Protein-Protein Interaction Landscape of Autism Spectrum Disorder (ASD)

by
Zun Zar Chi Naing

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Biological and Medical Informatics

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
Nevan Krogan
071F8FAE3EA8493...

Nevan Krogan
Chair

DocuSigned by:
Lauren Weiss
2F1E8F8D0A65...

Lauren Weiss

DocuSigned by:
Hao Li
C2CA172BE8684DD...

Hao Li

Committee Members

DEDICATION

This work is dedicated to my parents, U Hla Win and Daw Pheik Tin, whose sacrifice, love, and patience made this possible.

ACKNOWLEDGEMENTS

It is an absolute privilege and honor to be the first person in my family to pursue a PhD, and I cannot emphasize enough that I am where I am today only because of the love, support, and encouragement that my family, friends, and mentors have shown me. My journey of pursuing the PhD has not been the smoothest sailing one but it has led me to grow tremendously as a person and broadened my horizon for different scientific career possibilities.

I will first start off by thanking my PI, my boss, the ultimate science guru – Nevan Krogan for taking me under his wing and giving me endless opportunities to take on this complex monster of a project. I thank you, Nevan, for taking a chance on this bright-eyed student who was ready to dive in this wonderful world of systems biology using proteomics. And, also for introducing me to Ruth Huttenhain, who became my supervisor during my PhD training. Ruth, I owe everything I know about proteomics to you, and I thank you for your patience, kindness, mentorship, and compassion you have shown me throughout the years. You have always encouraged and motivated me during my difficult times. You truly are an inspirational person, I always marvel at how much passion, intensity, and consistency you show for everything that you take on, be it science, leadership, mentorship, or any outdoor activities. Many other Krogan lab members also play an instrumental role during my time in the lab as well, especially Kristen Obernier, Rasika Vartak, Jiewei Xu, Manon Eckhardt, Mehdi Bouhaddou, Yuan Zhou, Ben Polacco, Peggy Ackerberg, Vincenzo Pierotti, Kelsey Haas, and Reuben Hogan. Yuan and Ben, thanks for all the help with coding in R! Our weekly computational group meetings brought me so much joy and knowledge. Kelsey, I cannot thank you enough for giving me time and space to be vulnerable. Your kindness, and compassion truly have been a true source of warmth during the hard patches of my journey.

I also would like to thank my collaborators, Belinda Wang, Fima Zaltsman, Sam Drake, Vanessa Drury, Silvano Gonzalez and Jeremy Willsey from AJ Willsey and State labs, Micaela

Lasser and Helen Willsey from H. Willsey lab. You all have been very cordial, and it has been a pleasure to work with you all.

I am also grateful to have such wonderful mentors at UCSF. I want to thank my thesis committee members – Lauren Weiss and Hao Li. Laurie Weiss, having you on my thesis committee is one of the best decisions I made during my PhD career. I am truly grateful for your guidance on my project and for being an advocate. Hao Li, I have always enjoyed our conversations, every time I talk to you, I am reminded of how exciting basic science research is. I also want to thank Rong Wang and Sophie Dumont for being amazing mentors, lending an ear in my times of despair. It is always nice to know that there are people looking out for you.

I have built so many new friendships throughout my PhD career which have made the journey an enjoyable one. Michael Mobaraki, you have always been there for me throughout this entire process. You really are a supportive and encouraging friend. I treasure our long talks and walks all around San Francisco and I can always rely on you for a good dessert place. I am also grateful to have made strong friendships with Chris Mathy, Kelly Montgomery, Christa Caggiano, Matvei Khoroshkin, Yessica Gomez, and Nate Tran. Your friendships are one of the biggest highlights of my PhD journey.

The most important people I must thank is my family. Thank you, Papa and Mama, for daring and helping me to dream bigger than you have ever dreamt, making sacrifices so that I can pursue my scientific career goals. Thank you to all my sisters, Nay Chi Hla Win, Ei Ei Kyawt San and Moe Pwint Phyu, for being the kindest, smartest, cutest and most supportive sisters I can ever ask for. I am especially grateful for being able to be physically together with my family in the final years of my PhD training during COVID pandemic and coup d'état in Myanmar.

I am a better person and scientist because of all the help and support I received from everyone I met during my scientific journey. For that, I am blessed and grateful.

CONTRIBUTIONS

The work in this dissertation was performed under the supervision and guidance of Nevan Krogan and Ruth Huttenhain. This dissertation stems from a highly collaborative research effort within the members of National Institute of Mental Health (NIMH) U01 Psychiatric Cell Map Initiative (PCMI). Individuals contributing to this dissertation include members of the Krogan lab (UCSF, QBI, J. David Gladstone Institute), the Willsey lab (UCSF) and the State lab (UCSF).

The content in this dissertation is reformatted from a manuscript in preparation:

Zun Zar Chi Naing, Belinda Wang, Rasika Vartak, Fima Zaltsman, Jiewei Xu, Sam Drake, Vanessa Drury, Silvano Gonzalez, Micaela Lasser, Mehdi Bouhaddou, Helen Willsey, Kirsten Obernier, Jeremy Willsey*, Matt State*, Ruth Huttenhain*, Nevan J. Krogan*. Protein-Protein Interactions in Autism Spectrum Disorder. 2022. *Manuscript in preparation.*

*co-corresponding authors

Protein-Protein Interaction Landscape of Autism Spectrum Disorder (ASD)

Zun Zar Chi Naing

Abstract

Autism Spectrum Disorder (ASD) is a genetically complex and heterogenous neurodevelopmental disorder. As such, much effort has been put into uncovering the risk genes underlying ASD. A recent large-scale whole exome sequencing study focusing on *de novo* and case-control rare variants has identified 102 high-confidence ASD (hcASD) risk genes with a False Discovery Rate (FDR) ≤ 0.1(Satterstrom et al., 2020). Despite the advances in the discovery of ASD risk genes, we have yet to understand the molecular underpinnings of ASD pathobiology. To understand how hcASD risk genes contribute to ASD phenotypes, it is imperative to utilize integrative networks and systems biology approaches to unravel the molecular pathways connecting these hcASD risk genes. In this dissertation, I show how we used quantitative proteomics to systematically define the physical interaction landscape of proteins encoded by hcASD genes, and how these interactions are disrupted when we introduced *de novo* missense mutations as observed in the patients. The ASD protein-protein interaction network identifies 1024 unique proteins that interact with at least one of the 102 hcASD risk genes; of note, 82% of the interactions are novel. When we introduce patient-derived missense mutations in 35 out of 102 hcASD risk genes, we observed 133 protein interactions that are more specific to the mutants and 152 proteins that are more specific to the wild-type (WT) proteins. These differential protein interactions can be used to generate hypothesis regarding the molecular underpinnings of ASD etiology. Additionally, I present how we can elucidate biological processes, molecular pathways, and protein complexes from the generated protein-protein interaction network using network biology approaches and functional enrichment analyses, highlighting the convergent pathways that these high confident ASD risk genes may be involved in.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

ASD	Autism spectrum disorder
ASD_mut-PPI	Protein-protein interactions of *de novo* damaging missense mutations in autism spectrum disorder risk genes
ASD-PPI	Protein-protein interactions of autism spectrum disorder risk genes
FDR	False Discovery Rate
hcASD	High-confidence autism spectrum disorder risk genes
hcASD102	102 high-confident autism spectrum disorder risk genes at FDR ≤ 0.1 as identified in (Satterstrom et al., 2020)
hcASD255	255 high-confident autism spectrum disorder risk genes at FDR ≤ 0.1 as identified in (Fu et al., 2022)
HCIP	High-confidence interacting protein
MPC	Deleteriousness metric for missense mutations (Missense badness, PolyPhen-2, and Constraint) (Samocha et al., 2017)
OR	Odd ratio
pLI	Probability of being loss-of-function intolerant, a gene constraint metric for protein truncating variants (Lek et al., 2016)
PPI	Protein-protein interaction

PTV	Protein truncating variants
SCZ	Schizophrenia
WES	Whole exome sequencing

LIST OF SYMBOLS

ºC	Degrees Celsius
µg	Microgram
µL	Microliter
µm	Micrometer
x g	Times gravity (centrifugal force)
h	Hour(s)
min	minute(s)
s	second(s)

CHAPTER 1:

INTRODUCTION

1.1. Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder that affects 1-2% of the world's population, about 1 in 44 children, and accounts for the global burden of disease, particularly in developed countries (CDC, 2022). It is characterized by impairments in reciprocal social interactions as well as repetitive behaviors and/or restricted interests (American Psychological Association (APA), 2013). The symptoms of ASD may affect the patients' abilities to perform and function in routine areas of life, school, and work. As the name suggests, ASD is a *spectrum* disorder, meaning that two people diagnosed with ASD may have different symptoms and symptom severity. ASD can affect people of all races and genders, although there is a strong male sex bias with reports of 4:1 male to female ratio (Dougherty et al., 2022; Werling & Geschwind, 2013). Currently, there are no reliable biomarkers based diagnostic tests for ASD, and the diagnosis is based on a behavior and developmental evaluation by health care providers. ASD can be reliably diagnosed as early as the age of two (Hyman et al., 2020), although it can also be diagnosed later in life. ASD is a life-long disorder and as of this writing, no cure exists for it yet.

1.2. Uncovering the genetics underpinnings of ASD

Twin studies and sibling studies have consistently shown ASD to be one of the most heritable complex disorders among the neuropsychiatric disorders, with the heritability (h^2) between 70-90% (Sandin et al., 2017; Tick et al., 2016). This suggests that there is a strong genetic component contributing to the disorder and therefore, great efforts have been made to identify candidate genes for ASD. However, this has proven difficult due to complex genetic architecture exemplified by extreme locus and allelic heterogeneity (O'Roak & State, 2008), and

contribution from both common and rare variants (Gaugler et al., 2014; Klei et al., 2012). While common variants contribute the most risk for ASD, reproducible loci have not yet been identified, almost certainly due to insufficient sample sizes (Grove et al., 2017). Alternatively, identifying recurrent high-effect size rare structural and sequence variants—particularly *de novo* variants—has proved a reliable path for gene discovery (Buxbaum et al., 2012). Recent efforts in gene discovery studies, specifically exome sequencing studies focusing on rare variants in family studies, have resulted in the identification of 102 highly confident ASD risk genes at a false discovery rate (FDR) ≤ 0.1(Satterstrom et al., 2020). Excitingly, these genes and the *de novo* variants found in these genes are clear molecular clues that can be used to unravel the underlying biology. However, it is extraordinarily difficult to transition from a list of genetic variants to molecular understanding of ASD pathobiology. This is in part due to our incomplete knowledge of human brain development coupled with the genetic and phenotypic heterogeneity of ASD. Additionally, ASD risk genes are complex and pleiotropic, carrying out various biological functions that are developmental timeline and cell type dependent. A convergent framework is needed to pinpoint the conserved and converging function among multiple ASD risk genes (A. J. Willsey et al., 2018). Therefore, a critical step to understanding core ASD neuropathology and paving ways for therapies, is to delineate the roles of all the high confident ASD risk genes in brain development in a systemic manner.

1.3. Leveraging systems biology to delineate ASD etiology

Proteins never function on their own; all proteins interact with other proteins or themselves to carry out cellular functions. Studying proteins with their interacting partners can inform us which proteins are acting in concert to carry out a function. Generally, we can assume that if two proteins interact, they are physically located in the same subcellular compartment and are likely involved in a shared biological process (Bouhaddou et al., 2019; Swaney et al., 2021). If a protein has unknown localization and/or function, we can make inferences based on the known localization

or function of proteins with which it interacts. In addition, characterizing how disease-associated mutations alter the protein interaction landscape can help elucidate what cellular functions underlie disease pathology (Perica et al., 2021; Swaney et al., 2021).

In recent years, integrating systems biology approaches with protein-protein interaction data to understand complex diseases has proven fruitful in uncovering novel therapeutics. This was highlighted in the COVID pandemic when the situation called for urgent identification of druggable protein targets to ameliorate the effects of SARS-Cov2 infection; systems biology was leveraged to build host-pathogen interaction maps to systematically identify potential druggable sites at the host-pathogen interface (Gordon et al., 2020). Along the same ideology, we can use quantitative proteomics combined with systems biology to understand the functional roles of the high-confidence ASD (hcASD) risk genes.

The identification of ASD risk genes lays the foundation for uncovering ASD etiology. Here, using 102 hcASD risk genes (Satterstrom et al., 2020) as a starting point, we identified over 1000 high confident interactors in HEK293T cells using affinity-purification mass spectrometry (AP-MS) (Chapter 2). Applying various network ontology analyses to this protein-protein interaction (PPI) map, we identified functional relationships between the proteins of hcASD risk genes and transition from a list of risk genes to broader molecular mechanisms. Delineating the functional relationships among the proteins of hcASD risk genes allow us to generate testable hypotheses about functional convergence of these proteins. To this end, we pursued the function of two hcASD proteins that directly interact with each other in the ASD-PPI map, DYRK1A and KIAA0232. These two proteins also share multiple interaction partners as part of a protein complex. When knocking down *KIAA0232*, we also noted that it phenocopied the knock down of *DYRK1A*, resulting in the size reduction of telencephalon in *Xenopus tropicalis*. Lastly, we showed that we can quantitatively identify proteins that are gained or lost due to the introduction of *de novo* missense mutation. We focused on one of the ASD risk genes – *FOXP1* and its missense mutations. We found that missense mutations in FOXP1 leads to a loss in

interaction with FOXP4 and impaired neurite outgrowths. Collectively, this systematic molecular dissection of hcASD risk genes acts as a foundational resource to find target proteins that can potentially be targeted for biomarkers or therapies for ASD.

CHAPTER 2:

Defining the protein-protein interaction landscape of high confidence ASD risk genes

2.1. Introduction

Autism Spectrum Disorder (ASD) is a highly heritable complex neurodevelopmental disorder. A recent large-scale whole exome sequencing study focusing on *de novo* and rare variants identified 102 high-confidence ASD (hcASD) risk genes (Satterstrom et al., 2020). To reconcile the genetic diversity of ASD into convergent biological pathways, we can systematically define the protein-protein interaction landscape of proteins encoded by hcASD genes using affinity-purification mass spectrometry (AP-MS). The reasoning is that physical interactions between proteins often indicate shared biological function, and protein-protein interaction (PPI) maps can reveal shared functions of proteins, including the protein complexes and cellular pathways in which they participate. Here, we use AP-MS in HEK293T cells to generate a PPI network map of the proteins encoded by the 102 hcASD risk genes identified in (Satterstrom et al., 2020) (**Figure 2.1A**). From this PPI map, we can generate hypotheses regarding the functions of these different proteins, the cellular pathways and processes they participate in and how they are physically and functionally connected, reconciling the genetic diversity of ASD.

2.2. Materials and Methods

2.2.1. Cloning hcASD risk genes

The coding sequence of each hcASD risk gene was cloned into a pcDNA4 plasmid with either N- or C-terminal 2xStrep tags, which encode the bait proteins for the AP-MS study. The terminus position of the tags was determined so that the tag will not interfere with the protein function based on the prior reported plasmids that have been used in functional studies. The isoforms for the hcASD risk gene were chosen based on high brain expression levels as well as

the high frequency of mutations observed in ASD using Clonotator (http://ec2-52-91-98-53.compute-1.amazonaws.com/login/). All constructs were sequence validated.

2.2.2. Cell culture

HEK293T cells were cultured in Dulbecco's modified Eagle's medium (DMEM; Corning) supplemented with 10% fetal bovine serum (FBS; Gibco, Life Technologies) and 1% penicillin–streptomycin (Corning). All cells were maintained in a humidified incubator at 37 °C with 5% CO_2.

2.2.3. Transfection

Each transfection (102 baits, one GFP control and one empty vector control) was carried out in a 15cm dish with 10 million HEK293T cells (70-80% confluency), with three biological replicates per bait. Transfections were split in 15 batches, with three biological replicates of GFP and empty vector controls included in each batch. For each transfection, 15 µg of Strep-tagged plasmids was combined with PolyJet Transfection Reagent (SignaGen Laboratories) at a 1:3 µg:µL ratio of plasmid:transfection reagent, incubated at room temperature for 10 mins, and added dropwise to HEK293T cells. About 48h post transfection, cells were resuspended at room temperature using 10 ml Dulbecco's phosphate-buffered saline without calcium and magnesium (DPBS) supplemented with 10 mM EDTA, followed by centrifugation at 200g, 4 °C for 5 min. Cell pellets were frozen on dry ice and stored at −80 °C.

2.2.4. Affinity purification

The cell pellets were thawed on ice and then lysed with 1 ml ice-cold lysis buffer (IP buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1 mM EDTA, 0.5% Nonidet P40 substitute (NP40; Fluka Analytical), cOmplete mini EDTA-free protease and PhosSTOP phosphatase inhibitor cocktails (Roche)). Samples were then flash-frozen on dry ice for about 10 min and partially thawed at 37°C for 30-45s before incubation for 30 min at 4 °C on a tube rotator. Lysates were centrifuged

at 13,000g, 4 °C for 15 min to clarify lysate and pellet debris. A 50 µl lysate was reserved at this point for future experiments such as Western blot. The remaining lysates underwent automated affinity purification on the KingFisher Flex Purification System (Thermo Scientific), which was first equilibrated to 4°C in the cold room. First, MagStrep 'type3' beads (30 µl; IBA Lifesciences) were equilibrated twice with 1 ml wash buffer (IP buffer supplemented with 0.05% NP40) and incubated with 0.95 ml lysate for 2 h. Beads were washed 3 times with 1 ml wash buffer and once with 1 ml IP buffer, then resuspended in 50 µl denaturation–reduction buffer (2 M urea, 50 mM Tris-HCl pH 8.0, 1 mM DTT) and 50 µl 1× buffer BXT (IBA Lifesciences) and dispensed into a single 96-well KingFisher microtitre plate. Purified proteins were eluted at room temperature for 30 min with constant shaking at 1,100 rpm on a ThermoMixer C incubator.

2.2.5. On-bead digestion

Bead-bound proteins were denatured and reduced at 37 °C for 30 min, brought back to room temperature, alkylated in the dark with 3 mM iodoacetamide for 45 min and quenched with 3 mM DTT for 10 min. To offset evaporation, 15 µl 50 mM Tris-HCl, pH 8.0 were added and proteins were digested with 1.5 µl trypsin (0.5 µg/µl; Promega) with constant shaking at 1,100 rpm and incubation at 37 °C on a ThermoMixer C incubator for 4 h, and again for 1–2 h with 0.5 µl additional trypsin. Resulting peptides were combined with 50 µl 50 mM Tris-HCl, pH 8.0 to rinse beads and then, acidified with trifluoroacetic acid (0.5% final, pH < 2.0). Desalting was conducted in a BioPureSPE Mini 96-Well Plate (20 mg PROTO 300 C18; The Nest Group) according to standard protocols.

