Neuron-specific protein network mapping of autism risk genes identifies shared biological mechanisms and disease relevant pathologies

*Nadeem Murtaza[1,2], Annie A. Cheng[2], Chad O. Brown[1,2], Durga Praveen Meka[3], Shuai Hong[3], Jarryll A. Uy[2,4], Joelle El-Hajjar[1], Neta Pipko[1], Brianna K. Unda[2], Birgit Schwanke[3], Sansi Xing[1], Bhooma Thiruvahindrapuram[5], Worrawat Engchuan[5], Brett Trost[5], Eric Deneault[6], Froylan Calderon de Anda[3], Bradley W. Doble[7], James Ellis[5,8], Evdokia Anagnostou[9,10], Gary D. Bader[8], Stephen W. Scherer[5,8], Yu Lu[1], Karun K. Singh[2,4]**

[1]Department of Biochemistry and Biomedical Sciences, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada
[2]Krembil Research Institute, University Health Network, Toronto, ON, Canada
[3]Center for Molecular Neurobiology Hamburg (ZMNH), Research Group Neuronal Development, University Medical Center Hamburg-Eppendorf (UKE), Hamburg, Germany
[4]Department of Laboratory Medicine and Pathobiology, Faculty of Medicine, University of Toronto, Toronto, ON, Canada
[5]The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON, Canada
[6]Centre for Biologics Evaluation, Biologic and Radiopharmaceutical Drugs Directorate, Health Products and Food Branch, Health Canada, Ottawa, Ontario, Canada
[7]Department of Paediatrics, Department of Biochemistry and Medical Genetics, University of Manitoba, Winnipeg, MB, Canada
[8]Department of Molecular Genetics, Faculty of Medicine, University of Toronto, Toronto, ON, Canada
[9]Department of Neurology, The Hospital for Sick Children, Toronto, ON, Canada.
[10]Bloorview Research Institute, Holland Bloorview Kids Rehabilitation Hospital, Toronto, ON, Canada.

***Correspondence:** karun.singh@uhnresearch.ca

Main Highlights:

1. Neuron-specific protein interaction screening of 41 ASD-risk genes to identify new disease mechanisms at the protein level
2. High connectivity between multiple unrelated ASD-risk genes at the protein interaction level
3. PPI networks show disease-relevant pathways including synaptic transmission, metabolic pathways, Wnt signaling, ion channel activity, MAPK signaling
4. Metabolic pathways, such as TCA cycle and pyruvate metabolism, are altered in neurons by multiple ASD-risk genes not previously linked to this pathway
5. Novel localization of uncharacterized ASD-risk gene PPP2R5D at the synapse, which is disrupted by *de novo* mutations identified in patients
6. Clustering of ASD-risk genes based on PPI connectivity identifies multiple gene groups that show correlation between mutation-type and clinical behavior scores, revealing the importance of understanding PPI networks in ASD

Abstract

Autism spectrum disorder (ASD) is a genetically heterogeneous disorder. Sequencing studies have identified hundreds of risk genes for autism spectrum disorder (ASD), but the signaling networks of genes at the protein level remain largely unexplored, which can provide insight into previously unknown individual and convergent disease pathways in the brain. To address this gap, we used neuron-specific proximity-labeling proteomics (BioID) to identify protein-protein interaction (PPI) networks of 41 ASD-risk genes. Network analysis revealed the combined 41 risk gene PPI network map had more shared connectivity between unrelated ASD-risk genes than represented in existing public databases. We identified common pathways between established and uncharacterized risk genes, including synaptic transmission, mitochondrial/metabolic processes, Wnt signaling pathways, ion channel activity and MAPK signaling. Investigation of the mitochondrial and metabolic network using gene knockouts revealed a functional hub in neurons for multiple risk genes not previously associated with this pathway. Further, we identified that the uncharacterized ASD-risk gene PPP2R5D localizes to the synapse, which is disrupted by patient *de novo* missense mutations. Investigation of d*e novo* missense variants of additional synaptic ASD-risk genes demonstrated that changes in PPI networks can capture synaptic transmission deficits. The neuronal 41 ASD-risk gene PPI network map also revealed enrichment for an additional 112 ASD-risk genes and human brain cell types implicated in ASD pathology. Interestingly, clustering of ASD-risk genes based on their PPI network connectivity identified multiple gene groups that correlate mutation-type to clinical behavior scores. Together, our data reveal that using PPI networks to map ASD risk genes can identify previously unknown individual and convergent neuronal signaling networks, provide a method to assess the impact of patient variants, and reveal new biological insight into disease mechanisms.

Introduction

Autism spectrum disorders (ASD) are a heterogeneous group of neurodevelopmental conditions that manifest early in life, occurring in 1 in 66 children under the age of 8[1]. The risk of developing ASD has a strong genetic basis, including common and rare genetic risk variants[2–5]. As such, numerous large scale whole exome and genome sequencing studies have identified hundreds of genes associated with ASD risk[6,7,16,8–15]. While the mechanisms by which different risk genes lead to disease are poorly understood, one hypothesis is that they converge functionally within brain signaling networks. Understanding signaling convergence can help reveal the risk genes that work through common pathways and have functional relationships. In turn, this could help classify autism risk genes based on biological pathways, prioritize the discovery of new risk genes, and identify convergent pathways that could be harnessed for targeted therapy development.

The majority of convergent ASD-associated pathways discovered to date are based on exome and genome sequencing, transcriptomics, and gene co-expression analyses, including CRISPR/Cas9 knockout screens combined with single cell RNA sequencing[11,12,15,17–22]. These studies have implicated pathways such as synaptic transmission, translation, transcription, chromatin remodeling and splicing[19,23–26]. However, the majority of autism risk genes encode proteins, and protein-protein interactions (PPIs) are an essential mechanism of signaling[8,9]. Therefore, non-protein interaction-based networks, while important, lack information regarding which ASD-risk genes interact with each other or converge into common signaling networks at the protein level. Given that a large proportion of ASD genes have non-nuclear and non-gene expression regulating functions[15,27], assessment of PPIs provides an unbiased approach to gain insights into unknown convergent ASD disease processes[28,29]. Previous ASD sequencing studies have shown that that risk genes are part of core PPI networks[8,23,26,30], and large yeast-two-hybrid (Y2H) studies have identified PPI networks shared between ASD-risk genes[31,32]. However, these data are extracted from databases that are largely derived from non-neuronal cell lines and tissues, and do not represent brain-specific networks[33]. The lack of ASD risk-gene PPI networks in disease-relevant cell types represents a missing link towards understanding the biological mechanisms of ASD.

Multiple techniques can be used to identify PPIs, including affinity purification or proximity-labeling proteomics combined with mass spectrometry (reviewed in Richards *et al.*, 2021[34]). Both are powerful approaches to identify PPI networks in cells but have caveats that can be mitigated by using appropriate controls and validations. Further, many brain-expressed genes are large in size, including ASD-risk genes[35], which limits the systems that can be used for expression in cells and allow identification of their PPI networks. We took an approach that balances gene size limitations, while at the same

time captures strong and transient interactions to build comprehensive PPI networks for ASD risk genes. We developed a lentiviral *in vitro* proximity-labeling proteomics (BioID2) system that uses mouse primary neurons. Proximity-labeling proteomics has been used successfully to capture physiologically relevant interactomes in neural cell-types both *in vitro* and *in vivo*[36–39] or to map cellular compartments[40–42]. Given the implication of cortical neurons in ASD pathology[25,43], we captured PPI networks from cortical neurons, while allowing them to grow with their glial counterparts to promote proper maturation[44–46].

In the current study, we address the lack of brain cell type-specific PPI networks for ASD-risk genes. We designed a screen to identify the interactome of 41 ASD-risk proteins in neurons by using proximity-dependent biotinylation paired with mass spectrometry. We targeted non-nuclear proteins (e.g., cytosolic proteins, receptors, kinases, scaffold proteins, and intracellular signaling proteins) because nuclear proteins have a high level of endogenous biotinylation and categorically different functional pathways. Our screen found 1770 protein-level connections (direct and indirect) between the 41 genes in neurons, which was approximately 20-times that reported in the STRING database (at lowest confidence)[47]. Convergent protein networks included synaptic transmission, mitochondrial/metabolic processes and Wnt signaling. Further investigation of genes not previously linked to mitochondrial/metabolic processes, through gene-knockout approaches, revealed that multiple genes regulate mitochondrial cellular respiration in mouse and human neurons. To further demonstrate the value of applying PPI networks to study autism risk genes, we examined rare and *de novo* missense variants in synaptic or poorly characterized ASD-risk genes. We found disruption of key PPIs that led to functional deficits in synaptic transmission. Our PPI network in mouse cortical neurons was cross-referenced with human data to demonstrate its relevance to ASD pathology. Comparing the shared 41 ASD-risk gene PPI network map to human sequencing data revealed an enrichment of an additional 112 ASD risk genes and expression in human brain cell types associated with ASD pathology[11–15]. More importantly, comparing the PPI network to human clinical data from the MSSNG database (genome sequencing and clinical data from over 5,000 individuals with ASD)[12,48], we found that individuals with variants in risk genes with a high degree of shared interactions have similar adaptive behavior scores.

Taken together, we demonstrate that neuron-specific PPI networks provide a powerful approach to reveal novel individual and convergent disease mechanisms in ASD. Given the scalability of our method and its underutilization in ASD research, we believe our PPI network resource and screening system can be applied more broadly to additional autism risk genes to identify previously unknown or overlooked disease mechanisms that are not captured with current approaches.

Main

Development of a neuronal proximity-based proteomic system to identify PPI networks

To identify the PPI networks of 41 ASD-risk genes, we used mixed mouse cortical neurons and glia co-cultures infected with lentiviral constructs expressing BioID2 fusion proteins (pLV-hSyn-tGFP-P2A-POI-13xLinker-BioID2-3xFLAG) (Extended Fig. 1a). Neuron-specific expression of BioID2-tagged proteins was driven by a human Synapsin1 promoter, and neuron/glia co-cultures were used to promote synaptic maturation. A 13x Gly-Ser linker sequence was used to join proteins-of-interest (POIs) with the BioID2-3xFLAG, which increases the range of biotinylation around the fusion protein. To help monitor transduction efficiencies, TurboGFP (tGFP) was coexpressed with BioID2 fusion proteins in a bicistronic system employing a P2A "self-cleaving" peptide. We used a Luciferase-P2A-BioID2-3xFLAG construct as a negative control (Extended Fig. 1b). Since lentiviral (LV) systems can accommodate larger gene sizes than adeno-associated virus (AAV), we were able to perform BioID2 on larger proteins such as SHANK3, GRIN2B, MET, SYNGAP1 and CNTNAP2.

Embryonic age 16-17 (E16-17) mouse pup cortices were harvested and cultured until days in vitro (DIV) 14, then infected with the BioID2 fusion constructs by using a lentivirus at an MOI of 0.7 (Fig. 1a). Biotin was added on DIV17 and cells were lysed after 18 hours on DIV18 to allow maximal biotinylation time. To reduce variability between mass spectrometry runs, TMT10plex isobaric-labeling was used to combine at least 3 biological replicates per gene. One additional technical TMT-labeling replicate of luciferase control sample chosen at random was used to account for differences in labeling. Two statistical cut-offs were used to identify positive hits for the PPI networks of each POI: Biotinylated proteins in the POI sample with 1) a significant increase in Log2 abundance compared to the luciferase control (Student's t-test, $p<0.05$)[36] and 2) that were significant outliers when accounting for the overall protein abundance compared to the protein abundance ratio between the POI and control samples (SigB $p<0.05$)[49]. Protein abundances were normalized between biological replicates based on the sample with the highest total protein abundance. To reduce variability between each viral transduction, flow cytometry was used to determine the total abundance of GFP in the infected neurons between samples. The abundance levels of samples that had less total GFP (area under the curve in GFP intensity histogram) than the luciferase control were normalized by the factor needed to minimally equal the luciferase control GFP levels. To account for false positive hits due to variability in TMT-labeling between samples, the ratio of protein abundances between the luciferase control technical TMT-labeling replicates were used as the minimal required ratio between POI and control sample abundances. Proteins that did not have abundance ratios (POI/Luciferase control) higher than this minimal ratio were considered false positives and removed from

198 further analysis. Further, to promote high efficiency infections, we created an optimized
199 lentiviral production protocol to produce high-titer virus for small and large risk genes
200 (Fig. 1a). This BioID2 screen was used in five specific experimental outputs: to identify
201 1) shared molecular pathways, 2) the impact of patient genetic variants on the PPI
202 network, 3) correlation between ASD-risk genes, 4) enrichment of ASD-relevant cell
203 types in the shared PPI network map, and 5) correlation of clinical phenotypes with the
204 ASD-risk genes (Fig. 1a).

205 To validate the BioID2 screening system, we used the well characterized
206 excitatory synapse protein DLG4 (PSD95). Neurons expressing PSD95-BioID2
207 displayed punctate localization of BioID2-3xFLAG fusion proteins and biotinylated
208 proteins around the dendrites, suggesting appropriate synaptic expression and
209 biotinylation (Extended Fig. 1c). The Luciferase-P2A-BioID2 control showed non-
210 specific localization and biotinylation throughout the neuron, which is expected
211 (Extended Fig. 1c). We identified 74 proteins that interact with PSD95, and Reactome
212 pathway enrichment revealed neurotransmitter receptors and glutamatergic synapses,
213 as expected (Extended Fig. 1d and Supplementary Table 1). Enriched pathways also
214 include less directly associated networks, such as GABAergic synapses, Rho GTPase
215 signaling, and Wnt signaling. Comparison of our PSD95 PPI network with the previously
216 published PPI networks for PSD95, from *in vivo* mouse BioID and *in vivo* mouse tandem
217 affinity purification[36,50], revealed 25 shared proteins between all three datasets
218 (Extended Fig. 1e), highlighting that our BioID2 system captures relevant networks. The
219 distinct and partially-shared proteins from the other systems suggest differences
220 between proximity-labeling and affinity purification methods and/or *in vitro* and *in vivo*
221 approaches.

222 Cortical neurons are a major cell type associated with ASD[15]; however, scalable
223 BioID labeling approaches have been done primarily in cell lines, such as HEK293
224 cells[40]. To determine the necessity and importance of using neurons for the BioID2
225 screen of ASD-risk genes, we performed BioID2 in HEK293 cells using PSD95, and a
226 subset of ASD risk genes including *ETFB*, *SPAST*, *STXBP1*, *SYNGAP1*, and *TAOK2*
227 (Extended Fig. 2 and Supplementary Table 2). The PSD95 PPI network from HEK293
228 cells showed enrichment of many pathways, including EGF- and NTR-receptor signaling
229 and cell junction organization, but there was a complete absence of synaptic pathways
230 (Extended Fig. 2a). Furthermore, BioID2 of all six ASD risk proteins in HEK293 cells
231 revealed a significant loss of protein interactions localized in neuron-specific
232 compartments, and large differences in the PPI network between HEK293 cells
233 compared to mouse neurons (Extended Fig. 2b-g and Supplementary Table 3). While
234 HEK293 cells yield interaction networks for ASD risk genes, they may not have
235 relevance to pathways associated with brain-specific pathophysiology of
236 neurodevelopmental disorders.

To further validate the specificity of our neuron-specific BioID2 screening system, we targeted proteins associated with compartments[51], including microtubules (MAP2C), the endoplasmic reticulum network (CANX), plasma membrane (PDGFR transmembrane domain), trans-Golgi apparatus (TGOLN), the presynaptic terminal (SNCA), and the nuclei (MECP2). Cellular compartment analysis of each PPI network revealed enrichment of the compartments expected for MAP2C, MECP2, CANX, PDGFR-TM domain, and TGOLN (Extended Fig. 3, Extended Fig. 4d and Supplementary Table 1 and 4). SNCA did not have a strong enrichment of presynaptic compartments; however, it did identify enriched pathways involving axons, growth cones and the synapse (Extended Fig. 3e). BioID2 of MECP2, a nuclear protein, indicated localization to the nucleus (Extended Fig. 4a) and interaction with proteins enriched in nucleus-specific pathways, such as transcription regulation and mRNA splicing (Extended Fig. 4b). The MECP2 PPI network in mouse neurons had differences in protein interactions compared to the MECP2 network in HEK293 cells, but there was no enrichment for neuron-specific compartments (Extended Fig. 4c, d). The difference in identified proteins suggests that mouse neurons have differing MECP2 interactions that are localized to the nucleus. Further, the PPI network of MECP2 did not include some of the known protein interactions in mouse neurons (e.g., ATRX, CREB1, SIN3A, NCor, and TET1), suggesting that our system may not be optimized for nuclear proteins, possibly due to the presence of highly biotinylated endogenous proteins. The enrichment of proteins specific to each compartment provides additional validation that the BioID2 screen in mouse cortical neurons can provide relevant PPI networks.

Identification of a shared PPI network map and common pathways of 41 ASD-risk genes

To develop a shared PPI network map for ASD risk genes, we selected 41 ASD-risk genes that encode proteins with a range of molecular functions, including regulation of phosphorylation and ubiquitination, enzymatic control of metabolism, protein regulation and transport, and synapse formation and function (Fig. 1b). These genes were chosen from a combined list of ASD-risk genes from the SPARK, SFARI category 1, 2, and syndromic gene lists and previous sequencing studies[11–16]. For each gene, the human cDNA was cloned into a BioID2 lentiviral backbone and protein expression was confirmed with western blotting (Extended Fig. 5). All BioID2 fusion constructs were found to be the expected size through western blotting; however, some constructs showed a second larger size protein due to lower P2A efficiency or increased cleaved BioID2-FLAG (lowest band) due to increased degradation (Extended Fig. 5). As mentioned previously, the list includes large genes (>4kb) such as SHANK3 and SYNGAP1, allowing us to examine the PPI network of proteins from a range of sizes. All genes that were chosen for the screen have cytoplasmic functions. Nuclear genes were not selected because it has previously been shown there is a separation in function

between nuclear gene regulating proteins and cytoplasmic neural communication proteins[15]. We identified the individual PPI networks and enriched Reactome pathways, biological processes and cellular compartments for each of the 41 ASD-risk genes. These data can be found in Supplementary Table 1 and Supplementary Table 5, and are meant to be a resource for the research community.

The 41 ASD-risk gene PPI network consisted of 1109 proteins (41 ASD bait proteins and 1068 prey proteins) and 2349 connections. Half of the identified prey proteins were shared between 2-15 ASD bait proteins (489 prey proteins and 1770 connections), and of these, 15 prey proteins were shared between at least 10 different ASD bait proteins. Every ASD bait protein shared at least 4 interactions (direct or shared prey protein) with one other ASD bait protein, with up to 38 shared interactions between DLG4 and SYNGAP1 (Fig. 2a). The PPI network of 31 out of the 41 ASD bait genes showed direct interaction with at least 1 other ASD bait protein. Reciprocal identification was observed between DLG4 and CDKL5, SYNGAP1, GRIA1, or GRIA2 and between GRIA1 and GRIA2. The most identified ASD bait proteins were GRIN2B, PPP1R9B, GRIA2, and KCNQ2. Conversely, BioID2 of TAOK2, CDKL5, DLG4, LRRC4C, and SYNGAP1 identified the most ASD bait proteins, suggesting high connectivity between a subset of ASD bait proteins. To determine the utility of creating an ASD PPI network in neurons, we compared our results with physical interactions between the 41 ASD bait genes extracted from the STRING database (greater than or equal to medium confidence, 0.4). Our BioID2 ASD-risk gene PPI network had 245 connections (where each connection represents at least 5 shared protein interactions) between 36 of the 41 ASD bait proteins. The STRING database had 33 direct interactions between 23 of the bait proteins, revealing a near 50-fold increase in the number of connections within our ASD-risk gene PPI network (Fig. 2a, b). Current databases, such as STRING, are primarily derived from non-neuronal sources using gene co-expression or direct interaction data[33]. However, our PPIs were identified in neuronal cells and include both direct interacting proteins and shared interacting proteins that highlight important connections missed by traditional methods.

The most significant pathways in the shared 41 ASD-risk gene PPI network involve synaptic transmission, demonstrating that our system can identify the most frequently identified pathways in ASD pathophysiology (Fig. 2c and Supplementary Table 6). Other enriched pathways included TCA cycle and mitochondrial activity, Wnt signaling, potassium channel activity, and MAPK signaling (Fig. 2c and Extended Fig. 6a). These enriched pathways suggest that synaptic function plays a core role among non-nuclear ASD risk proteins, but it is not the only pathway involved between the 41 genes. The majority of the shared ASD-risk PPI network localized to specific cellular compartments including axons, dendrites and synapses (Extended Fig. 6b and Supplementary Table 6), while the majority of biological processes involve synaptic

signaling and organization, and protein transport (Extended Fig. 6c and Supplementary Table 6). Shared pathways in the ASD-risk gene PPI network reflect the major role of synaptic dysfunction in ASD, but also highlight that other, less well-studied pathways are important contributors to convergent ASD pathology.

The shared PPI network map identifies the tricarboxylic acid (TCA) cycle and pyruvate metabolism as a common signaling pathway in ASD

One rationale for constructing a PPI network map with ASD-risk genes was to identify novel or poorly characterized convergent signaling mechanisms. In this regard, one of the top pathways we identified was the TCA cycle and pyruvate metabolism (mitochondrial/metabolic processes), implicating dysregulation in mitochondrial function and cellular metabolism. This pathway has been associated with a few ASD associated genes[52–54], but the mechanisms are not well understood, and it is unknown whether other ASD risk genes regulate mitochondrial/metabolic processes. Interestingly, previous ASD clinical studies have identified abnormal mitochondrial function in patient lymphoblastoid cells[55–58], but whether this occurs in mammalian brain cells is unknown. TCA cycle and pyruvate metabolism proteins were highly enriched in the shared ASD-risk gene PPI network map (adj. p-value = 3.14×10^{-12}), even without the PPI network for the mitochondrial protein ETFB (adj. p-value = 1.35×10^{-7}) (Supplementary Table 6). 28 out of 41 ASD-risk genes were found to be interacting with at least one TCA cycle and pyruvate metabolism associated protein (Fig. 3a). Citrate synthase (CS), which is involved in turning acetyl-CoA into citrate early in the TCA cycle, was found to interact with eight ASD bait proteins (ERBIN, MET, NRXN1, SHANK3, SPAST, STXBP1, SYNGAP1, TAOK2). The TCA cycle and pyruvate metabolism are essential for proper cellular respiration. Therefore, we investigated this finding by focusing on a gene in our screen that was not previously associated with mitochondrial and metabolic processes in the brain, *TAOK2*, a gene in the 16p11.2 deletion/duplication region associated with ASD[59–62]. We measured cellular respiration using live-cell metabolic assays in a *Taok2* knockout (KO) mouse model, which we previously demonstrated has deficits in synapse formation and function[59]. *Taok2* heterozygous knockout (Het) cultured mouse cortical neurons showed a significant increase in maximal respiration, proton leak, non-mitochondrial respiration, and spare respiratory capacity, and a decrease in ATP coupling efficiency (Fig. 3b, c and Extended Fig. 7a-d) compared to wildtype (WT) neurons. These changes suggested the presence of less functional mitochondria, which was corroborated by proteomic analysis of post-synaptic density fractions isolated from *Taok2* WT and KO mouse cortices (Extended Fig. 7e). *Taok2* KO mice PSD fractions had significant downregulation of proteins involved in synaptic function and activity, and also in respiratory ETC complex proteins (Extended Fig. 7f and Supplementary Table 7). Analysis at the transcriptome level also revealed reduced mRNA levels of mitochondrial membrane proteins in *Taok2* KO mouse cortices (Extended Fig. 7g, h and

Supplementary Table 7), coinciding with the reduced protein levels of mitochondrial proteins (Extended Fig. 7e, f). Further investigation revealed that *Taok2* Het and KO neurons have a reduced proportion of active TMRM stained mitochondria (Fig. 3d, e and Extended Fig. 7i), and an overall increase in the amount or size of mitochondria labeled by TOMM20, an outer membrane protein (Fig. 3f, g). These data implicate dysregulated mitochondria in the absence of *Taok2*; therefore, we examined the morphology of mitochondria *in vivo* from electron microscopy (EM) images taken from WT and *Taok2* KO mouse cortical excitatory neurons[59]. We found that *Taok2* KO mouse neurons had altered mitochondrial morphology with a reduction in category 1 and 3 mitochondria, which show more typical mitochondria morphology, and an increase in category 2 mitochondria at their synapses (Fig. 3h, i)[63]. Category 2 mitochondria have enlarged non-contiguous mitochondrial cristae[63], which can cause reduced oxidative phosphorylation and prevent proper translation and insertion of inner membrane proteins[64,65]. We extended our mouse studies to examine whether TAOK2 regulates mitochondrial/metabolic processes in human induced pluripotent stem cell (iPSC)-derived NGN2-neurons. We used CRISPR/Cas9 to generate isogenic *TAOK2* homozygous KO and heterozygous knock-in *TAOK2 A135P* iPSC lines. A135P is a *de novo* missense variant which we previously demonstrated renders TAOK2 as kinase dead[59]. We generated human neurons through direct differentiation of iPSCs via NGN2 overexpression and found altered cellular respiration in *TAOK2* KO neurons (Extended Fig. 7j) similar to mouse neurons, and significant increases in the spare respiratory capacity of *TAOK2* KO and *A135P* neurons (Extended Fig. 7k). Human neurons transfected with Mito7-DsRed also displayed an increase in mitochondrial puncta size in *TAOK2* KO and *A135P* neurons, suggesting an increase in the number or size of the mitochondria, similar to that observed in the mouse neurons (Extended Fig. 7l, m). To determine if these changes were due to long-term developmental deficits caused by loss of TAOK2 function, we used acute shRNA knock-down through *in utero* electroporation and found that *Taok2* knock-down in cultured mouse neurons caused decreased mitochondrial membrane potential (Extended Fig. 8a-c) similar to that detected in the knockout mice (Fig. 3d, e). Taken together, using *TAOK2* as a validation gene from the identified mitochondrial/metabolic PPI network, we determined that mouse and human models with disruption of *TAOK2* have altered cellular respiration, likely caused by altered activity, size and number of mitochondria.

