Interaction studies of risk proteins in human induced neurons reveal convergent biology and novel mechanisms underlying autism spectrum disorders

Authors: Greta Pintacuda[1,2,†], Yu-Han H. Hsu[1,3,†], Kalliopi Tsafou[1,3], Ka Wan Li[4], Jacqueline M. Martín[1,2], Jackson Riseman[1,2], Miguel A. Gonzalez-Lozano[4], Shawn B. Egri[6], Jake Jaffe[6], August B. Smit[4], Nadine Fornelos[1,3], Kevin C. Eggan[1,2,*], Kasper Lage[1,3,5,*]

Affiliations:

[1]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA

[2]Harvard Stem Cell Institute, Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA

[3]Department of Surgery, Massachusetts General Hospital, Boston, MA, USA

[4]Department of Molecular and Cellular Neurobiology, CNCR, VU University, Amsterdam, The Netherlands

[5]Institute of Biological Psychiatry, Mental Health Center St. Hans, Mental Health Services Copenhagen, Denmark

[6]Proteomics Platform, Broad Institute of MIT and Harvard, Cambridge, MA, USA

*Correspondence to: lage.kasper@mgh.harvard.edu, keggan@mcb.harvard.edu

†Contributed equally

Abstract

Sequencing studies of autism spectrum disorders (ASDs) have identified numerous risk genes with enriched expression in the human brain, but it is still unclear how these genes converge into cell type-specific networks and how their encoded proteins mechanistically contribute to ASDs. To address this question, we performed brain cell type-specific interaction proteomics to build a protein-protein interaction network for 13 ASD risk genes in human excitatory neurons derived from iPS cells. The network contains many (>90%) reproducible interactions not reported in the literature and is enriched for transcriptionally perturbed genes observed in layer 2/3 cortical neurons of ASD patients, indicating that it can be explored for ASD-relevant biological discovery. We leveraged the network dataset to show that the brain-specific isoform of ANK2 is important for its interactions with synaptic proteins and characterized a PTEN-AKAP8L interaction that influences neuronal growth through the mTOR pathway. The IGF2BP1-3 complex emerges as a point of convergence in the network, and we showed that this complex is involved in a transcriptional circuit concentrating both common and rare variant risk of ASDs. Finally, we found the network itself enriched for ASD rare variant risk, indicating that it can complement genetic datasets for prioritizing additional risk genes. Our findings establish brain cell type-specific interactomes as an organizing framework to facilitate interpretation of genetic and transcriptomic data in ASDs and illustrate how both individual and convergent interactions lead to biological insights into the disease.

Introduction

Autism spectrum disorders (ASDs) comprise a group of heritable neurodevelopmental conditions associated with challenges in social interaction, cognition, and behavior. In recent

years, substantial progress has been made in understanding the genetic architecture of ASDs, establishing that their clinical manifestations reflect polygenic contributions from both common and rare genetic risk variants across a wide spectrum of allele frequencies (*1-9*). While common variants account for most of ASD heritability (*3*), rare variants have larger effect sizes, affect protein-coding genes (*10*), and are therefore more easily amenable to reductionistic experimental analyses and functional interpretation (*11*). However, even though some ASD risk genes implicated by rare variants immediately incriminate specific physiological processes and molecular mechanisms, such as chromatin remodeling and synaptic morphogenesis (*12, 13*), the majority do not individually offer a comprehensive picture of the biology of autism. More broadly, it remains to be determined to which extent ASD risk genes integrate into high-order pathways and networks (hereafter, we will collectively refer to these as networks) during neurodevelopment (*14, 15*) and which insights grouping genes in such cell type specific networks will offer.

Integration of genetic and protein-protein interaction (PPI) data has shown that genes implicated in immunological, metabolic, and cardiovascular diseases often converge into PPI networks (*16-19*). In the context of ASDs, early studies of *de novo* mutations and structural variation in risk genes also suggested that their encoded proteins form PPI networks (*20-22*). These studies establish that PPIs can provide a systematic framework for biological interpretation of ASD risk genes, however, they overwhelmingly relied on PPI data derived from public databases and thus lack cell type specificity, as experiments from highly proliferative and immortalized cell types (e.g., HEK or cancer cell lines) are vastly over-represented in these databases (*18, 23, 24*).

Recent developments in genetics, transcriptomics, and induced pluripotent stem cell (iPSC) technology have now opened up the possibility to address this limitation. Indeed, integration of genetic and transcriptomic datasets has shown that genes expressed in excitatory neurons are

particularly enriched for genetic risk in ASDs (*10, 25*). Furthermore, analyses of brain gene co-expression and regulatory networks (*26*), as well as single-cell RNA sequencing in *postmortem* brain tissues from patients and controls (*27-30*), have also identified cortical excitatory neurons as a key cell type in ASDs. In tandem, new iPSC protocols for producing large numbers of induced neurons (iNs) that resemble cortical excitatory neurons, provide the opportunity to systematically execute interaction proteomics in an *in vitro* neuronal cell model of key relevance to ASDs (*31, 32*).

In this study, we brought together emerging catalogs of risk genes, a cell model of human cortical excitatory neurons, proteomic approaches, and computational modeling to build a brain cell type-specific PPI network for ASD risk proteins. Specifically, through data integration of 26 immunoprecipitation (IP) experiments of 13 ASD risk proteins performed in iNs, we generated a combined brain cell type-specific PPI network in which >90% of the protein interactions are unreported in the literature, likely due to their selective expression in the brain. These data are nevertheless highly reproducible (>91% validation rate) and strongly overlap with interactions of the same proteins identified from IPs performed in human brain homogenates, indicating that iNs can capture PPIs present in analogous complex human primary tissues. As further support for the iN cell model, we found the network significantly enriched for transcriptionally perturbed genes in layer 2/3 cortical neurons of non-genetically profiled ASD patients, which agrees with previous evidence that specifically implicates this key cell type in ASD pathophysiology. To exemplify how perturbing specific PPIs in the network can lead to mechanistic insights and ASD-relevant cellular phenotypes, we showed that ablating a giant brain-specific ANK2 alternative exon, which concentrates mutations observed in ASD patients, resulted in loss of isoform-specific interactions with synaptic proteins, while disrupting a newly found PTEN-AKAP8L interaction affected

neuronal growth. Furthermore, convergent signals in the network can also provide important insights, as illustrated by the mRNA-binding complex IGF2BP1-3, which is connected to multiple ASD risk proteins in the network and is at the center of a transcriptional circuit enriched for both common and rare variant risks of ASDs. Finally, the network is also enriched for genes containing rare ASD-associated variants, indicating that it can complement large-scale genetic studies to prioritize additional risk genes.

In summary, our study demonstrates how brain cell type-specific interaction proteomics represents a rich and currently underutilized opportunity for biological discovery in ASDs, providing a rich community resource to catalyze interpretation of genetic and scRNA datasets. Our results lay a foundation for future discovery of biomarkers or targets for therapeutic intervention.

Results

ASD risk genes and proteins are expressed in induced neurons

A recent exome sequencing study of individuals on the autism spectrum identified over 100 ASD-associated genes using combined *de novo* and case-control analyses (*10*). 24 of these genes (i.e., *ADNP*, *ANK2*, *ANKRD11*, *ARID1B*, *ASH1L*, *CHD2*, *CHD8*, *CTNNB1*, *DEAF1*, *DSCAM*, *DYRK1A*, *FOXP1*, *GIGYF1*, *GRIN2B*, *KDM6B*, *KMT5B*, *MED13L*, *POGZ*, *PTEN*, *SCN2A*, *SHANK3*, *SLC6A1*, *SYNGAP1*, and *TLK2*) passed the statistical threshold of exome-wide significant enrichment for rare deleterious mutations. We used this set of 24 genes as the starting point of our workflow to generate an ASD PPI network in iNs (**Figure 1A**), and hereafter refer to them and their corresponding proteins as 'index genes' and 'index proteins', respectively.

As a substrate for cell type-specific analyses of the index genes and proteins, we adopted and further optimized NGN2 programming combined with developmental patterning as previously

described by Nehme *et al.* (*31*) to differentiate iPSCs into iNs. In particular, we used an iPS cell line (iPS3, **Methods**) where the neurogenic factor *Neurogenin 2* (*NGN2*) is integrated in the genome under a tetON promoter for rapid and controlled doxycycline-induced expression, resulting in highly homogeneous iN populations (**Figure 1B**).

We measured the expression of ASD index genes through RNA sequencing of cells during weeks 0, 3, and 6 of neuronal differentiation (**Figure 1C**). The expression patterns of these genes varied across time, with some genes showing decreased (e.g., *MED13L*) or increased (e.g., *CHD8*) expression levels at later time points. Generally, most index genes were expressed after three weeks of differentiation, including genes (i.e., *GRIN2B*, *SCN2A*, and *SHANK3*) with previously reported neuronal identity (*33*). Importantly, gene expression observed in iNs as early as week 3 (**Figure 1C**) resembled expression patterns seen in excitatory neuron populations of the human cortex (**Supplementary Figure 1A,B**) (*30*), supporting this cell model as a proxy for cell types in the human brain.

Next, we sought to identify immunoreagents to detect and immunoprecipitate index proteins in iNs. After testing ~60 antibodies (**Supplementary Table 1**), we concluded that there were no adequate commercial reagents for detection of seven proteins (ASH1L, CHD2, DSCAM, FOXP1, KDM6B, KMT5B, SLC6A1), based on western blots executed on a range of cells and tissues (i.e., mouse cortex and HEK293 cells; **Figure 1D**) included as positive controls. Furthermore, four proteins (ANKRD11, DEAF1, GRIN2B, POGZ) were successfully detected by immunoblotting (**Supplementary Figure 1C**), but no immunoreagent was IP-competent. The remaining 13 proteins could be detected by immunoblotting in the form of at least one protein isoform at week 3 of iN differentiation (**Figure 1E**) and were amenable to IP.

The protein expression analysis revealed that many index proteins show distinct isoform patterns across species (e.g., TLK2 in mouse cortex), neuronal differentiation time points (e.g., SHANK3 in iPSCs), or cell types (e.g., CHD8 in HEK293 cells; **Figure 1E**). This observation highlights the importance of using human iNs in order to capture cell type-specific biochemical interactions of relevance to ASDs.