2.2.6. Sequential Strep-FLAG purifications

Sequential purifications initially followed single-step purifications as described above, but are scaled to improve protein enrichment. Alterations are as follows: for each purification, five 14.5-cm dishes were co-transfected and subsequently combined for lysis with proportional

volumes of buffer > 40 hours post-transfection. Cell lysates were then divided between five separate Strep-tag purifications. Following wash steps, MagStrep beads were combined and incubated on a tube rotator for 30 minutes with 0.3 ml Buffer BXT supplemented with 0.38% NP40 to elute proteins. After removing eluted protein, MagStrep beads were rinsed with an additional 0.3 ml Buffer BXT supplemented with 0.38% NP40. Buffer BXT elution and wash steps were combined and added to 30 µl equilibrated FLAG beads. After incubating for 2 hours and washing beads as above, proteins were eluted in 50 µl 0.1 mg/ml 3xFLAG peptide in IP Buffer supplemented with 0.05% RapiGest by gentle shaking on a vortex mixer at room temperature for 30 minutes. 25 µl was submitted for mass spectrometry.

2.2.7. MS data acquisition and analysis

To prepare for mass spectrometry, samples were resuspended in 4% formic acid, 2% acetonitrile solution, and separated by a reversed-phase gradient over a Nanoflow C18 column (Dr Maisch). Each sample was directly injected via an Easy-nLC 1200 (Thermo Fisher Scientific) into a Q-Exactive Plus mass spectrometer (Thermo Fisher Scientific) and analyzed with a 75 min acquisition, with all MS1 and MS2 spectra collected in the orbitrap; data were acquired using the Thermo software Xcalibur (4.2.47) and Tune (2.11 QF1 Build 3006). For all acquisitions, QCloud was used to control instrument longitudinal performance. All proteomic data were searched against the human proteome (UP000005640_9606, downloaded in October 2021) using the default settings for MaxQuant software (version 1.6.12.0) (Cox & Mann, 2008) with match-between-runs (MBR) feature turned on. Briefly, the MBR algorithm annotates unidentified peaks by assessing and comparing the retention times of the identified peaks in an MS1 spectrum. Detected peptides and proteins were filtered to 1% false-discovery rate in MaxQuant, and identified proteins were then subjected to protein–protein interaction scoring with both SAINTexpress (v.3.6.3)(Teo et al., 2014) and CompPASS (version 0.0.0.9000) (Huttlin et al., 2015; Sowa et al., 2009).

2.2.8. Removal of carryover effect and identification of high-confidence interactors

A crucial step in our AP-MS study is the probabilistic scoring of all quantified proteins in the dataset to identify high-confidence interaction proteins (HCIP) of the ASD risk genes. To do this, the scoring outputs from existing computational tools – CompPASS (Comparative Proteomic Analysis Software Suite) (Huttlin et al., 2015; Sowa et al., 2009) and SAINTexpress (Teo et al., 2014) – were systematically combined. These two scoring algorithms are widely used in the proteomics community to score the quality of protein-protein interactions as distinct from background.

Input files – bait, prey and interaction files - for SAINTexpress were made using *artmsEvidenceToSaintExpress* function in an open-sourced R package, *artMS* (Jimenez-Morales et al., 2022). The interaction input file for SAINTexpress was reformatted to be used as a CompPASS input file. As CompPASS requires all baits to have the same number of preys, we defined spectral counts for the union of all identified proteins across all AP-MS experiments for each bait; if the bait-prey interaction was not detected, its spectral count was assigned to be zero. This bait-prey spectral count matrix was used to search for carryover effect, which was defined as proteins with a drastic and continuous decrease in spectral counts in replicates where such proteins were used as baits in the previous sample injections (**Figure S2.3**). The resulting bait-prey spectral count data table after carryover removal was reformatted and used as the input file for SAINTexpress and CompPASS.

To determine the scoring cutoffs that capture the highest number of true interactions, a gold standard set of protein-protein interactions was manually defined. Gold standard interactions were extracted from publicly available large-scale PPI databases, Hein et al (Hein et al., 2015)., InWeb (T. Li et al., 2017), CORUM (Core Corum Complexes, Corum 3.0) (Giurgiu et al., 2019a), BioPlex (Huttlin et al., 2015), and BioGrid (Oughtred et al., 2018), with some additional filtering steps for BioGrid and InWeb. As BioGrid contains interactions from various sources, including

both experimental and predicted, the dataset was filtered to only interactions that were attained using another experimental method in addition to the AP-MS method. Similarly, InWeb database were filtered for interactions that have high-confidence score of > 0.95 and are identified through an experimental method.

SAINTexpress (version 3.6.3) was run batch-wise, each batch with its own empty-vector and GFP replicates as controls. The outputs were then concatenated to a single SAINTexpress output. CompPASS was run using the R package, *cRomppass* (https://github.com/dnusinow/cRomppass). The scoring outputs from SAINTexpress and CompPASS were merged for each interaction (**Figure 2.1B**). The SAINTexpress Bayesian False Discovery Rate (BFDR) and CompPASS WD score (rank_WD, ranked from 0 to 1 across the dataset) were used as prediction confidence scores. Using the ranked CompPASS WD score (rank_WD) as a predictive value and interactions found in the previously described manually curated gold standard PPI database as true positives, the optimal rank_WD score cutoff for each SAINTexpress BFDR increment was determined by calculating Youden's index. The best performing composite score of SAINTexpress and CompPASS (out of 78 composite scores) was determined by calculating precision, recall and F1 scores (**Figure 2.1C**).

2.2.9. Pathway and molecular complex overrepresentation analysis

The ASD-PPI network was accessed for enrichment of GO biological process terms and CORUM core protein complexes at the bait (and associated prey proteins) level. CORUM complete complexes (CORUM 4.0 release) (Giurgiu et al., 2019b) were downloaded as text file and converted to GMT format. All the ontology terms for GO were obtained using the *get_GO_data* function from the R package, *clusterProfiler* (Yu et al., 2012) as GMT format. The two GMT files were then combined, and an enrichment test based on hypergeometric distribution was run via the *enricher* function in *clusterProfiler*. From the output of enricher function, gene sets for significant complex and GO terms were selected using a false discovery rate cutoff of

1%. To refine and select for non-redundant gene sets, a GO term tree was first built based on distances (1 − Jaccard similarity coefficients of shared genes) between the significant terms. The GO term tree was cut at h = 0.99 to identify clusters of non-redundant gene sets. The term with the largest gene set size was selected if a bait was enriched for multiple significant terms belonging to the same cluster.

2.2.10. ASD-PPI protein expression in HEK293T cells

We defined the HEK293T proteome to be the n=11,133 proteins with detected protein expression in a published global quantitative mass spectrometry dataset (Bekker-Jensen et al., 2017). We defined the protein expression level to be the average intensity based absolute quantification (iBAQ) scores for the two HEK293T experimental replicates reported in Supplementary Table 7. We compared the relative expression levels of ASD-PPI bait, prey (excluding hcASD102 genes), and other HEK293T expressed proteins using t-test, correcting p-values for multiple hypothesis testing (p.adj = p x 3 tests).

2.2.11. Prey expression in adult brain tissue (GTEx) and enrichment analysis

The Genotype-Tissue Expression (GTEx) project includes RNA-sequencing data from 54 non-diseased tissue sites across nearly 1000 individuals (Lonsdale et al., 2013). We downloaded GTEx v8 data (dbGaP Accession phs000424.v8.p2) from GTEx website (https://www.gtexportal.org/home/datasets), which contains median gene-level TPM for samples from 54 tissue types, including 13 brain tissue sites.

We assessed whether ASD-PPI genes are more highly expressed in GTEx brain tissues compared to other HEK293T expressed proteins. We defined HEK293T expressed proteins as the union of proteins detected by mass spectrometry dataset (Bekker-Jensen et al., 2017) and prey. We defined 3 gene sets of interest – bait, prey (excluding hcASD102), and HEK293T expressed proteins that are not bait or prey (**Figure S2.9B**). For each gene set, we found the

median gene set expression percentile within individual brain tissue samples and performed t-tests to assess whether these were significantly different across gene sets. P-values were adjusted for multiple hypothesis testing (p.adj = p x 3 tests).

We assessed whether prey genes are more highly expressed in GTEx brain samples compared to permuted gene sets from the HEK293T proteome. We created 100,000 permuted gene sets from a universe of n=10,984 genes that are in HEK293T proteome, measured in GTEx RNA-sequencing data, and not ASD-PPI bait. The probability of gene selection was weighed by HEK293T iBAQ scores (Lonsdale et al., 2013), and permuted gene sets were required to have a median HEK293T iBAQ rank within 1 quantile of that of ASD-prey genes. We calculated the median gene set expression percentile for prey genes and the 100,000 permuted gene sets in GTEx brain tissue samples, where higher rank reflects higher relative expression within a sample. We then calculated the permuted significance of the observed ASD-prey GTEx expression percentile within each tissue type (**Figure S2.9C**).

2.2.12. Prey expression in adult brain tissue (BrainSpan) and enrichment analysis

The BrainSpan developmental RNAseq dataset (bsRNAseq) profiles human brain tissue RNA expression across the full course of human brain development, from early fetal stages through late adulthood (Kang et al., 2011). BsRNAseq data (n=524 samples) was downloaded from brainspan.org. The original 52,376 genes were trimmed to a final set of 18,552 protein-coding genes associated with an HGNC symbol; if a HGNC symbol was associated with multiple Ensembl gene IDs, only one was kept. Expression values are reported in reads per kilobase million (RPKM).

We assessed whether ASD-PPI genes are expressed at relatively higher levels in BrainSpan prenatal samples compared to HEK293T background genes. We defined 3 gene sets of interest – bait, prey (excluding hcASD102), and HEK293T expressed proteins that are not bait or prey (n= 10,046). We restricted our analysis to the n=237 prenatal brain samples as the

prenatal period has been implicated in ASD risk (Ben-David & Shifman, 2013; Parikshak et al., 2013; A. J. Willsey et al., 2013). For each gene set, we calculated the median gene set expression percentile within individual brain tissue samples. We performed t-tests to assess whether the median genset expression percentiles are significantly different between different gene sets, adjusting p-values for multiple hypothesis testing (p.adj = p x 3 tests) (**Figure 2.4A**).

We assessed whether prey genes are more highly expressed in BrainSpan prenatal samples compared to permuted gene sets from the HEK293T proteome. We created 100,000 permuted gene sets from HEK293T proteome genes, matching the number of prey genes. To generate the permuted gene sets, we selected genes from a set of n=10,902 genes that are 1) in the HEK293T proteome (Bekker-Jensen et al., 2017), 2) measured in BrainSpan RNAseq, and 3) not an ASD-PPI bait. The probability of gene selection was weighed by HEK293T protein expression level (median iBAQ rank), and only gene sets with a median iBAQ rank within 1 quantile of the median iBAQ rank of ASD-prey were retained. We calculated the median gene set rank ('medRank') within each bsRNAseq brain tissue sample for prey and each of the 100,000 permuted gene sets (higher medRank reflects higher gene set expression within a sample). For each brain tissue sample, we calculated the 'normalized medRank', defined as the observed medRank – median (100K permuted medRanks). The normalized medRank reflects the sample-level expression enrichment of prey, with positive values reflecting higher than expected expression. As a comparator, we also calculated the sample-level expression enrichment of hcASD102 genes (Satterstrom et al., 2020) compared to permuted gene sets.

To assess whether the relative expression of prey mirror that of hcASD102 genes across samples, we calculated the Spearman and Pearson correlations of the normalized medRanks for prey versus hcASD102 genes across n=524 brain samples. To compare relative prey (or hcASD102) expression in prenatal versus postnatal brain tissue, we grouped bsRNAseq brain samples by prenatal versus postnatal status (n=237 prenatal, n=287 postnatal), and assessed whether the normalized medRank of prey (or hcASD102) were significantly different between

prenatal versus postnatal samples using t-test. We also grouped bsRNAseq brain samples by developmental period (Kang et al., 2011) where periods 1-7 reflect prenatal stages of development and periods 8-15 reflect late infancy through late adulthood, resulting in 13 sample groups ranging from period 2-14. For each period, we calculated the median (normalized medRank) across samples for prey and hcASD102; we subsequently calculated the Spearman and Pearson correlation of median (normalized medRank) for prey and hcASD102 across the 13 period groups (**Figure 2.4B, S2.9D, S2.9E**).

2.2.13. Prey enrichment for ASD genetic risk

We used *de novo* genetic variants previously identified from simplex family studies as a measure of ASD genetic risk (Satterstrom et al., 2020) and conducted enrichment tests using Fisher's exact test to assess whether prey are enriched for ASD genetic risk compared to other genes in the HEK293T proteome or other exome genes. We focused on de novo damaging variants identified from the Simon's Simplex Collection (Satterstrom et al., 2020) Supplementary Table 1). We defined "damaging" variants to be variants that resulted in a damaging missense mutation (PolyPhen Mis3 (damaging); or MPC (MPC>=2)) or PTV (frameshift, stop gained, or canonical splice site disruption).

We determined the association between having a damaging variant in different gene sets of interest with ASD status. We defined the following gene sets:

1) Prey: ASD-PPI prey (n=1,074)
2) HEK293T (-Prey, -hcASD102): HEK293T proteome genes, excluding prey and hcASD102 (n = 10,045)
3) Exome: all autosomal genes measured in Satterstrom 2020 (n= 17,332)
4) Exome (-hcASD102): autosomal genes measured in Satterstrom 2020, excluding hcASD102 (n= 17,230)

5) Exome (-Prey, hcASD102): autosomal genes measured in Satterstrom 2020, excluding hcASD102 and prey (n= 16,234).

For each Fisher's enrichment test, the universe of genes was defined to be the gene set members, and the 2x2 contingency table was: X, the number of ASD probands with a damaging variant in at least one gene in the gene set; Y; the number of control siblings with a damaging variant in at least one gene in the gene set; Z, the number of ASD probands with no damaging variants in any gene in the gene set; W, the number of control siblings with no damaging variants in any gene in the gene set. We calculated the enrichment odds ratio with Fisher's exact test (one sided, alternative = greater), and adjusted p-values for multiple hypothesis testing (p.adj = p x number gene sets assessed) (**Figure 2.4C, S2.9F**).

2.2.14. Downsampling analyses

To determine whether increasing the number of baits used to create ASD-PPI is associated with increased ability to identify prey associated with ASD genetic risk, we downsampled the ASD-PPI network. Specifically, we randomly selected sets of bait (ranging from n = 1-100 bait, 1000 iterations for each set size), and trimmed the ASD-PPI network to include only prey associated with the downsampled bait. For each downsampled network, we calculated the association between having a damaging variant in different gene sets of interest with ASD status as described above. We evaluated 2 gene sets: 1) Prey (-hcASD102), and 2) Exome (-Prey, -hcASD102), which allows us to compare the amount of ASD-associated genetic risk in prey versus the remaining genes in the human exome. For each bait set size, we calculated the median and standard deviation of ORs across the 1000 iterations, as well as the median and standard error of p-values across the 1000 iterations (**Figure 2.4E**).

We additionally assessed whether increasing the number of baits used to create ASD-PPI is associated with increased ability to identify prey that are novel hcASD genes. We defined novel hcASD genes as the n=255 genes with FDR $\leq$ 0.1 in the latest omnibus WES study (Fu et al.,

2022). We downsampled the ASD-PPI network as described above. For each downsampled network, we 1) counted the number of hcASD255 genes in the prey and 2) conducted a Fisher's exact test (one sided, greater) to assess for enrichment of hcASD255 in prey. The counts of the 2x2 contingency tables were: X, the number of hcASD255 that are prey; Y, the number of hcASD255 genes that are not prey; Z, the number of prey that are not in hcASD255; and W the number of genes that are not prey or hcASD255. We defined the universe of possible genes to consist of those with 1) proteins that are detected in the HEK293T proteome (Bekker-Jensen et al., 2017); 2) measured in Satterstrom et al., 2020; and 3) measured in Fu et al., 2022. For each bait set size, we calculated the median and standard deviation of OR, -log10(p-value), and number of hcASD255 genes among prey across the 1000 iterations. We repeated this analysis after excluding hcASD102 genes (Satterstrom et al., 2020, FDR $\leq$ 0.1) from the gene universe to assess the ability of ASD-PPI to identify truly novel hcASD genes (**Figure 2.4F**).

2.2.15. Prey enrichment for ASD, DD, or SCZ risk genes

We evaluated whether prey genes are enriched for risk genes that have been implicated in ASD, DD, or schizophrenia (SCZ). We defined the following sets of risk genes,

1) ASD, Fu 2022 (Fu et al., 2022): n=255 autosomal genes with TADA FDR $\leq$ 0.1
2) ASD, Zhou 2022 (Zhou et al., 2022): n=72 genes with study-wide significance (based on 5,754 constrained genes, p<8.69E-6)
3) ASD, SFARI (Abrahams et al., 2013): SFARI genes are a database that endeavors to include all genes associated with ASD risk, regardless of the level or nature of evidence. SFARI genes are further divided into categories, including syndromic, category 1 (high confidence), category 2 (strong candidate), and category 3 (suggestive evidence). SFARI genes were downloaded from gene.sfari.org (09/02/2021 release), and included a total of n=1020 genes, of which there were n=230 syndromic genes, n = 92 syndromic and category 1 genes, n=206 category 1 genes, n=219 category 2 genes, and n=514 category 3 genes.

4) DD, Kaplanis 2022 (Kaplanis et al., 2022): n=285 genes that are significantly associated with DD after one-sided Bonferroni correction
5) SCZ, Singh 2022 (Singh et al., 2022): n=34 genes with TADA FDR $\leq$ 0.1.

We defined the HEK293T expressed genes to be the union of prey and proteins that were detected by mass spectrometry (Bekker-Jensen et al., 2017). We conducted two sets of enrichment tests. For the first set of enrichment tests, we assessed only risk genes from WES studies that reported the set of genes that were evaluated as possible risk genes (Fu et al., 2022; Singh et al., 2022; Zhou et al., 2022). We defined the universe of possible genes to be those that were 1) HEK293T expressed genes, 2) measured in Satterstrom et al., 2020, and 3) measured in the WES that generated the risk gene set of interest (Fu et al., 2022; Singh et al., 2022; Zhou et al., 2022) (**Figure 2.4D**). For the second set of enrichment tests (**Figure S2.9G**), we simply defined the gene universe to be HEK293T expressed genes, as the exact gene universe considered is unavailable for the SFARI risk gene sets. For both sets of enrichment tests, we conducted Fisher's exact tests (one sided, greater), in which the contingency tables were set up by prey status (yes/no) and disease gene set status (yes/no). We corrected p-values for multiple hypothesis testing (p.adj = p x number of gene sets assessed).

2.2.16. Prey gene constraint

We evaluated whether ASD-PPI genes are more constrained than expected by assessing several metrics, including the pLI (probability of being intolerant of a single loss-of-function variant), misZ (missense Z score, measures gene intolerance to missense variation), synZ (synonymous Z score, measures gene intolerance to synonymous variation, used as negative control), and s_het (selective effect for heterozygous PTVs) (Cassa et al., 2017; Lek et al., 2016).

