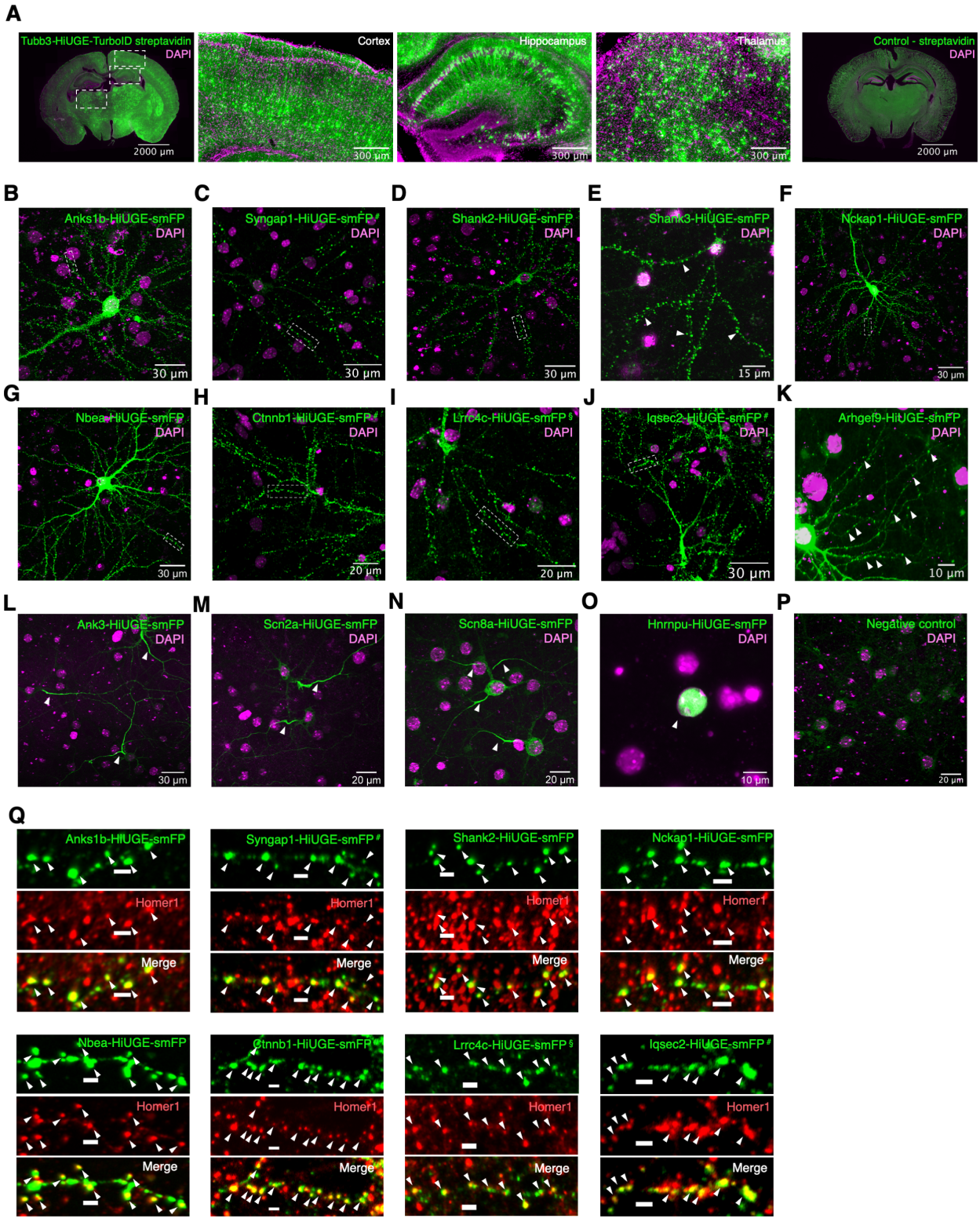
Supplementary Information

Proximity Analysis of Native Proteomes Reveals Phenotypic Modifiers in a Mouse Model of Autism and Related Neurodevelopmental Conditions

Gao *et al.*

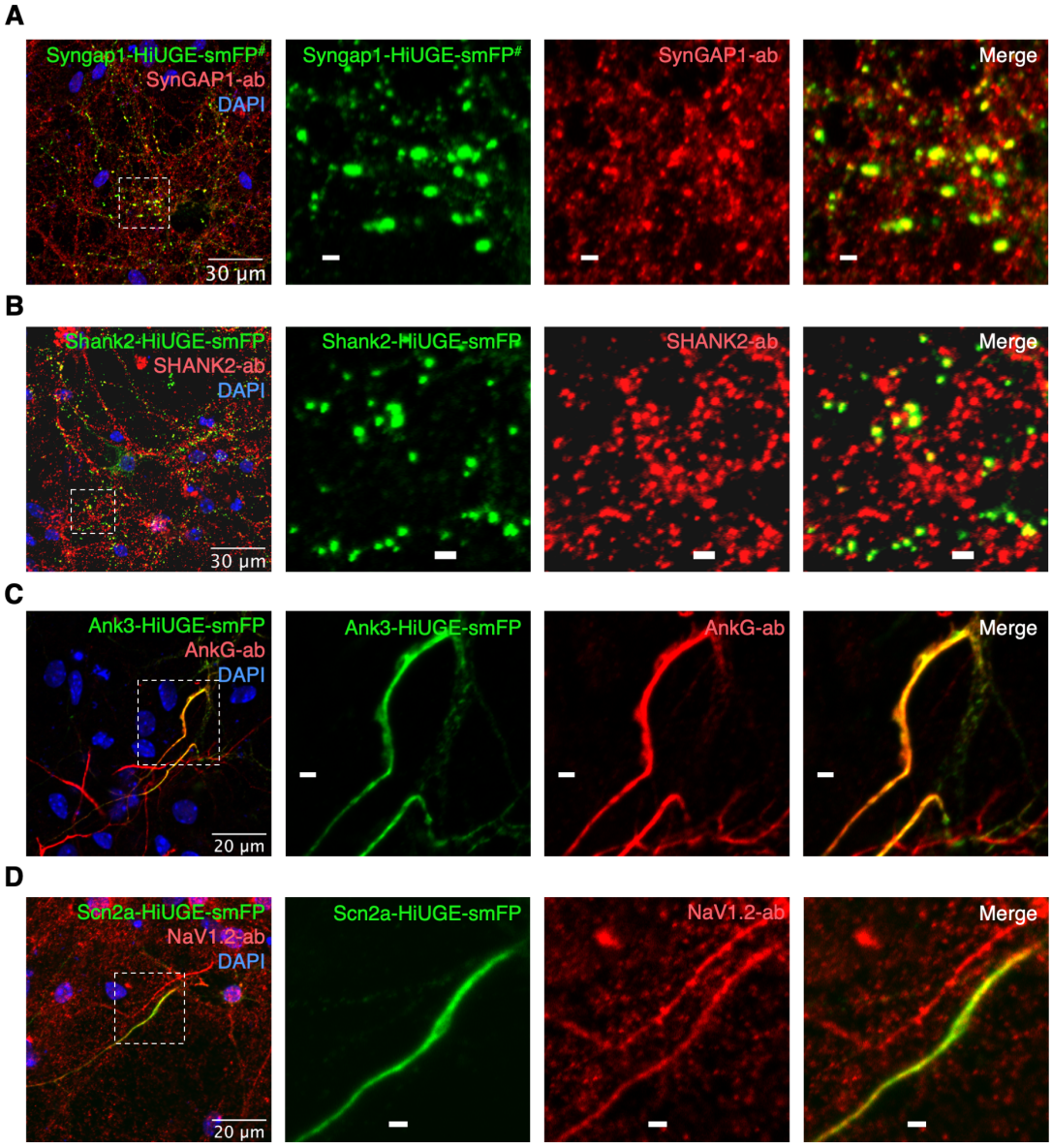
Document includes:

Supplementary Figures S1-28



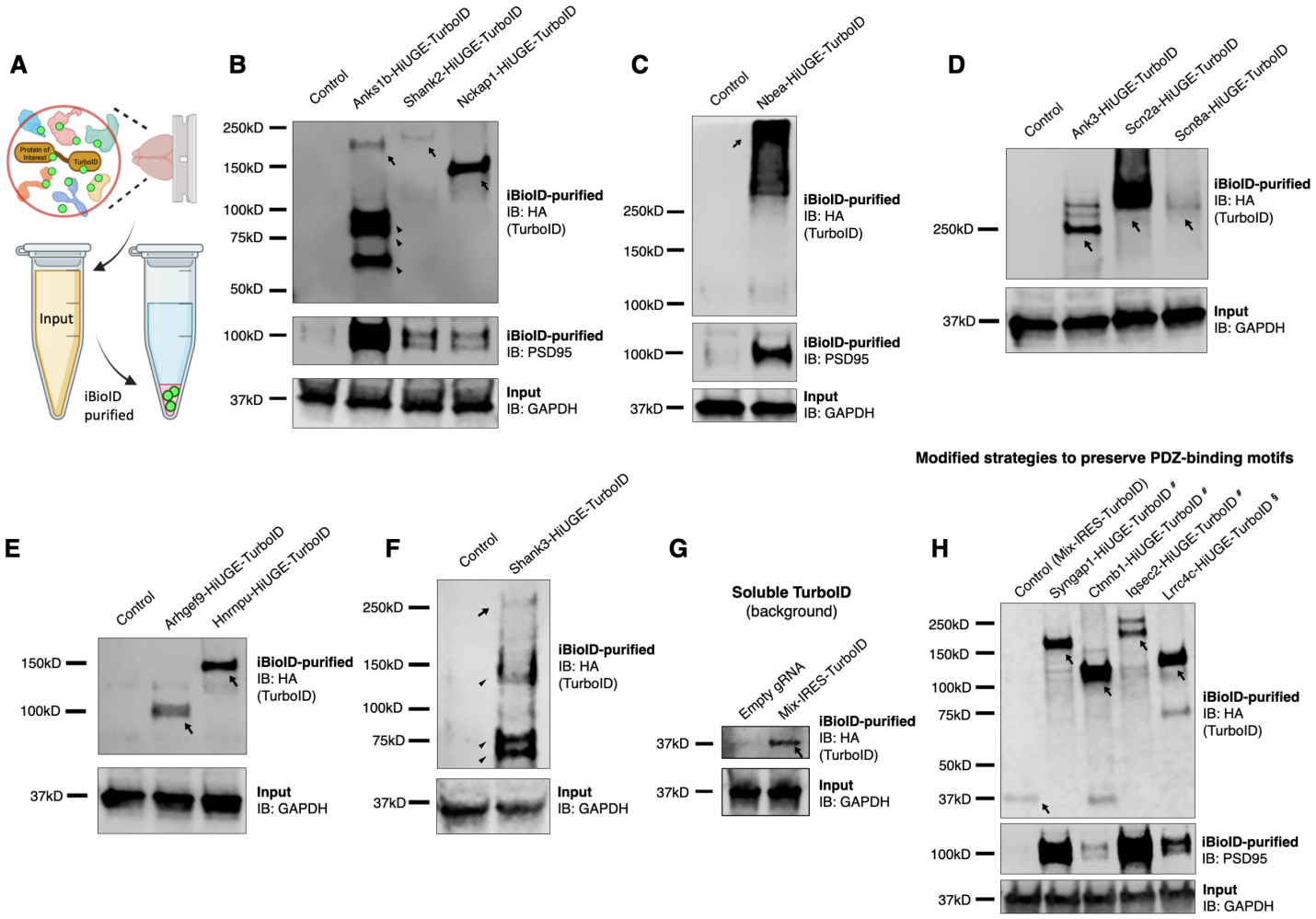
Supplementary Fig. S1. HiUGE labeling of 14 high-risk autism proteins.

(**A**) Representative images showing wide-spread TurboID-mediated biotinylation across the brain following labeling Tubb3 with HiUGE. (**B-O**) Representative images showing correct localization (arrowheads) of 14 high-risk autism proteins labeled with a highly antigenic "spaghetti monster" fluorescent protein (smFP), with boxed regions enlarged in (**Q**) to show colocalization with a synaptic marker Homer1. PDZ-binding motifs were preserved using #: intron-targeting strategy and §: modified donor incorporating native PDZ-binding motif. (**P**) Representative image of the negative control. Scale bar in the enlarged view represents 2µm.

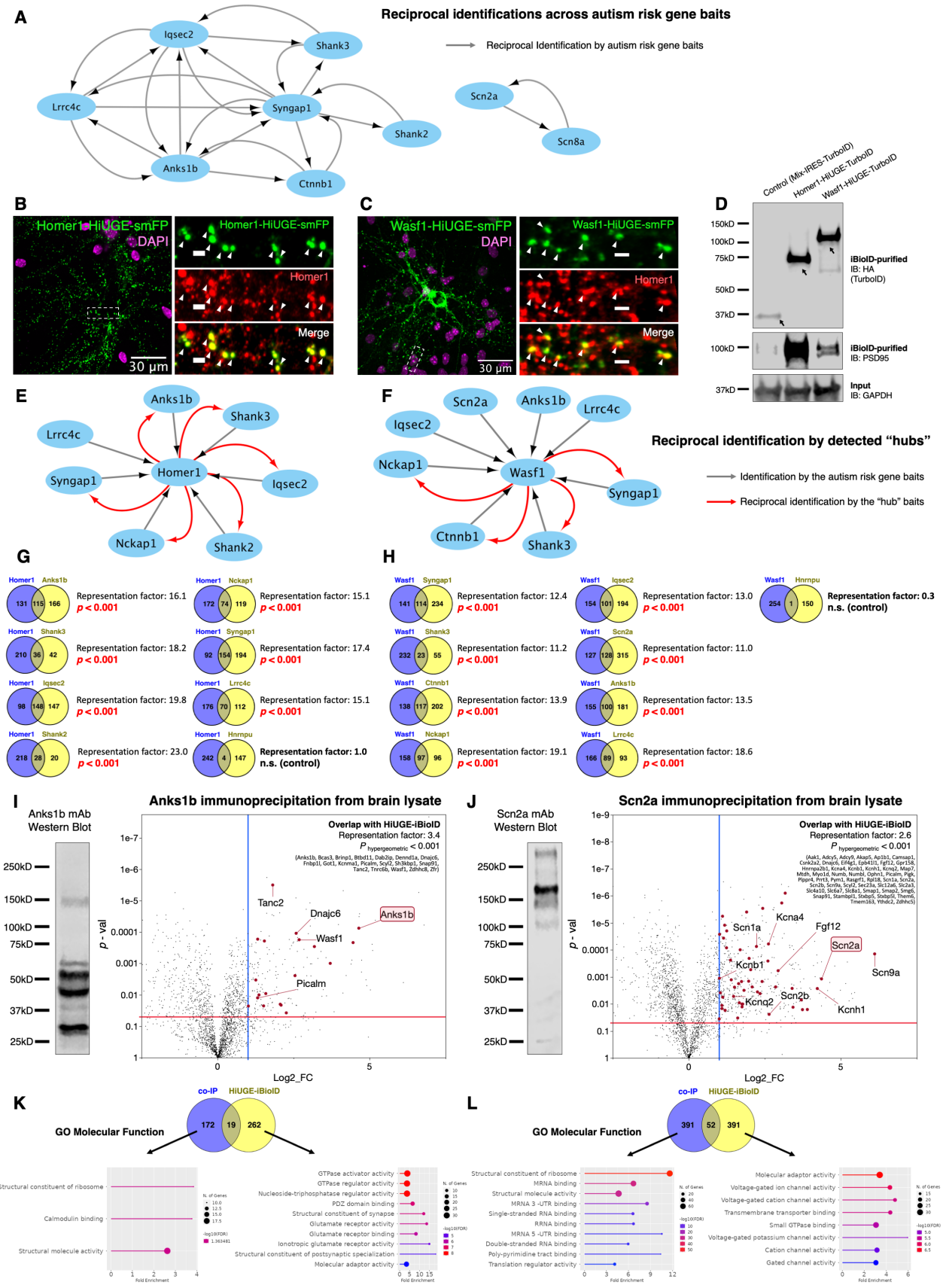


Supplementary Fig. S2. Immunofluorescence detection of representative bait proteins with HiUGE labeling.