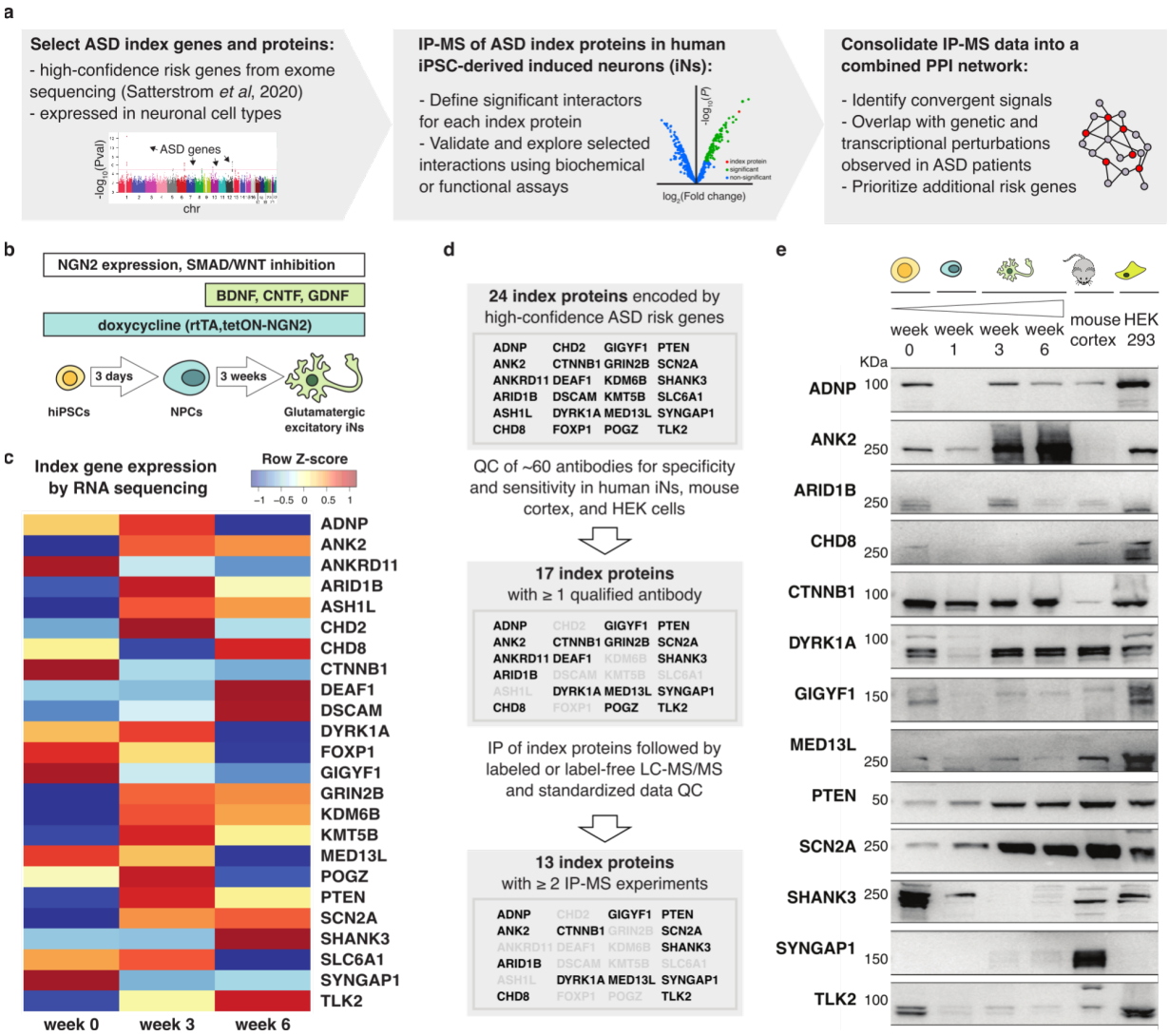

Figure 1. Study workflow and expression of ASD index genes and proteins in human iNs. (a) Schematic of the experimental and computational workflow to generate an ASD PPI network in

human iNs. **(b)** Schematic of the key steps in the glutamatergic-patterning protocol utilized to generate neural progenitor cells (NPCs) and excitatory iNs from human iPSCs. **(c)** Heatmap of ASD index gene expression over the course of neuronal maturation. Bulk RNA-seq analysis was performed on iPSCs (week 0) and after 3 and 6 weeks of neuronal maturation according to the protocol described in (b). **(d)** Flowchart representing the selection of index proteins studied in this work. Only proteins expressed in iNs for which IP-competent commercial antibodies are available were selected for downstream experiments. **(e)** Expression of 13 index proteins in differentiating iPSCs (as described in [b]), mouse cortex, and HEK293 cells detected by immunoblotting. The antibody utilized against ANK2 recognizes a human-specific epitope and SYNGAP1 is not expressed in HEK293 cells. Molecular weights (KDa) are marked on the left of each blot.

Interaction proteomics of index proteins in induced neurons

We executed IPs followed by mass spectrometry (IP-MS) of ADNP, ANK2, ARID1B, CHD8, DYRK1A, FOXP1, MED13L, POGZ, PTEN, SHANK3, SCN2A, SYNGAP1, and TLK2 in iNs at week 4 of differentiation (**Figure 2**, **Supplementary Figure 2**, and **Supplementary Table 2**). Each experiment was conducted in duplicates using ~15 million cells per replicate, and protein abundances in the IP samples were quantified using labeled or label-free liquid chromatography followed by tandem mass spectrometry (LC-MS/MS). In total, we conducted 37 IP-MS experiments for the 13 index proteins, generating 2-5 experimental datasets for each protein.

For each IP-MS experiment, we performed quality control (QC) and data analysis using Genoppi (*24*). Specifically, we calculated the $\log_2$ fold change (FC) and corresponding statistical significance of each protein identified in the index protein IPs compared to control IPs

(**Supplementary Table 2**). After filtering out 11 datasets that did not pass QC thresholds (i.e., $\log_2$ FC correlation between replicates was ≤ 0.6 or the index protein was not positively enriched at FDR ≤ 0.1), 26 high-quality datasets were used for all downstream analyses. In each dataset, we defined proteins with $\log_2$ FC > 0 and FDR ≤ 0.1 as 'significant interactors' of the index protein; the rest of the detected proteins were defined as 'non-interactors' (**Supplementary Table 3**). We also systematically intersected each dataset with known protein interactors of the index protein from the InWeb_InBioMap (InWeb) database (*23, 34*), which aggregates human PPI data from > 40k published articles (**Supplementary Table 3)**. This allowed us to assess if the identified interactions have been reported in the literature or are potentially novel.

We exemplify our analysis workflow using an IP of SHANK3: enrichment of SHANK3 in the IP was confirmed by immunoblotting (**Figure 2A**, left), and quantified in the IP-MS data ($\log_2$ FC = 1.2 and FDR = 0.012; **Figure 2B**). The IP-MS replicate $\log_2$ FC correlation is 0.79 (**Figure 2A**, right). Only two out of 104 significant interactors were previously reported in InWeb, suggesting that 98% of the interactions reported here are new (**Figure 2B**). Overall, the clear enrichment of SHANK3 both in the immunoblot of the IP and in the MS results, in combination with the high correlation between replicates, support the reproducibility and robustness of our data. The low number of known SHANK3 interactors supports the use of neuron-specific proteomics to discover new ASD risk gene relationships.

The ASD PPI network consists of many convergent, novel, and reproducible interactions

After analyzing each IP-MS dataset individually, we merged all the interactors identified across the 26 high-quality datasets to generate a combined PPI network for all 13 index proteins. This network comprises 1,021 proteins interacting with at least one index protein (**Figure 2C** and

Supplementary Table 3). The PPI network shows high interconnectivity, with > 20% of interactors being shared by more than one index protein (**Figure 2D** and **Supplementary Figure 3A**). Furthermore, whereas the interactors of several index proteins (ANK2, ARID1B, CTNNB1, DYRK1A, MED13L) are enriched for known InWeb interactors (**Supplementary Table 3**), > 90% of all interactions in the network have not been reported in InWeb (**Figure 2E**). This observation was not unexpected given the limited number of neuronal PPI datasets in the literature (*24*), and suggests the opportunity for biological discovery linked to further characterization of newly found interactions.

To confirm the robustness and reproducibility of MS-detected interactions, we extensively validated both unique (i.e., interacting with only one index protein) and shared (i.e., interacting with multiple index proteins) interactors in independent IPs by immunoblotting (**Supplementary Table 4** and **Supplementary Figure 3B-D)**. Overall, we validated 91.1% of the tested interactions, which is in line with the expected validation rate of ~90% based on the FDR ≤ 0.1 threshold used to define interactors, and with previously reported validation rates obtained using a similar experimental design in the same cell model (*24*).

To compare our data to protein interactions in the human brain, we also performed IP-MS experiments for a subset of index proteins (ANK2, DYRK1A, SCN2A, SHANK3, SYNGAP1) in *postmortem* brain homogenates (**Supplementary Table 2**). On average, brain IP-MS datasets have lower $\log_2$ FC correlations between replicates compared to the iN IP-MS data (average correlation = 0.63 and 0.72 for brain and iN IPs, respectively; **Supplementary Table 2**). This is likely reflecting noise and experimental artefacts of interaction proteomics executed in complex samples that comprise a mix of cell types and are not renewable. After applying the same QC metrics as previously described, we considered three of the brain IPs (for SCN2A, SHANK3, and

SYNGAP1) to be high-quality and compared the interactors identified in these IPs with the analogous iN-derived interactors (**Figure 2F** and **Supplementary Figure 3E,F**). Considering that the brain homogenates contain a mix of cells whereas the iNs are highly homogeneous, we observed a remarkable and statistically significant overlap between the interactors identified in the two sample types, with 42% of the brain interactors also identified as interactors in iNs (P = 2.6e-13). Throughout brain development (*35*) (www.brainspan.org), both brain- and iN-derived interactors have elevated gene expression in the cortex compared to random genes; in particular, interactors that were identified in both sample types have an expression profile that is comparable to ASD risk genes identified through exome sequencing (**Supplementary Figure 4A,B**). These results strongly support that our findings in iNs recapitulate those from a complex biological tissue in the human brain.