We obtained pLI, misZ, and synZ scores from the Genome Aggegation Database (ExAC dataset (https://storage.googleapis.com/gcp-public-data--gnomad/legacy/exac_browser/forweb_cleaned_exac_r03_march16_z_data_pLI_CNV-

final.txt.gz) and s_het scores from Cassa et al., 2017 Supplementary Table 1. We assessed whether the pLI, misZ, s_het, and synZ scores of ASD-PPI bait genes, prey genes, and HEK293T proteome genes (Bekker-Jensen et al., 2017) were significantly different fpr ASD-PPI bait, prey (excluding hcASD102), and other HEK293T proteome genes using t-tests, adjusting p-values for multiple hypothesis testing (p.adj = p-value x number of t-tests).

2.2.17. *Xenopus tropicalis* embryos and tadpoles

Ovulation was induced by injection of human chorionic gonadotropin (Sigma) into the dorsal lymph sac according to standard procedure (Sive et al., 2000) and in accordance with approved UCSF IACUC protocols. Natural matings and *in vitro* fertilizations were performed. Embryos and tadpoles were staged by (Nieuwkoop et al., 1994). Clutch mates were used as matched controls.

2.2.18. *Xenopus tropicalis* microinjections

Xenopus tropicalis embryonic microinjections were performed as in (Sive et al., 2000). Microinjections were performed at the 2-cell stage using a Narishige micromanipulator, Parker Picospritzer, and Zeiss Stemi microscopes. Injection volume was calibrated with an eye-piece micrometer. Plasmids (DYRK1A-strep and KIAA0232-flag) were injected at 10 pg per embryo and mRNA (α-tubulin-mCherry) at 400 pg per embryo. Embryos were grown between 22–25°C in 1/9 Modified Ringer's (MR) solution, which was refreshed daily. Male and female embryos were analyzed.

2.2.19. *Xenopus* CRISPR/Cas9 genome editing

High-efficiency sgRNAs were designed, synthesized, and validated as in (H. R. Willsey et al., 2021). For each embryo, 3 ng of purified Cas9-NLS protein (Macrolabs, UC Berkeley), 800 pg sgRNA, and a dextran dye conjugated with Alexa-555 (Invitrogen) were injected into 1 cell of a 2-

cell stage embryo. The day following injection at stages 14–20, embryos were sorted left from right according to the dye.

2.2.20. *Xenopus* whole mount immunofluorescence staining

Xenopus immunostaining was carried out according to Willsey et al., 2018. The following primary antibodies were used: anti-Flag (1:100, Cell Signaling, 14793), anti-Strep (1:100, IBA, 2-1507), and anti-acetylated α-Tubulin (1:500, Abcam, ab179484). Secondary fluorescence-conjugated antibodies were used at 1:250 (Life Technologies, A32732, A32723 and A32733).

2.2.21. *Xenopus* microscopy and brain size measurements

Xenopus colocalization images were acquired on a Zeiss LSM980 confocal with Aryscan and a 63X objective. *Xenopus* tadpole images were acquired on a Zeiss Axio Zoom.V16 microscope with apotome and a 1X objective. Images were processed using Fiji (Schindelin et al., 2012). Telencephalon size was calculated from stereoscope images of brain immunostainings using the freehand select and measure functions in Fiji (Schindelin et al., 2012). The injected side was compared to the non-injected side (internal control). These measurements were from two-dimensional images taken from a dorsal perspective and reflect relative size differences. GraphPad Prism software version 8.3 was used to graph data and determine statistical significance using a student's paired t-test. p-values less than 0.05 were considered significant.

2.3. Results

2.3.1 Removing the carryover effect and combining two scoring algorithms leads to a robust PPI network

Carryover effect is a type of sample contamination where analytes from the previous run remain in the LC-MS system and are detected in following runs (Vu et al., 2013; Yamagaki & Yamazaki, 2019). It can last for multiple sequential runs and can be identified by a half-life like

pattern in a sequence of consecutive runs. This effect can be more prominent in experiments where the baits are overexpressed (**Figure S2.1**, **Figure S2.2**). Removing this carryover contamination effect prior to the scoring analysis is critical in our analysis especially as carryover contamination may result in the false positive identification of an interactor.

We use two scoring algorithms, CompPASS (Huttlin et al., 2015; Sowa et al., 2009) and SAINTexpress (Teo et al., 2014), to define HCIPs. CompPASS uses a non-parametric filtering approach which identifies HCIPs based on the uniqueness of a prey interaction across different baits by comparing the abundance, frequency, and reproducibility of the prey interactors in the data set. HCIPs are then identified using the normalized weighted D-Score and Z-score. Because of this frequentist approach, CompPASS works best when implemented on a large dataset such as ours, which includes 102 baits. However, shared interactors are expected among baits, and in this circumstance, CompPASS will penalize true interactors identified across multiple different baits, as frequently identified interactors are regarded as “background” in CompPASS analysis, which could lead to high false negatives. SAINTexpress uses a parametric filtering approach in which a background distribution is built based on negative control affinity purifications datasets, which are comprised of GFP and empty vector constructs in our case. This Bayesian approach increases sensitivity but assumes a robust distribution of the negative datasets, which could be biased towards false positives. Combining these two divergent scoring strategies leverages the strengths of both methods to increase the overall identification of true positives. In addition to combining the scores, benchmarking our ASD-PPI dataset against recovery of a gold standard set of known portein interactions is essential to select optimal scoring cutoffs to identify HCIPs. Indeed, using SAINTexpress score (1-BFDR) ≥ 0.95 and CompPASS score (rank_WD) ≥ 0.971, we recovered the highest density of known interactions (**Figure 2.1D**) while also discovering novel interactions (**Figure 2.1F**).

2.3.2 ASD-PPI data provides the most coverage of the ASD interactome

Using SAINTexpress score (1-BFDR) ≥ 0.95 and CompPASS score (rank_WD) ≥ 0.971, we defined a high-confidence ASD-PPI network that involves 1074 unique prey interactors connecting 100 hcASD risk genes via 1881 interactions (**Figure 2.3**). Notably, 87% of the 1881 interactions are novel, that is, not previously reported in any of these large-scale PPI databases, Hein et al (Hein et al., 2015)., InWeb (T. Li et al., 2017), CORUM (Core Corum Complexes, Corum 3.0) (Giurgiu et al., 2019a), BioPlex (Huttlin et al., 2015), and BioGrid (Oughtred et al., 2018) (**Figure 2.1F**). Additionally, compared to other large-scale PPI databases, our ASD-PPI network has the most comprehensive coverage of having hcASD risk gene related interactions, defined as the number of hcASD proteins reported as an interactor and the number of interactions involving a hcASD protein (**Figure 2.1E**).

2.3.3 Proteins in ASD-PPI network are expressed in the human brain and enriched for ASD genetic risks

We evaluated whether the proteins identified in the ASD-PPI network are ASD relevant by assessing 1) whether they are expressed in human brain tissue, 2) whether prey proteins are co-expressed with hcASD risk gene baits, 3) whether prey proteins are expressed in the spatiotemporal contexts that have been previously implicated in ASD, and 4) whether prey proteins are enriched for ASD genetic risk.

As expected, we found prey proteins to be more highly expressed in HEK293T cells than the bait proteins or other HEK293T proteins (**Figure S2.9A**). To assess whether prey proteins are brain-relevant, we queried two large RNA-sequencing (RNAseq) datasets – the Genotype-Tissue Expression (GTEx) project (Aguet et al., 2017; Lonsdale et al., 2013) and the BrainSpan atlas of the developing human brain (M. Li et al., 2018). GTEx has RNAseq data from ~950 adult human donors from 54 tissues, including 13 brain regions. BrainSpan has RNAseq data from ~520 of samples from 16 brain regions of 42 neurotypical donors whose age span the prenatal and

postnatal periods (8 post-conceptual weeks to 40 years old). We found that in both prenatal brain tissue (from BrainSpan) and adult brain tissue (GTEx), bait and prey proteins are significantly more highly expressed than other HEK293T-expressed proteins (**Figure 2.4A, S2.9B**), suggesting that the ASD-PPI network genes are brain-relevant.

We next assessed whether prey genes are co-expressed with the 102 hcASD risk genes (baits) in the brain. To do this, we assessed the relative expression of 102 hcASD risk genes and the prey genes in BrainSpan RNAseq samples by calculating the median expression percentile of each gene set compared to the median expression percentile of 100,000 permuted gene sets with similar HEK293T protein expression levels. We found that the relative expression level of 102 hcASD risk genes and their prey genes are strongly correlated across the n=524 brain samples (**Figure 2.4B,** Pearson R^2 = 0.81, p<1e-15). We concluded that the expression pattern of the prey genes in the ASD-PPI network mirror that of 102 hcASD risk genes across brain samples from various brain regions and developmental periods.

We further evaluated whether prey genes are highly expressed in spatiotemporal contexts that have previously been implicated in ASD. A prior analysis of GTEx RNAseq data demonstrated that hcASD102 genes have enriched expression in the adult cortex and cerebellum (Satterstrom et al., 2020). We assessed the relative expression of 102 hcASD risk genes and the prey genes in GTEx brain samples by calculating the median expression percentile of prey compared to the median expression percentile of 100,000 permuted gene sets with similar HEK293T protein expression levels. We found that, like 102 hcASD risk genes, prey genes are significantly more highly expressed than expected in the adult cerebellum and cortex (**Figure S2.9C**). Multiple studies have demonstrated that ASD risk genes are highly expressed in prenatal brain tissue (Ben-David & Shifman, 2013; Parikshak et al., 2013; A. J. Willsey et al., 2013). We assessed the relative expression of 102 hcASD risk genes and the prey genes in BrainSpan, and found that, similar to 102 hcASD risk genes, prey proteins are significantly more highly expressed in prenatal compared to postnatal samples (**Figure S2.9D**). We further sub-categorized

BrainSpan samples by developmental period as defined in Kang et al., 2011 where periods 1-7 reflect prenatal stages of development and periods 8-15 reflect late infancy through late adulthood. We found that the relative expression levels of 102 hcASD risk genes and the prey genes have highly similar patterns of expression over development, with relatively high expression in prenatal periods and decreased expression in postnatal periods (**Figure S2.9E**).

We posited that if prey genes are ASD-relevant, they should be enriched for ASD genetic risk. Using data from the Simons Simplex Collection (Fischbach & Lord, 2010; Kang et al., 2011; Satterstrom et al., 2020), we found that prey genes are significantly enriched for *de novo* damaging variants associated with ASD; in comparison, other HEK293T-expressed proteins (excluding the proteins in ASD-PPI network) are not **(Figure 2.4C).** We confirmed that ASD probands have higher odds of having a *de novo* damaging variant in the human exome compared to control siblings (De Rubeis et al., 2014; Iossifov et al., 2014; Sanders et al., 2015; Satterstrom et al., 2020), but that these odds decrease significantly when we exclude hcASD102 and prey from the exome (**Figure S2.9F**), demonstrating that ASD-PPI network genes account for the majority of ASD genetic risk from *de novo* damaging variants.

We next assessed whether prey genes are enriched for high-confidence ASD or schizophrenia ('SCZ') risk genes that were identified in recent whole exome sequencing ('WES') studies (Fu et al., 2022; Singh et al., 2022; Zhou et al., 2022). We found that the prey genes are significantly enriched for both sets of hcASD genes, but not for hcSCZ genes (**Figure 2.4D**). We additionally found that the prey genes are significantly enriched for risk genes associated with developmental delay (DD) (Kaplanis et al., 2022), and ASD-associated risk genes curated by SFARI Gene (Banerjee-Basu & Packer, 2010) **(Figure S2.9G)**. We concluded that the ASD-PPI network is significantly enriched for risk genes associated with ASD and DD, but not for SCZ. Of note, SFARI Gene is a comprehensive database that includes any gene associated with ASD risk, allowing for various lines of evidence. SFARI Gene implements a ranking system that reflects the strength of evidence that a given gene is associated with ASD, including 4 categories

(syndromic, category 1 - high confidence, category 2 - strong candidate, and category 3 - suggestive evidence). We observe that prey are significantly enriched for the highest confidence sets of SFARI genes **(Figure S2.9G).**

We further tested whether the ASD-PPI network can be leveraged to identify additional ASD risk genes. To determine whether increasing the number of bait genes used to create ASD-PPI is associated with increased ability to identify prey genes associated with ASD genetic risk, we downsampled the ASD-PPI network by selecting random sets of bait (size 1-100 bait, 1000 iterations per size) and trimming the ASD-PPI network to include only these bait genes and their associated prey genes. For each downsampled network, we benchmarked with 1) enrichment for *de novo* damaging variants associated with ASD probands compared to control siblings in the prey genes, and 2) enrichment in the prey genes for an updated set of 255 ASD genes at FDR ≤ 0.1 (hcASD255) reported in the most recent ASD omnibus study (Fu et al., 2022). We found that increasing the number of bait genes used to construct ASD-PPI increases the amount of ASD-associated genetic risk in the network. Specifically, we found that enlarging the ASD-PPI dataset increases the amount of ASD-associated genetic risk captured by prey (**Figure 2.4E**, teal line) and as expected, the amount of ASD-associated genetic risk from *de novo* damaging variants remaining in the exome (excluding the genes identified in ASD-PPI network) decreases (**Figure 2.4E**, red line). We additionally found that enlarging the ASD-PPI dataset consistently increases the number of hcASD255 genes identified among prey **(Figure 2.4F**, red line)**,** even if we restrict to hcASD255 genes that were not previously identified in hcASD102 **(Figure 2.4F,** teal line**).** We note that on average, downsampled ASD-PPI network prey genes are significantly enriched for hcASD255 genes with just n=4 bait and are significantly enriched for hcASD255 (-hcASD102) genes with only n=11 bait. We concluded that increasing the number of baits used to construct ASD-PPI increases the ability to identify novel hcASD genes.

Finally, we evaluated whether, like ASD-associated risk genes, the prey genes tend to be highly constrained (De Rubeis et al., 2014; Fu et al., 2022; Iossifov et al., 2014; Samocha et al.,

2017; Satterstrom et al., 2020; Zhou et al., 2022). We evaluated several metrics, including pLI (probability of being intolerant of a single loss-of-function variant), misZ (missense Z score, measures gene intolerance to missense variation), synZ (synonymous Z score, measures gene intolerance to synonymous variation, used as negative control), and s_het (selective effect for heterozygous PTVs) (Cassa et al., 2017; Lek et al., 2016). We found that the pLI and misZ of prey genes (excluding hcASD102) are in between that of the baits and other HEK293T-expressed proteins, indicating that damaging mutations in prey genes would confer a smaller selective disadvantage than hcASD genes **(Figure S2.9H)**. The median s_het score of prey genes suggest they may act in an autosomal recessive or polygenic mode compared to hASD genes that are more likely to act in a autosomal dominant mode (Cassa et al., 2017**, Figure S2.9H**).

2.3.4 ASD-PPI shows functional convergence among hcASD risk genes

ASD-PPI contains a median of 11 interactors per bait (**Figure 2.1G**) and many of the preys identified are a part of a protein complex (**Figure 2.3)**. We can use network topology metrics to evaluate how tightly connected the network is and identify network 'hubs' which can be indicative of converging biological functions. Using the ASD-PPI, we calculated the geodesic distance between bait (length of the shortest path it would take to get from one hcASD risk gene to another hcASD risk gene). The median number of nodes it takes to get from one hcASD risk gene to the other is 3 (**Figure S2.5**), suggesting that ASD-PPI is tightly connected and have functional convergence. Indeed, when we performed enrichment analysis for GO terms and protein complexes at the level of bait and associated prey, we saw that many baits shared prey protein complexes and GO terms (**Figure 2.3**). Most notably, the highest shared enriched term was related to chromosomes, which was shared by 13 baits. As many of the ASD risk genes are annotated as transcription factors (Satterstrom et al., 2020), we expected that their interactors would localize to the nucleus. Indeed, when we assessed the subcellular localization of the baits

and their respective preys, we found that many of the baits and preys co-localize in the nucleus (**Figure S2.4**).

2.3.5 The hcASD risk genes - *DYRK1A* and *KIAA0232* directly interact in a protein complex

In our ASD-PPI network, we are particularly interested in interactions that directly connect hcASD risk genes as such interactions imply shared functions among hcASD risk genes. One such interaction was found between DYRK1A and KIAA0232 (**Figure 2.5**). DYRK1A is a highly conserved dual-specificity tyrosine phosphorylation-regulated kinase that is located on chromosome 21 and has been implicated in many neurological disorders, including Down's syndrome and intellectual disability (Dang et al., 2018; Roewenstrunk et al., 2019). KIAA0232 is an uncharacterized protein and not much is known about it. From our ASD-PPI network, we found DYRK1A and KIAA0232 directly interact and also shared many prey proteins, including DCAF7, which in turn interacts with additional hcASD risk proteins, CREBBP, DSCAM and SKI. To validate the interaction between DYRK1A and KIAA0232, we performed a co-immunoprecipitation with flag-tagged KIAA0232 and strep-tagged DYRK1A (**Figure 2.6**). We additionally used co-immunoprecipitation mass spectrometry (co-IP-MS) to confirm the interactions between DYRK1A and KIAA0232, and their shared interactors, CCDC8, DCAF7 and ATAD3B. These findings suggest that DYRK1A and KIAA0232 form a complex with CCDC8, DCAF7 and ATAD3B.

2.3.6 *DYRK1A* and *KIAA0232* knockdown leads to defects in neurodevelopment

It has been reported before that *DYRK1A* haploinsufficiency can interfere with proper brain development and result in a decrease in telencephalon size (Courcet et al., 2012; van Bon et al., 1993; H. R. Willsey et al., 2020). Since we found that DYRK1A and KIAA0232 directly interact in the ASD-PPI, we tested whether knockdown of *KIAA0232* would phenocopy *DYRK1A* knockdown

and disrupt brain development. Indeed, when we knocked down *kiaa0232* in *Xenopus tropicalis* at the two-cell stage, we observed that tadpoles had a significant reduction in telencephalon size, phenocopying the knockdown of *dyrk1a* (**Figure 2.7**).