To determine if other ASD risk genes converging on the mitochondrial/metabolic network regulate cellular respiration, we used the CRISPR/Cas9 system to knock out *Syngap1*, *Taok2*, and *Spast*. We also targeted *Etfb* and *Rheb*, which are both ASD risk genes known to localize to the mitochondrion or regulate neuronal mitochondrial function[66]. Combined gRNAs against BFP and Luciferase were used as a negative control[67,68], and we used 1-3 gRNAs targeting different genomic regions of the ASD-risk genes (Extended Fig. 8d). Mouse cortical neurons were infected with Cas9-EGFP and

gRNA-mCherry lentiviral constructs. Western blots of neurons infected with Taok2 gRNAs and Cas9 showed decreased expression by approximately 50%, suggesting a partial knockout (Extended Fig. 8e). CRISPR/Cas9 knockout of Etfb, a subunit of riboflavin required for proper electron transfer in the ETC, showed increased basal and maximal respiration, proton leakage, and no change in ATP synthase-dependent cellular respiration (Fig. 3j, k and Extended Fig. 8f, g). These changes may correspond to increased cellular respiration to counteract faulty ETC electron transfer. Mouse neurons with CRISPR knockout of *Taok2*, *Syngap1*, and *Rheb* also showed significant or trending changes in many aspects of cellular respiration (Fig. 3j, k and Extended Fig. 8f-i). CRISPR KO of *Spast* did not cause significant changes in cellular respiration, possibly due to subtle effects or a role in different aspects of mitochondrial function. The increase in basal respiration in *Taok2*, *Syngap1*, and *Etfb* KO neurons (Extended Fig. 8f) may be indicative of an acute effect, where altered cellular respiration has not yet reached homeostasis within the neuron[69]. These findings suggest that a subset of ASD risk genes regulate cellular respiration in neurons, and highlight the relevance of TCA cycle and pyruvate metabolism pathways in the developing brain as a risk factor for ASD when dysregulated.

PPI networks identify differences in signaling between missense variants in ASD risk genes

Next, we hypothesized that PPI networks could be used to study missense variants, which are a large and important class of genetic risk factors for ASD that have less obvious functional impacts compared to loss-of-function (LoF) variants. Sequencing of ASD individuals have identified many missense variants of unknown significance (VUS) and therefore, the biological impact of variants in the majority of risk genes remain unknown. Understanding the impact of a variant is important because it provides the affected individual and family with a possible causal explanation and, in some cases, it could help to assess clinical trajectory or treatments. Missense variants have been suggested to impact protein stability and protein-protein interaction networks[30]; however, these data were imputed from databases using primarily non-neuronal datasets, and were not tested in neurons. We used BioID2 to identify differences in severity and pathogenic mechanisms of *de novo* missense variants identified in individuals diagnosed with ASD. Due to the strong link between synaptic functional deficits and ASD pathophysiology, we chose two known synaptic genes (TAOK2β and GRIA1) and a less well-characterized risk gene with no specific cellular localization (PPP2R5D) (Fig. 4a-c, Supplementary Table 8, and Supplementary Table 9).

We used BioID2 to determine the change in the TAOK2β PPI network due to the A135P *de novo* missense variant, which was identified in an individual with ASD. The TAOK2β A135P PPI network had a reduced number of proteins associated with the

synaptic compartment, and simultaneously had increased dendritic and ribosomal proteins (Fig. 4d). The latter changes may be due to the loss of PPI network proteins in dendritic spines where TAOK2β localizes, and an increase in dendritic and ribosome translation complex protein interactions specific to the TAOK2β A135P (Fig. 4e), combined with the decreased expression of the A135P mutant (Extended Fig. 9a). To corroborate the possible synaptic deficits caused by the A135P variant, we performed patch-clamp electrophysiology on the isogenic iPSC-derived NGN2-neurons (Extended Fig. 9b)[70–72]. *TAOK2* KO and *TAOK2* A135P neurons had decreases in frequency and amplitude of spontaneous excitatory post-synaptic currents (sEPSCs) (Fig. 4f, g), corroborating the shift in interaction with synaptic proteins. The lack of change in the intrinsic firing properties or Synapsin1-positive punctae density in *TAOK2* A135P neurons, as opposed to the *TAOK2* KO neurons (Extended Fig. 9c-g), suggest that the shift in interaction and localization for the heterozygous A135P line has dissimilar phenotypes compared to the complete loss of TAOK2. In fact, *TAOK2* A135P neurons displayed increased size of Synapsin1 punctae, suggesting possible changes in the synaptic structure (Extended Fig. 9c, d). Taken together, the TAOK2β A135P variant showed significant decreases in synaptic pathway protein interactions, demonstrating that changes in PPI networks can be predictive of functional deficits.

We also asked whether PPI networks can distinguish the impact of missense variants based on their location within functional domains of a gene. We investigated *GRIA1* and two *de novo* missense variants, R208H and A636T[3,73,74], located in the extra-cellular ligand binding domain and the transmembrane domain, respectively (Fig. 4b). The *GRIA1* variants showed strong differential effects in their enriched cellular compartments (Fig. 4h) and the number of shared interacting proteins with the wildtype (Fig. 4i). GRIA1 R208H had a significant loss of proteins localizing to the AMPA receptor and post-synaptic density, which suggests functional changes in synapse function. GRIA1 A636T had a less severe impact, with small increases in the number of compartment-specific protein interactions and gains in membrane junction and ER proteins (Fig. 4h and Supplementary Table 8 and 9), suggesting possible trafficking issues. There were no changes in expression between the two variants (Extended Fig. 9h). To functionally corroborate the changes in PPI networks, we infected mouse cortical neurons with the GRIA1 WT and both variants to obtain whole-cell voltage clamp recordings. This revealed a trend towards decreased sEPSC frequency in neurons expressing the R208H variant, but not the A636T variant (Fig. 4j, k). Although the A636T mutant had no change in sEPSCs, we did observe large sEPSC bursts (Fig. 4j), which may be indicative of altered trafficking of AMPA receptors through the ER network and longer turnover rates[75,76]. Together, the stronger loss of interactions for R208H compared to A636T coincide with the electrophysiology results, demonstrating that BioID2 PPI networks can reveal functional differences in missense variants for receptor proteins.

Finally, we used BioID2 to test missense variants in the risk gene PPP2R5D, a regulatory subunit of phosphatase-2A[77]. This protein is not known to have multiple functional domains or a specific localization; therefore, BioID2 could help to first understand where it functions in neurons and then the impact of ASD missense variants. We selected three *de novo* PPP2R5D variants, P53S, E198K, and E420K, which are spread throughout the protein[77,78]. The PPI networks for the variants had both common and dissimilar effects (Fig. 4c, l, and m), with all three variants reducing interactions with synaptic and dendritic proteins enriched in the wildtype PPI network (Fig. 4l). This suggests that PPP2R5D has a potential role in dendrites and synapses based on PPI network. Additionally, all of the variants caused a loss and gain of diverse interactions (Fig. 4m), with no change in expression levels (Extended Fig. 9i). Interestingly, both the E198K and the E420K variants gained trans-Golgi compartment proteins (Fig. 4l and Supplementary Table 8 and 9), suggesting altered localization. Previous studies have described an overactive AKT pathway caused by the PPP2R5D E420K variant[79]. However, measurement of phospho-AKT levels in HEK293 cells expressing the variants revealed no difference (Extended Fig. 9j), suggesting that specific molecular assays may miss functional deficits. To probe E420K further, we performed imaging on neurons and found accumulation of E420K in the cell body, indicating possible trafficking deficits that cause increased interactions with trans-Golgi network proteins (Fig. 3n). Together, the BioID2 approach revealed dendritic and synaptic localization of PPP2R5D, which is lost in multiple missense variants that have their own subtle differences. The differences in the PPI network of wildtype proteins and their ASD-associated variants highlight the utility of the system to screen multiple disease variants within a gene.

The 41 ASD-risk gene PPI network map enriches for additional ASD risk genes, human disease cell types, and correlates with human behavioral phenotypes from clinical datasets

The complete PPI network map from the 41 ASD-risk genes demonstrates the importance of a neuron-specific network. The network ultimately contained significantly more connections than reported in databases such as STRING (Fig. 2a, b) and elucidated multiple convergent pathways (e.g., TCA and pyruvate signaling, Fig. 2c) linked to ASD that are poorly studied. To further demonstrate the utility of the 41 ASD-risk gene PPI network map resource, we used enrichment analysis to determine relevance to human ASD. We found a significant enrichment of 112 additional ASD-risk genes (Fisher's Exact test $p = 2.69 \times 10^{-30}$, OR = 3.45), highlighting the strong functional connectivity between ASD-risk genes at the protein level (Fig. 5a). Along with enrichment of ASD-risk genes from the original 41 ASD-risk protein baits, we found that gene lists reported from individual sequencing studies were enriched, especially when examining cytoplasmic (non-nuclear) proteins (Extended Fig. 10a). This suggests strong

connectivity of ASD protein signaling outside the nucleus. Gene lists with only nuclear proteins were not enriched (Extended Fig. 10b), providing evidence that there is less interaction between proteins localized to the nucleus and those in the cytoplasm. Of the 153 ASD-risk proteins in the network, 69 are interacting with 2 or more ASD bait proteins. Slitrk5, Gria2, Dlg4, Grin2b, and Shank2 were identified by more than eight of the ASD bait proteins, suggesting a potential central role for these genes in ASD pathology. Enrichment of multiple cytoplasmic ASD-risk proteins in the PPI network indicates functional connectivity between intracellular signaling proteins.

While the PPI network from 41 ASD-risk genes was generated using human genes, it was obtained in a background of mouse cortical neuron and glia co-cultures; therefore, it is unknown whether this network map is applicable to human brain cell types or differentially expressed genes (DEGs) implicated in ASD pathology. To address this, we examined the enrichment of specific cell types based on their single cell RNA-sequencing profiles[25,43]. We found that the 41 ASD-risk gene PPI network map strongly enriches for excitatory and inhibitory neuron cell types, along with neural progenitor cells, astrocytes and microglia (Fig. 5b), which have been associated with ASD pathophysiology[17,18,25,80,81]. When examining the ASD-specific DEGs of different cell types from human post-mortem brain samples[25], the shared PPI network was enriched for DEGs in layer 2/3 and 4 neurons, parvalbumin and VIP interneurons, and protoplasmic astrocytes (Fig. 5c). The enrichment of ASD DEGs of specific cell types highlights the human disease relevance of the 41 ASD-risk gene PPI network map.

Finally, we hypothesized a potential relationship between highly connected genes within the 41 ASD-risk gene PPI network map and human ASD behavioral phenotypes. This would link gene clusters to human phenotypes, and provide additional insight into the biological basis of ASD. We took the individual PPI networks of the 41 ASD-risk genes and identified 3 groups (labeled Group 1, 2 and 3) of highly connected ASD-risk genes, using the correlation between their individual PPI networks (Fig. 5d). Groups 1 and 2 showed high connectivity between the ASD-risk genes within each group, whereas connectivity was lower in Group 3. To determine if grouping the 41 ASD-risk genes is correlated with clinical ASD behavioral scores based on shared PPI networks, we obtained clinical data of individuals with rare variants in the 41 ASD-risk genes from the MSSNG database. The database contained the sequenced genomes of a total of 4,258 families and 5,102 ASD-affected individuals at the time of data extraction[12]. We obtained the adaptive behavior and socialization scores from up to 879 individuals who possess at least one rare missense/splicing/LoF variant in the 41 ASD-risk genes (data-explorer.mss.ng). Remarkably, we found that individuals with missense variants in Group 1 genes had lower adaptive behavior standard scores compared to Groups 2 or 3, suggesting that missense variants strongly impact the function of Group 1 genes in regards to adaptive behavior (Fig. 5e and Extended Fig. 10c). However,

individuals with variants impacting mRNA splicing in Groups 1 had significantly higher standard adaptive behavior and socialization scores compared to Group 2 or 3 (Extended Fig. 10d, e). Interestingly, the NRXN1 gene that is part of group 2 has been found to have alternative splicing in individuals with neuropsychiatric disorders[82]. This suggests that splice variants may play a more prominent role in these groups with respect to their effect on adaptive behavior and socialization scores. No significant differences were seen between individuals with frame shift or stop gain variants in genes from any group (Extended Fig. 10f, g), possibly due to the lower number of individuals in the analyses, or an equally detrimental impact of these variants on all ASD risk genes. The differences between Group 1 and Groups 2 or 3 suggest that PPI networks can be used to cluster ASD-risk genes, and individuals with variants in those genes. Group 1 genes were found to have the largest enrichment of ASD-risk genes (Extended Fig. 10h), suggesting that the highly interconnected PPI networks and shared pathways for this group of genes may be a core driver for the affected clinical phenotypes (Extended Fig. 10h). The functional grouping of ASD-risk genes highlights the potential of using PPI networks to correlate biological function with clinical phenotype. This could lead to a better approach in subdividing individuals with ASD and understanding the biological basis of these subgroups.

Discussion and Conclusion

ASD is a heterogeneous group of neurodevelopmental disorders that are largely caused by genetic variants in multiple risk genes[2–5]. A long-standing question in the field is how different risk genes contribute to ASD, and whether there are convergent signaling mechanisms that explain how a multitude of genes lead to a common, albeit heterogeneous, developmental brain disorder. Specific disease cell types or signaling pathways have been proposed as convergent mechanisms in ASD[9,19,24–26], but the bulk of these data are based on RNA expression, which does not take into account signaling at the protein level. To address this gap, we devised an *in vitro* neuron-specific proteomic screen to identify individual and shared PPI networks between 41 ASD-risk genes. Our screen identified links between risk genes and multiple convergent signaling pathways. In addition, PPI network mapping could predict the functional impact of disease-associated missense variants. Finally, PPI network mapping of ASD-risk genes revealed an important relevance to human ASD pathology as the network enriched for additional ASD risk genes and cell types implicated in ASD pathology. Cross-referencing the PPI network with human clinical data revealed a biological link between highly interacting ASD-risk genes and ASD diagnostic behavioral severity, demonstrating the clinical relevance of the network.

While other approaches for identifying PPI networks exist, such as Y2H or affinity purification coupled with mass spectrometry in cells lines, these methods can miss weak and transient interactions and signaling networks specific to neurons[31,32,83]. Our

use of BioID2 for the 41 ASD-risk genes revealed shared protein interactions which include direct and indirect interactions between ASD-risk proteins in a neuronal cell type, providing detailed insight into the relationship between the 41-risk proteins. However, unlike these studies we used single canonical isoforms of each gene and therefore some PPI networks may not encompass the full scope of possible interactions in the neuron. Further, the mouse system possesses glial cells required for synaptic maturation, and it is scalable; therefore, the system could be used to screen hundreds of genes. Some caveats of BioID2 include possible biotinylation inefficiencies, protein function impairments, and protein biotinylation selection biases, however, newer proximity-labeling tools could be used to extend the identification of PPI networks[84–87].

Previous genetic screening platforms have identified shared pathways between ASD-risk genes. CRISPR/Cas9 knockout screens have identified cell types and processes associated with groups of ASD-risk genes[20–22]. Since these knockout screens disrupted genes early in development, this may skew results towards neurogenesis deficits. Our BioID2 screen complement CRISPR/Cas9 approaches, given that they can be used to study earlier or later time points, and can be used to study disease-relevant variants. BioID2 can also help to understand the function or role of poorly characterized ASD risk genes using our PPI network pipeline and statistical cut-offs, where most previous studies rely on known compartment localization. Future studies could also be used to study changes in disease-relevant PPI networks in genetic mouse models or patient-derived iPSC neurons and organoids. Since changes in protein interaction complexes or synaptic networks in multiple ASD mouse models have been observed[32,88], this suggests that core ASD networks can reveal risk gene clusters or identify hub genes.

One of the main findings from our study is the identification of multiple convergent and shared pathways between 41 ASD-risk genes that are non-nuclear, which fall into categories pertaining to synaptic transmission, TCA cycle and mitochondrial activity, Wnt signaling, potassium channel activity, MAPK signaling, and other specific signaling pathways. Synaptic transmission and function is widely known in ASD pathophysiology, and Wnt and MAPK signaling have also been disrupted in ASD patient cell lines[89]. We focused on validating the TCA cycle and mitochondrial activity pathways because its dysfunction is indirectly associated with neurodevelopmental disorders[90] and our screen identified many uncharacterized ASD-risk genes associated with this pathway (Fig. 3a). Clinical studies have found mitochondrial and metabolic dysfunction or changes in metabolites in primary lymphocytes or brain tissue in individuals with ASD[55,56,91–96], but whether this is direct or indirect is not known. A mouse model expressing an mtDNA variant was shown to display autism associated behavioral deficits[97], but the variant is weakly associated with ASD. Some ASD associated syndromic disorders, co-morbid disorders and genetic ASD models have

shown deficits in mitochondrial and metabolic processes, however the specific proteins involved were unknown[52,53,105,54,98–104]. Our findings indicate that TCA cycle and mitochondrial activity proteins are interacting with multiple ASD-risk genes, including genes that were not previously connected to metabolic processes. While, two ASD-risk genes (RHEB and BCKDK) have been previously implicated directly in metabolic processes[66,106]. This highlights that our screen can identify relevant protein interactions and may even suggest a more direct connection between mitochondrial/metabolic processes and some genetic models (CDKL5 and KCTD13)[103,105].

Our CRISPR/Cas9 KO studies revealed that multiple ASD-risk genes are important for proper cellular respiration. Interestingly, these genes were all found to interact with citrate synthase (Fig. 3a), suggesting that upstream or downstream regulation may occur between ASD risk genes and TCA cycle function. Deficits in the TCA cycle can cause overreliance on glutaminolysis to produce energy and cause a decrease in synaptic vesicle glutamate levels[107–109]. This shift may help explain the deficits in synaptic transmission in neurons with disruption of synaptic ASD-risk genes, such as *Syngap1* and *Taok2*. The shared PPI network provides an important link between metabolic processes and ASD pathology. These data underscore the value of using PPI networks to map ASD-risk gene connectivity, and to pinpoint which risk genes are involved in convergent mitochondrial/metabolic dysregulation in ASD.

ASD-associated *de novo* missense mutations are enriched in hub genes of known protein interaction networks[30,110]. However, few studies have used proximity-labeling to study the impact of disease-relevant mutations on the PPI network of genes associated with neurodevelopmental or neurological disorders[111–113]. Our BioID2 screen provides functional evidence of the impact ASD-associated *de novo* missense variants have on the PPI network of three ASD-risk genes, as examples. Time- and resource-intensive studies have also investigated multiple variants in single genes in various animal models[114,115]. Additional bioinformatic approaches have been used to determine the pathogenicity of rare missense variants; however, the impact on biological pathways remains to be determined[116,117]. Using neuron-specific PPI networks allows the use of a relevant cell type, while being able to scale up the screen to test multiple single-gene variants in a nonbiased manner and reduced period of time. This approach could have potential applications for variants of unknown significance by providing important information on the severity of a given genetic variant.

The enrichment of an additional 112 ASD risk genes in the shared ASD-risk gene PPI network map and the enrichment of the network in ASD-associated cell types further emphasizes the interconnectedness of ASD risk proteins. Mid-fetal deep cortical projection neurons and superficial cortical glutamatergic neurons are enriched for ASD-risk genes and are associated with autism pathology[17,18]. The ASD-shared PPI network was highly enriched for genes expressed in excitatory and inhibitory neurons, and for

DEGs in individuals with ASD specific to Layers 2/3 excitatory neurons and VIP interneurons. The high connectivity and enrichment of ASD-risk genes within the network reflect its relevance to shared pathways associated with ASD pathology. Future studies will need to distinguish which ASD PPI networks are specific to each cell type, or possible subpopulations, to understand the subtle network changes that impact disease mechanisms.

Of great interest was the ability to group the 41 ASD-risk genes based on their PPI network, and the correlation of these groups to clinical scores in adaptive behavior and socialization. Although we focused specifically on missense/LoF variants, we found that the type of variant in each group of genes had large effects on the average score of the individuals within the group. To work through the complexity, it may be important to combine our analysis with other methods of categorizing mutations (e.g. gnomAD pLI, Polyphen-2) as higher or lower impact[118,119], which would reduce the number of people shared between groups. Based on our findings, we highlight the ability to group ASD-risk genes based on their PPIs and correlate the groups to differences in clinical scores related to ASD.

In conclusion, our neuron-specific 41 ASD-risk gene PPI network map demonstrates that protein signaling networks are relevant to ASD disease pathology, and are missing from transcriptome-based approaches. Our approach is scalable and to our knowledge, represents one of the largest protein network mapping studies for ASD risk genes. This resource containing the individual PPI networks of 41 ASD-risk genes will be valuable for future in-depth study of the genes, and has the potential to grow larger with PPI networks of additional risk genes. Furthermore, the comparison of PPI networks to large-scale human clinical and genetic datasets demonstrates a step towards grouping ASD individuals and risk genes based on biological evidence. Ultimately, the hope is that this approach may translate into a better understanding of wide-ranging ASD clinical phenotypes and the development of targeted therapies.

Material and Methods

Antibodies

The following antibodies were used for immunostaining and immunoblotting experiments: rabbit anti-FLAG (IB 1:2,000, MilliporeSigma, F7425), mouse anti-FLAG (IF 1:1,000, IB: 1:2000, MilliporeSigma, F3165), rabbit anti-turboGFP (IF 1:1,000, IB 1:1,000, Fisher, PA5-22688), chicken anti-MAP2 (IF 1:1,000, Cedarlane, CLN182), rabbit anti-β-actin (IB 1:1,000, Cell Signaling, 8457S), mouse anti-β-actin (IB 1:5,000, MilliporeSigma, A5316), goat anti-TAOK2α/β (IB 1:1,000, Santa Cruz Biotechnology, sc-47447), rabbit anti-TAOK2β (IB 1:1,000, Synaptic Signaling, 395 003), mouse anti-Synapsin1 (IF: 1:1000, Synaptic Systems, 106 001), mouse anti-TOMM20 (IF 1:100, US Biological, 134604), DAPI (IF 300mM, ThermoFisher, D21490), Hoechst (IF 1:10,000, Invitrogen, 1050083), Phalloidin-488 (IF 1:120, Cytoskeleton Inc., PHDG1), Anti-mouse-Cy3 (IF 1:500, Jackson Immunoresearch, 715-165-151), Cy3 anti-mouse (IF 1:500, Jackson Immunoresearch, 715-165-151), Alexa 488 anti-rabbit (IF 1:500, Jackson Immunoresearch, 711-545-152), Alexa 488 anti-chicken (IF 1:500, Jackson Immunoresearch, 703-545-155), 405 conjugated-streptavidin (IF 1:500, Jackson Immunoresearch, 016-470-084), 405 anti-chicken (IF 1:500, Jackson Immunoresearch, 703-475-155), Alexa 647 anti-mouse (IF 1:500, Jackson Immunoresearch, 715-605-150).