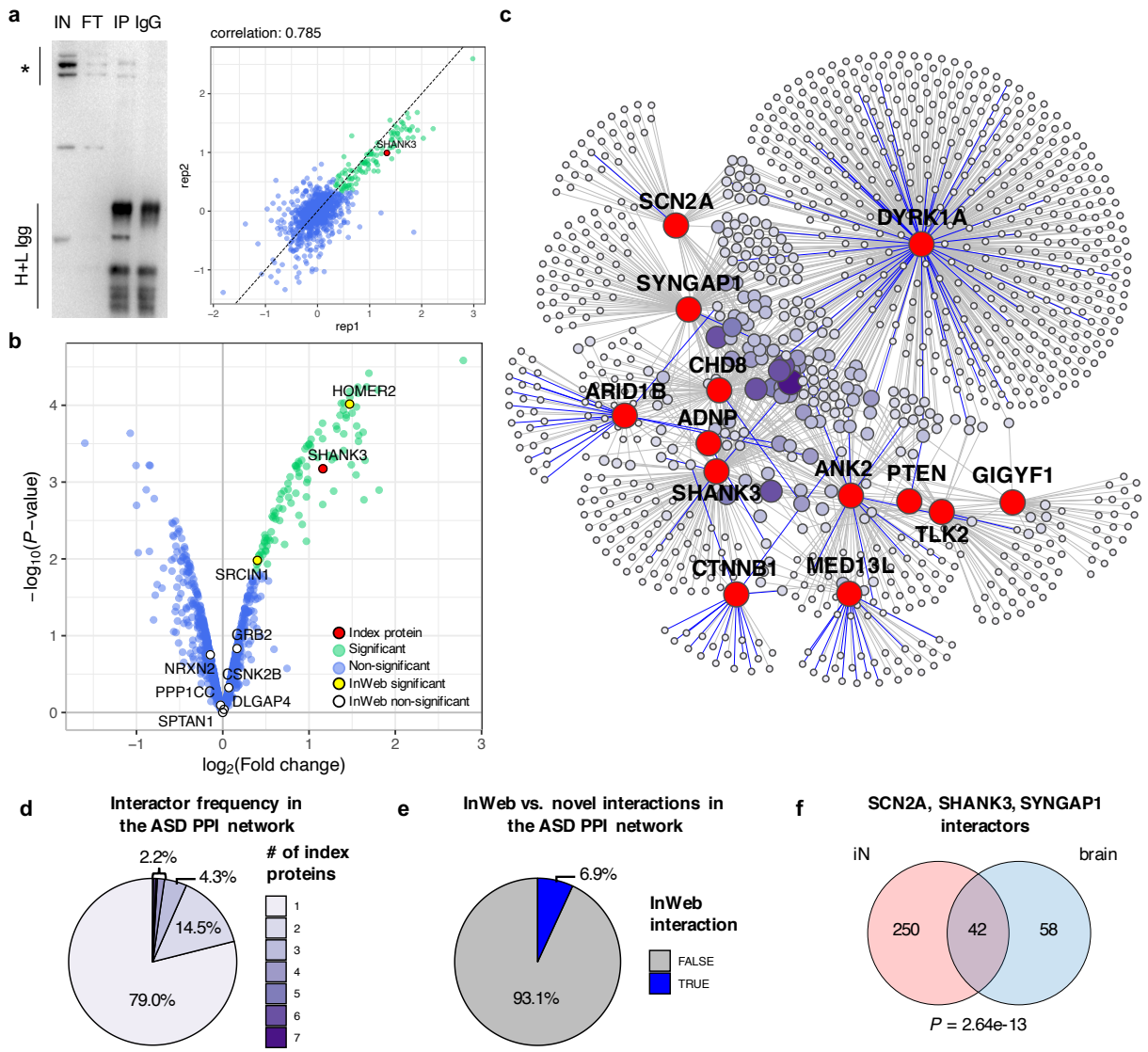

Figure 2. Generation of a combined PPI network for 13 ASD index proteins in iNs. (a) Example of a typical IP-MS experiment with SHANK3 as bait protein. Left: immunoblot of a SHANK3 IP with its main isoforms marked to the left by a vertical line and asterisk; IN=Input, FT=Flow-through, IP=Immunoprecipitation, IgG=IgG control, H+L Igg=Heavy and light IgG chains. Right: $\log_2$ fold change (FC) correlation between IP-MS replicates. **(b)** Volcano plot of IP-MS experiment from (a), showing SHANK3 in red, its significant interactors ($\log_2$ FC > 0 and FDR ≤ 0.1) in green, and non-interactors in blue; known InWeb interactors identified as interactors

or non-interactors in the experiment are highlighted in yellow or white, respectively. **(c)** Combined PPI network derived from 26 IP-MS experiments performed in iNs. Nodes represent index proteins (red) and their interactors (purple); color intensity and size of the interactor nodes scale with interactor frequency (i.e., how many index proteins are linked with each interactor). Edges indicate observed interactions, with known InWeb interactions highlighted in blue. **(d)** Pie chart showing distribution of interactor frequency in the combined network. **(e)** Pie chart showing distribution of InWeb vs. novel (i.e., non-InWeb) interactions in the combined network. **(f)** Venn diagram showing the overlap between interactors for SCN2A, SHANK3, and SYNGAP1 derived from IP-MS experiments performed in iNs vs. human brain homogenates. Overlap enrichment P-value was calculated using a one-tailed hypergeometric test that only considered the population of proteins detected in both iN and brain IPs.

The PPI network is transcriptionally perturbed in layer 2/3 cortical excitatory neurons of patients with ASDs

We assessed the tissue specificity of the network by calculating its overlap enrichment with GTEx tissue-specific genes (*36*) compared to the rest of the genome. As expected, the network showed statistically significant enrichment ($P < 0.05/53$, adjusting for 53 tissues) in the central nervous system, as well as in several other tissue categories (**Figure 3A** and **Supplementary Table 5**). When the analysis was repeated using brain region-specific genes, the network showed significant enrichment ($P < 0.05/13$, adjusting for 13 brain regions) in the frontal cortex, cerebellar hemisphere, anterior cingulate cortex, and amygdala (**Figure 3B** and **Supplementary Table 5**). In addition, we performed SynGO (*37*) gene set analysis to confirm that the PPI network is enriched for genes involved in biological processes in the synapse, such as "protein translation at

postsynapse" and "synapse organization" (**Supplementary Figure 4C** and **Supplementary Table 5**).

To link the network to specific cell types that are implicated in ASDs, we integrated our data with a recent single-cell RNA sequencing study (*30*), which measured gene expression in the *postmortem* cortex of ASD patients versus healthy controls and identified differentially expressed genes (DEGs) between the two groups in 17 annotated cell types. First, we performed a 'global' enrichment analysis to test whether the network genes show enriched overlap with commonly expressed genes in a cell type (i.e., expressed in >50% of cells in the cell type) compared to other genes in the genome. We also performed a more restricted 'conditional' enrichment analysis that compared the overlap of network genes to that of other genes whose proteins were also expressed in iNs (i.e., non-interactors detected in our IP-MS experiments, which did not show significant interaction with any of the index proteins; **Supplementary Table 3**). In both the global and conditional analyses, our network showed statistically significant enrichment ($P < 0.05/17$, adjusting for 17 cell types) in several neuronal cell types, including both excitatory neurons and inhibitory interneurons (**Figure 3C**, **Supplementary Figure 4D**, and **Supplementary Table 5**). Next, we tested the global and conditional enrichment of the network in terms of its overlap with cell type-specific DEGs in ASD patients compared to controls. In the global analysis, our network showed significant enrichment in several cell types including excitatory neurons, inhibitory interneurons, astrocytes, and microglia (**Supplementary Figure 4E** and **Supplementary Table 5**). Strikingly, in the more conservative conditional analysis, the layer 2/3 excitatory neurons stand out as the only significant cell type ($P < 0.05/17$; **Figure 3D** and **Supplementary Table 5**).

Taken together, these results firmly establish that the PPI network captures biochemical perturbations and pathways in ASD-relevant tissues and cell types, and specifically implicate layer 2/3 excitatory neurons as a key cell type for ASD pathophysiology.

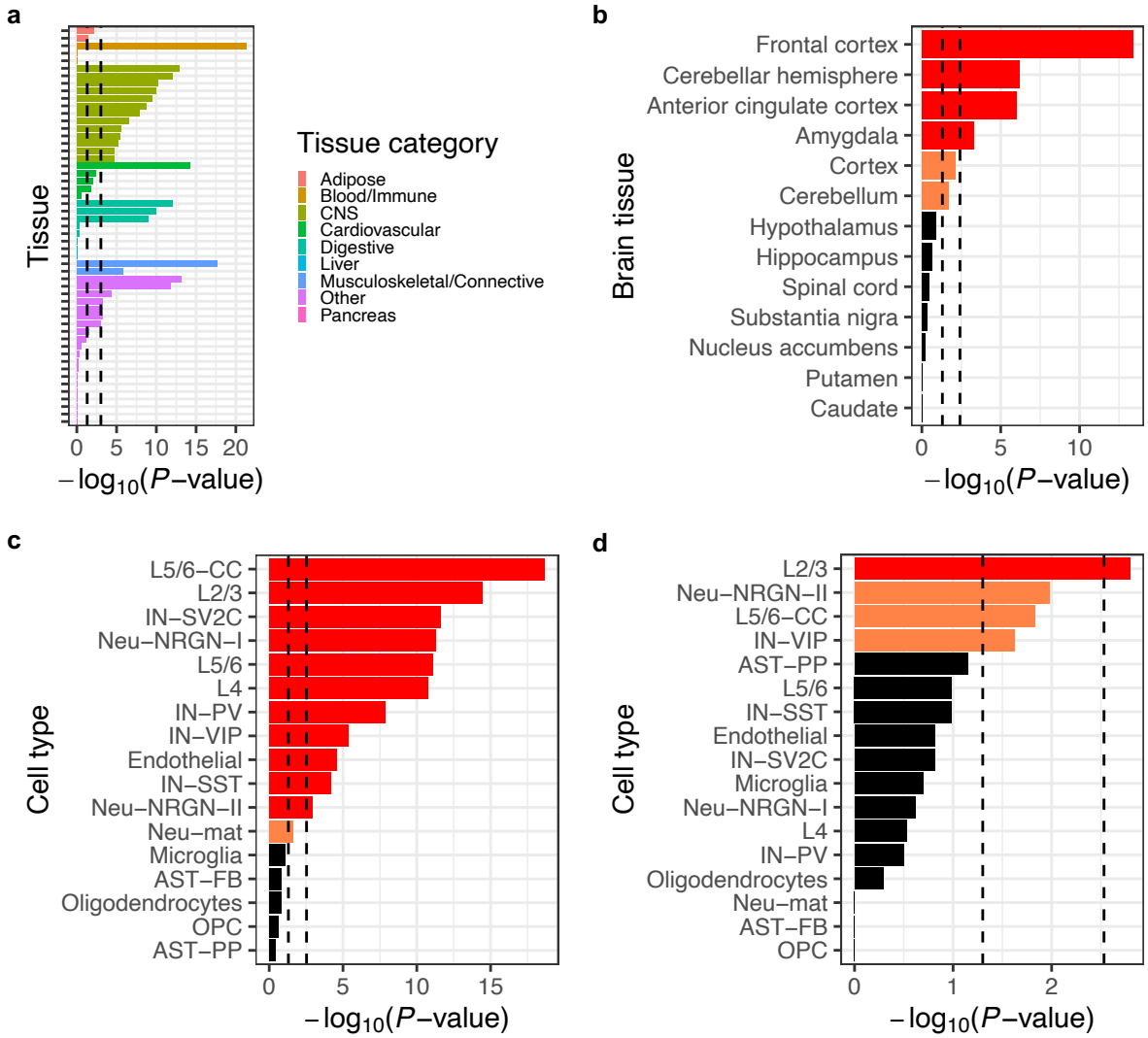

Figure 3. Tissue and cell type enrichment of the ASD PPI network. (a) Overlap enrichment between the network and GTEx tissue-specific genes. **(b)** Overlap enrichment between the network and GTEx brain region-specific genes. **(c)** Overlap enrichment between the network and genes expressed in > 50% of cells in each cell type in the *postmortem* cortex of ASD patients and controls. **(d)** Overlap enrichment between the network and differentially expressed genes in each

cell type in the *postmortem* cortex of ASD patients compared to controls. All P-values were derived from one-tailed hypergeometric tests; for (a) and (b), a genome-wide background (i.e., all genes in the GTEx data) was used as the 'population' in the test; for (c) and (d), a neuronal proteome background (i.e., genes encoding all proteins detected in iNs by IP-MS) was used. Left or right vertical dashed lines indicate nominal ($P < 0.05$) or Bonferroni-corrected ($P < 0.05$/number of tissues or cell types) significance thresholds, respectively; for (b)-(d), nominally or Bonferroni-significant results are shown in orange or red, respectively. Tissue or cell type abbreviations: CNS, central nervous system; AST-FB and AST-PP, fibrous and protoplasmic astrocytes; OPC, oligodendrocyte precursor cells; IN-PV, IN-SST, IN-SV2C, and IN-VIP, parvalbumin, somatostatin, SV2C, and VIP interneurons; L2/3 and L4, layer 2/3 and layer 4 excitatory neurons; L5/6 and L5/6-CC, layer 5/6 corticofugal projection and cortico-cortical projection neurons; Neu-mat, maturing neurons; Neu-NRGN1 and Neu-NRGN-II, NRGN-expression neurons.

Dissecting isoform-specific PPIs in the network: giant ANK2 knockout leads to loss of neuron-specific interactions in maturing neurons

A number of index genes in this study are disrupted in ASDs in an isoform-specific manner. This is the case for giant ankyrin-2 (ANK2) (*38*), a non-canonical and neurospecific alternatively spliced variant of the *ANK2* gene (*39*). This isoform is differentiated from canonical ANK2 due to inclusion of exon 37 (giant exon) encoding 9,240 amino acids, which contains the majority of the mutations identified in ASD patients (**Figure 4A**). When modeled in mice, human ASD-associated mutations of *ANK2* led to gain of axon branching in the central nervous system, a phenotype that is hypothesized to originate from loss of giant ANK2-specific interactions (*38*). While we validated numerous ANK2 interactors identified from the IP-MS experiments performed in iNs

(**Figure 4B**, **Supplementary Figure 5A**, and **Supplementary Table 4**), they are likely a mix of interactors of all ANK2 isoforms. Thus, to focus specifically on the ASD-relevant ANK2 isoform we increased the resolution of our interaction data by building an *in vitro* model of selective loss of giant ANK2 using CRISPR/Cas9 gene editing (**Figure 4C**).

