

Cell Genomics, Volume 3

Supplemental information

**Multiple genes in a single GWAS risk locus
synergistically mediate aberrant synaptic
development and function in human neurons**

Siwei Zhang, Hanwen Zhang, Marc P. Forrest, Yifan Zhou, Xiaotong Sun, Vikram A. Bagchi, Alena Kozlova, Marc Dos Santos, Nicolas H. Piguel, Leonardo E. Dionisio, Alan R. Sanders, Zhiping P. Pang, Xin He, Peter Penzes, and Jubao Duan

**Supplemental Items of
Multiple genes in a single GWAS risk locus synergistically mediate aberrant synaptic
development and function in human neurons.**

Include

Figure S1-S8

Table S1-S15 (as separate Microsoft Excel files)

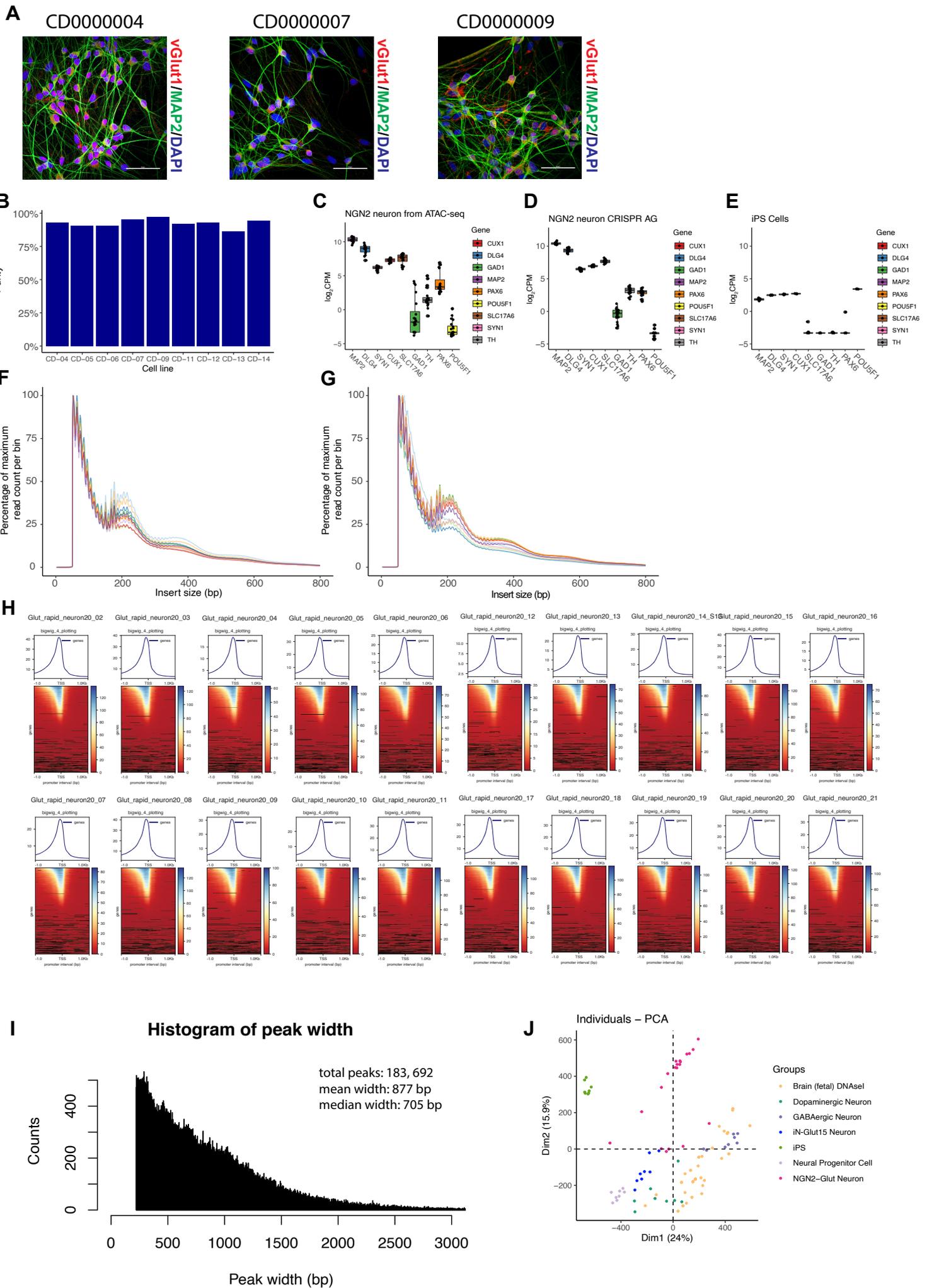


Figure S1: Morphological characterization and QC of hiPSC-derived NGN2-Glut neurons.

Related to Figure 1. (A) Representative immunofluorescence (IF) images of several NGN2-Glut lines stained with antibodies against glutamatergic cell-type-specific markers. Scale bar: 50 μm . (B) Cell line purity statistics show the percentages of NGN2-Glut cells stained positive with type-specific markers (vGLUT1+/MAP2+). Approximately ~200 cells per line were used for quantification. (C-E) The expression level of neuronal and iPSC signature genes in NGN2-Glut neurons used for ATAC-seq, CRISPR/Cas9 editing, as well as their corresponding donor iPS cells used to differentiate the NGN2-Glut neurons in the study from bulk RNA-seq data. (C) NGN2-Glut neurons used for ATAC-seq. (D) NGN2-Glut neurons from CRISPR/Cas9 editing. (E) Undifferentiated iPS cells. (F-G) The histogram of the insert sizes of all ATAC-seq samples shows the expected distribution pattern of transposase insertion at specific sizes (in bp; x-axis). 20 NGN2-Glut samples were plotted in two groups (10 each) for better visual resolution. (H) Chromatin accessibility in regions proximal to the TSS of core promoters (± 1 kb) in each sample, showing the library size-normalized overall ATAC-seq read intensity (top) and per-gene heatmap (bottom). The strong signal-to-noise ratio of the ATAC-seq peak intensity around the TSS site indicated a good quality of ATAC-seq data. Color bar indicates the density of reads at each position proximal to the TSS site. Black regions represent the absence of a signal at the corresponding coordination. (I) The histogram of all ATAC-seq peak distribution. X-axis: peak width in bp. (J) PCA of OCRs taken from NGN2-Glut, iPSC, the four iNs previously reported (Zhang *et al.*, 2020), and fetal brain DNaseI results, showing the relative distribution of NGN2-Glut neuron against other cell types.

