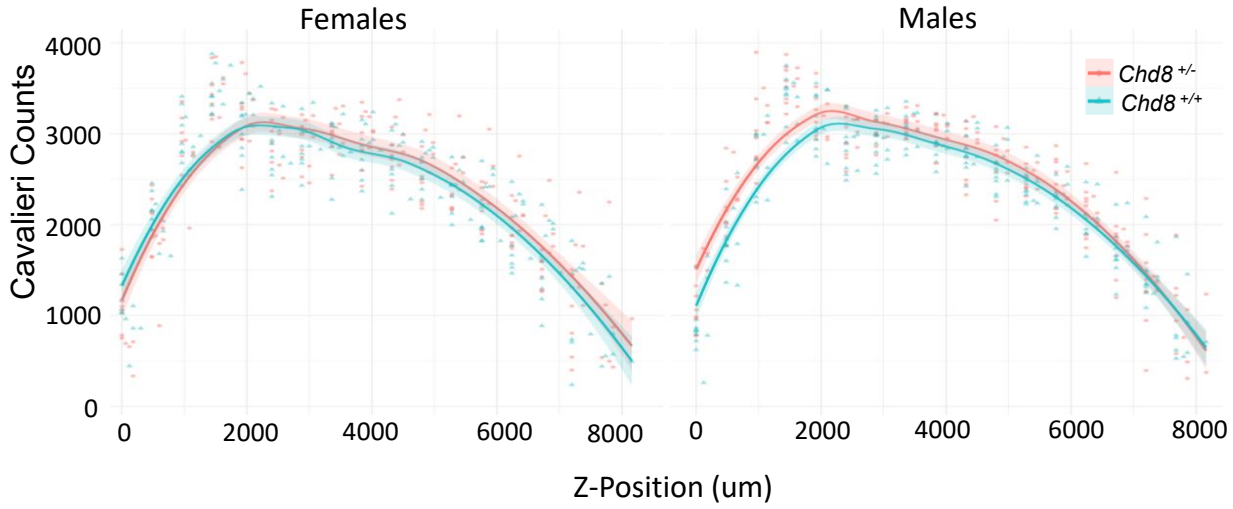
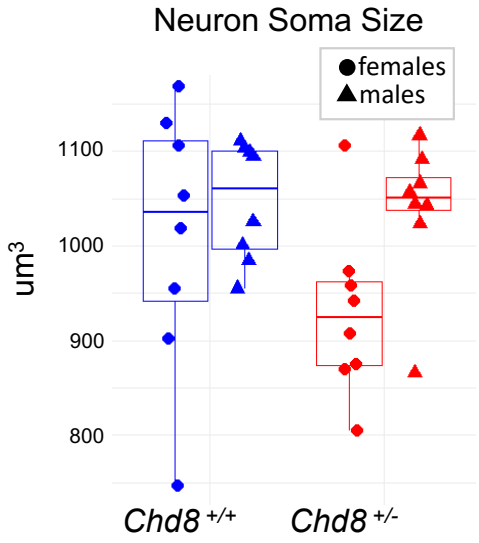
A


Smoothed Effect of Zpos on Cavalieri Counts
Females
Males
Cavalieri Counts
4000
3000
2000
1000
0
0
2000
4000
6000
8000
Chd8 +/-
Chd8 +/+
Z-Position (um)

