

Integrated multi-omic characterizations of the synapse reveal RNA processing factors and ubiquitin ligases associated with neurodevelopmental disorders

Highlights

- Proteomic profiling of fetal human brain and iNeurons
- Machine-learning model prioritizes high-confidence synapse proteins
- Model interpretation identifies synapse-specific molecular features
- Mapping DDX3X, YBX1, and CUL3 molecular networks uncovers their synaptic interactions

Authors

Yuan Mei, Maya L. Gosztyla, Xinzhu Tan, ..., Marcelo P. Coba, Gene W. Yeo, Trey Ideker

Correspondence

coba@usc.edu (M.P.C.),
geneyeo@ucsd.edu (G.W.Y.),
tideker@ucsd.edu (T.I.)

In brief

This study combines proteomes of multiple biological models with an integrated machine-learning approach to resolve the synapse architecture. By identifying key molecular features of synapse proteins, this study reveals high-confidence synaptic candidates and maps molecular networks of disease-linked proteins CUL3, DDX3X, and YBX1.

Article

Integrated multi-omic characterizations of the synapse reveal RNA processing factors and ubiquitin ligases associated with neurodevelopmental disorders

Yuan Mei,^{1,2} Maya L. Gosztyla,^{2,3,4,5} Xinzhu Tan,⁶ Lara E. Dozier,⁷ Brent Wilkinson,⁸ Justin McKetney,^{9,10,11} John Lee,¹ Michael Chen,¹ Dorothy Tsai,¹ Hema Kopalle,^{2,3} Marina A. Gritsenko,¹² Nicolas Hartel,¹³ Nicholas A. Graham,¹³ Ilse Flores,¹³ Stephen K. Gilmore-Hall,⁷ Shuhao Xu,^{2,3,4,5} Charlotte A. Marquez,¹ Sophie N. Liu,¹ Dylan Fong,¹ Jing Chen,¹ Kate Licon,¹ Derek Hong,¹⁴ Sarah N. Wright,¹ Jason F. Kreisberg,^{1,4} Alexi Nott,^{15,16} Richard D. Smith,¹² Wei-Jun Qian,¹² Danielle L. Swaney,^{9,10,11} Lilia M. Iakoucheva,¹⁴ Nevan J. Krogan,^{9,10,11} Gentry N. Patrick,⁷ Yang Zhou,⁶ Guoping Feng,^{17,18} Marcelo P. Coba,^{8,*} Gene W. Yeo,^{2,3,4,5,*} and Trey Ideker^{1,19,*}

¹Division of Genomics and Precision Medicine, Department of Medicine, University of California, San Diego, La Jolla, CA 92093, USA

²Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA 92023, USA

³Institute for Genomic Medicine, University of California, San Diego, La Jolla, CA 92023, USA

⁴Sanford Stem Cell Institute Innovation Center, University of California, San Diego, La Jolla, CA 92037, USA

⁵Center for RNA Technologies and Therapeutics, University of California, San Diego, La Jolla, CA, USA

⁶Department of Neurology and Neurosurgery, Montreal Neurological Institute-Hospital, Faculty of Medicine and Health Sciences, McGill University, Montreal, QC H3A 1A1, Canada

⁷Department of Neurobiology, School of Biological Sciences, University of California, San Diego, La Jolla, CA 92093, USA

⁸Zilkha Neurogenetic Institute, University of Southern California, Los Angeles, CA 90033, USA

⁹Gladstone Institute of Data Science and Biotechnology, Gladstone Institutes, San Francisco, CA 94158, USA

¹⁰University of California, San Francisco, Quantitative Biosciences Institute, San Francisco, CA 94158, USA

¹¹University of California, San Francisco, Department of Cellular and Molecular Pharmacology, San Francisco, CA 94143, USA

¹²Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99354, USA

¹³Mork Family Department of Chemical Engineering and Materials Science, University of Southern California, Los Angeles, CA 90089, USA

¹⁴Department of Psychiatry, University of California, San Diego, La Jolla, CA 92093, USA

¹⁵Department of Brain Sciences, Imperial College London, White City Campus, London W12 7RH, UK

¹⁶UK Dementia Research Institute, Imperial College London, White City Campus, London W12 0BZ, UK

¹⁷McGovern Institute for Brain Research, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

¹⁸Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

¹⁹Lead contact

*Correspondence: coba@usc.edu (M.P.C.), geneyeo@ucsd.edu (G.W.Y.), tideker@ucsd.edu (T.I.)

<https://doi.org/10.1016/j.cels.2025.101204>

SUMMARY

The molecular composition of the excitatory synapse is incompletely defined due to its dynamic nature across developmental stages and neuronal populations. To address this gap, we apply proteomic mass spectrometry to characterize the synapse in multiple biological models, including the fetal human brain and human induced pluripotent stem cell (hiPSC)-derived neurons. To prioritize the identified proteins, we develop an orthogonal multi-omic screen of genomic, transcriptomic, interactomic, and structural data. This data-driven framework identifies proteins with key molecular features intrinsic to the synapse, including characteristic patterns of biophysical interactions and cross-tissue expression. The multi-omic analysis captures synaptic proteins across developmental stages and experimental systems, including 493 synaptic candidates supported by proteomics. We further investigate three such proteins that are associated with neurodevelopmental disorders—Cullin 3 (CUL3), DEAD-box helicase 3 X-linked (DDX3X), and Y-box binding protein-1 (YBX1)—by mapping their networks of physically interacting synapse proteins or transcripts. Our study demonstrates the potential of an integrated multi-omic approach to more comprehensively resolve the synaptic architecture.

INTRODUCTION

As the site of communication between neurons, the synapse is a key building block of neural circuitry in the brain. It is also central

to disease pathogenesis, as numerous psychiatric diseases, including autism spectrum disorder (ASD), schizophrenia, and epilepsy, are closely associated with synapse dysfunction.^{1–4} Synapses come in distinct types, such as chemical synapses,

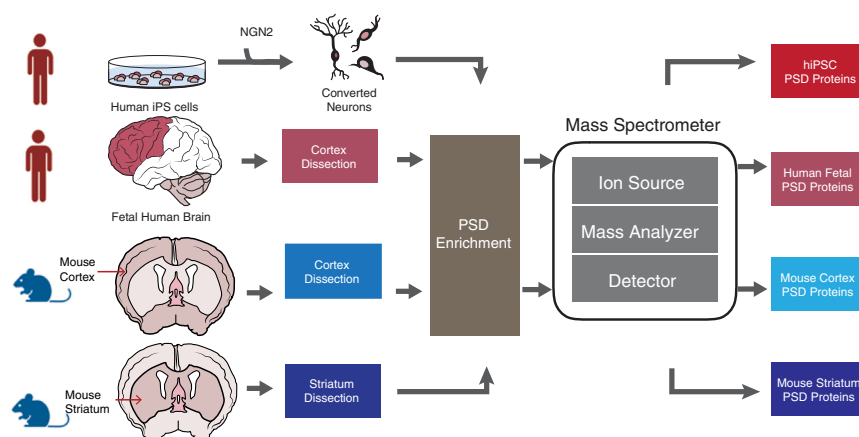


Figure 1. Synapse proteomic screens across diverse biological models

Experimental pipeline for human and mouse synapse proteomics. Human *in vitro* model: hiPSC cells were induced into neurons through NGN2 over-expression. Human *in vivo* model: cortical tissue was dissected from the human fetal brain at the end of the second trimester. From the adult mouse brain, the cortex and the striatum were dissected. These tissues were homogenized and biochemically enriched for PSD. Enriched PSD fractions were subsequently processed and analyzed using liquid chromatography-mass spectrometry (LC-MS).

which shuttle excitatory and/or inhibitory neurotransmitters from one cell to another, and gap junctions, which allow for direct conduction of electrical impulses. Each type is created from the combination of diverse assemblies of molecules, including receptors, signaling enzymes, and master scaffolding proteins. Discoveries and characterizations of proteins at the chemical excitatory synapse, such as the MAGUK protein scaffolds, AMPA/NMDA receptors, and calcium-sensitive vesicle-associated regulators, have laid the foundation for our modern understanding of synaptic transmission.^{5–16} Quantitative synapse proteomics screens as well as computational modeling of synaptic function in rodent brain samples have made marked strides toward achieving a systematic understanding of the synaptic protein network.^{13,17–24} In humans, parallel studies have been predominantly carried out in adult post-mortem brains, with recent investigations beginning to delve into the developing human synapse.²⁵

Due to differences in experimental approaches, diversity of synapses, and technical noise, the sets of proteins identified at the synapse have shown substantial variation across studies. Such differences have motivated major organized efforts by the research community to create unified expert-curated synapse databases, including SynaptomeDB, SynSysNet, Gene Ontology, and, most recently, SynGO.^{26–29} Remaining challenges include how to discover high-confidence synaptic proteins in previously under-investigated tissue types and developmental stages, how to capture poorly studied proteins without sacrificing specificity,³⁰ how to identify the distinct molecular patterns of synapse proteins across studies, and how to sustainably incorporate the wealth of new datasets that are emerging in ever greater numbers, particularly from omics pipelines.^{31–40}

Motivated by these challenges and opportunities, we embarked on a principled, systematic effort to create a more comprehensive map of the excitatory glutamatergic synapse. This effort included four diverse proteomic characterizations, focusing on relatively under-investigated synaptic contexts not as well investigated in previous studies, such as the fetal human synapse and human induced pluripotent stem cell (hiPSC)-derived neurons. We then evaluated the identified protein candidates against a large panel of multi-omics data, which we found harbors a diagnostic pattern of sequence, expression, and physical interaction features indicative of the excitatory synapse. As

we now describe, this integrated proteomic/multi-omic pipeline identified and corroborated hundreds of synapse proteins and candidates across tissues, providing a foundation for systematically resolving diverse synapse structures and functions.

RESULTS

Human fetal and hiPSC-derived neuron proteomic screens identify synapse candidates

To experimentally investigate synaptic proteins across diverse biological models, we used tandem mass spectrometry to characterize the synapse composition in the excitatory glutamatergic synapse in four complementary *in vitro* and *in vivo* contexts (Figure 1). We performed synapse proteomic screens using samples prepared from the adult mouse cortex and striatum. Both of these brain regions have been associated with ASD^{16,41} and shown to be affected by mutation of synapse-associated genes in ASD, such as *Shank3* and *Dlgap3*.^{16,42–44} We selectively enriched postsynaptic density (PSD) fractions prior to mass spectrometry as previously described.^{16,45,46} We hypothesized that fractionating the synaptosome could introduce numerous false positives when coupled with sensitive proteomic analysis; therefore, focusing on the PSD should reduce the number of false positives by yielding fewer proteins than the entire synaptosome fraction. This conservative approach to fractionation might reduce the abundance of some synaptic molecules, but it enhances sensitivity and minimizes contamination from non-synaptic proteins. Through this conservative analysis, we identified peptides that can be mapped to 2,613 genes in the cortex (Table S1) and 1,397 genes in the striatum (Table S2). Approximately 1,200 mouse synapse proteins were identified in both screens, the vast majority of which had also been reported in previous proteomics screens of the adult mouse synapse^{18,47–55} (94%, Figure 2A). This analysis demonstrated that our pipeline can capture synapse proteins consistent with previous studies.

Due to the association of synaptic proteins with neurodevelopmental disorders, including ASD,^{2,17,58,59} we investigated two experimental systems that capture the early human synapse: the human fetal cortex and cultured hiPSC-derived neurons (Figure 1).^{60–63} Others have shown that hiPSC-derived neurons resemble the fetal human brain in molecular composition.^{62,64–67} In our own hands, using established protocols,^{65–67} we found that

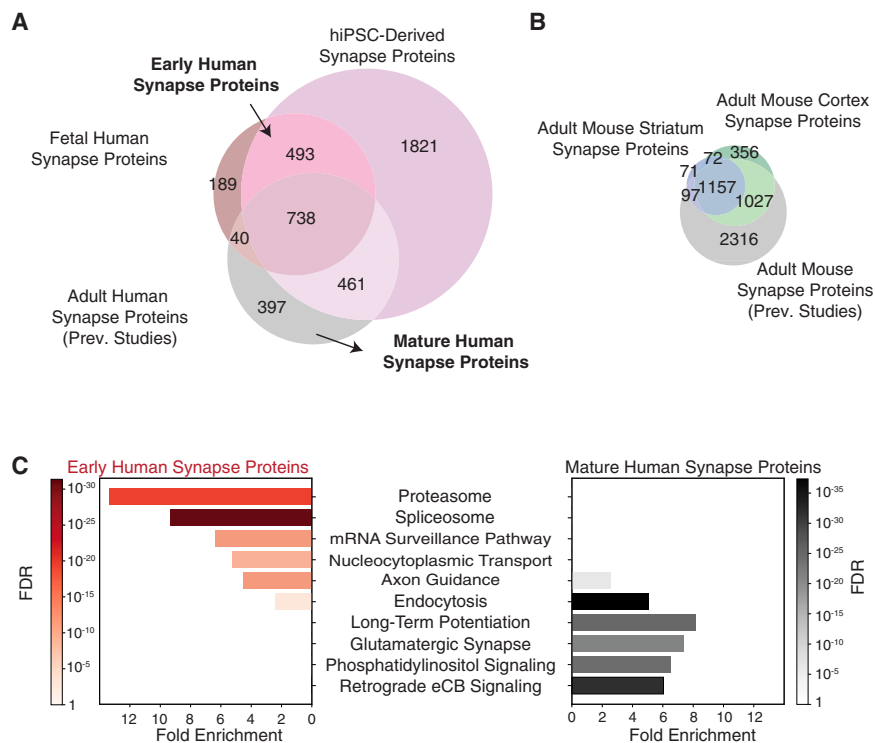


Figure 2. Early human synaptic proteins enrich for proteasomal and RNA processing factors

(A) Venn diagram showing the overlap of the mouse striatum and cortex synapse proteomic screens with previous mouse adult proteomic studies (Abul-Husn et al., 2009; Bayés et al., 2012; Bayés et al., 2017; Biesemann et al., 2014; Li et al., 2016; Collins et al., 2006; Filiou et al., 2010; Moczulski et al., 2014; Pandya et al., 2017; and Roy et al., 2018^{18,47–55}).

(B) Venn diagram showing the overlap of the synapse proteomics screens from the human fetal brain, hiPSC-derived neurons, and adult human tissues. The previous studies on adult human tissues include datasets from Bayés et al. (2011), Bayés et al. (2012), Chang et al. (2015), and Roy et al. (2018)^{13,48,55,56}. Arrows highlight sets of proteins specific to early and mature synapses.

(C) Pathway enrichment⁵⁷ analysis for early human synaptic proteins and mature human synaptic proteins. The color of the bars indicates the false discovery rate for each enrichment. Color scales indicate statistical significance (FDR, log₁₀ scale) with darker shades representing higher significance.

consistent with previous studies, these hiPSC-derived neurons at 3 weeks in culture indeed show morphology and neuronal activity that are indicative of the early stage nature of these neurons. For example, when comparing the field population spike between 3- versus 8-week-old hiPSC-derived neurons, we found that the 3-week hiPSC-derived neurons showed significantly less electrical activity than their 8-week-old counterparts (Figures S1A–S1C). In addition, they also showed a significantly reduced ratio of mushroom versus filopodia spines compared with at 8 weeks (Figures S1D and S1E). These measurements are consistent with the less mature state of their synapses. Furthermore, we also found that using these established protocols, we were able to obtain relatively pure glutamatergic neuron populations (Figure S2), which is consistent with previous studies.^{68,69} We confirmed the healthy karyotypes of these cells and the presence of classic synaptic proteins in the PSD fraction (Figure S3). We used mass spectrometry to more comprehensively analyze the synaptic protein composition in these 3-week-old hiPSC-derived neurons.

For both types of human samples, including fetal human brain and hiPSC-derived neurons, we again selectively enriched PSD fractions prior to mass spectrometry. Through this analysis, we identified peptides that could be mapped to 1,226 genes in the human fetal brain samples from two individuals (Tables S3 and S4) and 2,997 genes in the hiPSC-derived neuron samples from two additional individuals (Tables S5 and S6). Of the approximately 1,000 proteins identified across all four human synapse screens, we also found a substantial number (381, 36%) that appeared to be specific to the early synapse models, as they were not recurrently identified in a set of synapse proteomics data that had been previously collected by other studies in the adult human brain^{13,48,55,56} (Figure 2B). To further dissect the

functional composition of the human synapse candidate molecules, we compared and contrasted the pathway enrichments of the proteins between the early human synapse and mature human synapse. The 381 candidate proteins identified in the early human synapse models showed significant functional enrichments for the proteasome and spliceosome along with mRNA surveillance and nucleocytoplasmic transport functions (Figure 2C). While the core components of the synaptic protein network, such as SYNGAP1, NRXN1, and SHANK2, are present in the early human synapse (Tables S3, S4, S5, and S6), enrichment for synapse-related components and processes was notably absent. This is consistent with many of the identified proteins being synapse candidates not previously annotated as such by the Gene Ontology or in synapse-specific databases like SynGO. By contrast, the proteins found in the mature (adult-only) human synapse screens enriched for expected synapse annotations, including long-term potentiation, glutamatergic synapse, phosphatidylinositol signaling, and retrograde endocannabinoid signaling (Figure 2C). The enrichments from the mature synapse proteomics did not include the RNA and proteasome-related functions associated with the early synapse. This significant divergence in functional enrichments across the early and mature human synapse indicated that the synaptic proteome undergoes substantial remodeling during development. Our findings suggest that proteasomal and RNA processing factors are particularly critical during early development, perhaps due to their key roles in regulating and fine-tuning protein expression in the local translationalome, which then impacts downstream signaling pathways in neurite growth, migration, and survival.

Also notably, of the 1,049 proteins commonly identified across all four screens, 668 proteins were consistent between the early

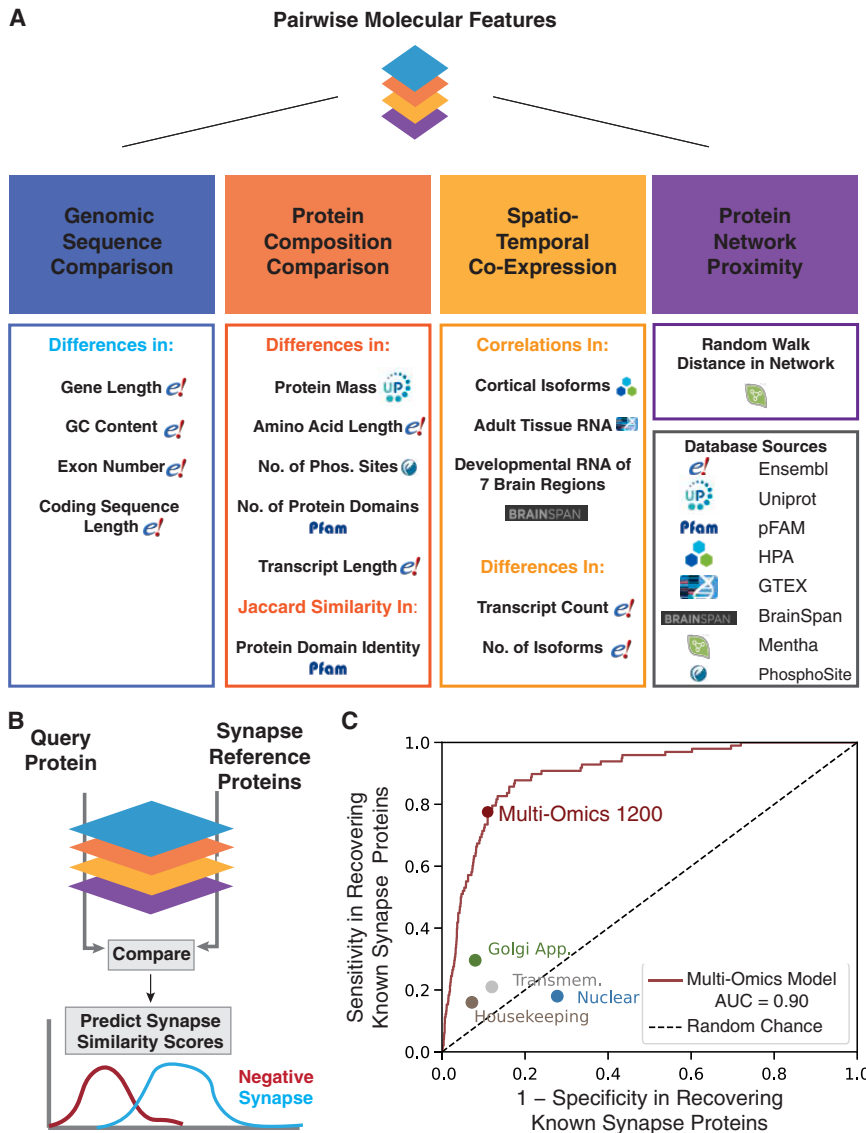


Figure 3. Multi-omics data identify synapse proteins

(A) Pairwise molecular features used for model training. Features are categorized into four broad types (colored columns). Database sources are indicated to the right of each feature. (B) Multi-omic machine-learning workflow. For any query protein, its molecular properties are compared with those of known synapse proteins. From these comparisons for each protein pair, a synapse similarity score is calculated. (C) Synapse similarity scores predicted by the multi-omics machine-learning model are evaluated against known synapse proteins (red curve; AUROC at 0.5 is random chance). The sensitivity and specificity of the model predictions in MO-1200 are compared with those of non-synapse protein sets, including Golgi apparatus proteins, nuclear proteins, transmembrane proteins, and conserved housekeeping proteins (colored points). Red point represents the model corresponding to the Multi-Omics 1200 database.

that simple accumulation of the results thus far identifies 10,947 distinct proteins at the synapse—representing greater than half of the gene products encoded by the human genome (Figures S5A and S5B). More than 4,000 of these proteins have been identified by individual datasets only, without further evidence in previous or subsequent reports (Figure S5C). These observations motivated a need for additional filtering mechanisms to identify higher-confidence hits from proteomic screens.