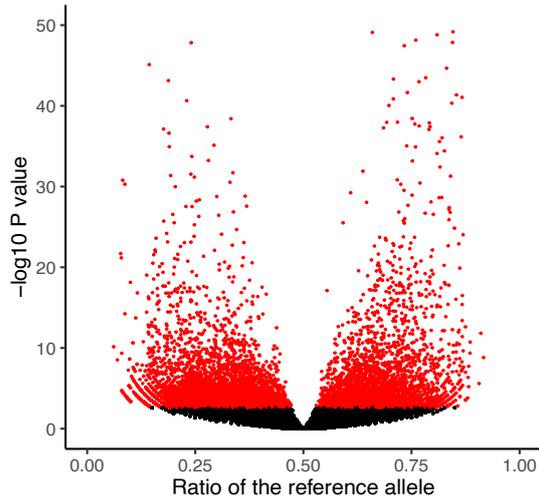
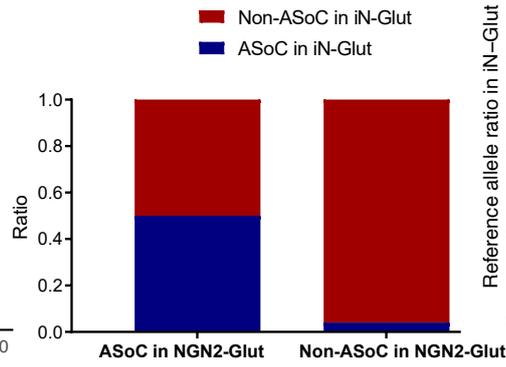
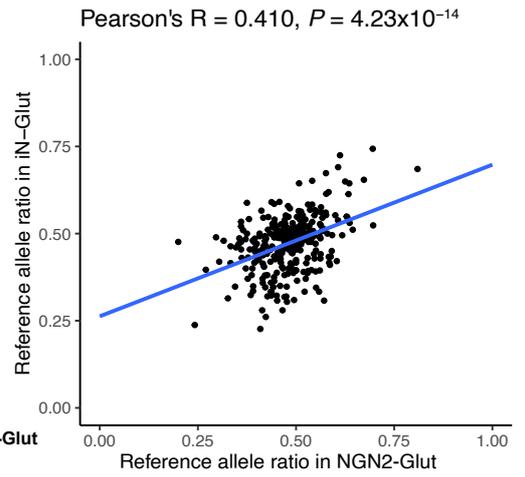
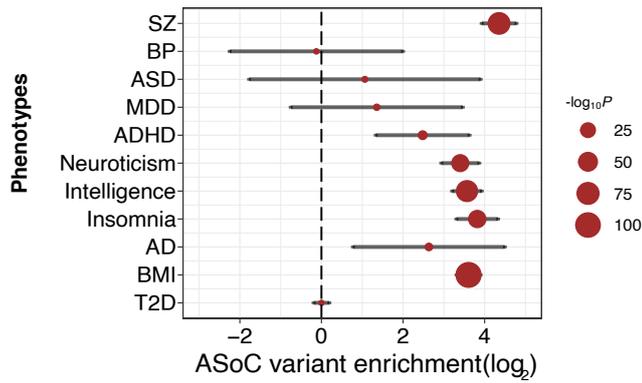
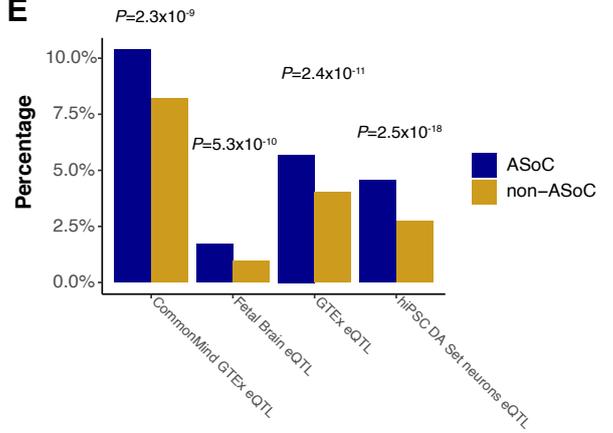
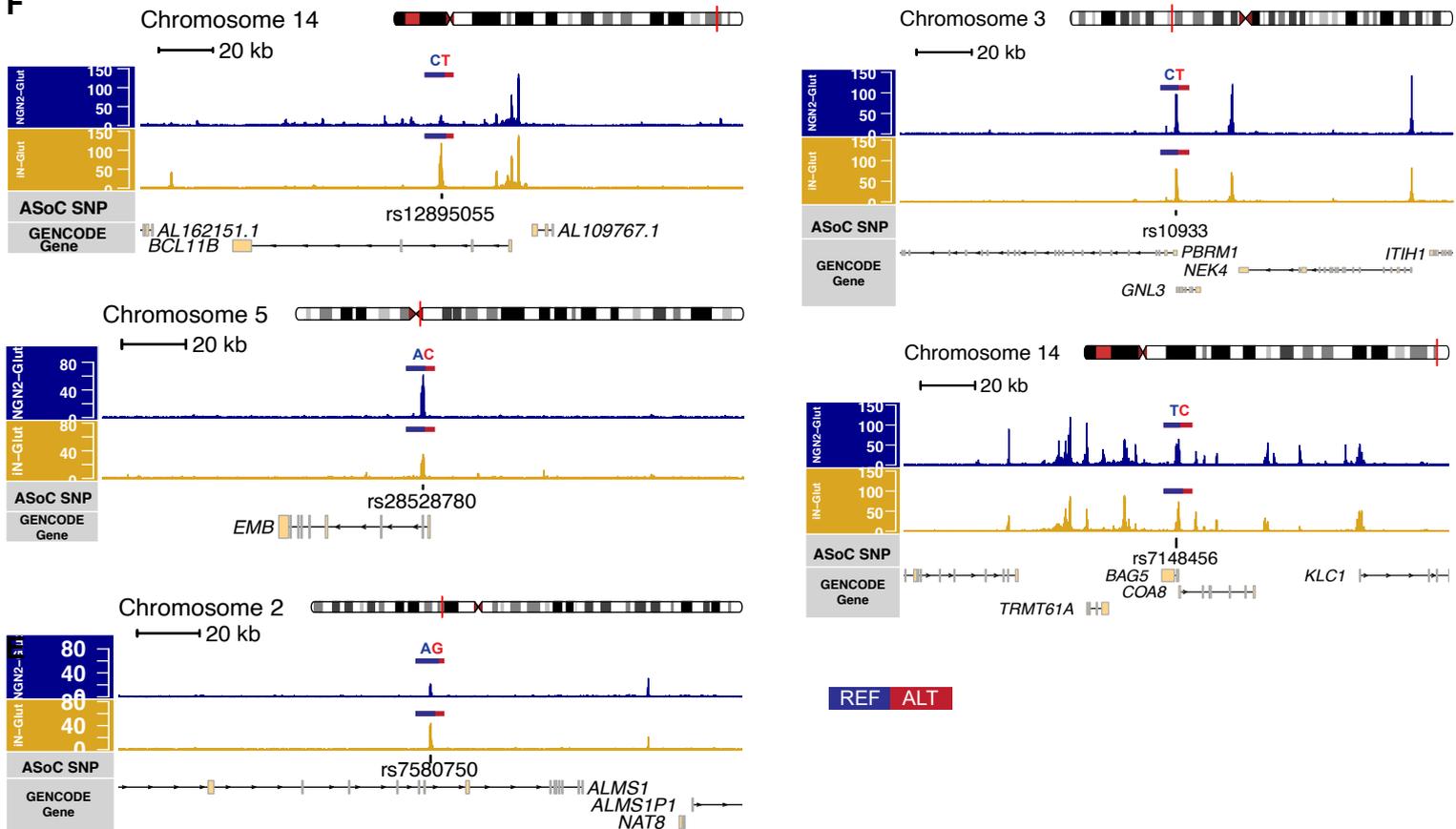
A**B****C****D****E****F**