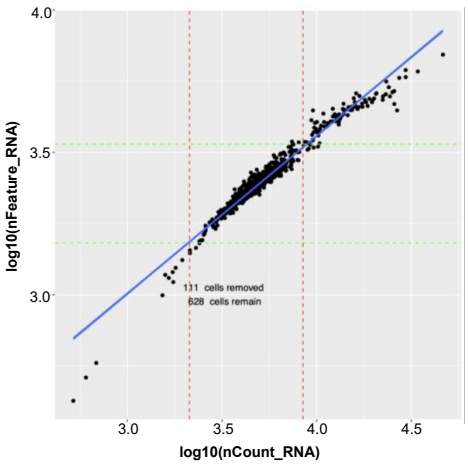
B


Neuron Soma Size
females
males
um3
1100
1000
900
800
Chd8 +/+
Chd8 +/-

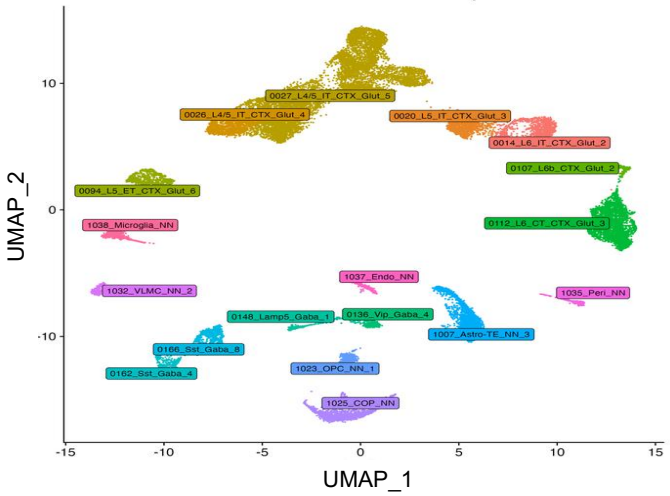
1

Figure S1: A) Cortical volume was quantified across rostral–caudal section in *Chd8*$^{+/-}$ and *Chd8*$^{+/+}$ mice using the Cavalieri method. In male *Chd8*$^{+/-}$ mice, increased volume was observed predominantly in more rostral regions, with clearer separation of genotype averages along the anterior Z-positions. Shaded portions of the plots indicate the more caudal regions (0–3000 µm), where volume differences are most evident. B) Neuronal soma size, measured by unbiased stereology in the same rostro-caudal range, did not differ significantly between genotypes. Data are stratified by sex and genotype. Error bars represent SEM. Statistical comparisons were performed using two-way ANOVA (unstratified: P = 0.23; males: P = 0.81; females: P = 0.19).

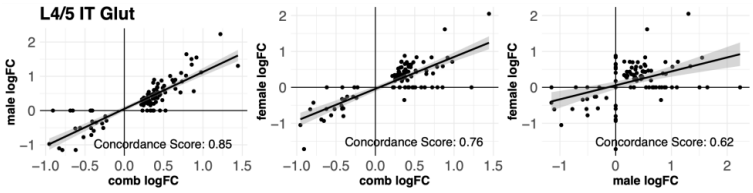
A Representative QC plot



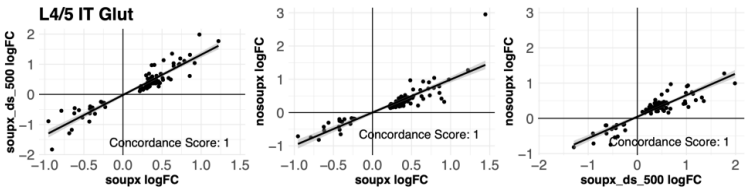
B Integration with BICCN data for cell type assignment



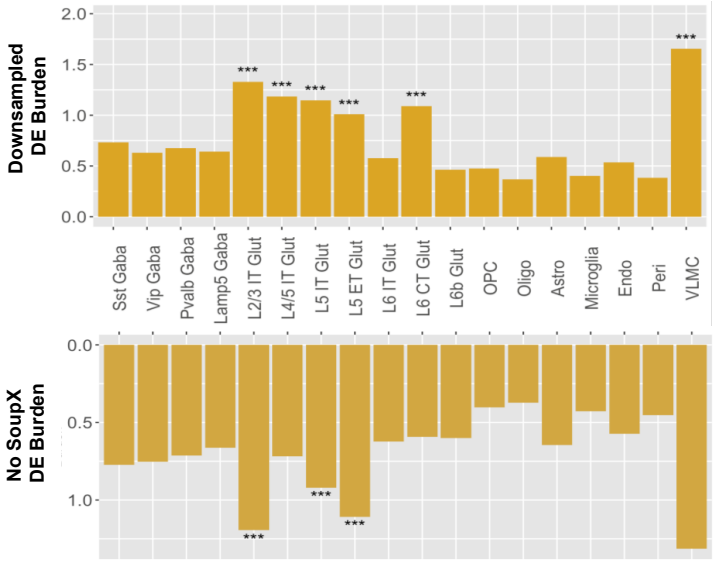
C Representative sex-stratified DE concordance