(**A-D**) Representative images showing colocalization of smFP-labeled proteins with immunofluorescence detected by specific antibodies against native proteins (Syngap1, Shank2, Ank3, Scn2a). Scale bar in the enlarged view represents 2μm.

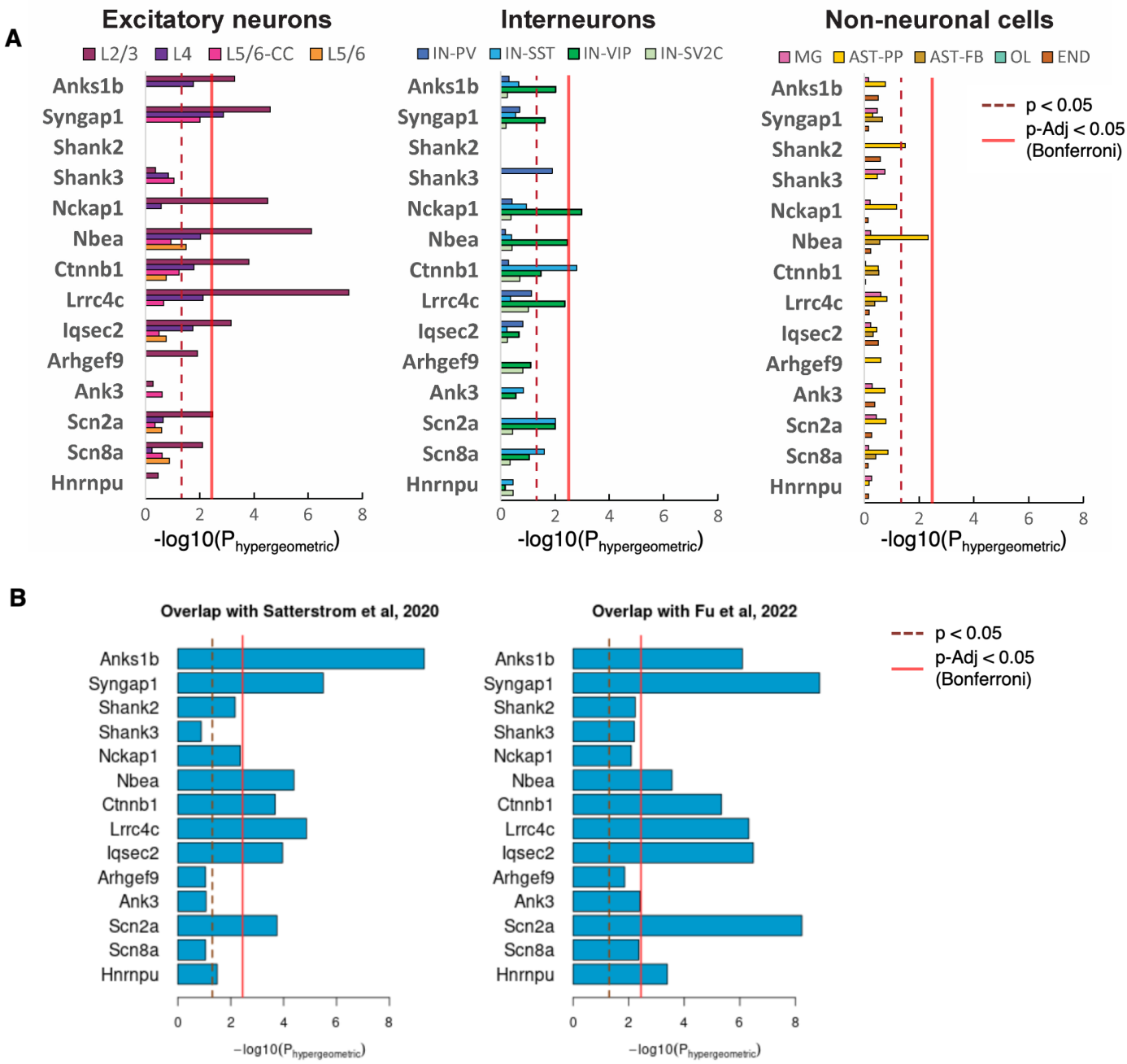


Supplementary Fig. S3. Western blot of HiUGE-iBioID purified samples following streptavidin pulldown. (**A**) Schematic illustration of enriching biotinylated proteins by streptavidin pulldown. (**B-H**) Western blot images showing detection of TurboID-HA fusion proteins at the expected molecular masses following purifications, (**B, C, H**) detection of an expected synaptic interactor (PSD95) is also confirmed. (**G, H**) Western blot images showing detection of soluble TurboID-HA (as a survey for background) by multiplexed insertion of IRES-TurboID-HA donor near 3'UTR. (**H**) #: intron-targeting strategy and §: modified donor incorporating native PDZ-binding motif. Target proteins are marked with arrows, and potential splice isoforms are marked with arrowheads.

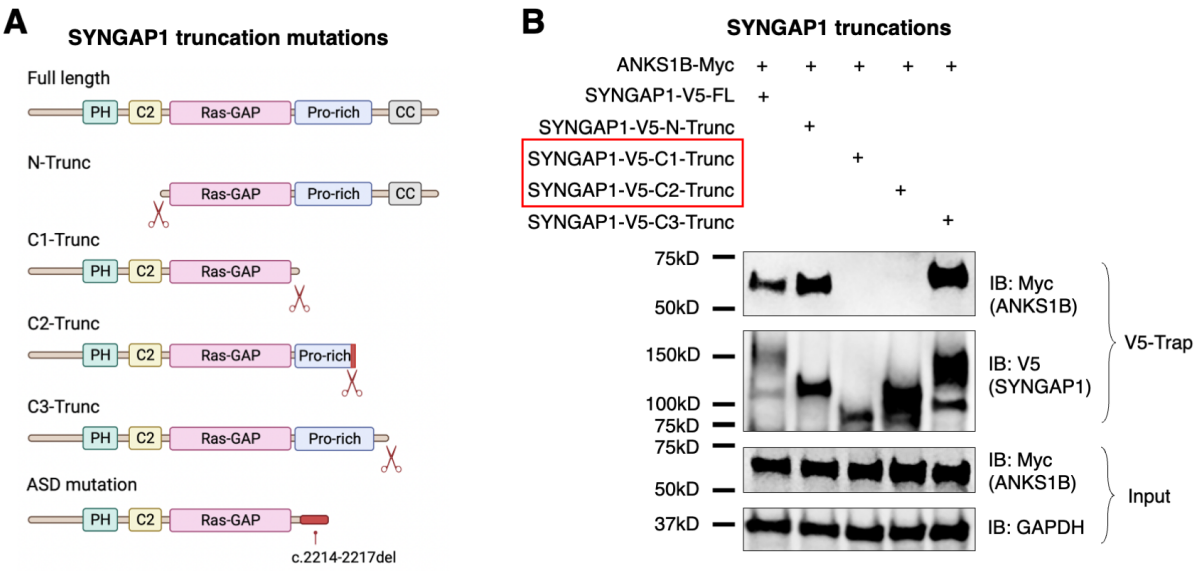


Supplementary Fig. S4. Reciprocal validation and comparison with immunoprecipitation.
(**A**) Network graph showing extensive reciprocal identifications amongst autism risk protein baits. (**B-D**) Immunofluorescence and Western blot validation of HiUGE-iBioID for Homer1 and Wasf1, two exemplary proteomic "hubs" detected by many baits. (**E, F**) Network graphs showing reciprocal identifications of the initial baits by these "hubs". (**G, H**) Hypergeometric analyses of overlap between the HiUGE-iBioID proteomes of the

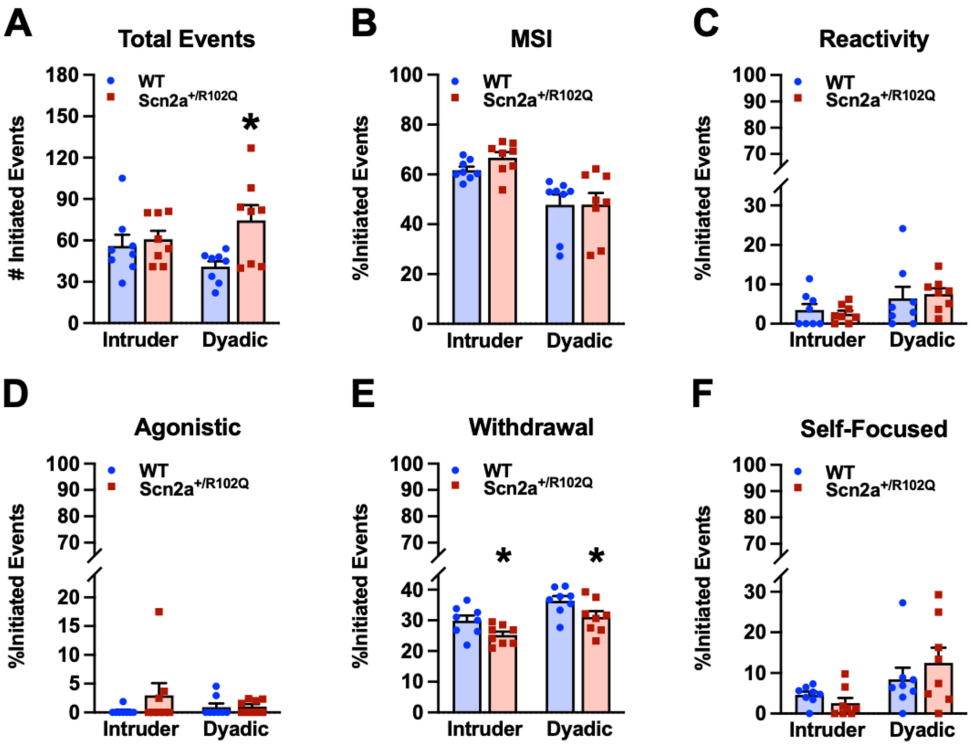
initial baits and the "hubs" show highly significant overlaps. No significant overlap is detected when comparing to the Hnrnpu dataset. **(I, J)** Comparison of HiUGE-iBioID results with proteomic detections following immunoprecipitation using monoclonal antibodies. Western blot images of brain lysate samples using these antibodies are shown. Genes that overlap with HiUGE-iBioID are highlighted on the volcano plot, and a few genes related to synaptic and ion channel functions are labeled. Significance of overlap is determined by hypergeometric test. (**K, L**) Comparative GO analyses of the gene sets that are exclusive to either the immunoprecipitation or the HiUGE-iBioID datasets. The statistical domain is the cumulative proteomic detections of brain-derived samples in our lab (Supplementary Data S5).



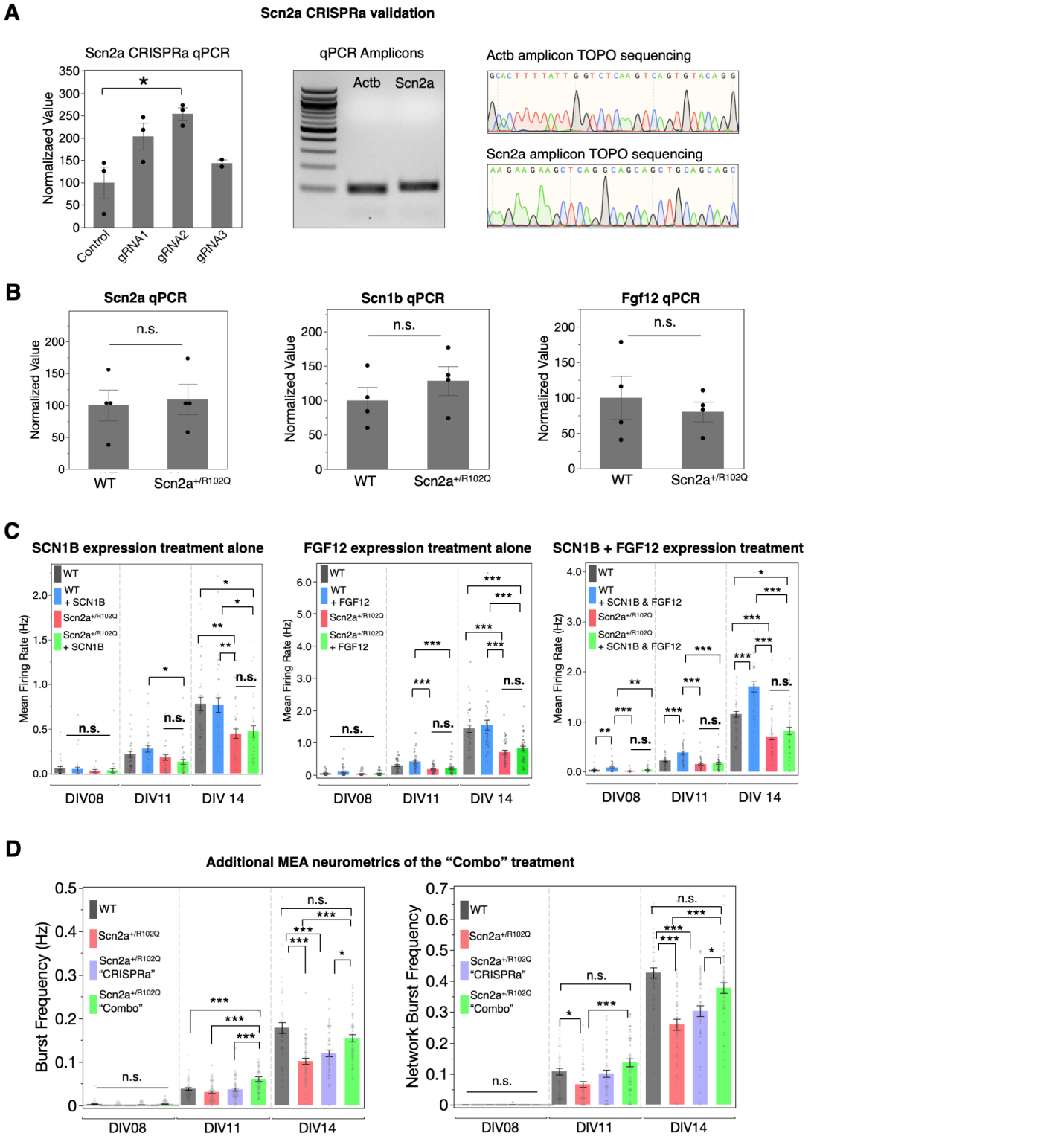
Supplementary Fig. S5. Overlap of HiUGE-iBioID proximity proteomes with published datasets.
(**A**) Significance of overlap with differentially expressed genes (DEGs) found in autistic individuals across excitatory neurons, interneurons, and non-neuronal cell populations. (**B**) Significance of overlap with autism risk genes identified by Satterstrom et al. [3], and Fu et al. [7]. Thresholds for statistical significance were delineated on the graphs.