Figure S2: Allele-specificity in NGN2-Glut neurons. Related to Figure 1D, Table S2, S3. (A) Volcano plot showing the reference allelic ratio of all heterozygous SNPs (*x*-axis) and their corresponding *p*-values of ASoC testing (*y*-axis; in $-\log_{10}$ scale) from 20 NGN2-Glut samples. Red: all ASoC SNPs (*FDR* < 0.05). (B) Bar graph depicting the enrichment of overlapping ASoC SNPs associated with SZ in NGN2-Glut and iN-Glut data sets. Chi-square test, *P* < 0.001. (C) Allelic ratio correlation of SZ credible SNPs between NGN2-Glut and iN-Glut neurons. (D) Multi-traits GWAS enrichment analysis of ASoC SNPs by TORUS. (E) Estimation of the overlapping and enrichment of ASoC/non-ASoC SNPs for different brain/neuron eQTL sets. (F) ATAC-seq read pileup and gene tracks proximal to five SZ-credible SNP sites in NGN2-Glut and iN-Glut neurons. The reference and alternate allele ratio have been marked over each SNP as the length of color bars, showing high concordance between the two neuron types.

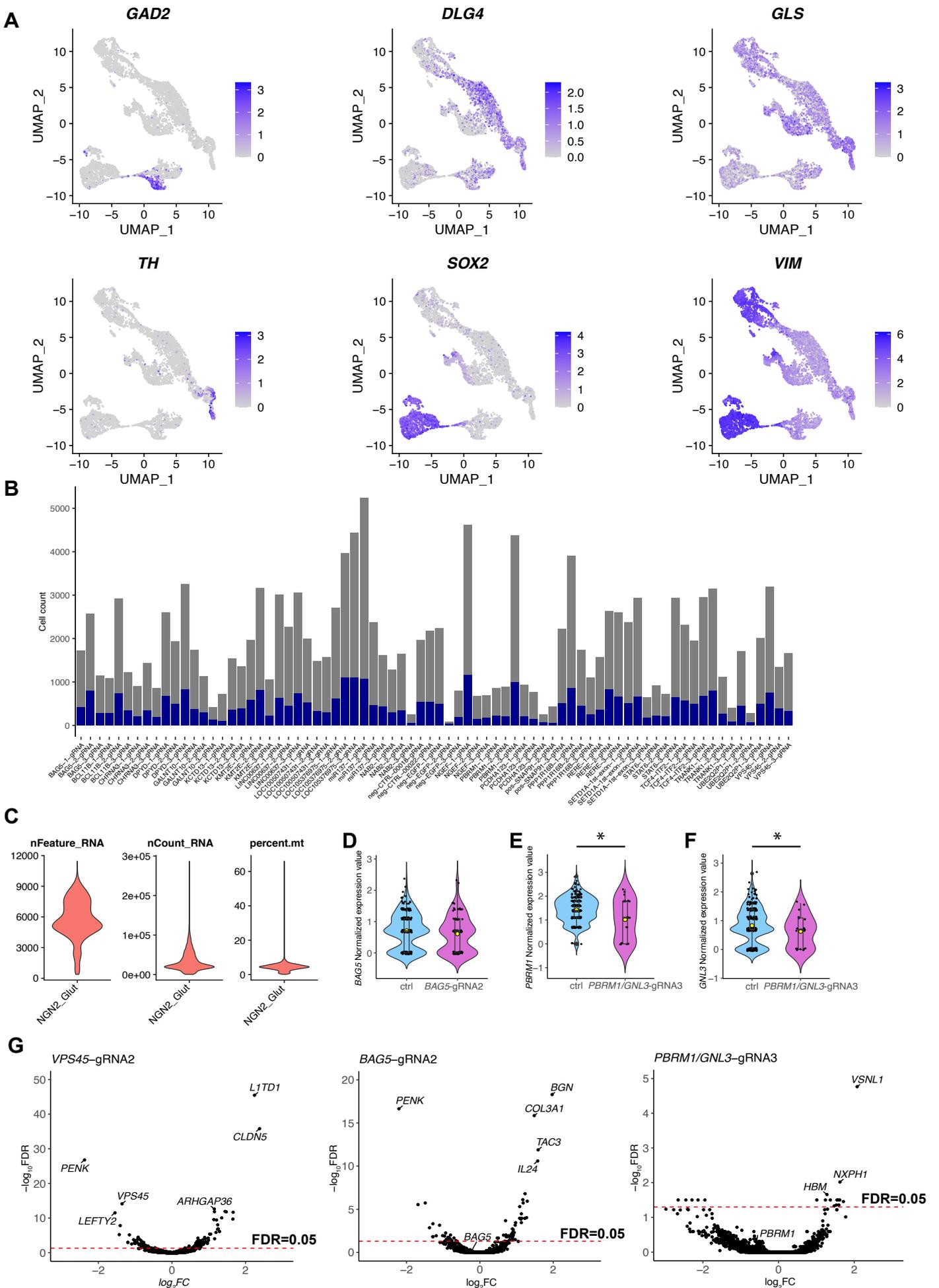


Figure S3: Quality control summary of CROP-seq experiment in NGN2-Glut neuron.