D Representative no soupx DE concordance



E Downsampled and nosoupx DE burden



F Pseudobulk DE GSEA

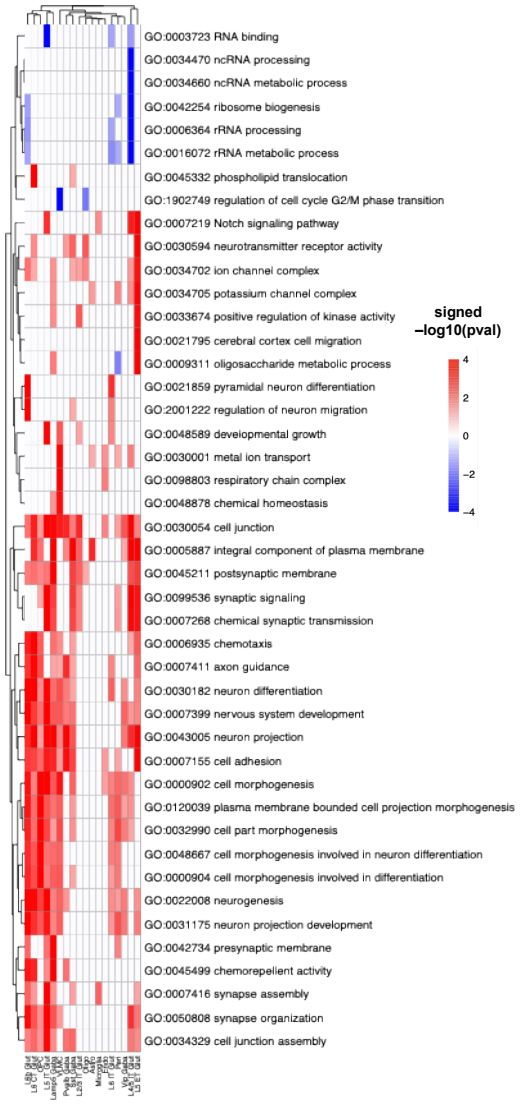
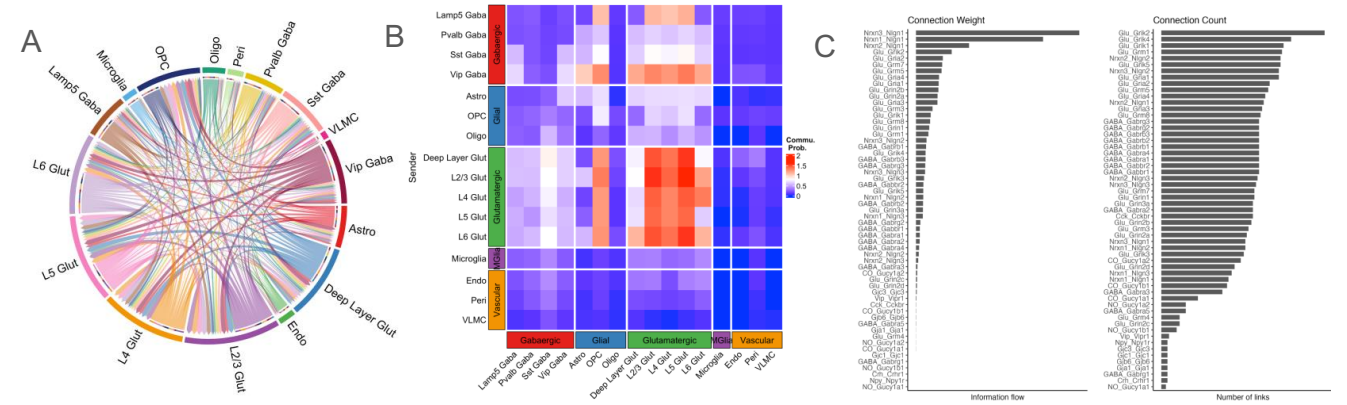
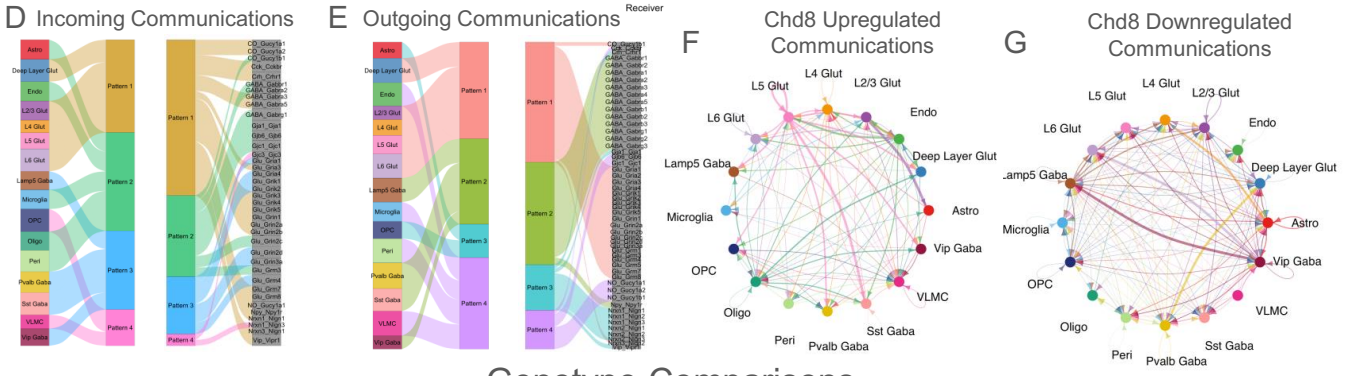


Figure S2: A) Scatterplot of log10-transformed total RNA counts (nCount_RNA) versus detected genes (nFeature_RNA) per cell in one batch. Custom MAD-based thresholds were chosen per batch to filter low-quality cells. Red and green dotted lines indicate nCount_RNA and nFeature_RNA cutoffs, respectively; cells outside these thresholds were excluded from downstream analysis. B) UMAP showing all *Chd8* mutant and wild-type nuclei integrated with BICCN data and annotated with predicted BICCN cell type labels, which are consistent with cell type identities derived from our independent annotation pipeline. C) Scatterplots comparing log fold changes (logFC) of the top 100 DEGs identified with full SoupX, non-downsampled DE analysis ("soupx") to their logFCs in the no-SoupX ("nosoupx") and downsampled to 500 nuclei + SoupX ("soupx_ds_500") analyses for a representative cell type. Concordance scores were calculated as the proportion of genes with the same direction of change between methods. D) Scatterplots comparing logFC of the top 100 DEGs identified with male-and-female-combined ("comb") DE analysis to their logFCs in the male and female DE analyses for a representative cell type. E) Bar plots of DEG burden across cell types (Burden = DE genes / total genes expressed * 100) for downsampled (top) and no-SoupX (bottom) DE analyses. A permutation test with 10,000 iterations was used to assess statistical significance and confirmed that the observed trends were similar to those in the full SoupX, non-downsampled analysis. F) Heatmap of normalized enrichment scores (NES) for a curated set of biologically relevant GO terms identified by GSEA performed on pseudobulk DEGs for each cell type.