To prioritize the identified proteins using an orthogonal large-scale approach, we looked beyond synapse proteomics to the plethora of other omics datasets in the public domain of potential relevance to synapse structure or function. We asked if synapse proteins show distinct molecular patterns across multi-omics data, comprising a synaptic signature, and if this signature can be used to systematically map proteins at the synapse. To this end, we limited our searchable pool to only proteins expressed in the human brain with a full panel of multi-omic characterizations (STAR Methods), and we compared each of these proteins with known synapse proteins through a multi-omics panel of pairwise molecular features. The features were drawn from four broad categories of molecular evidence, including genomic sequence, protein composition, spatiotemporal expression, and biophysical protein interactions (Figure 3A, STAR Methods). To identify high-confidence hits, we trained a random forest machine-learning model (Figure 3B, STAR Methods) to evaluate proteins based on the quantitative similarity of their multi-omics features to those of well-established synapse factors curated in the SynGO_CC (cellular component) database as published in 2019. We termed the resulting protein score the “synapse similarity score” and designated proteins with a high score as

and mature stages of the human synapse, indicating their persistent presence throughout development (Figure 2B). Comparing these persistently expressed proteins to those found in mouse samples revealed a high degree of conservation, with nearly all (except for seven proteins) also present in the mouse synapse (Figure S4A). Pathway enrichment analyses indicated that these conserved proteins are involved in fundamental cellular processes vital to synaptic function, such as vesicle-mediated transport, peptide biosynthetic process, and membrane organization (Figure S4B). These results suggest that these stable synapse proteins may be conserved across development and species.

A synaptic signature in multi-omics data

Numerous large-scale studies have been conducted over the past two decades to investigate the synapse proteome in adult mammalian brain tissues, which have served as the foundation for our knowledge of synaptic architecture. We collated the results from 31 of these datasets,^{13,18,47–49,51–56,70,71} observing

and mature stages of the human synapse, indicating their persistent presence throughout development (Figure 2B). Comparing these persistently expressed proteins to those found in mouse samples revealed a high degree of conservation, with nearly all (except for seven proteins) also present in the mouse synapse (Figure S4A). Pathway enrichment analyses indicated that these conserved proteins are involved in fundamental cellular processes vital to synaptic function, such as vesicle-mediated transport, peptide biosynthetic process, and membrane organization (Figure S4B). These results suggest that these stable synapse proteins may be conserved across development and species.

high-priority synapse candidates. The multi-omics model was trained and evaluated using a rigorous 48/12/40 cross-validation procedure, with 48% of SynGO_CC proteins used for training, 12% for validation, and 40% for testing (STAR Methods). Cross-validation insulates the evaluation of a model from feature selection, optimization, and parameter tuning, yielding conservative assessments of model performance and guarding against overfitting.⁷² Following this procedure, we examined the model's predictive power in recovering held-out sets of known synapse proteins. We found that the model achieved high predictive power in recovering the held-out SynGO_CC proteins in the test set, with an area under receiver operating characteristic curve (AUC) of 0.82, and generalized well to proteins documented in two other reference synapse databases, SynSysNet (AUC = 0.79) and SynDB (AUC = 0.78, Figure S6A). It performed especially well for held-out consensus proteins across all three of the synapse databases (AUC = 0.9; Figure 3C). We also examined the performance of the multi-omics model in recovering various negative control protein sets, including conserved housekeeping proteins,⁷³ nuclear proteins,⁷⁴ Golgi apparatus proteins,²⁷ and transmembrane proteins.⁷⁵ Recovery dropped significantly for these other protein lists (Figure S6B), supporting that the model's predictions are specific to molecules at the synapse.

To determine the set of high-priority candidates, we set the synapse similarity score threshold of the multi-omics model to roughly maximize the sensitivity/(1 – specificity) ratio, resulting in 78% sensitivity and 89% specificity. This selection yielded a set of 1,233 machine-learning predicted proteins, which we termed the Multi-Omics 1200 (MO-1200). This includes the proteins predicted beyond the original SynGO_CC training proteins. To determine if the model performance depended on a specific training and test split, we iterated the model with 50 additional random training and test selections. The results showed that the performance was similar across all iterations (Figure S6C).

Because the multi-omics model distinguished synapse proteins from other proteins expressed in the human brain (Figure 3C), it followed that certain biological properties must distinguish this class of proteins. We thus analyzed the random forest model to identify which multi-omic features were the most predictive of synapse proteins, allowing us to create a ranking of feature importance across functional categories (Figure S7; STAR Methods). The most predictive feature was protein network proximity—the random walk distance between protein pairs in the protein-protein interaction network (Figure S7A). Corroborating this finding, we noted that synapse proteins have significantly higher numbers of protein interaction partners (or degrees; $p < 0.0001$; Figure S7B) and shorter path lengths to each other compared with brain-expressed background proteins ($p < 0.0001$; Figure S7C). We also found that the protein proximity feature alone is able to recover synapse proteins and that this predictive strength stems from specific protein-protein interactions rather than network topology alone (Figure S8). Other highly predictive multi-omic features included expression patterns across a wide panel of adult tissue types ($p < 0.0001$; Figure S7D) and gene length ($p < 0.0001$; Figure S7E). Collectively, these results indicate that a data-driven, multi-omics approach is able to recognize the distinct patterns of interaction, expression, and genetic structure that are shared by synapse proteins.

Multi-omic integration increases specificity of synapse proteomics

We next used the orthogonal multi-omics screen to prioritize synapse candidates from the proteomics screens we conducted in the human fetal brain, hiPSC-derived neurons, mouse cortex, and mouse striatum. We found that all four proteomics screens enriched substantially and significantly for the MO-1200 proteins identified by the independent multi-omics analysis (Figure 4A; Figure S9; enrichments ranging from approximately 3- to 6-fold, $p < 0.001$ all screens). Furthermore, each of the 31 existing adult synapse screens we surveyed also enriched for the MO-1200 candidates, including those conducted in the presynapse, postsynaptic proteome, PSD, synaptic vesicle, and synaptic membrane (Figure 4A). To avoid potential bias, we excluded the SynGO_CC training proteins from all enrichment analyses, thereby focusing our evaluation solely on the accuracy of the predictions.

We also assessed the enrichment of the synapse proteomic screens for the sets of proteins that have been associated with various synaptic and non-synaptic diseases, before and after integration with MO-1200. After integration, the fold enrichment was increased substantially for diseases linked with defects in synaptic function, such as ASD,^{3,76,77} epilepsy,⁷⁸ schizophrenia,⁷⁹ intellectual disability,⁸⁰ and Alzheimer's disease.⁸¹ By contrast, for diseases with no known synaptic component, such as neuroblastoma, multiple sclerosis, height, Crohn's disease, congenital heart disease, and nasopharyngeal carcinoma,⁸¹ there was no enrichment before or after integration (Figure 4B). These analyses showed that the multi-omic screen, by identifying a synaptic signature, is able to capture proteins across synaptic compartments, tissue types, developmental stages, and synapse-associated diseases.

Thus, after demonstrating the ability of the multi-omic screen to capture synaptic proteins in an orthogonal approach, we used these results to prioritize the high-confidence hits from proteomic screens. Accordingly, we created a complete non-redundant list of proteins identified in our and previous synapse proteomics screens, then intersected this list with the MO-1200 candidates. Because we aimed to maximize specificity and minimize false discoveries, we chose stringent thresholds for the multi-omic predictions as mentioned above. This integration step resulted in a set of 1,025 omics-validated high-confidence synapse proteins beyond the training proteins, which we call Synaptic Signatures 1000 (SynSig-1000; Figure 5A). Notably, using these stringent criteria, we found that the number of supporting proteomics screens was significantly and substantially higher for SynSig-1000 proteins than for the entire cumulative set of proteins identified by synapse proteomics studies (Figure 5B, $p < 0.0001$ by Student's *t* test, median of 11 versus 2 supporting screens). These results suggested that the integration of synapse proteomics screens with multi-omic filtering can increase the specificity of synapse protein identification. Interestingly, we found that even though this integration removed many irreproducible proteins, it was still able to capture synaptic proteins in relatively understudied biological models such as the human fetal brain and hiPSC-derived neurons (Table S7) due to the presence of the synaptic signature across tissues. The full list of MO-1200, SynSig-1000, and the presence of each protein within each proteomic study is compiled in Table S7.

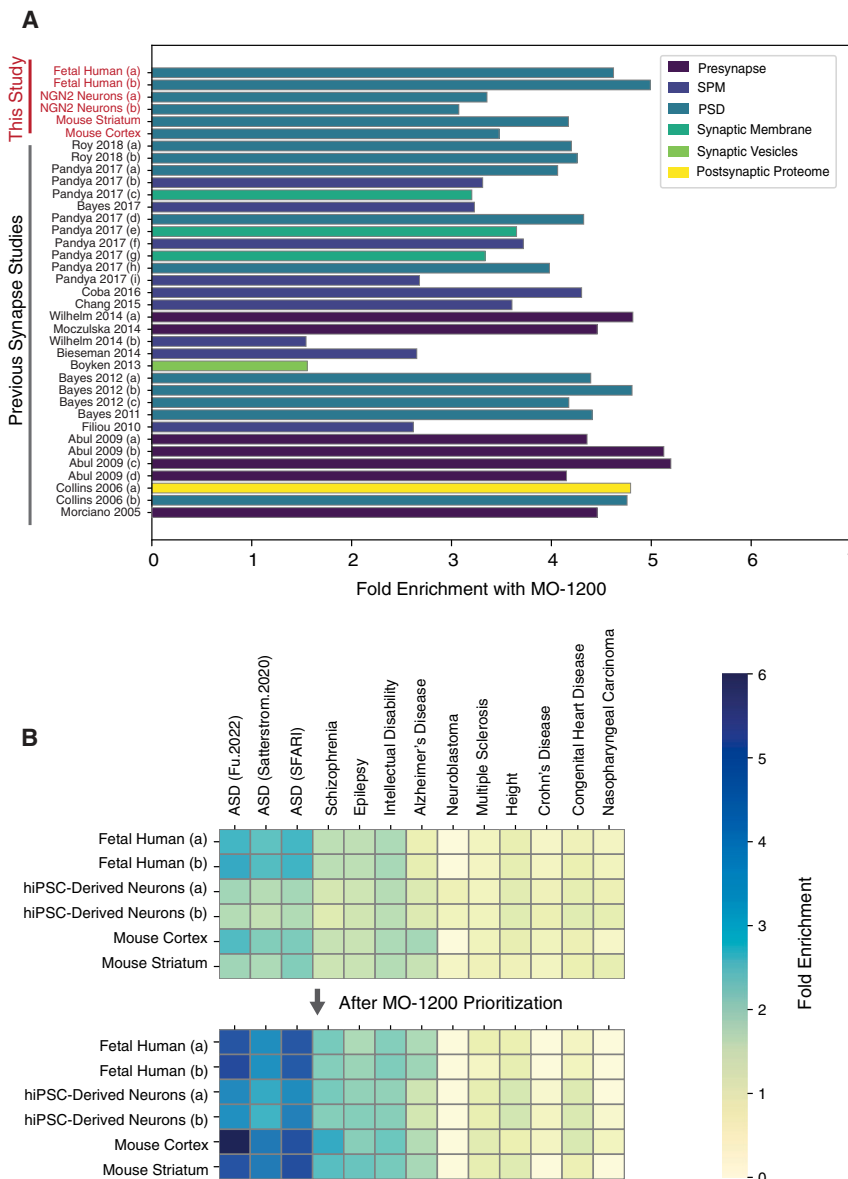


Figure 4. Integration of proteomics and multi-omics screens identifies diverse synapse proteins

(A) Graph shows significant enrichment of MO-1200 (linear scale) across our synaptic screens and 31 large-scale previous datasets. The color of each bar indicates the synaptic compartment focused on by each dataset ($***p < 0.001$ all screens; hypergeometric test with Benjamini-Hochberg multiple comparison correction). All analyses exclude training genes to avoid bias.

(B) The two heatmaps show the fold enrichment of human synapse proteomic datasets for a panel of disease gene sets. Color scale indicates fold enrichment (linear scale, ranging from 0 to 6), where darker blue represents higher fold enrichment and pale yellow represents low/no enrichment. Analysis was completed using hypergeometric test. The top heatmap shows the disease enrichment using proteomic results before any prioritization. The bottom heatmap shows the disease enrichment after prioritizing each proteomic dataset using MO-1200. All analyses exclude training genes to avoid bias.

omics results had not sacrificed coverage of the resulting dataset.

YBX1, an RNA-binding protein, localizes to the synapse and binds synaptic transcripts

We next sought to further study three prioritized SynSig-1000 candidates that showed preferential expression in the early human fetal brain (Figure S10) and are linked to neurodevelopmental disorders with unclear mechanisms. One of the SynSig-1000 candidates found in all four of our synapse proteomic screens was Y-box binding protein-1 (YBX1, also known as YB-1), a multifunctional nucleic-acid binding protein that has been reported to regulate transcription, translation, and pre-mRNA splicing.⁸² Although YBX1 had been associated with brain development,^{83,84} it had been studied predominantly in cancer due to its elevated expression in many tumor types.⁸⁵ It was not mapped to the synapse in the peer-reviewed SynGO_CC 20180731 release and was not part of our training genes. To investigate this predicted synapse protein, we used immunostaining and high-resolution microscopy to localize YBX1 within primary cultures of rat hippocampal neurons. We found that YBX1 has a punctal staining pattern along dendritic segments, and co-staining showed that YBX1 and DLG4 co-localize along neuronal projections (Figure 6A; Figure S11A). In addition to the proteomic results and multi-omic predictions, these results provided orthogonal evidence that YBX1 is part of the PSD.

Because YBX1 is an RNA-binding protein, we used enhanced UV cross-linking and immunoprecipitation (eCLIP) followed by deep sequencing⁹⁰ to discover the RNA targets of YBX1 in hiPSC-derived neurons (Figure 6B; STAR Methods; Table S8).

SynSig-1000 included 532 known synapse proteins and 493 new candidates not documented in previous peer-reviewed literature-curated databases, including SynGO_CC (release 20180731), SynaptomeDB (v.2020), or SynSysNet (v.2020). As with these databases, the SynSig-1000 candidates covered diverse molecular functions such as receptors/channels, vesicle-associated proteins, and scaffolds (Figure 5C). The relative gains in synapse protein identifications were highest for functions related to nucleic-acid binding, protein translation, and proteasome/ubiquitin, which include many RNA-binding proteins and ubiquitin ligases. The substantial expansions in these molecular categories were consistent with the enrichment we observed for the proteasome, spliceosome, and mRNA surveillance pathway in our earlier analysis of the proteomics data from the early human synapse (Figure 2C). Moreover, the persistent broad representation of diverse synapse functions suggested that use of the multi-omics signature to prioritize prote-

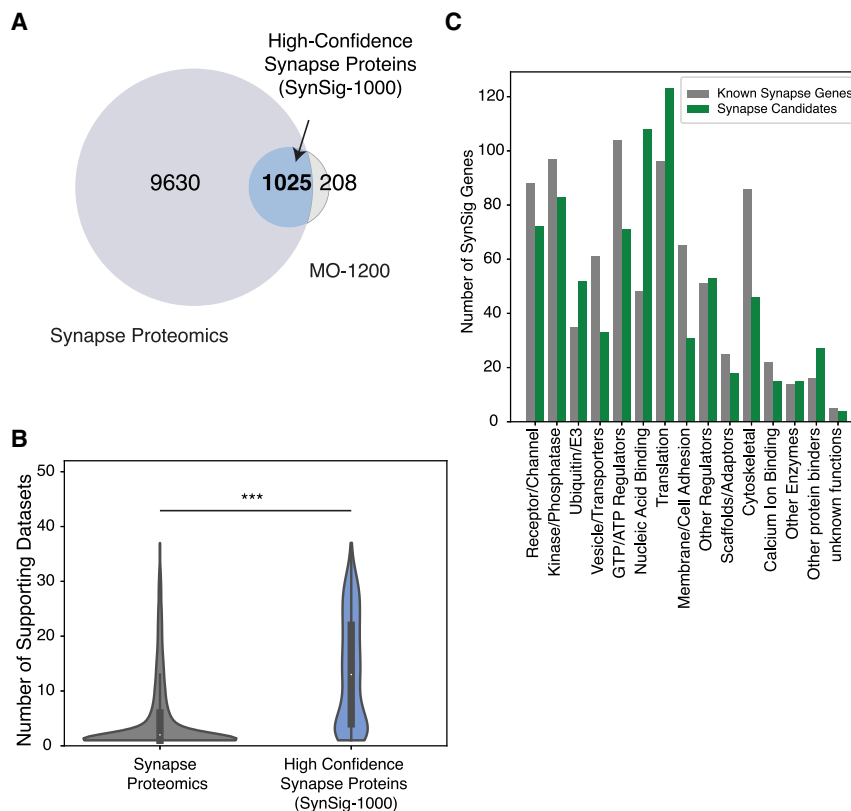


Figure 5. SynSig-1000 identifies experimentally reproducible proteins across diverse functions

(A) Venn diagram showing overlap between all proteins identified in experimental synaptic screens and the filtered set MO-1200. The intersection of 1,025 high-confidence synaptic proteins is called SynSig-1000.

(B) Distribution of the number of supporting experimental datasets for proteins in unprioritized synapse proteomics ($n = 9,630$) versus prioritized SynSig-1000 ($n = 1,025$). *** $p < 0.001$, two-tailed Student's *t* test.

(C) Grouped bar graph shows the molecular function of SynSig-1000 proteins, broken down into both known and candidate synapse proteins. Within each group, the proteins are further categorized according to their molecular function. Known synapse proteins are shown in gray, and new synapse protein candidates are shown in green.

images and protein-RNA interaction mapping further elucidate the potential role of YBX1 at the PSD.