Related to Figure 2. (A) Expression feature plots showing additional cell type-specific markers from the output of the scRNA-seq result. (B) Bar graph showing the number of cells (*y*-axis) that express a specific gRNA (*x*-axis) and their gRNA contents. Blue bar: cells were assigned with a unique gRNA and used for differential gene expression analysis. Five negative control gRNAs were used (neg_CTRL_0018, neg_CTRL_0022, neg_EGFP_1, neg_EGFP_2, and neg_EGFP_3). (C) General QC data output from Seurat after merging the two capture data sets and performing data normalization with SCTransform(). (D) Quantification on the knockdown efficiency of *BAG5* (rs7148456) CROP-seq gRNA on the expression level of *BAG5* gene. Student's *t*-test (two-tailed, non-parametric). (E-F) Quantification on the knockdown efficiency of *PBRM1/GNL3* (rs10933) CROP-seq gRNA on the expression level of either *PBRM1* (E) or *GNL3* (F) gene. Student's *t*-test (two-tailed, non-parametric). All five negative control gRNAs were used in quantification for the control group. (G) CRISPRi using gRNAs targeting specific ASoC loci and their transcriptomic effects as visualised by differentially expressed genes (DEGs).

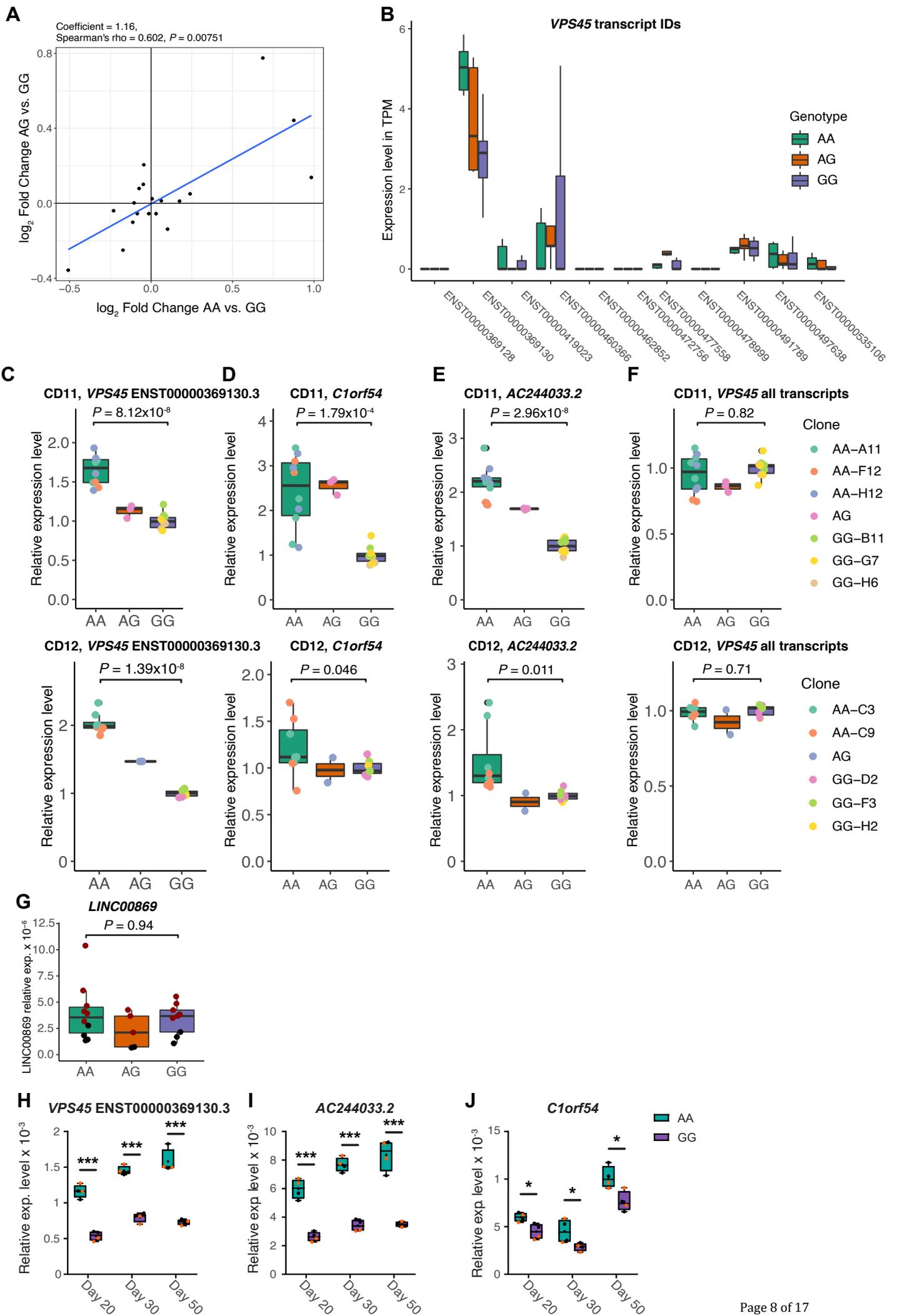


Figure S4: Cis-effects of SNP editing at the *VPS45* locus in NGN2-Glut of the rs2027349-edited isogenic lines. Related to Figure 3. (A) Cis-effect with 500 kb interval of the edited SNP site (rs2027349), demonstrating a high correlation on the DE gene expression patterns between AA/GG vs. AG/GG. (B) Comparison of the expression level of different *VPS45* transcripts in rs2027349 edited NGN2-Glut neurons from bulk RNA-seq data (2 cell lines, 2-3 clones per genotype). Note the high expression level of main transcript ENST00000369130.3. (C-F). qPCR validation of *VPS45*-proximal DE genes and transcripts identified from RNA-seq results. Points of different colors indicate biological replicates from different clones. (C) *VPS45* transcript ENST00000369130.3. (D) *C1orf54*. (E) *AC244033.2*. (F) All *VPS45* transcripts. Two donor lines (CD0000011 and CD0000012) and 2-3 biological replicates per clone per genotype were used for each donor line. (G) qPCR validation of control gene *LINC00869*. Points of different colors indicate two donor lines. 2-3 replicates per clone (1-2 clones per genotype) were used. (H-J) The expression level of *VPS45* transcript ENST00000369130.3, *AC244033.2*, and *C1orf54* in day 20, 30, and 50 NGN2-Glut neurons of different rs2027349 genotypes (AA/GG). Points of different colors indicate two clones (CD0000011). Two biological replicates per clone were used. Kruskal-Wallis test (two-tailed, non-parametric).