Genotype Comparisons

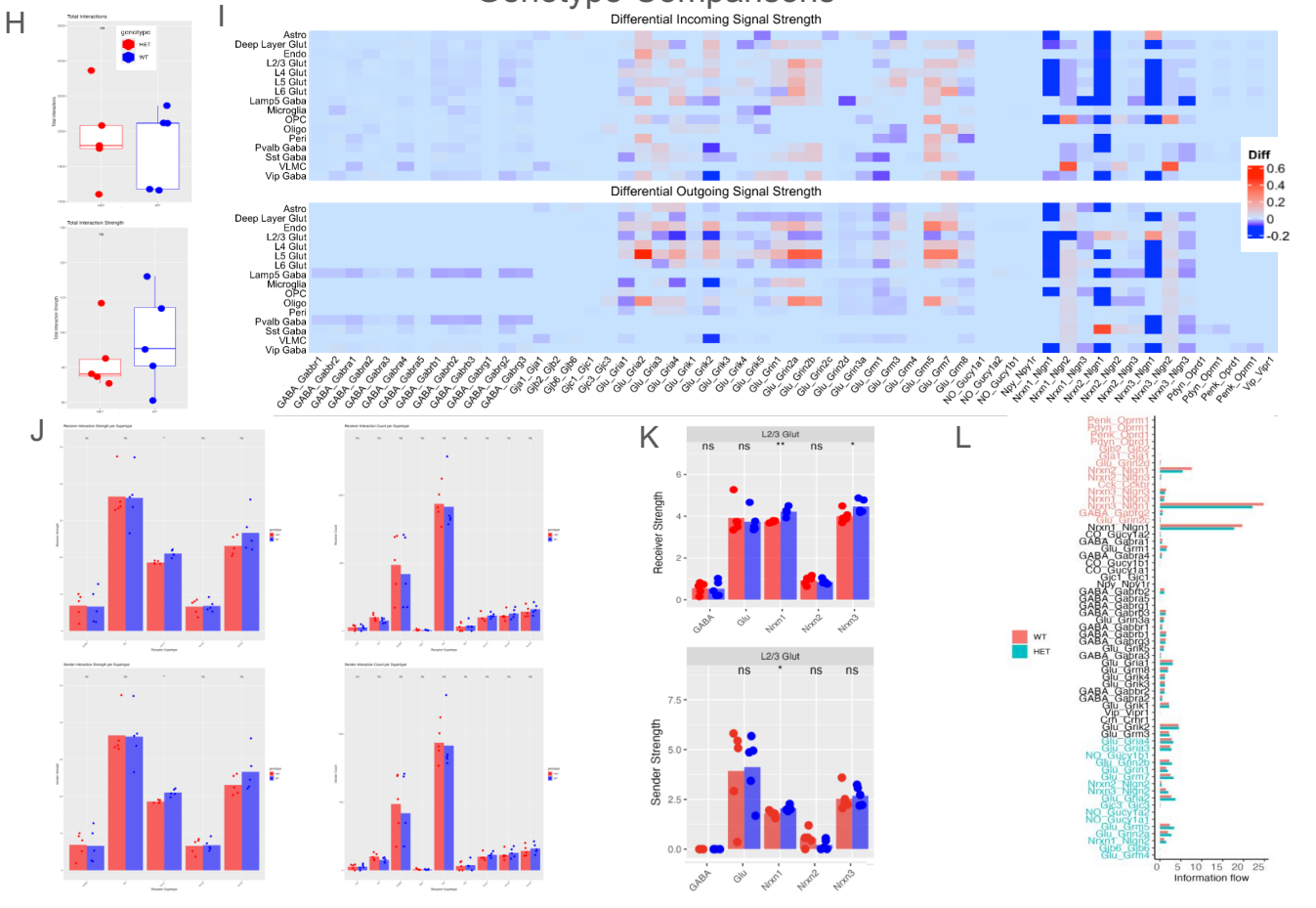


Figure S3: A) Overview of all inferred connections by cell type shows generally balanced co-directional signaling for each cell type. B) The communication probability is strongest within Glutamatergic cell types and between Glutamatergic and GABAergic populations. C) The strength and count of connections anchored by the listed receptor highlight networking drivers. D, E) Cell types are grouped based on the similarity of their incoming or outgoing connections showing similarities correlated with similar cell identities and neurotransmitter types. F, G) Circle plots showing the upregulated (F) and downregulated (G) intercellular communications in *Chd8*$^{+/-}$ compared to control. The width of each link indicates the absolute difference between Chd8 and control, in the sum of communication strength values over all interaction pairs. H) The relative strength (top) of all network connections is similar in both WT and *Chd8*$^{+/-}$ ("HET") animals but the number of connections (bottom) shows some elevation in the *Chd8*$^{+/-}$ animals. I) Heatmap showing the differential outgoing (upper panel) and incoming (lower panel) signal strength between *Chd8*$^{+/-}$ and control. Given the cell group and interaction pair, the outgoing (or incoming) signal strength is defined as the sum of communication strength over the links from (or to) the cell group. The color bar indicates the difference in the outgoing (or incoming) signal strength values between *Chd8*$^{+/-}$ and control. Interaction pairs with signal strength of 0 for all cell types were omitted from plotting. J) Replicate level quantification of receptor supertype sender (top) and receiver (bottom) differences shows dysregulation only in the Nrxn1 strength metrics on the level of the entire dataset. K) Further cell type specific inspection indicates that the L2/3 population is the only significantly affected population driving the dysregulation. L) Bar chart comparing the information flow between *Chd8*$^{+/-}$ and control for each interaction pair at the full dataset level. While the HET samples show both decreased and increased information flow in some interaction pairs, those that are increased have a vastly greater magnitude thus dominating the effects. In contrast, the increased information flow seen in the glutamate interactions is present in connections with decreased overall magnitude but is considerably consistent across each glutamate-receptor-ligand pair and reveals the glutamate network dominating the rankings of information flow increases in the HET.