Generation of constructs

All cloning was accomplished using the In-Fusion HD cloning kit (Takara). To create the BioID2 fusion constructs, we obtained an expression construct containing a 198bp (13x “GGGGS” repeat) linker sequence upstream of a C-terminal 3xFLAG-tagged BioID2 sequence with BioID2 (Genscript). For lentiviral expression, 13xlinker-BioID2-3xFLAG was amplified and cloned into the lentiviral backbone pLV-hSYN-RFP (Addgene #22909)[120]. For ease of visualization and to create a bicistronic construct, the RFP in the pLV-hSYN-RFP backbone was replaced with the TurboGFP(tGFP)-P2A from pCW57-GFP-2A-MCS (Addgene #71783)[121]. NheI digest sites were added after the P2A sequence and before the 13xLinker to allow easy insertion of ASD-risk bait genes. The final construct being pLV-hSyn-tGFP-P2A-Bait-13xLinker-BioID2-3xFLAG (referred to as the BioID2 fusion construct). For the control luciferase construct a second P2A was cloned in between the luciferase ORF and the 13xLinker, creating the pLV-hSyn-tGFP-P2A-Luciferase-P2A-13xLinker-BioID2-3xFLAG construct (referred to as the Luciferase control construct). ASD-risk genes open reading frames (ORFs) were purchased from Addgene and Genscript or amplified from human adult and fetal brain RNA (Takara) (see Supplementary Table 10)[122,123,132–139,124–131]. For mouse electrophysiology experiments, the GRIA1, GRIA1 R208H and GRIA1 A636T ORFs were inserted between the GFP-P2A and 3xFLAG. The pLV-CMV-Cas9-T2A-EGFP plasmid was made by replacing the UBC promoter-rTetR in the FUW-M2rtTA plasmid (Addgene #20342)[140] with CMV-Cas9-T2A-EGFP from PX458 (Addgene #48138)[141]. All generated constructs are available upon request.

Animal housing

Taok2 Het (*Taok2* +/-) and KO (*Taok2* -/-) mice were created by Kapfhamer *et al.*[142]. The E15-16 or E18 mouse embryo brains were used for cortical neuronal cultures. P21-P23 mice were used for mass spectrometry or RNA sequencing experiments. Animals housed at the Central Animal Facilities at McMaster University were approved for experiments and procedures by the Animal Research Ethics Board (AREB) at McMaster University. Animals housed at the University Medical Center Hamburg-Eppendorf, Hamburg were approved for experiments and procedures by local authorities of the city-state Hamburg (Behörde für Gesundheit und Verbraucherschutz, Fachbereich Veterinärwesen) and the animal care committee of the University Medical Center Hamburg-Eppendorf. All procedures were performed according to the German and European Animal Welfare Act. Animals housed at the Animal Resource Center at University Health Network were approved for experiments and procedures by the University Health Network animal care committees.

Mouse Cortical Neuron Cultures

E15-E16 CD1 mice embryo cortices were harvested using a dissecting microscope and kept in HBSS. Cortices were then digested in 300 µg/ml of papain (Worthington) and 2 U/ml of DNase (Thermo) for 20 minutes at 37 °C. Cortices were then washed three times with mouse plating media (Neurobasal media supplemented with 2 mM GlutaMAX (Thermo), Pen-Strep (Thermo), and 10% FBS(Gibco)). Digested cortices were triturated and put through 40 µm strainer. Cells were counted, suspended in plating media, and plated at 600,000 cells per well of a 12-well plate. Plates were coated with 100 µg/ml poly-D-lysine (mol wt > 300,000, Sigma) and 3 µg/ml Laminin (Sigma). For immunostaining, 12 mm coverslips (Fisher) were placed in the well prior to coating. The cells were incubated at 37 °C (with 5 % CO_2) for one hour, after which plating media was removed and replaced with mouse culturing media (Neurobasal media supplemented with 2 mM GlutaMAX, Pen-Strep, and B27). Cells were grown at 37 °C (with 5 % CO_2) and half media changes were done on day 7 and every 3-4 days onwards.

CRISPR/Cas9 editing of human induced pluripotent stem cells (iPSCs)

All work with the human iPSCs was performed with the approval of the Hamilton Integrated Research Ethics Board. Human iPSCs were maintained on Matrigel (Corning) coated plates using mTeSR1 media (Stem Cell Technologies) and passaged every 3-4 days using ReLeSR (Stem Cell Technologies). Human iPSCs were edited for homozygous knockout of *TAOK2* or heterozygous knock-in of the A135P mutation as described in Deneault *et al.*[70]. MGB probes were ordered from ThermoFisher scientific and ssODN were designed on Benchling.com (Biology Software) and ordered from Integrated DNA Technologies. For the A135P mutation a mutant and wildtype ssODN containing the A135P (G to C) mutation and a PAM site mutation or just the PAM site mutation, respectively, were used to create a heterozygous knock-in.

Human iPSC to neuron differentiation via NGN2 induction

Human iPSCs were cultured on Matrigel (Corning) coated plates using mTeSR1 media (Stem Cell Technologies) and passaged every 3-4 days using ReLeSR (Stem

Cell Technologies) until neural induction. A modified NGN2 induction protocol (Zhang *et al.* 2013) was used to differentiate human iPSCs into excitatory NGN2 neurons[72]. Human iPSCs were dual infected with pTet-O-NGN2-P2A-EGFP and FUW-M2rtTA lentiviruses for dox-inducible expression and were titered for > 90% infection efficiency. On Day -1 iPSCs were singularized using Accutase (Stem Cell Technologies) and plated with mTeSR1 media (supplemented with 10 µM Y-27632) on Matrigel at 400,000 cells per well in a 6-well plate. On Day 0, media exchanged and supplemented with Doxycycline (1 µg/ml). On Day 1 and 2, media was replaced with iNPC media (DMEM/F12 media (Gibco) supplemented with N2 (Gibco), MEM NEAA (Thermo), 2mM GlutaMAX, and Pen-Strep) with Doxycycline and Puromycin (2µg/ml). On Day 3, media was then replaced with iNi media (Neurobasal media with SM1 (Stem Cell Technologies), 2mM GlutaMAX, Pen-Strep, 20 ng/ml BDNF, 20 ng/ml GDNF, and 1 µg/ml Laminin) with Doxycycline. On day 4, differentiated neurons were singularized using Accutase and re-plated at 100,000 cells per well in a 24-well plate in only iNi media. Plates were pre-coated with 20 µg/ml Laminin and 67 ug/ml Poly-ornithine (Sigma). Mouse glial cells were plated on top of the differentiated neurons after 24 hours at a density of 50,000 cells per well. Half-media changes were carried out every other day, and iNi media was supplemented with 2.5 % FBS on Day 9 and onwards. Neurons were grown until day 28 post NGN2-induction.

Generation of high-titer lentivirus

All viruses were made using the 2nd generation lentiviral packaging systems in Lenti-X HEK293 FTT cells (Takara). Lenti-X cells were passaged maximum 3 times before being used for virus production in HEK media (High glucose DMEM with 4 mM GlutaMAX, 1 mM Sodium Pyruvate, and 10 % FBS). Lenti-X cells were passaged once with 500µg/ml Gentamycin (Thermo) to increase T antigen expressing cells. Cells were plated into T150 flasks and each flask was transfected with the BioID2 lentiviral plasmid and the packaging plasmids, pMD2.G and pPAX2 (Addgene #12259 and #12260), using Lipofectamine 2000 in a 3:5 Opti-MEM: HEK media mix. Media was exchanged for fresh media after 5.5-6 hours. Media was harvested twice, first at 48 hours and then at 72 hours post-transection and spun at 100,000xG for 2 hours (maximum acceleration and deceleration). The virus was resuspended in PBS and kept at -80°C until they were used. Larger and unstable viruses were spun at 20,000xG for 4 hours in a table top centrifuge using a 20 % sucrose cushion[143]. See nature exchange protocol for detailed procedure.

Infection of mouse cortical neurons for BioID2 screen

One plate of 7.2 million mouse cortical neurons was considered as one biological replicate. Each cortical neuron culture produced at least 5 plates for four separate BioID2 bait gene samples and one luciferase control sample. Three separate cultures were done in a 3 days span in one week to get 3 biological replicates per protein-of-interest (POI). On days *in vitro* (DIV) 14, the conditioned media from the mouse neuron cultures were removed, leaving only 0.5 ml of media per well. Extra wells with and without coverslips were infected at the same MOI for flow cytometry measurements of GFP positive neurons and immunostaining, respectively. On DIV14, lentivirus with BioID2 fusion constructs were added to each well at an MOI of 0.7 and on DIV17 each

well was supplemented with 50 µM of Biotin. After 18-20 hours, cells for mass spectrometry were lysed with RIPA buffer (1 % NP40, 50 mM Tris-HCl, 150 mM NaCl, 0.1 % SDS, 0.5 % deoxycholic acid, and protease inhibitor cocktail (PIC)) and flash frozen in liquid nitrogen. Cells for flow cytometry were dissociated with 0.25 % Trypsin-EDTA (Fisher) and resuspended in PEF media (PBS with 2 mM EDTA and 5 % FBS) (See flow cytometry section). Cells for immunostaining were fixed with 4 % PFA for 20 minutes, washed with PBS, and kept at 4 °C for staining.

Transfection of HEK293 FT cells for BioID2 screen

10 million HEK293 FT cells were plated in a 10 cm culture dish and transfected 24 hours later with the BioID2 fusion construct plasmids using Lipofectamine 2000. Media was changed 6 hours after transfection and 50µM biotin was added 48 hours post-transfection. Cells were lysed 72 hours post-transfection in RIPA buffer and flash frozen in liquid nitrogen. Each individual plate was considered as biological replicate and three plates were used for each gene and the luciferase control. An extra plate was used for flow cytometry measurements of GFP positive cells.

Processing of mouse cortical neuron and HEK293 FT cell BioID2 samples

Lysed cells were thawed and DNA was digested using benzonase (Sigma). Lysates were than sonicated at high speed for 5 seconds and centrifuged at 20,000xG for 30 minutes. The lysate supernatants were incubated with streptavidin Sepharose beads (GE Healthcare) at 4 °C for 3 hours. Following the incubation, the supernatant was spun down at 100xG for 2 minutes and the supernatant was removed. The beads were than washed once with RIPA buffer, and then six times with 100 mM triethylammonium bicarbonate (TEAB) with centrifugation between each wash. After the final wash, the beads are then resuspended in 100 mM TEAB and sequencing-grade trypsin (Promega) was added to digest the biotinylated proteins on the beads into peptides. The beads were incubated at 37 °C for 16 hours while rotating, and additional trypsin was added and incubated for a further 2 hours. The beads were then pelleted and the supernatant was transferred to a new tube. The beads were washed twice with 100 mM TEAB and each wash was added to the supernatant. The supernatant was then transferred to a 1.5 mL screw cap tub and speed vacuum dried. The dried peptides were stored at 4 °C for TMT-labeling.

Multiplex TMT-labeling of BioID2 samples

Dried peptides were resuspended in 100 mM TEAB. Each sample was TMT-labeled using the TMT 10plex Isobaric Mass Tagging Kit (Thermo). The four genes (proteins-of-interest, POI) were divided into two separate batches and the luciferase control samples were divided between the batches. Each batch had three biological replicates of the two genes and the luciferase control. One luciferase sample chosen at random was divided and labeled with two different labels to determine variance due to labeling efficiencies. In brief, TMT-label resuspended in acetonitrile was added to each sample and incubated at room temperature for one hour. To stop the reaction, 5 % hydroxylamine was then added to the samples and incubated for 15 minutes at room temperature. All ten samples were combined into one tube and divided into two

samples. Both samples were then speed vacuum dried. One sample was kept at -80 °C for storage and the second sample was kept at 4 °C to be run in the mass spectrometer.

Identification of biotinylated proteins from BioID2 screen samples using LC-MS/MS

Peptide samples were resuspended in 0.1% Trifluoroacetic acid (TFA) and loaded for liquid chromatography, which was conducted using a home-made trap-column (5 cm x 200 µm inner diameter; POROS 10 µm 10R2 C18 resin) and a home-made analytical column (50 cm x 50 µm inner diameter; Monitor 5 µm 100A C18 resin), running a 120min (label free) or 180min (TMT) reversed-phase gradient at 70nl/min on a Thermo Fisher Ultimate 3000 RSLCNano UPLC system coupled to a Thermo QExactive HF quadrupole-Orbitrap mass spectrometer. A parent ion scan was performed using a resolving power of 120,000 and then up to the 20 most intense peaks were selected for MS/MS (minimum ion count of 1000 for activation), using higher energy collision induced dissociation (HCD) fragmentation. Dynamic exclusion was activated such that MS/MS of the same m/z (within a range of 10 ppm; exclusion list size = 500) detected twice within 5 seconds were excluded from analysis for 30 seconds. Data were analyzed using Proteome Discoverer 2.2 (Thermo). For protein identification, search was against the Swiss-Prot mouse proteome database (55,366 protein isoform entries)[144], while the search parameters specified a parent ion mass tolerance of 10 ppm, and an MS/MS fragment ion tolerance of 0.02 Da, with up to two missed cleavages allowed for trypsin. Dynamic modification of +16@M was allowed.

Analysis for the identification of ASD-risk and cellular compartment protein PPI networks

Only proteins identified with two unique peptides were used for analysis. Flow cytometry was used to calculate the total GFP in infected neuron samples. If the POI sample had less GFP than the luciferase control sample, the factor needed to equalize the amount of GFP was applied to the protein abundances of the POI samples. Protein abundances were also normalized to the highest total protein count sample for each set of biological replicates. Unpaired one-tailed student's test was used to determine significantly enriched biotinylated proteins in the POI sample using the Log2 abundances of the three biological replicates of the POI samples compared to the luciferase control samples ($p<0.05$)[36]. Significance B outlier test was used to identify significantly biotinylated proteins in the POI sample compared to the luciferase control sample using the average abundance and protein abundance ratio between POI and luciferase samples (SigB $p<0.05$). Only proteins that were found to be significant from both analyses were included in the PPI network. The protein abundance ratio between the luciferase control replicate samples, which were labeled with different TMT labels, was considered to be the minimal ratio required for significance. Any protein that did not surpass this ratio was considered to be a false positive, even if statistically significant, and not included in the PPI network.

Pathway enrichment analyses

All pathway enrichment analysis was done using the g:Profiler GOst functional profiling tool (https://biit.cs.ut.ee/gprofiler/gost)[145]. We used internal sources without

electronic GO annotations for GO biological processes and GO cellular component (compartment), and curated Reactome pathway gene sets from the Bader lab (http://download.baderlab.org/EM_Genesets/)[146]. All three sources were used for the shared ASD-risk gene network proteins. Only GO cellular component enrichment was used for the HEK293 FT cell BioID2 PPI networks, neuron cellular compartment BioID2 PPI networks, and *de novo* missense mutation network BioID2 PPI network comparisons. We compared the protein lists against a custom statistical domain of proteins identified through fractionated mass spectrometry of the mouse brain[147] and combined with any additional proteins identified in the BioID2 screen. The final mouse brain proteome background had a total of 11992 proteins after removing multiple isoforms of the same protein. HEK293 BioID2 PPI networks were compared to all annotated gene lists. The g:Profiler Benjamini-Hochberg FDR multiple correction was used and only pathways with an adj. p-value < 0.05 were considered significantly enriched. For cellular component enrichment for *de novo* missense variant BioID2 PPI networks, the ggplot package in R was used to create the dot plots. For *de novo* missense variant BioID2 bait genes, all proteins identified in the wildtype samples were used for analysis, while for the shared PPI network map only proteins found in all wildtype samples were used for pathway enrichment analysis.

Virus titering and GFP normalization for BioID2 screen

Mouse cortical neurons were cultured as described above and infected on DIV3 at three dilutions of virus (1:100, 1:333, 1:1000). On DIV 5, infected mouse cortical neurons were singularized using 0.25 % Trypsin-EDTA (Fisher) and resuspended in PEF media (PBS with 2 mM EDTA and 5 % FBS). For GFP normalization, DIV18 mouse neurons infected with the BioID2 lentiviruses were dissociated with Trypsin and resuspended in PEF media. CytoFLEX-LX or CANTO II flow cytometers were used to measure the percentage of GFP-positive cells with the 488 laser and 525/40 or 525/50 filters, respectively, using CytExpert software (Beckman Coulter). Functional titers were calculated based on the linear relationship between virus amount and percent of GFP positive cells. Mouse cortical neurons were infected at an MOI of 0.7, where 70 percent of cells were expected to be infected with the BioID2 lentiviral constructs. For normalization, the total GFP per 20,000 GFP-positive cells were quantified by taking the area under the GFP intensity histogram. GFP percentage and total amount was calculated using FlowJo software.

Western blots

HEK293 FT cells were transfected with the BioID2 constructs using Lipofectamine 2000 (Invitrogen) in Opti-MEM: HEK media. Cells were harvested 48 hours post-transfection and lysed with RIPA buffer (with fresh PIC). Lysates were either snap-frozen in liquid nitrogen or taken directly for western blot sample preparation. Thawed or fresh lysates were sonicated at high frequency for 5 seconds and centrifuged at 20,000xG for 5 minutes at 4 °C. Lysates were than quantified using the Bio-Rad Bradford protein assay (Bio-Rad) by measuring absorbance with the SPECTROstar Nano machine and MARS Data analysis software (BMG LABTECH) and diluted to equal concentrations with RIPA buffer. 30-50 µg of protein were run on 8 % or 10 % SDS-PAGE Tris-Glycine gels (depending on the size of the proteins) at 100V for

initial stacking and then 140V for 1-1.25 hours in a Tris-Glycine running buffer. Proteins were then transferred onto PVDF membrane using a Tris-Glycine buffered wet transfer system for 2 hours at constant 200 mA. Blots were then blocked with 5% milk in TBS-T (Tris buffered saline pH 7.4 with 0.1 % Tween). Blots were incubated with primary antibodies overnight in 5 % milk/TBS-T. The next day, membranes were washed three times with TBS-T for 5 minutes each and then incubated with secondary antibodies in 5 % milk/TBS-T for 1 hour. Blots were imaged by incubating them with the Amersham ECL western blotting detection reagent (VWR) for 1 minute and then imaging every 10 seconds for 5 minutes on the ChemiDoc XRS+ machine (Bio-Rad). ImageLab (Bio-Rad) was used for band intensity quantification.

In vitro whole-cell patch clamp recordings of human iPSC-derived neurons and mouse cortical neurons

Human iPSC-derived neurons were used for electrophysiology experiments between days 21-24 of the neural differentiation protocol. Whole-cell patch-clamp recordings were performed at room temperature using Multiclamp 700B amplifier (Molecular Devices) from borosilicate patch electrodes (P-97 puller and P-1000 puller; Sutter Instruments) containing a potassium-based intracellular solution (123 mM K-gluconate, 10 mM KCl, 10 mM HEPES; 1 mM EGTA, 1 mM MgCl2, 0.1 mM CaCl2, 1 mM Mg-ATP, and 0.2mM Na4GTP, pH 7.2). 0.06% sulpharhodamine dye was added to the intracellular solution to confirm the selection of multipolar neurons. The extracellular solution consisted of 140 mM NaCl, 5 mM KCl, 1.25 mM NaH_2PO_4, 1 mM $MgCl_2$, 2 mM $CaCl_2$, 10 mM glucose, and 10 mM HEPES (pH 7.4). Data was digitized at 10 – 20 kHz and low-pass filtered at 1 - 2 kHz. Recordings were omitted if access resistance was >30 MΩ. Whole-cell recordings were clamped at -70 mV and corrected for a calculated -10mV junction potential. Rheobase was determined by a step protocol with 5 pA increments, where the injected current had 25 ms duration. Action potential waveform parameters were all analyzed in reference to the threshold. Repetitive firing step protocols ranged from -20 pA to +50 pA with 5 pA increments. No more than two neurons per coverslip were used to reduce the variability. Data were analyzed using the Clampfit software (Molecular Devices), while phase-plane plots were generated in the OriginPro software (Origin Lab). For GRIA1 overexpression experiments, mouse neurons were infected with GRIA1, GRIA1 R208H, and GRIA1 A636T lentiviral constructs at DIV11 and recorded on DIV14-15. Mouse neurons for electrophysiology experiments were cultured in Neurobasal media (supplemented with an additional 0.3 % (w/v) glucose and 0.22 % (w/v) NaCl). The same intracellular solution was used as the human neuron recordings, with a mouse artificial cerebrospinal fluid extracellular solution (125 mM NaCl, 2.5 mM KCl, 2 mM CaCl2, 1 mM $MgCl_2$, 5 mM HEPES, 33 mM Glucose, pH 7.2). Whole-cell recordings of mouse neurons were clamped at -80 mV and corrected for a calculated -10mV junction potential.

Staining and imaging of mouse cortical neurons and human iPSC-derived neurons

Mouse cortical neurons and human iPSC-derived neurons on coverslips were fixed with 4% paraformaldehyde (PFA) for 10 minutes at room temperature, washed once with PBS, and stored in PBS at 4 °C protected from light. Fixed coverslips were

then blocked and permeabilized in BP solution (PBS with 10% donkey serum and 0.3% Triton-X) for 45 minutes at room temperature. Coverslips were then incubated with primary antibodies at 4 °C overnight. The following day, coverslips were washed three times with PBS for eight minutes each. Coverslips were then incubated with secondary antibodies for one hour at room temperature, followed by three washes with PBS. For human iPSC-derived neuron Synapsin1 staining, coverslips were incubated with 300 mM of DAPI for 15 minutes, before the third wash with PBS. Excess liquid was then removed from the coverslips and they were mounted onto VistaVision glass microscope slides (VWR) with 10 µL of Prolong Gold Anti-Fade mounting medium (Life Technologies). For TOMM20 staining, mouse neurons were fixed with 4% PFA at 37°C for 10 min and then permeabilized with 0.5 % Triton X-100 for 10 minutes. Non-specific binding was blocked by incubation with 5 % donkey serum in PBS for 50 minutes at room temperature, followed by primary antibody incubation. The secondary antibody was added for 50 minutes at room temperature. Primary and secondary antibodies were diluted in PBS with 0.5 % BSA, 2.5 % Donkey-serum, and 0.15 % Triton X-100. After primary and secondary antibody incubation, three washing steps with PBS were performed. Then, coverslips were incubated with Phalloidin-488, for F-actin labeling, and Hoechst dye for 45 minutes at room temperature followed by three PBS washes. Coverslips were mounted onto slides using Fluoromount-G® (Southern Biotech) and were stored protected from light. Synapsin1 and BioID2 stained images were taken on the Zeiss LSM 700 confocal microscope with63x or 40x oil objective, respectively. Mito-dsRed images were taken on the Echo Revolve microscope with a 20x objective.

Synapsin1 puncta analysis in human iPSC-derived neurons

Synapsin1 stained images were processed and analyzed with ImageJ software. The Synapsin1 antibody was co-immunostained with MAP2 to determine dendrites with presynaptic puncta. Five biological replicates, which represent five separate neural inductions, were used for synaptic analysis. 5-10 neurons per genotype per replicate were used. Imaging settings were kept the same between images and synapsin1 images were analyzed at the same threshold. Dendrites were traced using ImageJ and the measure tool was used to quantify the number and size of the puncta within the traced region.

Proteomic profiling of *Taok2* KO mice cortical post-synaptic density fraction through LC-MS/MS

The right cortical lobes of three P21-23 *Taok2* KO mice and five P21-23 wildtype littermates were harvested and differential centrifugation was used to obtain the crude post-synaptic density fraction[148]. PSD fractionations were validated by western blot for PSD-95 and synaptophysin (data not shown). Final post-synaptic density pellets were resuspended using 8 M urea and 100 mM ammonium bicarbonate. Protein samples were then reduced with 10 mM Tris(2-carboxyethyl)phosphine for 45 min at 37 °C, alkylated with 20 mM iodoacetamide for 45 min at room temperature, and digested by trypsin (Promega) (1:50 enzyme-to-protein ratio) overnight at 37 °C. The peptides were desalted with the 10 mg SOLA C18 Plates (Thermo Scientific), dried, and labeled with Multiplex 10-plex TMT labels (Thermo) in 100 mM triethylammonium bicarbonate, and quenched with 5% hydroxylamine before combined. 40 µg of the pooled sample was

separated into 60 fractions by high-pH reverse-phase liquid chromatography (RPLC) using a homemade C18 column (200 µm × 30 cm bed volume, Waters BEH 130 5 µm resin) running a 70 min gradient from 11 to 32% acetonitrile− 20 mM ammonium formate (pH 10) at a flow rate of 5 µL/min. Each fraction was then loaded onto a homemade trap column (200 µm × 5 cm bed volume) packed with POROS 10R2 10 µm resin (Applied Biosystems), followed by a homemade analytical column (50 µm × 50 cm bed volume) packed with Reprosil-Pur 120 C18-AQ 5 µm particles (Dr. Maisch) with an integrated Picofrit nanospray emitter (New Objective). LC-MS experiments were performed on a Thermo Fisher Ultimate 3000 RSLCNano UPLC system that ran a 3 h gradient (11− 38% acetonitrile−0.1% formic acid) at 70 nL/min coupled to a Thermo QExactive HF quadrupole-Orbitrap mass spectrometer. A parent ion scan was performed using a resolving power of 120 000; then, up to 30 of the most intense peaks were selected for MS/MS (minimum ion counts of 1000 for activation) using higher energy collision-induced dissociation (HCD) fragmentation. Dynamic exclusion was activated such that MS/MS of the same m/z (within a range of 10 ppm; exclusion list size = 500) detected twice within 5 seconds was excluded from the analysis for 30 seconds. Data were analyzed using Proteome Discoverer 2.2 (Thermo). For protein identification, search was against the SwissProt mouse proteome database (55,366 protein isoform entries), while the search parameters specified a parent ion mass tolerance of 10ppm, and an MS/MS fragment ion tolerance of 0.02Da, with up to two missed cleavages allowed for trypsin. Dynamic modification of +16@M was allowed. Only proteins with two unique peptides were used for further analysis. Differentially expressed proteins (DEPs) were calculated through Significance B outlier test using the Perseus software[149], and only proteins that had adj. p-value < 0.05 were considered as DEPs.