DDX3X is an RNA helicase at the excitatory synapse

Another candidate synapse protein from our integrated pipeline was DEAD-box helicase 3 X-linked, or DDX3X. It is an

Two replicates of immunoprecipitation (IP) and two replicates of size-matched input (IN) samples were processed and sequenced to ~ 30 million reads on average. We obtained $\sim 25,000$ binding sites discovered by the CLIPPer algorithm,⁹¹ satisfying ENCODE statistical thresholds of adjusted $p < 0.001$, 8-fold higher than input.⁹² We found that YBX1 binds predominantly to the coding sequence (CDS) and 3' untranslated regions (UTRs) of 7,768 protein-coding genes (Figure 6C), consistent with its reported role in translational regulation.⁹³ The most significantly enriched binding motif was AUCAUC ($p < 0.0001$), present in 58.2% of the targets (Figure S12A), consistent with previous reports of a UYAUC motif for YBX1.^{94,95} The most highly significant RNA interactors enriched for genes annotated to the PSD and other synapse-related locations compared with the background of genes expressed in the hiPSC-derived neurons (Figure 6D). The high-confidence RNA targets also enriched significantly for genes annotated to the synapse in databases including SynGO, SynSysNet, and SynDB, as well as our MO-1200 resource and synapse proteomic screens (Figures S12B and S12C). An example YBX1 binding target is CNIH2, which encodes cornichon family AMPA receptor auxiliary protein 2—a protein with known synaptic localization.⁹⁶ Irreproducible discovery rate analysis (IDR)⁹⁰ showed significantly enriched binding peaks of YBX1 across multiple CNIH2 exons compared with the IN controls (IN; Figure 6E). Finally, analysis of YBX1 binding targets showed significant enrichment for autism risk factors across multiple studies and risk factor databases, including the binding targets of autism risk genes CHD8 and FMR1^{3,4,76,77,87–89} (Figure 6F). Collectively, the co-localization

ATP-dependent helicase that has been found to regulate RNA splicing, export, and translation in non-neuronal contexts.⁹⁷ While mutations in DDX3X have been associated with neurodevelopmental disorders including ASD and intellectual disability,^{98,99} its localization to the synapse and interactions with synaptic molecules have not been extensively studied. It was not found in SynGO_CC release 20180731 and was not part of our training genes. Like YBX1, DDX3X was found in all our proteomic results from the fetal human, hiPSC-derived neurons, mouse cortex, and mouse striatum (Tables S1, S2, S3, S4, S5, and S6). To further validate the presence of DDX3X at the synapse using orthogonal approaches, we used targeted western blotting in biochemically enriched synaptosome (SPM) fractions in mouse brains and found abundant DDX3X expression (Figure S13).

To determine whether and how DDX3X interacts with other synaptic proteins, we mapped the DDX3X interactome in the endogenous mouse brain tissue using affinity-purification mass spectrometry (AP-MS; Figure S14A; STAR Methods). This experiment identified 25 interaction partners (Table S9) that were significantly enriched for canonical synapse proteins in SynGO, including CAMK2A, CYFIP2, DLG1, and DLG2. The interaction partners were also enriched for ASD risk factors such as CSNK2A1^{3,100} and STXBP1^{3,76,101} (Figures S14B and S15A), indicating that DDX3X mutations may act in signaling pathways dysregulated in ASD. Next, we used immunostaining in primary rat hippocampal neurons and found that DDX3X has a subcellular spatial localization pattern similar to that of known synapse markers. In particular, high-resolution confocal imaging showed

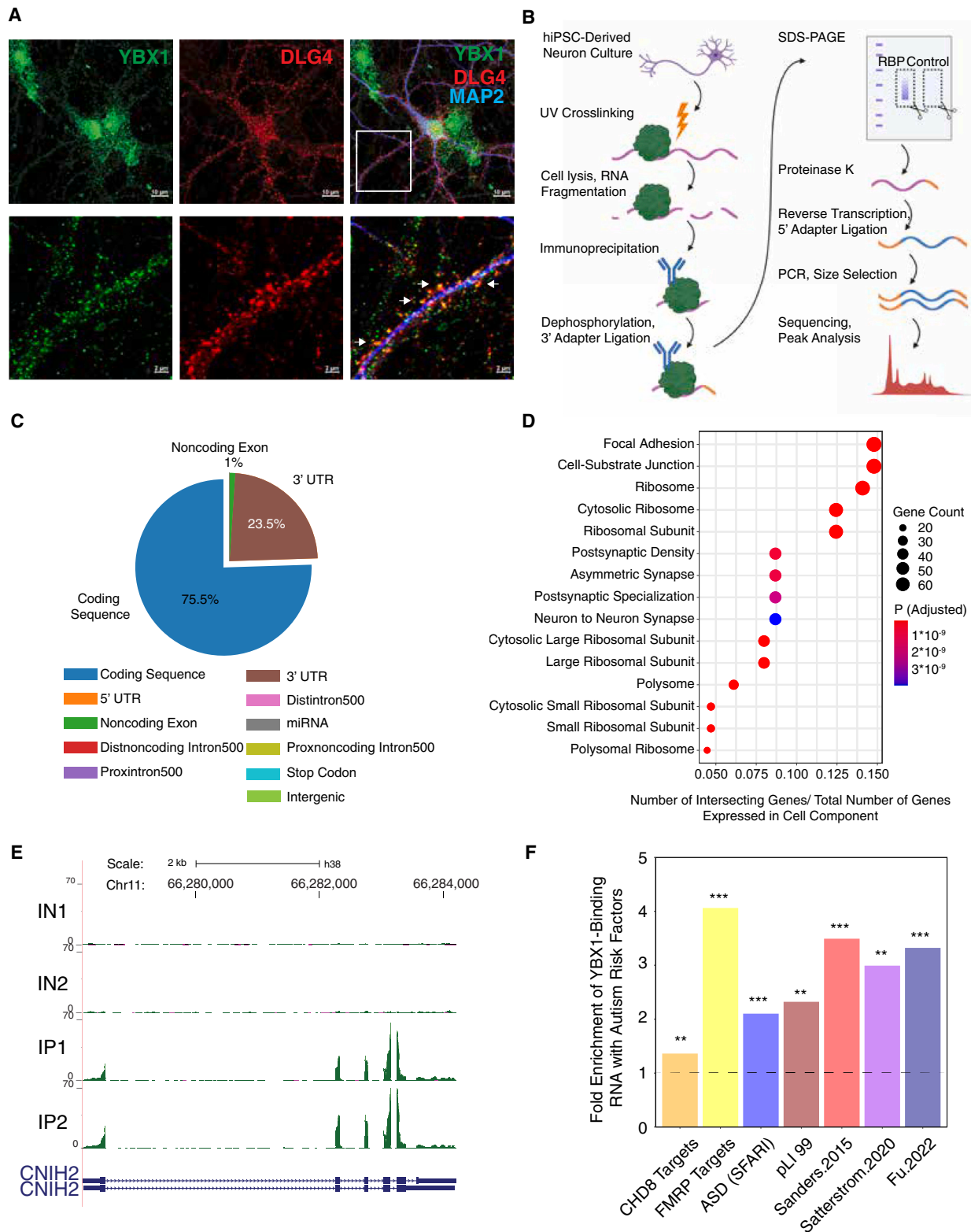


Figure 6. YBX1 binds synapse and ASD-associated transcripts

(A) Representative confocal images of YBX1 in primary rat hippocampal neurons showing co-localization of synapse candidate protein YBX1 and known synapse marker DLG4. Top row: YBX1 (green), DLG4 (red), and MAP2 (blue) staining across the neuron. Scale bar, 10 μ m. Bottom row: magnified images. Magnified (legend continued on next page)

that, along neuronal projections, DDX3X co-localizes with synapse marker DLG4 (Figures S11B and S14C), demonstrating its presence within dendritic spines. We also sought to investigate the expression pattern of DDX3X within human neurons due to its presence in our human fetal and iPSC-derived synapse proteomics and because recent reports had associated DDX3X mutations with human neurodevelopmental disorders.^{102,103} Using hiPSC-derived neurons, we again immunostained for DDX3X and found that it co-localized with Synaptophysin along dendritic segments (Figure S16A). These results showed that DDX3X localizes to the synapse and that such localization is conserved across rodents and humans.

To determine the RNA targets of DDX3X, we again performed an eCLIP experiment in hiPSC-derived neurons. Approximately 30 million reads on average were obtained from two replicates of IP and IN samples. We found that DDX3X interacts with RNAs from 8,377 genes through a total of ~65,000 binding sites in (adjusted $p < 0.001$, 8-fold higher than input; Table S10). These experiments indicated that DDX3X binds predominantly to CDS regions and 5' UTRs (Figure S14D). We identified multiple statistically significant binding motifs, the top being the Kozak translational initiation sequence CCAUGG ($p = 10^{-84}$), consistent with DDX3X's previously reported role in translational control^{103,104} (Figure S16B). The most significant RNA-binding targets enriched for genes associated with the PSD in both literature-curated databases and our experimental synapse screens (Figures S14E, S14F, and S16C; STAR Methods). An example RNA target is CACNG4, which encodes calcium voltage-gated channel auxiliary subunit gamma 4 at the synapse.^{105–107} We confirmed this target via IDR analysis, which showed significantly enriched binding peaks across CACNG4 exons in the DDX3X pull-down conditions compared with IN controls (Figure S14G). Furthermore, analyses of both the AP-MS protein interactors and eCLIP RNA-binding targets showed significant enrichment for autism risk factors,^{3,4,76,77,87–89} suggesting that DDX3X plays a role in multiple disease pathways through various molecular interactions (Figures S14B, S15B, and S16D). Collectively, our AP-MS, microscopy, and eCLIP results consistently indicated that DDX3X is a part of the synaptic cellular compartment and interacts with many known synaptic molecules.

CUL3 interacts with ASD risk factors at the synapse

As a third exploration of synapse proteins identified by the SynSig-1000 resource, we investigated Cullin 3 (CUL3), an E3 ubiquitin ligase and known genome-wide significant risk factor for ASD.⁷⁶ We had previously demonstrated that CUL3 haploinsufficiency leads to autism-like phenotypes and neuronal cytoskeleton dysfunction in animal models.^{108,109} It is not a member of the cellular components of SynGO and was not part of our

training genes. Whether CUL3 physically localizes to the synapse and is part of the synaptic protein network remains unclear.

We found evidence that CUL3 is part of the synaptic subcellular compartment in our synapse proteomics screens in adult mouse striatum, adult mouse cortex, and hiPSC-derived neurons (Tables S1, S2, S5, and S6). To further investigate the spatial localization of CUL3 in the synapse, we used immunofluorescence staining in primary rat hippocampal cultures, finding that CUL3 is abundant along neuronal projections, with partial co-localization with synaptic marker DLG4 along dendritic segments (Figure 7A; Figure S11C).

To determine whether and how CUL3 interacts with synaptic proteins, we used AP-MS to map its protein binding partners in primary mouse brain tissue. Using endogenous CUL3 as bait for IP in the mouse cortex (Figure 7B; STAR Methods), we found that CUL3 interacts with 136 proteins, including multiple components of E3 ubiquitin ligase complexes such as KLH22, KLHL9, and KLHDC5 (Table S11). The interactome of CUL3 significantly enriched for known synapse proteins, including classic synapse markers such as SHANK3, SYNGAP1, NRXN1, and NMDA receptor subunits (15.85-fold enrichment with SynGO_CC; p value = 6.36×10^{-111} ; hypergeometric test). Notably, we found that many of the interaction partners of CUL3 are also known risk factors for ASD,⁷⁷ including SHANK3, NRXN1, NRXN3, SYNGAP1, and our SynSig-1000-predicted DDX3X protein.

To determine which interacting proteins are also modulated by CUL3, we investigated the differentially expressed proteins (DEPs) at the synapse in our previously generated CUL3 haploinsufficient mouse.¹⁰⁸ Using biochemical fractionation of cell lysates, cytoplasm, and SPM from mouse brains, we confirmed that, as expected, CUL3 expression is decreased in haploinsufficient mutants compared with wild type (Figure 7C; Figures S17A and S17B). Western blot analysis in the synaptosome fraction showed that CUL3 is significantly decreased by about 40% in the mutant samples (Figure 7C; Student's t test, $p < 0.05$). We then used quantitative proteomics to identify DEPs between wild-type and CUL3-haploinsufficient mouse cortex, with separate analyses performed for whole-cell and synaptosome fractions. We found a much greater number of DEPs at the synapse (Figure 7D; Table S12) compared with the whole cell (Figure S17C; Table S13), suggesting a profound effect of CUL3 mutation specific to the synapse (249 DEPs in the synaptosome, 36 DEPs in the whole cell). In the synaptosome fraction, we found a significant reduction in many synapse proteins, including SYNGAP1, HPCAL4, DLG2, and VIM (Figure 7D; Table S12). Pathway analysis of the DEPs showed significant enrichment for multiple processes, including synaptic membrane, anterograde synaptic signaling, glutamatergic synapse, and neurotransmitter transport (Figure S17D).

regions focus on a single dendritic segment indicated by white box in the top row image. Scale bar, 2 μ m. White arrowheads indicate representative positions of YBX1 and DLG4 co-localization. Magnified regions focus on a single dendritic segment indicated by white box in the top row image.

(B) Experimental workflow of the eCLIP assay.

(C) Pie chart showing distribution of YBX1 binding sites on detected RNA species.

(D) GO cellular component enrichment⁹⁵ of the most significant RNA species (approximately 500) bound to YBX1.

(E) RNA-sequencing read density tracks along CNIH2 for both experimental duplicates (IP1 and IP2) compared with size-matched input controls (IN1 and IN2). Bottom tracks in blue indicate the two isoforms of the CNIH2 gene, with the taller blue boxes representing exons and the shorter blue boxes representing UTRs.

(F) Fold enrichment of the top YBX1-bound RNAs (approximately 500) for autism risk gene sets drawn from literature and genome-wide association studies^{3,4,76,77,87–89} (** $p < 0.0001$; hypergeometric test with Benjamini-Hochberg multiple comparison correction).

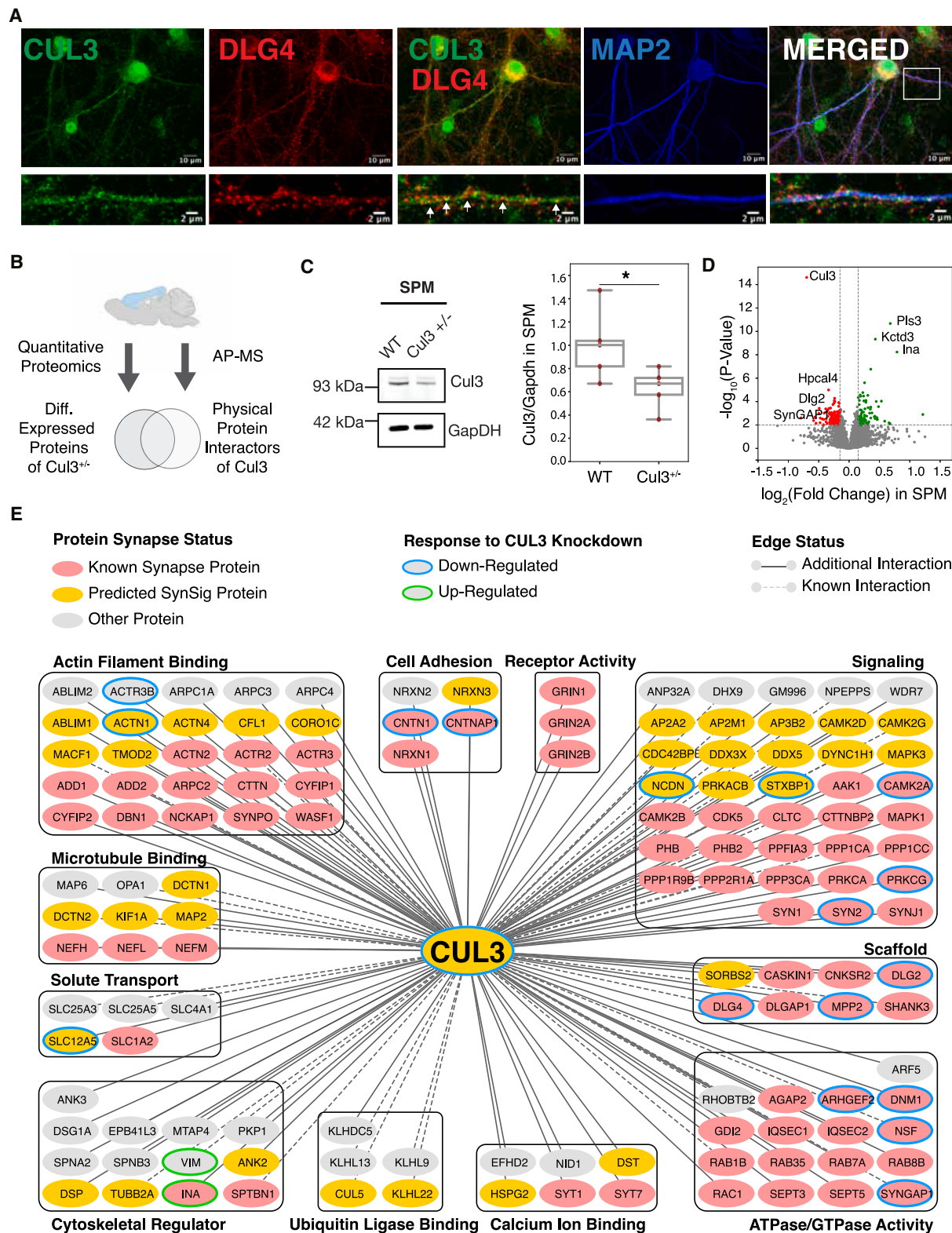


Figure 7. CUL3 interacts with and modulates proteins at the synapse

(A) Representative confocal images of CUL3 in primary rat hippocampal neurons showing the co-localization of CUL3 and synapse marker DLG4 (or PSD-95). Top row shows CUL3 (green), DLG4 (red), and MAP2 (blue) staining across the rat hippocampal neuron. Bottom row shows the magnified images of the neuron (legend continued on next page)

The integration of the AP-MS results and the DEPs culminated in a functional protein interactome map of the CUL3 that shows both the physical interactors of CUL3 as well as the differentially regulated proteins (Figure 7E). One of the CUL3 protein interactors, SYNGAP1, was significantly reduced in the CUL3-mutant synaptosome fraction, suggesting that mutations in CUL3 may impact other ASD risk factors like SYNGAP1 through physical protein-protein interactions. Together, the multi-species results from biochemistry, interactome mapping, and imaging demonstrated consistent support for the spatial localization and functional impact of CUL3 at the excitatory synapse.

DISCUSSION

Our proteomic and multi-omic efforts characterizing SynSig-1000 constitute an archive of synapse proteins that spans across diverse biological functions (Figure S18). To ensure that this resource captures synapse proteins relevant to the full human lifespan, we first characterized the composition of human prenatal synapses isolated from two distinct sources, the fetal human cortex and hiPSC-derived neurons. These screens are some of the first proteomic studies of the excitatory synapse in the early human brain. Through integration with multi-omics data, we found that many of the synapse candidates identified in these proteomic screens could be reinforced by multiple lines of independent evidence drawn from sequence, structure, expression, or physical interaction data.

SynSig-1000 identified 493 candidate synapse proteins (Table S7), including many RNA-binding proteins and proteasomal factors associated with developmental disorders. While RNA-binding proteins have been previously associated with synaptic regulation,^{111–113} it was unclear whether our predicted protein candidates were physical components of the synaptic molecular network. To help validate the synapse protein candidates YBX1 and DDX3X, both RNA-binding proteins, we profiled their RNA targets in hiPSC-derived neurons. To our knowledge, this is one of the first studies to systematically characterize the protein-RNA interactions in human neurons. Our eCLIP analysis uncovered distinct RNA-binding patterns for DDX3X and YBX1, revealing that while both proteins predominantly bind to the coding region, DDX3X shows increased binding to the 5' UTR (Figure S14D), and YBX1 exhibits a preference for the 3' UTR (Figure 6C). This divergence in binding preferences suggests that they may have distinct functional roles in mRNA translation initiation and localization within the PSD.