FIGURES AND LEGENDS

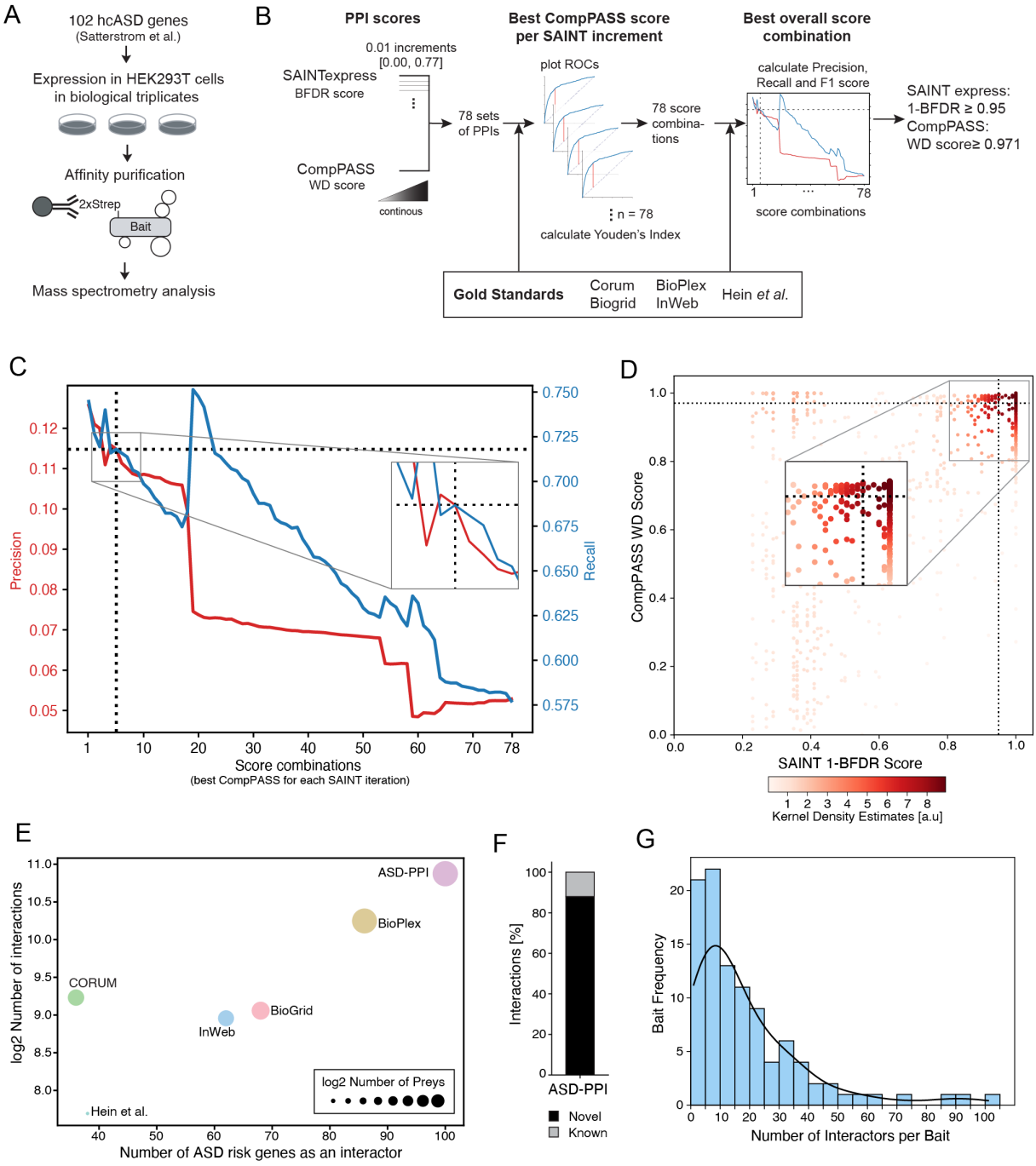


Figure 2.1. Generation of protein-protein interaction (PPI) map in Autism Spectrum Disorder (ASD). (A) Schematic overview of the affinity-purification mass spectrometry (AP-MS) pipeline. (B) Workflow to determine SAINTexpress and CompPASS scoring cutoffs using public databases. (C) Precision (red) and recall (blue) analysis of known interactions using different combinations of SAINTexpress and CompPASS cutoffs. The dotted lines show the precision and recall values at determined scoring cutoffs combination. (D) The kernel density plot displaying the scoring distribution of known interactions observed in the unfiltered PPI dataset. The dotted lines show the determined scoring cutoffs and highlight the high density of known interactions at the

determined scoring cutoffs. (E) Comparison of the number of hcASD proteins reported as an interactor and the number of their associated protein-protein interactions in different large-scale PPI datasets. Different colored bubbles represent different public PPI databases. The size of the bubble shows the size of the unique interactors (preys) in each study. (F) Fraction of novel and known interactions in the ASD-PPI. (G) The distribution of the number of interactors per bait. The median number of interactors for the baits is 11.

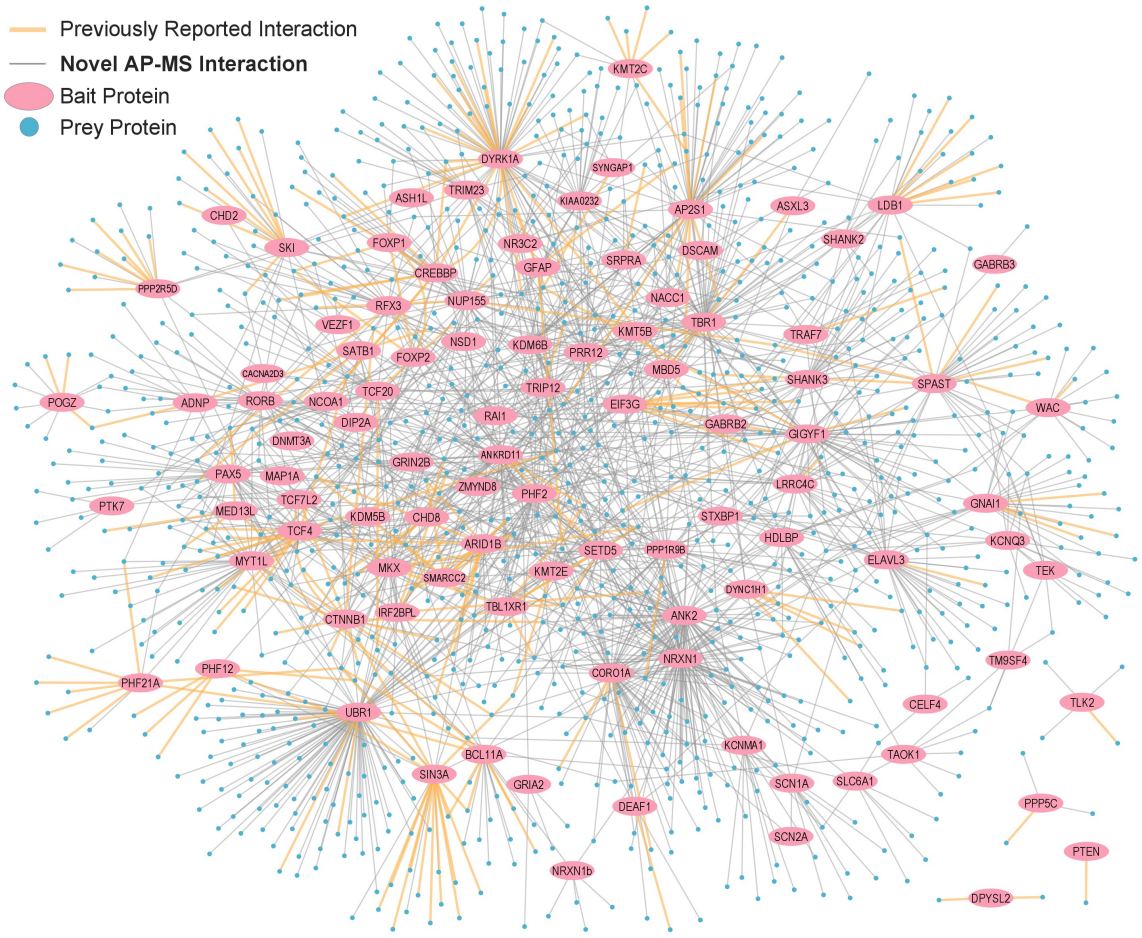


Figure 2.2. Protein-Protein interaction map connecting the 100 hcASD risk genes. A stringent scoring cutoff results in a protein interaction map with 1074 unique prey interactors connecting 100 hcASD risk genes via 1881 interactions. The salmon-colored node denotes the bait, or one of the 100 hcASD risk genes. The blue nodes represent the preys, interactors of the baits. Interactions between the nodes or proteins are represented by an edge. Yellow edges represent interactions that have been previously reported and gray edges represent novel interactions from this study.

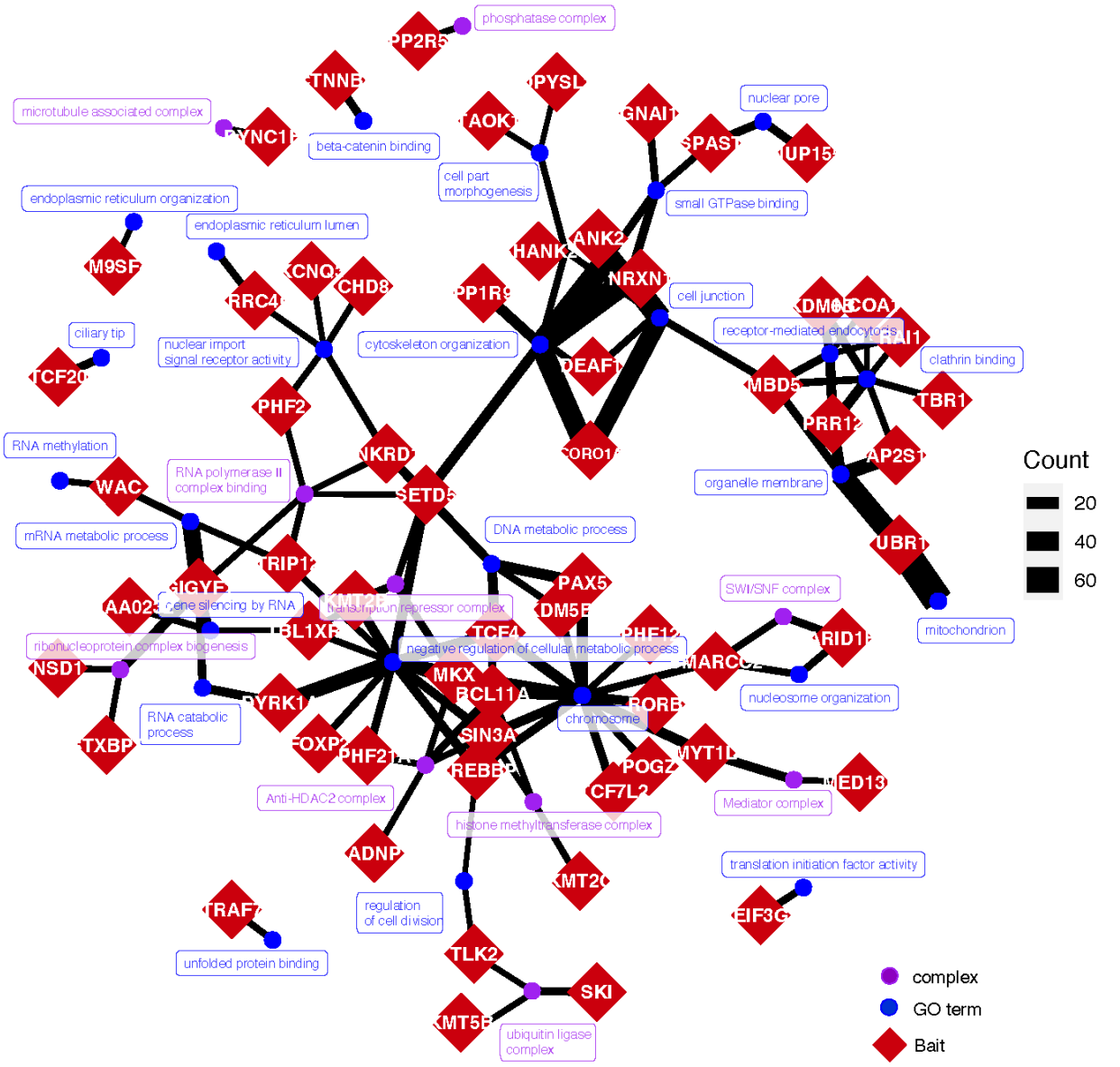


Figure 2.3. Protein complex and gene ontology enrichment analysis in ASD-PPI highlights functional convergence among hcASD risk genes. The red diamonds denote the baits, and the purple and blue circles represent significantly enriched CORUM complex and GO term respectively. The width of the edges corresponds to the number of preys that are found in the enriched CORUM complex or GO term.

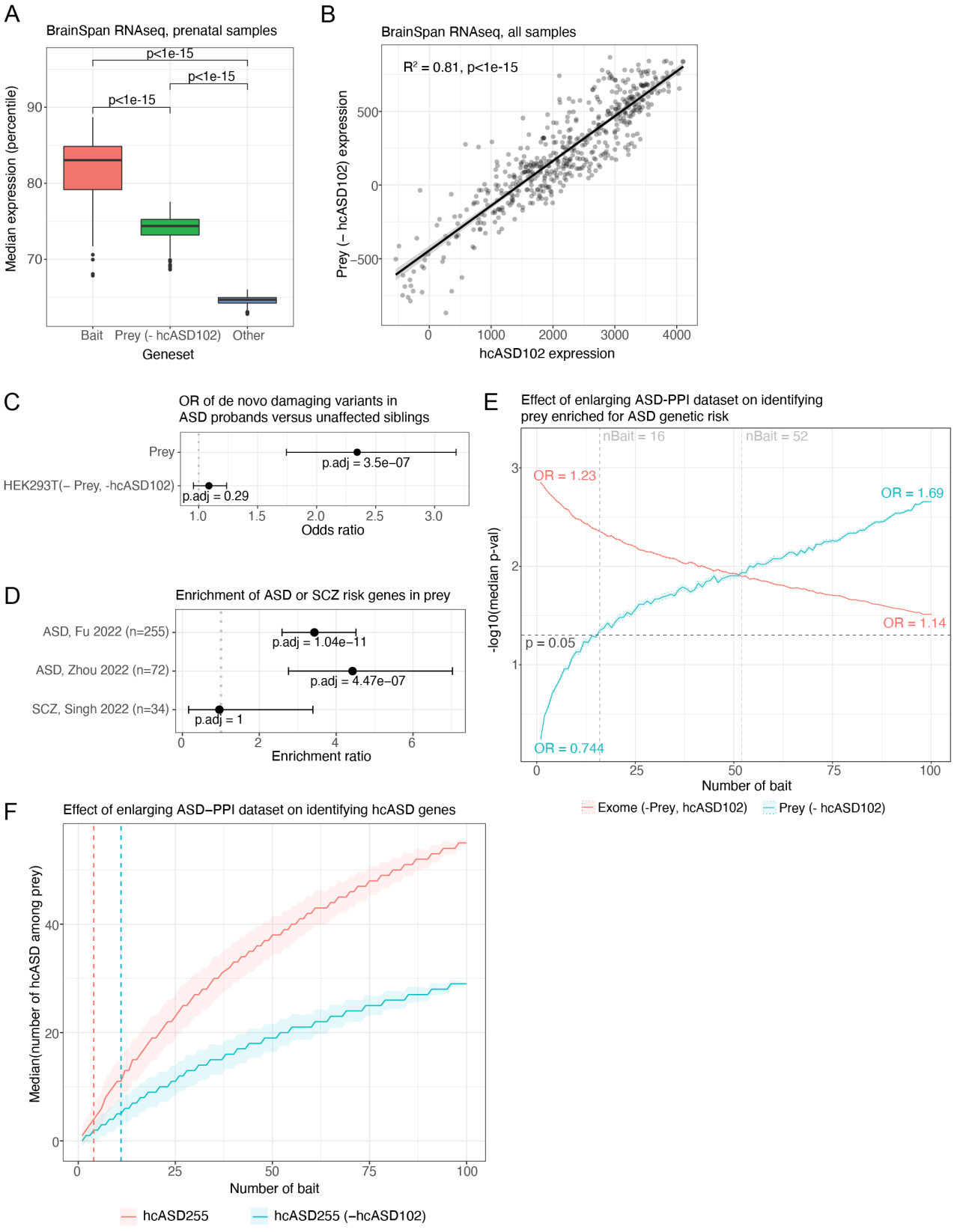

A
BrainSpan RNAseq, prenatal samples
p<1e-15
p<1e-15
p<1e-15
Median expression (percentile)
90
80
70
Bait
Prey (- hcASD102)
Other
Geneset
B
BrainSpan RNAseq, all samples
R² = 0.81, p<1e-15
Prey (– hcASD102) expression
500
0
−500
0
1000
2000
3000
4000
hcASD102 expression
C
OR of de novo damaging variants in
ASD probands versus unaffected siblings
Prey
HEK293T(– Prey, -hcASD102)
p.adj = 3.5e−07
p.adj = 0.29
1.0
1.5
2.0
2.5
3.0
Odds ratio
D
Enrichment of ASD or SCZ risk genes in prey
ASD, Fu 2022 (n=255)
ASD, Zhou 2022 (n=72)
SCZ, Singh 2022 (n=34)
p.adj = 1.04e−11
p.adj = 4.47e−07
p.adj = 1
0
2
4
6
Enrichment ratio
E
Effect of enlarging ASD-PPI dataset on identifying
prey enriched for ASD genetic risk
nBait = 16
nBait = 52
OR = 1.23
OR = 1.69
OR = 1.14
p = 0.05
OR = 0.744
-log10(median p-val)
3
2
1
0
25
50
75
100
Number of bait
Exome (-Prey, hcASD102)
Prey (- hcASD102)
F
Effect of enlarging ASD–PPI dataset on identifying hcASD genes
Median(number of hcASD among prey)
40
20
0
0
25
50
75
100
Number of bait
hcASD255
hcASD255 (–hcASD102)

Figure 2.4. ASD-PPI network genes are expressed in the human brain and enriched for ASD genetic risk.

(A) The median gene set expression percentile across n=237 BrainSpan RNAseq prenatal brain samples (Kang et al., 2011) is displayed for 3 gene sets – bait, prey (excluding hcASD102, Prey – hcASD102)), and all other proteins expressed in HEK293T cells ('Other'). Differences in median expression percentile between different gene sets was assessed by t-test; p-values were adjusted for 3 tests.

(B) The relative expression levels of hcASD102 compared to prey (excluding hcASD102, Prey – hcASD102)) within each of n=524 BrainSpan RNAseq samples are strongly correlated (Pearson R^2 = 0.81, p<1e-15). The relative expression within each brain sample was quantified by the difference between the median gene set rank of observed versus the median of 100,000 permuted gene sets.

(C) Odd ratios (ORs) for de novo damaging variants in ASD probands compared with unaffected siblings from the Simons Simplex Collection (Satterstrom et al., 2020). ORs were calculated using Fisher's exact test (one sided, greater), and p-values were adjusted for 2 tests.

(D) Enrichment of ASD- and SCZ- assocated risk genes in prey calculated using Fisher's exact test (one sided, greater), with p-values adjusted for 3 tests. Two sets of ASD-associated risk genes and one set of SCZ-associated risk genes were obtained from recent WES studies (Fu et al., 2022): n=255 ASD risk genes with FDR ≤ 0.1; (Zhou et al., 2022): n=72 ASD risk genes with study-wide significance; (Singh et al., 2022): n=34 SCZ risk genes with TADA FDR ≤ 0.1.

(E) We assessed the effect of increasing the number of baits used to construct the ASD-PPI network on the ability to capture prey associated with ASD genetic risk (de novo damaging variants from the Simons Simplex Collection). We downsampled the ASD-PPI network by selecting random sets of n=1 to 100 bait and trimming the ASD-PPI network to include only the selected bait and associated prey. For each downsampled network, we calculated the ORs for de novo damaging variants in ASD probands compared with unaffected siblings for prey (teal) and for all other genes in the exome (red), excluding hcASD102 from the analysis. Solid lines depict the median p-values across 1000 iterations; shaded regions indicate the median p-value +/- 1 standard error; the threshold for significance (p=0.05) is labelled with a dashed black line. The median ORs for the gene sets are labelled for ASD-PPI networks constructed using one bait (left) or 100 bait (right). ORs were calculated using Fisher's exact test (one sided, greater), and p-values were not corrected for multiple hypothesis testing. Gray dashed lines indicate the ASD-PPI size at which the prey capture a significant amount of ASD genetic risk (n=16 bait) and when the prey capture more ASD genetic risk than the remaining genes in the human exome (n=52 bait).