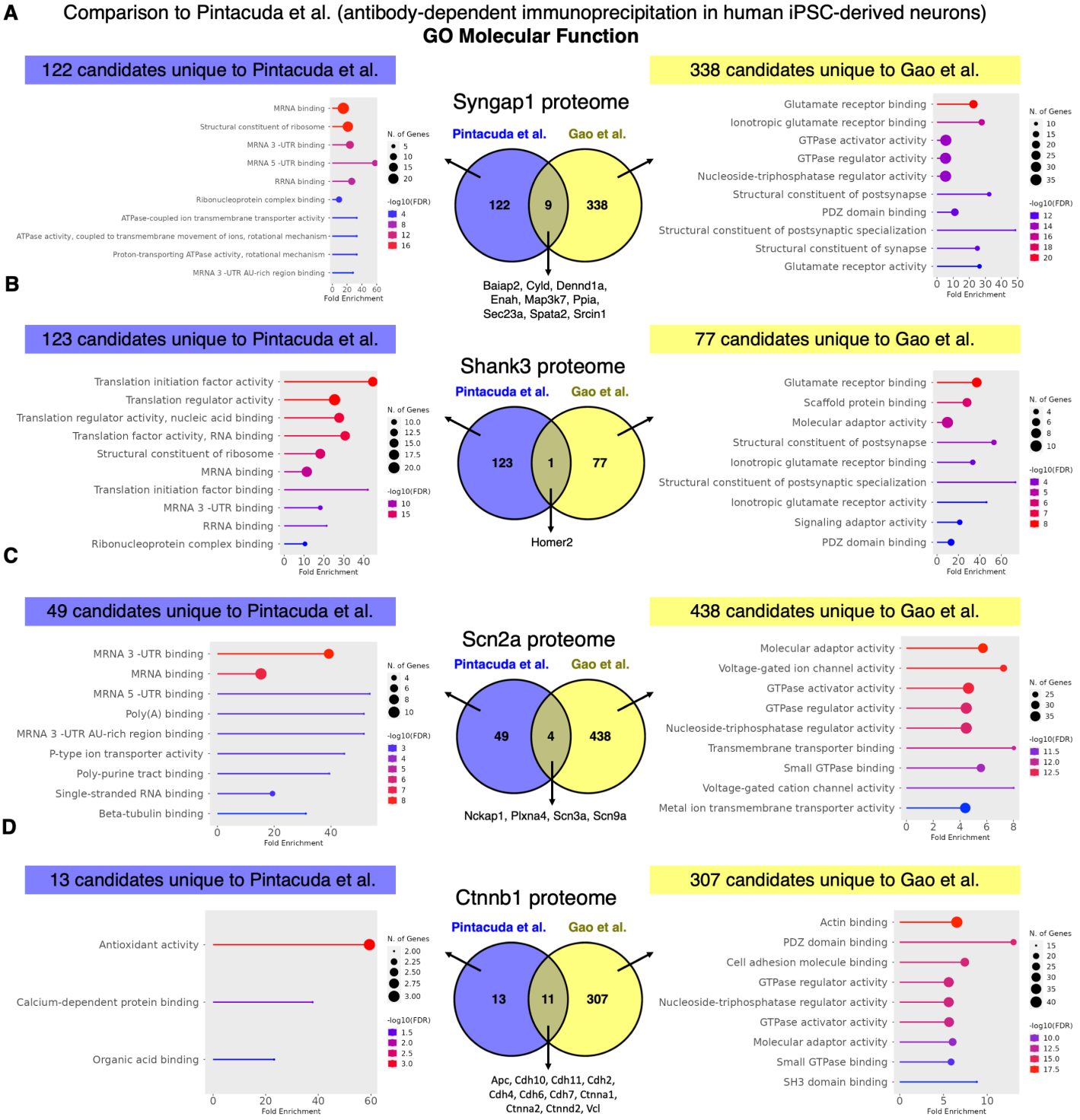
Supplementary Fig. S6. Structure-function analysis of SYNGAP1-ANKS1B interaction.
(**A**) Schematic illustration of assessing the ANKS1B interaction with SYNGAP1 truncations using human cDNA constructs expressed in HEK293T cells. (**B**) Co-immunoprecipitation results showing the interaction with ANKS1B ablated in C1- and C2- SYNGAP1 truncations while retained in C3- and N- truncations.



Supplementary Fig. S7. Additional social behavior data for the Scn2a$^{+/R102Q}$ mice.
(**A**) Scn2a$^{+/R102Q}$ males initiated more overall social events in the dyadic assay with C3H/HeJ males than WT males (p = 0.013). (**B**) No genotype effect was detected for the percent of mild social interactions (MSI). (**C**) No genotype effect was detected for the percent of reactivity events. (**D**) The numbers of agonistic events were very low and were not distinguished by social test or genotype. (**E**) The percent of withdrawal events from C3H/HeJ partners was lower in the Scn2a$^{+/R102Q}$ males than WT males (genotype effect, p = 0.020). (**F**) No genotype effect was detected for the self-focused behaviors. The data are presented as means ± SEMs and were analyzed with RMANOVA with Bonferroni corrections, n = 8 mice / genotype. *: $p < 0.05$, WT vs. Scn2a$^{+/R102Q}$ mice. Additional statistics are summarized in Supplementary Data S6.

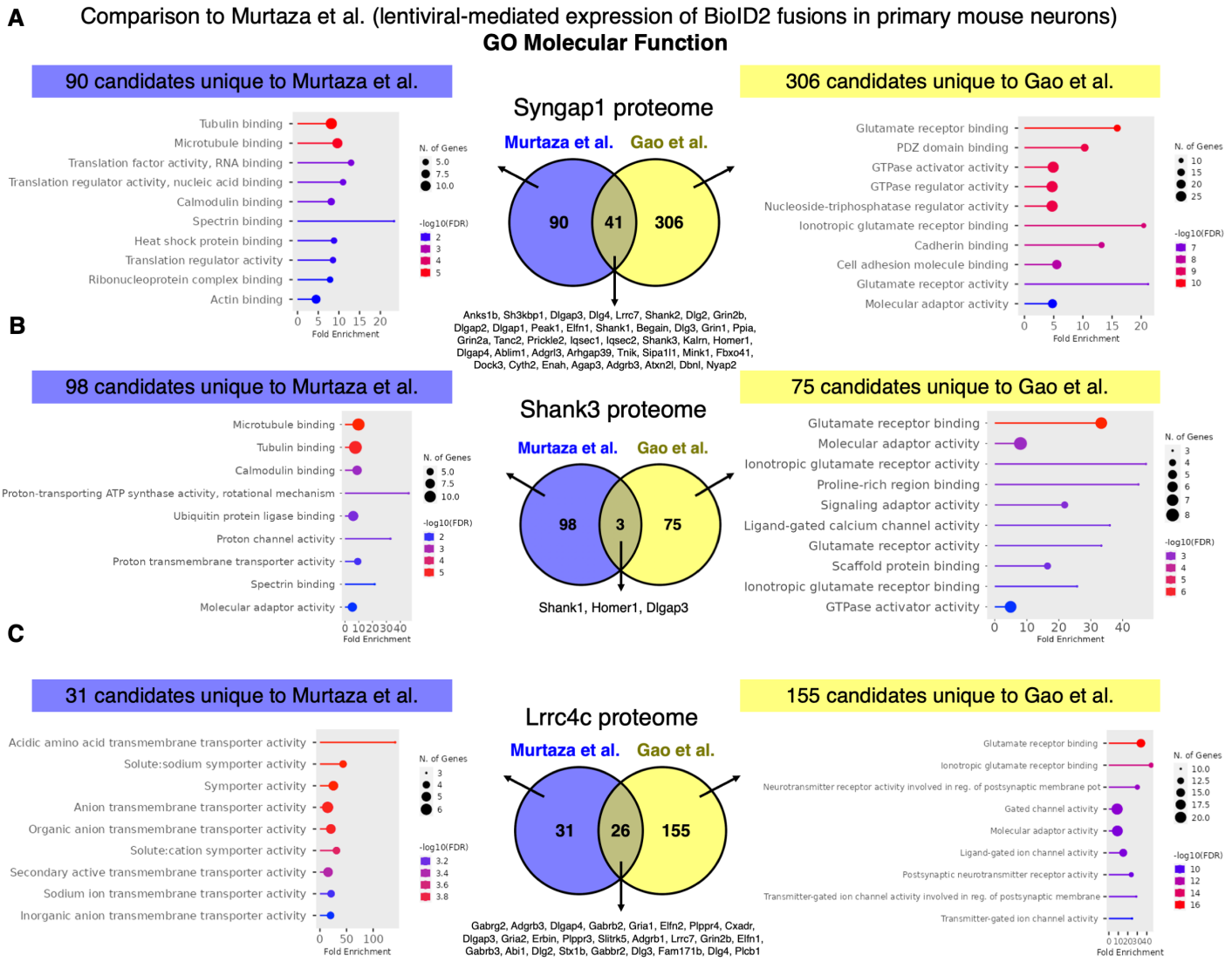


Supplementary Fig. S8. Additional data of the Scn2a$^{+/R102Q}$ phenotypic rescue experiment.
(**A**) Quantitative PCR screening of three different CRISPRa gRNAs targeting Scn2a *in vitro*. The gRNA2 shows the best performance in upregulating Scn2a expression, and is used for subsequent experiments (ANOVA followed by Dunnett's test, n = 3 wells). Specificity of the qPCR assay is validated by sequencing the amplicon. (**B**) mRNA expression levels of Scn2a, Scn1b, and Fgf12 in cultured Scn2a$^{+/R102Q}$ mutant neurons are comparable to that of WT (two-tailed *t*-test, n = 4 wells). (**C**) MEA neurometrics following overexpression of SCN1B, FGF12, or combined showing ineffective rescue of the Scn2a$^{+/R102Q}$ phenotype, One-way ANOVA followed by *post-hoc* Tukey HSD tests (n= 36 or 48 wells). (**D**) Additional neurometrics from the MEA recording of the "Combo" treatment experiment. One-way ANOVA followed by *post-hoc* Tukey HSD tests (n= 48 wells). *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$; n.s.: non-significant. Plots are mean ± SEM.



Supplementary Fig. S9. Comparative GO analysis with a recently published autism risk protein immunoprecipitation dataset using human iPSC-derived neurons.

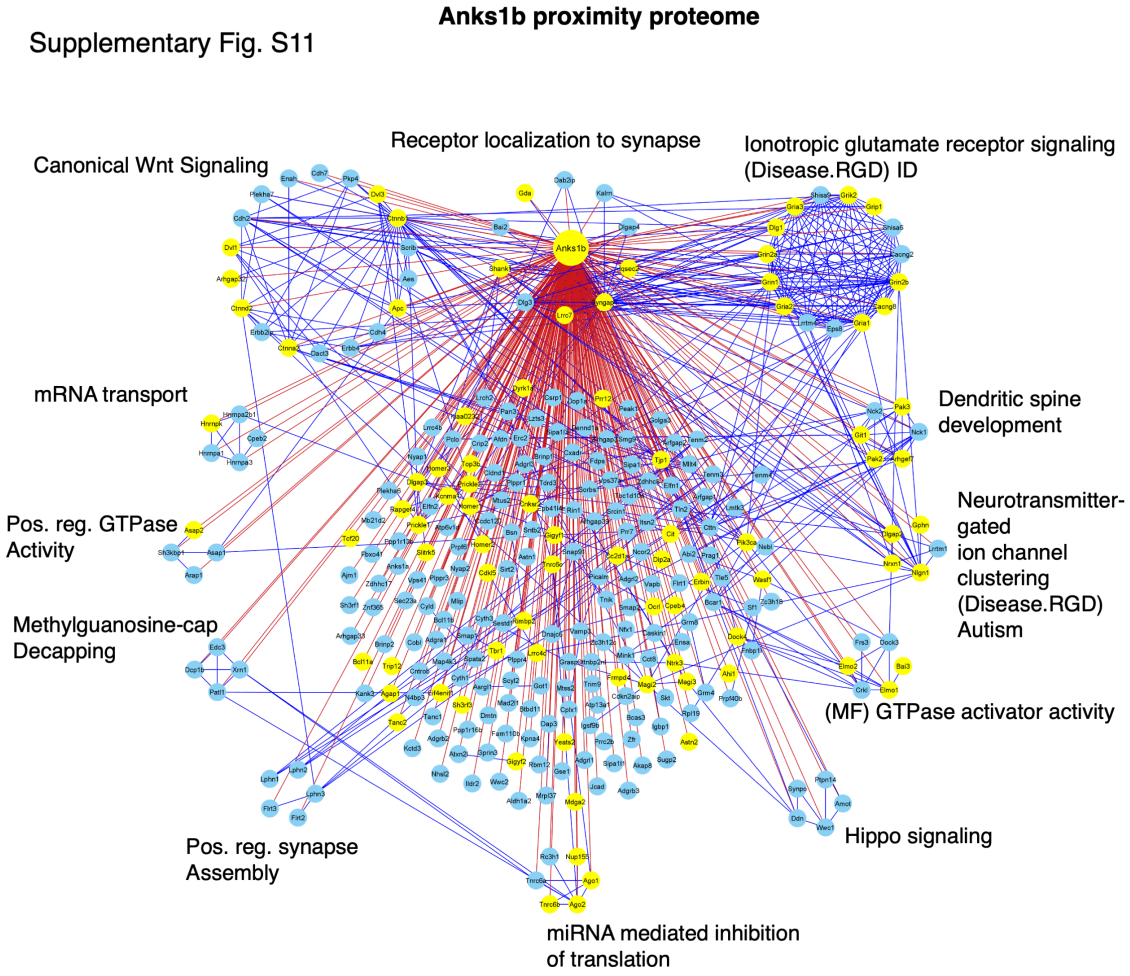
(**A-D**) Comparative GO analyses of genes that are exclusive to either the immunoprecipitation dataset using human iPSC-derived neurons [28] or the HiUGE-iBioID dataset for Syngap1, Shank3, Scn2a and Ctnnb1. The human genes were converted to mouse orthologs before intersecting with HiUGE-iBioID dataset. Bait self-IDs were excluded for this analysis. Due to the lack of a consensus statistical domain, mouse genome was used as a non-biased background. Top molecular function GO terms are shown.

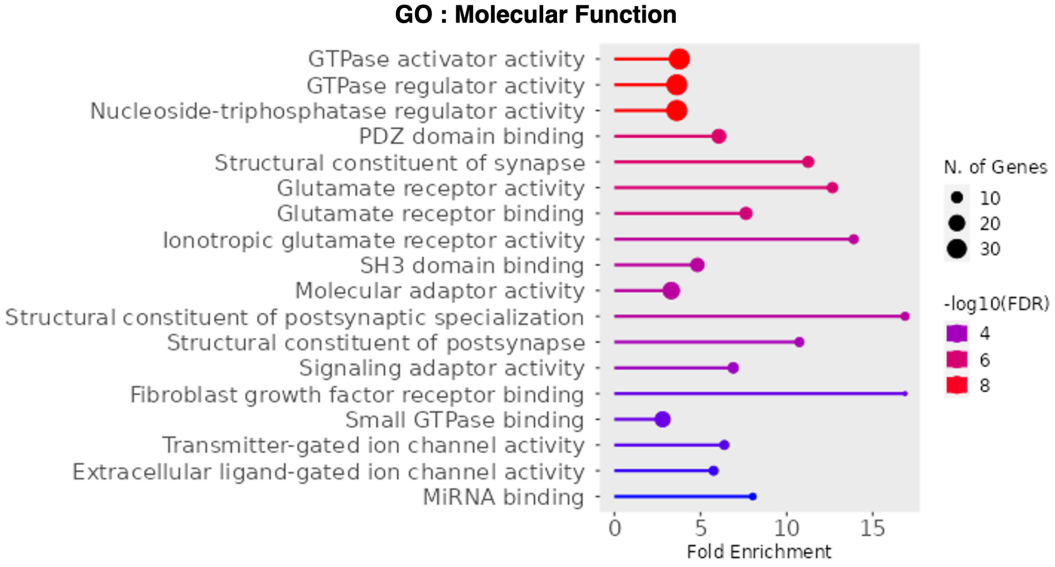


Supplementary Fig. S10. Comparative GO analysis with a recently published autism risk protein recombinant BioID dataset using cultured neurons.