Complementing the RNA-binding work, our endogenous protein-protein interaction mapping of DDX3X and CUL3 provided valuable insights into not only their physical interactions with other synaptic proteins but also known autism risk factors. For example, through both AP-MS and quantitative proteomics, we showed that SYNGAP1, an autism risk factor,^{76,114} co-purifies with CUL3 and is differentially regulated at the synapse (Figure 7E). Notably, when we had previously performed AP-MS with SYNGAP1 as bait in the endogenous mouse cortex, this pull-down had reciprocally identified CUL3.¹⁷ These symmetric results strongly argue that SYNGAP1 may be a direct physical interactor of CUL3, raising the possibility that this interaction between two genome-wide significant autism risk factors at the synapse may be a key contributor to autism pathogenesis.

We also investigated the intersection of interacting proteins between DDX3X and CUL3 and identified 13 shared protein interactors. Remarkably, these overlapping interactors were enriched for multiple pathways integral to synaptic function, including chemical synaptic transmission, neurotransmitter transport, and anterograde trans-synaptic signaling (Figure S19). Several of these shared interactors, including SYN1,¹¹⁵ DLG2,¹¹⁶ DLG4,¹¹⁷ DYNC1H1,¹¹⁸ and STXB1,¹¹⁹ have previously been associated with neurodevelopmental disorders. This not only strengthens the case for the physical presence of DDX3X and CUL3 within the synaptic molecular network but also underscores their potential relevance in the pathogenesis of neurodevelopmental disorders. Furthermore, our analysis revealed that DDX3X shows binding affinity toward CUL3. Altogether, these observations suggest that DDX3X and CUL3 may contribute to similar pathways in synapse function and disease pathogenesis.

While the vast majority of proteins identified by the multi-omics approach (MO-1200) were supported by proteomic screens, 208 did not have experimental evidence (Figure 5A). These 208 proteins may have been missed previously due to potential biases in proteomics screens against low-abundance or hydrophobic molecules.¹²⁰ In this case, orthogonal approaches such as large-scale immunocytochemistry, flow cytometry screens, AP-MS, or neuronal eCLIP (as performed here in Figures 6B–6F; Figures S14C–S14G) are promising follow-up strategies. Naturally, each new screening method presents a distinct set of challenges, and immunocytochemistry, flow cytometry, AP-MS, and eCLIP all require the development of target-specific antibodies, which may not be readily available for all molecules.

In addition to identifying synapse candidates, the multi-omic aspect of our study also gained insights into the molecular

(as indicated by white box in the top row in the merged image), focusing on a single dendritic segment. White arrowheads indicate representative CUL3 and DLG4 co-localization.

(B) Experimental workflow showing the mapping of the protein networks of CUL3 using endogenous mouse brain tissue. Quantitative proteomics identifies the DEPs regulated by CUL3, and AP-MS maps the physical interactors of Cul3. The intersection represents proteins that interact with CUL3 and are also regulated by CUL3.

(C) Representative western blot showing protein expression of CUL3 and loading control GAPDH from both wild-type and CUL3 mutant mouse brain in the synaptosome (SPM). Boxplot shows densitometry quantification of CUL3 and GAPDH protein expression in the SPM from wild-type and CUL3 mutant brains ($n = 5$). Statistical significance is denoted by asterisk ($*p < 0.05$, Student's t test, two-tailed).

(D) Volcano plot of the quantitative proteomics results from the SPM fraction of wild-type versus CUL3 mutant mice. Upregulated proteins are denoted in green, downregulated proteins are in red, and unchanged proteins are in gray.

(E) Cytoscape¹¹⁰ network graph showing the protein interaction partners of CUL3 from affinity-purification mass spectrometry (AP-MS) in endogenous mouse cortex (Table S11), grouped by primary function. Nodes of interaction partners are colored according to whether they are categorized as known synapse proteins (pink), new synapse proteins in SynSig-1000 (orange), or other (gray). Color of the node border denotes DEPs via quantitative proteomics in the mouse cortex SPM: upregulation (green) and downregulation (blue).

features that define the essence of synaptic proteins. In this regard, our results highlight protein interaction neighborhoods as highly predictive features (Figure S7A). Since a comprehensive brain-specific protein interaction network is not yet available, this particular analysis relied on the large cell-type-general network from the Mentha database.¹²¹ Despite its generality, this network contributed highly to neuronal synapse protein prediction. These results, together with the high correlation in synapse expression patterns across non-brain adult tissues (Figure S7D), suggest that synapse proteins may have distinctive roles outside of the brain. They also suggest that more extensive interactome mapping may improve the identification of synaptic molecules. For instance, our subsequent interaction maps of synapse candidates CUL3 and DDX3X in endogenous brain tissue supported and expanded upon the initial multi-omic predictions (Figures 7E and S14B). A key advantage of our multi-omic machine-learning strategy is the ability to progressively add to both the proteomics input and the multi-omics datasets, thereby increasing both coverage and accuracy of synapse factors through continued research. Complex cellular structures with limited existing literature, such as the inhibitory synapse or activity-stimulated synapses, could particularly benefit from such a progressive framework moving forward.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Trey Ideker (tideker@ucsd.edu).

Materials availability

This study did not generate new materials.

Data and code availability

- Proteomics data have been deposited at ProteomeXchange and are publicly available as of the date of publication. Accession numbers are listed in the [key resources table](#). eCLIP protein-RNA-binding profiles have been deposited at GEO and are publicly available as of the date of publication. Accession numbers are listed in the [key resources table](#). All other data reported in this paper will be shared by the [lead contact](#) upon request.
- All original code has been deposited at Zenodo and is publicly available as of the date of publication. DOIs are listed in the [key resources table](#). Original code is also available through https://github.com/KarenYuanMei/SynSig_Updated.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

ACKNOWLEDGMENTS

We gratefully acknowledge helpful discussion and comments from Ben Andreone, Michael F. Wells, members of the Ideker laboratory, and the anonymous referees. We thank Dr. Phuong Uyen Le, Dr. Kevin Petrecca, and Dr. Guy A. Rouleau for their generous support with image experiments. This work was supported by the National Institutes of Health (NIH) under grants MH115005 to the Coba and Ideker labs, MH115747 (Psychiatric Cell Mapping Initiative) to the Ideker and Krogan labs, NS098505 to the Feng lab, GM103493 to the Smith and Qian labs, MH108528 and MH128827 to the lakoucheva lab, and HG004659 and HG009889 to the Yeo lab. G.W.Y. was additionally supported by an Allen Distinguished Investigator Award. Y.M. was supported by the Hartwell Postdoctoral Foundation Fellowship and the F32 Postdoctoral Fellowship. G.N.P. was supported by the Cure Alzheimer's Fund and the

Whitehall Foundation. Y.Z. and X.T. were supported by the NSERC Discovery grants (RGPIN-2020-04872 and DGECR-2020-00030). M.L.G. was supported by a Myotonic Dystrophy Foundation Doctoral Research Fellowship and by the National Science Foundation Graduate Research Fellowship Program under grant no. DGE-2038238. Portions of the MS proteomics experiments were performed in the Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by the Department of Energy and located at the Pacific Northwest National Laboratory, which is operated by Battelle Memorial Institute for the DOE under contract DE-AC05-76RL0 1830. A.N. is supported by the UK Dementia Research Institute, which is funded by the UK Medical Research Council, Alzheimer's Society, and Alzheimer's Research UK, and by the Edmond and Lily Safra Early Career Fellowship Program.

AUTHOR CONTRIBUTIONS

Y.M. and T.I. conceived and designed the study. Y.M. and T.I. designed the machine-learning approach. Y.M., Y.Z., G.W.Y., G.N.P., L.M.I., D.L.S., and M.P.C. designed the experimental approach. J.M., D.L.S., Y.Z., M.A.G., R.D.S., W.-J.Q., and Y.M. conducted the adult mouse proteomics experiments. Members of the M.P.C. lab, including B.W., N.H., N.A.G., and M.P.C., conducted the human fetal and iPSC proteomics experiments, MEA, and western blot in human tissues. L.E.D., G.N.P., X.T., Y.Z., M.C., S.K.G.-H., and Y.M. contributed to the co-localization studies in primary rodent neurons. H.K. generated the iPSC-derived neurons for eCLIP. M.L.G. performed and analyzed the eCLIP assays. B.W., I.F., and M.P.C. contributed to the immunostaining and imaging of the hiPSC cells. L.M.I. and D.H. provided the Cul3 haploinsufficient animals and helped design and complete the study of Cul3 function. B.W., N.H., N.A.G., and M.P.C. completed the AP-MS experiments. J.M., D.L.S., and N.J.K. completed and analyzed the proteomics studies in Cul3 haploinsufficient mice. M.C., K.L., and Y.M. contributed to the SPM westerns in mice. All co-authors contributed to data analysis and interpretation. C.A.M. and S.N.W. edited data figures and contributed to figure illustrations. S.N.L., D.F., and J.C. constructed the integrated website. Y.M., T.I., M.P.C., and G.W.Y. co-wrote the manuscript with edits from all collaborators.

DECLARATION OF INTERESTS

T.I. is a co-founder of Data4Cure, Inc., is on the scientific advisory board, and has an equity interest. T.I. is on the scientific advisory board of Ideaya Biosciences, Inc. and has an equity interest. T.I. is on the advisory board of the *Cell Systems* journal. G.W.Y. is a co-founder, member of the board of directors, on the SAB, equity holder, and paid consultant for Eclipse Biolnnovations. The terms of these arrangements have been reviewed and approved by the University of California, San Diego, in accordance with its conflict-of-interest policies. The Krogan Laboratory has received research support from the Journal of Virology, Biotechnology, F. Hoffmann-La Roche, and Rezo Therapeutics. N.J.K. has financially compensated consulting agreements with the Icahn School of Medicine at Mount Sinai, New York; Maze Therapeutics; Interline Therapeutics; Rezo Therapeutics; GEN1E Lifesciences, Inc.; and Twist Bioscience Corp. He is on the board of directors of Rezo Therapeutics and is a shareholder in Tenaya Therapeutics, Maze Therapeutics, Rezo Therapeutics, and Interline Therapeutics. D.L.S. has financially compensated consulting agreements with Maze Therapeutics and Rezo Therapeutics.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the authors used ChatGPT and Claude to improve the writing of the paper as well as to improve graph visualizations. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Rodent
 - Fetal Human Brain Tissue
- METHOD DETAILS
 - Adult Mouse Proteomics Screens
 - Fetal Human Proteomics Screens
 - Rat Hippocampal Neuron Culture
 - Confocal Imaging of Rat Hippocampal Neurons
 - Immunofluorescence and Confocal Imaging of hiPSC-Derived Neurons
 - Affinity Purification Mass Spectrometry
 - Synaptosomal Preparation and Western Analysis
 - Abundance Proteomics of Synaptosome and Whole Cell Lysate
 - Human iPSC derived neurons for eCLIP analysis
 - Assessment of bound RNA by eCLIP analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
 - Statistical Significance
 - Modeling of Gene-Gene Similarity in Multi-Omics Data
 - Molecular Features of Gene-Gene Similarity
 - Synapse Similarity Score Calculation
 - Synaptic Similarity Score to Synapse Gene Identification
 - Network Analysis of Synapse Genes
 - Random Forest Feature Importance
 - Quantification of Co-localization in Immunofluorescence Images

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cels.2025.101204>.

Received: May 3, 2024

Revised: November 26, 2024

Accepted: February 4, 2025

Published: March 6, 2025

REFERENCES

1. Bourgeron, T. (2015). From the genetic architecture to synaptic plasticity in autism spectrum disorder. *Nat. Rev. Neurosci.* *16*, 551–563. <https://doi.org/10.1038/nrn3992>.
2. Zoghbi, H.Y., and Bear, M.F. (2012). Synaptic dysfunction in neurodevelopmental disorders associated with autism and intellectual disabilities. *Cold Spring Harb. Perspect. Biol.* *4*, a009886. <https://doi.org/10.1101/cshperspect.a009886>.
3. Satterstrom, F.K., Kosmicki, J.A., Wang, J., Breen, M.S., De Rubeis, S., An, J.-Y., Peng, M., Collins, R., Grove, J., Klei, L., et al. (2020). Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell* *180*, 568–584.e23. <https://doi.org/10.1016/j.cell.2019.12.036>.
4. Sanders, S.J., He, X., Willsey, A.J., Ercan-Sencicek, A.G., Samocha, K.E., Cicek, A.E., Murtha, M.T., Bal, V.H., Bishop, S.L., Dong, S., et al. (2015). Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* *87*, 1215–1233. <https://doi.org/10.1016/j.neuron.2015.09.016>.
5. Cole, A.J., Saffen, D.W., Baraban, J.M., and Worley, P.F. (1989). Rapid increase of an immediate early gene messenger RNA in hippocampal neurons by synaptic NMDA receptor activation. *Nature* *340*, 474–476. <https://doi.org/10.1038/340474a0>.
6. Biederer, T., Sara, Y., Mozhayeva, M., Atasoy, D., Liu, X., Kavalali, E.T., and Südhof, T.C. (2002). SynCAM, a synaptic adhesion molecule that drives synapse assembly. *Science* *297*, 1525–1531. <https://doi.org/10.1126/science.1072356>.
7. Miller, S.G., and Kennedy, M.B. (1986). Regulation of brain type II Ca²⁺/calmodulin-dependent protein kinase by autophosphorylation: a Ca²⁺-triggered molecular switch. *Cell* *44*, 861–870. [https://doi.org/10.1016/0092-8674\(86\)90008-5](https://doi.org/10.1016/0092-8674(86)90008-5).
8. Lee, H.K., Barbarosie, M., Kameyama, K., Bear, M.F., and Huganir, R.L. (2000). Regulation of distinct AMPA receptor phosphorylation sites during bidirectional synaptic plasticity. *Nature* *405*, 955–959. <https://doi.org/10.1038/35016089>.
9. Kang, H., and Schuman, E.M. (1996). A requirement for local protein synthesis in neurotrophin-induced hippocampal synaptic plasticity. *Science* *273*, 1402–1406. <https://doi.org/10.1126/science.273.5280.1402>.
10. Ehlers, M.D. (1999). Synapse structure: glutamate receptors connected by the shanks. *Curr. Biol.* *9*, R848–R850. [https://doi.org/10.1016/s0960-9822\(00\)80043-3](https://doi.org/10.1016/s0960-9822(00)80043-3).
11. Ferguson, S.M., Brasnjo, G., Hayashi, M., Wölfel, M., Collesi, C., Giovedi, S., Raimondi, A., Gong, L.-W., Ariel, P., Paradise, S., et al. (2007). A selective activity-dependent requirement for dynamin 1 in synaptic vesicle endocytosis. *Science* *316*, 570–574. <https://doi.org/10.1126/science.1140621>.
12. Blasi, J., Chapman, E.R., Link, E., Binz, T., Yamasaki, S., De Camilli, P., Südhof, T.C., Niemann, H., and Jahn, R. (1993). Botulinum neurotoxin A selectively cleaves the synaptic protein SNAP-25. *Nature* *365*, 160–163. <https://doi.org/10.1038/365160a0>.
13. Bayés, A., van de Lagemaat, L.N., Collins, M.O., Croning, M.D.R., Whittle, I.R., Choudhary, J.S., and Grant, S.G.N. (2011). Characterization of the proteome, diseases and evolution of the human postsynaptic density. *Nat. Neurosci.* *14*, 19–21. <https://doi.org/10.1038/nn.2719>.
14. Chen, X., Nelson, C.D., Li, X., Winters, C.A., Azzam, R., Sousa, A.A., Leapman, R.D., Gainer, H., Sheng, M., and Reese, T.S. (2011). PSD-95 is required to sustain the molecular organization of the postsynaptic density. *J. At. Mol. Phys.* *31*, 6329–6338.
15. Kim, M.J., Futai, K., Jo, J., Hayashi, Y., Cho, K., and Sheng, M. (2007). Synaptic accumulation of PSD-95 and synaptic function regulated by phosphorylation of serine-295 of PSD-95. *Neuron* *56*, 488–502. <https://doi.org/10.1016/j.neuron.2007.09.007>.
16. Welch, J.M., Lu, J., Rodriguez, R.M., Trotta, N.C., Peca, J., Ding, J.-D., Feliciano, C., Chen, M., Adams, J.P., Luo, J., et al. (2007). Cortico-striatal synaptic defects and OCD-like behaviours in Sapap3-mutant mice. *Nature* *448*, 894–900. <https://doi.org/10.1038/nature06104>.
17. Li, J., Zhang, W., Yang, H., Howrigan, D.P., Wilkinson, B., Souaiaia, T., Evgrafov, O.V., Genovese, G., Clementel, V.A., Tudor, J.C., et al. (2017). Spatiotemporal profile of postsynaptic interactomes integrates components of complex brain disorders. *Nat. Neurosci.* *20*, 1150–1161. <https://doi.org/10.1038/nn.4594>.
18. Bayés, À., Collins, M.O., Reig-Viader, R., Gou, G., Goulding, D., Izquierdo, A., Choudhary, J.S., Emes, R.D., and Grant, S.G.N. (2017). Evolution of complexity in the zebrafish synapse proteome. *Nat. Commun.* *8*, 14613. <https://doi.org/10.1038/ncomms14613>.
19. McMahon, A.C., Barnett, M.W., O’Leary, T.S., Stoney, P.N., Collins, M.O., Papadia, S., Choudhary, J.S., Komiyama, N.H., Grant, S.G.N., Hardingham, G.E., et al. (2012). SynGAP isoforms exert opposing effects on synaptic strength. *Nat. Commun.* *3*, 900. <https://doi.org/10.1038/ncomms1900>.
20. Cizeron, M., Qiu, Z., Koniaris, B., Gokhale, R., Komiyama, N.H., Fransén, E., and Grant, S.G.N. (2020). A brainwide atlas of synapses across the mouse life span. *Science* *369*, 270–275. <https://doi.org/10.1126/science.aba3163>.
21. Zhu, F., Cizeron, M., Qiu, Z., Benavides-Piccione, R., Kopanitsa, M.V., Skene, N.G., Koniaris, B., DeFelipe, J., Fransén, E., Komiyama, N.H., et al. (2018). Architecture of the Mouse Brain Synaptome. *Neuron* *99*, 781–799.e10. <https://doi.org/10.1016/j.neuron.2018.07.007>.
22. van Oostrum, M., Blok, T.M., Giandomenico, S.L., Tom Dieck, S., Tushev, G., Fürst, N., Langer, J.D., and Schuman, E.M. (2023). The proteomic landscape of synaptic diversity across brain regions and cell types. *Cell* *186*, 5411–5427.e23. <https://doi.org/10.1016/j.cell.2023.09.028>.
23. Samavat, M., Bartol, T.M., Harris, K.M., and Sejnowski, T.J. (2024). Synaptic Information Storage Capacity Measured With Information