Transcriptome profiling of *Taok2* KO mice cortices through RNA sequencing

Cortices from *Taok2* KO and wildtype littermates were also harvested for RNA at post-natal day 21-23, with 3 males and 3 females from each genotype. The RNA was extracted using Trizol and was sent for total RNA sequencing at the Center for Applied Genomics (TCAG). mRNA was purified using poly(A) selection to avoid contamination of ribosomal RNAs and miRNAs. All samples were run on one lane resulting in ~31-34 million of read pairs per sample. All analysis was carried out using the open-source platform Galaxy (usegalaxy.org)[150]. RNA reads were checked for good quality using the FastQC tool. The trimmomatic tool was used to identify and trim off known adaptors and remove any bases that have a Phred score of less than 20. FastQC was used again to ensure that adaptor sequences were removed and that the quality of the reads was not affected. We next used the HISAT2 alignment program for alignment of the RNA sequences to the mouse genome GRCm38 (NCBI). On average 85% of reads from mouse samples were aligned once and 5% were aligned more than once to distinct genome locations. Moving on, the featureCounts tool was used to count the number of reads per gene using the same reference genome as the HISAT2 tool. The DESeq2 tool was used to determine the significant differentially expressed genes (DEGs) between *Taok2* WT and KO mouse cortices. Genes were considered as DEGs if they had an adjusted p-value lower than 0.05.

Gene set enrichment analysis (GSEA) of *Taok2* KO mouse proteome and transcriptome

DEGs and DEPs were ranked based on the equation $-\log_{10}$(adj. p-value)*Ln(fold change). GSEA 4.1.0 (Broad Institute)[151,152] was used to run the GSEA preranked test. Tests were run with 1000 permutations, weighted enrichment statistics, and excluding gene sets smaller than 15 and larger than 500 genes. All other settings were kept as default. All mouse GO term gene sets without electronic GO annotations (http://download.baderlab.org/EM_Genesets/) were used for the analysis[146]. Visualization of the enriched gene sets was done on Cytoscape 3.8.2 using the EnrichmentMap app and the AutoAnnotate app was used for clustering similar gene sets[153–155]. All visualized gene sets had an FDR < 0.1.

Seahorse assay of *in vitro* mouse and human iPSC-derived neurons

Mouse cortical neurons were cultured as described above at a density of 30,000 cells/well in the Seahorse XF96 cell culture microplate. CRISPR/Cas9 KO mouse neurons were infected at DIV 7 and assayed at DIV 14. Human iPSC-derived NGN2 neurons were plated on day 4 of dox induction at a density of 50,000 cells per well in the Seahorse XF96 cell culture microplates (Agilent), pre-coated with 20 µg/ml Laminin and 67 µg/ml Poly-ornithine (Sigma). Mouse glia was plated on top of the neurons at a density of 25,000 cells per well, 24 hours later. Cells were used for the Seahorse assay on day 7. The day prior to the seahorse assay the Seahorse XFe96 sensor cartridge was filled with Calibrant XF solution and incubated at 37 °C (without CO_2) overnight. On the day of the assay the Seahorse XF96 microplates were washed twice with 200 µl per well of pre-warmed MST media (Seahorse XF DMEM pH 7.4 media supplemented with 1 mM sodium pyruvate, 2.5 mM GlutaMAX, and 17.5 mM Glucose). The plate was then filled with 180 µl per well MST media and incubated at 37 °C (without CO2) for 1 hour. During the incubation, the mitochondrial stress test drugs were added to the XFe96 sensor cartridge (1 µM Oligomycin for mouse neurons and 3 µM Oligomycin for human neurons, 1 µM FCCP, and 1 µM Rotenone/Antimycin A resuspended in MST media). The cartridge plate with the drug compounds were then put in the Seahorse XFe96 analyzer for calibration. After calibration, the microplate was placed into the Seahorse XFe96 analyzer for the pre-set mitochondrial stress test protocol. Oxygen consumption rates (OCR) were recorded every seven minutes and the drug compounds were added in 21-minute intervals. Oligomycin was used to inhibit ATP-synthase to measure ATP-synthase dependant respiration, FCCP was added to decouple the inner membrane to measure maximal respiration, and Rotenone and Antimycin A were added together to measure non-mitochondrial respiration. After the assay, microplates were frozen at -80°C overnight and cell content was measured using the Cyquant cell proliferation assay (Thermo) by measuring fluorescence with the CLARIOStar machine and MARS data analysis software (BMG LABTECH). Cellular respiration analysis was performed using the Wave software (Agilent) and OCR values were normalized to the number of cells per well.

CRISPR/Cas9 knockout in mouse cortical neurons

Mouse cortical neurons were infected at DIV7 with the pLV CMV-Cas9-T2A-EGFP (MOI 1) and pLV U6-sgRNA/EF1a-mCherry (MOI 3) lentiviruses. Cultures were

allowed to recover until DIV14 and were then taken for the seahorse assay. The GeneArt genomic cleavage detection kit (Thermo) was used to detect insertions or deletions in the targeted sites.

Measuring mitochondrial activity in shRNA knockdown mouse neurons

Embryonic age E15 C57BL6/J mouse pups were *in utero* electroporated with Taok2 shRNA and control shRNA. Electroporated mouse embryo cortices were than harvested and cultured at E18. Mouse neuron cultures were imaged at DIV5 after incubation with 2nM TMRM (Thermo). Images were analyzed on ImageJ. Soma regions were delineated and integrated density in the soma (soma area x mean intensity) was measured. For background correction, mean background intensity was obtained from the neighbouring region.

Measuring mitochondrial activity and content in mouse cortical neurons

DIV 6 mouse cortical neurons cultured from *Taok2* WT, Het, and KO mouse embryos were incubated with 2 nM TMRM (Invitrogen, #T668) and/or 100 nM MitoTracker Green (Cell Signaling Technology, #9074P) were directly added to the conditioned medium, and incubated for 15 minutes. Cells were then imaged within 30 minutes after the incubation time. Images were loaded onto ImageJ, background mean intensity was measured from the region without TMRM and MitoTracker signals inside the cell, then the cell was delineated and the background was removed. After background correction, using the JACoP plugin the TMRM-MitoTracker signal colocalization was analyzed using Manders' correlation coefficients. For Manders' correlation coefficients, threshold values for TMRM (red channel) and MitoTracker (green channel) were set to 335±55 and 640±50 respectively. 16-bit wide field images were taken on a Nikon EclipseTi2 inverted spinning disk microscope equipped with 60X oil (NA 1.4) objective, an LED light source (Lumencor® from AHF analysentechnik AG, Germany), and a digital CMOS camera (ORCA-Flash4.0 V3 C13440-20CU from Hamamatsu) controlled with NIS-Elements software. The microscope imaging chamber is equipped to maintain 37 °C temperature and 5 % CO_2. Illumination, exposure and gain settings were kept the same across different conditions for imaging TMRM and MitoTracker signals.

TOMM20 staining analysis for mitochondria content in mouse cortical neurons

DIV 7-8 mouse cortical neurons cultured from *Taok2* WT, Het, and KO mouse embryos were fixed and stained for TOMM20. 16-bit Z-series images with a step size of 300 nm Images were acquired on confocal spinning disk microscope using a 60X oil (NA 1.4) objectives. Illumination, exposure and gain settings were kept the same across the conditions. The images were loaded onto ImageJ and z-projection (sum slices) for the entire cell in z-axis was performed on the confocal images. Using ImageJ, soma region was carefully delineated and total intensity, also known as integrated density, in the soma (soma area * mean intensity) was measured. For background correction, mean intensity (background mean intensity) was obtained from the neighbouring region (out of the cell). Using the following equation, we obtained the corrected values. Corrected value = total intensity in the soma – (background mean intensity * soma area).

Electron microscopy of synaptic mitochondria from mouse brain cortices

Coronal vibratome sections of the cingulate cortex (cg1 and cg2) and the prelimbic cortex (PL) of the PFC, the primary somatosensory regions S1HL, S1Fl, S1BF, and the intermediate HC were collected and prepared for electron microscopy as described in Richter *et al.*[59]. Semithin sections (0.5 µm) were prepared for light microscopy mounted on glass slides and stained for 1 min with 1% Toluidine blue. Ultrathin sections (60 nm) were examined in an EM902 (Zeiss, Munich, Germany). Pictures were taken with a MegaViewIII digital camera (A. Tröndle, Moorenweis, Germany). EM images that were collected and analyzed for synapse formation on the dendritic spines or shafts from Richter *et al.* were reanalyzed for mitochondrial morphology. Mitochondria morphology from the EM images obtained from Taok2 Wt and Taok2 KO genotypes were analyzed manually using ImageJ. based on their morphology the mitochondria are and categorized to Category 1 - Normal mitochondria with well stacked Cristae, Category 2 - mitochondria with enlarged Cristae, Category 3 - mitochondria with condensed Cristae.

Mito7-dsRed puncta analysis in human iPSC-derived neurons

TAOK2 KO, A135P and wildtype human iPSC-derived NGN2 neurons were transfected with 0.8 µg of Mito7-dsRed (Addgene #55838) and 0.2 µg of pCAG-Venus at day 5, with 2 µl of Lipofectamine 2000 (Thermo). Venus was used to trace neuron projections.10 neurons per genotype from two separate neural inductions were used for analysis. Imaging settings were kept the same between images and Mito7-dsRed images were analyzed at the same threshold. Dendrites were traced using ImageJ and the measure tool was used to quantify the size of the puncta within the traced region

Correlation of 41 ASD-risk gene PPI networks

Corrplot (R package) was used to create the correlation plot. The normalized biotinylation score to the bait protein was used to calculate the correlation between ASD-risk gene PPI networks. The Silhouette and Within cluster sum of squares methods were used to calculate the optimal kmeans number for clustering. Genes were ordered by hierarchal clustering.

Cell type/DEG/ASD gene list enrichment analysis

Human cell type gene expression and ASD DEGs and ASD gene lists were obtained from their respective publications[11–16,25,43]. For the enrichment analysis we used the Fisher exact test comparing each gene list with the shared ASD-risk gene PPI network in the mouse brain background protein list, which was used for pathway enrichment analysis. P-values and ODDs ratios were calculated for each comparison. To account for multiple comparisons, Bonferroni correction thresholds were calculated as $p = 0.05$ divided by the number of comparisons.

Clinical score analysis

Rare variants of individuals diagnosed with ASD were extracted from the MSSNG database (research.mss.ng)[12], which has whole genome sequences of 4,258 families and 5,102 ASD-affected individuals. Only variants with estimated high or medium impact strengths were used for analysis, and variants were categorized into

three groups (missense variants, splicing variants, and frame shift/premature stop codon variants). Adaptive behavior and socialization standard scores of affected individuals was extracted from the MSSNG associated Metabase (data-explorer.mss.ng). Individuals were grouped based on the presence of mutations in the 41 ASD-risk genes that were clustered into three groups. Individuals that had variants in genes between multiple groups were not included in the analysis. Separate analyses were carried out between individuals grouped by missense, splicing or frame shift/premature stop codon variants. Clinical data was considered as non-parametric and the Kruskal-Wallis ranked test with post hoc Dunn's test was used for comparison between the adaptive behavior and socialization standard scores of each group.

Data representation and figure generation

Networks and gene set enrichment maps were created on Cytoscape v3.8.2. Graphs were created on GraphPad Prism 7. Representative electrophysiology traces were extracted onto CorelDRAW. Microscopy images were prepared using ImageJ. Dot plots, correlation plots, and heat maps were created on R Studio. Flowcharts were created on and exported from BioRender.com (SD235B8ORF, KW235KT7TM, RZ235KTA0S). Final figures were organized and created using Adobe Illustrator CC.

Statistics analysis

Data are expressed as mean ± s.e.m, except the clinical analysis which is shown as a box and whisker plot showing the minimum, median, and maximum scores. A minimum of three biological replicates were used for all experiments, where separate HEK cell transfections, iPSC dox-inductions, mouse neuron cultures, or littermates are considered as individual replicates. All statistical analysis was done on GraphPad Prism 7. All comparisons were assumed to be parametric, except for the clinical score analyses. ROUT's outlier test was used to identify possible outliers, with a Q value of 0.1 %. For statistical analysis unpaired t-test, or One-Way ANOVA and Two-Way ANOVA with *post hoc* Holm-Sidak tests were used to compare all experimental conditions to the control condition. All unpaired t-tests were two-sided, except for the one-sided t-test used for identification of BioID2 prey proteins. Clinical scores were assumed to be non-parametric and the Kruskal-Wallis H test with post hoc Dunn's test was used to compare all groups to each other. Any variation from the described statistical analyses is described and explained in the figure legends. The p-values are defined in the figure legends and $p < 0.05$ are considered statistically significant.

Data availability

Mass spectrometry datasets consisting of raw files and results files with statistical analysis to identify PPI networks or significant DEPs will be deposited into ProteomeXchange through the Proteomics Identification Database. Individual PPI networks and shared ASD-risk gene PPI network map protein lists and enriched pathways can be found in Supplementary Tables 1-9. The Mouse_Human_Reactome and Mouse_GO_ALL_no_GO_iea gene sets used for overrepresentation and gene set enrichment analyses were downloaded on 13 August 2021 from http://download.baderlab.org/EM_Genesets/[146]. RNA sequencing raw sequence files and results files with statistical analysis to identify significant DEGs will be deposited

into the Gene Expression Omnibus. ASD proband variant information and clinical scores are available through the MSSNG database (research.mss.ng)[12] and the associated Metabase (data-explorer.mss.ng), respectively.

References

1. Ofner, M. *et al. Autism Spectrum Disorder among Children and Youth in Canada 2018*. (2018).
2. Gaugler, T. *et al.* Most genetic risk for autism resides with common variation. *Nat. Genet.* **46**, 881–885 (2014).
3. Iossifov, I. *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515**, 216–221 (2014).
4. Robinson, E. B. *et al.* Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nat. Genet.* **48**, 552–555 (2016).
5. Weiner, D. J. *et al.* Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. *Nat. Genet.* **49**, 978–985 (2017).
6. Glessner, J. T. *et al.* Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* **459**, 569–573 (2009).
7. Iossifov, I. *et al.* De Novo Gene Disruptions in Children on the Autistic Spectrum. *Neuron* **74**, 285–299 (2012).
8. Neale, B. M. *et al.* Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* **485**, 242–245 (2012).
9. O'Roak, B. J. *et al.* Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* **485**, 246–250 (2012).
10. Sanders, S. J. *et al.* De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 1–6 (2012).
11. Sanders, S. J. *et al.* Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* **87**, 1215–1233 (2015).
12. Yuen, R. K. C. *et al.* Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nat. Neurosci.* **20**, 602–611 (2017).
13. Feliciano, P. *et al.* Exome sequencing of 457 autism families recruited online provides evidence for novel ASD genes. *Genomic Med.* **4**, https://doi.org/10.1038/s41525-019-0093-8 (2019).
14. Ruzzo, E. K. *et al.* Inherited and De Novo Genetic Risk for Autism Article Inherited and De Novo Genetic Risk for Autism Impacts Shared Networks. *Cell* **178**, 850–866 (2019).
15. Satterstrom, F. K. *et al.* Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell* **180**, 568-584.e23 (2020).
16. Wilfert, A. B. *et al.* Recent ultra-rare inherited variants implicate new autism candidate risk genes. *Nat. Genet.* **53**, 1125–1134 (2021).
17. Parikshak, N. N. *et al.* Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* **155**, 1008 (2013).

18. Willsey, A. J. *et al.* Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* **155**, 997 (2013).
19. Ramaswami, G. *et al.* Integrative genomics identifies a convergent molecular subtype that links epigenomic with transcriptomic differences in autism. *Nat. Commun.* **11**, 1–14 (2020).
20. Jin, X. *et al.* In vivo Perturb-Seq reveals neuronal and glial abnormalities associated with autism risk genes. *Science (80-.).* **370**, eaaz6063 (2020).
21. Cederquist, G. Y. *et al.* A Multiplex Human Pluripotent Stem Cell Platform Defines Molecular and Functional Subclasses of Autism-Related Genes. *Cell Stem Cell* **27**, 35-49.e6 (2020).
22. Willsey, H. R. *et al.* Parallel in vivo analysis of large-effect autism genes implicates cortical neurogenesis and estrogen in risk and resilience. *Neuron* **109**, 788–804 (2021).
23. O'Roak, B. J. *et al.* Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science (80-.).* **338**, 1619–1622 (2012).
24. Voineagu, I. *et al.* Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474**, 380–384 (2011).
25. Velmeshev, D. *et al.* Single-cell genomics identifies cell type–specific molecular changes in autism. *Science (80-.).* **364**, 685–689 (2019).
26. Chang, J., Gilman, S. R., Chiang, A. H., Sanders, S. J. & Vitkup, D. Genotype to phenotype relationships in autism spectrum disorders. *Nat. Neurosci.* **18**, 191–198 (2015).
27. De Rubeis, S. *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209–215 (2014).
28. Kuzmanov, U. & Emili, A. Protein-protein interaction networks: Probing disease mechanisms using model systems. *Genome Med.* **5**, 1–12 (2013).
29. Murtaza, N., Uy, J. & Singh, K. K. Emerging proteomic approaches to identify the underlying pathophysiology of neurodevelopmental and neurodegenerative disorders. *Mol. Autism* **11**, 27 (2020).
30. Chen, S. *et al.* De novo missense variants disrupting protein-protein interactions affect risk for autism through gene co-expression and protein networks in neuronal cell types. *Mol. Autism* **11**, 1–16 (2020).
31. Corominas, R. *et al.* Protein interaction network of alternatively spliced isoforms from brain links genetic risk factors for autism. *Nat. Commun.* **5**, 1–12 (2014).
32. Sakai, Y. *et al.* Protein interactome reveals converging molecular pathways among autism disorders. *Sci. Transl. Med.* **3**, 86ra49 (2011).
33. Lage, K. Protein-protein interactions and genetic diseases: The interactome. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1842**, 1971–1980 (2014).
34. Richards, A. L., Eckhardt, M. & Krogan, N. J. Mass spectrometry-based protein–protein interaction networks for the study of human diseases. *Mol. Syst. Biol.* **17**, 1–18 (2021).
35. Casanova, E. L. *et al.* Autism risk genes are evolutionarily ancient and maintain a unique feature landscape that echoes their function. *Autism Res.* **12**, 860–869 (2019).
36. Uezu, A. *et al.* Identification of an elaborate complex mediating postsynaptic inhibition. *Science (80-.).* **353**, 1123–1129 (2016).

37. Spence, E. F. *et al.* In vivo proximity proteomics of nascent synapses reveals a novel regulator of cytoskeleton-mediated synaptic maturation. *Nat. Commun.* **10**, 1–16 (2019).
38. Loh, K. H. *et al.* Proteomic Analysis of Unbounded Cellular Compartments : Synaptic Clefts Resource Proteomic Analysis of Unbounded Cellular Compartments : Synaptic Clefts. *Cell* **166**, 1295-1307.e21 (2016).
39. Hamdan, H. *et al.* Mapping axon initial segment structure and function by multiplexed proximity biotinylation. *Nat. Commun.* **11**, (2020).
40. Go, C. D. *et al.* A proximity-dependent biotinylation map of a human cell. *Nature* **595**, (2021).
41. Markmiller, S. *et al.* Context-Dependent and Disease-Specific Diversity in Protein Interactions within Stress Granules. *Cell* **172**, 590–604 (2018).
42. Hung, V. *et al.* Proteomic mapping of cytosol-facing outer mitochondrial and ER membranes in living human cells by proximity biotinylation. *Elife* **6**, e24463 (2017).
43. Nowakowski, T. J. *et al.* Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science (80-.).* **358**, 1318–1323 (2017).
44. Barres, B. A. The Mystery and Magic of Glia: A Perspective on Their Roles in Health and Disease. *Neuron* **60**, 430–440 (2008).
45. Stogsdill, J. A. & Eroglu, C. The interplay between neurons and glia in synapse development and plasticity. *Curr. Opin. Neurobiol.* **42**, 1–8 (2017).
46. Wilton, D. K., Dissing-Olesen, L. & Stevens, B. Neuron-Glia Signaling in Synapse Elimination. *Annu. Rev. Neurosci.* **42**, 107–127 (2019).
47. Snel, B., Lehmann, G., Bork, P. & Huynen, M. A. String: A web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res.* **28**, 3442–3444 (2000).
48. Trost, B. *et al.* Genome-wide detection of tandem DNA repeats that are expanded in autism. *Nature* **586**, 80–86 (2020).
49. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
50. Fernández, E. *et al.* Targeted tandem affinity purification of PSD-95 recovers core postsynaptic complexes and schizophrenia susceptibility proteins. *Mol. Syst. Biol.* **5**, (2009).
51. Fazal, F. M. *et al.* Atlas of Subcellular RNA Localization Revealed by APEX-Seq. *Cell* **178**, 473-490.e26 (2019).
52. Bülow, P., Zlatic, S. A., Wenner, P. A., Bassell, G. J. & Faundez, V. FMRP attenuates activity dependent modifications in the mitochondrial proteome. *Mol. Brain* **14**, 1–11 (2021).
53. Licznerski, P. *et al.* ATP Synthase c-Subunit Leak Causes Aberrant Cellular Metabolism in Fragile X Syndrome. *Cell* **182**, 1170–1185 (2020).
54. Shulyakova, N., Andreazza, A. C., Mills, L. R. & Eubanks, J. H. Mitochondrial dysfunction in the pathogenesis of rett syndrome: Implications for mitochondria-targeted therapies. *Front. Cell. Neurosci.* **11**, 1–9 (2017).
55. Rose, S. *et al.* Butyrate enhances mitochondrial function during oxidative stress in

cell lines from boys with autism. *Transl. Psychiatry* **8**, (2018).
56. Rose, S. *et al.* Oxidative stress induces mitochondrial dysfunction in a subset of autistic lymphoblastoid cell lines. *Transl. Psychiatry* **4**, 1–8 (2014).
57. Shen, L., Feng, C., Zhang, K., Chen, Y. & Gao, Y. Proteomics Study of Peripheral Blood Mononuclear Cells (PBMCs) in Autistic Children. *Front. Cell. Neurosci.* **13**, 1–16 (2019).
58. Frye, R. E. Mitochondrial Dysfunction in Autism Spectrum Disorder : Unique Abnormalities and Targeted Treatments. *Semin. Pediatr. Neurol.* **35**, 100829 (2020).
59. Richter, M. *et al.* Altered TAOK2 activity causes autism-related neurodevelopmental and cognitive abnormalities through RhoA signaling. *Mol. Psychiatry* **24**, 1329–1350 (2019).
60. Ultanir, S. K. K. *et al.* MST3 Kinase Phosphorylates TAO1/2 to Enable Myosin Va Function in Promoting Spine Synapse Development. *Neuron* **84**, 968–982 (2014).
61. Calderon de Anda, F. *et al.* Autism spectrum disorder susceptibility gene TAOK2 affects basal dendrite formation in the neocortex. *Nat. Neurosci.* **15**, 1022–1031 (2012).
62. Yadav, S. *et al.* TAOK2 Kinase Mediates PSD95 Stability and Dendritic Spine Maturation through Septin7 Phosphorylation. *Neuron* **93**, 1–15 (2017).
63. Da Costa, R. M., Karmirian, K. & Rehen, S. K. Deformation of mitochondrial cristae in human neural progenitor cells exposed to valproic acid. *An. Acad. Bras. Cienc.* **90**, 2223–2232 (2018).
64. John, G. B. *et al.* The Mitochondrial Inner Membrane Protein Mitofilin Controls Cristae Morphology. *Mol Biol Cell* **16**, 1543–1554 (2005).
65. Schmidt, O., Pfanner, N. & Meisinger, C. Mitochondrial protein import: From proteomics to functional mechanisms. *Nat. Rev. Mol. Cell Biol.* **11**, 655–667 (2010).
66. Yang, W. *et al.* Rheb mediates neuronal-activity-induced mitochondrial energetics through mTORC1-independent PDH activation. *Dev. Cell* **56**, 811-825.e6 (2021).
67. Hart, T. *et al.* High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell* **163**, 1515–1526 (2015).
68. Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol.* **34**, 339–344 (2016).
69. Ruggiero, A., Katsenelson, M. & Slutsky, I. Trends in Mitochondria : new players in homeostatic regulation of fi ring rate set points. *Trends Neurosci.* **44**, 605–618 (2021).
70. Deneault, E. *et al.* Complete Disruption of Autism-Susceptibility Genes by Gene Editing Predominantly Reduces Functional Connectivity of Isogenic Human Neurons. *Stem Cell Reports* **11**, 1211–1225 (2018).
71. Deneault, E. *et al.* CNTN5-/+or EHMT2-/+human iPSC-derived neurons from individuals with autism develop hyperactive neuronal networks. *Elife* **8**, 1–26 (2019).
72. Zhang, Y. *et al.* Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron* **78**, 785–798 (2013).
73. Geisheker, M. R. *et al.* Hotspots of missense mutation identify

neurodevelopmental disorder genes and functional domains. *Nat. Neurosci.* **20**, 1043–1051 (2017).