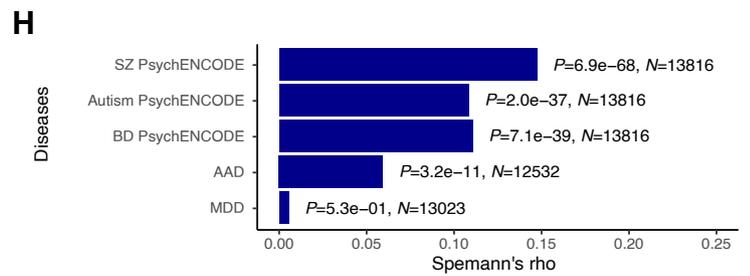
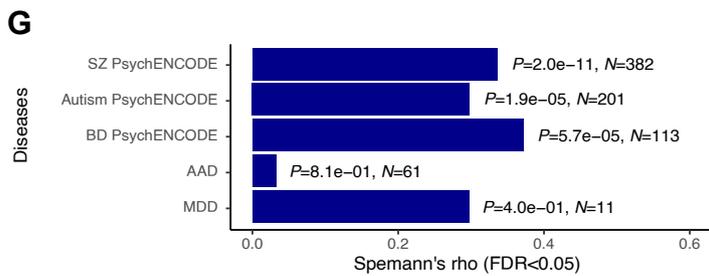
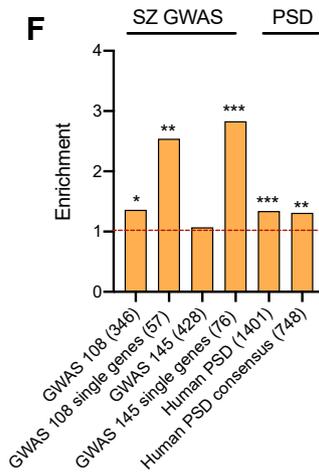
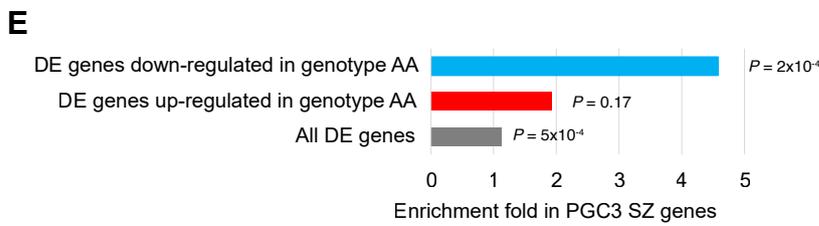
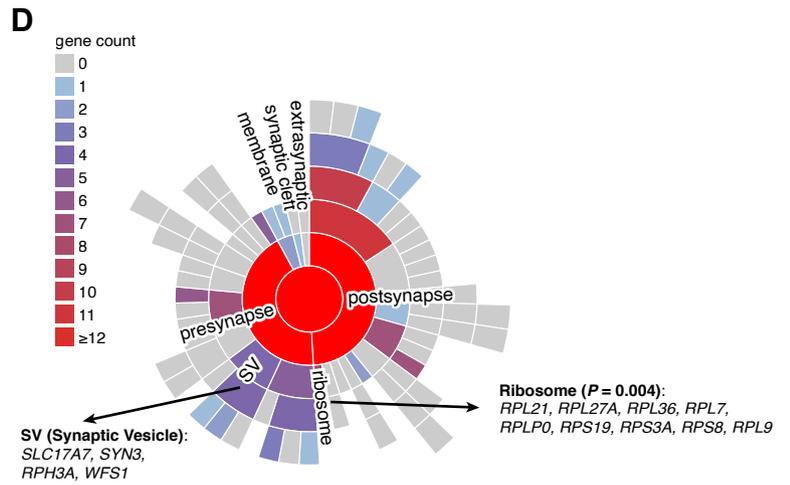
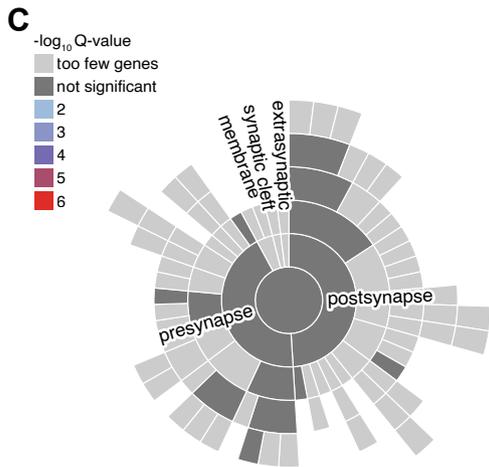
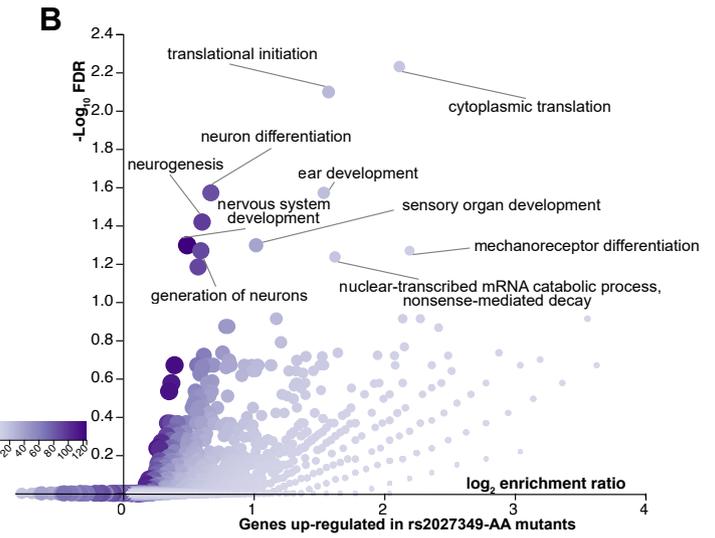
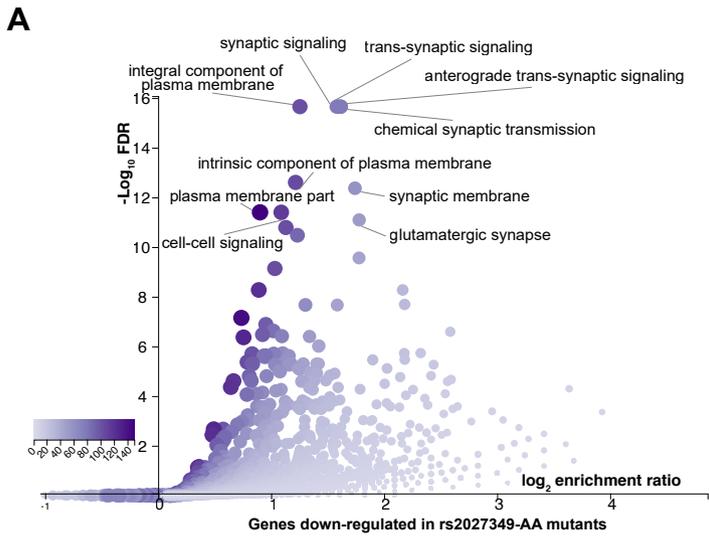