- Theory. *Neural Comput.* 36, 781–802. https://doi.org/10.1162/neco_a_01659.
24. Garcia, J.W., Bartol, T.M., and Sejnowski, T.J. (2022). Multiscale modeling of presynaptic dynamics from molecular to mesoscale. *PLOS Comput. Biol.* 18, e1010068. <https://doi.org/10.1371/journal.pcbi.1010068>.
 25. Wang, L., Pang, K., Zhou, L., Cebrián-Silla, A., González-Granero, S., Wang, S., Bi, Q., White, M.L., Ho, B., Li, J., et al. (2023). A cross-species proteomic map reveals neoteny of human synapse development. *Nature* 622, 112–119. <https://doi.org/10.1038/s41586-023-06542-2>.
 26. Koopmans, F., van Nierop, P., Andres-Alonso, M., Byrnes, A., Cijssouw, T., Coba, M.P., Cornelisse, L.N., Farrell, R.J., Goldschmidt, H.L., Howrigan, D.P., et al. (2019). Syngo: an evidence-based, expert-curated knowledge base for the synapse. *Neuron* 103, 217–234.e4. <https://doi.org/10.1016/j.neuron.2019.05.002>.
 27. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25, 25–29. <https://doi.org/10.1038/75556>.
 28. von Eichborn, J., Dunkel, M., Gohlke, B.O., Preissner, S.C., Hoffmann, M.F., Bauer, J.M.J., Armstrong, J.D., Schaefer, M.H., Andrade-Navarro, M.A., Le Novere, N., et al. (2013). SynSysNet: integration of experimental data on synaptic protein–protein interactions with drug-target relations. *Nucleic Acids Res.* 41, D834–D840. <https://doi.org/10.1093/nar/gks1040>.
 29. Pirooznia, M., Wang, T., Avramopoulos, D., Valle, D., Thomas, G., Hugarir, R.L., Goes, F.S., Potash, J.B., and Zandi, P.P. (2012). SynaptomeDB: an ontology-based knowledgebase for synaptic genes. *Bioinformatics* 28, 897–899. <https://doi.org/10.1093/bioinformatics/bts040>.
 30. Kustatscher, G., Collins, T., Gingras, A.-C., Guo, T., Hermjakob, H., Ideker, T., Lilley, K.S., Lundberg, E., Marcotte, E.M., Ralsler, M., et al. (2022). Understudied proteins: opportunities and challenges for functional proteomics. *Nat. Methods* 19, 774–779. <https://doi.org/10.1038/s41592-022-01454-x>.
 31. Finkbeiner, S., Frumkin, M., and Kassner, P.D. (2015). Cell-based screening: extracting meaning from complex data. *Neuron* 86, 160–174. <https://doi.org/10.1016/j.neuron.2015.02.023>.
 32. Havugimana, P.C., Hart, G.T., Nepusz, T., Yang, H., Turinsky, A.L., Li, Z., Wang, P.I., Boutz, D.R., Fong, V., Phanse, S., et al. (2012). A census of human soluble protein complexes. *Cell* 150, 1068–1081. <https://doi.org/10.1016/j.cell.2012.08.011>.
 33. Pourhaghighi, R., Ash, P.E.A., Phanse, S., Goebels, F., Hu, L.Z.M., Chen, S., Zhang, Y., Wierbowski, S.D., Boudeau, S., Moutaoufik, M.T., et al. (2020). BrainMap Elucidates the Macromolecular Connectivity Landscape of Mammalian Brain. *Cell Syst.* 333–350, e14.
 34. Williams, E.G., Wu, Y., Jha, P., Dubuis, S., Blattmann, P., Argmann, C.A., Houten, S.M., Amariuta, T., Wolski, W., Zamboni, N., et al. (2016). Systems proteomics of liver mitochondria function. *Science* 352, aad0189. <https://doi.org/10.1126/science.aad0189>.
 35. Youn, J.-Y., Dunham, W.H., Hong, S.J., Knight, J.D.R., Bashkurov, M., Chen, G.I., Bagci, H., Rathod, B., MacLeod, G., Eng, S.W.M., et al. (2018). High-Density Proximity Mapping Reveals the Subcellular Organization of mRNA-Associated Granules and Bodies. *Mol. Cell* 69, 517–532.e11. <https://doi.org/10.1016/j.molcel.2017.12.020>.
 36. Gandal, M.J., Zhang, P., Hadjichimichael, E., Walker, R.L., Chen, C., Liu, S., Won, H., van Bakel, H., Varghese, M., Wang, Y., et al. (2018). Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* 362, eaat8127. <https://doi.org/10.1126/science.aat8127>.
 37. Loo, L., Simon, J.M., Xing, L., McCoy, E.S., Niehaus, J.K., Guo, J., Anton, E.S., and Zylka, M.J. (2019). Single-cell transcriptomic analysis of mouse neocortical development. *Nat. Commun.* 10, 134. <https://doi.org/10.1038/s41467-018-08079-9>.
 38. Linsley, J.W., Tripathi, A., Epstein, I., Schmunk, G., Mount, E., Campioni, M., Oza, V., Barch, M., Javaherian, A., Nowakowski, T.J., et al. (2019). Automated four-dimensional long term imaging enables single cell tracking within organotypic brain slices to study neurodevelopment and degeneration. *Commun. Biol.* 2, 155. <https://doi.org/10.1038/s42003-019-0411-9>.
 39. Go, C.D., Knight, J.D.R., Rajasekharan, A., Rathod, B., Hesketh, G.G., Abe, K.T., Youn, J.-Y., Samavarchi-Tehrani, P., Zhang, H., Zhu, L.Y., et al. (2021). A proximity-dependent biotinylation map of a human cell. *Nature* 595, 120–124. <https://doi.org/10.1038/s41586-021-03592-2>.
 40. Aebersold, R., and Mann, M. (2003). Mass spectrometry-based proteomics. *Nature* 422, 198–207. <https://doi.org/10.1038/nature01511>.
 41. Gandal, M.J., Haney, J.R., Wamsley, B., Yap, C.X., Parhami, S., Emani, P.S., Chang, N., Chen, G.T., Hoftman, G.D., de Alba, D., et al. (2022). Broad transcriptomic dysregulation occurs across the cerebral cortex in ASD. *Nature* 611, 532–539. <https://doi.org/10.1038/s41586-022-05377-7>.
 42. Peça, J., Feliciano, C., Ting, J.T., Wang, W., Wells, M.F., Venkatraman, T.N., Lascola, C.D., Fu, Z., and Feng, G. (2011). Shank3 mutant mice display autistic-like behaviours and striatal dysfunction. *Nature* 472, 437–442. <https://doi.org/10.1038/nature09665>.
 43. Willsey, A.J., Sanders, S.J., Li, M., Dong, S., Tebbenkamp, A.T., Muhle, R.A., Reilly, S.K., Lin, L., Fertuzinhos, S., Miller, J.A., et al. (2013). Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* 155, 997–1007. <https://doi.org/10.1016/j.cell.2013.10.020>.
 44. Parikshak, N.N., Luo, R., Zhang, A., Won, H., Lowe, J.K., Chandran, V., Horvath, S., and Geschwind, D.H. (2013). Integrative Functional Genomic Analyses Implicate Specific Molecular Pathways and Circuits in Autism. *Cell* 155, 1008–1021. <https://doi.org/10.1016/j.cell.2013.10.031>.
 45. Mei, Y., Monteiro, P., Zhou, Y., Kim, J.-A., Gao, X., Fu, Z., and Feng, G. (2016). Adult restoration of Shank3 expression rescues selective autistic-like phenotypes. *Nature* 530, 481–484. <https://doi.org/10.1038/nature16971>.
 46. Welch, J.M., Wang, D., and Feng, G. (2004). Differential mRNA expression and protein localization of the SAP90/PSD-95-associated proteins (SAPAPs) in the nervous system of the mouse. *J. Comp. Neurol.* 472, 24–39. <https://doi.org/10.1002/cne.20060>.
 47. Abul-Husn, N.S., Bushlin, I., Morón, J.A., Jenkins, S.L., Dolios, G., Wang, R., Iyengar, R., Ma’ayan, A., and Devi, L.A. (2009). Systems approach to explore components and interactions in the presynapse. *Proteomics* 9, 3303–3315. <https://doi.org/10.1002/pmic.200800767>.
 48. Bayés, A., Collins, M.O., Croning, M.D.R., van de Lagemaat, L.N., Choudhary, J.S., and Grant, S.G.N. (2012). Comparative study of human and mouse postsynaptic proteomes finds high compositional conservation and abundance differences for key synaptic proteins. *PLOS One* 7, e46683. <https://doi.org/10.1371/journal.pone.0046683>.
 49. Biesemann, C., Grønborg, M., Luquet, E., Wichert, S.P., Bernard, V., Bungers, S.R., Cooper, B., Varoqueaux, F., Li, L., Byrne, J.A., et al. (2014). Proteomic screening of glutamatergic mouse brain synaptosomes isolated by fluorescence activated sorting. *EMBO J.* 33, 157–170. <https://doi.org/10.1002/emboj.201386120>.
 50. Li, J., Wilkinson, B., Clementel, V.A., Hou, J., O’Dell, T.J., and Coba, M.P. (2016). Long-term potentiation modulates synaptic phosphorylation networks and reshapes the structure of the postsynaptic interactome. *Sci. Signal.* 9, rs8. <https://doi.org/10.1126/scisignal.aaf6716>.
 51. Collins, M.O., Husi, H., Yu, L., Brandon, J.M., Anderson, C.N.G., Blackstock, W.P., Choudhary, J.S., and Grant, S.G.N. (2006). Molecular characterization and comparison of the components and multiprotein complexes in the postsynaptic proteome. *J. Neurochem.* 97, 16–23. <https://doi.org/10.1111/j.1471-4159.2005.03507.x>.
 52. Filiou, M.D., Bisle, B., Reckow, S., Teplytska, L., Maccarrone, G., and Turck, C.W. (2010). Profiling of mouse synaptosome proteome and

- phosphoproteome by IEF. *Electrophoresis* 31, 1294–1301. <https://doi.org/10.1002/elps.200900647>.
53. Moczulska, K.E., Pichler, P., Schutzbier, M., Schleiffer, A., Rumpel, S., and Mechtler, K. (2014). Deep and precise quantification of the mouse synaptosomal proteome reveals substantial remodeling during postnatal maturation. *J. Proteome Res.* 13, 4310–4324. <https://doi.org/10.1021/pr500456t>.
54. Pandya, N.J., Koopmans, F., Slotman, J.A., Paliukhovich, I., Houtsmuller, A.B., Smit, A.B., and Li, K.W. (2017). Correlation profiling of brain subcellular proteomes reveals co-assembly of synaptic proteins and subcellular distribution. *Sci. Rep.* 7, 12107. <https://doi.org/10.1038/s41598-017-11690-3>.
55. Roy, M., Sorokina, O., Skene, N., Simonnet, C., Mazzo, F., Zwart, R., Sher, E., Smith, C., Armstrong, J.D., and Grant, S.G.N. (2018). Proteomic analysis of postsynaptic proteins in regions of the human neocortex. *Nat. Neurosci.* 21, 130–138. <https://doi.org/10.1038/s41593-017-0025-9>.
56. Chang, R.Y.K., Etheridge, N., Nouwens, A.S., and Dodd, P.R. (2015). SWATH analysis of the synaptic proteome in Alzheimer's disease. *Neurochem. Int.* 87, 1–12. <https://doi.org/10.1016/j.neuint.2015.04.004>.
57. Ge, S.X., Jung, D., and Yao, R. (2020). ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics* 36, 2628–2629. <https://doi.org/10.1093/bioinformatics/btz931>.
58. Spooren, W., Lindemann, L., Ghosh, A., and Santarelli, L. (2012). Synapse dysfunction in autism: a molecular medicine approach to drug discovery in neurodevelopmental disorders. *Trends Pharmacol. Sci.* 33, 669–684. <https://doi.org/10.1016/j.tips.2012.09.004>.
59. King, I.F., Yandava, C.N., Mabb, A.M., Hsiao, J.S., Huang, H.-S., Pearson, B.L., Calabrese, J.M., Starmer, J., Parker, J.S., Magnuson, T., et al. (2013). Topoisomerases facilitate transcription of long genes linked to autism. *Nature* 501, 58–62. <https://doi.org/10.1038/nature12504>.
60. Zhang, Y., Pak, C., Han, Y., Ahlenius, H., Zhang, Z., Chanda, S., Marro, S., Patzke, C., Acuna, C., Covy, J., et al. (2013). Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron* 78, 785–798. <https://doi.org/10.1016/j.neuron.2013.05.029>.
61. Nehme, R., Zuccaro, E., Ghosh, S.D., Li, C., Sherwood, J.L., Pietilainen, O., Barrett, L.E., Limone, F., Worringer, K.A., Komminen, S., et al. (2018). Combining NG2 Programming with Developmental Patterning Generates Human Excitatory Neurons with NMDAR-Mediated Synaptic Transmission. *Cell Rep.* 23, 2509–2523. <https://doi.org/10.1016/j.celrep.2018.04.066>.
62. Ho, S.-M., Hartley, B.J., Tcw, J., Beaumont, M., Stafford, K., Slesinger, P.A., and Brennand, K.J. (2016). Rapid Ngn2-induction of excitatory neurons from hiPSC-derived neural progenitor cells. *Methods* 101, 113–124. <https://doi.org/10.1016/j.ymeth.2015.11.019>.
63. Boulting, G.L., Kiskinis, E., Croft, G.F., Amoroso, M.W., Oakley, D.H., Wainger, B.J., Williams, D.J., Kahler, D.J., Yamaki, M., Davidow, L., et al. (2011). A functionally characterized test set of human induced pluripotent stem cells. *Nat. Biotechnol.* 29, 279–286. <https://doi.org/10.1038/nbt.1783>.
64. Miskinyte, G., Grønning Hansen, M., Monni, E., Lam, M., Bengzon, J., Lindvall, O., Ahlenius, H., and Kokaia, Z. (2018). Transcription factor programming of human ES cells generates functional neurons expressing both upper and deep layer cortical markers. *PLOS One* 13, e0204688. <https://doi.org/10.1371/journal.pone.0204688>.
65. Fink, J.J., and Levine, E.S. (2018). Uncovering True Cellular Phenotypes: Using Induced Pluripotent Stem Cell-Derived Neurons to Study Early Insults in Neurodevelopmental Disorders. *Front. Neurol.* 9, 237. <https://doi.org/10.3389/fneur.2018.00237>.
66. Carter, J.L., Halmaj, J.A.N.M., and Fink, K.D. (2020). The iNs and Outs of Direct Reprogramming to Induced Neurons. *Front. Genome Ed.* 2, 7. <https://doi.org/10.3389/fgeed.2020.00007>.
67. Lähnemann, D., Köster, J., Szczurek, E., McCarthy, D.J., Hicks, S.C., Robinson, M.D., Vallejos, C.A., Campbell, K.R., Beerwinkler, N., Mahfouz, A., et al. (2020). Eleven grand challenges in single-cell data science. *Genome Biol.* 21, 31. <https://doi.org/10.1186/s13059-020-1926-6>.
68. Hulme, A.J., Maksour, S., St-Clair Glover, M., Miellat, S., and Dottori, M. (2022). Making neurons, made easy: The use of Neurogenin-2 in neuronal differentiation. *Stem Cell Rep.* 17, 14–34. <https://doi.org/10.1016/j.stemcr.2021.11.015>.
69. Ang, C.E., Olmos, V.H., Vodehnal, K., Zhou, B., Lee, Q.Y., Sinha, R., Narayanaswamy, A., Mall, M., Chesnov, K., Dominicus, C.S., et al. (2024). Generation of human excitatory forebrain neurons by cooperative binding of proneural NGN2 and homeobox factor EMX1. *Proc. Natl. Acad. Sci. USA* 121, e2308401121. <https://doi.org/10.1073/pnas.2308401121>.
70. Wilhelm, B.G., Mandad, S., Truckenbrodt, S., Kröhnert, K., Schäfer, C., Rammner, B., Koo, S.J., Claßen, G.A., Krauss, M., Haucke, V., et al. (2014). Composition of isolated synaptic boutons reveals the amounts of vesicle trafficking proteins. *Science* 344, 1023–1028. <https://doi.org/10.1126/science.1252884>.
71. Morciano, M., Burré, J., Corvey, C., Karas, M., Zimmermann, H., and Volkandt, W. (2005). Immunolocalization of two synaptic vesicle pools from synaptosomes: a proteomics analysis. *J. Neurochem.* 95, 1732–1745. <https://doi.org/10.1111/j.1471-4159.2005.03506.x>.
72. Cawley, and Talbot. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *J. Mach. Learn. Res.* 11, 2079–2107.
73. Hounkpe, B.W., Chenou, F., de Lima, F., and De Paula, E.V. (2021). HRT Atlas v1.0 database: redefining human and mouse housekeeping genes and candidate reference transcripts by mining massive RNA-seq datasets. *Nucleic Acids Res.* 49, D947–D955. <https://doi.org/10.1093/nar/gkaa609>.
74. Thul, P.J., Åkesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., Alm, T., Asplund, A., Björk, L., Breckels, L.M., et al. (2017). A subcellular map of the human proteome. *Science* 356, eaal3321. <https://doi.org/10.1126/science.aal3321>.
75. Apweiler, R., Bairoch, A., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., et al. (2004). UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res.* 32, D115–D119. <https://doi.org/10.1093/nar/gkh131>.
76. Fu, J.M., Satterstrom, F.K., Peng, M., Brand, H., Collins, R.L., Dong, S., Wamsley, B., Klei, L., Wang, L., Hao, S.P., et al. (2022). Rare coding variation provides insight into the genetic architecture and phenotypic context of autism. *Nat. Genet.* 54, 1320–1331. <https://doi.org/10.1038/s41588-022-01104-0>.
77. Abrahams, B.S., Arking, D.E., Campbell, D.B., Mefford, H.C., Morrow, E.M., Weiss, L.A., Menashe, I., Wadkins, T., Banerjee-Basu, S., and Packer, A. (2013). SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol. Autism* 4, 36. <https://doi.org/10.1186/2040-2392-4-36>.
78. Ran, X., Li, J., Shao, Q., Chen, H., Lin, Z., Sun, Z.S., and Wu, J. (2015). EpilepsyGene: a genetic resource for genes and mutations related to epilepsy. *Nucleic Acids Res.* 43, D893–D899. <https://doi.org/10.1093/nar/gku943>.
79. Trubetskoy, V., Pardiñas, A.F., Qi, T., Panagiotaropoulou, G., Awasthi, S., Bigdeli, T.B., Bryois, J., Chen, C.-Y., Dennison, C.A., Hall, L.S., et al. (2022). Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* 604, 502–508. <https://doi.org/10.1038/s41586-022-04434-5>.
80. Vissers, L.E.L.M., Gilissen, C., and Veltman, J.A. (2016). Genetic studies in intellectual disability and related disorders. *Nat. Rev. Genet.* 17, 9–18. <https://doi.org/10.1038/nrg3999>.
81. Piñero, J., Ramírez-Anguita, J.M., Saüch-Pitarch, J., Ronzano, F., Centeno, E., Sanz, F., and Furlong, L.I. (2020). The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res.* 48, D845–D855. <https://doi.org/10.1093/nar/gkz1021>.