We edited an iPS cell line (iPS3, **Methods**) to exclude retention of exon 37, which differentiates the giant ANK2 from canonical (short) ANK2 isoforms (**Figure 4D**, top). Fully characterized giant ANK2 knockout (KO) iPSCs (**Figure 4D**, bottom) were viable, but showed visible morphological defects upon neuronal differentiation, defective ANK2 distribution, and resulted in high cell death (data not shown). We could not collect sufficient giant ANK2 KO iNs for IP-MS due to high cell death during neuronal maturation. Nonetheless we performed immunoblot analysis in KO iNs to test for presence of selected ANK2 interactors identified by IP-MS in corresponding wild-type (WT) neurons. KO iNs were still expressing proteins essential for various aspects of neuronal physiology, including the calcium channel subunit CACNA2D and the complement protein C3, but these proteins were no longer associated with ANK2 (**Figure 4E**), suggesting that their localization might rely on ANK2-dependent cytoskeletal architecture.

We also performed IP-MS experiments of ANK2 in WT and giant ANK2 KO isogenic neural progenitor cells (NPCs) after one week of neuronal induction, and identified a number of ANK2 interactors that are specific to either WT (which express both canonical and giant ANK2 isoforms) or KO cells (**Supplementary Table 2** and **Supplementary Figure 5B**). GO analysis of differential interactors shows that the most enriched cellular component for the WT-specific interactors is “glutamatergic synapses”, while the KO-specific interactors is most enriched for “microtubules” (**Supplementary Table 6**). These results suggest the giant ANK2 isoform plays a central role in localizing the protein to form cell projections of maturating NPCs and in guiding

neuronal morphology. This hypothesis is further supported by the observation that ARPC5L is amongst the ANK2-WT-specific interactors. ARPC5L is a component of the ARP2/3 complex that regulates the synthesis of branched actin networks and is essential for both structural and functional maturation of the growth cone in maturating neurons (*40*).

These observations collectively suggest that the wide range of functional defects previously observed in giant ANK2 loss-of-function (LoF) models first result from mis-localization of giant ANK2-specific interactors. Severe defects in cell morphogenesis could be a secondary effect. This vignette demonstrates the potential of using cell type-specific and isoform-specific PPI networks to gain detailed mechanistic insights into ASDs.

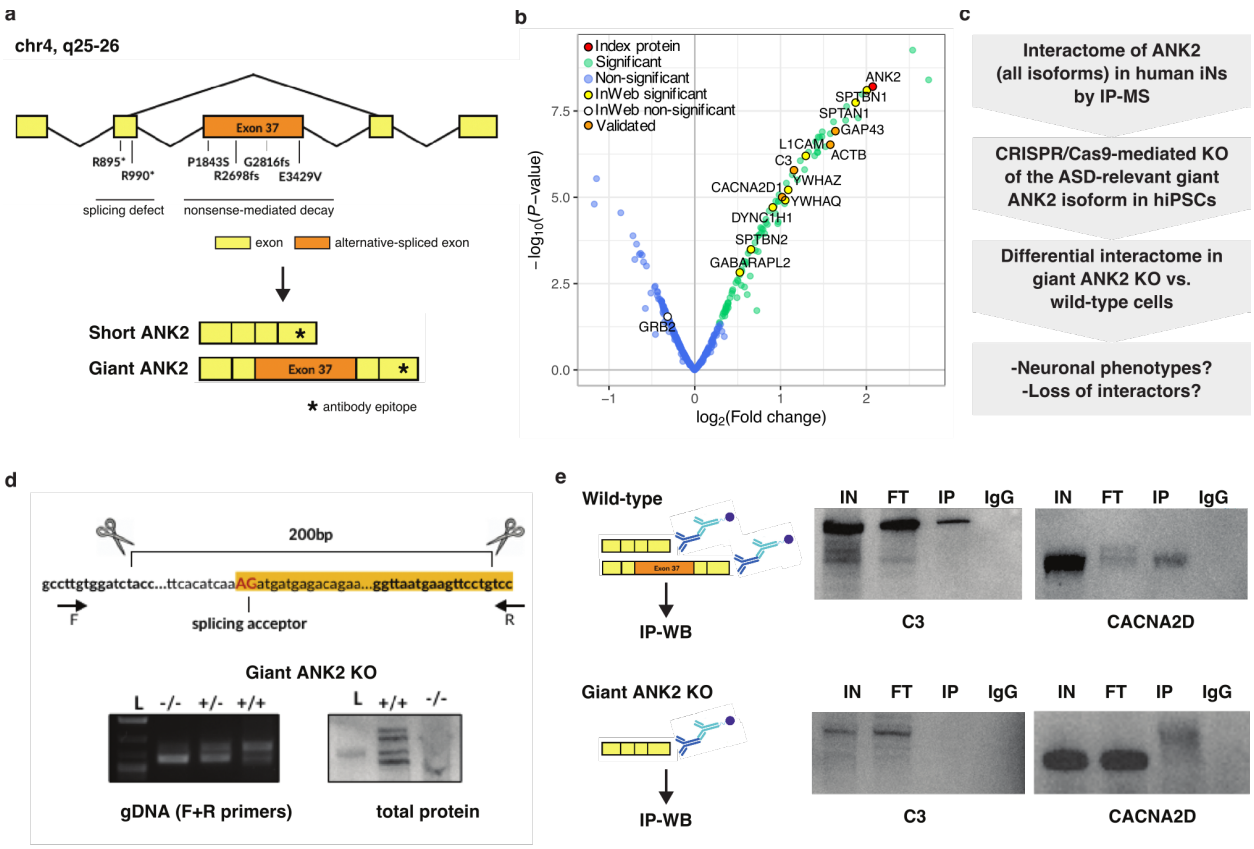

Figure 4. Giant ANK2 is responsible for brain-specific PPIs. (a) Depiction of the ANK2 gene with mutations observed in ASD patients leading to splicing defects or early decay. Giant ANK2

includes exon 37, which is alternatively spliced to produce short ANK2. The epitope recognized by the ANK2 antibody used in this study is marked with an asterisk. **(b)** Volcano plot of an IP-MS experiment for ANK2 in iNs, showing ANK2 in red, significant interactors ($\log_2$ FC > 0 and FDR ≤ 0.1) in green, and non-interactors in blue; known InWeb interactors identified as interactors or non-interactors in the experiment are highlighted in yellow or white, respectively; interactors validated by immunoblotting are highlighted in orange. **(c)** Experimental workflow to identify giant ANK2-specific PPIs. **(d)** Top: CRISPR/Cas9 editing strategy to generate a cell line exclusively expressing the short ANK2 isoform, obtained by deletion of the giant ANK2 exon splicing acceptor. Bottom: validation of giant ANK2 KO cell line. The deletion was confirmed by PCR on genomic DNA using primers flanking the deletion site (left), and by immunoblotting (right) showing the absence of the giant ANK2 isoform (-/-). **(e)** C3 and CACNA2D western blots on ANK2 IPs in wild-type and giant ANK2 KO iNs. L=Ladder, IN=Input, FT=Flow-through, IP=Immunoprecipitation, IgG=IgG control.

Perturbing a newly found PPI in the network: AKAP8L knockout phenocopies PTEN's cellular phenotype in iNs

We next performed functional follow-up experiments to investigate a newly identified interaction found in the PTEN IPs. PTEN encodes a phosphatase that acts as a tumor suppressor by antagonizing PI3K/AKT/mTOR signaling (*41*). PTEN is mutated at high frequency in a large number of cancers, but germline mutations of PTEN have also been described in a subset of patients with autism and macrocephaly (*42*). However, it is unclear whether cell proliferation caused by disruption of mTOR signaling upon PTEN mutation can also be linked to these

developmental disorders (*43*). To address this question, we further explored the PTEN interactors found by IP-MS in iNs.

Surprisingly, the most robust interactor we identified for PTEN in iNs is the A-kinase anchor protein AKAP8L (**Figure 5A**), which has not been associated with PTEN in previous biochemical studies (that were almost exclusively conducted in cancer cell models or tissues), suggesting a neuron-specific mTOR circuitry of PTEN that acts through AKAP8L. The PTEN-AKAP8L interaction was replicated using two independent PTEN antibodies for IP-MS (data not shown) and consistently validated by immunoblotting and reverse IP (**Figure 5B, Supplementary Figure 5C**).

We generated AKAP8L KO (-/-) iPS cell lines through CRISPR/Cas9 gene editing (**Figure 5C**) to test if they phenocopy neurodevelopmental phenotypes observed in individuals with PTEN-associated autism and macrocephaly. We obtained isogenic lines with both heterozygous (+/-) and homozygous (-/-) AKAP8L mutations, with the homozygous KO line resulting in complete loss of AKAP8L (**Figure 5D**). Interestingly, the AKAP8L homozygous KO line showed a statistically significant increase in growth rate at the NPC stage, compared to isogenic wild-type and heterozygous lines ($P = 0.0287$ and 0.0162, respectively; **Figure 5E**).

To gain deeper mechanistic insights we tested whether this cellular phenotype is linked to a neuron-specific circuitry that includes the mTOR pathway. Although AKAP8L KO did not result in changes of PTEN levels in iNs as measured by immunoblotting, we did observe a small, but detectable increase in the levels of phosphorylated ribosomal protein S6 (**Figure 5F**). This is a marker of hyperactivation of the mTOR pathway and suggests an AKAP8L-mediated defect in the PTEN signaling cascade in ASDs. Taken together, our results illustrate a neuronal variation on the PTEN-mTOR circuitry that includes AKAP8L and specifically suggest that the PTEN-AKAP8L

interaction impacts downstream signalling and cell proliferation without impacting PTEN protein levels directly (**Figure 5G**).