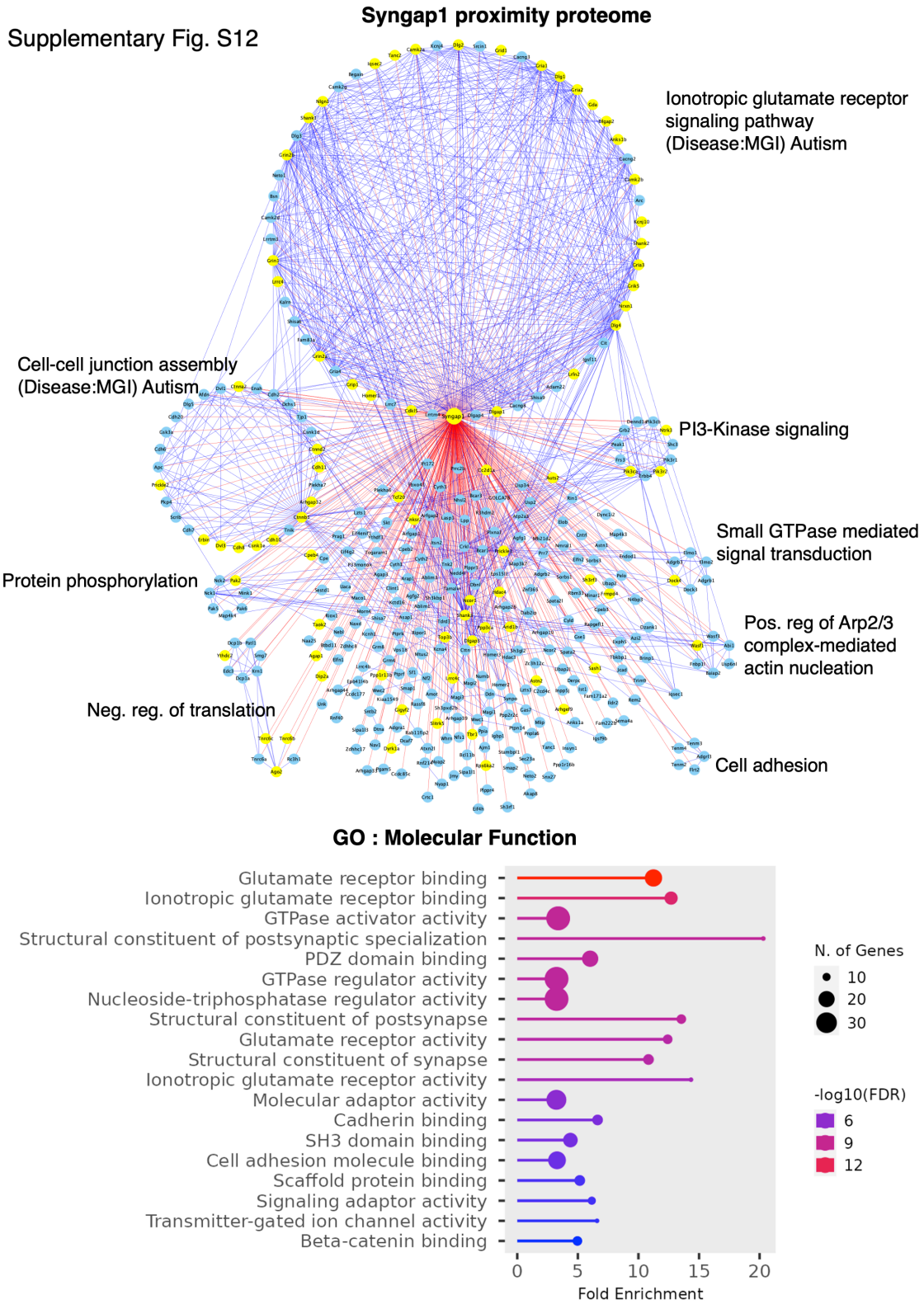
(**A-C**) Comparative GO analyses of genes that are exclusive to either the recombinant BioID expression dataset using primary mouse neurons [27] or the HiUGE-iBioID dataset for Syngap1, Shank3, and Lrrc4c. Bait self-IDs were excluded for this analysis. Due to the lack of a consensus statistical domain, mouse genome was used as a non-biased background. Top molecular function GO terms are shown.

Supplementary Fig. S11



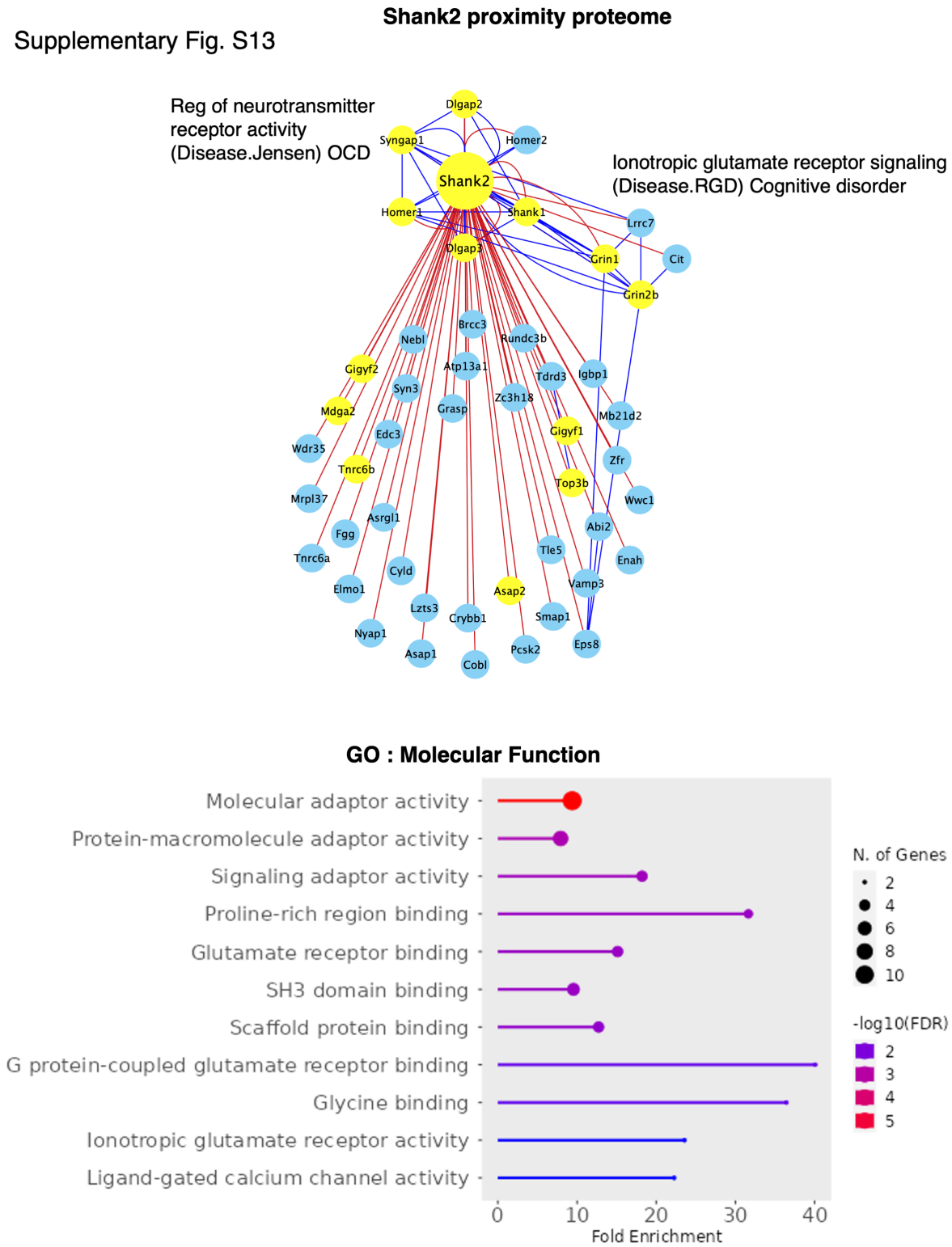


Supplementary Fig. S11. Proximity proteomic networks associated with bait protein Anks1b.
The proximity proteomic network associated with the bait protein is shown. Blue lines denote STRING interactions and red lines signify identified HiUGE-iBioID interactions. Yellow nodes highlight proteins encoded by SFARI gene orthologs. Annotations denote exemplary significant gene ontology (GO) terms associated with the protein clusters segregated by MCL. Unless otherwise specified, the Biological Process pathway database was used. CC: Cellular Component pathway database. Charts of GO results of the overall bait proximity proteomes using the Molecular Function (MF) pathway database are shown as well, below each network plot.

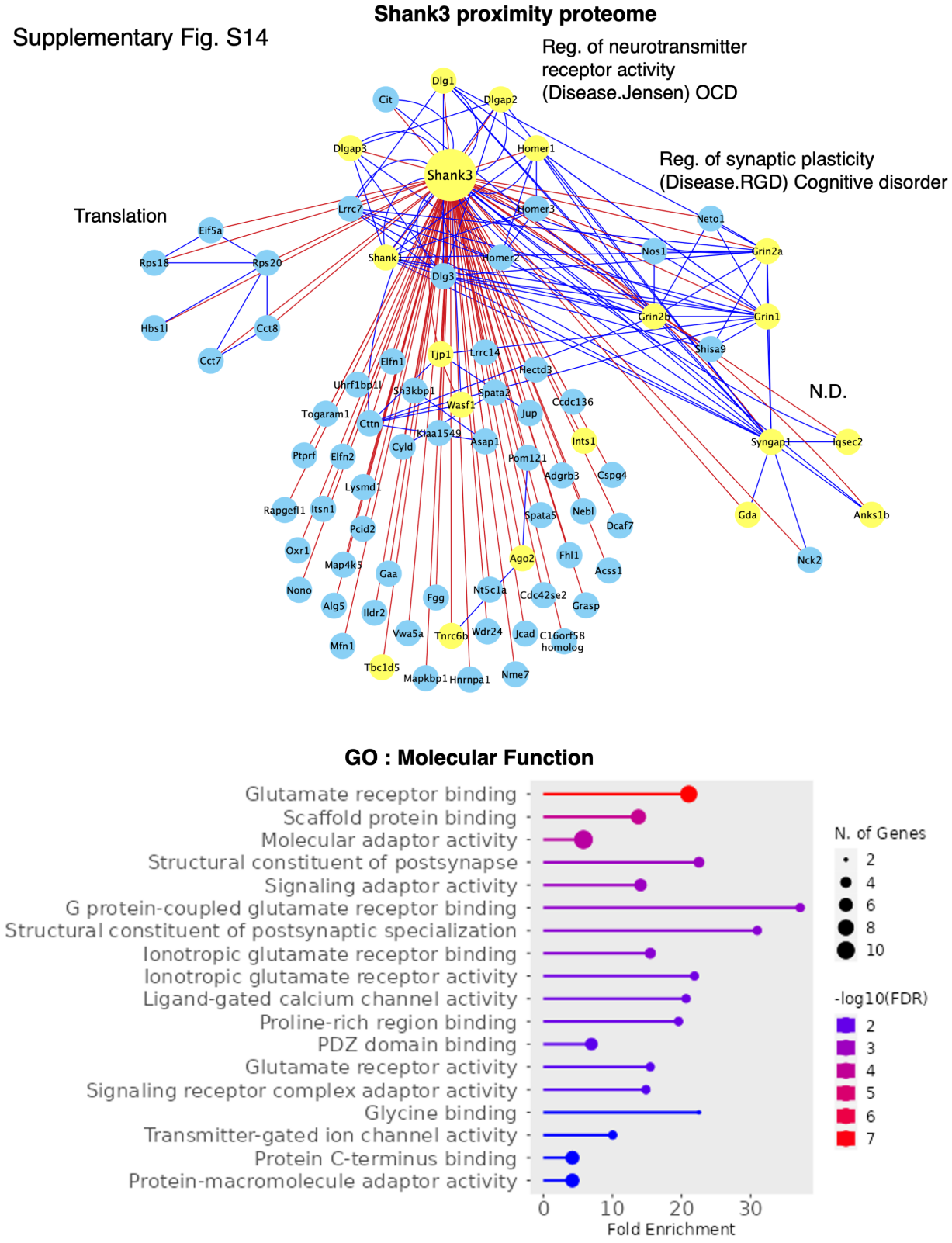


Supplementary Fig. S12. Proximity proteomic networks associated with bait protein Syngap1.
The proximity proteomic network associated with the bait protein is shown. Blue lines denote STRING interactions and red lines signify identified HiUGE-iBioID interactions. Yellow nodes highlight proteins encoded by SFARI gene orthologs. Annotations denote exemplary significant gene ontology (GO) terms associated with the protein clusters segregated by MCL. Unless otherwise specified, the Biological Process pathway database was used. CC: Cellular Component pathway database. Charts of GO results of the overall bait proximity proteomes using the Molecular Function (MF) pathway database are shown as well, below each network plot.