Figure S5: Transcriptomic gene pathway and disease relevance analyses in NGN2-Glut of the rs2027349-edited isogenic lines. Related to Figure 3, Table S8, Table S10. (A-B) GO term analysis of DE genes in rs2027349-edited neurons. Genotype GG was used as the expression baseline. (A) Genes downregulated in rs2027349-edited AA neurons. (B) Genes upregulated in rs2027349 AA neurons. (C) Sunburst plot showing the enrichment, albeit insignificant, of synaptic-related terms among the upregulated DE genes in rs2027349 AA neurons. (D) Further dissection of (C) using the gene count as the coloring scheme, shows an excess number of synaptic vesicle and ribosome-related terms. (E) GSEA result of DE genes in rs2027349 AA neurons for PGC3 SZ GWAS risk genes. (F) Enrichment of DE genes in rs2027349 AA neurons for several SZ-related gene lists. Fisher's Exact Test was used for the enrichment analyses in (E) and (F). (G) Correlation of the \log_2FC of DE genes ($FDR < 0.05$) from rs2027349 editing (AA vs. GG neurons) in NGN2-Glut neurons against the overlapping DE genes ($FDR < 0.05$) in post-mortem brains of multiple neuropsychiatric disorders. (H) Similar to (G), except using all the expressed genes. Spearman's ρ test, two-tailed.