74. de Ligt, J. *et al.* Diagnostic Exome Sequencing in Persons with Severe Intellectual Disability. *N. Engl. J. Med.* **367**, 1921–1929 (2012).
75. Pick, J. E. & Ziff, E. B. Regulation of AMPA receptor trafficking and exit from the endoplasmic reticulum. *Mol. Cell. Neurosci.* **91**, 3–9 (2018).
76. Schwenk, J. *et al.* An ER Assembly Line of AMPA-Receptors Controls Excitatory Neurotransmission and Its Plasticity. *Neuron* **104**, 680-692.e9 (2019).
77. Shang, L. *et al.* De novo missense variants in PPP2R5D are associated with intellectual disability, macrocephaly, hypotonia, and autism. *Neurogenetics* **17**, 43–49 (2016).
78. Houge, G. *et al.* B56δ-related protein phosphatase 2A dysfunction identified in patients with intellectual disability. *J. Clin. Invest.* **125**, 3051–3062 (2015).
79. Papke, C. M. *et al.* A disorder-related variant (E420K) of a PP2A-regulatory subunit (PPP2R5D) causes constitutively active AKT-mTOR signaling and uncoordinated cell growth. *J. Biol. Chem.* **296**, 100313 (2021).
80. Xu, Z. X. *et al.* Elevated protein synthesis in microglia causes autism-like synaptic and behavioral aberrations. *Nat. Commun.* **11**, (2020).
81. Tang, G. *et al.* Loss of mTOR-Dependent Macroautophagy Causes Autistic-like Synaptic Pruning Deficits. *Neuron* **83**, 1131–1143 (2014).
82. Flaherty, E. *et al.* Neuronal impact of patient-specific aberrant NRXN1α splicing. *Nat. Genet.* **51**, 1679–1690 (2019).
83. Li, J. *et al.* Identification of Human Neuronal Protein Complexes Reveals Biochemical Activities and Convergent Mechanisms of Action in Autism Spectrum Disorders. *Cell Syst.* **1**, 361–374 (2015).
84. May, D. G., Scott, K. L., Campos, A. R. & Roux, K. J. Comparative Application of BioID and TurboID for Protein-Proximity Biotinylation. *Cells* **9**, (2020).
85. Minde, D. P., Ramakrishna, M. & Lilley, K. S. Biotin proximity tagging favours unfolded proteins and enables the study of intrinsically disordered regions. *Commun. Biol.* **3**, 1–13 (2020).
86. Niinae, T., Imami, K., Sugiyama, N. & Ishihama, Y. Proximity Labeling-assisted Identification of Endogenous Kinase Substrates. *BioRxiv* (2020). doi:10.1101/2020.06.09.143370
87. Branon, T. C. *et al.* Efficient proximity labeling in living cells and organisms with TurboID. *Nat. Biotechnol.* **36**, 880–898 (2018).
88. Brown, E. A. *et al.* Clustering the autisms using glutamate synapse protein interaction networks from cortical and hippocampal tissue of seven mouse models. *Mol. Autism* **9**, 1–16 (2018).
89. Gazestani, V. H. *et al.* A perturbed gene network containing PI3K-AKT, RAS-ERK and WNT-B-catenin pathways in leukocytes is linked to ASD genetics and symptom severity. *Nat. Neurosci.* **22**, (2019).
90. Iwata, R., Casimir, P. & Vanderhaeghen, P. Mitochondrial dynamics in postmitotic cells regulate neurogenesis. *Science (80-.).* **862**, 858–862 (2020).
91. Rossignol, D. A. & Frye, R. E. Mitochondrial dysfunction in autism spectrum disorders: A systematic review and meta-analysis. *Mol. Psychiatry* **17**, 290–314 (2012).

92. Frye, R. E. *et al.* Redox metabolism abnormalities in autistic children associated with mitochondrial disease. *Transl. Psychiatry* **3**, 1–12 (2013).
93. Frye, R. E. & Rossignol, D. A. Mitochondrial dysfunction can connect the diverse medical symptoms associated with autism spectrum disorders. *Pediatr. Res.* **69**, 41–47 (2011).
94. Anitha, A. *et al.* Brain region-specific altered expression and association of mitochondria-related genes in autism. *Mol. Autism* **3**, 1–12 (2012).
95. Wang, Y., Picard, M. & Gu, Z. Genetic Evidence for Elevated Pathogenicity of Mitochondrial DNA Heteroplasmy in Autism Spectrum Disorder. *PLoS Genet.* **12**, 1–24 (2016).
96. Kurochkin, I. *et al.* Metabolome signature of autism in the human prefrontal cortex. *Commun. Biol.* **2**, 1–10 (2019).
97. Yardeni, T., Cristancho, A. G., Mccoy, A. J., Schaefer, P. M. & Mcmanus, M. J. An mtDNA mutant mouse demonstrates that mitochondrial deficiency can result in autism endophenotypes. *Proc. Natl. Acad. Sci. U. S. A.* **118**, 1–11 (2021).
98. Shen, M. *et al.* Reduced mitochondrial fusion and Huntingtin levels contribute to impaired dendritic maturation and behavioral deficits in Fmr1-mutant mice. *Nat. Neurosci.* **22**, 386–400 (2019).
99. Ebrahimi-fakhari, D. *et al.* Impaired Mitochondrial Dynamics and Mitophagy in Neuronal Models of Tuberous Sclerosis Complex. *Cell Rep.* **17**, 1053–1070 (2016).
100. Kanellopoulos, A. K. *et al.* Aralar Sequesters GABA into Hyperactive Article Aralar Sequesters GABA into Hyperactive Mitochondria , Causing Social Behavior Deficits. *Cell* **180**, 1178-1197.e20 (2020).
101. Menzies, C. *et al.* Distinct basal metabolism in three mouse models of neurodevelopmental disorders. *eNeuro* **8**, 1–14 (2021).
102. Li, J. *et al.* Mitochondrial deficits in human iPSC-derived neurons from patients with 22q11.2 deletion syndrome and schizophrenia. *Transl. Psychiatry* **9**, (2019).
103. Jagtap, S. *et al.* Aberrant mitochondrial function in patient-derived neural cells from CDKL5 deficiency disorder and Rett syndrome. *Hum. Mol. Genet.* **28**, 3625–3636 (2019).
104. Fernandez, A. *et al.* Mitochondrial Dysfunction Leads to Cortical Under-Connectivity and Cognitive Impairment 22q11. *Neuron* **102**, 1127-1142.e3 (2019).
105. Madison, J. M. *et al.* Regulation of purine metabolism connects KCTD13 to a metabolic disorder with autistic features. *iScience* **24**, 101935 (2021).
106. Heinemann-Yerushalmi, L. *et al.* BCKDK regulates the TCA cycle through PDC in the absence of PDK family during embryonic development. *Dev. Cell* **56**, 1182-1194.e6 (2021).
107. Namba, T., Nardelli, J., Gressens, P. & Huttner, W. B. Perspective Metabolic Regulation of Neocortical Expansion in Development and Evolution. *Neuron* **109**, 408–419 (2021).
108. Fendt, S. M. & Verstreken, P. Neurons eat glutamate to stay alive. *J. Cell Biol.* **216**, 863–865 (2017).
109. Divakaruni, S. S. *et al.* Long-Term Potentiation Requires a Rapid Burst of Dendritic Mitochondrial Fission during Induction. *Neuron* **100**, 1–16 (2018).
110. Chen, S. *et al.* An interactome perturbation framework prioritizes damaging

missense mutations for developmental disorders. *Nat. Genet.* **50**, 1032–1040 (2018).
111. Chou, C. *et al.* TDP-43 pathology disrupts nuclear pore complexes and nucleocytoplasmic transport in ALS/FTD. *Nat. Neurosci.* **21**, (2018).
112. Tracy, T. E. *et al.* Tau interactome mapping reveals dynamic processes in synapes and mitochondria associated with neurodegenerative disease. *bioRxiv* https://doi.org/10.1101/2021.06.17.448349 (2021).
113. Pintacuda, G. *et al.* Interaction studies of risk proteins in human induced neurons reveal convergent biology and novel mechanisms underlying autism spectrum disorders. *medRxiv* https://doi.org/10.1101/2021.10.07.21264575 (2021).
114. Meili, F. *et al.* Multi-parametric analysis of 57 SYNGAP1 variants reveal impacts on GTPase signaling, localization, and protein stability. *Am. J. Hum. Genet.* **108**, 148–162 (2021).
115. Post, K. L. *et al.* Multi-model functionalization of disease-associated PTEN missense mutations identifies multiple molecular mechanisms underlying protein dysfunction. *Nat. Commun.* **11**, (2020).
116. Koire, A. *et al.* A method to delineate de novo missense variants across pathways prioritizes genes linked to autism. *Sci. Transl. Med.* **13**, eabc1739 (2021).
117. Wu, Y., Li, R., Sun, S., Weile, J. & Roth, F. P. Improved pathogenicity prediction for rare human missense variants. *Am. J. Hum. Genet.* **108**, 1891–1906 (2021).
118. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
119. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
120. Nathanson, J. L., Yanagawa, Y., Obata, K. & Callaway, E. M. Preferential labeling of inhibitory and excitatory cortical neurons by endogenous tropism of adeno-associated virus and lentivirus vectors. nsc. *NSC* **161**, 441–450 (2009).
121. Barger, C. J., Branick, C., Chee, L. & Karpf, A. R. Pan-cancer analyses reveal genomic features of FOXM1 overexpression in cancer. *Cancers (Basel).* **11**, 251 (2019).
122. Johannessen, C. M. *et al.* COT drives resistance to RAF inhibition through MAP kinase pathway reactivation. *Nature* **468**, 968–972 (2010).
123. Kim, E. *et al.* Systematic functional interrogation of rare cancer variants identifies oncogenic alleles. *Cancer Discov.* **6**, 714–726 (2016).
124. Braun, J. E., Huntzinger, E., Fauser, M. & Izaurralde, E. GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets. *Mol. Cell* **44**, 120–133 (2011).
125. Butko, M. T. *et al.* Fluorescent and photo-oxidizing TimeSTAMP tags track protein fates in light and electron microscopy. *Nat. Neurosci.* **15**, 1742–1751 (2012).
126. Ohno, H. *et al.* Localization of p0071-interacting proteins, plakophilin-related armadillo-repeat protein-interacting protein (PAPIN) and ERBIN, in epithelial cells. *Oncogene* **21**, 7042–7049 (2002).
127. Wang, S., Tukachinsky, H., Romano, F. B. & Rapoport, T. A. Cooperation of the ER-shaping proteins atlastin, lunapark, and reticulons to generate a tubular membrane network. *Elife* **5**, 1–29 (2016).
128. Hiday, A. C. *et al.* Mechanisms and Consequences of Dopamine Depletion-

Induced Attenuation of the Spinophilin/Neurofilament Medium Interaction. *Neural Plast.* **2017**, (2017).

129. Seeling, J. M. *et al.* Regulation of β-catenin signaling by the B56 subunit of protein phosphatase 2A. *Science (80-.).* **283**, 2089–2091 (1999).
130. Urano, J. *et al.* Point mutations in TOR confer Rheb-independent growth in fission yeast and nutrient-independent mammalian TOR signaling in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 3514–3519 (2007).
131. Solowska, J. M., Garbern, J. Y. & Baas, P. W. Evaluation of loss of function as an explanation for SPG4-based hereditary spastic paraplegia. *Hum. Mol. Genet.* **19**, 2767–2779 (2010).
132. Braganza, A. *et al.* UBE3B is a calmodulin-regulated, mitochondrion-associated E3 ubiquitin ligase. *J. Biol. Chem.* **292**, 2470–2484 (2017).
133. Sowa, M. E., Bennett, E. J., Gygi, S. P. & Harper, J. W. Defining the Human Deubiquitinating Enzyme Interaction Landscape. *Cell* **138**, 389–403 (2009).
134. Cummins, J. M. & Vogelstein, B. HAUSP is required for p53 destabilization. *Cell Cycle* **3**, 689–692 (2004).
135. Howarth, M., Takao, K., Hayashi, Y. & Ting, A. Y. Targeting quantum dots to surface proteins in living cells with biotin ligase. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 7583–7588 (2005).
136. Furlong, R. A., Narain, Y., Rankin, J., Wyttenbach, A. & Rubinsztein, D. C. α-Synuclein overexpression promotes aggregation of mutant huntingtin. *Biochem. J.* **346**, 577–581 (2000).
137. Małecki, J., Ho, A. Y. Y., Moen, A., Dahl, H. A. & Falnes, P. Human METTL20 is a mitochondrial lysine methyltransferase that targets the β subunit of electron transfer flavoprotein (ETFβ) and modulates its activity. *J. Biol. Chem.* **290**, 423–434 (2015).
138. Alford, S. C., Ding, Y., Simmen, T. & Campbell, R. E. Dimerization-dependent green and yellow fluorescent proteins. *ACS Synth. Biol.* **1**, 569–575 (2012).
139. Lu, G. *et al.* The myeloma drug lenalidomide promotes the cereblon-dependent destruction of ikaros proteins. *Science (80-.).* **343**, 305–309 (2014).
140. Hockemeyer, D. *et al.* A Drug-Inducible System for Direct Reprogramming of Human Somatic Cells to Pluripotency. *Cell Stem Cell* **3**, 346–353 (2008).
141. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
142. Kapfhamer, D. *et al. Taok2* controls behavioral response to ethanol in mice. *Genes, Brain Behav.* **12**, 87–97 (2013).
143. Yacoub, N. al, Romanowska, M., Haritonova, N. & Foerster, J. Optimized production and concentration of lentiviral vectors containing large inserts. *J. Gene Med.* **9**, 579–584 (2012).
144. Bateman, A. *et al.* UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).
145. Raudvere, U. *et al.* G:Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* **47**, W191–W198 (2019).
146. Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G. D. Enrichment map: A network-based method for gene-set enrichment visualization and interpretation.

PLoS One **5**, e13984 (2010).

147. Sharma, K. *et al.* Cell type- and brain region-resolved mouse brain proteome. *Nat. Neurosci.* **18**, 1819–1831 (2015).
148. Kwan, V. *et al.* DIXDC1 Phosphorylation and Control of Dendritic Morphology Are Impaired by Rare Genetic Variants. *Cell Rep.* **17**, 1892–1904 (2016).
149. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740 (2016).
150. Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* **46**, W537–W544 (2018).
151. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* **102**, 15545–15550 (2005).
152. Daly, M. J. *et al.* PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **34**, 267–273 (2003).
153. Shannon, P. *et al.* Cytoscape: A Software Environment for Integrated Models. *Genome Res.* **13**, 2498–2504 (2003).
154. Bader, G. D., Isserlin, R., Merico, D. & Voisin, V. Enrichment Map - a Cytoscape app to visualize and explore OMICs pathway enrichment results. *F1000Research* **3**, 1–8 (2014).
155. Kucera, M., Isserlin, R., Arkhangorodsky, A. & Bader, G. D. AutoAnnotate: A Cytoscape app for summarizing networks with semantic annotations [version 1; referees: 2 approved]. *F1000Research* **5**, 1–12 (2016).

Acknowledgements

We thank C.I. and P.dG. for proof reading the manuscript. We also thank K.J.B. and M.B.F. for providing the NRXN1 cDNA. Flowcharts were created with BioRender.com (SD235B8ORF, KW235KT7TM, RZ235KTA0S). Work in the Singh Lab was supported by the Canadian Institute of Health Research (CIHR), the Ontario Brain Institute (OBI), the Network for European Funding for Neuroscience Research (NEURON ERA-NET), and the Donald K. Johnson Eye Institute at University Health Network.

Author contributions

N.M. and K.K.S. conceived the project. N.M. and K.K.S. wrote the paper with input from A.A.C., B.T., W.E., B.T., and B.W.D. A.A.C and B.K.U created initial BioID2 lentiviral construct backbone. N.M. generated all of subsequent DNA constructs and performed all experiments and data analysis unless otherwise specified. N.M and A.A.C. generated all lentiviruses. S.X. and Y.L. ran samples through the mass spectrometer and helped with data acquisition. C.O.B. performed all electrophysiology recordings. J.A.U, J.E.H., and N.P. helped perform western blots. D.P.M., S.H., B.S., and F.C.dA. performed and analysed mitochondrial activity and content experiments in mouse cortical neurons. E.D., J.E, and S.W.S helped to create the human *TAOK2* KO and A135P iPSC lines. E.A. advised on clinical score analysis and G.D.B. advised on pathway analyses used in the project. K.S.S supervised the project.

Competing interests

The authors declare no competing interests

Materials & Correspondence

Correspondence and material requests to Karun K. Singh

Supplementary Tables

Supplementary Table 1. BioID2 PPI networks of 41 ASD-risk genes and cellular compartment genes

Supplementary Table 2. Comparison of BioID2 PPI networks identified in HEK293 cells and mouse cortical neurons

Supplementary Table 3. Comparison of BioID2 PPI network enriched cellular components identified in HEK293 cells and mouse cortical neurons

Supplementary Table 4. BioID2 PPI network enriched cellular components of compartment specific genes

Supplementary Table 5. BioID2 PPI network enriched pathways of 41 ASD-risk genes

Supplementary Table 6. 41 ASD-risk gene PPI network map enriched pathways

Supplementary Table 7. Differentially expressed genes and proteins and dysregulated pathways in *Taok2* KO mouse cortices

Supplementary Table 8. Comparison of BioID2 PPI networks between ASD-risk genes and their variants

Supplementary Table 9. BioID2 PPI network enriched pathways of ASD-risk genes and their variants

1756 Supplementary Table 10. List of sources for 41 ASD-risk genes and cellular
1757 compartment genes

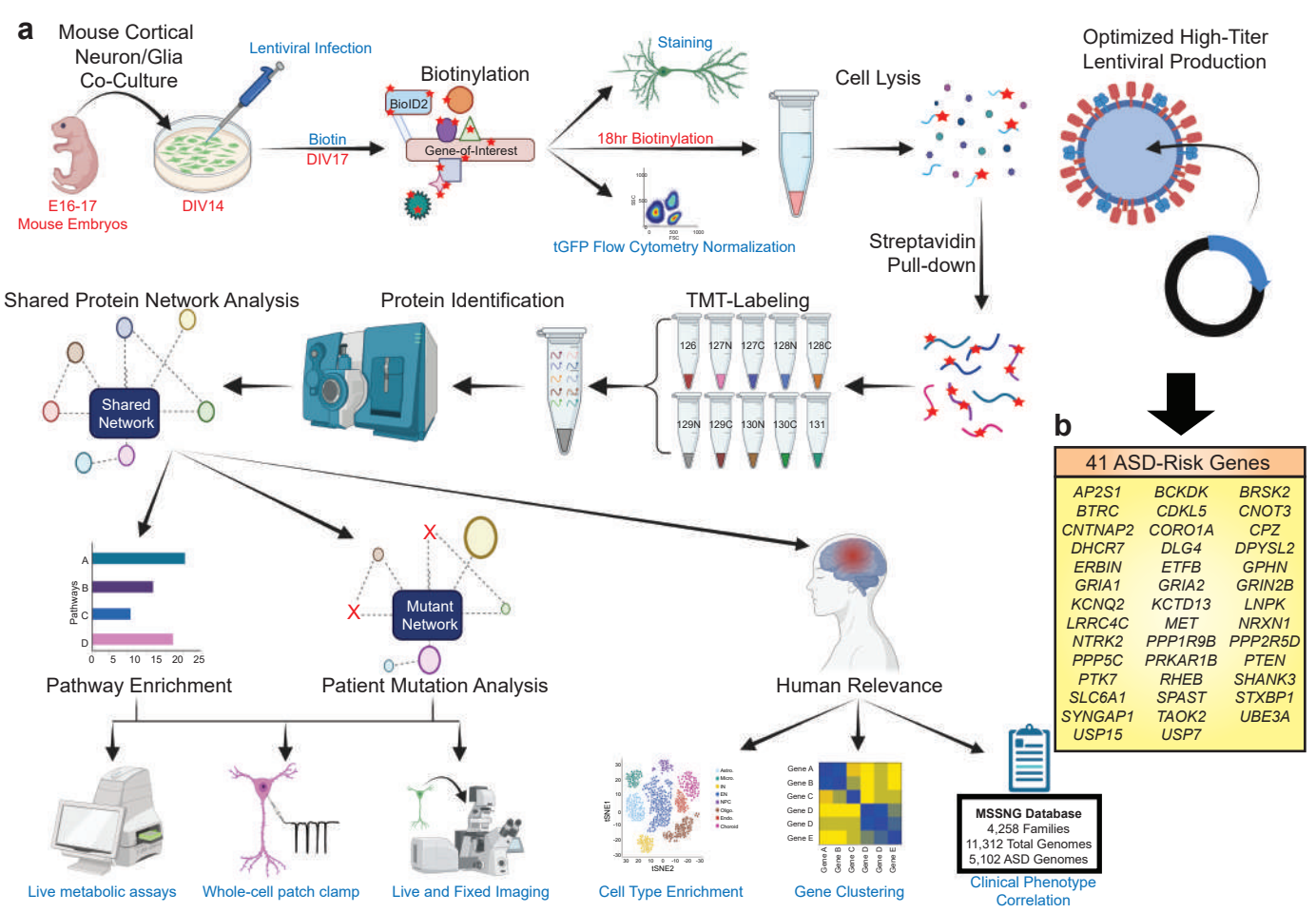


Figure 1. Development of a proximity-based proteomics screen to identify neuronal PPI networks for ASD-risk genes

a, Workflow of neuron-specific BioID2 screen for identification of ASD-risk gene PPI networks. **b,** List of 41 ASD-risk gene used in the BioID2 screen

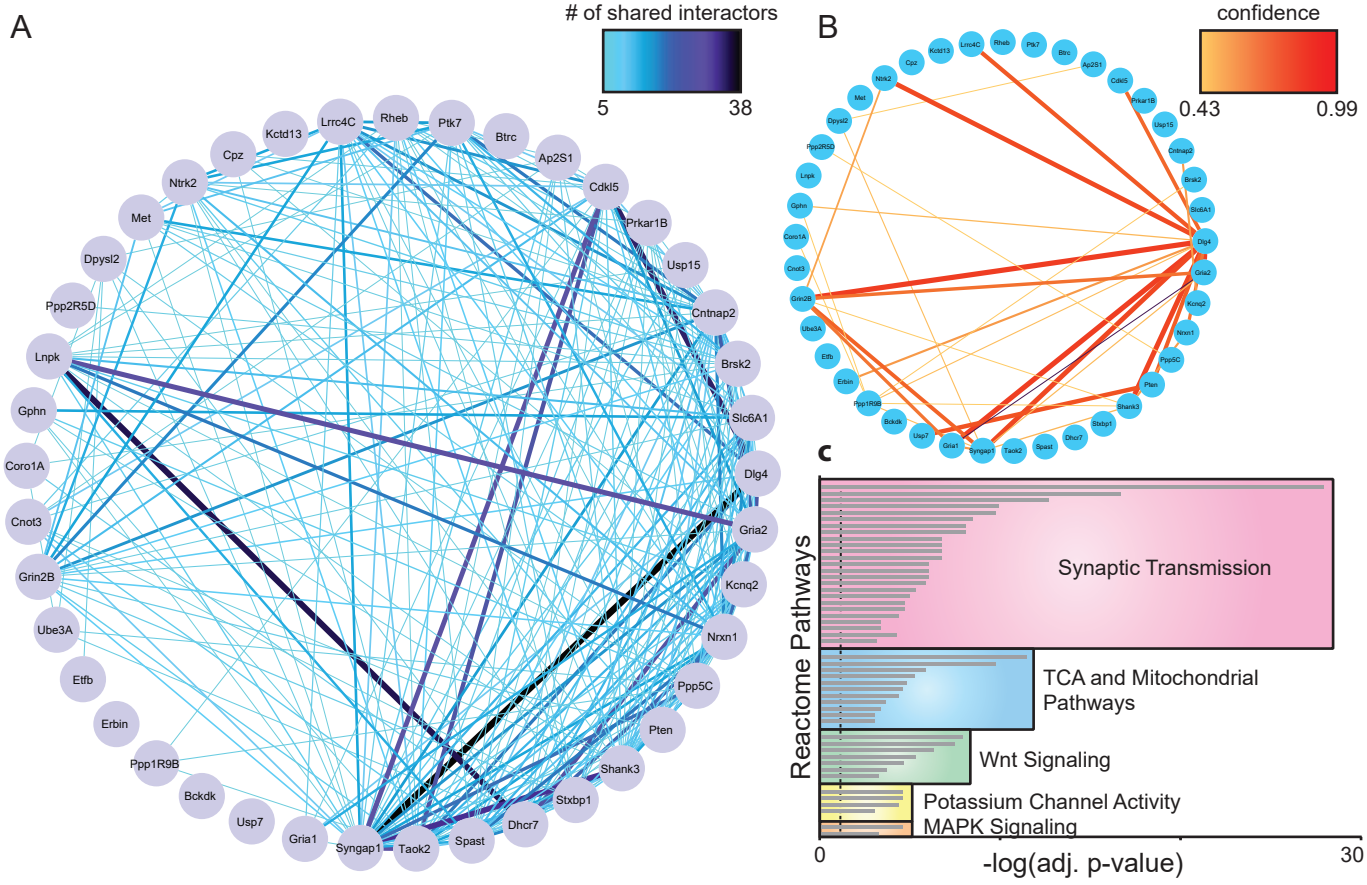


Figure 2. Neuron-specific PPI network map identifies convergent signaling pathways between 41 ASD-risk genes
a, Shared PPI network map of 41 ASD-risk bait genes. Large red nodes represent bait proteins. Color and increased thickness of connecting lines represent number of interactions (direct or shared prey proteins) between bait genes. **b,** Known physical interactions of the 41 ASD-risk genes from the STRING database. Color and increasing thickness of the line represents the confidence of the interaction starting at medium interaction confidence (0.4) **c,** Top 50 enriched Reactome pathways of the shared 41 ASD-risk gene PPI network map. Individual pathways are grouped by functional similarity (g:Profiler, Bejamini-Hochberg FDR adj. p<0.05).