(F) We evaluated the effect of increasing the number of baits on the ability to identify prey that are high-confidence ASD risk genes ('hcASD') as defined by the latest ASD WES omnibus study (Fu et al., 2022). We define 'hcASD255', teal, to be the n=255 ASD risk genes with FDR ≤ 0.1 and 'hcASD255 (-hcASD102)', red, to be the n=174 hcASD255 genes that are not among the previously identified set of hcASD102 genes (Satterstrom et al., 2020). We downsample the ASD-PPI network as described above. For each downsampled network, we calculate the median number of hcASD genes among the prey, with shaded regions reflecting median (number of hcASD genes) +/- 1 standard error. We additionally calculated the the OR of prey being enriched for hcASD genes (Fisher's eact test, one sided, greater), and indicate the threshold for significance (median p<0.05) with a dashed line (n=4 bait for hcASD255, n=11 bait for hcASD255 (-hcASD102)).

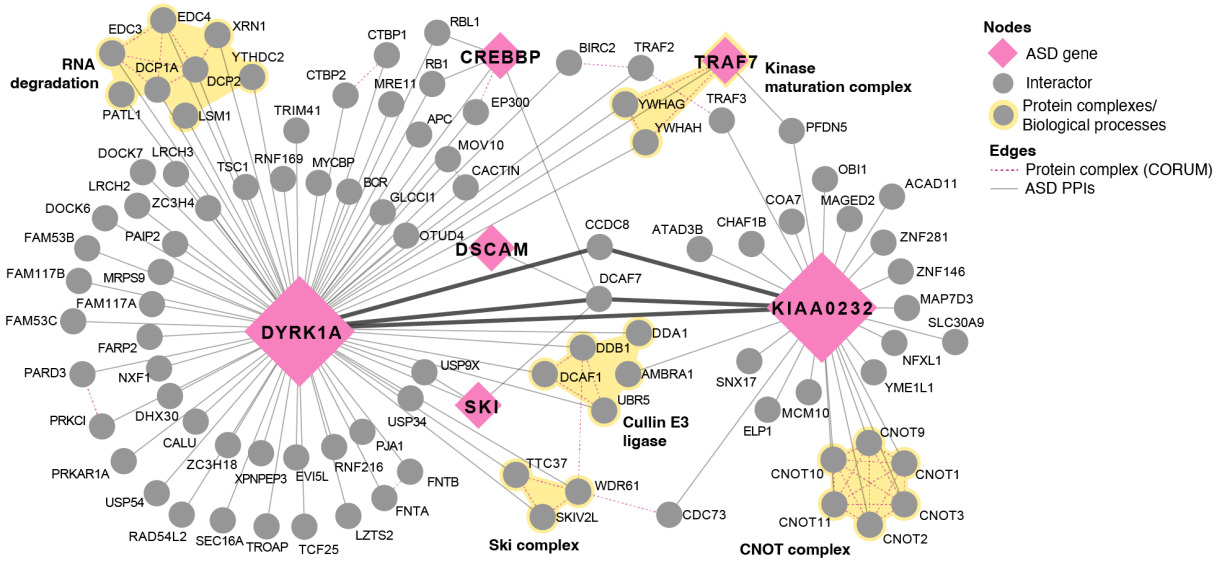
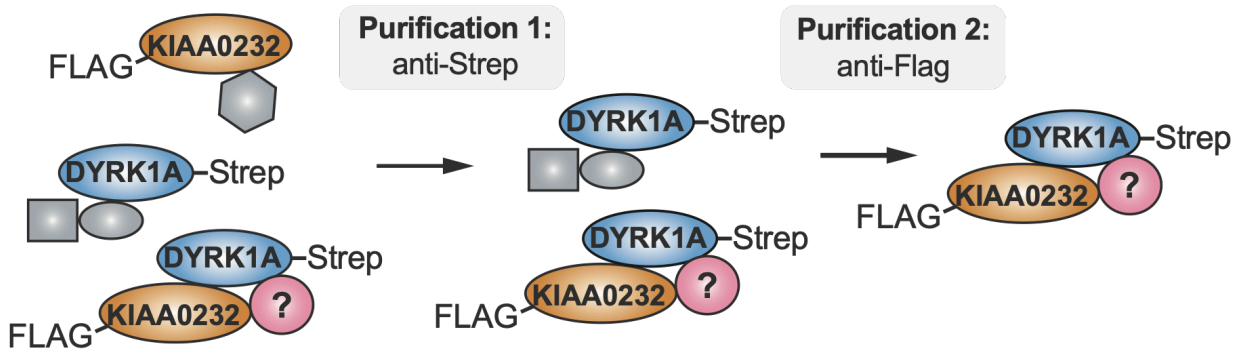


Figure 2.5. The hcASD risk genes, *DYRK1A* and *KIAA0232* from a protein complex. The sub-network of *DYRK1A* and *KIAA0232* was visualized using Cytoscape (Shannon et al., 2003). CORUM protein complexes among the prey proteins are highlighted in yellow. Direct and shared interactions between *DYRK1A* and *KIAA0232* are bolded. ASD risk genes are denoted in pink diamonds and their interactors or preys are in gray circles.

A



B

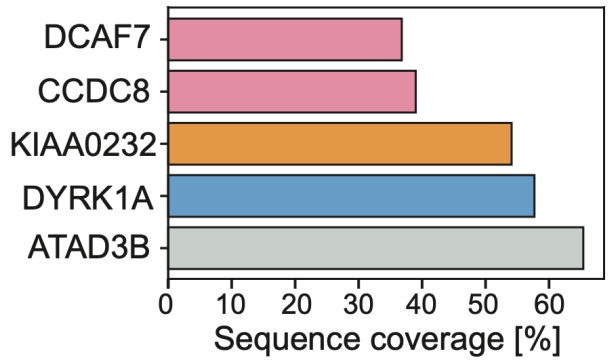


Figure 2.6. Co-IP-MS confirms the direct physical interaction between ASD risk genes, *DYRK1A* and *KIAA0232*. (A) Schematic overview of the sequential immunoprecipitation workflow. (B) Co-IP-MS assay confirms DYRK1A and KIAA0232 to be interacting and shows that they may be in a protein complex with their shared interactors, CCDC8, DCAF7 and ATAD3B.

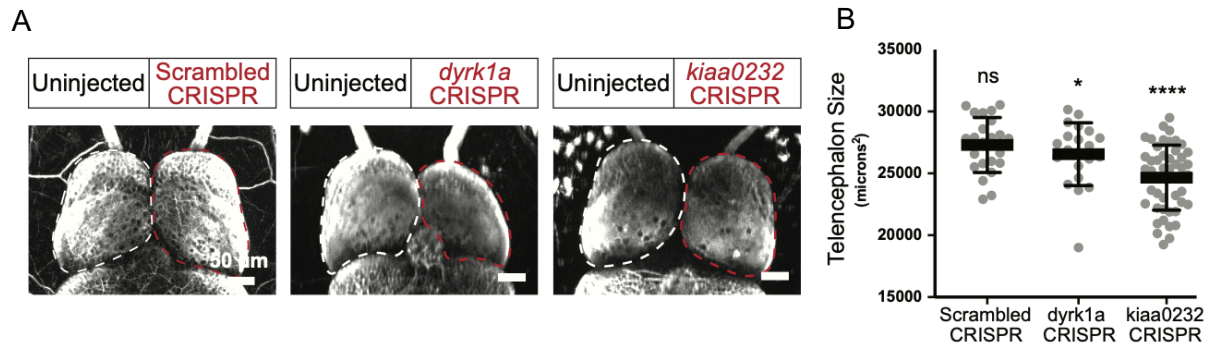


Figure 2.7. *dyrk1a* and *kiaa0232* knockout leads to reduced telencephalon size in Xenopus model. (**A**) Loss of *kiaa0232* phenocopies loss of *dyrk1a in vivo*. Scale bar: 50µm. The telencephalon region for each half is outlined by a dotted line (white, non-injected control; red, CRISPR targeted gene). Non-targeting scrambled CRISPR has a symmetric brain, while mutating *dyrk1a* or *kiaa0232* decreased telencephalon size. (**B**) Quantification of telencephalon size by targeted gene. Measurements normalized by within-animal control side. p-values from student's paired t-test compared to non-injected contralateral control. "n.s.", not significant, indicates $p > 0.05$. (*) $p < 0.05$, (**) $p < 0.01$, (***) $p < 0.001$, and (****) $p < 0.0001$.

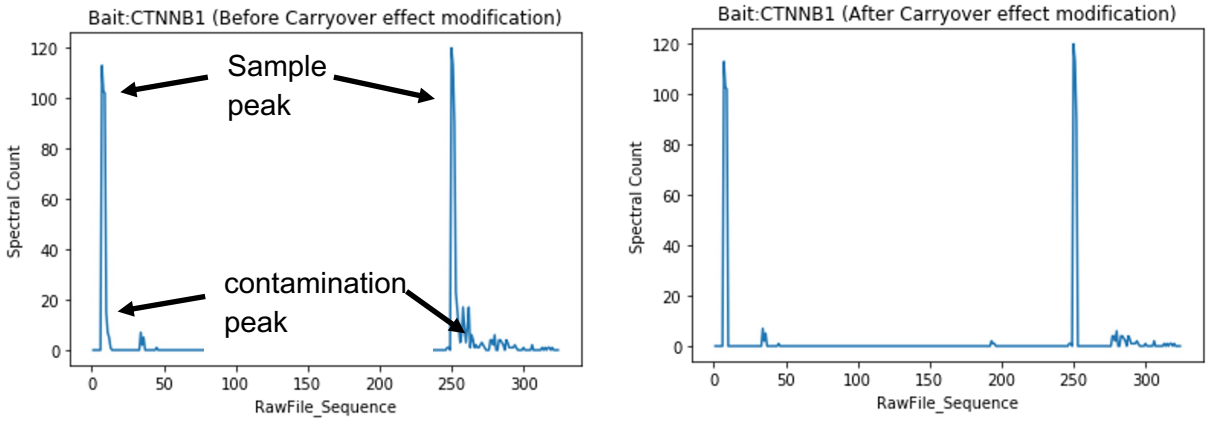


Figure S2.1. An example of the correction for carryover contamination effect. On the left panel, carryover contamination can be observed in the decreasing spectral count presence of the CTNNB1 sample peaks in sequential runs. On the right panel, the carryover contamination has been removed.

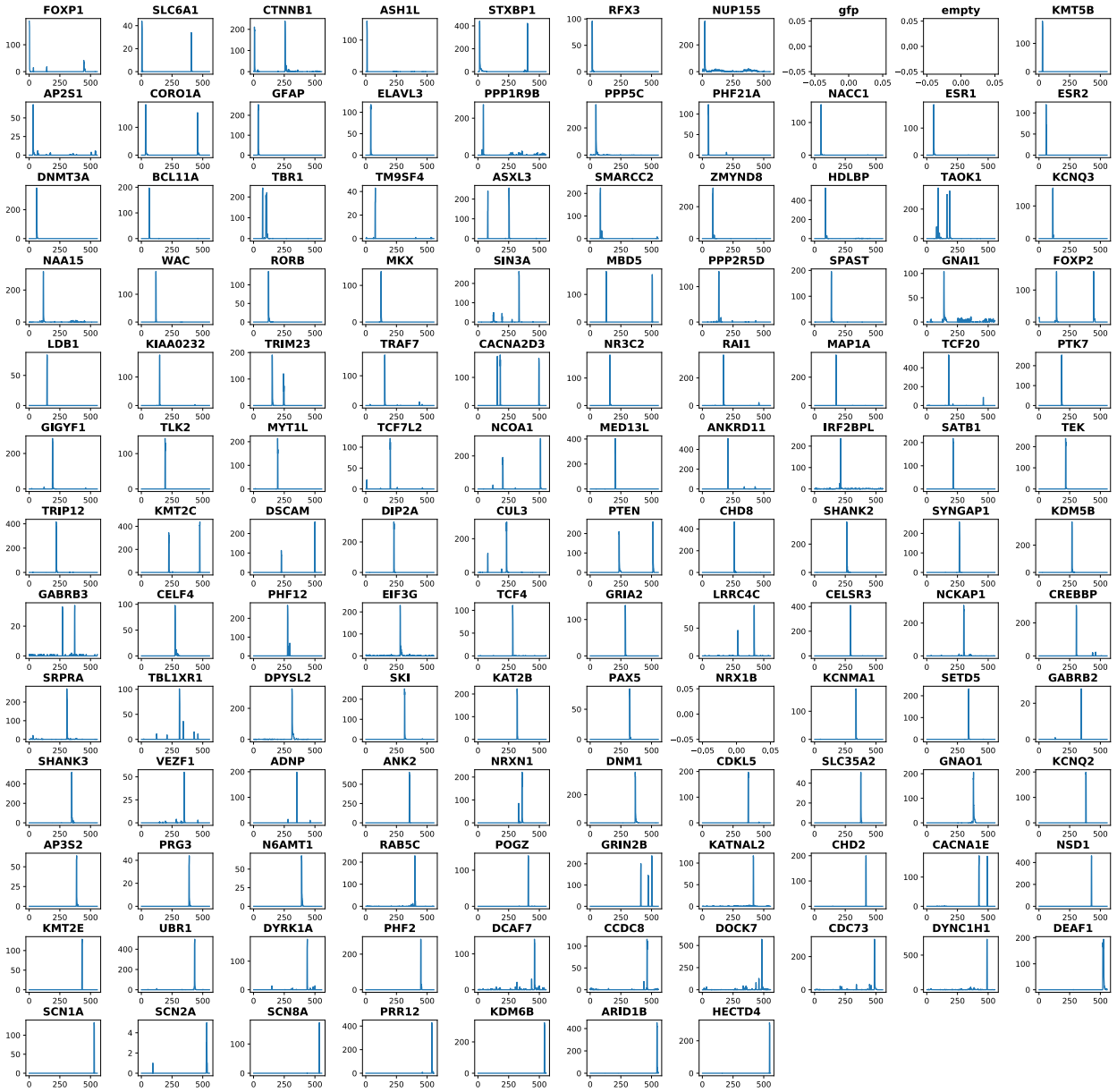


Figure S2.2. Sample peaks before correcting for carryover contamination. Each panel shows the spectral count of the bait observed in the order in which the MS experiment was run. Carryover contamination can be identified by the decreasing presence of bait spectral counts in sequential runs after the injection of the bait samples.

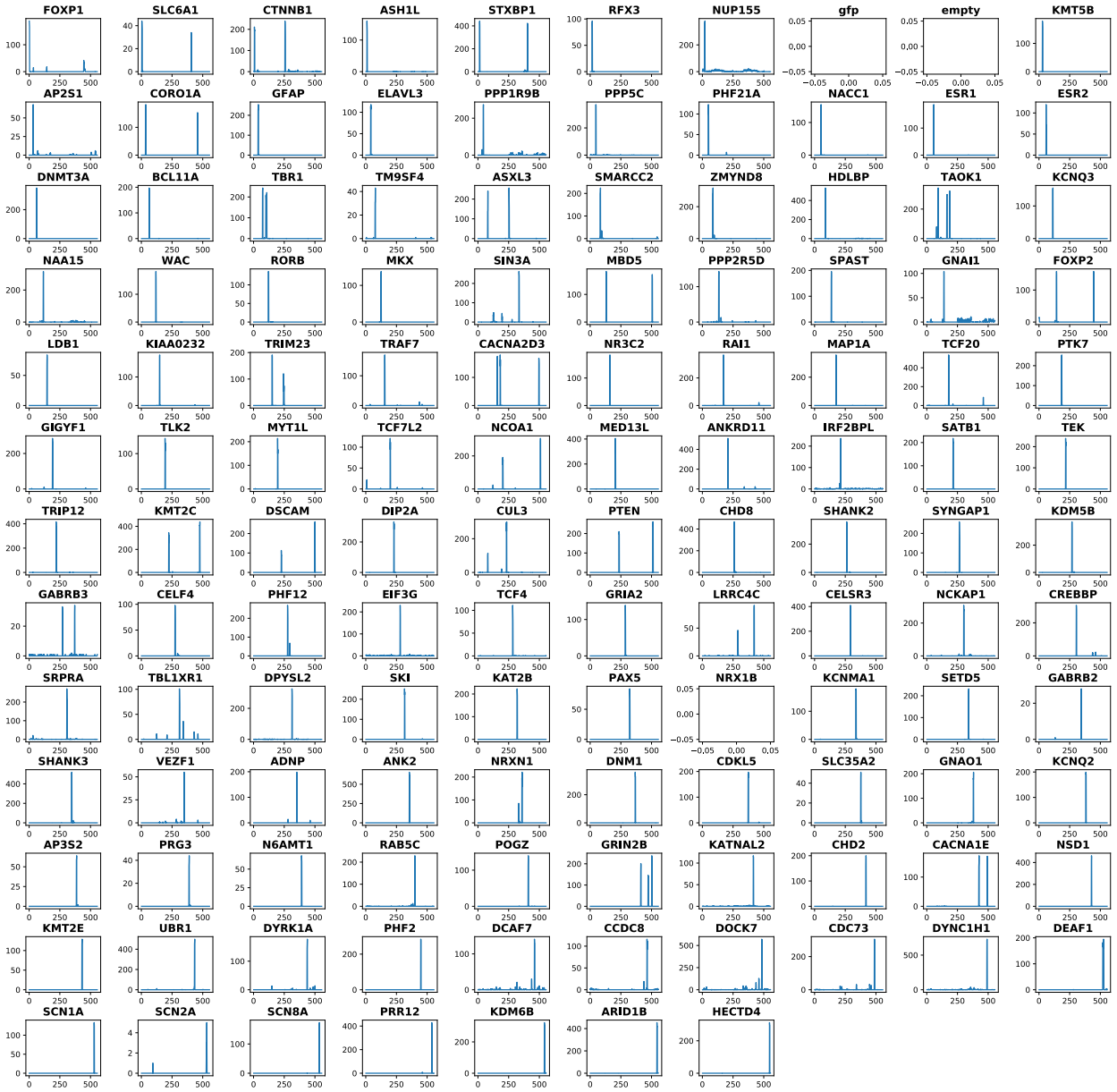


Figure S2.3. Sample peaks after correcting for carryover contamination. Each panel shows the spectral count of the bait observed in the order in which the MS experiment was run, after the removal of carryover contamination.