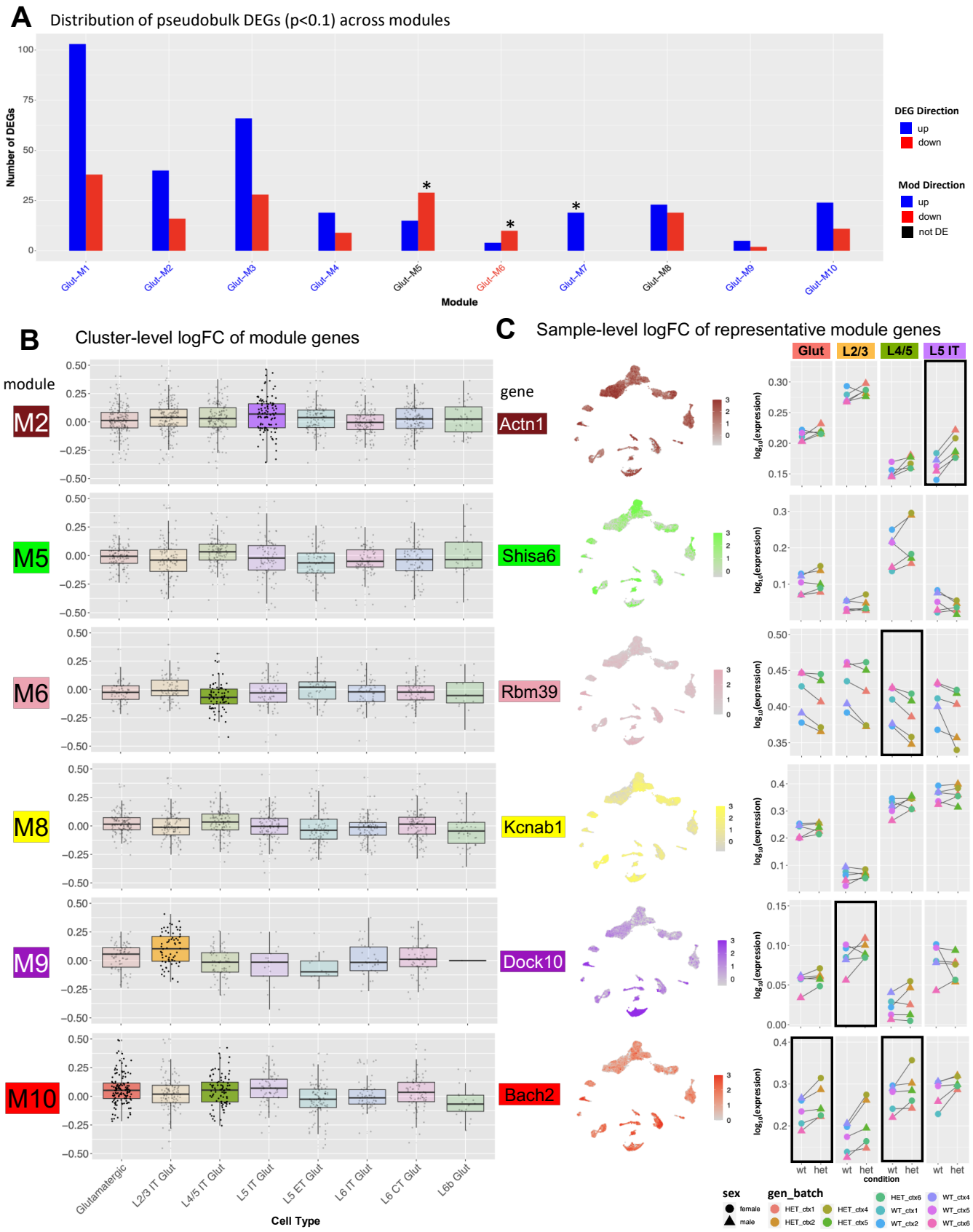
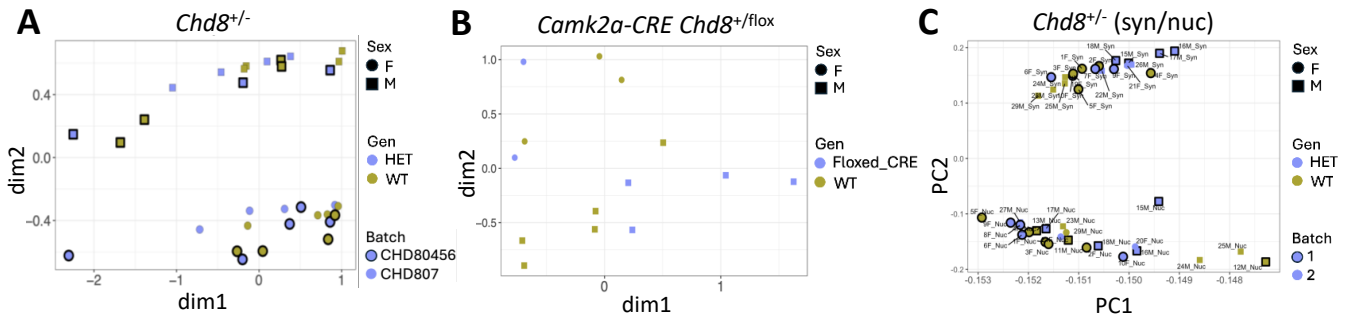
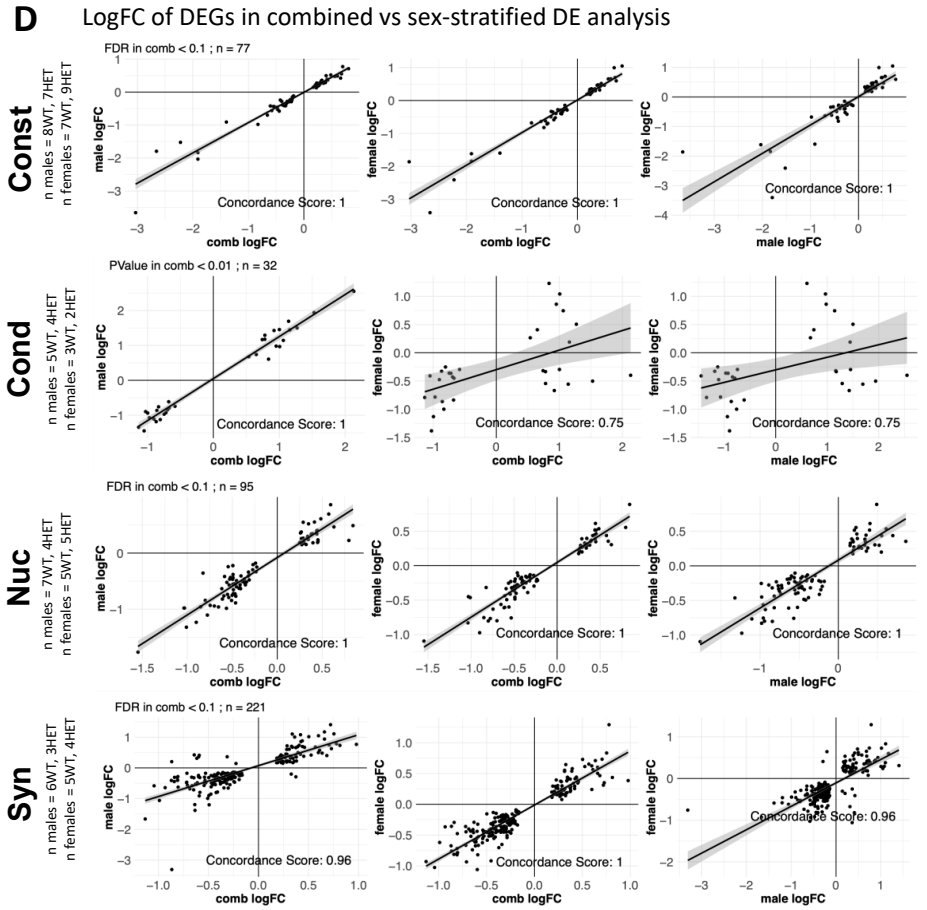