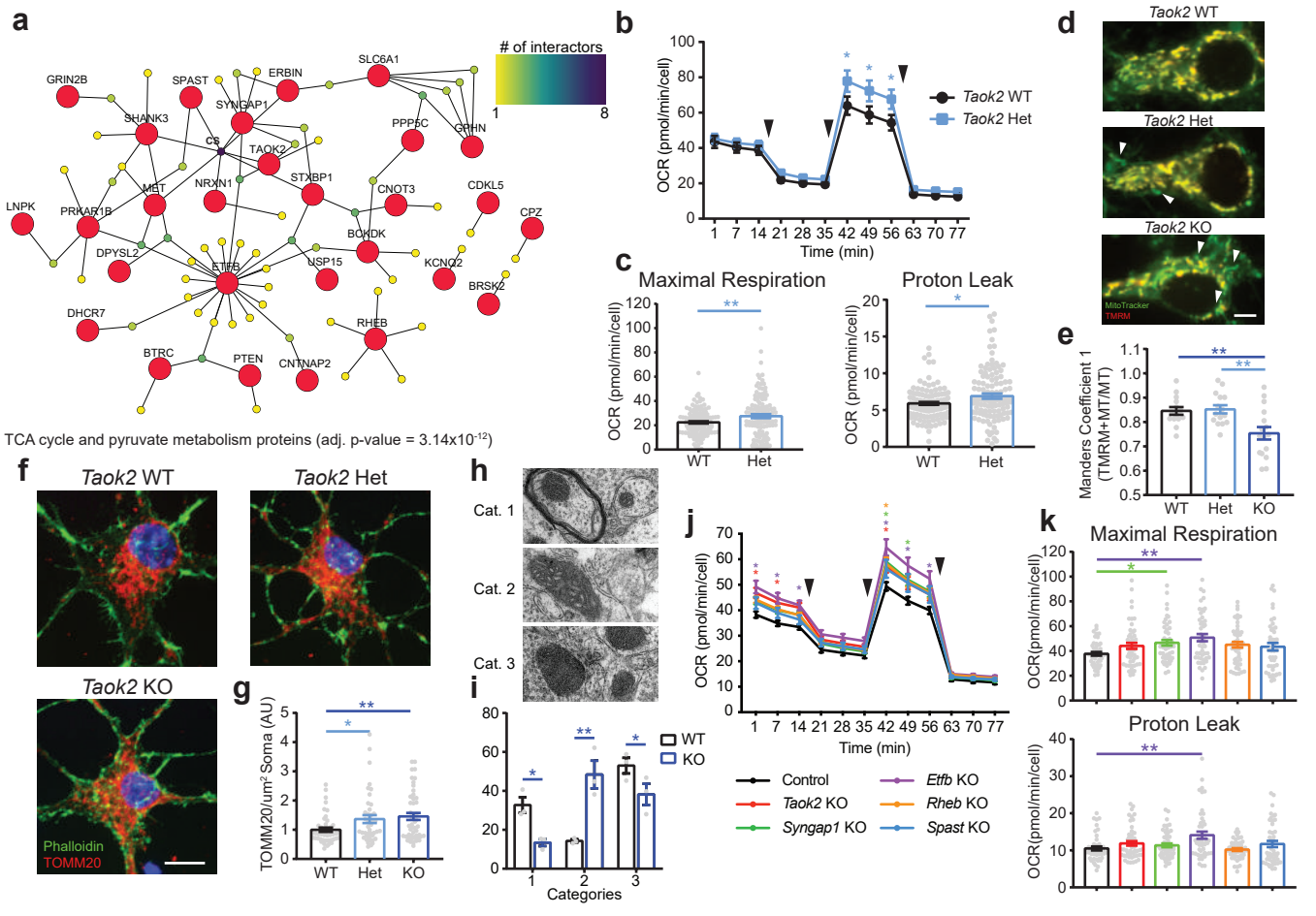


Figure 3. Regulation of cellular respiration and mitochondrial activity by ASD-risk genes

a, Network of TCA cycle and pyruvate metabolism proteins enriched in the shared ASD-risk gene PPI network map (g:Profiler, Benjamini-Hochberg FDR adj. p<0.05). **b,** Loss of *Taok2* alters cellular respiration in DIV14 mouse cortical neurons (Two-Way ANOVA, F(1,336) = 18.22, p<0.0001 between genotypes, *post hoc* Holm-Sidak test; WT = 118 wells and Het = 122 wells from three separate cultures). **c,** Loss of Taok2 increases maximal respiration (*left*) and proton leakage (*right*) in DIV14 mouse cortical neurons (ROUT outlier test, Q = 0.1%; Two-tailed unpaired t-test, maximal respiration: t = 3.015, df = 233, p = 0.0029, proton leak: t = 2.374, df = 225, p = 0.0184; WT = 118 wells, Het = 122 wells from three separate cultures). **d,** Representative images of *Taok2* WT, Het, and KO neurons stained with MitoTrack (green) and TMRM (red). Arrows indicate Mitotracker-labeled mitochondria with no TMRM staining. Scale bar is 5 µm. **e,** *Taok2* Het and KO neurons have decreased number of active (TMRM) Mitotracker-labeled mitochondria (One-Way ANOVA, F(2, 42) = 7.47, p = 0.0017, *post hoc* Holm-Sidak test; WT = 14, Het = 15, KO = 15 neurons from 1-3 separate pups per genotype from one culture). **f,** Representative images of Taok2 WT, Het, and KO neurons stained with TOMM20 (red) and Phalloidin (green). Scale bar 5 µm. **g,** *Taok2* Het and KO neurons have increased TOMM20 staining (One-Way ANOVA, F(2,132) = 4.633, p = 0.0111; WT = 45, Het = 42, KO = 48 neurons from 1-4 separate pups per genotype from two cultures). **h**, Representative images of synaptic mitochondria morphological categories. **i**, *Taok2* KO neurons have increased proportion of category 2 mitochondria with enlarged cristae. (Two-way ANOVA, F(1, 12) < 0.0001, p = 0.9998 between genotypes, *post hoc* Holm-Sidak test. 19-25 images per animal and three animals per genotype). **j,** CRISPR/Cas9 KO of *Taok2*, *Syngap1*, *Etfb*, *Rheb*, and *Spast* differentially alters cellular respiration in DIV14 mouse cortical neurons (Two-Way ANOVA, F(5,280) = 3.492, p = 0.0044, *post hoc* Holm-Sidak test; Control = 45 wells, *Taok2* KO = 51 wells, *Syngap1* KO = 50 wells, *Etfb* KO = 47 wells, *Rheb* KO = 48 wells, *Spast* KO = 45 wells from five separate cultures). **k,** Significant increases in maximal respiration (*top*) and proton leakage (*bottom*) show signficant increase in *Syngap1* and *Etfb* KO neurons (ROUT outlier test, Q = 0.1%; maximal respiration: One-Way ANOVA, F(5, 280) = 2.927, p = 0.0136, *post hoc* Holm-Sidak test, proton leak: One-Way ANOVA, F(5, 277) = 4.138, p = 0.0012, *post hoc* Holm-Sidak test). Mean ± s.e.m. *p<0.05, **p<0.01, ***p<0.001.

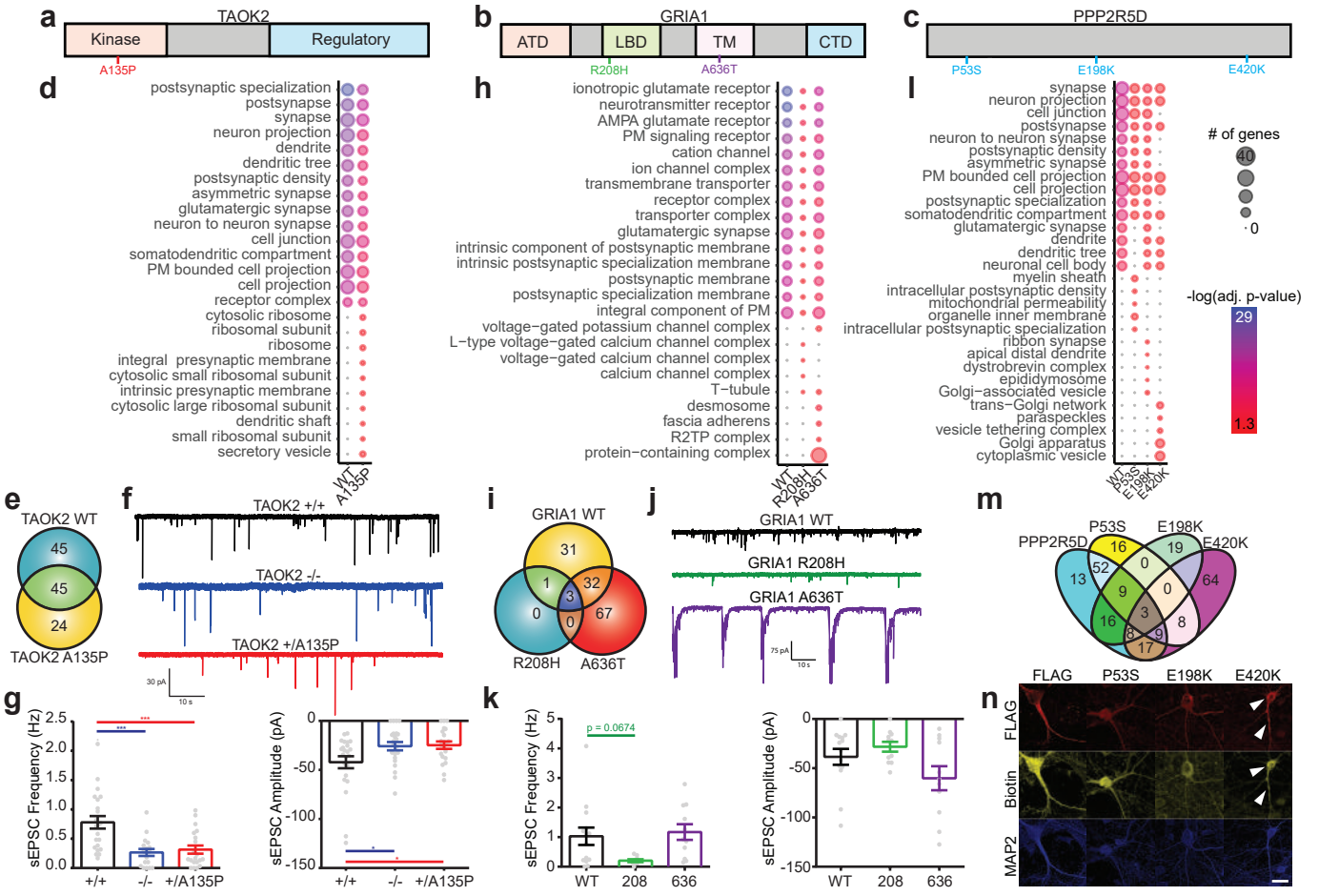


Figure 4. De novo missense variants alter the PPI networks of uncharacterized and established ASD-risk genes that correspond to functional deficits. a-c, Diagram of TAOK2, GRIA1, and PPP2R5D and location of *de novo* missense variants. ATD: amino-terminal domain, LDB: ligand-binding domain, TM: transmembrane domain, CTD: carboxy-terminal domain. **d,** Dot plot of top 15 cellular compartment gene sets and top 10 variant-specifc gene sets identified in TAOK2 variants (g:Profiler, Benjamini-Hochberg FDR adj. p<0.05). PM: plasma mebrane. Size of dots indicate protein number and the color represents the significance. **e,** Venn diagram of PPI network proteins of TAOK2 WT and A135P. **f,** Representative traces of sEPSC recordings of DIV21 *TAOK2* WT, KO, and A135P human iPSC-derived NGN2 neurons. **g,** *TAOK2* KO and A135P neurons show decreased sEPSC fequency (*left*) and amplitude (*right*) (ROUT Outlier Test, Q = 0.1%; frequency: One-Way ANOVA, F(2,57) = 11.63, p<0.0001, amplitude: One-Way ANOVA, F(2,63) = 4.027, p = 0.0226, *post hoc* Holm- Sidak test; WT = 23, KO = 22, and A135P = 21 neurons from three separate transductions). **h,** Dot plot of top 15 cellular compartment gene sets and top five variant-specific gene sets identified in GRIA1 variants (g:Profiler, Benjamini-Hochberg FDR adj. p<0.05). **i,** Venn diagram of PPI network proteins of GRIA1 WT and variants. **j,** Representative traces of sEPSC recordings of mouse cortical neurons expressing GRIA1 or variants. **k,** R208H variant shows trend in decrease sEPSC frequency *left*) and no change in amplitude (*right*). (ROUT Outlier Test, Q = 0.1%; frequency: One-Way ANOVA, F(2,31) = 3.506, p = 0.0424, amplitude: One-Way ANOVA, F(2,33) = 3.147, p = 0.0561, *post hoc* Holm-Sidak test; WT = 14, R208H = 11, and A636T= 11 neurons from three separate infections) Dot plot of top 15 cellular compartment gene sets and top five variant-specific gene sets identified in PP2R5D variants (g:Profiler, Benjamin Hochberg FDR adj. p<0.05). **m,** Venn diagram of PPI network proteins of PPP2R5D WT and variants. **n,** Representative images of neurons infected with PPP2R5D WT and variants show mislocalization of E420K variant (white arrows). Scale bar is 25μm. Mean ± s.e.m. *p<0.05, **p<0.01, ***p<0.001.

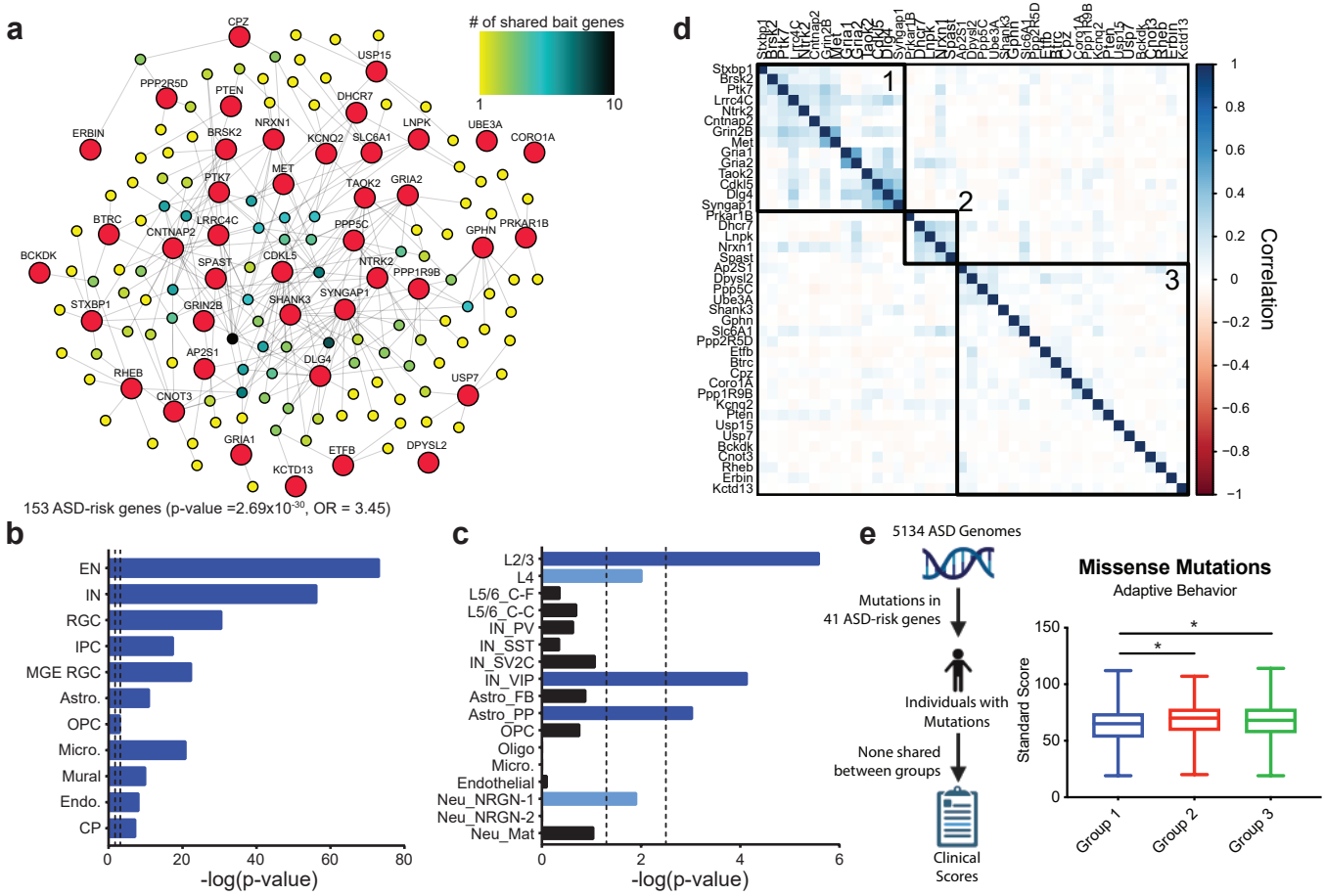
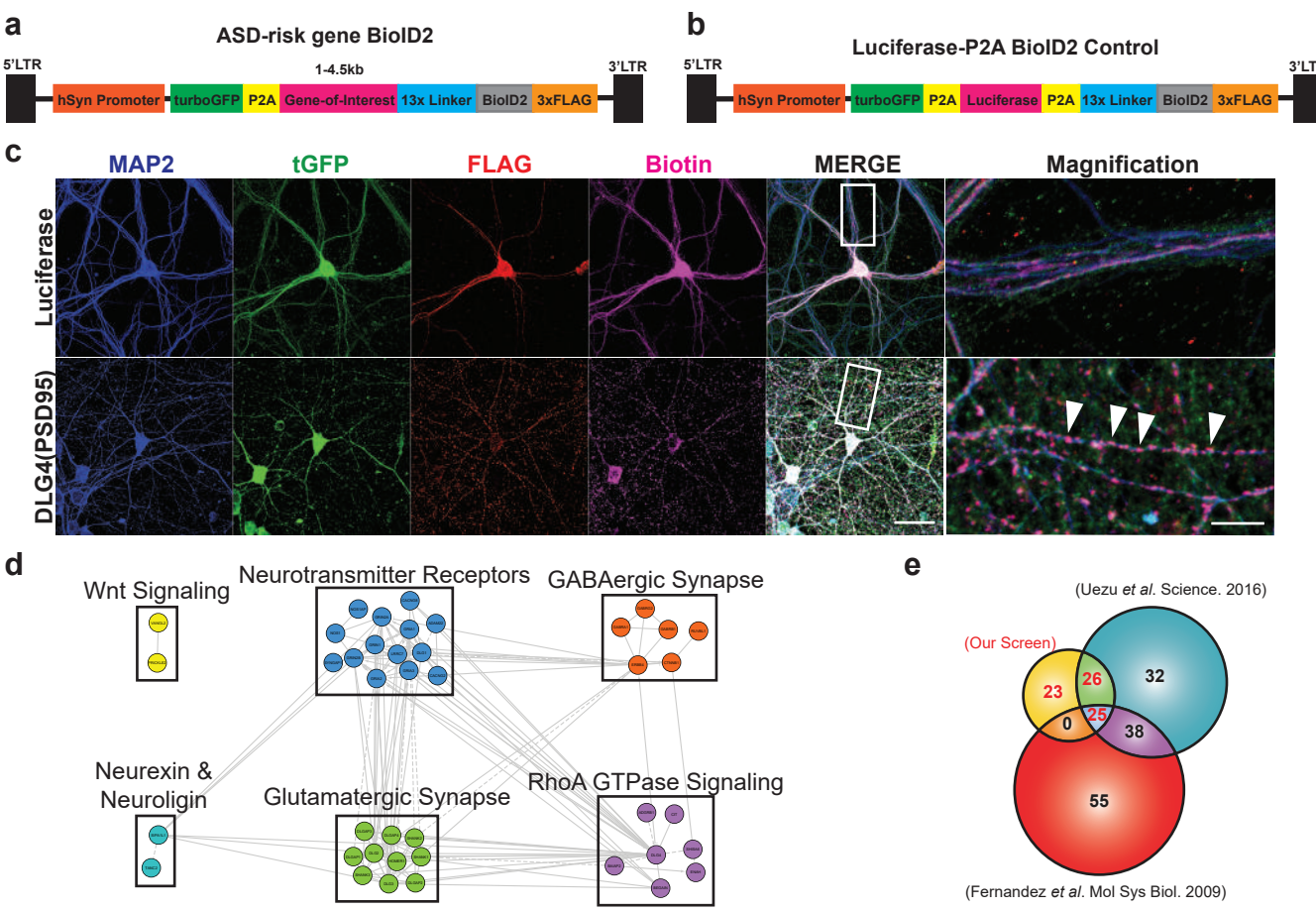