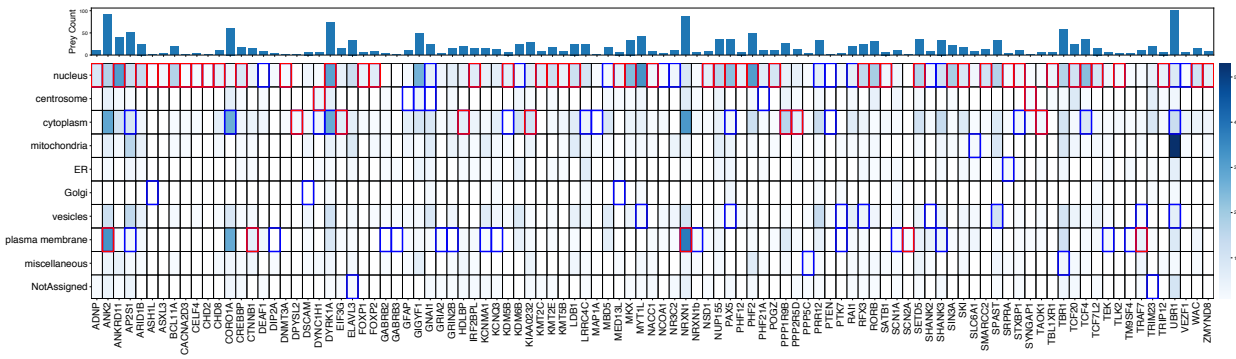


Figure S2.4. Co-localization of baits and preys. The barplot on the top panel indicates the number of preys associated with each hcASD risk gene or bait (x-axis). The heatmap indicates the organelle localization of the preys from each bait. The blue scale denotes the number of preys, the darker the blue, the higher the number of preys. Blue box indicates the localization of the bait. Red box is drawn when the organelle localization for the bait and most of its preys is the same. The protein localization information were first gathered from HEK293T cell data: OpenCell (Cho et al., 2022) and Human Cell Map (Go et al., 2021), which were then supplemented with data from the Human Protein Atlas (Uhlén et al., 2015).

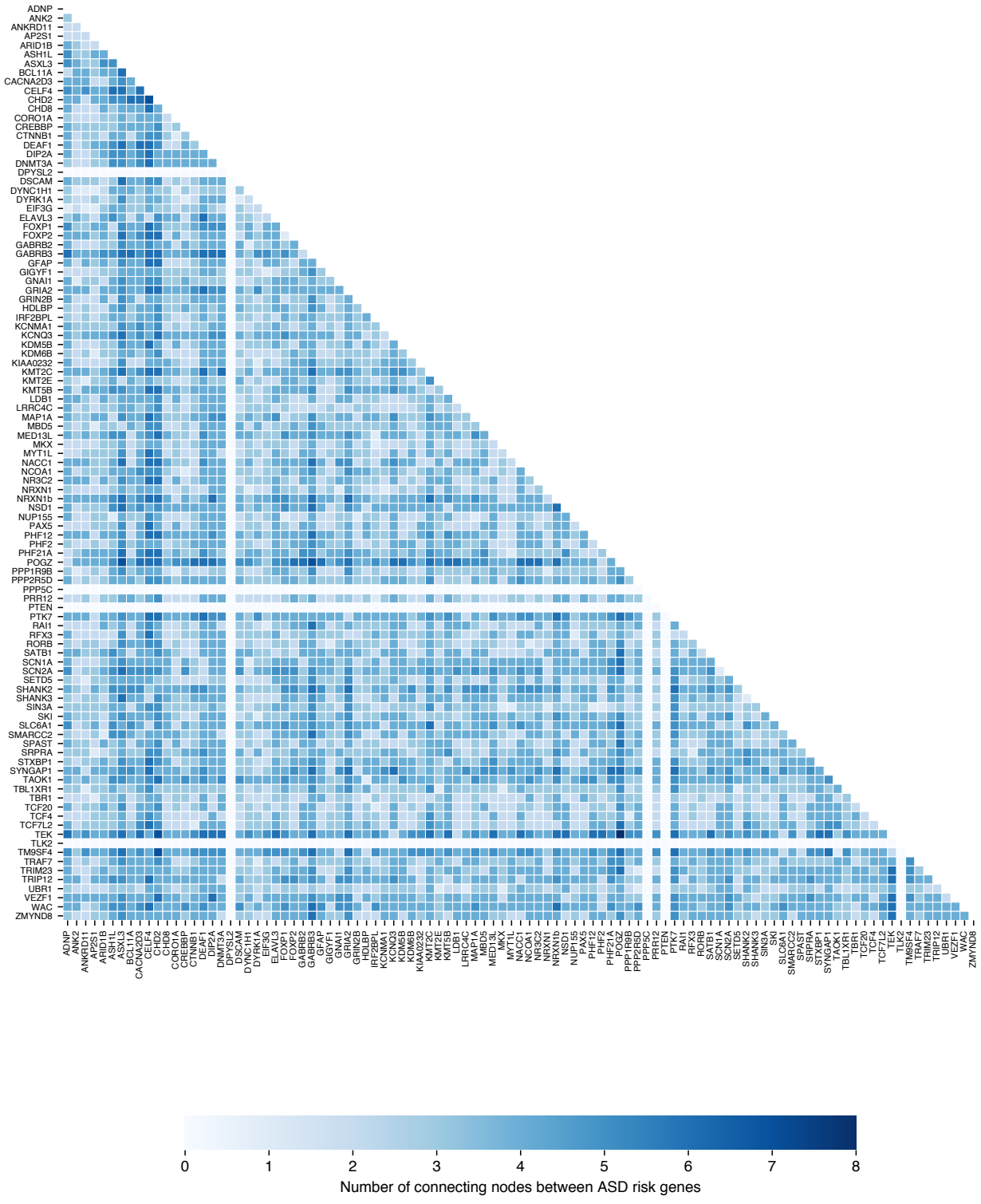


Figure S2.5. Connectivity between hcASD risk genes. Using the ASD-PPI, we calculated the length of the shortest path it would take to get from one ASD risk gene to another hcASD risk gene. The minimum number of nodes connecting 2 hcASD risk genes was 0, i.e., there are no

connection between those 2 ASD risk genes. The median number of nodes it takes to get from one ASD risk gene to the other is 3, and the maximum number is 8.

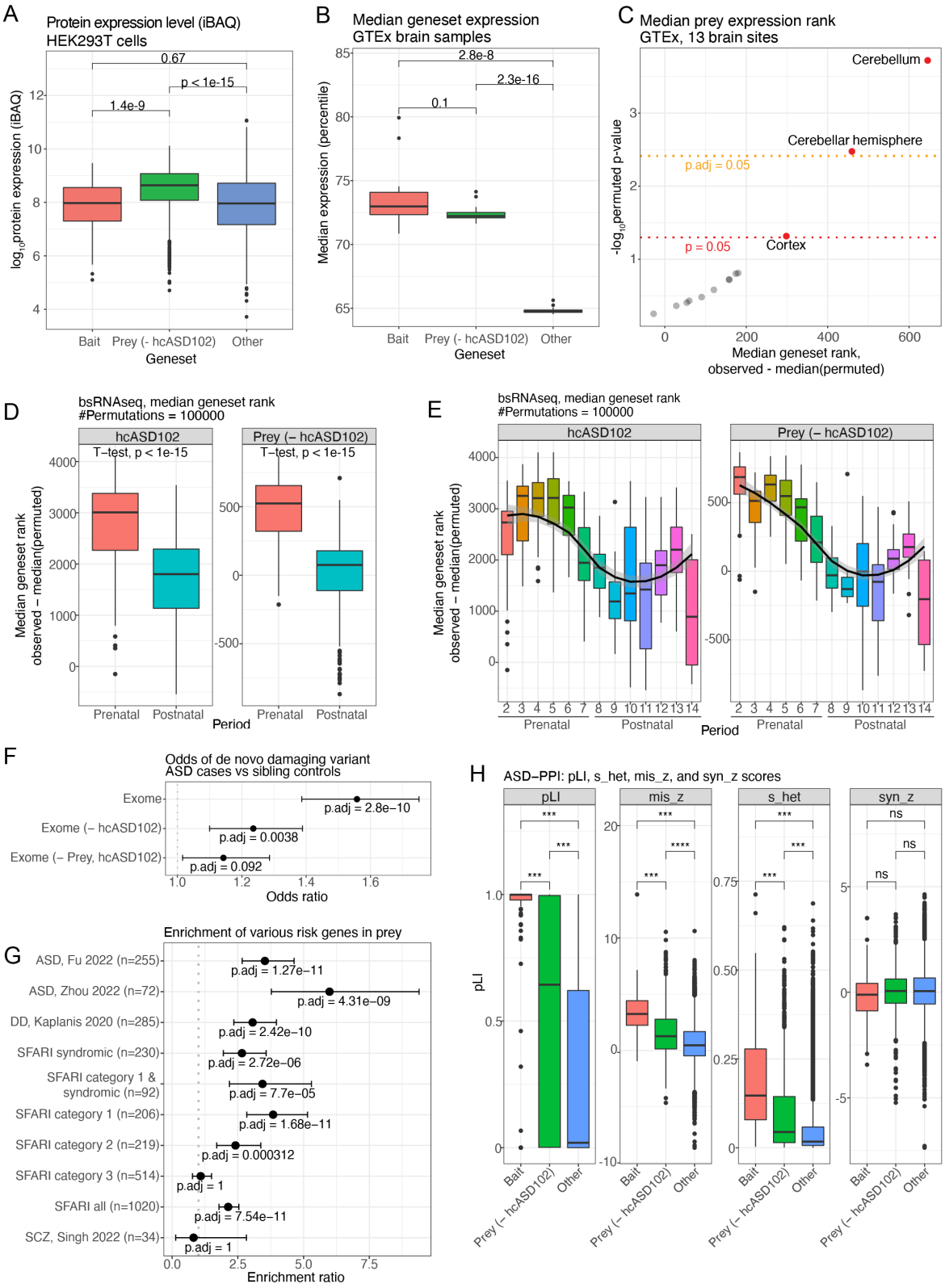


Figure S2.9. ASD-PPI preys are ASD relevant.

(A) The HEK293T protein expression level is shown for 3 gene sets – bait, prey (excluding hcASD102, Prey – hcASD102)), and all other proteins expressed in HEK293T cells ('Other').

HEK293T protein expression was defined using data from (Bekker-Jensen et al., 2017) which profiled HEK293T protein expression by quantitative mass spectrometry (intensity based absolute quantification, iBAQ). Differences in the distribution of protein expression levels between gene sets was assessed by t-test; p-values were adjusted for 3 tests

(B) The median gene set expression percentile for adult brain tissue samples from n=13 brain regions in GTEx (Lonsdale et al., 2013) is displayed for 3 gene sets – bait, prey (excluding hcASD102, Prey – hcASD102)), and all other proteins expressed in HEK293T cells ('Other'). Differences in the median expression percentile between different gene sets was assessed by t-test; p-values were adjusted for 3 tests.

(C) The relative expression levels of prey (excluding hcASD102) in GTEx brain samples across 13 brain regions (Lonsdale et al., 2013) compared to that of permuted gene sets selected from the HEK293T proteome. The relative expression within each brain region was quantified by the difference between the median gene set rank of observed versus median of 100,000 permuted gene sets. Red dashed line shows nominally significant p value of 0.05, and orange dashed line shows significance adjusted for 13 tests.

(D) The relative expression levels of hcASD102 compared to prey (excluding hcASD102, Prey – hcASD102)) in prenatal versus postnatal brain samples from BrainSpan RNAseq (Kang et al., 2011). The relative expression within each brain sample was quantified by difference between the median gene set rank of observed versus median of 100,000 permuted gene sets. Differences in relative expression between prenatal and postnatal samples were assessed by t-test.

(E) We can further categorize BrainSpan RNAseq samples by developmental period as defined in (Kang et al., 2011) where periods 1-7 reflect prenatal stages of development and periods 8-15 reflect late infancy through late adulthood. The relative expression levels of hcASD102 compared t to prey (excluding hcASD102, Prey – hcASD102)) across different periods of development are shown with an overlying Loess regression line, where gray shading reflects 1 standard error. The Spearman's rho of the median rank difference of hcASD102 versus Prey-hcASD102 across periods is 0.946.

(F) ORs for de novo damaging variants in ASD probands compared with unaffected siblings from the Simons Simplex Collection in any gene in the human exome ('Exome'), the exome excluding hcASD102 genes ('Exome (-hcASD102)', or the exome excluding prey and hcASD102 ('Exome (-Prey, hcASD102)'). ORs were calculated using Fisher's exact test (one sided, greater), and p-values were adjusted for 3 tests.

(G) Enrichment of various sets of ASD, DD, or SCZ asssociated risk genes in prey. Enrichment was calculated using Fisher's exact test (one sided, greater), with p-values adjusted for 10 tests. The 8 sets of ASD-associated risk genes were obtained from two recent WES studies (Fu et al., 2022) : n=255 genes; (Zhou et al., 2022): n=72 genes; and SFARI(Abrahams et al., 2013): n=230 syndromic genes, n=92 syndromic & category 1 genes, n=206 category 1, n=219 category 2, n=514 category 3, and n=1020 SFARI genes ('SFARI all')). SCZ-associated risk genes were obtained from(Singh et al., 2022): n=34 genes. Risk genes associated with developmental disorders (DD) were obtained from (Kaplanis et al., 2020): n=285 genes.

(H) The distribution of mutation constraint metrics (pLI, mis_z, s_het, and syn_z) for bait, prey (excluding hcASD102, 'Prey (-hcASD102)'), and all other proteins expressed in HEK293T cells ('Other'). The difference in score distribution between different gene sets was assessed by t-test, with p-values corrected for 3 tests. ***p.adj <= 1e-7; ns = not significant (p.adj>0.05).

CHAPTER 3:

Delineating the differential protein interaction networks of *de novo* ASD missense variants

3.1. Introduction

ASD is a highly genetically heterogeneous neurodevelopmental disorder. High confidence risk genes for ASD were identified based on focusing on the *de novo* and case-control rare variants in the exome (Satterstrom et al., 2020). Protein-coding missense mutations can affect protein function in a wide variety of ways, including the induction of loss- as well as gain-of-function phenotypes. It has been estimated that missense mutations contribute to at least 10% of ASD cases and are found at higher frequencies in probands than in their siblings (Petryshen et al., 2018; Ronemus et al., 2014). Given this burden of missense mutations in ASD cases, understanding how these missense mutations affect protein function, specifically protein-protein interactions, has the potential to reveal much insight into ASD pathobiology. In this chapter, I will be focusing on understanding the protein interaction changes introduced by patient-derived *de novo* missense mutations in ASD risk genes as found in the patients. The overall objective in this chapter is to determine how key interactors of hcASD genes are affected by missense mutations with the goal of deducing molecular pathways and cellular processes dysregulated by ASD variants. We achieved this goal by carrying out affinity purification coupled to mass spectrometry (AP-MS) of proteins encoded by ASD genes in the presence and absence of patient derived *de novo* missense mutations.

3.2. Mutations in hcASD risk genes

Mutations in hcASD genes can be categorized into 3 types – protein-truncating variants (PTVs), missense mutations and synonymous mutations. PTVs encompass different types of mutations such as nonsense, frameshift, and essential splice site variants. Missense mutations

are point mutations in which a single nucleotide change results in a different amino acid. Synonymous mutations are a type of mutation in which a single nucleotide change leads to a codon change but results in the same amino acid. Each type of mutation can be assigned a probability score of how deleterious it is likely to be for gene function. For PTVs, this probability score is called Probability of Loss-of-function Intolerance (pLI) (Kosmicki et al., 2017; Lek et al., 2016; Satterstrom et al., 2020) and for missense mutations, it is an integrated score of Missense badness, PolyPhen-2, and Constraint (MPC) (Samocha et al., 2017).

3.3. Materials and Methods

3.3.1. Selection and cloning of damaging missense mutations in hcASD risk genes

We selected patient-derived *de novo* missense mutations with MPC ≥ 2 in 102 hcASD risk genes that were identified in (Satterstrom et al., 2020). This resulted in 87 *de novo* missense mutations in 43 hcASD risk genes. The missense mutations were introduced in the wild-type version of the 2x-Strep-tagged pcDNA4 hcASD gene construct (as described in Chapter 2) using Q5 site-directed mutagenesis. All constructs were sequence validated.

3.3.2. Affinity-purification and mass spectrometry (AP-MS)

We used the same methods of cell culture, transfection, affinity purification, on-bead digestion, MS acquisition and analysis as described in Chapter 2 to perform AP-MS using mutant constructs.

3.3.3. PPI scoring

Following the identification and quantification of proteins using MaxQuant, high confidence interacting proteins (HCIP) were identified by running both SAINTexpress and CompPASS after carryover effects were removed (See Chapter 2: Materials and Methods for carryover effect removal). SAINTexpress was run batch-wise, using empty vector and GFP constructs as controls.

The SAINTexpress output for the ASD-PPI was then concatenated with the output for the ASD_mut-PPI. To have better specificity in identifying HCIPs using CompPASS, a larger input dataset was created by combining the ASD-PPI CompPASS input with ASD_mut-PPI input. CompPASS was then run using the CompPASS input for ASD-PPI WT network as the stats table. The two scoring outputs from SAINTexpress and CompPASS were then merged and HCIPs were identified using the scoring cutoffs previously optimized from the wild-type ASD-PPI network - SAINTexpress score (1-BFDR) ≥ 0.95 and CompPASS score (rank_WD) ≥ 0.971 (See Chapter 2 for the details on how the scoring cutoffs were determined). We removed new HCIPs identified for ASD-PPI and retained only HCIPs from the mutant AP-MS experiments to build the ASD_mut-PPI interactome.

3.3.4. Differential interaction scoring

To account for the variabilities in the intensities of interacting proteins due to the variability in bait expression, intensity values were first normalized to the bait intensity level within each bait AP-MS run using the Tukish Median Polish normalization method. To quantify the changes in interactions between wild-type and mutant baits, we used an R package for statistical analysis, MSstats (Choi et al., 2014). Using the default parameters in the *dataProcess* function, the normalized intensity values were used as an input to run differential analysis between wild-type and mutant groups. Using Python, the fold-change values along with the p-values for the HCIPs in ASD_mut-PPI were used to plot hierarchically clustered heatmaps so that differential prey proteins are easily visualized (**Figure S3.2**, **S3.3**).

3.3.5. Differentiation of neural progenitor cells into neurons

Neural progenitor cells (NPCs) were maintained in DMEM/F12 medium containing Glutamax (1X), non-essential amino acids (NEAA, 1X), N2 supplement (1X), B27 supplement (1X), FGF2 (10ng/ml), EGF (10ng/ml). NPCs were differentiated into neurons as reported

previously (Qi et al., 2017). Briefly, NPCs were cultured in 1:1 DMEM/F12 and Neurobasal Media containing N2 (1X), B27 -Vitamin A (1X), GlutaMAX (1X), 0.5 mM dibutyryl cAMP, 0.2 mM ascorbic acid, BDNF (20 ng/mL) along with small molecules PD0325901 (1μM), SU5402 (5 μM), DAPT (10 μM) and grown on PDL/Laminin coated tissue culture plates. Media was refreshed every other day.