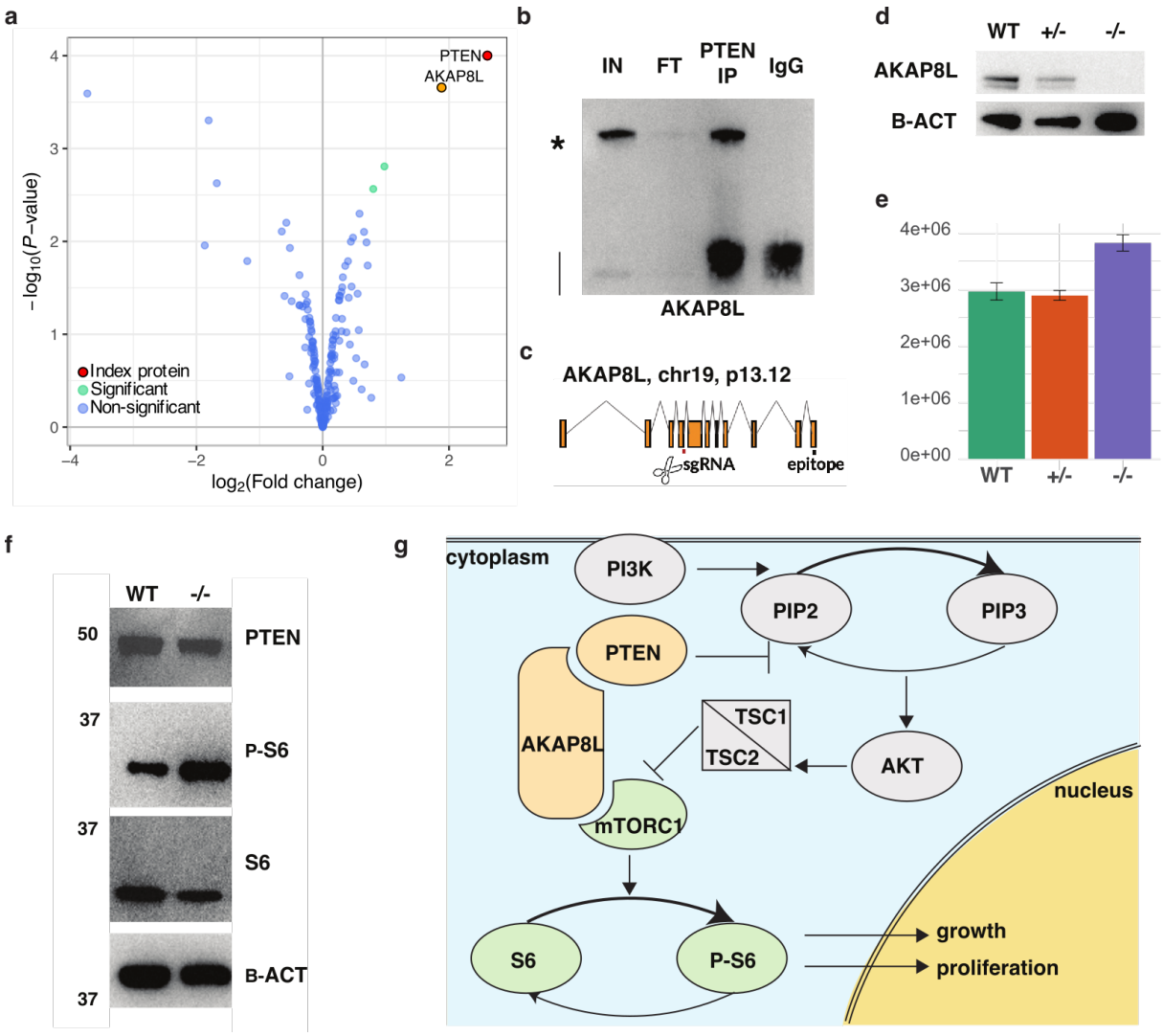

Figure 5. The neuronal PTEN-AKAP8L interaction regulates cell proliferation. (a) Volcano plot of an IP-MS experiment for PTEN in iNs, showing PTEN in red, significant interactors ($\log_2$ FC > 0 and FDR ≤ 0.1) in green, and non-interactors in blue; AKAP8L is marked in orange. **(b)** AKAP8L immunoblot on the PTEN IP. AKAP8L is marked with an asterisk, IgG heavy chains are marked with a vertical line; IN=Input, FT=Flow-through, IP=Immunoprecipitation, IgG=IgG

control. **(c)** Schematic of the CRISPR/Cas9 editing strategy to generate AKAP8L KO iPS cell lines. **(d)** Immunoblot validation of AKAP8L heterozygous (+/-) and homozygous (-/-) KO lines generated using the strategy described in (c). B-actin is used as a loading control. **(e)** Comparison of wild-type (WT), AKAP8L heterozygous KO (+/-), and AKAP8L homozygous KO (-/-) neural progenitor cell growth over four days upon seeding cells at identical confluences. Error bars are standard deviations of the mean of three biological replicates. P-values were calculated using a two-tailed t-test. **(f)** Immunoblot analysis of PTEN, phosphorylated S6 (P-S6), and S6 in WT and AKAP8L homozygous KO iNs. Molecular weights (KDa) are marked to the left of each blot and B-actin is used as loading control. **(g)** Pathway depiction of the AKAP8L interaction with both PTEN and mTORC1. Proteins shown in green have been associated with increased cell growth and proliferation while proteins in orange have been linked to decreased cell growth and proliferation.

Network convergence implicates the IGF2BP1-3 complex in an ASD-relevant transcriptional and protein interaction circuitry

Since the combined PPI network derived from all IP-MS datasets identified a number of interactors linked to multiple index proteins (**Figure 2C,D; Supplementary Table 3**), we hypothesized that the degree of convergence in the network could be used to discover biological themes of relevance to ASDs. By ranking proteins in descending order based on the number of index proteins they interact with (**Supplementary Table 3**), we noticed that three insulin-like growth factor 2 mRNA-binding proteins, IGF2BP1, 3, and 2, are all among the top ten recurring interactors and are associated with seven, six, and five of the index proteins, respectively. All three

proteins interact with CHD8, DYRK1A, SCN2A, SHANK3, and SYNGAP1; IGF2BP1 and IGF2BP3 additionally interact with TLK2, and IGF2BP1 further interacts with ANK2.

IGF2BP1-3 belong to a conserved family of single-stranded RNA-binding proteins. Together, they form a N6-methyladenosine (m6A)-reader complex broadly involved in post-transcriptional regulation by impacting mRNA stability and translation (*44*). We therefore set out to test the hypothesis that this complex is implicated in ASDs through a circuit connecting transcription of ASD risk genes to the protein network that we have identified. We analyzed published RNA targets of IGF2BP1-3 identified in HEK cells (*44*) and found that targets of each protein individually, as well as their combined targets, are all significantly enriched for ASD risk genes identified through exome sequencing (*10*) (P = 3.9e-10 for the combined target list; **Figure 6A**, **Supplementary Table 7**); among these risk genes are also nine of the index genes in this study.

Genetic risk enrichment analysis using ASD GWAS data (*45*) and MAGMA (*46*) also found the IGF2BP1-3 targets to be enriched for common variant risk of ASDs (P = 1.4e-3 for the combined target list; **Figure 6B**, **Supplementary Table 7**). Furthermore, we performed a conditional enrichment analysis to test whether genes in our combined PPI network and index protein-specific sub-networks are enriched for IGF2BP1-3 targets compared to the non-interactors detected in our neuronal proteomic data (**Figure 6C**, **Supplementary Table 7**). At a Bonferroni-corrected significance threshold ($P < 0.05/56$, adjusting for 14 networks and 4 IGF2BP target lists), the combined and DYRK1A networks both show significant enrichment for IGF2BP1-3 targets (P = 1.7e-17 and 7.6e-9 for the combined target list, respectively).

DYRK1A is a ubiquitously expressed protein kinase encoded by a gene localized in the Down syndrome critical region of chromosome 21 and is strongly associated with learning defects

associated with Down syndrome (*47*) and ASDs (*48*). DYRK1A has been implicated in post-transcriptional regulation of different cellular processes involved in brain development and function, ranging from early embryogenesis through late aging (*49*). Thus, given the enriched overlap observed between the IGF2BP1-3 targets and the DYRK1A network, we hypothesized that DYRK1A might be part of an ASD-relevant feedback circuitry acting upstream of the IGF2BP1-3 complex. To test this hypothesis, we first verified that the interactions between DYRK1A and IGF2BP1-3 identified by IP-MS (**Figure 6D**) are not mediated by nucleic acids, hence indicating a direct regulatory relationship between DYRK1A and the IGF2BP1-3 complex (**Figure 6E**). Next, we performed siRNA knockdown of DYRK1A in iPSCs, and observed that partial knockdown was sufficient to trigger increased expression of IGF2BP1-3 (**Figure 6F**).

Overall, our results suggest that the IGF2BP1-3 complex regulates an ASD-relevant transcriptional circuitry, and that the newly found interactions between DYRK1A and the complex might be key in modulating the expression of numerous ASD-relevant genes.

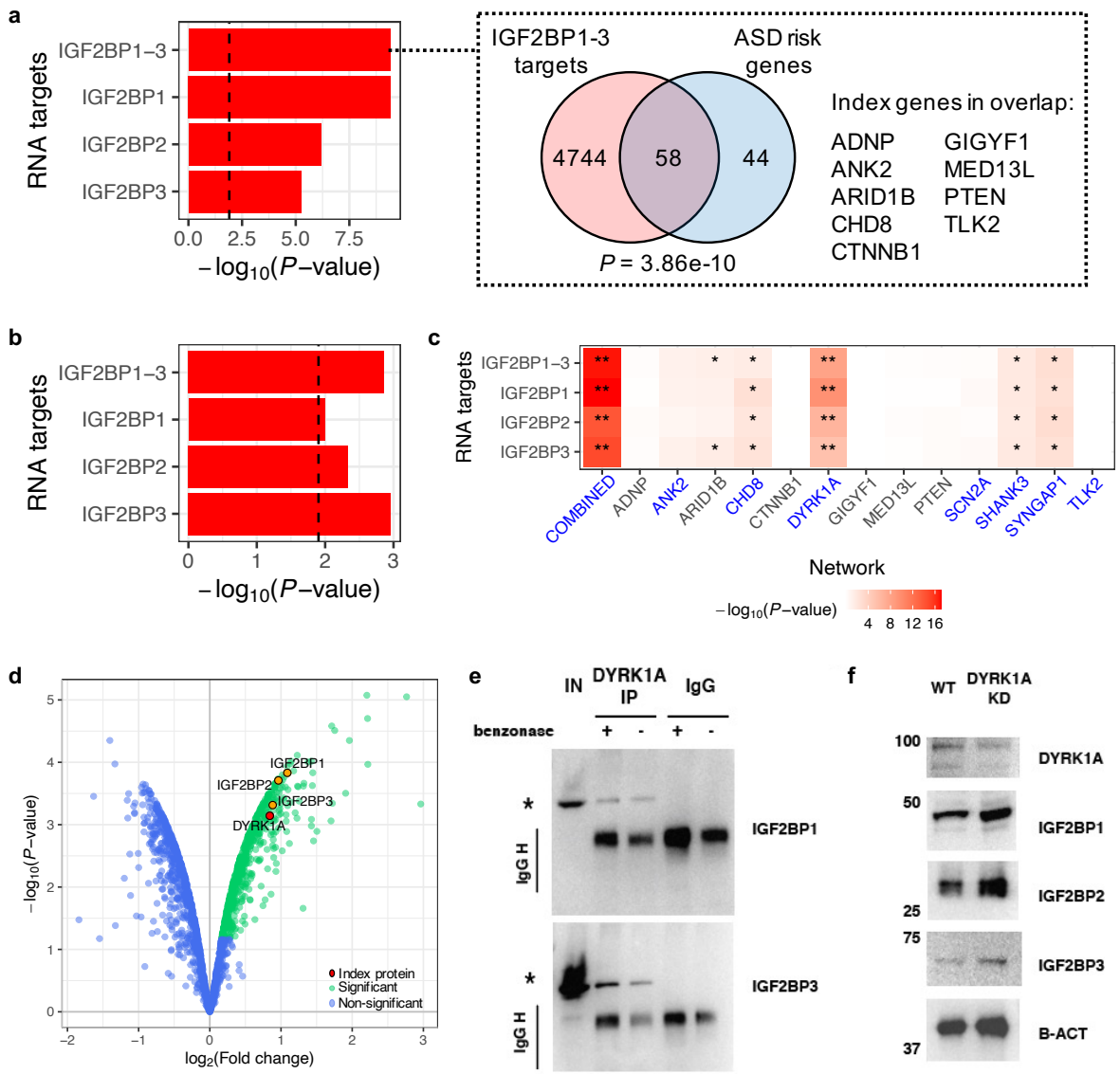

Figure 6. The IGF2BP1-3 complex interacts with ASD-related transcripts and proteins. **(a)** Left: overlap enrichment between RNA targets of IGF2BP1-3 and ASD risk genes (FDR ≤ 0.1 in exome sequencing data). 'IGF2BP1-3' indicates the combined target list (i.e., union of the targets for each IGF2BP). Enrichment P-values were calculated using one-tailed hypergeometric tests; vertical dashed line indicates Bonferroni-corrected ($P < 0.05/4$, adjusting for number of target lists) significance threshold. Right: Venn diagram showing the overlap between the IGF2BP1-3 combined targets and ASD risk genes. Index genes in the overlap are listed on the right. **(b)**