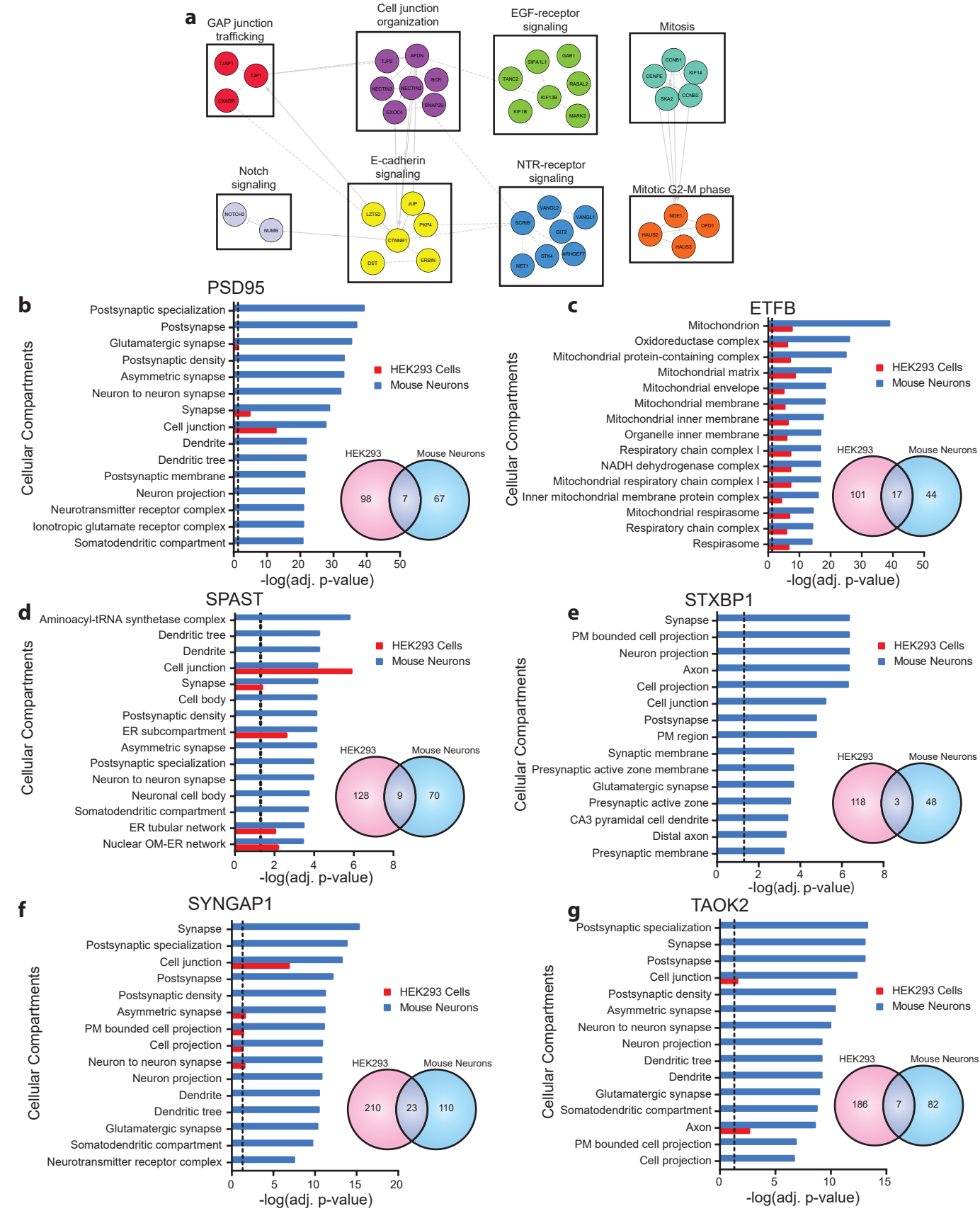
Figure 5. Shared ASD PPI network map network correlates to human brain development and disease pathology

a, Network of ASD-risk genes enriched in the shared ASD-risk gene PPI network. Large red nodes represent bait genes, while smaller colored nodes represent sharedness of ASD-risk genes between bait genes, between 1 (yellow) and 10 (blue) shared bait genes (g:Profiler, Benjamini-Hochberg FDR adj. p<0.05). **b,** The shared ASD-risk gene PPI network enriches for human neuron cell types (Fisher's exact test). Dashed lines represent nominal (p = 0.05, *left*) and Bonferroni corrected (p = 0.05/number of cell types, *right*) significance thresholds. **c,** Enrichment of ASD differentially expressed genes (DEGs) in specific cell types. (Fisher's exact test). Dashed lines represent nominal (p = 0.05, *left*) and Bonferroni corrected (p = 0.05/number of cell types, *right*) significance thresholds. EN = excitatory neurons, IN = inhibitory neurons, RGC = radial glial cells, MGE RGC = medial ganglionic eminence, IPC = intermediate progenitor cells, Astro. = astrocyte, OPC = oligodendrocyte progenitor cells, Micro. = microglia, Endo. = endothelial cells, CP = choroid plexus cells, C-F = cortico-fugal, C-C = cortico-cortico, PV = paravalbumin, SST = somatostatin, VIP = vasoactive intestinal peptide, FB = fibrous, PP = protoplasmic, Neu_NRGN = neurogranin-expressing. Light blue bars have nominal p-value significance, while dark blue bars have Bonferroni corrected signficance. **d,** Correlation plot of 41 ASD-risk genes through individual PPI networks. Genes were ordered by hierarchical clustering and clustered using kmeans (k = 3). **e,** Significant decrease in the average standard scores of individuals diagnosed with ASD, who have rare inherited missense mutations in Group 1 genes compared to Groups 2 and 3 (Non-parametric Kruskal-Wallis test, p = 0.0103, *post hoc* Dunn's test, Group 1 = 350, Group 2 = 113, and Group 3 = 416 probands). Box and whisker plot (minimum,1st quartile, median, 3rd quartile, maximum).

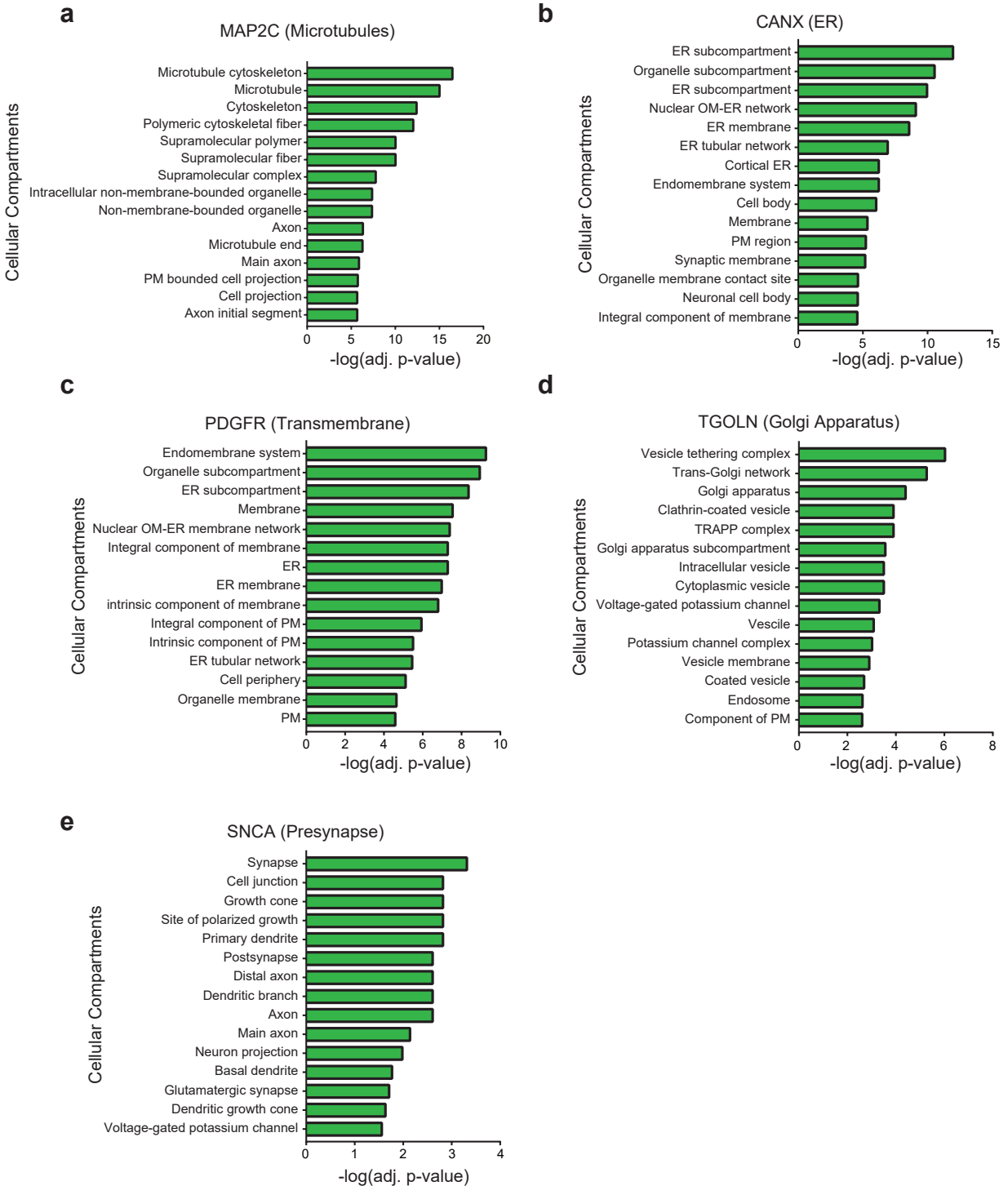


Extended Figure 1. BioID2 of DLG4 in mouse cortical neurons

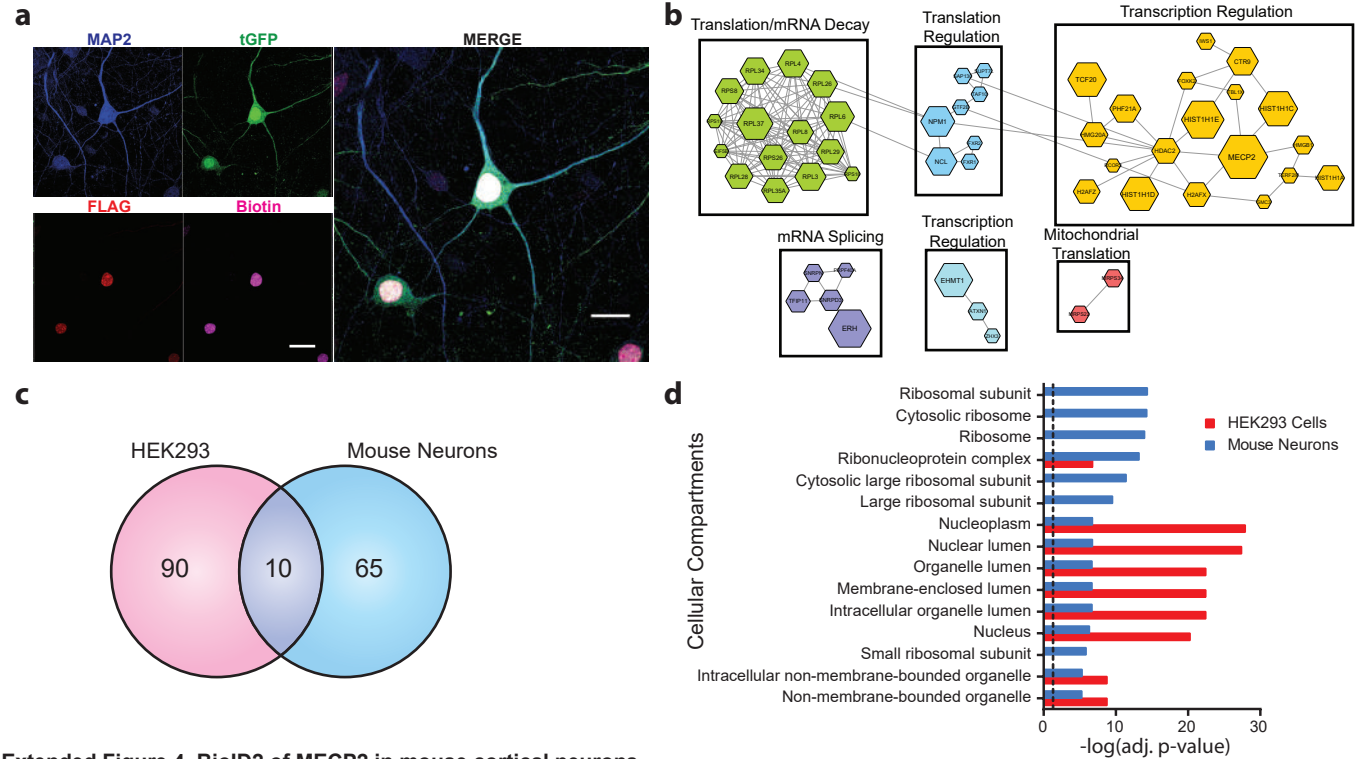
a, Diagram of the BioID2 fusion construct for the 41 ASD-risk genes. **b,** Diagram of the control Luciferase control fusion construct **c,** Representative images of cortical neurons infected with the PSD95-BioID2 and Luciferase-P2A-BioID2 constructs. Scale bar is 20μm. Magnified images are shown on the right. White arrows point to synaptic localization of PSD95-BioID2. Scale bar is 5μm. **d,** Reactome pathways enriched in the PSD95 PPI network. Clusters created using the Reactome FI app on Cytoscape and labeled with most signifcantly enriched pathways for each cluster (adj. p<0.05). **e,** Venn diagram of shared protein interactors between our *in vitro* PSD95 PPI network and proteins identified by mouse *in vivo* PSD95 BioID (Uezu *et al.* 2016) or mouse *in vivo* tandem affinity purification of PSD95 (Fernandez *et al.* 2009).



Extended Figure 2. Comparison of ASD-risk gene PPI networks from BioID2 in HEK293 cells and mouse cortical neurons
a, Reactome pathways enriched in the PSD95 PPI network from HEK293 cells. Clusters created using the Reactome FI app on Cytoscape and labeled with the most significantly enriched pathways (adj. p<0.05). **b-g,** Top 15 cellular compartments from mouse cortical neurons (blue) enriched in the PPI networks of PSD95, ETFB, SPAST, STXBP1, SYNGAP1, and TAOK2 compared to enrichment in HEK293 cells (red) (g:Profiler, Benjamini-Hochberg FDR adj. p<0.05). Adjacent Venn diagrams show shared protein interactors identified by BioID2 in HEK293 cells vs mouse cortical neurons. PM: Plasma membrane, OR: Outer membrane, ER: endoplasmic reticulum.

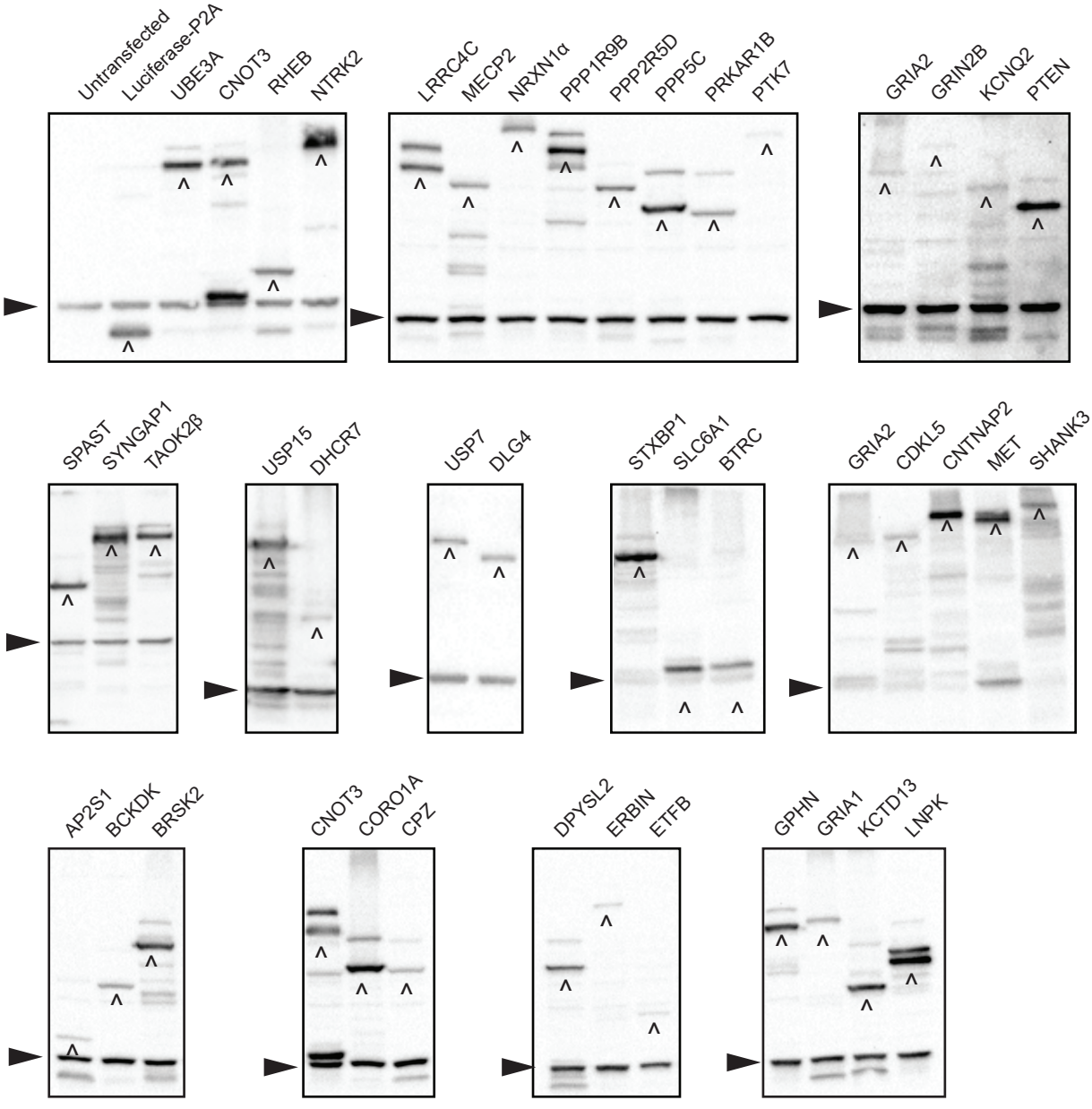


Extended Figure 3. Validaton of BioID2 using genes localized to specific compartments in mouse cortical neurons
BioID2 of cellular compartment proteins MAP2C (**a**), CANX (**b**), PDGFR (transmembrane domain) (**c**), TGOLN (**d**), and SNCA (**e**). g:Profiler pathway enrichment was used to identify significantly enriched cellular compartments (g:Profiler, Benjamini-Hochberg FDR adj. p<0.05). PM: plasma membrane, OM: outer membrane, ER: endoplasmic reticulum.

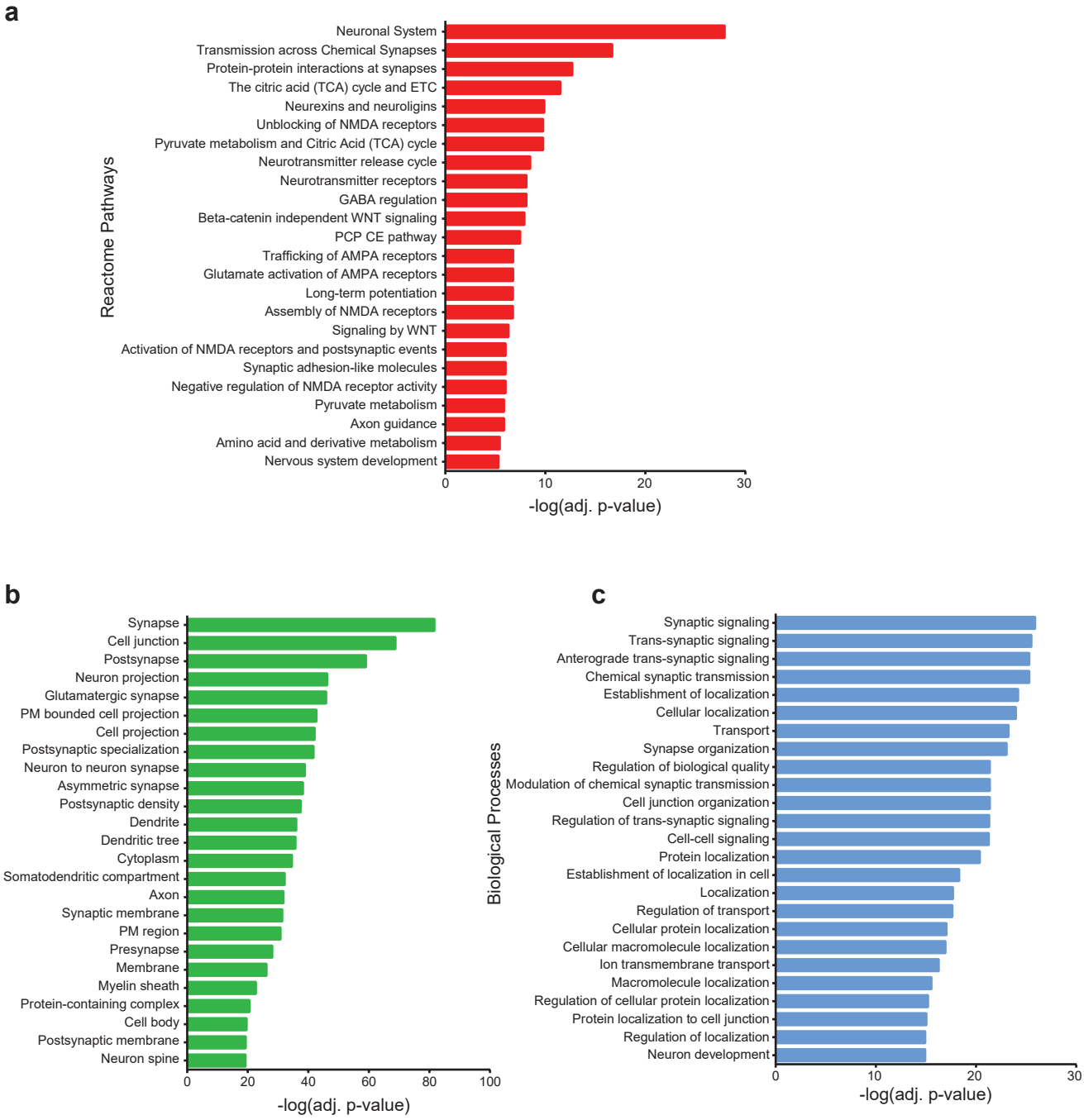


Extended Figure 4. BioID2 of MECP2 in mouse cortical neurons

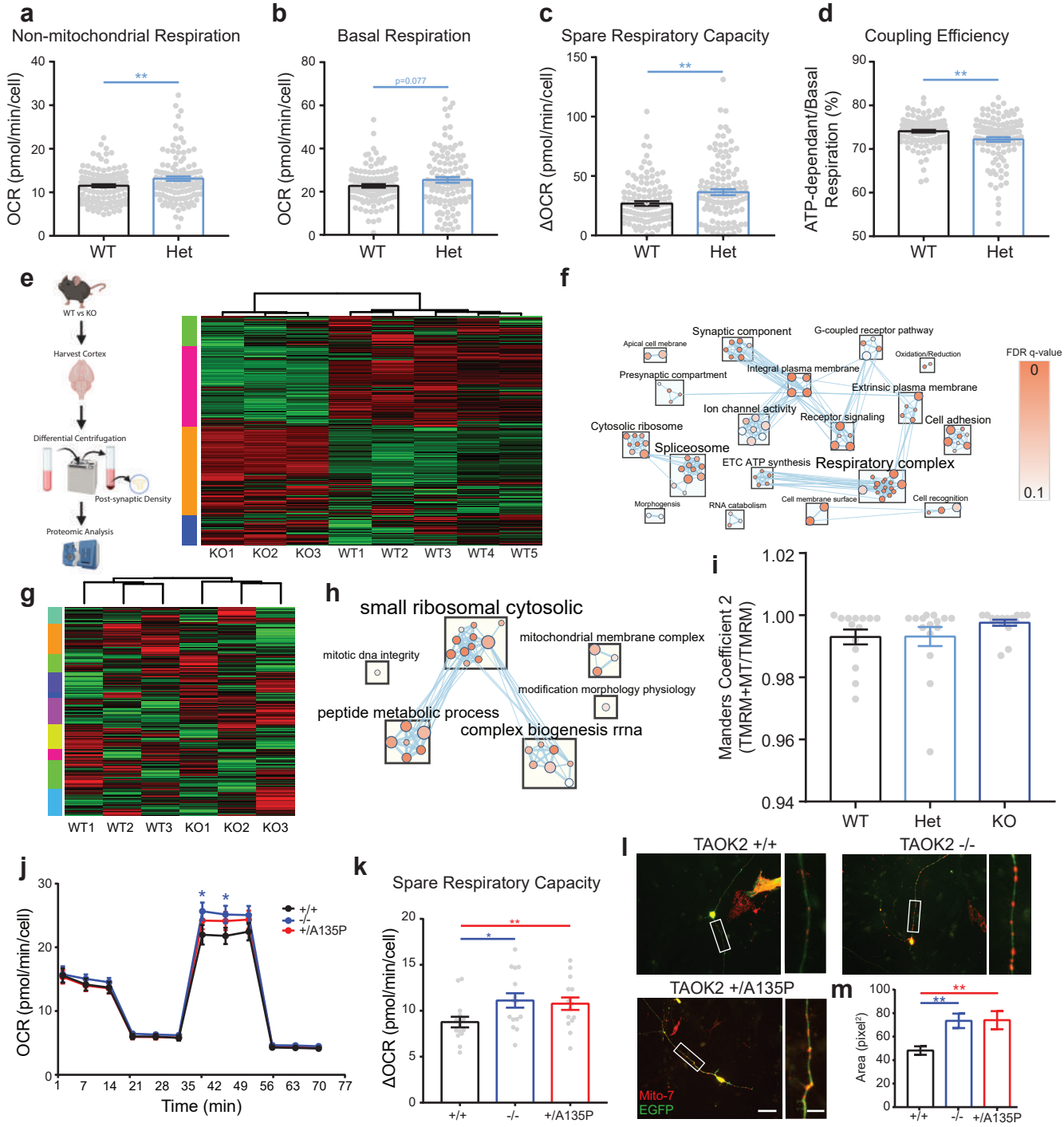
a, Representative images of cortical neurons infected with the MECP2-BioID2 construct. Scale bar is 20µm. **b,** Reactome pathways enriched in the MECP2 PPI network. Clusters created using the Reactome FI app on Cytoscape and labeled with most signifcantly enriched pathways (adj. p<0.05). **c,** Venn diagram shows shared protein interactors identified by BioID2 in HEK293 cells vs mouse cortical neurons. **d,** Top 15 cellular compartments from mouse cortical neurons (blue) enriched in the MECP2 PPI network compared to enrichment in HEK293 cells (red) (g:Profiler, Benjamini-Hochberg FDR adj. p<0.05).