82. Taylor, L., Kerr, I.D., and Coyle, B. (2021). Y-Box Binding Protein-1: A Neglected Target in Pediatric Brain Tumors? *Mol. Cancer Res.* 19, 375–387. <https://doi.org/10.1158/1541-7786.MCR-20-0655>.
83. Braunschweig, D., Krakowiak, P., Duncanson, P., Boyce, R., Hansen, R.L., Ashwood, P., Hertz-Picciotto, I., Pessah, I.N., and Van de Water, J. (2013). Autism-specific maternal autoantibodies recognize critical proteins in developing brain. *Transl. Psychiatry* 3, e277. <https://doi.org/10.1038/tp.2013.50>.
84. Jones, K.L., Pride, M.C., Edmiston, E., Yang, M., Silverman, J.L., Crawley, J.N., and Van de Water, J. (2020). Autism-specific maternal autoantibodies produce behavioral abnormalities in an endogenous antigen-driven mouse model of autism. *Mol. Psychiatry* 25, 2994–3009. <https://doi.org/10.1038/s41380-018-0126-1>.
85. Kuwano, M., Shibata, T., Watari, K., and Ono, M. (2019). Oncogenic Y-box binding protein-1 as an effective therapeutic target in drug-resistant cancer. *Cancer Sci.* 110, 1536–1543. <https://doi.org/10.1111/cas.14006>.
86. Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omic* 16, 284–287. <https://doi.org/10.1089/omi.2011.0118>.
87. Darnell, J.C., Van Driesche, S.J., Zhang, C., Hung, K.Y.S., Mele, A., Fraser, C.E., Stone, E.F., Chen, C., Fak, J.J., Chi, S.W., et al. (2011). FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* 146, 247–261. <https://doi.org/10.1016/j.cell.2011.06.013>.
88. Wilkinson, B., Grepo, N., Thompson, B.L., Kim, J., Wang, K., Evgrafov, O.V., Lu, W., Knowles, J.A., and Campbell, D.B. (2015). The autism-associated gene chromodomain helicase DNA-binding protein 8 (CHD8) regulates noncoding RNAs and autism-related genes. *Transl. Psychiatry* 5, e568. <https://doi.org/10.1038/tp.2015.62>.
89. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536, 285–291. <https://doi.org/10.1038/nature19057>.
90. Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y., Sundaraman, B., Blue, S.M., Nguyen, T.B., Surka, C., Elkins, K., et al. (2016). Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods* 13, 508–514. <https://doi.org/10.1038/nmeth.3810>.
91. Lovci, M.T., Ghanem, D., Marr, H., Arnold, J., Gee, S., Parra, M., Liang, T.Y., Stark, T.J., Gehman, L.T., Hoon, S., et al. (2013). Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. *Nat. Struct. Mol. Biol.* 20, 1434–1442. <https://doi.org/10.1038/nsmb.2699>.
92. Van Nostrand, E.L., Freese, P., Pratt, G.A., Wang, X., Wei, X., Xiao, R., Blue, S.M., Chen, J.-Y., Cody, N.A.L., Dominguez, D., et al. (2020). A large-scale binding and functional map of human RNA-binding proteins. *Nature* 583, 711–719. <https://doi.org/10.1038/s41586-020-2077-3>.
93. Kretov, D.A., Clément, M.-J., Lambert, G., Durand, D., Lyabin, D.N., Bollot, G., Bauvais, C., Samsonova, A., Budkina, K., Maroun, R.C., et al. (2019). YB-1, an abundant core mRNA-binding protein, has the capacity to form an RNA nucleoprotein filament: a structural analysis. *Nucleic Acids Res.* 47, 3127–3141. <https://doi.org/10.1093/nar/gky1303>.
94. Wei, W.-J., Mu, S.-R., Heiner, M., Fu, X., Cao, L.-J., Gong, X.-F., Bindereif, A., and Hui, J. (2012). YB-1 binds to CAUC motifs and stimulates exon inclusion by enhancing the recruitment of U2AF to weak polypyrimidine tracts. *Nucleic Acids Res.* 40, 8622–8636. <https://doi.org/10.1093/nar/gks579>.
95. Wu, S.-L., Fu, X., Huang, J., Jia, T.-T., Zong, F.-Y., Mu, S.-R., Zhu, H., Yan, Y., Qiu, S., Wu, Q., et al. (2015). Genome-wide analysis of YB-1-RNA interactions reveals a novel role of YB-1 in miRNA processing in glioblastoma multiforme. *Nucleic Acids Res.* 43, 8516–8528. <https://doi.org/10.1093/nar/gkv779>.
96. Shanks, N.F., Cais, O., Maruo, T., Savas, J.N., Zaika, E.I., Azumaya, C.M., Yates, J.R., 3rd, Greger, I., and Nakagawa, T. (2014). Molecular dissection of the interaction between the AMPA receptor and cornichon homolog-3. *J. Neurosci.* 34, 12104–12120. <https://doi.org/10.1523/JNEUROSCI.0595-14.2014>.
97. Valentin-Vega, Y.A., Wang, Y.-D., Parker, M., Patmore, D.M., Kanagaraj, A., Moore, J., Rusch, M., Finkelstein, D., Ellison, D.W., Gilbertson, R.J., et al. (2016). Cancer-associated DDX3X mutations drive stress granule assembly and impair global translation. *Sci. Rep.* 6, 25996. <https://doi.org/10.1038/srep25996>.
98. Tang, L., Levy, T., Guillory, S., Halpern, D., Zweifach, J., Giserman-Kiss, I., Foss-Feig, J.H., Frank, Y., Lozano, R., Belani, P., et al. (2021). Prospective and detailed behavioral phenotyping in DDX3X syndrome. *Mol. Autism* 12, 36. <https://doi.org/10.1186/s13229-021-00431-z>.
99. Snijders Blok, L., Madsen, E., Juusola, J., Gilissen, C., Baralle, D., Reijnders, M.R.F., Venselaar, H., Helmsmoortel, C., Cho, M.T., Hoischen, A., et al. (2015). Mutations in DDX3X Are a Common Cause of Unexplained Intellectual Disability with Gender-Specific Effects on Wnt Signaling. *Am. J. Hum. Genet.* 97, 343–352. <https://doi.org/10.1016/j.ajhg.2015.07.004>.
100. Trinh, J., Hüning, I., Budler, N., Hingst, V., Lohmann, K., and Gillissen-Kaesbach, G. (2017). A novel de novo mutation in CSNK2A1: reinforcing the link to neurodevelopmental abnormalities and dysmorphic features. *J. Hum. Genet.* 62, 1005–1006. <https://doi.org/10.1038/jhg.2017.73>.
101. Stamberger, H., Nikanorova, M., Willemsen, M.H., Accorsi, P., Angriman, M., Baier, H., Benkel-Herrenbrueck, I., Benoit, V., Budetta, M., Caliebe, A., et al. (2016). STXBP1 encephalopathy: A neurodevelopmental disorder including epilepsy. *Neurology* 86, 954–962. <https://doi.org/10.1212/WNL.0000000000002457>.
102. Dai, Y., Yang, Z., Guo, J., Li, H., Gong, J., Xie, Y., Xiao, B., Wang, H., and Long, L. (2022). Expansion of Clinical and Genetic Spectrum of DDX3X Neurodevelopmental Disorder in 23 Chinese Patients. *Front. Mol. Neurosci.* 15, 793001. <https://doi.org/10.3389/fnmol.2022.793001>.
103. Hoyer, M.L., Calviello, L., Poff, A.J., Ejimogu, N.-E., Newman, C.R., Montgomery, M.D., Ou, J., Floor, S.N., and Silver, D.L. (2022). Aberrant cortical development is driven by impaired cell cycle and translational control in a DDX3X syndrome model. *eLife* 11, e78203. <https://doi.org/10.7554/eLife.78203>.
104. Calviello, L., Venkataramanan, S., Rogowski, K.J., Wyler, E., Wilkins, K., Tejera, M., Thai, B., Krol, J., Filipowicz, W., Landthaler, M., et al. (2021). DDX3 depletion represses translation of mRNAs with complex 5' UTRs. *Nucleic Acids Res.* 49, 5336–5350. <https://doi.org/10.1093/nar/gkab287>.
105. Letts, V.A., Mahaffey, C.L., Beyer, B., and Frankel, W.N. (2005). A targeted mutation in *Cacng4* exacerbates spike-wave seizures in stargazer (*Cacng2*) mice. *Proc. Natl. Acad. Sci. USA* 102, 2123–2128. <https://doi.org/10.1073/pnas.0409527102>.
106. Chen, L., Chetkovich, D.M., Petralia, R.S., Sweeney, N.T., Kawasaki, Y., Wenthold, R.J., Bredt, D.S., and Nicoll, R.A. (2000). Stargazin regulates synaptic targeting of AMPA receptors by two distinct mechanisms. *Nature* 408, 936–943. <https://doi.org/10.1038/35050030>.
107. Tomita, S., Chen, L., Kawasaki, Y., Petralia, R.S., Wenthold, R.J., Nicoll, R.A., and Bredt, D.S. (2003). Functional studies and distribution define a family of transmembrane AMPA receptor regulatory proteins. *J. Cell Biol.* 161, 805–816. <https://doi.org/10.1083/jcb.200212116>.
108. Amar, M., Pramod, A.B., Yu, N.K., Herrera, V.M., Qiu, L.R., Moran-Losada, P., Zhang, P., Trujillo, C.A., Ellegood, J., Urresti, J., et al. (2021). Autism-linked Cullin3 germline haploinsufficiency impacts cytoskeletal dynamics and cortical neurogenesis through RhoA signaling. *Mol. Psychiatry* 26, 3586–3613. <https://doi.org/10.1038/s41380-021-01052-x>.
109. Dong, Z., Chen, W., Chen, C., Wang, H., Cui, W., Tan, Z., Robinson, H., Gao, N., Luo, B., Zhang, L., et al. (2020). CUL3 Deficiency Causes Social Deficits and Anxiety-like Behaviors by Impairing Excitation-Inhibition

- Balance through the Promotion of Cap-Dependent Translation. *Neuron* 105, 475–490.e6. <https://doi.org/10.1016/j.neuron.2019.10.035>.
110. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. <https://doi.org/10.1101/gr.1239303>.
111. Vuong, C.K., Wei, W., Lee, J.-A., Lin, C.-H., Damianov, A., de la Torre-Ubieta, L., Halabi, R., Otis, K.O., Martin, K.C., O'Dell, T.J., et al. (2018). Rbfox1 Regulates Synaptic Transmission through the Inhibitory Neuron-Specific vSNARE Vamp1. *Neuron* 98, 127–141.e7. <https://doi.org/10.1016/j.neuron.2018.03.008>.
112. Zheng, S., Gray, E.E., Chawla, G., Porse, B.T., O'Dell, T.J., and Black, D.L. (2012). PSD-95 is post-transcriptionally repressed during early neural development by PTBP1 and PTBP2. *Nat. Neurosci.* 15, 381–388. <https://doi.org/10.1038/nn.3026>.
113. Zhang, M., Ergin, V., Lin, L., Stork, C., Chen, L., and Zheng, S. (2019). Axonogenesis Is Coordinated by Neuron-Specific Alternative Splicing Programming and Splicing Regulator PTBP2. *Neuron* 101, 690–706.e10. <https://doi.org/10.1016/j.neuron.2019.01.022>.
114. Berryer, M.H., Hamdan, F.F., Klitten, L.L., Møller, R.S., Carmant, L., Schwartztruber, J., Patry, L., Dobrzyniecka, S., Rochefort, D., Neugnot-Cerioni, M., et al. (2013). Mutations in SYNGAP1 cause intellectual disability, autism, and a specific form of epilepsy by inducing haploinsufficiency. *Hum. Mutat.* 34, 385–394. <https://doi.org/10.1002/humu.22248>.
115. Parenti, I., Leitão, E., Kuechler, A., Villard, L., Goizet, C., Courdier, C., Bayat, A., Rossi, A., Julia, S., Bruel, A.-L., et al. (2022). The different clinical facets of SYN1-related neurodevelopmental disorders. *Front. Cell Dev. Biol.* 10, 1019715. <https://doi.org/10.3389/fcell.2022.1019715>.
116. Yang, E., Scott, D., and Nithianantharajah, J. (2021). A novel missense variant in DLG2 /PSD-93 disrupts protein folding and binding in neurodevelopmental disorders. *FASEB J.* 35. <https://doi.org/10.1096/fasebj.2021.35.S1.04406>.
117. Xing, J., Kimura, H., Wang, C., Ishizuka, K., Kushima, I., Arioka, Y., Yoshimi, A., Nakamura, Y., Shiino, T., Oya-Ito, T., et al. (2016). Resequencing and Association Analysis of Six PSD-95-Related Genes as Possible Susceptibility Genes for Schizophrenia and Autism Spectrum Disorders. *Sci. Rep.* 6, 27491. <https://doi.org/10.1038/srep27491>.
118. More, R.P., Warrier, V., Brunel, H., Buckingham, C., Smith, P., Allison, C., Holt, R., Bradshaw, C.R., and Baron-Cohen, S. (2023). Identifying rare genetic variants in 21 highly multiplex autism families: the role of diagnosis and autistic traits. *Mol. Psychiatry* 28, 2148–2157. <https://doi.org/10.1038/s41380-022-01938-4>.
119. Uddin, M., Woodbury-Smith, M., Chan, A., Brunga, L., Lamoureux, S., Pellecchia, G., Yuen, R.K.C., Faheem, M., Stavropoulos, D.J., Drake, J., et al. (2017). Germline and somatic mutations in STXBP1 with diverse neurodevelopmental phenotypes. *Neuro. Genet.* 3, e199. <https://doi.org/10.1212/NXG.000000000000199>.
120. Chandramouli, K., and Qian, P.-Y. (2009). Proteomics: challenges, techniques and possibilities to overcome biological sample complexity. *Hum. Genomics Proteomics* 2009, 239204. <https://doi.org/10.4061/2009/239204>.
121. Calderone, A., Castagnoli, L., and Cesareni, G. (2013). mentha: a resource for browsing integrated protein-interaction networks. *Nat. Methods* 10, 690–691. <https://doi.org/10.1038/nmeth.2561>.
122. Gore, A., Li, Z., Fung, H.-L., Young, J.E., Agarwal, S., Antosiewicz-Bourget, J., Canto, I., Giorgetti, A., Israel, M.A., Kiskinis, E., et al. (2011). Somatic coding mutations in human induced pluripotent stem cells. *Nature* 471, 63–67. <https://doi.org/10.1038/nature09805>.
123. Hurley, M.J., Urra, C., Garduno, B.M., Bruno, A., Kimbell, A., Wilkinson, B., Marino-Buslje, C., Ezquer, M., Ezquer, F., Aburto, P.F., et al. (2022). Genome Sequencing Variations in the Octodon degus, an Unconventional Natural Model of Aging and Alzheimer's Disease. *Front. Aging Neurosci.* 14, 894994. <https://doi.org/10.3389/fnagi.2022.894994>.
124. Wilkinson, B., Evgrafov, O.V., Zheng, D., Hartel, N., Knowles, J.A., Graham, N.A., Ichida, J.K., and Coba, M.P. (2019). Endogenous cell type-specific disrupted in schizophrenia 1 interactomes reveal protein networks associated with neurodevelopmental disorders. *Biol. Psychiatry* 85, 305–316. <https://doi.org/10.1016/j.biopsych.2018.05.009>.
125. Polpitiya, A.D., Qian, W.-J., Jaitly, N., Petyuk, V.A., Adkins, J.N., Camp, D.G., 2nd, Anderson, G.A., and Smith, R.D. (2008). DANTE: a statistical tool for quantitative analysis of -omics data. *Bioinformatics* 24, 1556–1558. <https://doi.org/10.1093/bioinformatics/btn217>.
126. Kong, A.T., Leprevost, F.V., Avtonomov, D.M., Mellacheruvu, D., and Nesvizhskii, A.I. (2017). MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat. Methods* 14, 513–520. <https://doi.org/10.1038/nmeth.4256>.
127. da Veiga Leprevost, F., Haynes, S.E., Avtonomov, D.M., Chang, H.-Y., Shanmugam, A.K., Mellacheruvu, D., Kong, A.T., and Nesvizhskii, A.I. (2020). Philosopher: a versatile toolkit for shotgun proteomics data analysis. *Nat. Methods* 17, 869–870. <https://doi.org/10.1038/s41592-020-0912-y>.
128. Demichev, V., Messner, C.B., Vernardis, S.I., Lilley, K.S., and Ralser, M. (2020). DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat. Methods* 17, 41–44. <https://doi.org/10.1038/s41592-019-0638-x>.
129. Cox, J., Hein, M.Y., Luber, C.A., Paron, I., Nagaraj, N., and Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* 13, 2513–2526. <https://doi.org/10.1074/mcp.M113.031591>.
130. Resnik, P. (1999). Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. *J. Artif. Intell. Res.* 11, 95–130. <https://doi.org/10.1613/jair.514>.
131. Yu, M.K., Ma, J., Ono, K., Zheng, F., Fong, S.H., Gary, A., Chen, J., Demchak, B., Pratt, D., and Ideker, T. (2019). DDOT: A Swiss Army Knife for Investigating Data-Driven Biological Ontologies. *Cell Syst.* 8, 267–273.e3. <https://doi.org/10.1016/j.cels.2019.02.003>.
132. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
133. Kim, S., and Pevzner, P.A. (2014). MS-GF+ makes progress towards a universal database search tool for proteomics. *Nat. Commun.* 5, 5277. <https://doi.org/10.1038/ncomms6277>.
134. Pepper, J.-P., Wang, T.V., Hennes, V., Sun, S.Y., and Ichida, J.K. (2017). Human Induced Pluripotent Stem Cell-Derived Motor Neuron Transplant for Neuromuscular Atrophy in a Mouse Model of Sciatic Nerve Injury. *JAMA Facial Plast. Surg.* 19, 197–205. <https://doi.org/10.1001/jamafacial.2016.1544>.
135. Djakovic, S.N., Schwarz, L.A., Barylko, B., DeMartino, G.N., and Patrick, G.N. (2009). Regulation of the proteasome by neuronal activity and calcium/calmodulin-dependent protein kinase II. *J. Biol. Chem.* 284, 26655–26665. <https://doi.org/10.1074/jbc.M109.021956>.
136. Doellinger, J., Schneider, A., Hoeller, M., and Lasch, P. (2020). Sample Preparation by Easy Extraction and Digestion (SPEED) - A Universal, Rapid, and Detergent-free Protocol for Proteomics Based on Acid Extraction. *Mol. Cell. Proteomics* 19, 209–222. <https://doi.org/10.1074/mcp.TIR119.001616>.
137. Markmiller, S., Sathe, S., Server, K.L., Nguyen, T.B., Fulzele, A., Cody, N., Javaherian, A., Broski, S., Finkbeiner, S., Bennett, E.J., et al. (2021). Persistent mRNA localization defects and cell death in ALS neurons caused by transient cellular stress. *Cell Rep.* 36, 109685. <https://doi.org/10.1016/j.celrep.2021.109685>.
138. Zhang, J., Lee, D., Dhiman, V., Jiang, P., Xu, J., McGillivray, P., Yang, H., Liu, J., Meyerson, W., Clarke, D., et al. (2020). An integrative ENCODE