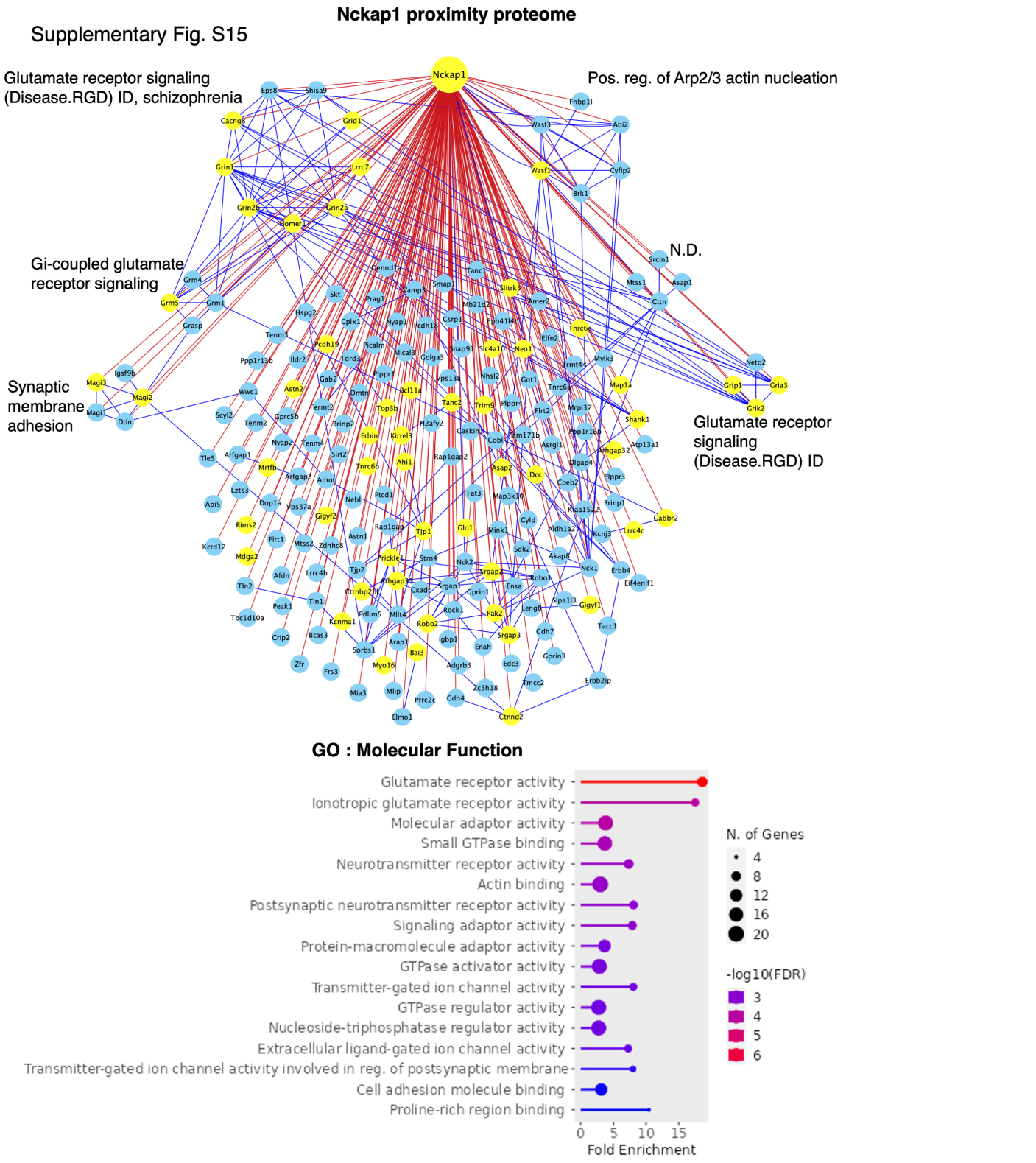
Supplementary Fig. S13



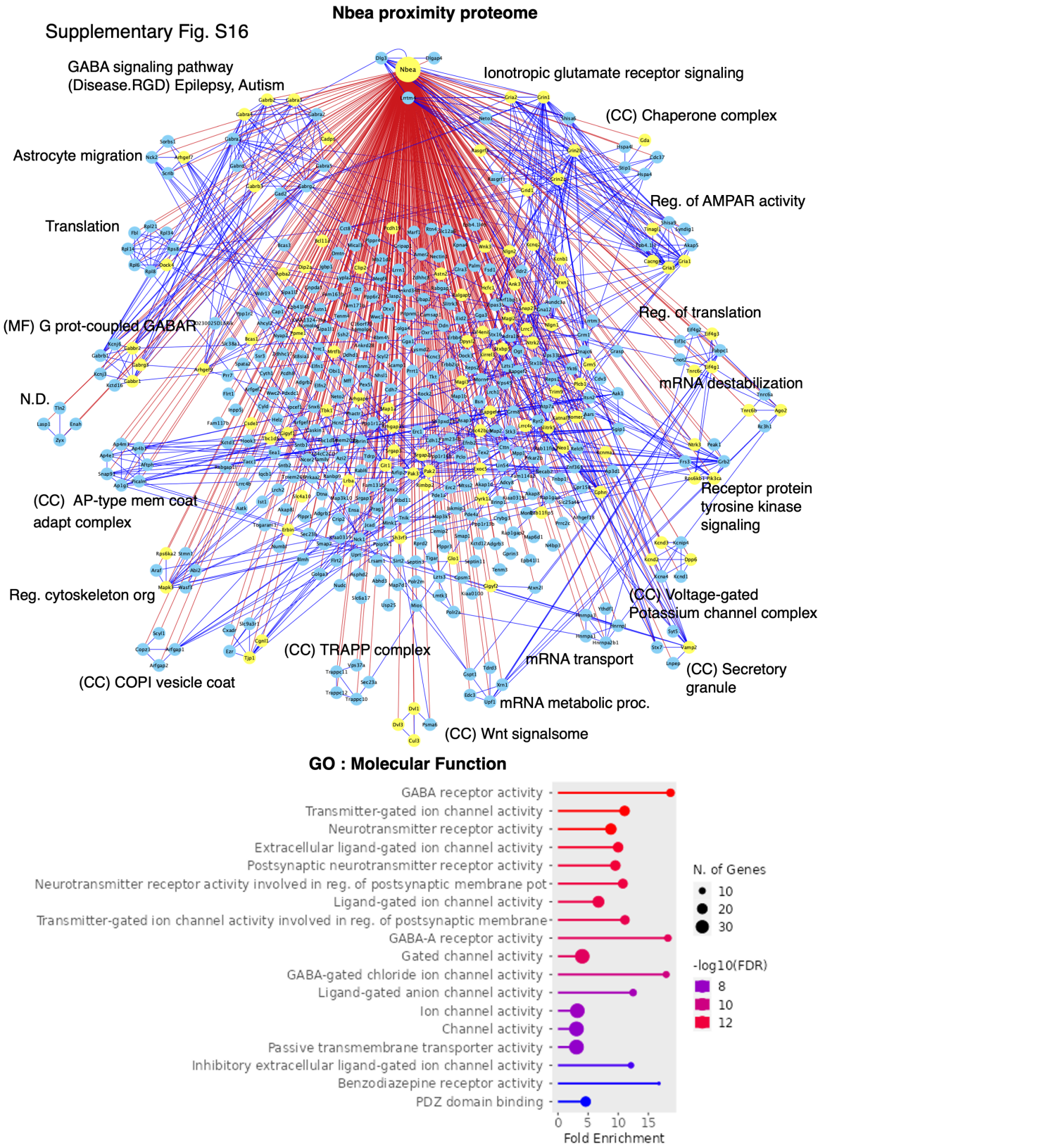
Supplementary Fig. S13. Proximity proteomic networks associated with bait protein Shank2.
The proximity proteomic network associated with the bait protein is shown. Blue lines denote STRING interactions and red lines signify identified HiUGE-iBioID interactions. Yellow nodes highlight proteins encoded by SFARI gene orthologs. Annotations denote exemplary significant gene ontology (GO) terms associated with the protein clusters segregated by MCL. Unless otherwise specified, the Biological Process pathway database was used. CC: Cellular Component pathway database. Charts of GO results of the overall bait proximity proteomes using the Molecular Function (MF) pathway database are shown as well, below each network plot.



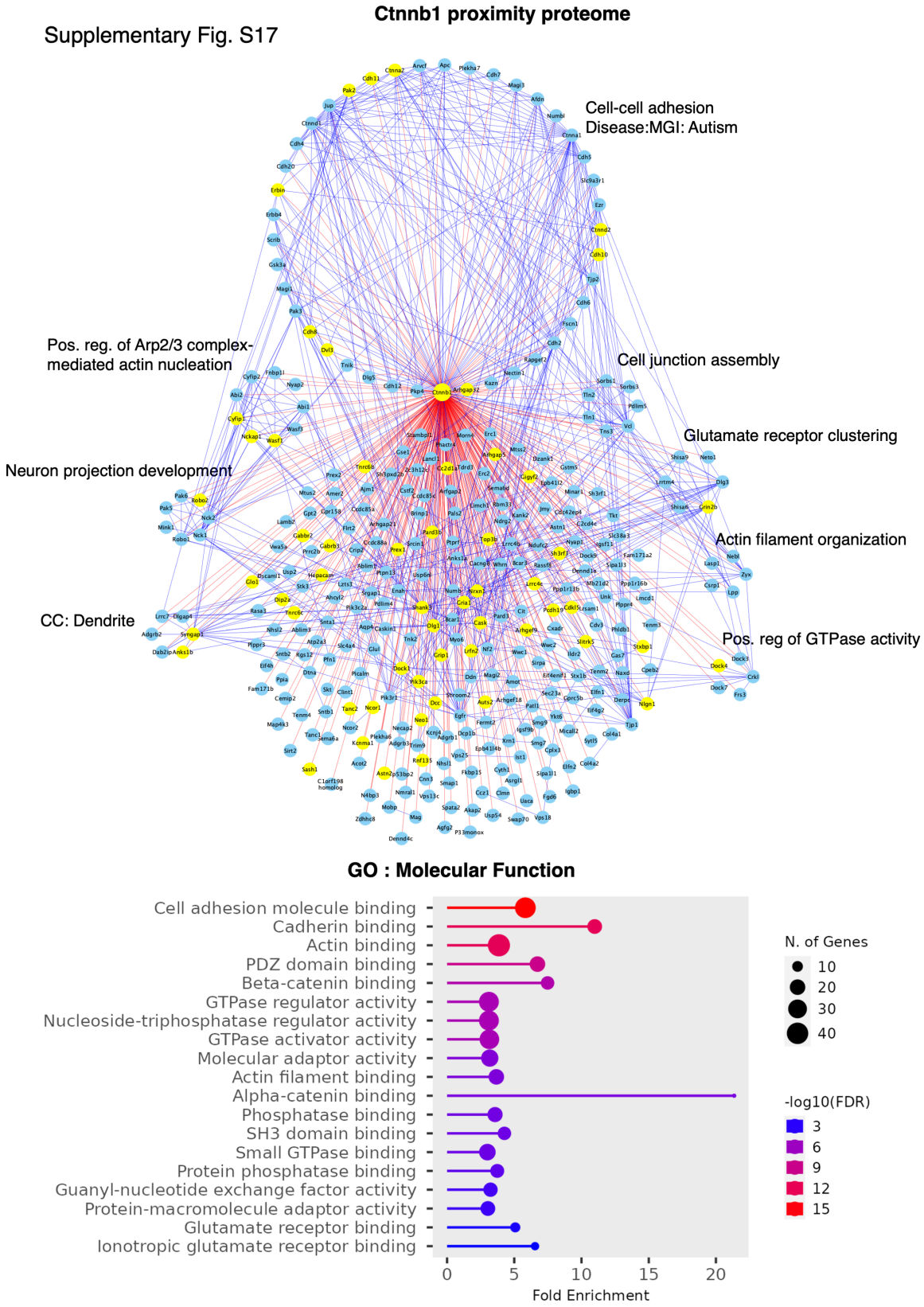
Supplementary Fig. S14. Proximity proteomic networks associated with bait protein Shank3.
The proximity proteomic network associated with the bait protein is shown. Blue lines denote STRING interactions and red lines signify identified HiUGE-iBioID interactions. Yellow nodes highlight proteins encoded by SFARI gene orthologs. Annotations denote exemplary significant gene ontology (GO) terms associated with the protein clusters segregated by MCL. Unless otherwise specified, the Biological Process pathway database was used. CC: Cellular Component pathway database. Charts of GO results of the overall bait proximity proteomes using the Molecular Function (MF) pathway database are shown as well, below each network plot.



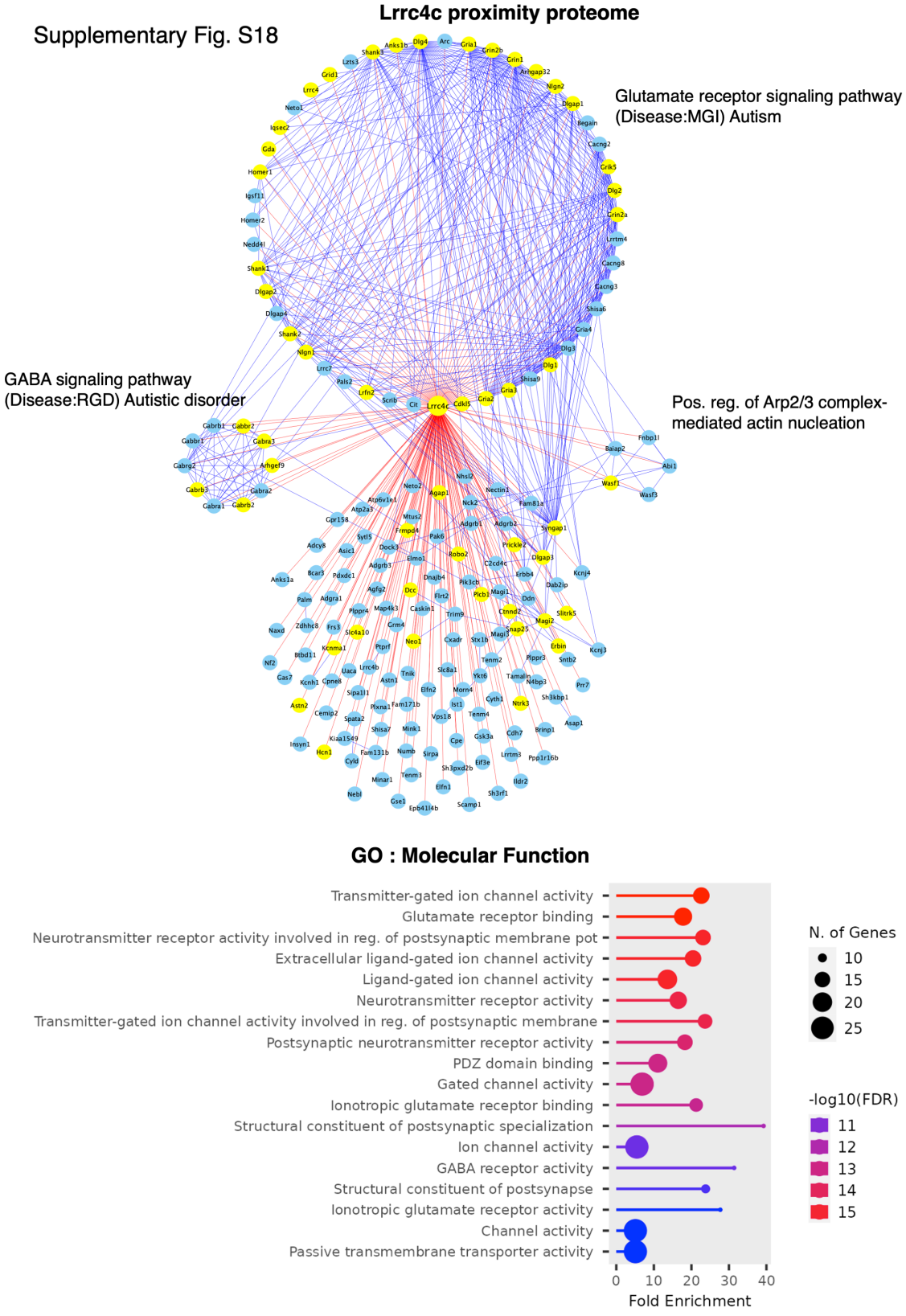
Supplementary Fig. S15. Proximity proteomic networks associated with bait protein Nckap1.
The proximity proteomic network associated with the bait protein is shown. Blue lines denote STRING interactions and red lines signify identified HiUGE-iBioID interactions. Yellow nodes highlight proteins encoded by SFARI gene orthologs. Annotations denote exemplary significant gene ontology (GO) terms associated with the protein clusters segregated by MCL. Unless otherwise specified, the Biological Process pathway database was used. CC: Cellular Component pathway database. Charts of GO results of the overall bait proximity proteomes using the Molecular Function (MF) pathway database are shown as well, below each network plot.