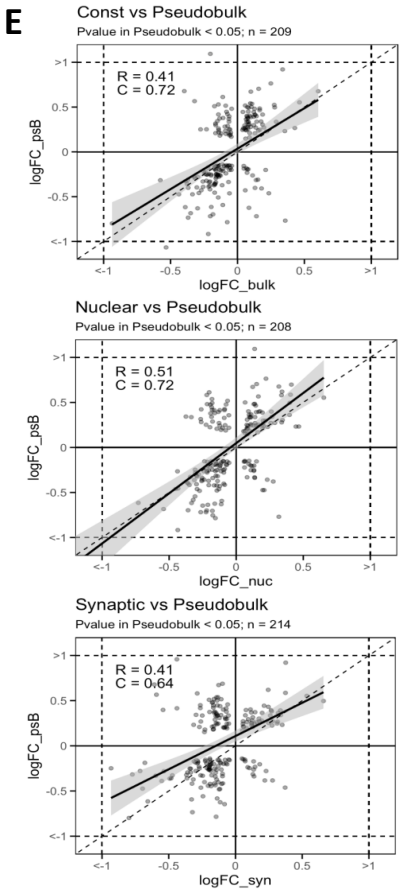
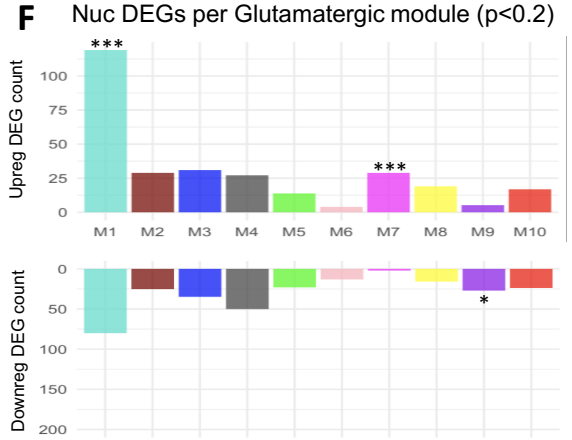