Common variant risk enrichment of IGF2BP1-3 targets derived using ASD GWAS data and MAGMA. A vertical dashed line indicates Bonferroni-corrected ($P < 0.05/4$) significance threshold. **(c)** Overlap enrichment between IGF2BP1-3 targets and the ASD PPI network or index protein-specific sub-networks (compared to non-interactors). Enrichment P-values were calculated using one-tailed hypergeometric tests; nominally ($P < 0.05$) or Bonferroni-significant ($P < 0.05/4$) results are indicated by single or double asterisks, respectively. Networks that contain any of the IGF2BPs are highlighted in blue on the x-axis. **(d)** Volcano plot of an IP-MS experiment for DYRK1A in iNs, showing DYRK1A in red, significant interactors ($\log_2$ FC > 0 and FDR ≤ 0.1) in green, and non-interactors in blue; IGF2BP1-3, which were validated by western blot, are marked in orange. **(e)** Immunoblot on an DYRK1A IP with and without addition of 25U benzonase, as stated at the top. An asterisk marks the expected band of each protein named on the right. L=Ladder, IN=Input, IP=Immunoprecipitation, Igg=IgG control. **(f)** Immunoblot showing protein levels (named on the right) upon DYRK1A siRNA-mediated knockdown for 48 hours in iPSCs. B-actin is used as loading control. Un=Untreated, KD=knockdown.

The network is enriched for rare variant risks of ASDs and developmental disorders

To evaluate whether the combined PPI network and index protein-specific sub-networks are enriched for genetic risks of ASDs, we performed global and conditional enrichment analyses using rare variant association statistics derived from exome sequencing of ASD patients and controls (*10*). In the global analysis, we tested whether genes in the networks are more significantly associated with ASDs compared to other protein-coding genes in the genome. In the conditional analysis, we compared the association scores of the network genes against that of the non-interactors detected in our neuronal proteomic data. Five networks (combined, ANK2, CHD8,

DYRK1A, SHANK3) reached Bonferroni significance ($P < 0.05/14$, adjusting for 14 networks) in the global analysis (**Supplementary Figure 6A**), and only the combined network was significant (P = 1.1e-5) in the conditional analysis (**Figure 7A** and **Supplementary Table 8**). Notably, the conditional enrichment was robust even after we removed known InWeb interactors from the combined network (P = 5.5e-5; **Supplementary Figure 6B**), supporting that this signal is not driven by the proteome of the neuronal cell model or the known interaction partners of the index proteins. Rather, the novel, neuron-specific interactions identified in this study are enriched for rare variant risk of ASDs over and above a background set of proteins expressed in iNs. We also performed analogous enrichment analyses using exome sequencing data for developmental disorders (DD) (*50*) and schizophrenia (SCZ) (*51*) and gnomAD (*52*) pLI scores, which represent the probability of a gene being loss-of-function intolerant (**Figure 7A**, **Supplementary Figure 6A,B**, and **Supplementary Table 8**). In the conditional analysis, we observed that the combined network is significantly enriched for DD genes (P = 6.2e-7) and genes with high pLI scores (P = 2.6e-31); five index protein-specific sub-networks (ANK2, ARID1B, CHD8, DYRK1A, SYNGAP1) are also enriched for high pLI scores (P = 1.1e-12 to 3.4e-4). In contrast, there was no significant SCZ enrichment for any of the networks. Overall, these results indicate that the interactors that we have identified contribute to rare variant risks of both ASDs and DD more than one would expect for genes generally expressed in neurons. The interactors are also more likely to be intolerant to loss-of-function mutations, suggesting that they might be essential for cell fate specification or survival.

Next, in order to assess the common variant risk enrichment of the PPI networks, we performed global and conditional analyses using ASD GWAS data (*45*) and MAGMA (*46*). Only the SYNGAP1 network reached nominal significance in both the global and conditional analyses

(conditional P = 3.71e-3, which is just shy of the Bonferroni-corrected threshold of $P < 0.05/14$; **Supplementary Figure 6C** and **Supplementary Table 9**). We also repeated the MAGMA analyses using GWAS data of other psychiatric disorders, including attention deficit hyperactivity disorder (ADHD) (*53*), bipolar disorder (BIP) (*54*), major depressive disorder (MDD) (*55*), and SCZ (*56*), as well as height (*57*) as an additional control trait. While several networks were nominally significant for some of these phenotypes, none reached Bonferroni significance (**Supplementary Figure 6C** and **Supplementary Table 9**). Given that the GWAS for ASDs and other psychiatric disorders had relatively modest sample sizes (~46k to 77k samples, except for MDD which had 500k samples), repeating these analyses when better-powered GWAS become available in the future may allow us to assess whether the suggestive common variant enrichment results reported here reflect true biology related to psychiatric conditions.

The network can prioritize additional ASD risk genes from genetic studies

Since the combined PPI network is enriched for rare variants associated with ASDs, we used it to prioritize additional ASD risk genes that did not meet stringent statistical significance cutoffs in the exome sequencing study (*10*). Specifically, we generated a 'social Manhattan plot' to highlight protein interactions in the network between index genes and other ASD risk genes with FDR ≤ 0.25 in the exome sequencing data (**Figure 7C**). Consistent with the rare variant enrichment results, the overlap between our network and these ASD risk genes is statistically significant (P = 4.9e-4), with DYRK1A being linked to the most risk genes out of all the index genes. In total, we prioritized 32 ASD risk genes (13 with FDR ≤ 0.1 and 19 with FDR > 0.1 and ≤ 0.25) that are connected to at least one index gene in the network. Interestingly, these prioritized genes have elevated gene expression throughout cortical development that mirrors the high-

confidence, exome-wide significant genes identified from exome sequencing (**Supplementary Figure 6D**). While follow-up investigation on these genes is necessary to better validate and evaluate our prioritization approach, this analysis showcases how the ASD PPI network is a unique resource that can be used to corroborate findings from recent GWAS and sequencing studies of ASDs, both in terms of building an exhaustive list of ASD risk genes and generating hypotheses of how these genes might contribute to ASDs mechanistically.

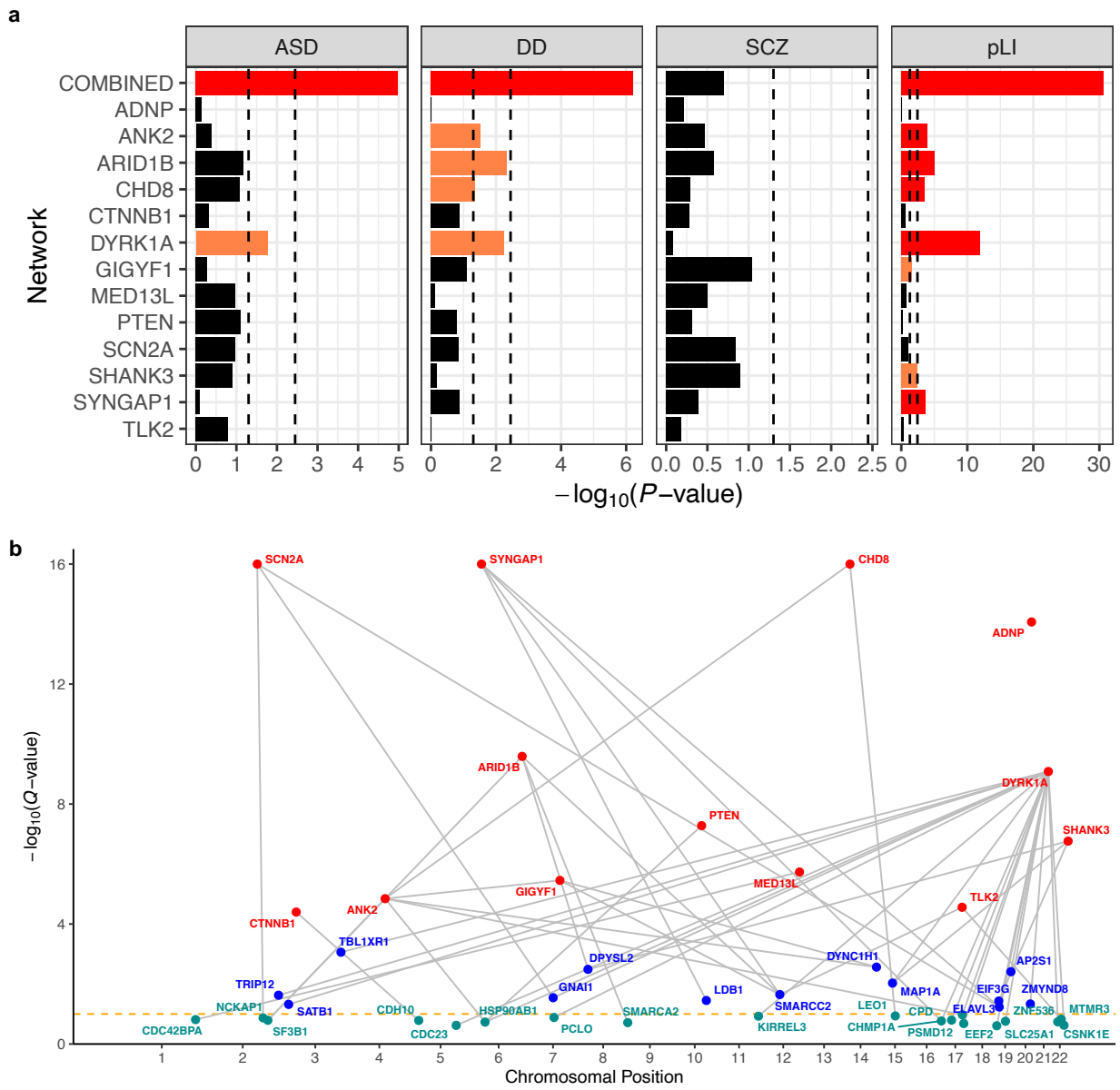

Figure 7. Genetic risk enrichment in the ASD PPI network. (a) Rare variant or pLI score enrichment of the network or index protein-specific sub-networks compared to non-interactors, derived using one-tailed Kolmogorov-Smirnov tests and gene-based association statistics from ASD, DD, or SCZ exome sequencing data or gnomAD pLI scores. Left or right vertical dashed lines indicate nominal ($P < 0.05$) or Bonferroni-corrected ($P < 0.05/14$, adjusting for 14 networks) significance thresholds, respectively; nominally or Bonferroni-significant results are shown in

orange or red, respectively. **(b)** Social Manhattan plot of index genes and interactors that are suggestive ASD risk genes in exome sequencing data. Index genes are in red, FDR ≤ 0.1 risk genes are in blue, and FDR ≤ 0.25 risk genes are in cyan. Genes connected by grey lines indicate observed interactions in the ASD PPI network.