Supplementary Fig. S16. Proximity proteomic networks associated with bait protein Nbea.
The proximity proteomic network associated with the bait protein is shown. Blue lines denote STRING interactions and red lines signify identified HiUGE-iBioID interactions. Yellow nodes highlight proteins encoded by SFARI gene orthologs. Annotations denote exemplary significant gene ontology (GO) terms associated with the protein clusters segregated by MCL. Unless otherwise specified, the Biological Process pathway database was used. CC: Cellular Component pathway database. Charts of GO results of the overall bait proximity proteomes using the Molecular Function (MF) pathway database are shown as well, below each network plot.

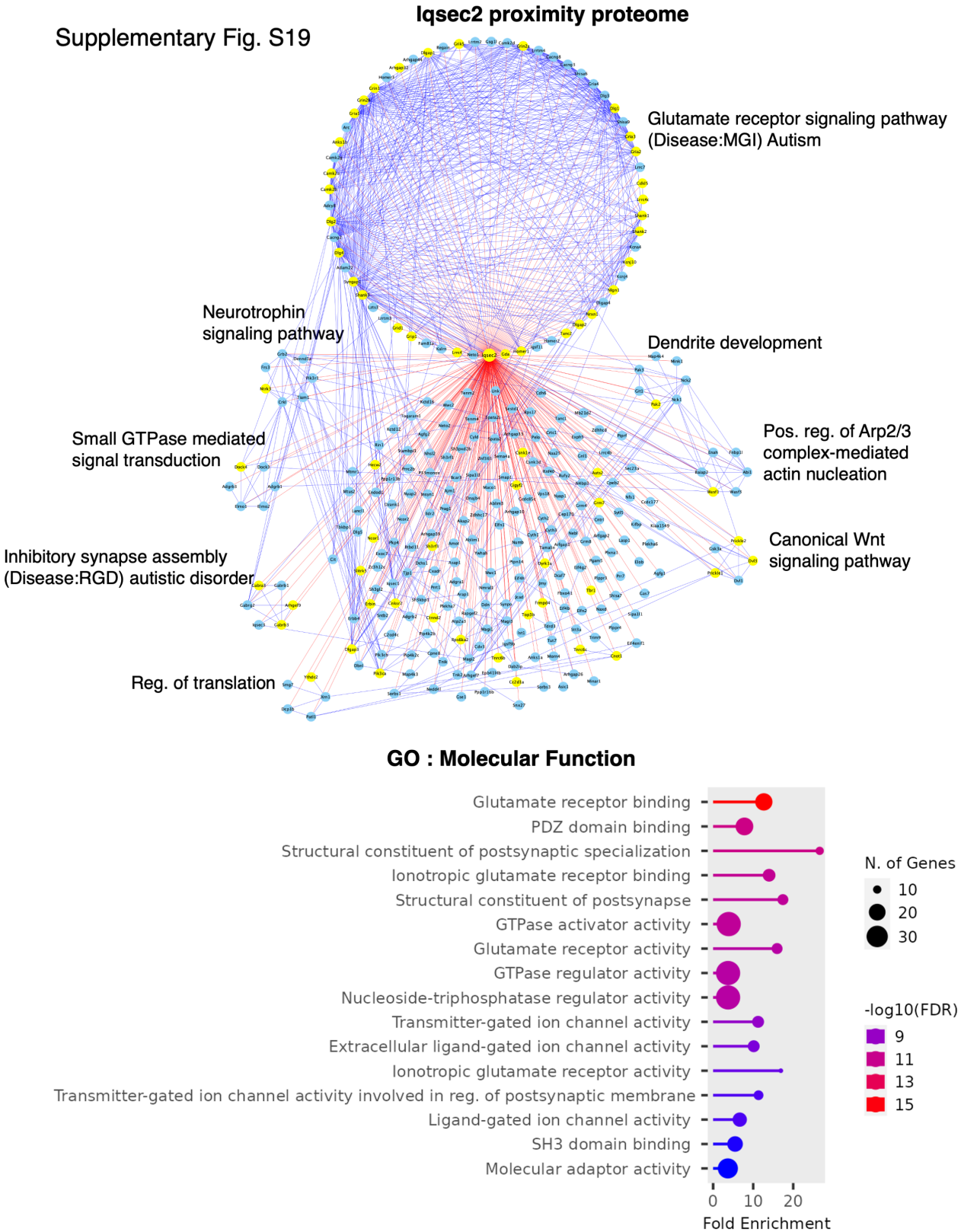


Supplementary Fig. S17. Proximity proteomic networks associated with bait protein Ctnnb1.
The proximity proteomic network associated with the bait protein is shown. Blue lines denote STRING interactions and red lines signify identified HiUGE-iBioID interactions. Yellow nodes highlight proteins encoded by SFARI gene orthologs. Annotations denote exemplary significant gene ontology (GO) terms associated with the protein clusters segregated by MCL. Unless otherwise specified, the Biological Process pathway database was used. CC: Cellular Component pathway database. Charts of GO results of the overall bait proximity proteomes using the Molecular Function (MF) pathway database are shown as well, below each network plot.



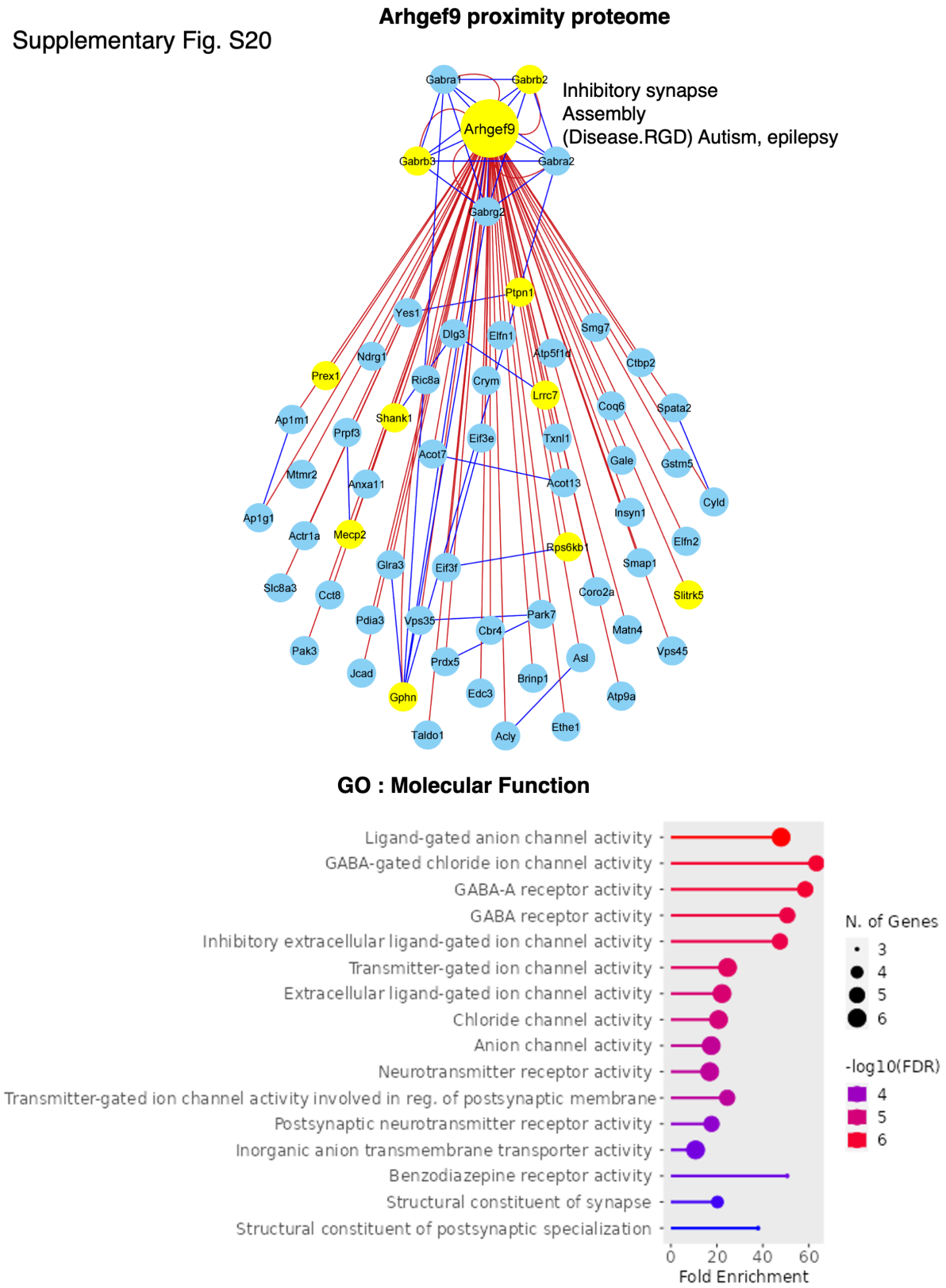
Supplementary Fig. S18. Proximity proteomic networks associated with bait protein Lrrc4c.
The proximity proteomic network associated with the bait protein is shown. Blue lines denote STRING interactions and red lines signify identified HiUGE-iBioID interactions. Yellow nodes highlight proteins encoded by SFARI gene orthologs. Annotations denote exemplary significant gene ontology (GO) terms associated with the protein clusters segregated by MCL. Unless otherwise specified, the Biological Process pathway database was used. CC: Cellular Component pathway database. Charts of GO results of the overall bait proximity proteomes using the Molecular Function (MF) pathway database are shown as well, below each network plot.

Supplementary Fig. S19

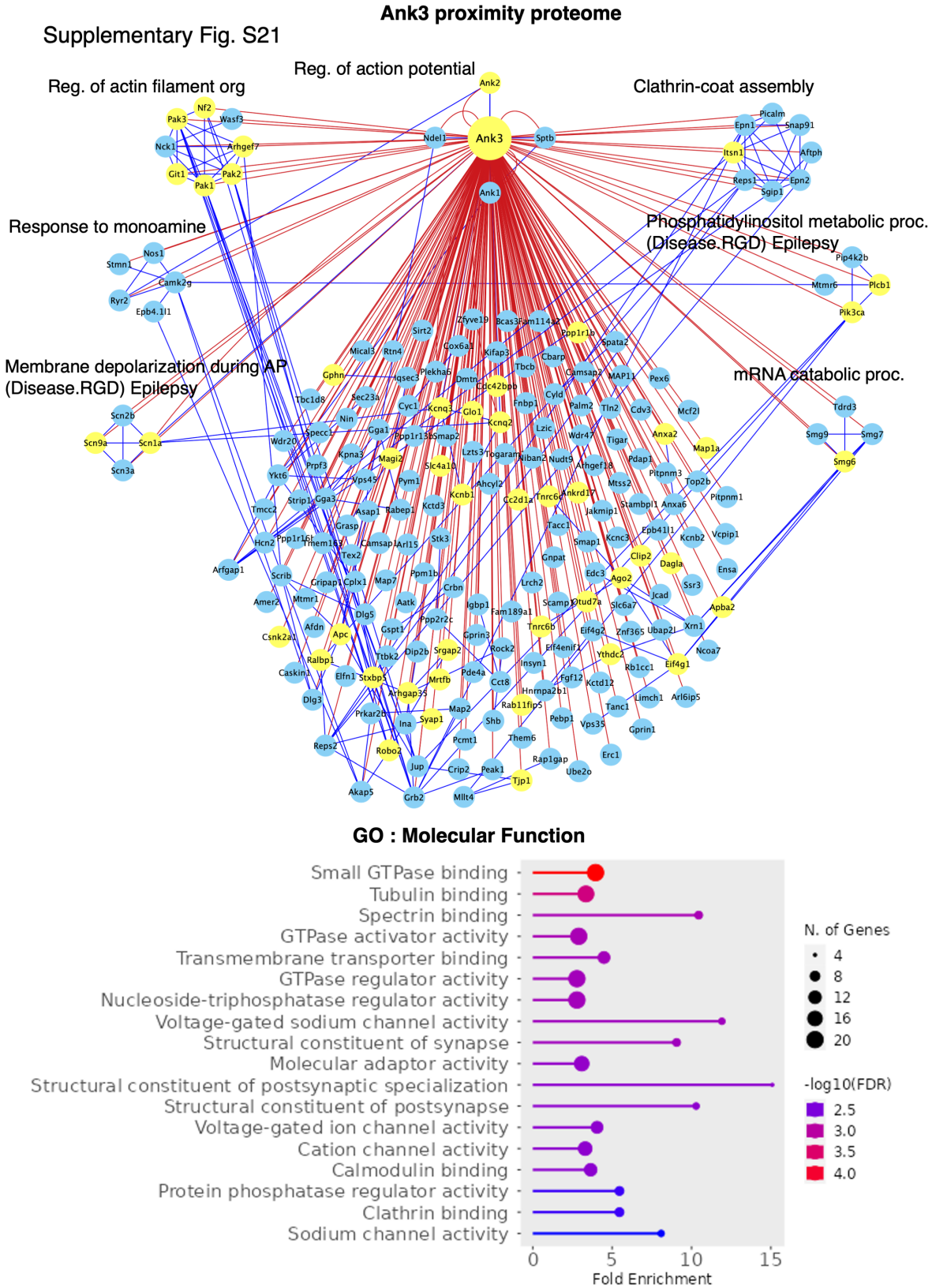


Supplementary Fig. S19. Proximity proteomic networks associated with bait protein Iqsec2.
The proximity proteomic network associated with the bait protein is shown. Blue lines denote STRING interactions and red lines signify identified HiUGE-iBioID interactions. Yellow nodes highlight proteins encoded by SFARI gene orthologs. Annotations denote exemplary significant gene ontology (GO) terms associated with the protein clusters segregated by MCL. Unless otherwise specified, the Biological Process pathway database was used. CC: Cellular Component pathway database. Charts of GO results of the overall bait proximity proteomes using the Molecular Function (MF) pathway database are shown as well, below each network plot.