- resource for cancer genomics. *Nat. Commun.* *11*, 3696. <https://doi.org/10.1038/s41467-020-14743-w>.
139. Blue, S.M., Yee, B.A., Pratt, G.A., Mueller, J.R., Park, S.S., Shishkin, A.A., Starner, A.C., Van Nostrand, E.L., and Yeo, G.W. (2022). Transcriptome-wide identification of RNA-binding protein binding sites using seCLIP-seq. *Nat. Protoc.* *17*, 1223–1265. <https://doi.org/10.1038/s41596-022-00680-z>.
140. Van Nostrand, E.L., Nguyen, T.B., Gelboin-Burkhart, C., Wang, R., Blue, S.M., Pratt, G.A., Louie, A.L., and Yeo, G.W. (2017). Robust, Cost-Effective Profiling of RNA Binding Protein Targets with Single-End Enhanced Crosslinking and Immunoprecipitation (seCLIP). In *mRNA Processing: Methods and Protocols*, Y. Shi, ed. (Springer), pp. 177–200.
141. Breiman, L. (2001). Random forests. *Mach. Learn.* *45*, 5–32. <https://doi.org/10.1023/A:1010933404324>.
142. Bolte, S., and Cordelières, F.P. (2006). A guided tour into subcellular colocalization analysis in light microscopy. *J. Microsc.* *224*, 213–232. <https://doi.org/10.1111/j.1365-2818.2006.01706.x>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
anti-YBX1 (ICC)	Abcam	Ab76149; RRID AB_2219276
anti-YBX1 (Western)	Novus Biological	NBP1-97572; RRID AB_11188537
anti-YBX1 (eCLIP)	Bethyl	A303-230A
anti-CUL3 (ICC and Western)	Thermo Fisher	11107-1-AP; RRID:AB_2086429
anti-CUL3 (AP-MS)	Bethyl Laboratories	A301-109A; RRID:AB_873023
anti-DDX3X (rodent ICC and Western)	Abcam	ab235940
anti-DDX3X (human iPSC ICC, rodent AP-MS, eCLIP)	Bethyl Laboratories	A300-474A; RRID:AB_451009
anti-DLG4 (ICC and Western)	Novus Biological	NB300-556; RRID:AB_2092366
anti-Synaptophysin (ICC)	Novus Biologicals	NBP1-19222; RRID:AB_1643135
anti-Synaptophysin (ICC)	Millipore	MAB5258-I
anti-MAP2 (ICC)	Novus Biological	NB300-213; RRID:AB_2138178
anti-Tubulin (Western)	Abcam	ab196583
anti-GAPDH (Western)	Cell Signaling	3683S; RRID:AB_1642205
anti-GRIN2A	Cell Signaling	4205s
anti-GRIN2B	UC DAVIS/NIH Neuromab Facility	75-097
anti-GRM5	Cell Signaling	55920s
anti-SYP	Cell Signaling	9020s
anti-GRIA1	UC DAVIS/NIH Neuromab Facility	75-327
anti-HOMER1	Cell Signaling	8231S
anti-DLGAP1/4	Neuromab	75-236
anti-ERK2	Cell Signaling	9102S
anti-GSK3B	Cell Signaling	9315
anti-CAMK2A	Invitrogen	137300
anti-SYNGAP1	Cell Signaling	5540
anti-DLG2	Neuromab	75-057
anti-MAP2 (ICC)	Cell Signaling	8707S
anti-GFAP (ICC)	UC DAVIS/NIH Neuromab Facility	75-240
anti-GAD67/1	EMD Millipore	MAB5406
anti-VGLUT2	Sigma	SAB5200276
Biological Samples		
Human fetal brain tissue	This Paper	HS-12-00474
Deposited data		
Early human proteomics data	This Paper	ProteomeXchange accession number is PXD020933 (Username: reviewer_pxd020933@ebi.ac.uk ; Password: vHyONzzE)
Adult mouse proteomics data	This Paper	ProteomeXchange accession number is PXD021009 (Username: MSV000085977; Password: PSD30#)
Cul3 differential protein expression	This Paper	ProteomeXchange accession number is PXD041108 (Username: reviewer_pxd041108@ebi.ac.uk ; Password: moOf5n3m)

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
eCLIP protein-RNA binding profiles	This Paper	GEO: GSE228444 (reviewer token: yvslekakrdkvxhux)
Experimental Models: Cell Lines		
CV-B iPSCs (46, XY)	Gore et al. ¹²²	N/A
H9 iPSCs	Hurley et al. ¹²³	N/A
ND03231	NINDS repository ¹²⁴	N/A
Experimental Models: Organisms/Strains		
C57BL/6J mice	Jackson Labs	000664
Cul3 Mutant mice	Amar et al. ¹⁰⁸	N/A
Sprague Dawley rats	Envigo	002 - US
Software and algorithms		
Mentha	Calderone et al. ¹²¹	https://mentha.uniroma2.it/
DAnTE	Polpitiya et al. ¹²⁵	https://pnml-comp-mass-spec.github.io/
Proteome Discoverer 1.4	Thermo Fisher Scientific	https://www.thermofisher.com/us/en/home/industrial/mass-spectrometry/proteomics-mass-spectrometry.html
MSFragger version 3.5	Kong et al. ¹²⁶	https://www.nesvilab.org/software
Philosopher version 4.4.0	Leprevost et al. ¹²⁷	https://philosopher.nesvilab.org/
EasyPQP version 0.1.30	Python Package Index	https://pypi.org/project/easypqp/
DIA-NN	Demichev et al. ¹²⁸	https://github.com/vdemichev/diann
maxLFQ ¹²⁹	Cox et al. ¹²⁹	https://www.rdocumentation.org/packages/iq/versions/1.9.10/topics/maxLFQ
dplyr version 1.0.7	CRAN Sources	https://github.com/tidyverse/dplyr/releases/tag/v1.0.7
ggplot2 version 3.3.6	CRAN Sources	https://github.com/tidyverse/ggplot2/releases/tag/v3.3.6
clusterProfiler version 4.2.2	CRAN Sources	https://github.com/YuLab-SMU/clusterProfiler
org.Mm.eg.db version.3.8.2	CRAN Sources	https://bioconductor.org/packages/release/data/annotation/html/org.Mmu.eg.db.html
Resnik semantic similarity score	Resnik ¹³⁰	https://arxiv.org/abs/1105.5444
Data-Driven Ontology (DDOT)	Yu et al. ¹³¹	https://github.com/idekerlab/ddot
NetworkX Python library	NetworkX	https://networkx.org/
Scikit-learn	Pedregosa et al. ¹³²	https://scikit-learn.org/stable/
Manuscript code	This paper	https://github.com/KarenYuanMei/SynSig_Updated ; Zenodo: 10.5281/zenodo.14219149

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Rodent

All animal experiments using mice and rats were conducted according to the NIH Health Guide for Care and Use of Laboratory Animals and were approved by the University of California San Diego and Massachusetts Institute of Technology Institutional Animal Care and Use Committees.

Fetal Human Brain Tissue

De-identified human fetal cortical tissues were obtained from second trimester elective termination. Collection of this tissue was performed under approval by the Institutional Review Boards of both Children’s Hospital Los Angeles and the Keck School of Medicine of the University of Southern California. University of Southern California institutional review board approvals HS-12-00474 and HS-13-00399 were obtained for this work.

METHOD DETAILS

Adult Mouse Proteomics Screens

Reagents

All chemicals and reagents were purchased from Sigma-Aldrich (St. Louis, MO) unless stated otherwise. Ammonium bicarbonate and acetonitrile were purchased from Fisher Scientific (Pittsburgh, PA), and Sequencing Grade Modified Trypsin was purchased from Promega (Madison, WI). Bicinchoninic acid (BCA) assay reagents and standards were obtained from Pierce (Pierce ThermoFisher Scientific, Waltham, MA). TMTsixplex Isobaric Mass Tagging Kit was purchased from ThermoFisher (ThermoFisher Scientific, Waltham, MA). Ni-NTA-agarose beads were obtained from QIAGEN (Valencia, CA). Purified, deionized water, >18 MW, (Nanopure Infinity ultrapure water system, Barnstead, Dubuque, IA) was used to make all aqueous buffers and solutions.

Preparation of PSD Fractions

Cortex and dorsal striatum tissues were dissected from wildtype mice aged at least 2 months old. PSD enrichment was processed as previously described.¹⁶ Three replicates of PSD fractions from wildtype mice for each dorsal striatum and cortex samples were processed for mass spectrometry analysis. 50 μ L of ice-cold lysis buffer comprised of 8 M Urea, 100 mM Ammonium Bicarbonate pH 7.8, 75 mM Potassium Chloride and cComplete Mini EDTA-free Protease Inhibitor Cocktail (1 tablet/10mL; Roche Diagnostics, Indianapolis, IN, USA) was added to each PSD pellet. Samples were homogenized with the assistance of two sessions of short (30 sec) sonication in an ultrasonic bath (Branson, Danbury, CT) filled with ice water, with manual pipetting. Assessment of lysate protein concentration was performed through a bicinchoninic acid (BCA) Protein Assay. Proteins were denatured and intra- and inter-molecular disulfide bonds were reduced with 5 mM Dithiothreitol (DTT) for 1 hour at 37°C with constant 1200 rpm shaking in a Thermomixer R (Eppendorf, Westbury, NY). Subsequently, free sulfhydryl groups were alkylated with 10 mM Iodoacetamide for one hour at 25°C in the dark with constant Thermomixer shaking at 1200 rpm. Samples were diluted 2-fold with nano-pure water containing 1 mM CaCl₂, and trypsin digestion was achieved with Sequencing Grade Modified Trypsin. Prepared trypsin was added to the protein samples at a 1:50 (wt/wt) trypsin-to-protein ratio. Following a 4-hour incubation at 37°C in the Thermomixer, samples were 4-fold diluted with nano-pure water containing 1 mM CaCl₂, and the same amount of trypsin was added to the samples. After 14 hours of 25°C incubation, the digestion reaction was stopped by acidifying samples to pH 2.5 with 10% TFA stock solution. Digested samples were centrifuged for 10 min at 9000 x g centrifugation, transferred to fresh tubes and stored at -80°C until the next processing step. Digested peptides were submitted for solid-phase extraction (SPE) via appropriately sized tC18 SepPak cartridges (Waters, Milford, MA). Desalted peptides were eluted from the SPE column with 80% acetonitrile and concentrated in a SpeedVac vacuum concentrator. 50 μ g of peptides (as measured by peptide-level BCA Assay) from each sample, was removed and transferred to a new sample vial for the drying in SpeedVac concentrator (ThermoFisher Scientific). Dry samples were temporarily stored at -80°C awaiting subsequent steps. Selected samples were labeled with 6-plex tandem mass tags (TMT, ThermoFisher Scientific). Dry sample aliquots were reconstituted in 30 μ L of 500 mM triethylammonium bicarbonate (TEAB) pH 8.5, and 0.8 mg of each TMT reagent, rehydrated in 70 μ L of anhydrous ACN, was added to each sample. After 1 hr incubation at 21°C, all individual TMT-labeling reactions were quenched with 8 μ L of 5% hydroxylamine in 200 mM TEAB pH 8.5, followed by incubation at 21°C for 15 min. Individually labeled samples were combined in a new microcentrifuge tube, and the pooled sample was acidified to pH 4.0 with 10% TFA. Labeled samples were dried in the SpeedVac then resuspended in 3% ACN, 0.1% TFA and desalted using the same SPE C18 desalting method as before.

Mass Spectrometry Peptide fractions were analyzed on a high resolution, reversed-phase capillary LC system coupled with a Thermo Fisher Scientific LTQ-Orbitrap Velos MS. The automated LC system was custom built using two Agilent 1200 nanoflow pumps and one Agilent 1200 capillary pump with a PAL autosampler. Capillary reversed-phase columns were prepared in-house by slurry packing 3- μ m Jupiter C18 (Phenomenex, Torrance, CA) into 35-cm x 360 μ m o.d. x 75 μ m i.d. fused silica (Polymicro Technologies Inc., Phoenix, AZ). Trapping columns were prepared similarly but using 3.6- μ m Aeris Widepore XB-C18 packed into a 4 cm length of 150 μ m i.d. fused silica. Mobile phases consisted of 0.1% formic acid in water and 0.1% formic acid acetonitrile operated at constant flow of 300nL/min with a gradient profile over the course of 100min gradient. MS analysis was performed on a LTQ-Orbitrap Velos mass spectrometer with heated capillary temperature and spray voltage of 350°C and 2.2 kV, respectively. Full MS spectra were recorded at a resolution of 100K (for ions at m/z 400) over the range of m/z 400–2000 with an automated gain control (AGC) value of 1×10^6 . MS/MS was performed in the data-dependent mode with an AGC target value of 3×10^4 . The ten most abundant parent ions, excluding single charge states, were selected for MS/MS using high-energy collisional dissociation (HCD) with a normalized collision energy setting of 40%. A dynamic exclusion time of 45 sec was used.

Data Analysis

TMT-labeled peptides were identified based on tandem MS/MS spectra by searching against a human protein database (UniprotKB, released 2010-05) using the MSGF+ algorithm¹³³ and abundance information across 6-plex TMT labeling was extracted from the report ion intensities within a given spectra. All peptides were identified with <0.1% False Discovery Rate by using a MSGF probability score $< 1 \times 10^{-10}$ and a decoy database searching strategy. The reporter ion intensities for each peptide were summed for all identified spectra for each channel in each biological condition. Relative abundances at the peptide level were rolled-up to the protein level using the software tool DaNTe¹²⁵ with the abundances being log₂ transformed and normalized by the central tendency approach.

Fetal Human Proteomics Screens **Human iPSC-Derived Induced Neurons**

The control 03231 iPSC line was generated from a lymphoblastoid cell line derived from a healthy 56-year-old male (NINDS repository, ND03231) as previously described.¹³⁴ For general maintenance, iPSCs were cultured on Geltrex (Thermo Fisher Scientific, Waltham, MA) coated plates in mTeSR-1 medium (Stemcell Technologies, Vancouver, Canada) at 37°C with 5% CO₂. Karyotyping was carried out by the Center for Personalized Medicine at the Children's Hospital of Los Angeles. Induced neurons (iNs) were differentiated directly from iPSCs using a modified protocol derived from previous work.⁶⁰ When confluent, iPSCs were split via Accutase and seeded into Geltrex coated 6-well plates at a density of 3×10^5 cells per well in mTeSR supplemented with 10 μM rock inhibitor. The following day, media were changed to mTeSR supplemented with 4 μg/ml polybrene, and iPSCs were infected with hNGN2 (Addgene plasmid # 79049) and rtTA (Addgene plasmid # 19780) lentiviruses at sufficient concentration to obtain >90% infection efficiency. Media was changed daily until iPSCs were ready for passaging. iPSCs were then passaged directly into neural induction media, consisting of DMEM/F12 basal media, 1X N2 supplement, 1X NEAA, 10 ng/ml BDNF (Shenandoah Biotechnology, Warwick, PA), 10 ng/ml NT-3 (Shenandoah Biotechnology), and 1 μg/ml Doxycycline (Enzo Life Sciences, Farmingdale, NY) supplemented with 10 μM Rock Inhibitor. Cells were plated in 10 cm tissue culture dishes at a density of 1.2×10^6 cells per dish. The following day, cells were changed to neural induction media supplemented with 0.7 μg/ml puromycin. 48 hr later, media was changed to B27 (Neurobasal, 1X B27 supplement, 1X Glutamax, 10 ng/ml BDNF, 10 ng/ml NT-3, and 1 μg/ml doxycycline). 48 hr later, media was changed to B27 supplemented with 2 μM Ara-C (MilliporeSigma, Burlington, MA). Half media changes were carried out every other day until neurons were harvested at 3 weeks post initiation of differentiation.

Preparation of PSD Fractions

Postsynaptic density preparations were performed as previously described.¹⁷ Either 3-week-old induced neurons from 1, 10 cm dish or second-trimester fetal cortex was homogenized in sucrose buffer, consisting of 0.32 M sucrose, 10 mM Hepes buffer (pH 7.4), 2 mM EDTA, 30 mM NaF, 20 mM β-glycerol phosphate, 5 mM sodium orthovanadate, and Roche cOmplete protease inhibitor cocktail. Samples were centrifuged at 500 x g for 6 minutes, after which supernatant was collected and spun at 10,000 x g for 10 min. The resulting pellet was solubilized in triton buffer, consisting of 50 mM Hepes (pH 7.4), 2 mM EGTA, 2 mM EDTA, 50 mM NaF, 20 mM β-glycerol phosphate, 5 mM sodium orthovanadate, Roche complete, and 1% Triton X-100. The solubilized pellet was centrifuged at 30,000 rpm for 30 min and supernatant was collected for non-PSD fractions. The pellet was solubilized in DOC buffer, consisting of 50 mM Tris (pH 9), 30 mM NaF, 5 mM sodium orthovanadate, 20 mM β-glycerol phosphate, 20 μM ZnCl₂, Roche cOmplete, and 1% sodium deoxycholate, yielding the PSD fraction.

Mass Spectrometry

25 μg of protein obtained from the postsynaptic density preparations was loaded onto 4–12% Bis-Tris [1] gels and separated at 135V for 1.5 hours. Following separation, gels were stained using InstantBlue (Expedeon, San Diego, CA) and then destained in 25% ethanol overnight. The following day, gel lanes were cut, individual gel slices were placed into perforated 96-well plates for destaining, and peptide digestion via trypsin was completed at 37°C overnight. Peptides were extracted with acetonitrile and dried. Postsynaptic density preparations were separated into four individual fractions per preparation and desalted using stage tips. LC-MS experiments were performed using a nanoscale UHPLC system (EASY-nLC1200, Thermo Fisher Scientific) connected to a Q Exactive Plus Hybrid quadrupole-Orbitrap mass spectrometer for reverse-phase separation and data acquisition.

Data Analysis

Proteome Discoverer 1.4 (Thermo Fisher Scientific) was used to process MS data which was analyzed using both Sequest HT and Mascot (Matrix Science, Boston, MA) against the Uniprot mouse and human databases combined with its decoy database. Peptide identification settings were as follows: the mass tolerance was set at 10 ppm for precursor ions and 0.8 Daltons for fragment ions; no more than two missed cleavage sites were allowed; static modification was set as cysteine carboxyamidation; and dynamic modification was set as methionine oxidation. False Discovery Rates were automatically calculated by the Percolator node of Proteome Discoverer; a peptide FDR = 0.01 was used to threshold results. Peptides with scores better than this threshold were considered as true identifications.

Rat Hippocampal Neuron Culture

Sprague Dawley rat dissociated hippocampal neurons from postnatal day 1 pups of either sex were plated at a density of 45,000 cells/cm² onto poly-d-lysine-coated coverslips and were maintained in B27 supplemented neurobasal medium (Invitrogen) until day in vitro (DIV) 16, as described previously.¹³⁵ Overall neuronal health was monitored visually weekly and throughout the experimental process.

Confocal Imaging of Rat Hippocampal Neurons

Neurons older than DIV 16 were fixed in 4% paraformaldehyde with 4% sucrose in 1X PBS for 10 or 15 minutes. Cells were permeabilized using 0.1% Triton X-100 at room temperature for 30 minutes, or 0.25% Triton X-100 at room temperature for 20 minutes. Afterwards, they were incubated with a blocking buffer, which consisted of 15% Normal Goat Serum and 5% BSA in 1X PBS for one hour at room temperature, or 5% BSA in 1X PBS for 7 hours at 4°C. After overnight incubation with primary antibodies in the same blocking buffer, or 2% BSA in 1X PBS at 4 degree Celsius, the neurons were washed three times with 1X PBS. They were then incubated with secondary antibodies in the blocking buffer, or 2% BSA in 1X PBS for 1 hour at room temperature, before being washed three times with 1X PBS and DAPI staining. Coverslips were mounted onto microscope slides using Fluoromount or Aqua-Poly/Mount. For

details on primary and secondary antibodies used for immunofluorescence, see [key resources table](#). After immunostaining, confocal images for cultured rat neurons were obtained with an LSM700 laser-scanning confocal microscope (Carl Zeiss, Jena, Germany), using a 63×oil immersion Plan-Apochromat objective (N.A. = 1.4). Images of neurons and branches were acquired with 1X optical zoom. Z-stack images or single-plane images of local dendrite and spines were acquired with 4X optical zoom. Brightness and contrast within linear ranges were adjusted for acquired images, and exported using the standard ZEISS Software ZEN 3.3 (blue edition). Images were also collected using a Leica DMI6000 inverted microscope equipped with a Yokogawa Nipkow Spinning disk confocal head, Orca ER High Resolution black and white cooled CCD camera (6.45 μm/pixel at 1 ×) (Hamamatsu), Plan Apochromat 63×/1.4 numerical aperture objective, Andor ILE laser merge module for 100mW 405nm, 150mW 488nm, 100mW 561nm, and 140mW 637nm lasers, 8-bit acquisition (0–255 pixel intensity units, respectively) with Volocity (PerkinElmer) imaging software

Immunofluorescence and Confocal Imaging of hiPSC-Derived Neurons

Eight-week-old induced neurons were fixed in cold 4% paraformaldehyde (Alpha Aesar, Haverhill, MA, catalog # A11313)/ 4% sucrose at room temperature for 15 minutes, washed three times with PBS (Lonza, Walkersville, MD, catalog # 17-516F), and then permeabilized with 0.5% PBS-Tween (Tween 20, VWR, Radnor, PA, catalog # 0777-1L) (PBST) for 15 minutes. Cells were blocked with blocking solution (10% fetal bovine serum (Genesee Scientific, San Diego, CA, catalog # 25-514) in 0.1% PBST for 1 hour at room temperature and then incubated with their respective primary antibodies diluted in blocking buffer overnight at 4 degrees Celsius. The next day, the cells were washed three times with 0.1% PBST and then incubated with their respective secondary antibodies diluted in blocking buffer for 1 hour at room temperature followed by DAPI staining. Coverslips were mounted onto microscope slides using ProLong Diamond Antifade Mountant (Thermo Fisher Scientific, Waltham, MA, catalog # P36965) and then imaged using a Zeiss LSM 800 confocal laser scanning microscope. Images of neuronal dendrites were acquired using a 63X oil immersion objective and processed using imageJ. For details on primary and secondary antibodies used for immunofluorescence, see [key resources table](#).