3.3.6. Electroporation in NPCs

About 16 million NPCs were grown per biological replicate (about 1.5-2 confluent 10 cm dishes). The dishes were coated with 2.5 mL of Matrigel before transfection. After the media was aspirated and the plates were washed with PBS, NPCs were dissociated with 2 mL of Accutase per dish with periodic agitation and incubation for 10 mins. Cells were then pipetted up and down a few times to fully break up the cell clumps. One plate was transferred to a 15 cm conical tube which was then used for cell counting and the rest of the plates were transferred into a second conical tube. The volume of Accutase was doubled with DMEM and centrifuged at 300 g for 5 mins. The first pellet was then resuspended in 1 mL of MaxCyte buffer; from which, 100 μL was mixed with 100 μL of Accumax in a tube and incubated at room temperature for 5 mins. 10 μL of the MaxCyte buffer and Accumax mixture was then put on the cell counter to count the number of cells per 10 cm dish. The second pellet was resuspended in 5 mL of MaxCyte buffer. Both conical tubes were then combined into one tube and centrifuged at 300 g for 5 mins. The pellet was resuspended in 6 mL of MaxCyte buffer and centrifuged again at 300 g for 5 mins. The pellet was resuspended in 1 mL of MaxCyte buffer, and 16 million cells were aliquoted into an Eppendorf tube for each replicate, which were centrifuged at 300 g for 5 mins. The supernatant was aspirated and 70 μL of MaxCyte buffer was added to the pellet. The cell pellet was then fully resuspended by pipetting up and down using a wide bore 200 μL pipette tip gently so as not to generate bubbles. Plasmids were then added to result in a concentration of 150 ng/μL. The cells were

resuspended using a wide bore tip and transferred carefully into a well of OC-100x2 processing assembly. The processing assembly was then inserted into MaxCyte instrument for electroporation. After the electroporation process was completed, the processing assembly was then placed in the incubator for 20 mins. The cells were then replated in NPC media.

3.3.7. Neurite outgrowth assay

NPCs were dissociated with Accutase and 50,000 cells were plated on PDL/Laminin coated plates in the differentiation medium. On day 3, the cells were washed with 1X PBS and fixed with 4% formaldehyde in PBS for 10 mins at room temperature. Cells were permeabilized with 1X PBS with 0.2% triton X-100 and treated with antibodies against neuronal marker TUBB3 (Invitrogen, cat#MA-1 19187). Cell nuclei were stained with DAPI. The cells were imaged on Nikon Ti2-E microscope and images were processed in Fiji (Image J) software. Neurite length was analyzed using NeuronJ while percentage of cells with neurite outgrowth and branching was calculated using Cell Counter. Analysis shown is for three independent experiments with 100 cells analyzed per experiment. ANOVA with *post hoc* Bonferroni test was used to calculate significance between different groups.

3.4. Results

3.4.1. *De novo* missense mutations alters hcASD risk genes interactions

There are three tiers for MPC score (≥2, 1-2, 0-1) which are in the order of decreasing functional impact. De novo missense variants with high MPC scores (MPC ≥2) are significantly enriched in ASD cases compared to controls (Fu et al., 2022; Satterstrom et al., 2020; Zhou et al., 2022). To study the effect of *de novo* missense mutations in the protein-protein interaction network, we focused on hcASD risk genes with missense mutations that have MPC ≥ 2 (87

missense mutations in 43 hcASD risk genes) as they are significantly enriched in the ASD individuals and predicted to have the most deleteriousness effect.

We successfully generated AP-MS data for 64 missense mutations in 35 hcASD risk genes. Stringent scoring analysis of these AP-MS data resulted in 929 high-confident interactors. Depending on the bait, the number of interactors varied among wild-type and mutant AP-MS across different baits (**Figure S3.1**). To quantify the abundance of the interactors between wild-type and mutant proteins, we focused on high-confidence interactors that were identified for both wild-type and mutant proteins per hcASD risk gene. While many of the interactors show no change in affinity between wild-type and mutant proteins, we found that 133 proteins have more affinity for mutant proteins and 152 proteins for wild-type proteins (**Figure 3.1C**, **3.1D**).

Hierarchically clustering the interactors based on their log2 fold-change levels in each bait, we observe that different mutants of the same gene tend to result in a similar pattern of decrease or increase in interaction affinity, suggesting that the mutant proteins may have convergent molecular and functional consequences (**Figure S3.2**, **S3.3**). Indeed, when we superimposed CORUM protein complex, we see a common loss or gain of the interactors from the same protein complexes among the mutants in the same hcASD protein and sometimes shared across mutants from different hcASD proteins. This is evident by the gain of interactors involved in chromosome organization in both missense mutants of *MKX*, and loss of interactors involved in spliceosome in both missense mutants of *STXBP1* and the missense mutant of *NSD1* (**Figure 3.2**).

3.4.2. Functional characterization of *de novo* missense mutations in *FOXP1*

FOXP1 and *FOXP2* are both hcASD risk genes that are transcriptional repressors of the forkhead-box (FOX) family. They play an important role in the development of the brain, heart, lung, esophagus, immune system, and spinal motor neurons. To regulate transcription, FOXP1 forms either homo- and hetero-dimers with itself or other FOXP proteins (Co et al., 2020; S. Li et al., 2004; Sollis et al., 2017). In addition to being ASD risk genes, both *FOXP1* and *FOXP2* have

been implicated in different neurodevelopmental disorders (Bacon & Rappold, 2012; Johnson et al., 2018; Sin et al., 2015). There were 3 missense mutations in FOXP1 and 1 missense mutation in FOXP2 with MPC $\geq$ 2 identified from the ASD cohort in Satterstrom et al., 2020. The two missense mutations in FOXP1 and one missense mutation in FOXP2 lie in the same conserved forkhead domain (**Figure 3.3**). From our ASD_mut-PPI data, we observed that the *de novo* FOXP1 and FOXP2 missense mutants lose interaction with FOXP4 (**Figure 3.4**). This loss of FOXP4 interaction was confirmed by Western blot (**Figure 3.5**). We hypothesized that this loss of FOXP4 interaction in FOXP1 mutant that we observed in HEK293T can be translated to neural progenitor cells (NPCs) and neurons. Indeed, we found that overexpression of $FOXP1^{R513C}$ in NPCs resulted in nuclear aggregates (**Figure 3.6**), which is consistent with previous literature (Sollis et al., 2017). We also observed aggregates of FOXP2 and FOXP4, co-localizing with $FOXP1^{R513C}$ (**Figure 3.6**). Interestingly, when we differentiated NPCs overexpressing $FOXP1^{R513C}$ into neurons, we saw a significant decrease in neurite length compared to the neurons with wild-type FOXP1 proteins (**Figure 3.7**), suggesting that FOXP1 plays a role in normal neurite growth.

FIGURES AND LEGENDS

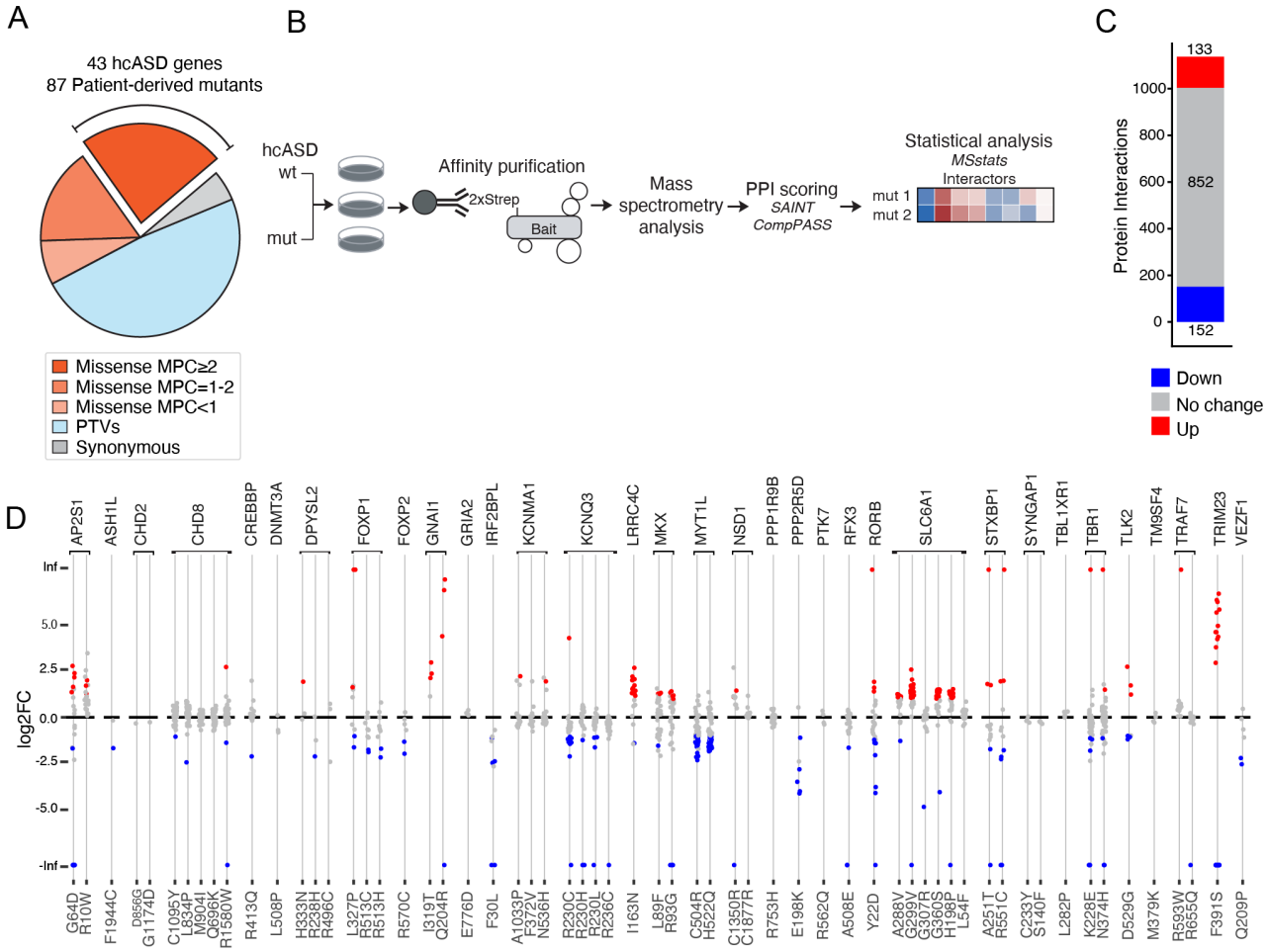


Figure 3.1. Differential Interaction analysis of patient-derived *de novo* missense mutations in hcASD risk genes. (A) Distribution of the types of mutations observed in the ASD cohort analyzed in (Satterstrom et al., 2020), highlighting *de novo* missense mutations with MPC ≥ 2, that are used to construct the ASD mutant interactome. (B) A schematic overview to study the differential protein interactions in the ASD mutant interactome. (C) Quantification of the differential protein interactors identified 133 interactors (in red) to have higher affinity for the mutant proteins and 152 interactors (in blue) for the wild-type proteins. (D) Dot-plot visualization of the loss and gain of interactors in hcASD risk genes with *de novo* missense mutations, grouped by hcASD risk genes (baits). Each dot is a high-confident interactor in either the wild-type or the mutant of the risk gene.

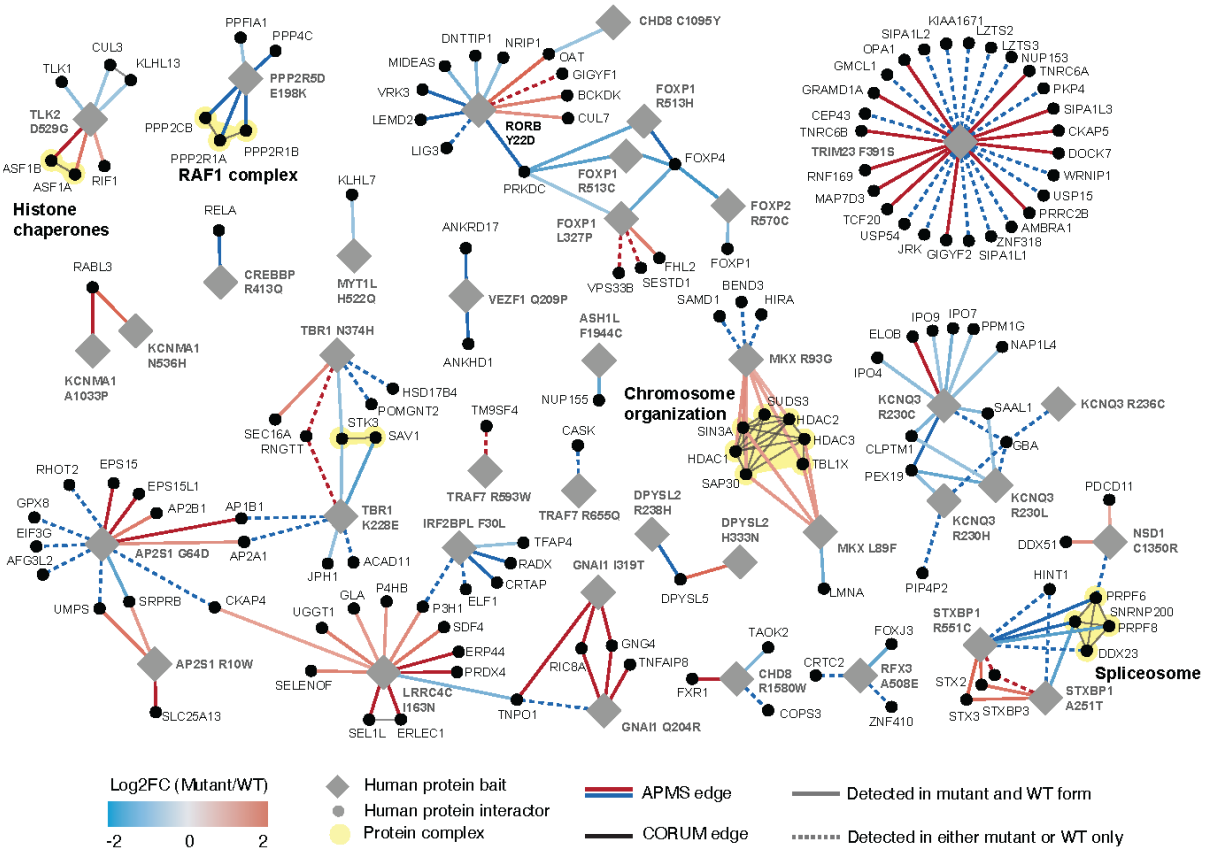


Figure 3.2. ASD mutant interactome (ASD_mut-PPI). Differential scoring analysis of the interactome of missense mutant vs wild-type hcASD risk genes, highlighting the loss and/or gain of CORUM protein complexes shared by different mutants. Diamond shapes denote the mutant baits and circles denote their respective interactors, or preys. The color scale of the edges corresponds to the specificity of the interactions, where blue edges have stronger affinity for the wild-type protein and red edges have higher affinity to the mutant protein. The prey proteins connected by the blue and red dotted lines are detected only in wild-type or mutant proteins respectively. CORUM protein complexes among the prey proteins are shaded in yellow.

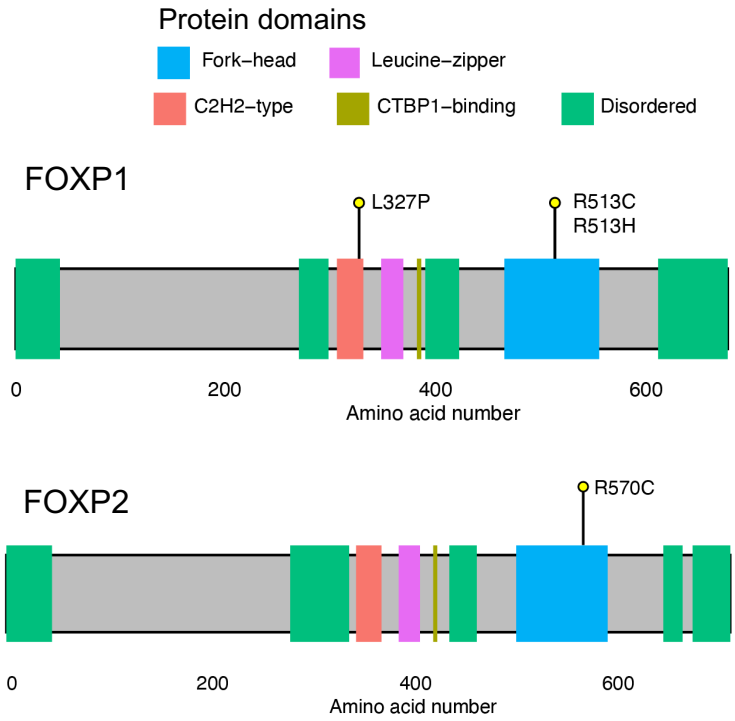


Figure 3.3. Patient-derived *de novo* missense mutations in FOXP1 and FOXP2. Lollipop plot visualizing the *de novo* missense mutations from ASD patients in FOXP1 and FOXP2.

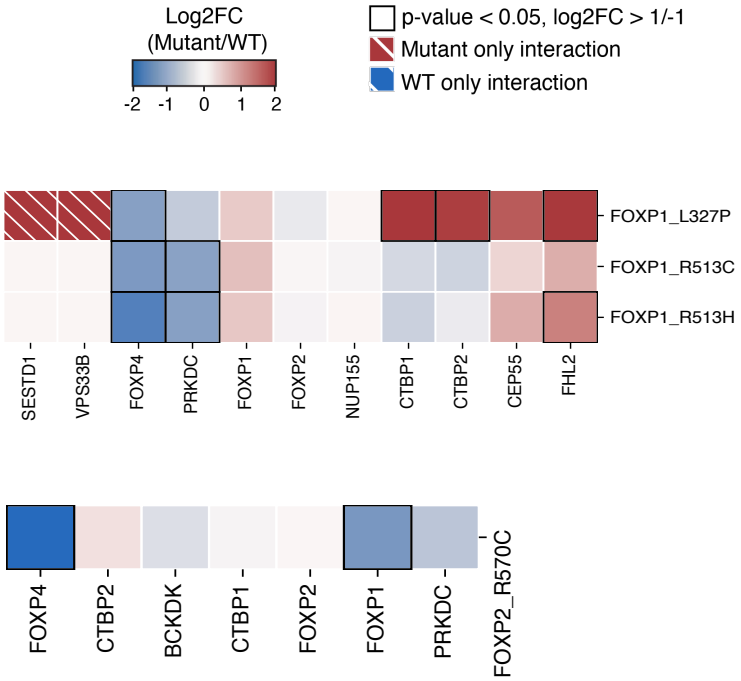


Figure 3.4. Differential interaction analysis of patient-derived missense mutations in FOXP1 and FOXP2. Heatmap visualizing changes in prey abundance in the interactome of mutant versus wild-type FOXP1 and FOXP2.