Supplementary Fig. S20

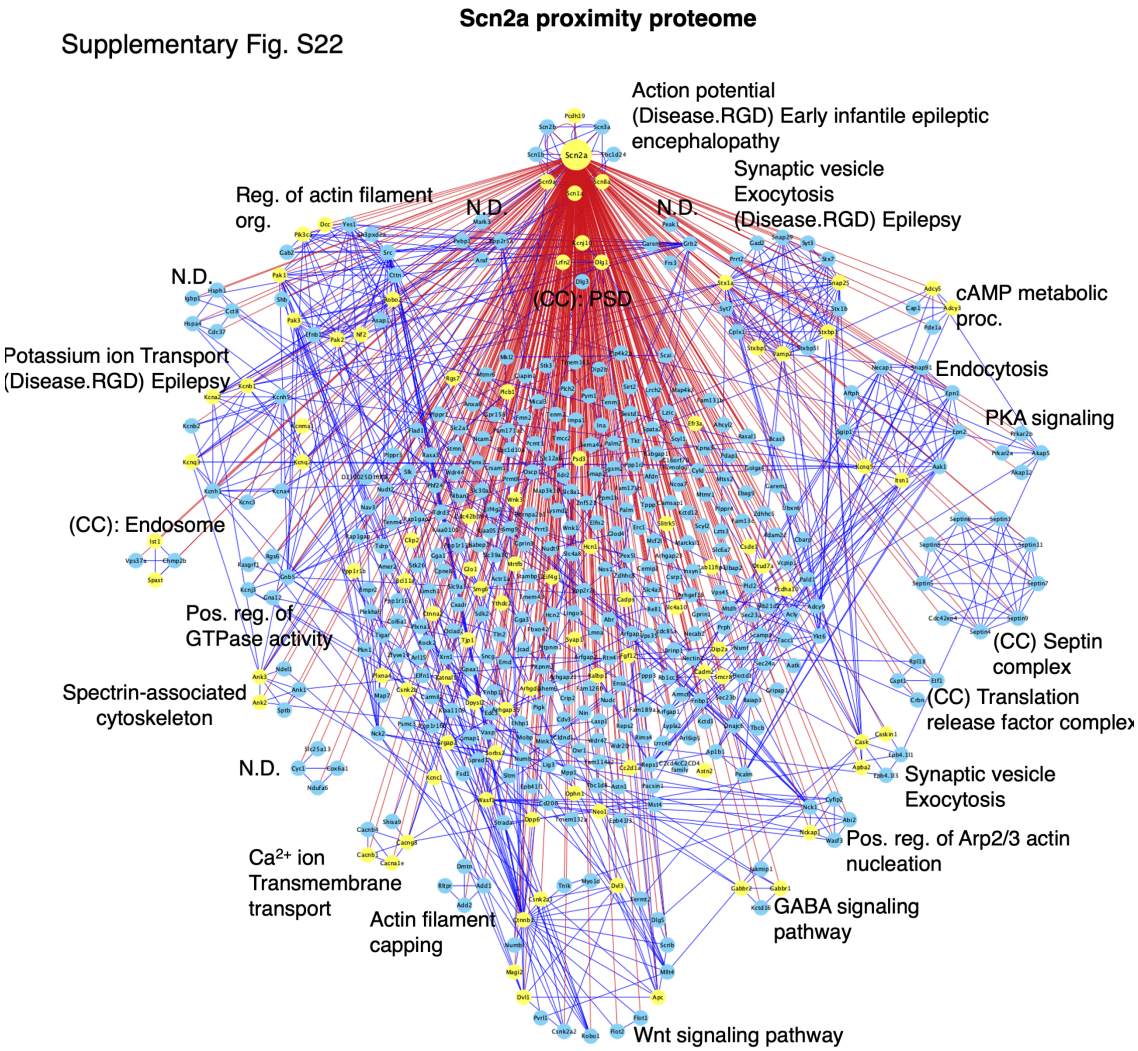


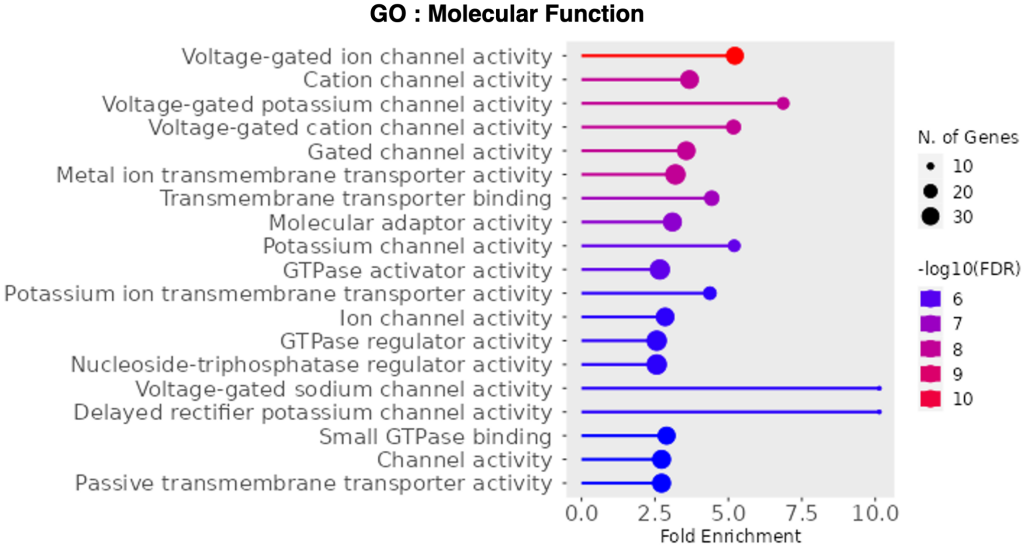
Supplementary Fig. S20. Proximity proteomic networks associated with bait protein Arhgef9.
The proximity proteomic network associated with the bait protein is shown. Blue lines denote STRING interactions and red lines signify identified HiUGE-iBioID interactions. Yellow nodes highlight proteins encoded by SFARI gene orthologs. Annotations denote exemplary significant gene ontology (GO) terms associated with the protein clusters segregated by MCL. Unless otherwise specified, the Biological Process pathway database was used. CC: Cellular Component pathway database. Charts of GO results of the overall bait proximity proteomes using the Molecular Function (MF) pathway database are shown as well, below each network plot.



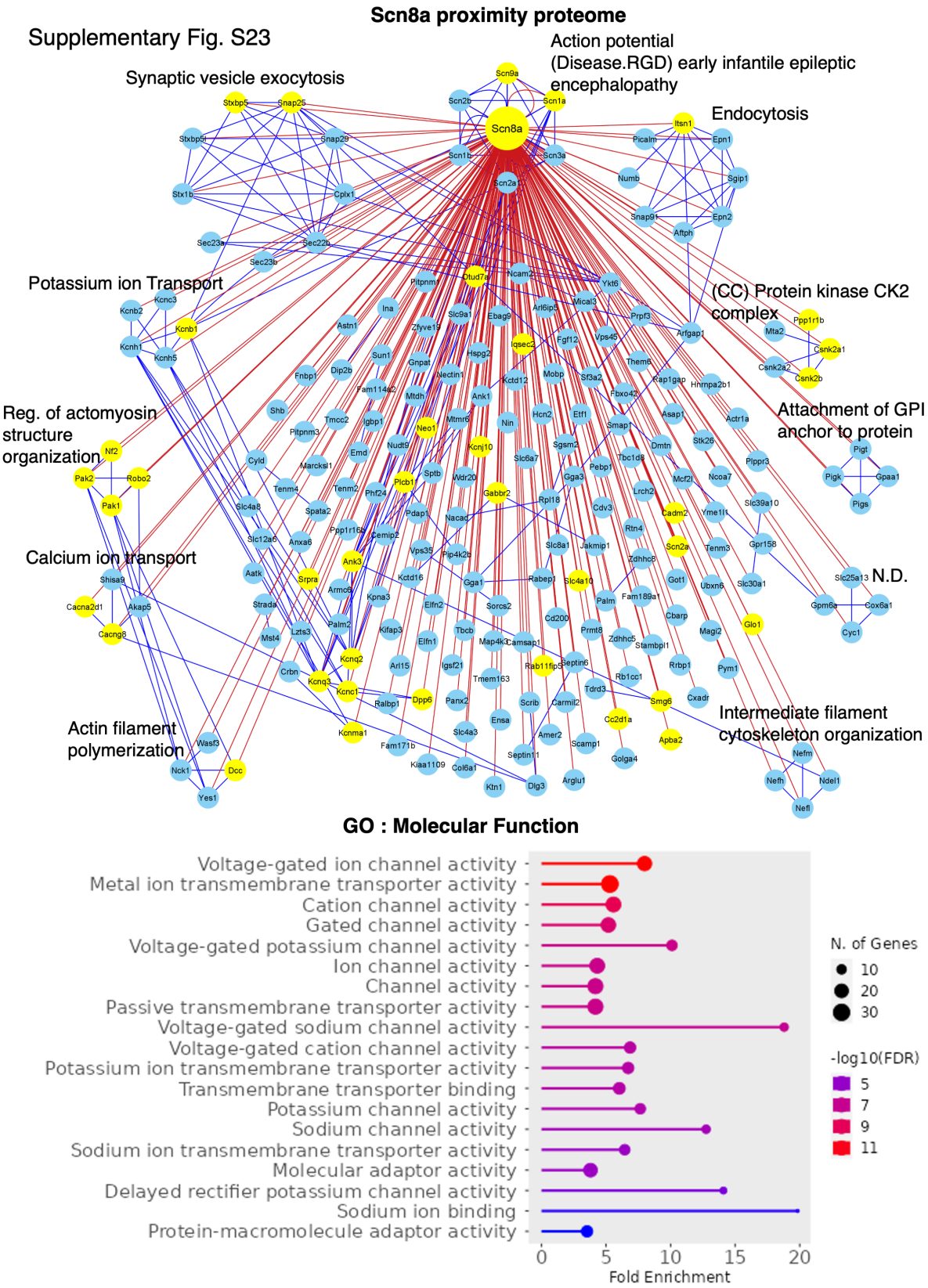
Supplementary Fig. S21. Proximity proteomic networks associated with bait protein Ank3.
The proximity proteomic network associated with the bait protein is shown. Blue lines denote STRING interactions and red lines signify identified HiUGE-iBioID interactions. Yellow nodes highlight proteins encoded by SFARI gene orthologs. Annotations denote exemplary significant gene ontology (GO) terms associated with the protein clusters segregated by MCL. Unless otherwise specified, the Biological Process pathway database was used. CC: Cellular Component pathway database. Charts of GO results of the overall bait proximity proteomes using the Molecular Function (MF) pathway database are shown as well, below each network plot.

Supplementary Fig. S22



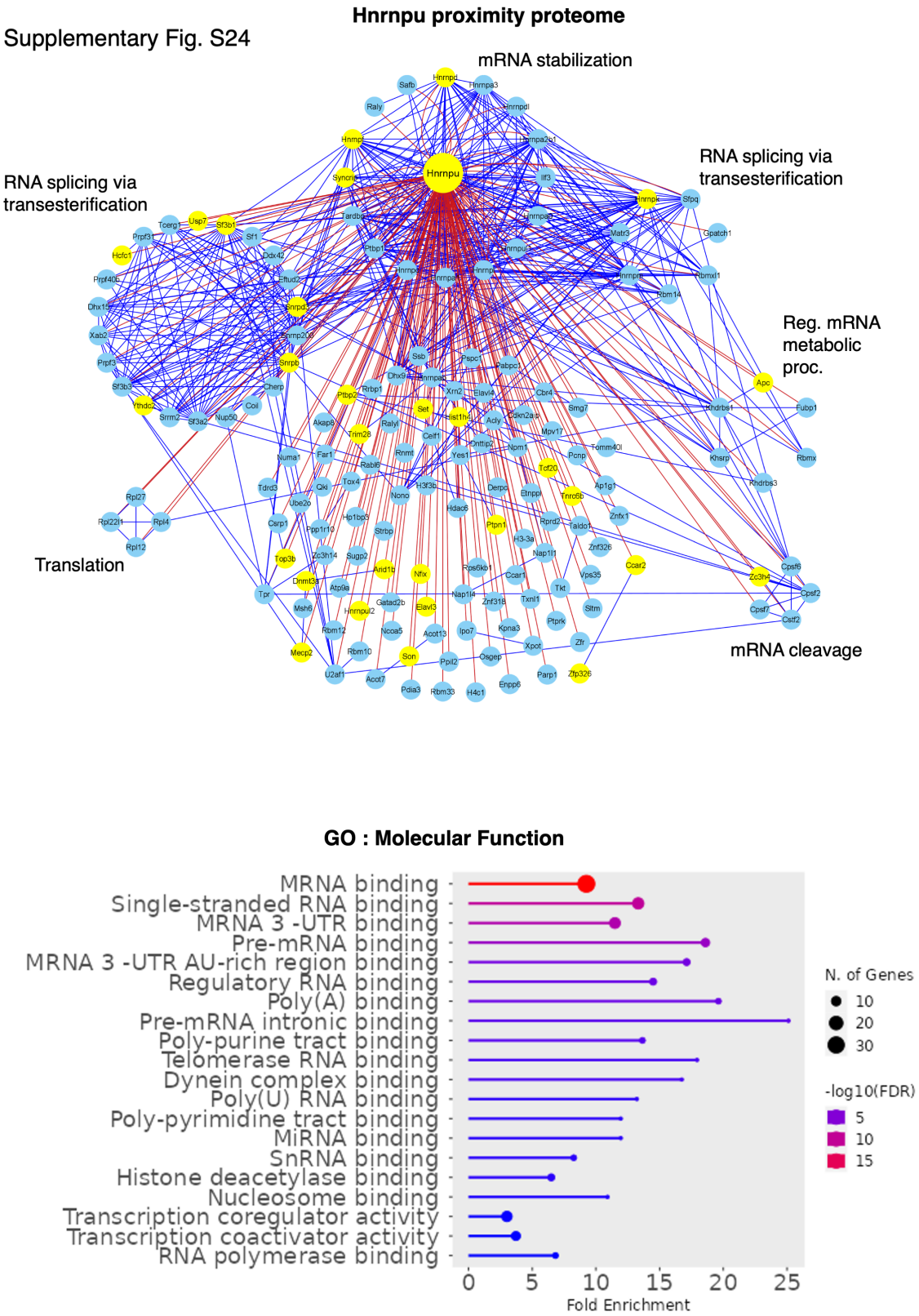


Supplementary Fig. S22. Proximity proteomic networks associated with bait protein Scn2a.
The proximity proteomic network associated with the bait protein is shown. Blue lines denote STRING interactions and red lines signify identified HiUGE-iBioID interactions. Yellow nodes highlight proteins encoded by SFARI gene orthologs. Annotations denote exemplary significant gene ontology (GO) terms associated with the protein clusters segregated by MCL. Unless otherwise specified, the Biological Process pathway database was used. CC: Cellular Component pathway database. Charts of GO results of the overall bait proximity proteomes using the Molecular Function (MF) pathway database are shown as well, below each network plot.