A
Distribution of pseudobulk DEGs (p<0.1) across modules
Number of DEGs
Module
Glut-M1
Glut-M2
Glut-M3
Glut-M4
Glut-M5
Glut-M6
Glut-M7
Glut-M8
Glut-M9
Glut-M10
DEG Direction
up
down
Mod Direction
up
down
not DE
B
Cluster-level logFC of module genes
module
M2
M5
M6
M8
M9
M10
Glutamatergic
L2/3 IT Glut
L4/5 IT Glut
L5 IT Glut
L5 ET Glut
L6 IT Glut
L6 CT Glut
L6b Glut
Cell Type
C
Sample-level logFC of representative module genes
gene
Actn1
Shisa6
Rbm39
Kcnab1
Dock10
Bach2
Glut
L2/3
L4/5
L5 IT
log10(expression)
wt
het
condition
sex
female
male
gen_batch
HET_ctx1
HET_ctx2
HET_ctx4
HET_ctx5
HET_ctx6
WT_ctx1
WT_ctx2
WT_ctx4
WT_ctx5
WT_ctx6

Figure S4: A) Bar plot showing the distribution of DEGs ($p < 0.1$) across glutamatergic modules. Bars indicate the number of genes upregulated (blue) or downregulated (red) in at least one cell type per module. Modules Glut-M5 and Glut-M6 showed significant enrichment for downregulated DEGs (Fisher's exact test, FDR-adjusted $*p < 0.05$), while Glut-M7 showed significant enrichment for upregulated DEGs. Overall, the direction of module-level change tends to agree with the predominant direction of DE for the genes within each module. B) Box plots for the remaining glutamatergic modules, showing logFCs of module genes across cell types. Opaque boxes indicate subtypes where the module is significantly differentially expressed (Student's t-test comparing mean logFC of each module to grey module, Bonferroni-adjusted $p < 0.05$). C) Feature plots (left) show relative expression of representative genes from each module in UMAP space. Dot plots (right) display median sample-level expression of the same genes. Black boxes highlight subtypes where the module is differentially expressed or where consistent directional changes are observed across batches.


A
Chd8+/-
dim1
dim2
Sex
F
M
Gen
HET
WT
Batch
CHD80456
CHD807
B
Camk2a-CRE Chd8+/flox
dim1
dim2
Sex
F
M
Gen
Floxed_CRE
WT
C
Chd8+/- (syn/nuc)
PC1
PC2
Sex
F
M
Gen
HET
WT
Batch
1
2


D
LogFC of DEGs in combined vs sex-stratified DE analysis
Const
n males = 8WT, 7HET
n females = 7WT, 9HET
FDR in comb < 0.1 ; n = 77
Concordance Score: 1
Concordance Score: 1
Concordance Score: 1
Cond
n males = 5WT, 4HET
n females = 3WT, 2HET
PValue in comb < 0.01 ; n = 32
Concordance Score: 1
Concordance Score: 0.75
Concordance Score: 0.75
Nuc
n males = 7WT, 4HET
n females = 5WT, 5HET
FDR in comb < 0.1 ; n = 95
Concordance Score: 1
Concordance Score: 1
Concordance Score: 1
Syn
n males = 6WT, 3HET
n females = 5WT, 4HET
FDR in comb < 0.1 ; n = 221
Concordance Score: 0.96
Concordance Score: 1
Concordance Score: 0.96
comb logFC
male logFC
female logFC