Discussion

ASDs are a heterogeneous set of heritable neurodevelopmental conditions with contributions from both rare and common genetic variants, and their genetic and biological complexity is reflected in the clinical diversity of ASD patients (*9*). In the past decade, rare variants with large effect sizes in protein-coding genes have been shown to significantly shift individual risk of ASDs during development and early childhood (*11*). In particular, a recent exome sequencing study identified 24 genes that contain an excess of rare deleterious mutations in ASD patients versus controls (*10*). However, how such ASD-associated mutations act singularly or in ensemble remains largely unknown (*14, 15*). For instance, there is evidence that mutations in several of these genes result in similar mild impairments in social and communication skills (*58*). In other cases, specific mutations are associated with altered brain size and other synaptopathies (*59, 60*), but it is still unclear whether their function may converge on a single or restricted number of pathways and networks (*14*).

In this study, we combined human genetics, iPSC technology, computational modeling, and interaction proteomics to shed light on how a subset of ASD risk genes (index genes) may functionally converge in disease-relevant and brain cell type-specific pathways. Starting with the 24 high-confidence ASD index genes identified by exome sequencing (*10*), we performed extensive screening of immunoreagents amenable for IP of their encoded proteins, and

successfully carried out 26 IP-MS experiments of 13 index proteins (ADNP, ANK2, ARID1B, CHD8, CTNNB1, DYRK1A, GIGYF1, MED13L, PTEN, SCN2A, SHANK3, SYNGAP1, TLK2) in human iPSC-derived excitatory neurons. While excitatory neurons were challenging to obtain *in vitro* in the past (*61*), recent iPSC protocols have enabled large-scale production of homogeneous, electrophysiologically active, and synaptically connected iNs that resemble cortical excitatory neurons in approximately three weeks (*31, 62*), allowing to systematically execute interaction proteomics in an *in vitro* neuronal cell model of key relevance to ASDs according to recent genetic and transcriptomic studies (*63*). We consolidated the iN-derived IP-MS data into a combined PPI network consisting of the 13 index proteins and 1,021 interactor proteins that represents, to our knowledge, the most extensive cell type-specific ASD PPI network thus far. We found that over 90% of the interactions in the network consists of interactions that have not been reported in the literature but are nevertheless highly reproducible (91.1% validation rate). This finding reflects the potential for biological discovery of neuronal-specific interactome proteomics, as opposed to relying on publicly available PPI datasets that are mostly populated by data collected in highly proliferative or immortalized cell types (e.g., HEK cells, cancer cell lines) (*23, 34*). We also performed IP-MS for a subset of index proteins in human brain samples and found that a significant portion of the iN-derived PPIs recapitulate interactions observed in the brain. Furthermore, the lower statistical power in the brain IP-MS data (i.e., lower replicate correlations and fewer significant interactions) supports our choice of executing interaction proteomics in more homogeneous, *in vitro* iNs instead of less abundant and non-renewable brain samples consisting of a complex mix of cell types. The neuronal relevance and specificity of the PPI network is further supported by its enrichment for genes expressed in brain tissues (*36*) and neuronal cell types (*30*). We also observed a significant overlap between the network and differentially expressed genes in

the *postmortem* cortex of idiopathic ASD patients versus controls, specifically in layer 2/3 cortical excitatory neurons (*30*). This finding is of particular relevance given that such patients were not genetically characterized, suggesting that PPIs can be envisioned as a tool to stratify ASD patients in the future.

Our findings also illustrate how novel, brain-specific interactions within the PPI network can link ASD-associated genetic signals to relevant neuronal phenotypes. The *ANK2* gene, for instance, encodes for two major protein isoforms as a result of alternative splicing. One protein isoform has a molecular weight of 220 kDa and is expressed in multiple tissues, while the other (commonly referred to as giant ANK2) has a molecular weight of 440 kDa due to the inclusion of exon 37 and is expressed only in the nervous system (*39*). Three human ANK2 ASD mutations (P1843S, R2608 frameshift, and E3429V) are located in exon 37 and thus only affect giant ANK2, whereas others (R895*, R990*, Q3683fs) affect both isoforms (*64*). In particular, a mouse model for human ASD mutation of giant ANK2 shows increased axonal branching and ectopic connectivity in cultured neurons, as well as a transient increase in excitatory synapses during postnatal development (*38*). In order to identify isoform-specific PPIs that may contribute to ANK2 function in axon guidance, we designed a giant ANK2 KO iPSC model and compared ANK2 IPs performed in the wild-type versus giant ANK2 KO NPCs and iNs. Our results suggest that giant ANK2 and its PPIs are essential for proper localization of cytoskeletal components to cell projections and maintenance of neuronal morphology. We were also able to recapitulate cellular phenotypes observed in mice, and concluded that they are the result of both primary loss-of-function of neuronal proteins (likely due to mis-localization), and secondary effects linked to severely impaired cell morphogenesis.

Another example of how the PPI network can be used to reconcile genetic signals with specific molecular mechanisms leading to cellular phenotypes resides in the newly characterized interaction between PTEN and AKAP8L. *PTEN* is frequently mutated in cancers due to its antagonizing role in PI3K/AKT/mTOR signaling (*41*), but germline mutations of *PTEN* have also been described in a subset of patients with autism combined with macrocephaly (*42*). However, there is no consensus on whether PTEN-mediated cell proliferation defects are tied to the neuronal phenotypes observed in ASDs. We found that knockout of the pseudo-kinase AKAP8L, the most robust interaction partner of PTEN in iNs, phenocopies its cellular phenotype. This discovery in iNs, together with the recently reported interaction between AKAP8L and mTORC1 (*65*), indicate that AKAP8L could be implicated in ASDs by bridging PTEN to key players in the mTOR pathway in human neurons. A previous study of a family with autism and macrocephaly has also identified AKAP8L as a likely candidate gene affected by the disease-associated genomic alteration at 19p13.12 (*66*), further supporting the hypothesis that the PTEN-AKAP8L interaction may influence neurodevelopmental phenotypes through the mTOR signaling cascades.

Our approach of generating a combined PPI network of multiple ASD index proteins also allowed us to experimentally identify functional convergence of these proteins based on their shared interactions (*16, 34, 67*). The hypothesis that ASD risk genes may act in a small number of networks has been long debated (*15, 68*), and previous studies have only identified broad functional categories (e.g., "gene expression regulation", "neuronal communication", and "cytoskeleton") to be enriched among known ASD risk genes (*10*)**.** In this study, we identified a number of proteins involved in RNA metabolism as being highly interconnected to multiple index proteins. Intriguingly, we found three components of the same complex (i.e., IGF2BP1, IGF2BP2, and IGF2BP3) to be among the most recurrent proteins in the network, each interacting with 5-7

index proteins. Insulin-like growth factor 2 mRNA-binding proteins 1-3 (IGF2BP1-3) belong to a conserved family of single-stranded RNA-binding proteins, and were recently characterized as part of a N6-methyladenosine (m6A)-reader complex broadly involved in post-transcriptional regulation by impacting mRNA stability and translation rates (*44*). Despite that their role in modulating individual neural stem cell growth and division has been explored in *Drosophila* (*69*), there is no previous evidence of their genetic associations to ASDs, which may be due to that fact at least two components of the complex, IGF2BP1 and IGF2BP3, are intolerant of LoF mutations in humans (*52*). Furthermore, when we analyzed published RNA targets of IGF2BP1-3 (*44*), we found that they are significantly enriched for ASD risk genes identified through exome sequencing, common variant risks of ASDs, as well as genes in our PPI network. In particular, we found the DYRK1A interactome to be strongly enriched for IGF2BP1-3 targets, and showed that partial knockdown of DYRK1A led to increased expression of IGF2BP1-3. These results suggest that the IGF2BP1-3 complex and DYRK1A might be part of a feedback mechanism reinforcing the expression of ASD-relevant transcripts. Importantly, these findings also showcase how convergent signals in the PPI network can be used to identify ASD-relevant genes and pathways that large-scale genetic studies may fail to nominate due to either early embryonic lethality or sampling/statistical power limitations.

Finally, we show that the network is indeed statistically enriched for rare variant risks of ASDs and developmental disorders captured in exome sequencing data (*10, 50*). Interestingly, the enrichment appears to be specific to these developmental phenotypes, as the same degree of enrichment was not observed for SCZ (*51*). Given these results, we integrated the network with ASD exome sequencing data in a 'social Manhattan plot' to demonstrate that it can be used to prioritize risk genes with suggestive significance in the genetic analysis, based on their interactions

with ASD index proteins in the network. Although we did not observe an enrichment for common variant risk of ASDs in the network, it is possible that this is due to statistical power limitations of the most recent ASD GWAS; we expect that with more data in the future, our network will also enable the prioritization of risk genes from GWAS data. This expectation is supported by our findings in a parallel study (co-submitted manuscript), where we applied the same brain cell type-specific interaction proteomics framework to generate an iN-derived PPI network for SCZ risk genes prioritized from GWAS loci. As well-powered, multi-ancestry GWAS datasets are available for SCZ, we found that the SCZ PPI network is enriched for common variant risk of SCZ in both Europeans and East Asians, and that the network can complement conventional strategies such as statistical fine-mapping and eQTL co-localization analyses to prioritize additional risk genes from GWAS loci. Repeating similar analyses for ASDs when larger GWAS datasets become available in the future may allow us to better determine if rare and common variant risks of ASDs generally converge onto shared pathways represented in the ASD PPI network, as we have already seen specifically for the transcriptional circuit regulated by IGF2BP1-3.

In conclusion, we built a PPI network of 13 index proteins encoded by high-confidence ASD risk genes in human iPSC-derived excitatory neurons. This network, which contains > 90% newly characterized interactions, is enriched for genes expressed in brain tissues and neuronal cell types, transcriptionally perturbed genes in the brains of idiopathic ASD patients, as well as genes containing rare deleterious mutations associated with ASDs and developmental disorders. We performed follow-up experiments and analyses to demonstrate that the network contains rich information for uncovering ASD-related biology, both in terms of investigating the individual PPIs (e.g., giant ANK2-specific PPIs and PTEN-AKAP8L interaction) and the convergent signals (e.g., IGF2BP1-3 complex) in the network. Overall, our findings not only bring clarity on how certain

risk genes contribute to ASDs, but also suggest convergent biology and shared pathways that may inform therapeutic intervention or patient stratification strategies. Going forward, the approach described in this study and the parallel SCZ study (co-submitted manuscript) may be adapted and expanded to generate PPI networks for additional disease risk genes in various brain cell types in order to compile a comprehensive cell type-specific PPI atlas for psychiatric disorders.

References

1. E. B. Robinson, B. M. Neale, S. E. Hyman, Genetic research in autism spectrum disorders. *Curr Opin Pediatr* **27**, 685-691 (2015).
2. B. S. Abrahams, D. H. Geschwind, Advances in autism genetics: on the threshold of a new neurobiology. *Nat Rev Genet* **9**, 341-355 (2008).
3. T. Gaugler *et al.*, Most genetic risk for autism resides with common variation. *Nat Genet* **46**, 881-885 (2014).
4. I. Iossifov *et al.*, The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515**, 216-221 (2014).
5. J. Sebat *et al.*, Strong association of de novo copy number mutations with autism. *Science* **316**, 445-449 (2007).
6. S. J. Sanders *et al.*, Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* **87**, 1215-1233 (2015).
7. D. J. Weiner *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).

8. E. B. Robinson *et al.*, Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nat Genet* **48**, 552-555 (2016).

9. D. H. Geschwind, Genetics of autism spectrum disorders. *Trends Cogn Sci* **15**, 409-416 (2011).

10. F. K. Satterstrom *et al.*, Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell* **180**, 568-584 e523 (2020).

11. X. Jin *et al.*, In vivo Perturb-Seq reveals neuronal and glial abnormalities associated with autism risk genes. *Science* **370**, (2020).