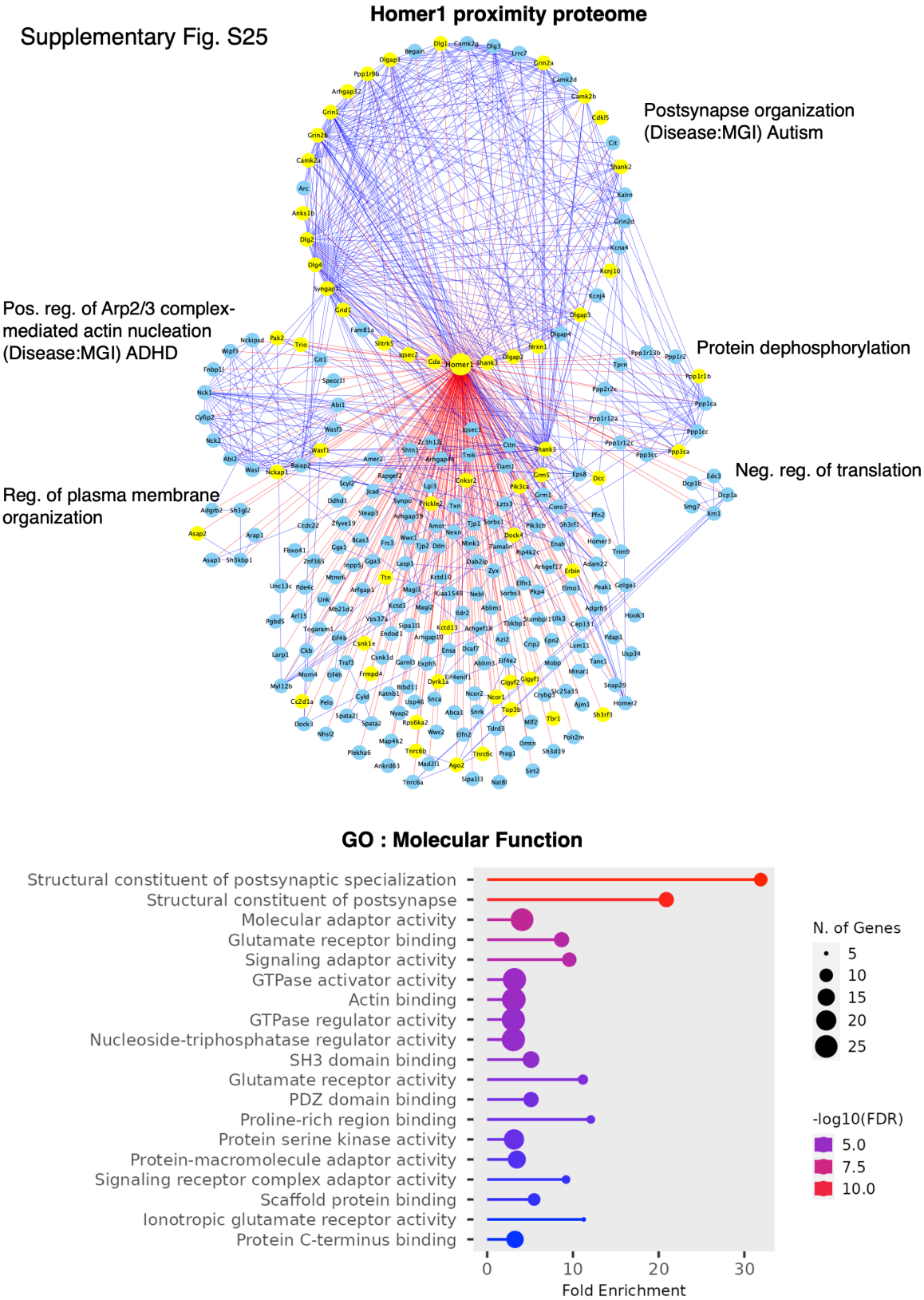
Supplementary Fig. S23. Proximity proteomic networks associated with bait protein Scn8a.
The proximity proteomic network associated with the bait protein is shown. Blue lines denote STRING interactions and red lines signify identified HiUGE-iBioID interactions. Yellow nodes highlight proteins encoded by SFARI gene orthologs. Annotations denote exemplary significant gene ontology (GO) terms associated with the protein clusters segregated by MCL. Unless otherwise specified, the Biological Process pathway database was used. CC: Cellular Component pathway database. Charts of GO results of the overall bait proximity proteomes using the Molecular Function (MF) pathway database are shown as well, below each network plot.

Supplementary Fig. S24



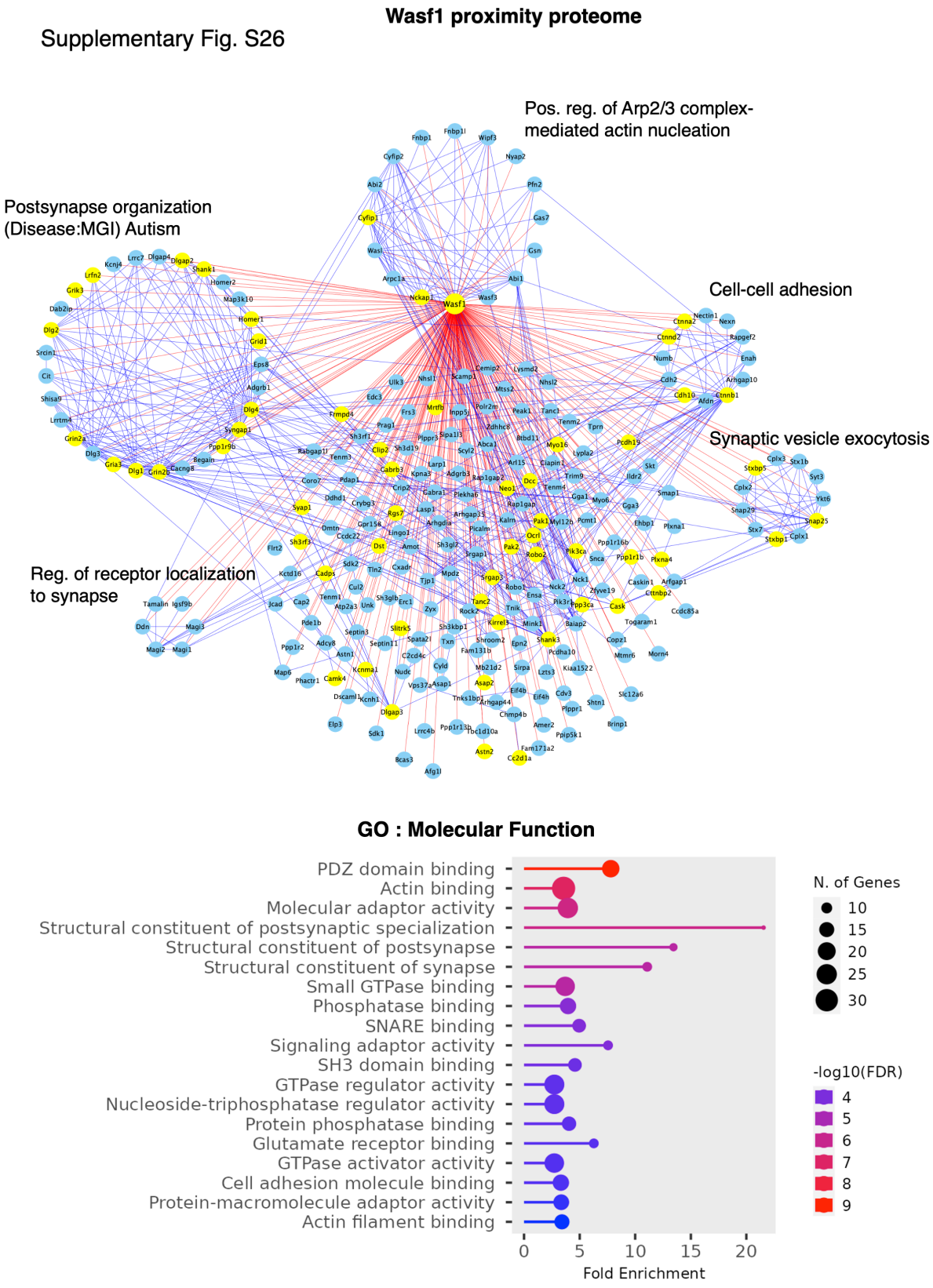
Supplementary Fig. S24. Proximity proteomic networks associated with bait protein Hnrnpu.
The proximity proteomic network associated with the bait protein is shown. Blue lines denote STRING interactions and red lines signify identified HiUGE-iBioID interactions. Yellow nodes highlight proteins encoded by SFARI gene orthologs. Annotations denote exemplary significant gene ontology (GO) terms associated with the protein clusters segregated by MCL. Unless otherwise specified, the Biological Process pathway database was used. CC: Cellular Component pathway database. Charts of GO results of the overall bait proximity proteomes using the Molecular Function (MF) pathway database are shown as well, below each network plot.

Supplementary Fig. S25

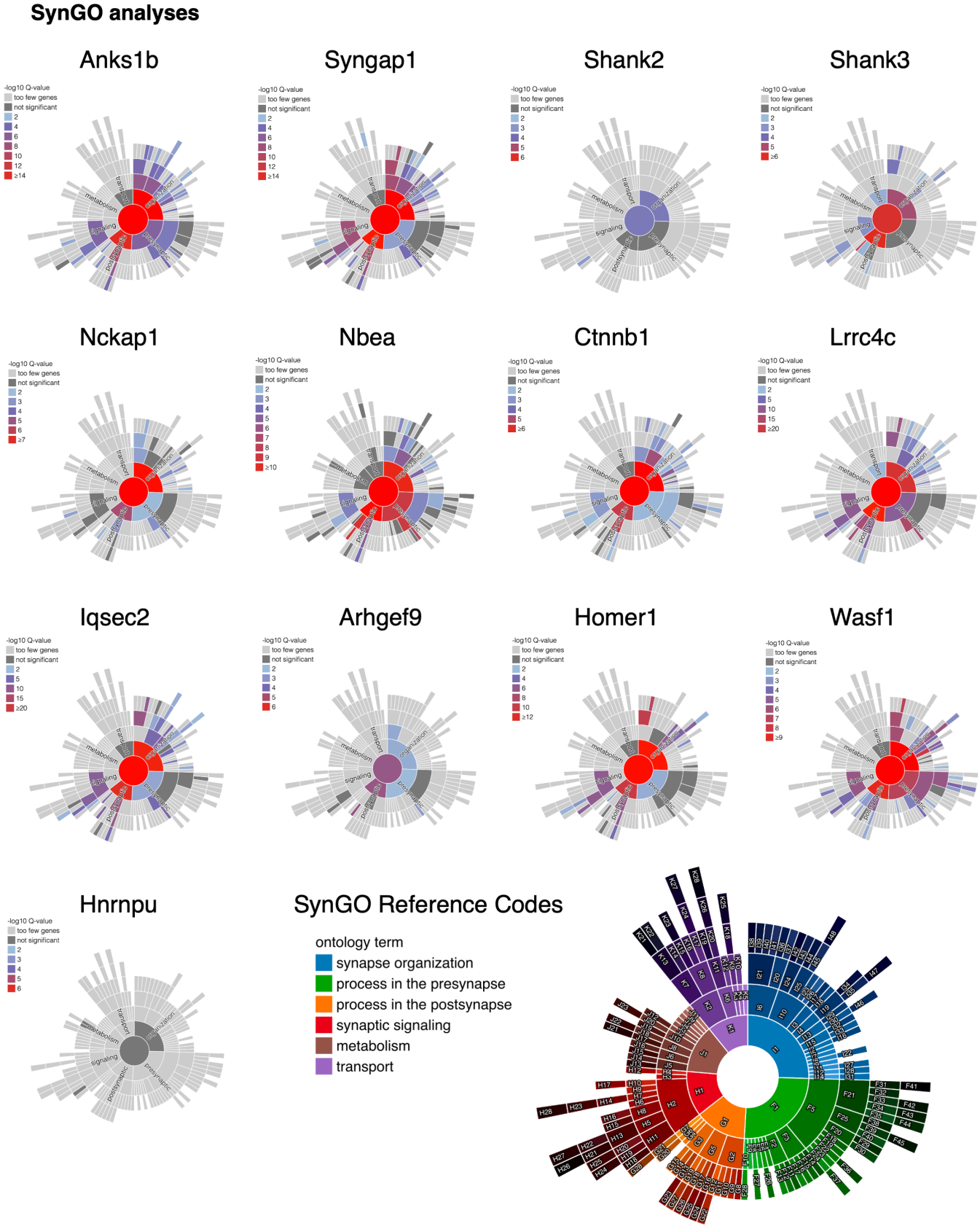


Supplementary Fig. S25. Proximity proteomic networks associated with bait protein Homer1.
The proximity proteomic network associated with the bait protein is shown. Blue lines denote STRING interactions and red lines signify identified HiUGE-iBioID interactions. Yellow nodes highlight proteins encoded by SFARI gene orthologs. Annotations denote exemplary significant gene ontology (GO) terms associated with the protein clusters segregated by MCL. Unless otherwise specified, the Biological Process pathway database was used. CC: Cellular Component pathway database. Charts of GO results of the overall bait proximity proteomes using the Molecular Function (MF) pathway database are shown as well, below each network plot.

Supplementary Fig. S26

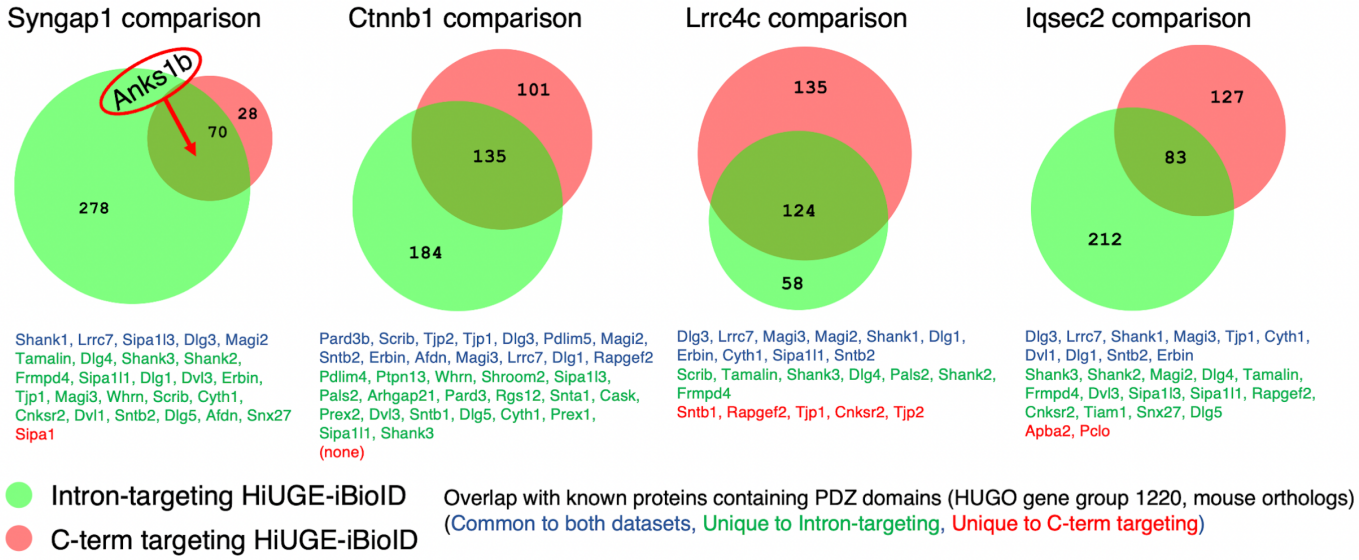


Supplementary Fig. S26. Proximity proteomic networks associated with bait protein Wasf1.
The proximity proteomic network associated with the bait protein is shown. Blue lines denote STRING interactions and red lines signify identified HiUGE-iBioID interactions. Yellow nodes highlight proteins encoded by SFARI gene orthologs. Annotations denote exemplary significant gene ontology (GO) terms associated with the protein clusters segregated by MCL. Unless otherwise specified, the Biological Process pathway database was used. CC: Cellular Component pathway database. Charts of GO results of the overall bait proximity proteomes using the Molecular Function (MF) pathway database are shown as well, below each network plot.



Supplementary Fig. S27. SynGO analyses of the synaptic bait proximity proteomes versus the nucleus bait proximity proteome.

SynGO analyses showing the expected functions of the synaptic bait proximity proteomes mainly within the domains of "synapse organization" and "process in the postsynapse". In contrast, no significant SynGO enrichment was detected for the nucleus Hnrnpu proximity proteome.



Supplementary Fig. S28. Proportional Venn diagram comparisons of different tagging sites on baits. The diagram is color-coded as green = intron tag (preserving the PDZ-binding motif); pink = C-term tag. N represents the number of proteins in each population. The lists of proteins below each bait diagram are of PDZ domain-containing proteins, color-coded as blue = common to both tags; green = intron tag (preserving the PDZ-binding motif); red = C-term tag. Results demonstrate that preservation of the PDZ-binding motif by intron-mediated tag resulted in better identification of interacting proteins with PDZ domains. Proportional Venn diagrams were made using BioVenn [118].