E
Const vs Pseudobulk
Pvalue in Pseudobulk < 0.05; n = 209
R = 0.41
C = 0.72
logFC_psB
logFC_bulk
Nuclear vs Pseudobulk
Pvalue in Pseudobulk < 0.05; n = 208
R = 0.51
C = 0.72
logFC_psB
logFC_nuc
Synaptic vs Pseudobulk
Pvalue in Pseudobulk < 0.05; n = 214
R = 0.41
C = 0.64
logFC_psB
logFC_syn


F
Nuc DEGs per Glutamatergic module (p<0.2)
Upreg DEG count
Downreg DEG count
M1
M2
M3
M4
M5
M6
M7
M8
M9
M10

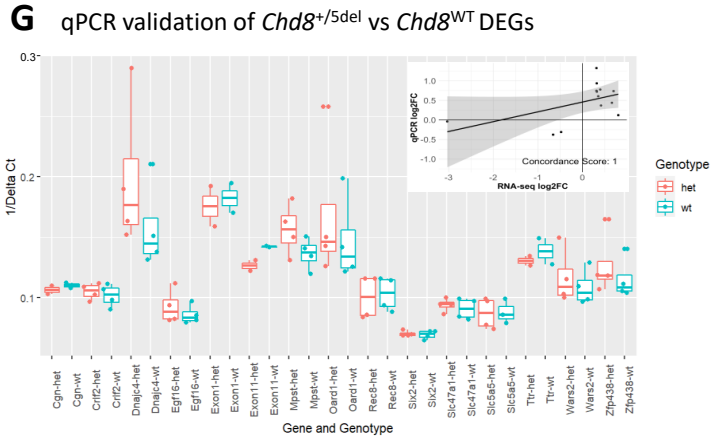

G
qPCR validation of Chd8+/5del vs Chd8WT DEGs
1/Delta Ct
Concordance Score: 1
qPCR logFC
RNA-seq log2FC
Genotype
het
wt
Gene and Genotype

Figure S5: A-B) MDS plot of cortical *Chd8* mutant (A: *Chd8*$^{+/-}$, B: *Camk2a-CRE Chd8*$^{+/flox}$) and WT samples, with the the second dimension ("dim2") reflecting biological sex. C) PCA plot of cortical *Chd8*$^{+/-}$ and *Chd8*^{WT} synaptic and nuclear samples, where leading dimensions distinguish sample fraction. D) Log fold change (logFC) concordance scatterplots comparing logFC of DEGs from the combined ("comb") analysis to their logFCs in each of the sex-stratified analyses ("male", "female"). Only genes passing FDR < 0.1 or P < 0.01 in combined datasets were used for comparisons. Each row represents one of 4 bulk RNA-seq datasets: constitutive *Chd8*$^{+/-}$ ("Const"), conditional *Camk2a-CRE Chd8*$^{+/flox}$ ("Cond"), nuclear fraction of *Chd8*$^{+/-}$ ("Nuc"), and synaptic fraction of *Chd8*$^{+/-}$ ("Syn"). Plots display gene-level logFC correlations between pairs of analyses with linear regression fits and concordance scores indicating the fraction of genes regulated in the same direction. E) LogFC scatterplots comparing constitutive *Chd8*$^{+/-}$ ("Const") to pseudobulk snRNA-seq DEGs (top), nuclear fraction *Chd8*$^{+/-}$ ("Nuc") to pseudobulk DEGs (middle), and synaptic fraction *Chd8*$^{+/-}$ ("Syn") to pseudobulk DEGs. Genes were selected based on significance in pseudobulk analysis (P < 0.05) and differential expression in each respective bulk RNA-seq analysis. Pearson correlation coefficients (R) and concordance scores (C) are shown on each plot. F) Bar plots showing counts of nuclear differentially expressed genes (DEGs) per module, separated into upregulated (top) and downregulated (bottom). Enrichment was assessed using Fisher's exact test, comparing the proportion of up- vs. downregulated DEGs in each module relative to all other modules, with multiple testing correction (Benjamini-Hochberg). M1 and M7 were significantly enriched for upregulated DEGs (p*** < 0.001), while M9 was enriched for downregulated DEGs (p* < 0.05). G) Box plots showing the median 1/ΔCt (relative to GAPDH) across biological replicates for 13 constitutive *Chd8*$^{+/-}$ bulk RNA-seq DEGs. *Chd8* Exons 1 and 11 ("Exon1", "Exon 11") were included as positive controls. A Wilcoxon text showed nonsignificant differences in 1/ΔCt between *Chd8* mutants and wild-types, all 13 genes tested showed the same direction of change as RNA-seq DE analysis. Inset shows scatterplot of RNA-seq logFC versus qPCR logFC, with concordance calculated as in (D-E).