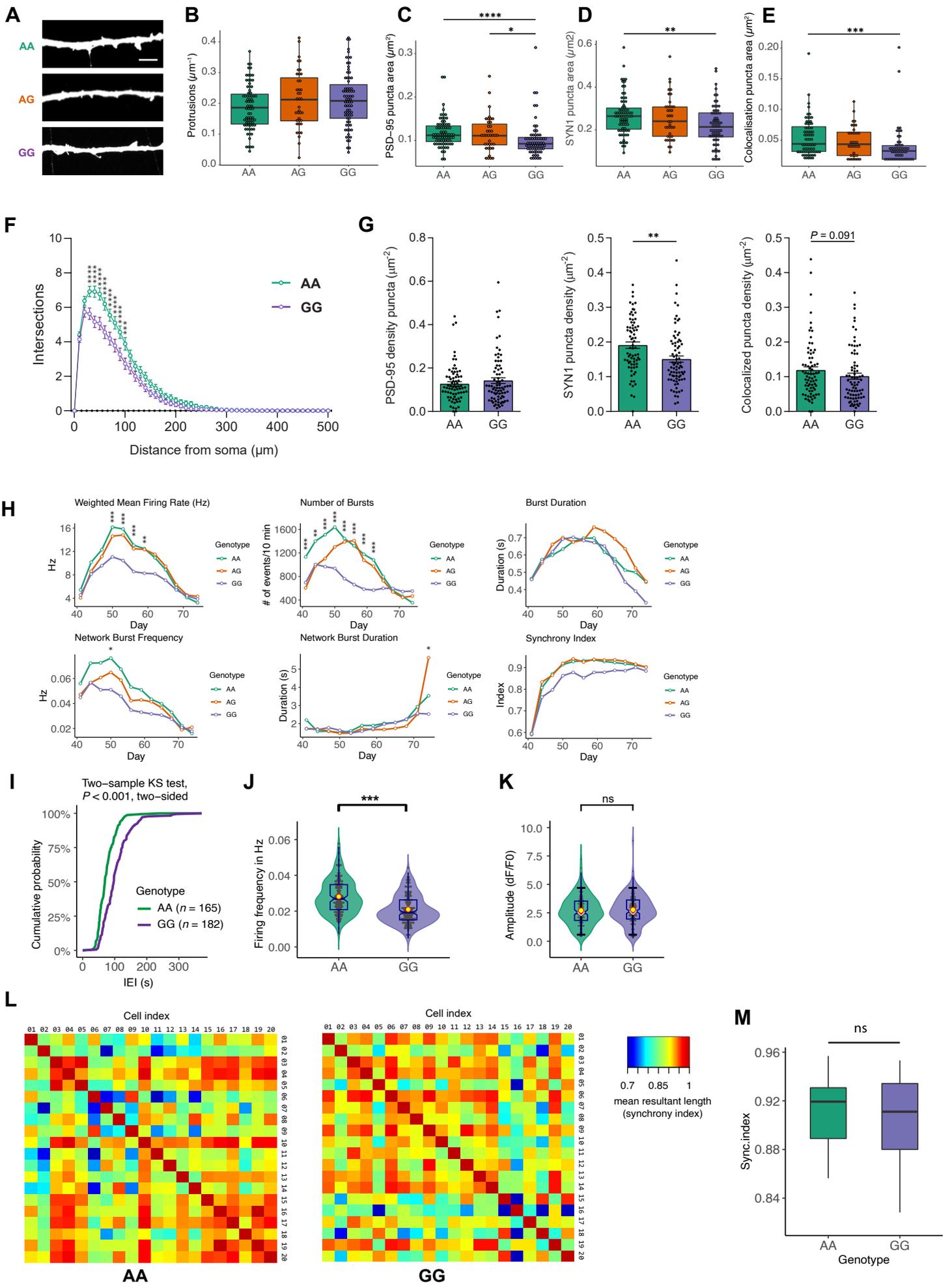


Figure S6: Protrusion density, puncta size, and Ca²⁺ signaling analysis in NGN2-Glut of the rs2027349-edited isogenic lines. Related to Figure 4, Figure 5. (A) Representative immunofluorescence images of dendrites and their associated neural protrusions in excitatory neurons; scale bar: 5 μ m. (B) Statistics of protrusion density in three genotypes show that genotype does not affect protrusion density. (C-E) Puncta size analysis. Data were generated from line CD0000011 of three independent experiments. Two clones per genotype were used. (C) Puncta size in different genotypes as measured by PSD-95 signal. (D) Puncta size in different genotypes as measured by SYN1 signal. (E) Puncta density in different genotypes as measured by the colocalization of PSD-95:SYN1 signal, showing the AA risk allele is associated with larger neural puncta size. In all cases, Kruskal-Wallis (two-tailed, non-parametric) test was used. (F-G) Similar to Figure 4D-F. Data were generated from a second donor line (CD0000012) and collated from two independent experiments. Two clones per genotype were used. (H) Complete time-series (days 41-74) data from MEA assay of all six parameters measured. The significance level between AA and GG genotypes was indicated on the graphs. Two-way ANOVA with *post-hoc* correction for multiple testing. 2-3 clones per genotype derived from CD0000011 were used. (I-K) Similar to Figure 5H-J. Data were generated from a second donor line (CD0000012) and collated from two independent experiments. Two clones per genotype were used. (L) Representative correlation matrix heatmap of neural activity synchrony index of rs2027349 AA and GG neurons. For each matrix heatmap, 20 neurons were randomly sampled from Ca²⁺ imaging data and the spike train distance model was applied to their detected peak series to generate the synchrony index heatmap (represented by the calculated mean resultant length). (M) The box-whisker plot of the correlation matrix in (L) with orthogonal elements removed, ns: not significant.

Figure S7: Additional data showing that *VPS45*, *AC244033.2*, and *C1orf54* interactively contribute to the altered neural phenotypes at the rs2027349 locus. Related to Figure 6. (A) shRNA KD of any of the three targeted genes led to a significant decrease in overall *VPS45* (all transcript isoforms) expression, with shRNA_*VPS45* showing the most robust reduction of *VPS45*. Data were generated from line CD0000011, 1-2 clones per condition and 3 biological replicates per clone were used. Kruskal-Wallis (non-parametric) test with Dunn's multiple comparisons and adjusted *P*-values. (B) Sample-wise split graph of Figure 6D, showing the Sholl analysis results of rs2027349 AA line under different KD conditions in contrast with rs2027349 GG line. Data were generated from CD0000011 and collated from two independent experiments. Two clones per genotype were used. Two-way ANOVA with *posthoc* correction for multiple testing was used to test the significance of the differences. (C) Principal component analysis (PCA) plot of all RNA-seq samples in individual KD experiments showing the clustering of different rs2027349 AA knockdown lines and rs2027349 GG control lines. 2-3 biological replicates per clone. (D) Scatter plot showing the strong correlation of shRNA KD effects between the two clones (A11 and H12). A union of genes found in at least one clonal line were used. (E) Scatter plot showing the correlation and linear regression result between the \log_2 fold changes (FC) of DE genes in SNP-edited GG neurons (GG vs. AA) and the \log_2 FC in different shRNA KD conditions in AA neurons. (F-H) Venn diagrams and bubble plots show gene count overlap between different KD conditions. In each figure, the right-panel bubble plot matches the left-panel Venn diagram. Only the top 40 GO terms (ranked by the enrichment ratio) were shown in each bubble plot and only conditions (rows) with corresponding values in these terms were plotted. Circle size: enrichment ratio; fill: $-\log_{10}FDR$.