12. S. De Rubeis *et al.*, Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209-215 (2014).

13. D. Pinto *et al.*, Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *Am J Hum Genet* **94**, 677-694 (2014).

14. D. H. Geschwind, Autism: many genes, common pathways? *Cell* **135**, 391-395 (2008).

15. J. M. Berg, D. H. Geschwind, Autism genetics: searching for specificity and convergence. *Genome Biol* **13**, 247 (2012).

16. N. Safari-Alighiarloo, M. Taghizadeh, M. Rezaei-Tavirani, B. Goliaei, A. A. Peyvandi, Protein-protein interaction networks (PPI) and complex diseases. *Gastroenterol Hepatol Bed Bench* **7**, 17-31 (2014).

17. S. Choobdar *et al.*, Assessment of network module identification across complex diseases. *Nat Methods* **16**, 843-852 (2019).

18. K. Lage, Protein-protein interactions and genetic diseases: The interactome. *Biochim Biophys Acta* **1842**, 1971-1980 (2014).

19. A. Lundby *et al.*, Annotation of loci from genome-wide association studies using tissue-specific quantitative interaction proteomics. *Nat Methods* **11**, 868-874 (2014).

20. B. J. O'Roak *et al.*, Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* **485**, 246-250 (2012).

21. B. M. Neale *et al.*, Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* **485**, 242-245 (2012).

22. T. Li *et al.*, GeNets: a unified web platform for network-based genomic analyses. *Nat Methods* **15**, 543-546 (2018).

23. T. Li *et al.*, A scored human protein-protein interaction network to catalyze genomic interpretation. *Nat Methods* **14**, 61-64 (2017).

24. G. Pintacuda *et al.*, Genoppi is an open-source software for robust and standardized integration of proteomic and genetic data. *Nat Commun* **12**, 2580 (2021).

25. N. N. Parikshak *et al.*, Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* **155**, 1008-1021 (2013).

26. C. Hartl *et al.*, The architecture of brain co-expression reveals the brain-wide basis of disease susceptibility. *bioRxiv*, (2020).

27. X. Xu, A. B. Wells, D. R. O'Brien, A. Nehorai, J. D. Dougherty, Cell type-specific expression analysis to identify putative cellular mechanisms for neurogenetic disorders. *J Neurosci* **34**, 1420-1431 (2014).

28. I. Voineagu *et al.*, Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474**, 380-384 (2011).

29. G. Ramaswami *et al.*, Integrative genomics identifies a convergent molecular subtype that links epigenomic with transcriptomic differences in autism. *Nat Commun* **11**, 4873 (2020).

30. D. Velmeshev *et al.*, Single-cell genomics identifies cell type-specific molecular changes in autism. *Science* **364**, 685-689 (2019).

31. R. Nehme *et al.*, Combining NGN2 Programming with Developmental Patterning Generates Human Excitatory Neurons with NMDAR-Mediated Synaptic Transmission. *Cell Rep* **23**, 2509-2523 (2018).

32. Y. Zhang *et al.*, Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron* **78**, 785-798 (2013).

33. A. Comella-Bolla *et al.*, Human Pluripotent Stem Cell-Derived Neurons Are Functionally Mature In Vitro and Integrate into the Mouse Striatum Following Transplantation. *Mol Neurobiol* **57**, 2766-2798 (2020).

34. K. Lage *et al.*, A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat Biotechnol* **25**, 309-316 (2007).

35. H. J. Kang *et al.*, Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483-489 (2011).

36. H. K. Finucane *et al.*, Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat Genet* **50**, 621-629 (2018).

37. F. Koopmans *et al.*, SynGO: An Evidence-Based, Expert-Curated Knowledge Base for the Synapse. *Neuron* **103**, 217-234 e214 (2019).

38. R. Yang *et al.*, ANK2 autism mutation targeting giant ankyrin-B promotes axon branching and ectopic connectivity. *Proc Natl Acad Sci U S A* **116**, 15262-15271 (2019).

39. M. J. Gandal *et al.*, Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**, (2018).

40. E. F. Spence, D. J. Kanak, B. R. Carlson, S. H. Soderling, The Arp2/3 Complex Is Essential for Distinct Stages of Spine Synapse Maturation, Including Synapse Unsilencing. *J Neurosci* **36**, 9696-9709 (2016).

41. A. Carracedo, P. P. Pandolfi, The PTEN-PI3K pathway: of feedbacks and cross-talks. *Oncogene* **27**, 5527-5541 (2008).

42. F. Zahedi Abghari, Y. Moradi, M. Akouchekian, PTEN gene mutations in patients with macrocephaly and classic autism: A systematic review. *Med J Islam Repub Iran* **33**, 10 (2019).

43. A. K. Tilot, T. W. Frazier, 2nd, C. Eng, Balancing Proliferation and Connectivity in PTEN-associated Autism Spectrum Disorder. *Neurotherapeutics* **12**, 609-619 (2015).

44. H. Huang *et al.*, Recognition of RNA N(6)-methyladenosine by IGF2BP proteins enhances mRNA stability and translation. *Nat Cell Biol* **20**, 285-295 (2018).

45. J. Grove *et al.*, Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet* **51**, 431-444 (2019).

46. C. A. de Leeuw, J. M. Mooij, T. Heskes, D. Posthuma, MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* **11**, e1004219 (2015).

47. A. Duchon, Y. Herault, DYRK1A, a Dosage-Sensitive Gene Involved in Neurodevelopmental Disorders, Is a Target for Drug Development in Down Syndrome. *Front Behav Neurosci* **10**, 104 (2016).

48. T. Dang *et al.*, Autism-associated Dyrk1a truncation mutants impair neuronal dendritic and spine growth and interfere with postnatal cortical development. *Mol Psychiatry* **23**, 747-758 (2018).

49. P. Fernandez-Martinez, C. Zahonero, P. Sanchez-Gomez, DYRK1A: the double-edged kinase as a protagonist in cell growth and tumorigenesis. *Mol Cell Oncol* **2**, e970048 (2015).

50. J. Kaplanis *et al.*, Integrating healthcare and research genetic data empowers the discovery of 28 novel developmental disorders. *bioRxiv*, (2020).

51. T. Singh, B. M. Neale, M. J. Daly, Exome sequencing identifies rare coding variants in 10 genes which confer substantial risk for schizophrenia. *medRxiv*, (2020).

52. K. J. Karczewski *et al.*, The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434-443 (2020).

53. D. Demontis *et al.*, Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet* **51**, 63-75 (2019).

54. E. A. Stahl *et al.*, Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat Genet* **51**, 793-803 (2019).

55. D. M. Howard *et al.*, Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat Neurosci* **22**, 343-352 (2019).

56. Schizophrenia Working Group of the Psychiatric Genomics Consortium, Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-427 (2014).

57. L. Yengo *et al.*, Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. *Hum Mol Genet* **27**, 3641-3649 (2018).

58. M. Losh, D. Childress, K. Lam, J. Piven, Defining key features of the broad autism phenotype: a comparison across parents of multiple- and single-incidence autism families. *Am J Med Genet B Neuropsychiatr Genet* **147B**, 424-433 (2008).

59. R. Keller, R. Basta, L. Salerno, M. Elia, Autism, epilepsy, and synaptopathies: a not rare association. *Neurol Sci* **38**, 1353-1361 (2017).

60. R. M. Busch *et al.*, Neurobehavioral phenotype of autism spectrum disorder associated with germline heterozygous mutations in PTEN. *Transl Psychiatry* **9**, 253 (2019).

61. D. V. Hansen, J. L. Rubenstein, A. R. Kriegstein, Deriving excitatory neurons of the neocortex from pluripotent stem cells. *Neuron* **70**, 645-660 (2011).

62. V. Busskamp *et al.*, Rapid neurogenesis through transcriptional activation in human stem cells. *Mol Syst Biol* **10**, 760 (2014).

63. F. B. Russo *et al.*, The use of iPSC technology for modeling Autism Spectrum Disorders. *Neurobiol Dis* **130**, 104483 (2019).

64. V. Bennett, K. Walder, Evolution in action: giant ankyrins awake. *Dev Cell* **33**, 1-2 (2015).

65. C. H. Melick, D. Meng, J. L. Jewell, A-kinase anchoring protein 8L interacts with mTORC1 and promotes cell growth. *J Biol Chem* **295**, 8096-8105 (2020).

66. R. A. Nebel *et al.*, Reciprocal Relationship between Head Size, an Autism Endophenotype, and Gene Dosage at 19p13.12 Points to AKAP8 and AKAP8L. *PLoS One* **10**, e0129270 (2015).

67. V. Janjic, N. Przulj, Biological function through network topology: a survey of the human diseasome. *Brief Funct Genomics* **11**, 522-532 (2012).

68. M. J. Gandal *et al.*, Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science* **359**, 693-697 (2018).

69. T. J. Samuels, A. I. Jarvelin, D. Ish-Horowicz, I. Davis, Imp/IGF2BP levels modulate individual neural stem cell growth and division through myc mRNA stability. *Elife* **9**, (2020).

70. A. Kechin, U. Boyarskikh, A. Kel, M. Filipenko, cutPrimers: A New Tool for Accurate Cutting of Primers from Reads of Targeted Next Generation Sequencing. *J Comput Biol* **24**, 1138-1143 (2017).

71. A. D. Yates *et al.*, Ensembl 2020. *Nucleic Acids Res* **48**, D682-D688 (2020).

72. D. Kim, J. M. Paggi, C. Park, C. Bennett, S. L. Salzberg, Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**, 907-915 (2019).

73. M. Pertea *et al.*, StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**, 290-295 (2015).

74. S. Anders, P. T. Pyl, W. Huber, HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169 (2015).

75. M. D. Robinson, A. Oshlack, A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **11**, R25 (2010).

76. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).

77. T. Stuart *et al.*, Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902 e1821 (2019).

78. S. Tyanova *et al.*, The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* **13**, 731-740 (2016).

79. M. E. Ritchie *et al.*, limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).

80. M. Ashburner *et al.*, Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29 (2000).

81. The Gene Ontology Consortium, The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res* **47**, D330-D338 (2019).

82. The 1000 Genomes Project Consortium, A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).

Acknowledgements

We would like to thank Hailiang Huang and Ruize Liu for fruitful discussions and analytical advice, and Steve Hyman for helping conceptualize this study and critically reading the manuscript. We thank the Broad Proteomics Platform, Whitehead Proteomics Core Facility, and Harvard Center for Mass Spectrometry for discussions and generation of IP-MS data.

Funding

This work was supported by grants from the Stanley Center for Psychiatric Research, the US National Institute of Mental Health (R01 MH109903 and U01 MH121499), the Simons Foundation Autism Research Initiative (awards 515064 and 735604), the Lundbeck Foundation

(R223-2016-721 and R350-2020-963), the US National Institute of Diabetes and Digestive and Kidney Diseases (U01 DK078616 and T32 DK110919), and a Broad Next10 grant.

Author Contributions

GP and JR carried out tissue culture, GP, JMM and JR tested antibodies and ran immunoblots. GP and KWL executed IP experiments. KWL and MAG ran MS experiments and analysis at CNCR. GP performed the rest of the experiments. KT performed bulk RNA-seq, IP-MS, GTEx, and BrainSpan analyses. YHH performed IP-MS and the rest of the analyses. ABS, NF, KCE, and KL supervised and managed the study. KCE and KL designed the study. KL initiated and led the study. GP, YHH, KT, NF, KCE, and KL wrote the manuscript with input from co-authors.

Declaration of Interests

KCE is a co-founder of Q-State Biosciences, Quralis and Enclear, and currently employed at BioMarin Pharmaceutical. Other authors declare no competing interests.

Data Availability

Raw RNA-seq and IP-MS data will be available through GEO and MassIVE, respectively, upon publication of the manuscript.