Affinity Purification Mass Spectrometry

Immunoprecipitation experiments were performed as previously described.¹⁷ Cortical tissues from adult wildtype mice were homogenized in DOC buffer while e14 fetal brain was homogenized in DDM buffer 50 mM Tris pH 7.4, 2 mM EDTA, 10 mM NaVO₄, 30 mM NaF, 20 mM β-glycerophosphate, 1% n-Dodecyl-β-Maltopyranoside (Anatrace, Maumee, OH, catalog # D310S), supplemented with cOmplete Protease Inhibitor Cocktail Tablets. Lysate was incubated at 4 degrees Celsius with rotation for 40 minutes and then centrifuged at 35,000 RPM for 30 minutes at 4 degrees Celsius. Following the determination of protein concentration via the BCA assay, lysate containing 2 mg of total protein was incubated with the indicated primary antibody at a concentration of 1 - 2 μg/mg of lysate at 4 degrees Celsius overnight with rotation. The following day, IPs were incubated with Dynabeads protein G (Thermo Fisher Scientific, Waltham, MA, catalog # 10004D) for 2 hours at 4 degrees Celsius with rotation. IPs were washed three times with IP wash buffer (25 mM Tris (pH 7.4), 150 mM NaCl (VWR, Radnor, PA, catalog # BDH9286-500G), 1 mM EDTA, and 1% Triton X-100). IPs were re-suspended in 2X LDS sample buffer and incubated at 95 degrees Celsius for 15 minutes to elute protein complexes. The eluant was incubated with DTT at a final concentration of 1 mM at 56 degrees Celsius for 1 hour followed by incubation with Iodoacetamide at a final concentration of 20 mM at room temperature for 45 minutes. Protein Interactions were considered positive if a minimum of two peptides were present in at least two assays and absent in anti-IgG controls.

Synaptosomal Preparation and Western Analysis

Synaptosomal preparation was completed using a protocol based on previous studies.^{16,42,46} Briefly, brain tissues from wildtype adult mice were homogenized in 4 mM HEPES buffer with 0.32M sucrose and Protease Inhibitor (Roche Diagnostics #11873580001). Samples collected from the homogenates were labeled as cell lysate. The homogenates were centrifuged at 900 rcf for 15 minutes and the supernatant was collected for another round of centrifugation at 900 rcf. After the pellet was discarded, aliquots taken from the supernatant were labeled as cytoplasmic fraction. The remaining supernatant was centrifuged at 18,000 rcf for 15 minutes, resulting in a pellet that was again centrifuged with the same settings. The final pellet was saved as the washed crude synaptosome. All centrifugations were done at 4 degrees Celsius. For Western analysis, the cell lysate, cytoplasmic, and synaptosomal fractions were recovered in M-PER Mammalian Protein Extraction Reagent (ThermoFisher Scientific #78501) with protease inhibitors. Enrichment of synaptosomal protein markers was verified by quantification of Western blots targeting known synaptic markers such as DLG4 (Figures S11A and S11B).

The protein samples were then run out on the XCell SureLock Mini-Cell Electrophoresis System (EI0001) using 10% Bis-Tris gels (NP0301) from and different Novex running buffers from Life Technologies. The gels were then transferred (Trans-Blot Semi-Dry Transfer Cell #1703940) onto nitrocellulose membranes (Bio-Rad #1620115). The membranes were blocked with 5% milk in TBST before probing with primary antibodies overnight at 4 degrees Celsius. After washing with 1X TBST, the membranes were probed with secondary antibodies at room temperature for 1 hour. After washing, the membrane was then developed on Blue Devil Autoradiography Film from Genesee Scientific (30-100) using the Pierce ECL Western Blotting Substrate from ThermoFisher (32106).

For Western analysis of YBX1, protein lysates were analyzed via the WES system from ProteinSimple. Protein lysates were prepared for the WES machine via manufacturer's specifications using their Jess/Wes Separation (12 - 230 kDa) 8 x 25 Capillary Cartridges (REF SM-W004-1) and related WES reagents. Images of the WES run were taken with ProteinSimple's Program "Compass for SW."

Abundance Proteomics of Synaptosome and Whole Cell Lysate

Sample Preparation

Both synaptic and whole cell lysate were prepared for mass spectrometry using a modified version of the SPEED method.¹³⁶ Whole cell lysate samples were brought to 80% MeOH before spinning at 16,000 g for 10 minutes. Supernatant was disposed of before the pellet was resuspended in a volume of >99% TFA, three times that of the pellet. Synaptosome sample pellets were washed twice with PBS before addition of TFA at the same volume proportion. Pellets were allowed to lyse for 5 minutes at room temperature, before acid was quenched with 2M Tris at 10x volume of acid added. Proteins were reduced and alkylated by adding 1.1x acid volume of 100 mM TCEP, 400 mM CAA, bringing the lysates to a final concentration 10mM TCEP and 40 mM CAA, before incubating at 95 C for 5 minutes. Samples were diluted 1:2 with water before protein concentration was determined using the Protein 660 assay (Pierce Thermo Fisher). 70 ug of protein was aliquoted from each sample, and trypsin and lys-C were added, each at a ratio of 1:100 protease to protein. Samples were digested with shaking overnight at 37C. Digested samples were desalted using a 96-well C18 plate (NEST Group, HNS S18V 20MG PROTO 300 C18), before dried down using a SpeedVac Concentrator. Samples were resuspended in 0.1% formic acid at a concentration of 1 mg/ml before injection onto the LC-MS/MS system.

Liquid Chromatography Mass Spectrometry

Digested peptides were separated using Thermo EASY-nLC 1200 nano liquid chromatography setup. Synaptosome samples utilized an attached 15 cm long fused silica column filled with 1.7 um bridged ethylene hybrid (BEH) C18 particles. Whole cell samples utilized an attached 25 cm long fused silica column filled with 1.9 um Reprosil-Pur C18 particles. Mobile phase A was composed of a 0.1% formic acid solution, while mobile phase B was composed of 0.1% formic acid with 80% acetonitrile. Sample injections of 1 ul, equating to approximately 1 ug of material, were loaded onto the column at a flow rate of ~1 ul/min (maximum pressure of 650 bar). After loading, the gradient began with mobile phase B at 4% for 2 minutes, before increasing to 30% over 60 mins. Mobile phase B then increased to 45% over 10 minutes before increasing to 95% B over 2 minutes and finishing with a wash at 95% B for 6 minutes. The gradient required a total time of 80 minutes. Eluting peptides were ionized by electrospray ionization and analyzed using an Orbitrap Exploris 480 Mass Spectrometer. Four data-dependent acquisition runs were collected for both the whole cell and synaptosome samples, two WT and two heterozygous mutants, to generate a spectral library. Survey scans were taken from 350-1100 m/z at a resolution of 60,000, with a normalized AGC of 300%.

Maximum injection time set to "Auto" and Advanced Peak Determination was utilized. Precursors were isolated for tandem MS using a 1.6 m/z isolation window and fragmented using HCD with a normalized collision energy of 30%. Tandem MS scans were taken at a resolution of 15,000 with a normalized AGC of 200% and maximum injection time of 22 ms. Tandem MS scan range was set to "Auto" and isolated precursors were excluded for 45 s with a 10 ppm mass tolerance. Only precursors with a charge state from 2-6 were selected for fragmentation. For data independent acquisition, precursors were isolated for tandem MS using 20 m/z windows with a 2 m/z overlap between windows covering the range of 350-1100 m/z. All other parameters for full scans and tandem MS scans were the same between DDA and DIA runs, with full scans being taken each complete loop of the DIA windows.

Data Search and Analysis

Spectral libraries were generated from DDA runs using the MSFragger (MSFragger,¹²⁶ version 3.5), Philosopher,¹²⁷ version 4.4.0), and EasyPQP (version 0.1.30) components inside of the FragPipe graphic user interface. Raw files were searched against a full mouse proteome, downloaded from Uniprot on 12/29/2021. Reversed sequences were appended to the original as decoys and common contaminants were added. For the MSFragger search enzyme specificity was set to trypsin with 2 missed cleavages allowed and precursor and fragment tolerances both set to 20 ppm. Mass calibration and parameter optimization were both used with an isotope error of 0/1/2. Allowable Peptide length was set from 7 to 50 residues and with an acceptable peptide mass range of 500 to 5,000 Da. Carbamidomethylation of cysteine was set as a fixed modification, while methionine oxidation, N-terminal acetylation, pyroglutamate on N-terminal glutamine or cysteine, and pyroglutamate at N-terminal glutamate were all set as variable modifications. The spectral library generated from Fragpipe was used to search DIA files using DIA-NN.¹²⁸ DIA-NN search was run in double-pass mode with match between runs engaged, and protein inference turned off. The quantification strategy was set to robust LC with high precision. Mass accuracy for both full scans and tandem scans was inferred based on the first DIA run of the batch. Search relied only on the spectral library and so no FASTA file was included. Precursor-level results from DIA-NN were converted into label-free protein abundances using the maxLFQ¹²⁹ algorithm within the diann R package. Significance testing for proteomics data relied on base R and the dplyr package (version 1.0.7), with plotting performed using ggplot2 (version 3.3.6).

Proteins with significant up or down effects (p -value < 0.02) were tested for enrichment of Gene Ontology (GO Biological Process, Molecular Function and cellular Component) terms. The over-representation analysis (ORA) was performed using the enricher function from R package clusterProfiler (version 4.2.2) (<https://pubmed.ncbi.nlm.nih.gov/22455463/>). The Gene Ontology terms and annotations were obtained from the R annotation package org.Mm.eg.db (version 3.8.2). In an effort to select non-redundant GO terms, we first constructed a term tree based on distances (1-Jaccard Similarity Coefficients of shared genes in GO database) between the significant terms using the R function hclust. The term tree was cut at a specific level (R function cutree, $h = 0.99$) to identify clusters of redundant gene sets. For results with multiple significant terms belonging to the same cluster, we selected the most significant term (i.e., minimum adjusted p -value).

Human iPSC derived neurons for eCLIP analysis

Due to the scalability of hiPSC-derived motor neurons for obtaining sufficient samples for downstream eCLIP analysis, we differentiated motor neurons from induced pluripotent stem cells as previously described.¹³⁷ Briefly, human iPSCs (CV-B) were

plated into 1x matrigel-coated 10cm culture dishes and grown for 2-3 days in StemFlex medium until they reached 70%–90% confluency. On day one of the differentiation protocol, medium was changed to N2B27 medium (DMEM/F12+Glutamax, 1:200 N2 supplement, 1:100 B27 supplement, 150mM ascorbic acid and 1% Penicillin/Streptomycin) supplemented with 1 μ M Dorsomorphin, 10 μ M SB431542 and 3 μ M CHIR99021. Cells were maintained with daily medium changes in the same medium for six days. On day seven, the medium was changed to N2B27 medium supplemented with 1 μ M Dorsomorphin, 10 μ M SB431542, 200nM Smoothened Agonist (SAG), and 1.5 μ M retinoic acid (RA). Medium was changed daily to compensate for increasing cell density until day 15 of differentiation when cells reached the Motor Neuron Progenitor (MNP) stage. On day 15 MNP cells were dissociated using Accutase and expanded on 1x matrigel-coated 15cm culture dishes. Medium was changed on days 15 and 17 to Maturation Medium (N2B27, 20ng/ml each of Brain Derived Neurotrophic Factor, Glial cell line Derived Neurotrophic Factor, and Ciliary Neurotrophic Factor) supplemented with 200nM SAG, 1.5 μ M RA, and 10 μ M Y-27632 (ROCK inhibitor). On day 19 medium was changed to Maturation Medium supplemented with 2 μ M DAPT. Medium was changed to Maturation Medium without additional small molecules on day 22, and cells were maintained with 75% medium changes every 2 days thereafter. Motor neurons reached synaptic maturity on day 24, and were harvested for eCLIP assays on day 28.

Assessment of bound RNA by eCLIP analysis

The eCLIP experiments were performed as previously described.^{90,138–140} Briefly, >40 million cells were collected for UV crosslink (400 mJoules/cm², 254 nm), followed by cell lysis, sonication, and RNase I treatment. Anti-DDX3 antibody (Bethyl #A300-474A) or anti-YBX1 antibody (Bethyl #A303-230A) were validated for eCLIP according to ENCODE standards.⁹⁰ Antibodies were incubated with cell lysates at 4°C overnight. 2% of samples were saved as size-matched input before immunoprecipitation (IP) steps. IP samples were washed, followed by RNA dephosphorylation (FastAP, Thermo Fisher; T4 PNK, NEB) and 3' RNA adaptor ligation (T4 RNA ligase, NEB). IP and input samples were run on a PAGE Bis-Tris protein gel and subsequently transferred to a nitrocellulose membrane. Region starting from the protein size up to 75 kDa above was excised from the membrane for proteinase K (NEB) treatment and column purification (Zymo). RNA obtained from input samples was also dephosphorylated and ligated to 3' RNA adaptors as performed previously to IP samples. Final RNA samples were reverse transcribed, ligated to a 3' DNA adaptor (T4 RNA ligase, NEB), and PCR amplified to obtain the final library for next generation sequencing. Following sequencing, raw reads were aligned to GRCh38 and analyzed following a previously published pipeline.^{138,139} Consistent with the ENCODE standard,⁹⁰ reads aligning to artifact-enriched or repetitive genomic regions were removed, and reproducible and significant peaks of aligned reads were defined as IDR cutoff of 0.01, $P \leq 0.001$, and fold enrichment ≥ 8 . Genic regions of eCLIP peaks were annotated based on overlap with GENCODE v26 transcripts following the priority order consistent with the previous study.^{138,139} For gene ontology analysis, ClusterProfiler⁸⁶ was used to analyze the top ~500 genes (ranked by P-value of the most significant IDR peak), using expressed genes from the size-matched input samples as the background set. For enrichment analysis of the RNA targets for synapse databases and autism risk factors, hypergeometric test was used with the expressed genes from the size-matched input samples as the background set.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical Significance

We denote statistical significance with * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Modeling of Gene-Gene Similarity in Multi-Omics Data

Selection of Genes for Prediction

In predicting synapse genes, we adopted a conservative approach by including only genes expressed in the human brain that are equivalently characterized. To achieve this, we used only protein-coding genes according to Ensembl, excluded genes that showed no expression across human brain regions in BrainSpan, and excluded genes that had missing data in any of the features. This resulted in a searchable pool of 11,082 human-brain expressed genes for each of which our machine learning pipeline was able to assign a synapse similarity score.

From this pool of 11,082 genes, we identified the genes that were annotated as synapse genes in the synapse database SynGO_CC (20180731 release),²⁶ and randomly selected 487 training positive genes (or 60% of SynGO_CC) from this subset. We used the remaining 40% of the SynGO_CC as held out for subsequent evaluation of model performance. For negative examples, we randomly selected the same number of genes, 487, from the pool of genes not contained in SynGO_CC. SynGO annotations were downloaded in October 2019. To determine if the model performance depended on a specific training/test split, we iterated the model with 50 additional random training/test selections. The results showed that the performance was similar across all iterations (Figure S6C). We also tested if using all of SynGO_CC simultaneously would lead to different predictions, and found that the predicted synapse similarity scores from using part of SynGO compared to all of SynGO were almost identical (Spearman rho = 0.92, Figure S6D). Because we used training examples from SynGO, we adopted the same definition of synapse proteins as SynGO to include both proteins that are enriched and also proteins that are simply present at the synapse without enrichment.

Molecular Features of Gene-Gene Similarity

The pairwise similarity for all pairs of human genes expressed in the brain was scored using a molecular features matrix (see also Figure 3A). A diverse panel of 60 molecular features was assembled from public data (https://github.com/KarenYuanMei/SynSig_Updated/tree/main/full_feature_files). Using cross-validation, 22 features were selected for the final random forest model; these features maximized performance in recovering known synapse genes on the cross-validation sets. See Figure 3A for the final feature list.

Synapse Similarity Score Calculation

The multi-omics screen was implemented as a random forest regression model¹⁴¹ trained to predict Gene Ontology (GO) Resnik semantic similarity scores between gene pairs. The Resnik semantic similarity score¹³⁰ provides a measure of gene-gene proximity in GO, and is calculated by the following formula:

$$Sim_{Resnik}(g_1, g_2) = \max(IC(t)), t \in S(g_1, g_2)$$

where g_1, g_2 are two genes in the ontology, $S(g_1, g_2)$ is the set of all common ancestor terms shared by the two genes, and $IC(t)$ is the information content as defined by:

$$IC(t) = -\log(anno(t) / anno(c))$$

where $anno(t)$ is the number of genes annotated to the term or its descendants and $anno(c)$ is the number of genes annotated to the root or its descendants, i.e. the total number of genes covered by the ontology. For each gene pair, the GO semantic similarity score was calculated using the Data-Driven Ontology package (DDOT).¹³¹

Using cross-validation, the random forest model was optimized after sweeping multiple parameters including number of trees, tree depth, and the minimum sample split. The random forest regressor model was then trained to predict semantic similarity scores with the molecular features for each gene pair (described above). Random forest was implemented using *sklearn* package in Python with `RandomForestRegressor` using the optimized parameters (`n_estimators=100`; `depth=50`, `min_sample_split=2`; See https://github.com/KarenYuanMei/SynSig_Updated).

Synaptic Similarity Score to Synapse Gene Identification

Model predictions of gene-gene similarity were used to classify genes as synapse or non-synapse, as follows (see also Figure 3B). To score the synaptic potential of a target gene, we averaged the similarity scores of this target gene versus all of the positive training examples, defining a “synaptic similarity” score. Genes were classified as synapse versus non-synapse depending on whether their synaptic similarity score was above a defined threshold. Receiver Operator Characteristic (ROC) curves were computed by varying this threshold using the consensus synapse genes as the reference. Consensus synapse genes were synapse genes confirmed by all three databases including SynGO, SynDB, and SynSysNet. SynaptomeDB annotations were downloaded in December 2018; SynSysNet annotations were downloaded in March 2019. Based on this ROC analysis (AUC= 0.9), we selected a score threshold with a sensitivity value of approximately 78% and specificity of approximately 88% (Figure 3C, see also results). Genes with a synaptic similarity score above this threshold were classified as a synapse gene, resulting in the Multi-Omic network of 1233 genes.

Network Analysis of Synapse Genes

Protein proximity was calculated using the human physical interaction network compiled by Mentha.¹²¹ Random walk with restart was run on this network, resulting in proximity scores for each pair of proteins. The random walk measures both direct and indirect interactions in the physical interaction network. Degree and eigenvector centrality were calculated using the NetworkX Python library.

Random Forest Feature Importance

Feature importance (Figure S7A) was calculated using Gini importance, as implemented by the Scikit-learn software package.¹³² Gini importance measures the amount of variance reduction that each feature contributes to the random forest model. Features that are used frequently in the random forest decision trees, and that improve prediction accuracy, receive high Gini importance.

Quantification of Co-localization in Immunofluorescence Images

Following image acquisition, the degree of colocalization between the individual channels was determined using the JACoP¹⁴² (Just Another Colocalization Plugin) plugin in ImageJ/Fiji. Prior to analysis, images were processed identically, and consistent min/max values were applied to all images within the experiment. Following processing, images were analyzed using JACoP and the resulting Mander's coefficient of colocalization was determined. Data shown reflect at least 3 independent experiments (N=3 for Ybx1; N=3 for Ddx3x; and N=10 for Cul3).