S3). SFARI genes were enriched in L2/3/4 excitatory neurons and VIP-expressing and SST-expressing interneurons. Fig S6 differentiates sporadic epilepsy DEGs from ASD-specific DEGs.

Gandal 2022 performed RNA sequencing on 49 idiopathic or dup15q11-q13 ASD patients and 54 matched controls across frontal, parietal, temporal, and occipital cortex represented by 725 samples spanning 11 distinct brain regions. 3 of these brain regions were sequenced using single nucleus RNA-seq. They discovered 4.2k differentially expressed genes represented by 9.5k transcripts

[Arlotta, *et al.* 2022](https://www.nature.com/articles/s41586-021-04358-6) does scRNA-seq and whole proteome mass spectrometry on 745,000 cells from cortical organoids heterozygous for a known ASD-associated mutation—either *SUV420H1* (also known as *KMT5B*), *ARID1B*, or *CHD8*. No excitability perturbation apparently, so this is more of an atlas dataset. scRNA-seq, scATAC-seq, and proteomics data is available in unprocessed form.

Human post-mortem data

Lower priority

[Arlotta, *et al.* 2020](https://pubmed.ncbi.nlm.nih.gov/33243861/) does in vivo perturb-seq on autism related genes. Basically another atlas, I think.

**Analysis outline**

Note how much overlap in the genes *measured* there is between datasets, in addition to the overlap in the differentially expressed genes across the different datasets.

The overlap of the DEGs will constitute a crude homeostatic plasticity network of genes.

*What are more sophisticated ways of grouping together these genes*?

Look at overlap of genes across datasets with similar perturbations.

Look at overlap between datasets with similar biological models (anatomical location, in vivo, in vitro, species, etc.)*.*

Look at overlap between datasets with similar timepoints.

*Other dimensions of overlap*?

* GSEA on individual gene sets