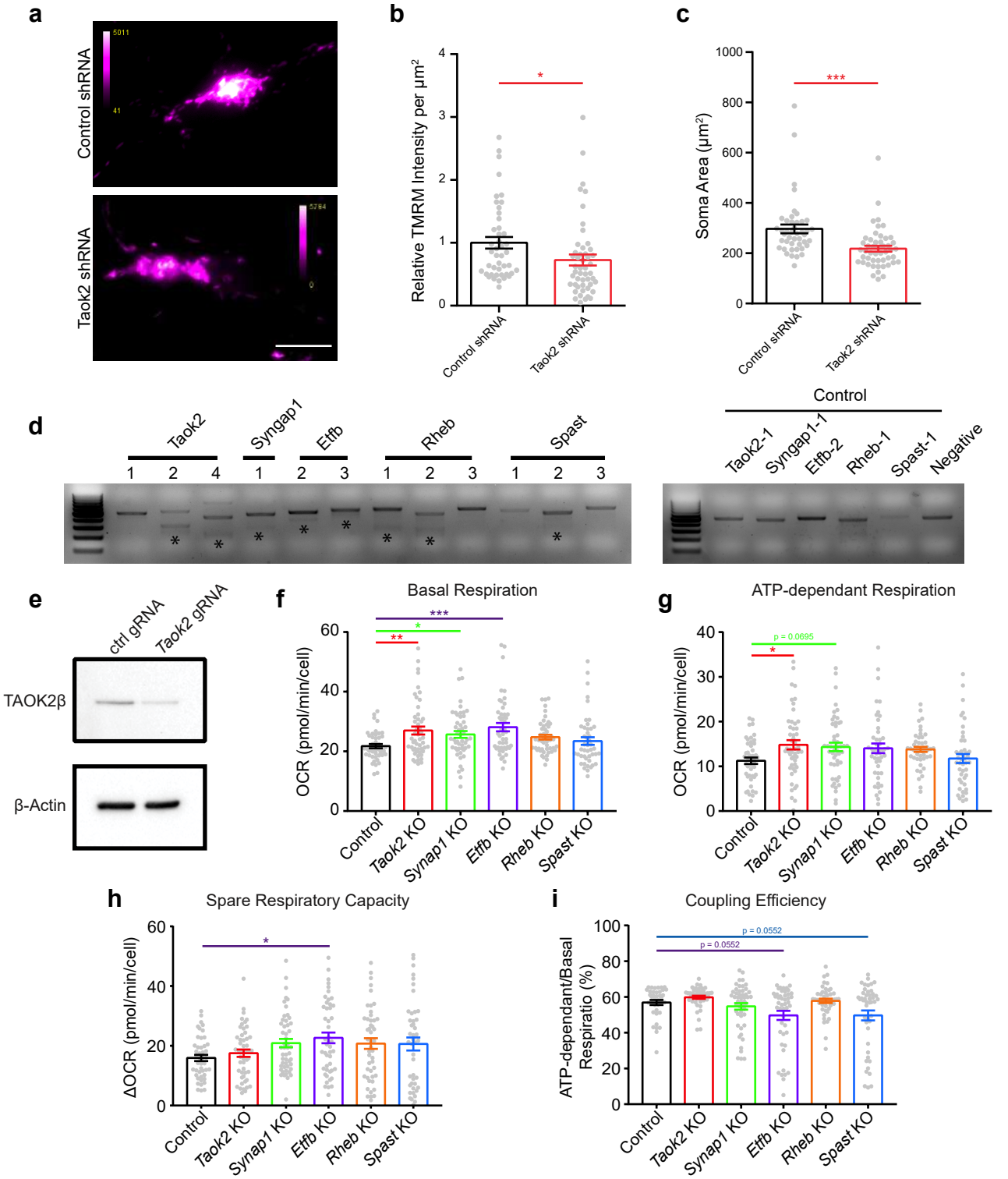
Extended Figure 5. Western blots of 41 ASD-risk gene BioID2 constructs
Western blots of ASD-risk gene BioID2 constructs transfected in HEK293 cells and immunoblotted for FLAG and β-actin as the loading control. ^ denotes expected BioID2 fusion protein size. Bands higher than the caret indicated bands are possible tGFP fusion proteins due to P2A inefficiency. Arrow denotes β-actin loading control. Bands lower than β-actin are possible degraded BioID2-FLAG proteins.



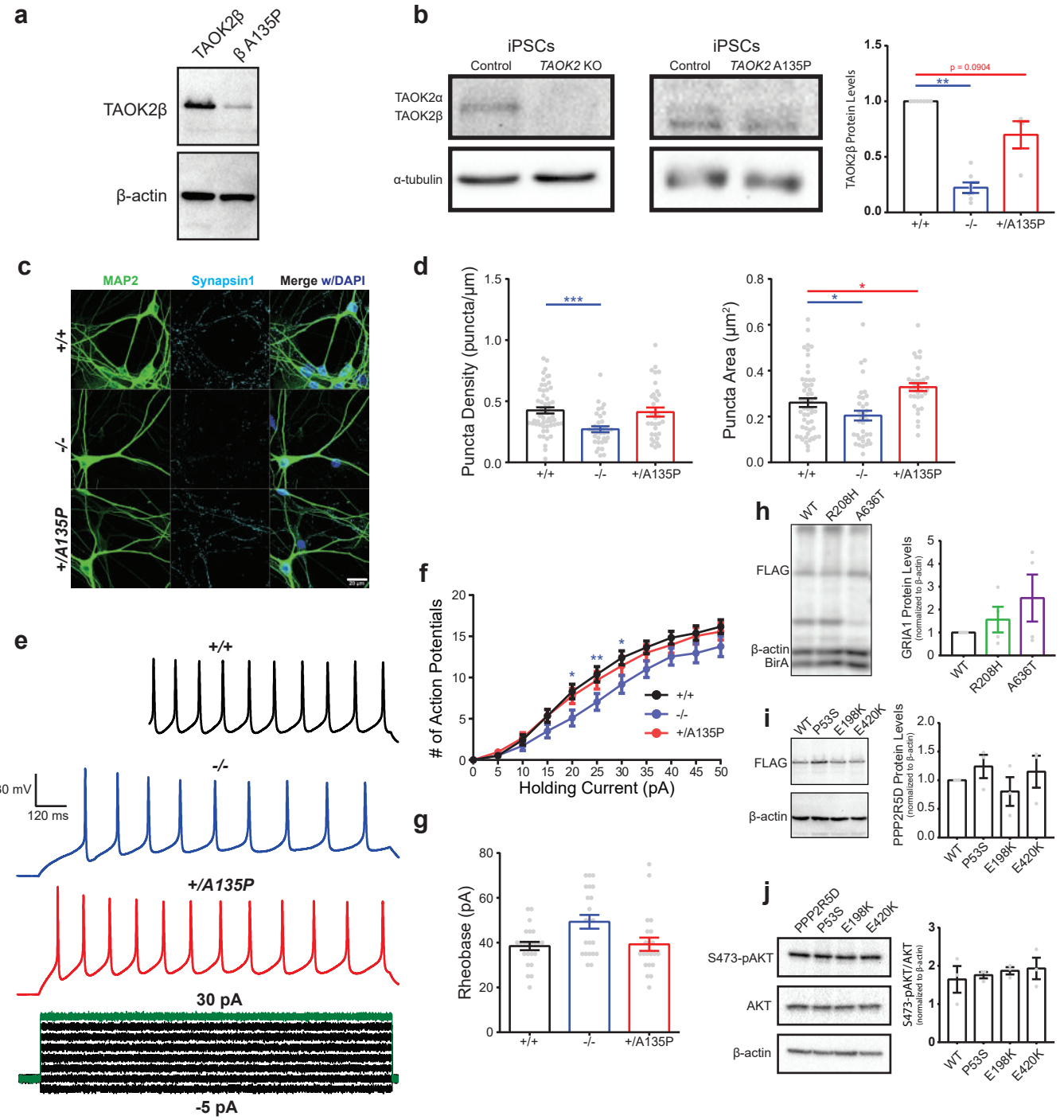
Extended Figure 6. Enriched pathways in the shared ASD-risk gene PPI network map identified using BioID2
Pathway enrichment was used to identify significantly enriched Reactome pathways (**a**), cellular compartments (**b**), and biological processes (**c**) (g:Profiler, Benjamini-Hochberg FDR adj. p<0.05). The top 25 pathways are shown for each graph. PM: plasma membrane, OM: outer membrane, ER: endoplasmic reticulum.



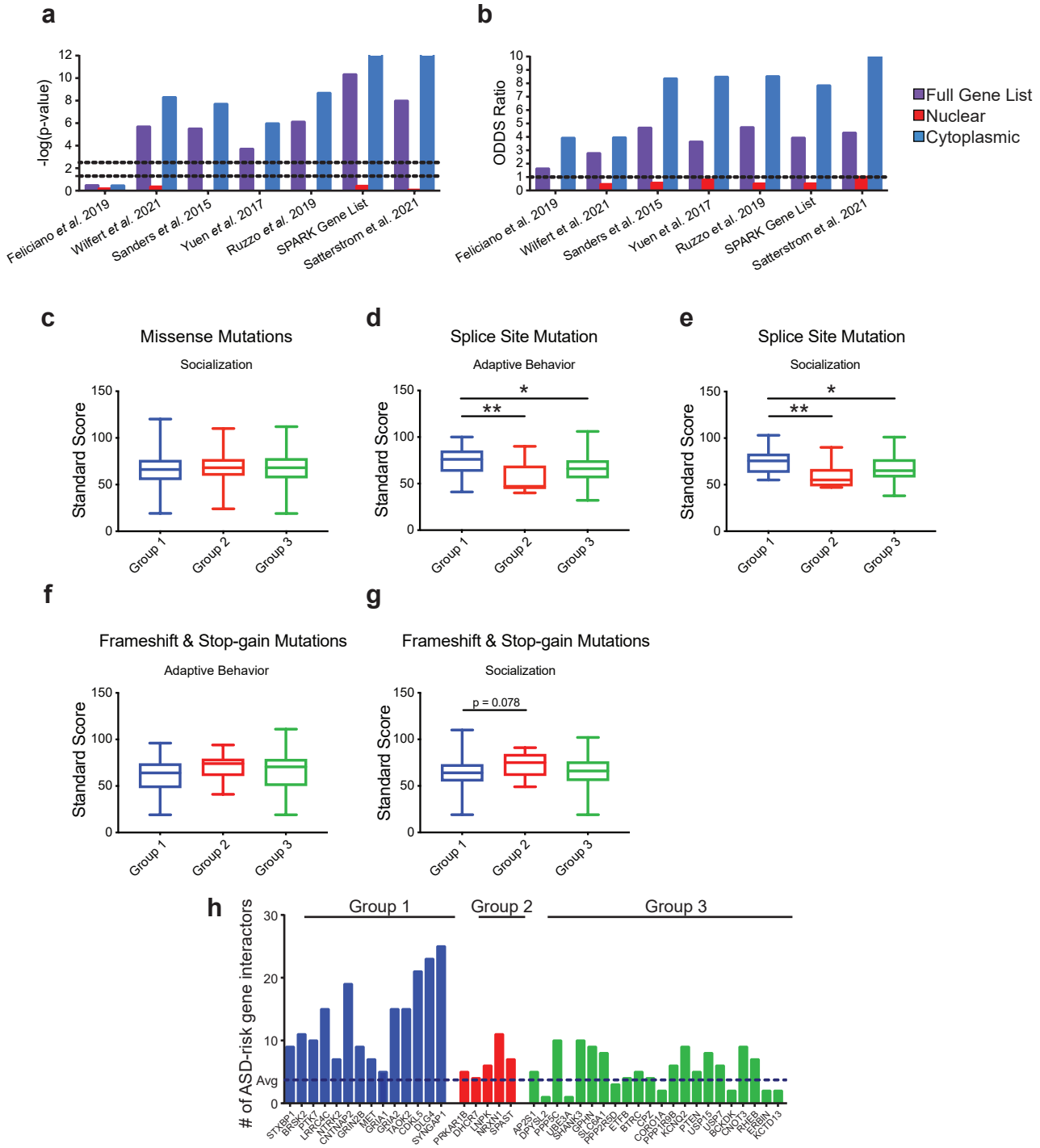
Extended Figure 7. Loss or disruption of *TAOK2* causes alterations in cellular respiration and mitochondrial proteins.
a-d, *Taok2* Het neurons have increased non-mitochondrial respiration and spare respiratory capacity, and decreased coupling efficiency (ROUT outlier test, Q = 0.1%; Two-tailed unpaired t-test, non- mitochondrial respiration: t = 2.668, df = 228, p = 0.0082, basal respiration: t = 1.776, df = 227, p = 0.0770, spare respiratory capacity: t = 3.007, df = 234, p = 0.0029, coupling efficiency: t = 3.146, df = 220, p = 0.0019; WT = 118 wells, Het = 122 wells from three separate cultures). **e,** Shotgun proteomics of post-synaptic density fraction from *Taok2* WT and KO mouse cortices. **f,** *Taok2* KO PSD fractions have significant decrease in synaptic and mitochondrial protein gene sets (GSEA, FDR<0.1.; five *Taok2* WT and three *Taok2* KO mice littermates). Size of nodes represents number of proteins and color represents FDR q-value. **f,** RNA sequencing of *Taok2* WT and KO mouse cortices. **g-h,** *TAOK2* KO mouse cortices have altered mRNA levels of mitochondrial membrane proteins. (GSEA, FDR<0.1, three *Taok2* WT and KO mice littermates each). Size of nodes represents number of proteins and color represents FDR q-value. **i,** All active mitochondria stained by MitoTracker (One-Way ANOVA, F(2, 42) = 1.355, p = 0.2689, *post hoc* Holm-Sidak test; WT = 14, Het = 15, KO = 15 neurons from three separate cultures). **j-k,** DIV7 *TAOK2* KO human neurons have signficantly increased maximal respiration (Two-Way ANOVA, F(2,42) = 0.659, p = 0.5226 between genotypes, *post hoc* Holm-Sidak test; WT = 118 wells and Het = 122 wells from three separate cultures). *TAOK2* KO and A135P neurons have significantly increased spare respiratory capacity (One-Way ANOVA, F(2,42) = 3.409, p = 0.0424, *post hoc* Holm-Sidak test; WT = 118 wells, Het = 122 wells from three separate cultures). **l,** Representative images of human neurons transfected with Mito7-DsRed constructs at DIV7 and fixed and imaged at DIV9. Scale bar is 20 µm. Magnification of boxed areas shown on the right. Scale bar is 5 µm. **m,** *TAOK2* KO and A135P neurons have larger Mito7-DsRed punctae size compared to wildtype neurons (One-Way ANOVA, F(2, 1309) = 5.032, p = 0.0067, *post hoc* Holm-Sidak test; WT = 371, KO = 520, A135P = 421 punctae from 15-16 neurons per geneotype). Mean ± s.e.m. *p<0.05, **p<0.01.



Extended Figure 8. Acute knockout of ASD-risk genes in mouse cortical neurons alters mitochondrial activity and cellular respiration
a, Representative images of mouse cortical neurons infected with control shRNA or *Taok2* shRNA showing TMRM intensity. Scale bar is 10µm. **b,** Mouse cortical neurons with acute knockout of *Taok2* have decreased relative TMRM activity, even with decreased soma size (**c**) (TMRM activity: Unpaired t-test, t = 2.156, df = 91, p = 0.0337; WT = 44, Taok2 KD = 49 neurons from two separate cultures). **d,** GeneArt cleavage assay kit shows indel insertion in at least 1 gRNA target region in mouse cortical neurons based on presence of multiple bands. *Indicates secondary band due to digested indel. **e,** Reduced Taok2 protein expression in mouse cortical neurons infected with Cas9 and *Taok2* KO gRNAs. Infected at DIV7 and harvested at DIV18 for western blot of TAOK2β. **f-i,** Significant changes in different aspects of cellular respiration (Basal respiration (**f**),ATP-dependent respiration (**g**), spare respiratory capacity (**h**), and coupling efficiency (**i**) in mouse cortical neurons with CRISPR/Cas9 KO of *Taok2*, *Syngap1*,or *Etfb* (ROUT Outlier Test, Q = 0.1%; basal respiration: One-Way ANOVA, F(5,279) = 3.994, p = 0.0016, ATP-dependent respiration: One-Way ANOVA, F(5,279) = 2.517, p = 0.0300, spare respiratory capacity: One-Way ANOVA, F(5,276) = 2.381, p = 0.0389, coupling efficiency: One-Way ANOVA, F(5,263) = 4.894, p = 0.0003, *post hoc* Holm-Sidak test; *Taok2* KO = 51 wells, *Syngap1* KO = 50 wells, *Etfb* KO = 47 wells, *Rheb* KO = 48 wells, *Spast* KO = 45 wells from five separate cultures). Mean ± s.e.m. *p<0.05, **p<0.01, ***p<0.001.



Extended Figure 9. ***De novo*** **mutations in TAOK2 caused altered synaptic transmision and neuron firing. a,** Western blot of TAOK2 WT and A135P BioID2 constructs. **b**, Western blot of CRISPR/Cas9-edited iPSCs and neurons showing loss of TAOK2 expression in TAOK2 KO (-/-) and A135P (+/A135P) lines (One-sample t-test, WT vs KO: t = 16, df = 6, p<0.0001, WT vs A135P: t = 2.465, df = 3, p < 0.0904; WT = 7, KO = 7, A135P = 4 wells from separate iPSC cultures). **c,** Representative images of TAOK2 WT (+/+), KO, and A135P human neurons, stained with MAP2, Synapsin 1, and DAPI 21 days after NGN2 induction. **d,** Reduced synapsin punctae density and size in TAOK2 KO neurons (*Left*) and increased synapsin puncta size in TAOK2 A135P neurons (*Right*) (ROUT outlier test, Q = 0.1%, One-Way ANOVA, density: F(2,120) = 8.104, p = 0.0005, area: F(2, 119) = 8.207, p = 0.0005, post hoc Holm-Sidak test; WT = 55, KO = 35, A135P = 34 neurons from five separate transductions). **e,** Representative traces of repetitive firing (*top*) and current injection (*bottom*). **f**, TAOK2 KO neurons have reduced repetitive firing (Two-Way ANOVA, F(2, 693) = 17.6, p<0.0001 between genotypes; WT = 23, KO = 22, A135P = 21 neurons from 3 separate transductions) increased rheobase (**g**) (One-Way ANOVA, F(2, 63) = 5.229, p = 0.0079; WT = 23, KO = 22, A135P = 21 neurons from 3 separate transductions). **h,** Representative western blot of GRIA1 WT, R208H, and A636T BioID2 constructs (*left*) and quantification (*right*) showing no significant difference (One-sample t-test, WT vs R208H: t = 1.008, df = 3, p = 0.3877, WT vs A636T: t = 1.466, df = 3, p = 0.2388; four separate transfections). **i,** Representative western blot of PPP2R5D WT, P53S, E198K, and E420K BioID2 constructs (*left*) and quantification (*right*) showing no difference in expression (One-sample t-test, WT vs P53S: t = 1.185, df = 2, p = 0.3577, WT vs E198K: t = 0.7838, df = 2, p = 0.5152, WT vs E420K: t = 0.5371, df = 2, p = 0.6449; three separate transfections). **j,** Representative western blot of p-AKT and AKT in HEK293 cells expressing PPP2R5D WT or variant constructs (*left*) and quantification (*right*) showing no difference in the pAK/AKT ratio (One-way ANOVA, F(3,8) = 0.2948, p = 0.8283; three separate transfections). Mean ± s.e.m. *p<0.05, **p<0.01, ***p<0.001.



Extended Figure 10. Enrichment of ASD risk-genes in the shared ASD PPI network map and grouping of clinical phenotypes
a, Enrichment of full gene list, cytoplasmic gene only lists, and nuclear gene only lists from published works and SPARK, in the shared ASD-risk gene PPI network shown by signficance. (Fisher's exact test). Dashed lines represent nominal (p = 0.05, *left*) and Bonferroni corrected (p = 0.05/number of cell types, *right*) significance thresholds. **b,** ODDs ratio of full gene list, cytoplasmic gene only lists, and nuclear gene only lists enriched in the shared ASD-risk gene PPI network. **c,** Individuals with missense mutations in Cluster 1, 2 and 3 genes show no significant differences in socialization standard scores (Non-parametric Kruskal-Wallis test, p = 0.1765, *post-hoc* Dunn's test; Group 1 = 351, Group 2 = 114, and Group 3 = 416 probands). **d-e,** Individuals with splice site mutations in Cluster 1 have significantly higher adaptive behavior and socialization standard scores than Cluster 2 and 3 (Non-parametric Kruskal-Wallis test, adaptive behaviour: p = 0.0036, Group 1 = 32, Group 2 = 9, and Group 3 = 72 probands; socialization: p = 0.0021, Group 1 = 32, Group 2 = 9, and Group 3 = 71 probands; *post hoc* Dunn's test). **f-g,** Individuals with frameshift or stop gain mutations in Cluster 1, 2 and 3 genes show no significant differences in adaptive behavior and socialization standard scores (Non-parametric Kruskal-Wallis test, adaptive behaviour: p = 0.1069, Group 1 =51, Group 2 = 19, and Group 3 = 60 probands; socialization: p = 0.0803, Group 1 = 51, Group 2 = 19, and Group 3 = 60 probands; *post hoc* Dunn's test). **h,** Number of ASD-risk genes identified in each of the 41 ASD-risk gene protein-protein interactions. Dashed line represents average expected risk genes. Box and whisker plot (minimum, 1st quartile, median, 3rd quartile, maximum). *p<0.05, **p<0.01.

Table S1. BioID2 PPI networks of 41 ASD-risk genes and cellular compartment genes

Table S2. Comparison of BioID2 PPI networks identified in HEK293 cells and mouse cortical neurons

Table S3. Comparison of BioID2 PPI network enriched cellular components identified in HEK293 cells and mouse cortical neurons

Table S4. BioID2 PPI network enriched cellular components of compartment specific genes

Table S5. BioID2 PPI network enriched pathways of 41 ASD-risk genes
Table S6. 41 ASD-risk gene PPI network map enriched pathways

Table S7. Differentially expressed genes and proteins and dysregulated pathways in *Taok2* KO mouse cortices

Table S8. Comparison of BioID2 PPI networks between ASD-risk genes and their variants

Table S9. BioID2 PPI network enriched pathways of ASD-risk genes and their variants

Table S10. List of sources for 41 ASD-risk genes and cellular compartment genes
Table S1-S10 are posted online as Excel Files