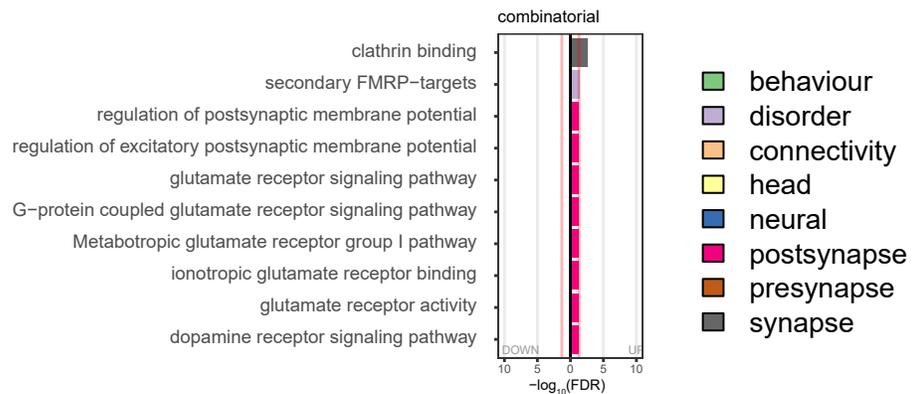
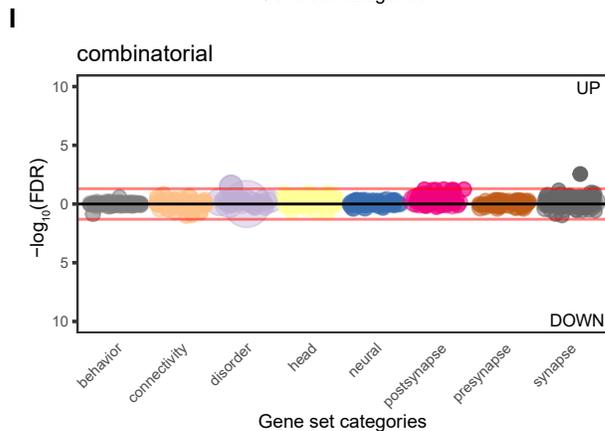
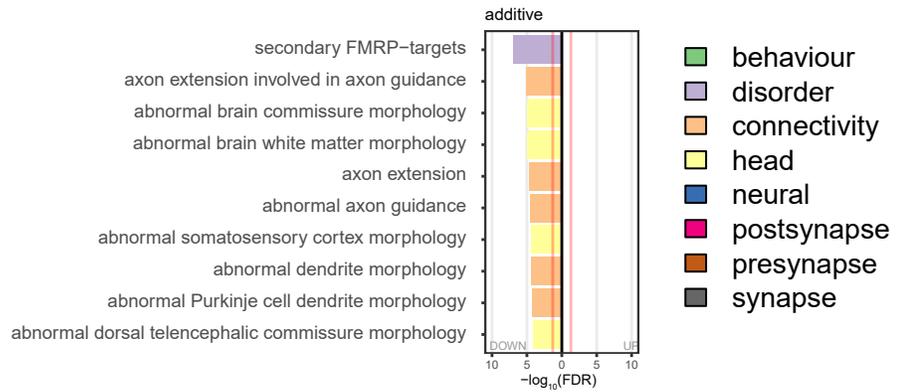
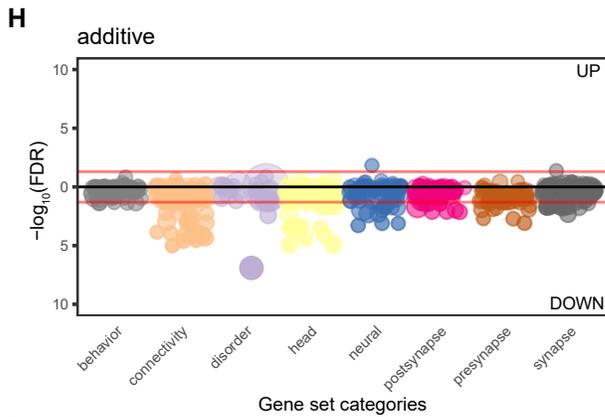
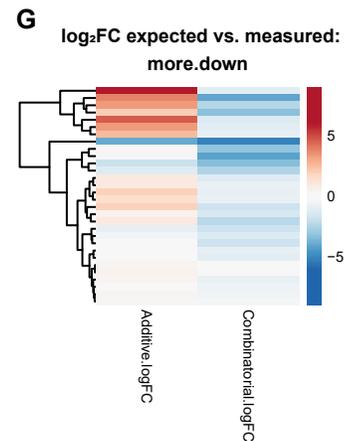
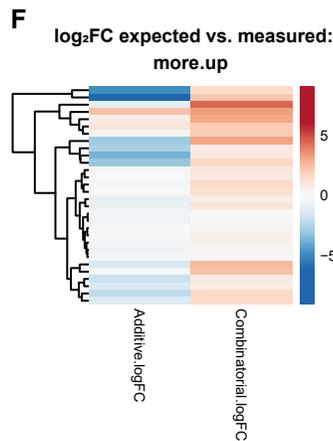
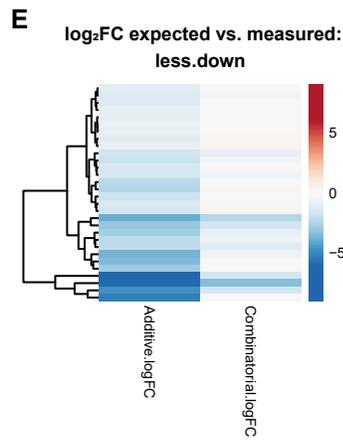
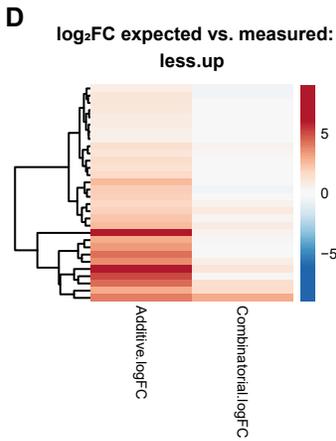
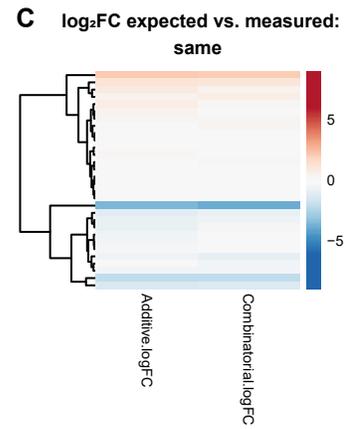
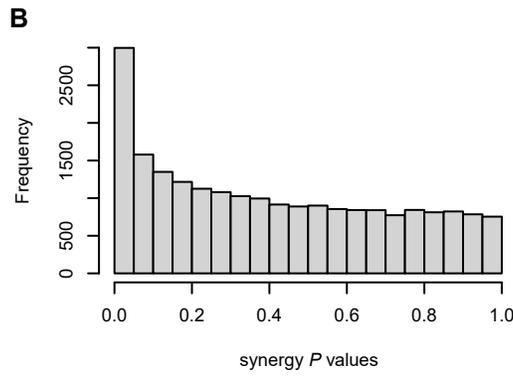
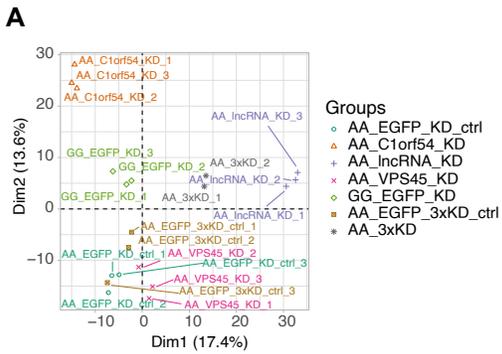


Figure S8: Additional results on synergy-driving gene expression analysis. Related to Figure 7. (A): Principal component analysis (PCA) plot of all RNA-seq samples used in synergy-driving gene expression analysis, showing the clustering of different rs2027349 AA knockdown and rs2027349 GG control replicates (CD0000011, 2-3 replicates per clone). (B) Histogram and statistical data show the distribution of P values. (C-G) DE value (in \log_2FC) and hierarchical clustering of all synergy categories in the additive model vs. the combinatorial perturbation comparisons. (H-I) competitive GSEA of DE in the additive (H) and the combinatorial (I) comparisons based on 698 curated neural gene sets stratified by eight neural categories and their corresponding bar plots as ranked by FDR value. Red lines denote enrichment $\pm \log_{10}(FDR) = 0.05$.