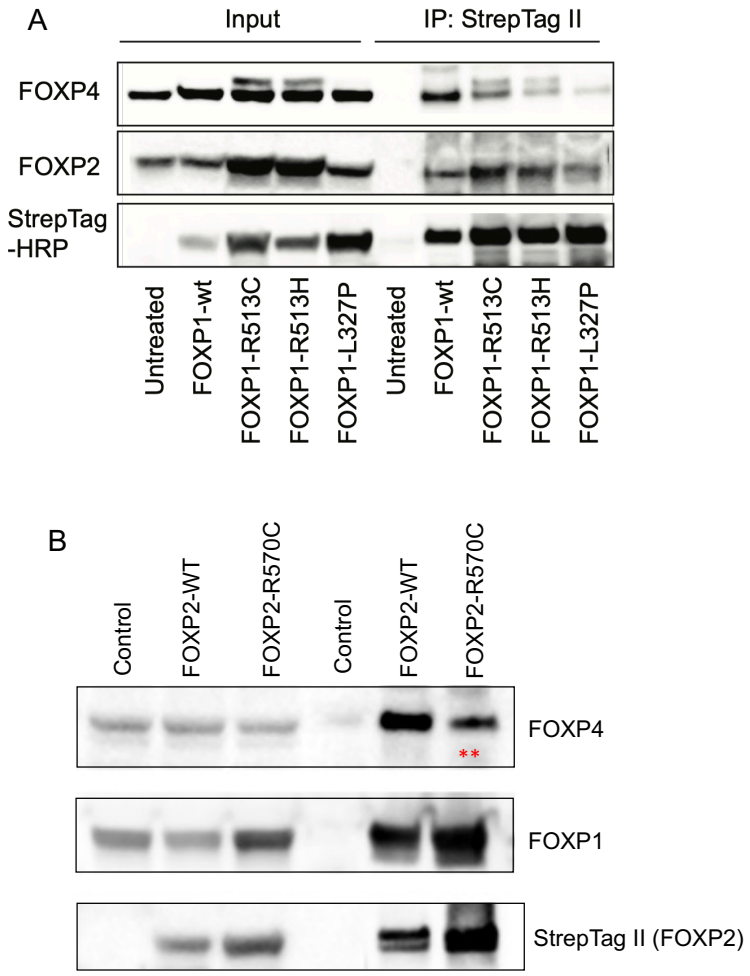


Figure 3.5. Immunoblot analysis of missense mutations in FOXP1 and FOXP2. (A) Western blot evaluating expression of FOXP4 and FOXP2 in HEK239T cells overexpressing WT or mutant FOXP1. (B) Western blot evaluating expression of FOXP4 and FOXP1 in HEK239T cells overexpressing WT or mutant FOXP2.

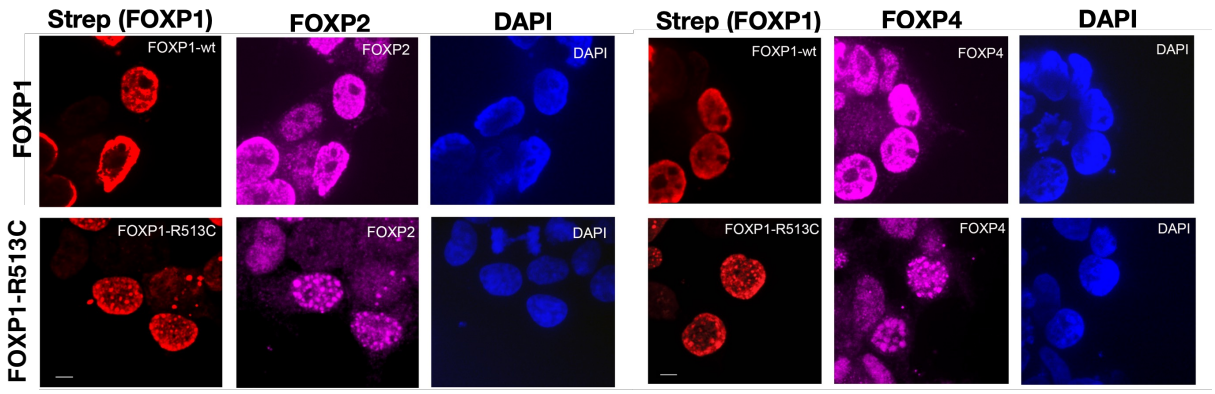


Figure 3.6. Mutant FOXP1 forms nuclear condensates in neural progenitor cells (NPCs). Fluorescence imaging of NPCs expressing Strep-tagged wild-type FOXP1 protein (top panels) and FOXP1^{R513C} (bottom panels). Nuclei were stained with DAPI (blue). FOXP2 and FOXP4 are depicted in magenta. Scale bar, 10μm.

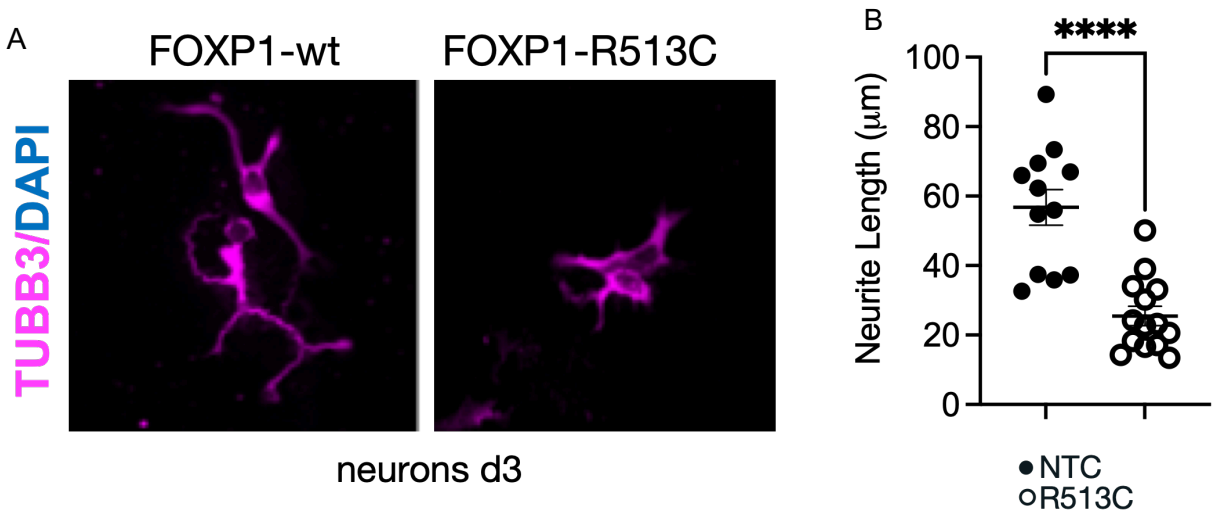


Figure 3.7. Mutant *FOXP1* impairs dendrite formation in neurons. (A) Fluorescence imaging of cortical glutaminergic neurons overexpressing Strep-tagged wild-type FOXP1 and FOXP1^{R513C}. Nuclei were stained with DAPI (blue). Tubulin (TUBB3) were depicted in magenta. Scale bar, 10μm. (B) Quantification of the neurite length. P-values from student's paired t-test compared to the non-targeting control. (*) $p < 0.05$, (**) $p < 0.01$, (***) $p < 0.001$, and (****) $p < 0.0001$.

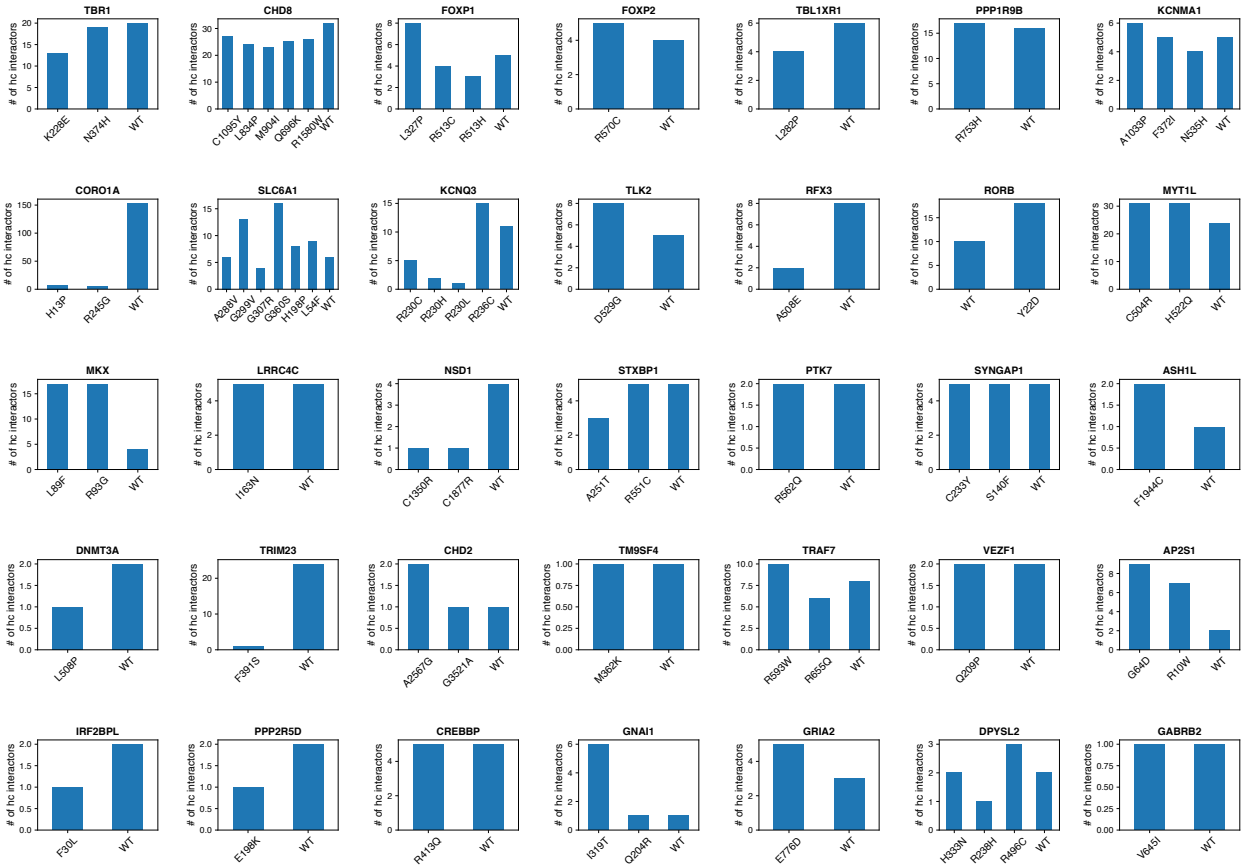


Figure S3.1. Number of interactors vary among WT and mutant hcASD proteins. Bar charts depict the number of interactors per bait, grouped by ASD risk gene.

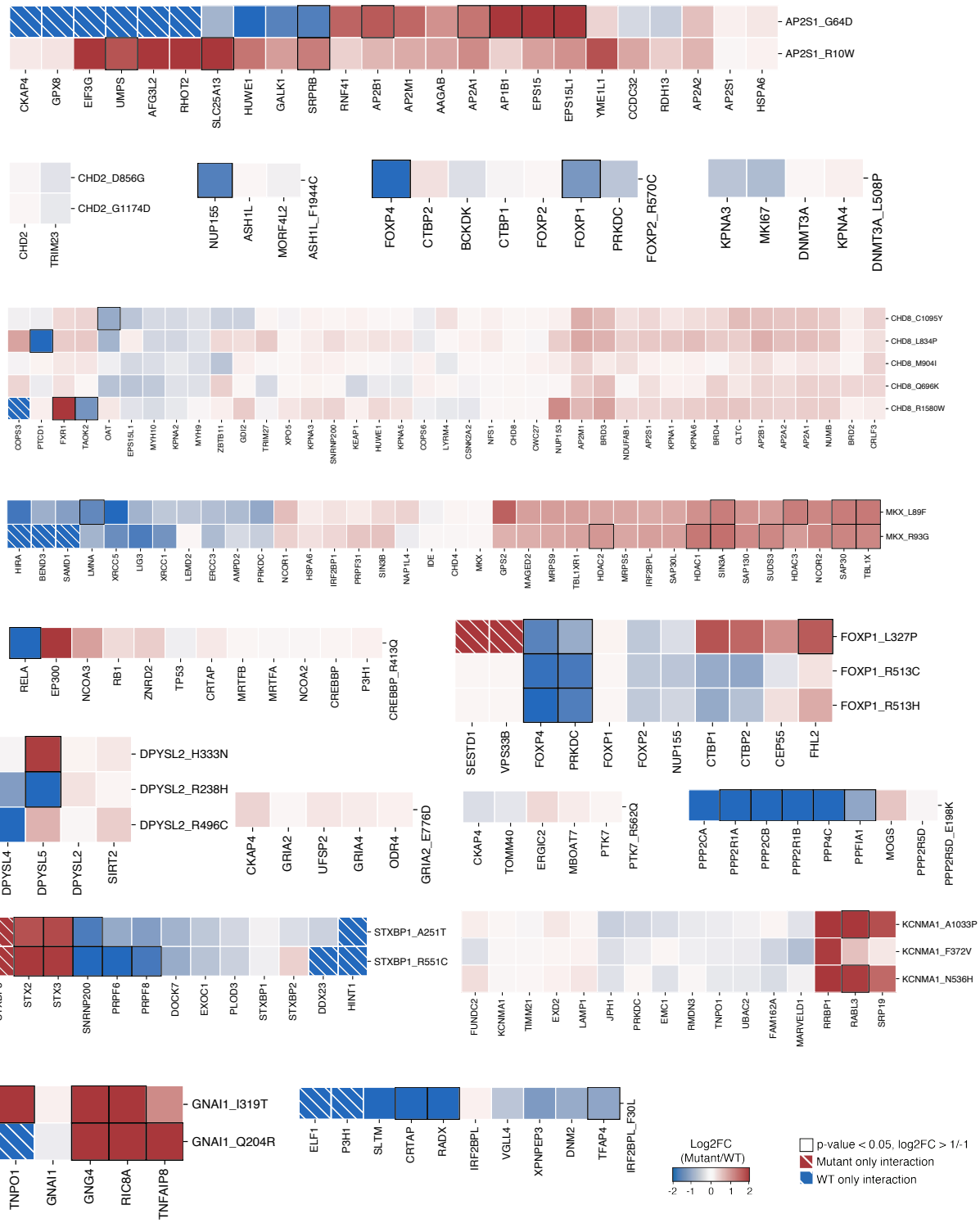


Figure S3.2. Heatmap visualizing the differential interactors in mutant hcASD proteins (Part 1). Quantification of the high-confidence PPIs for the subset of ASD missense mutants.

Protein interactors are ordered by hierarchical clustering of the log2 fold-change values. The dendrograms are hidden for simplicity.

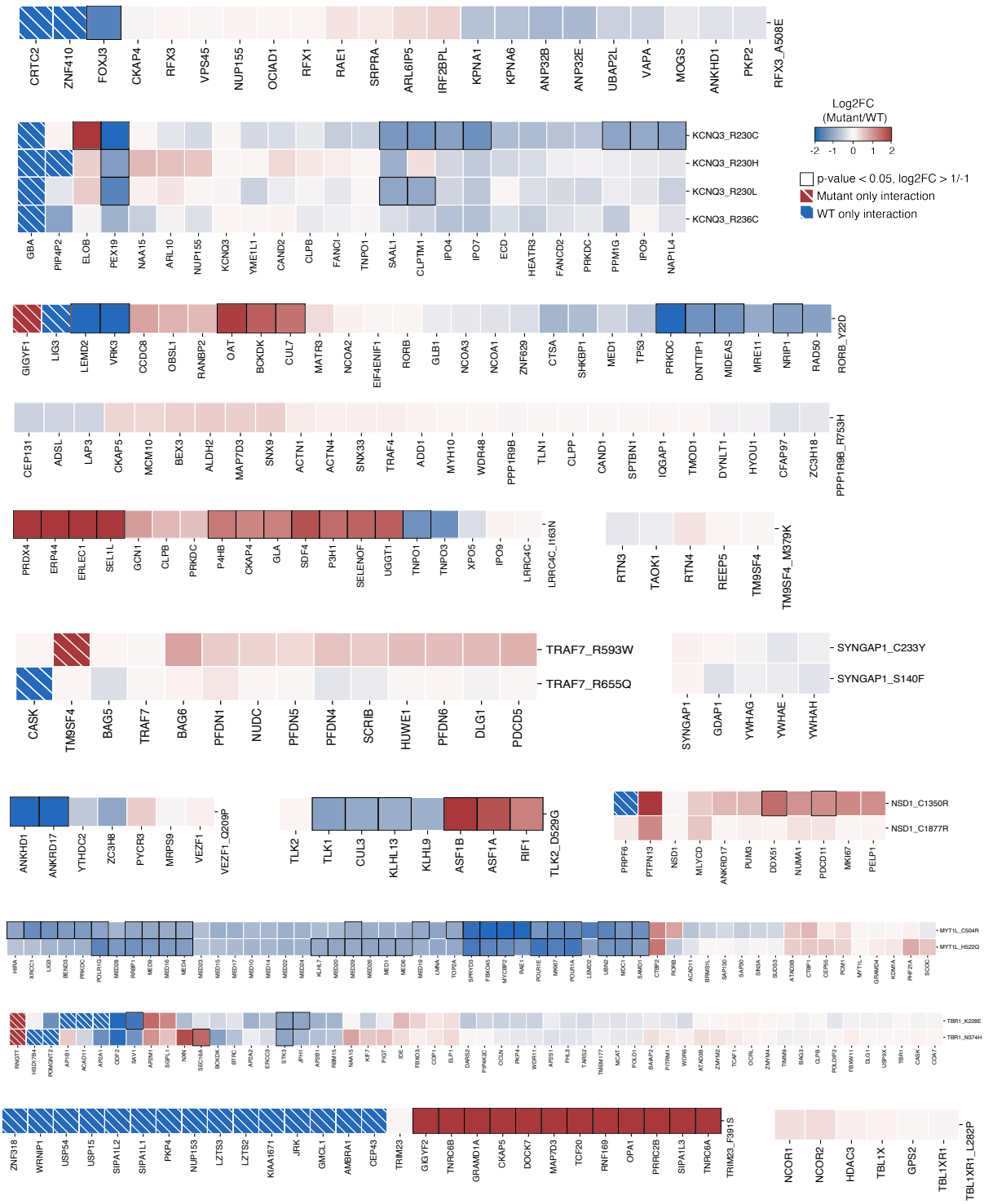


Figure S3.2. Heatmap visualizing the differential interactors in mutant hcASD proteins (Part 2). Quantification of the high confident PPIs for the subset of ASD missense mutants.

Protein interactors are ordered by hierarchical clustering of the log2 fold-change values. The dendrograms are hidden for simplicity.

CHAPTER 4:

DISCUSSION

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder with a complex and highly heterogeneous genetic architecture that affects 1 in 44 children in the United States (Maenner, 2021). There have been tremendous advances in the understanding of the ASD genetic architecture and the identification of large-effect ASD risk genes in the past decade and yet, the translation of this knowledge into molecular mechanisms underlying ASD is still lagging. To understand how these risk genes play a role in ASD neurobiology, it is imperative to first comprehend how the proteins of these genes are interacting with other proteins to carry out molecular functions. A protein-protein interaction map can highlight shared functional relationships between different proteins and pinpoint the pathways which are disrupted with the introduction of mutations. In this study, we mapped the protein-protein interaction landscape of ASD in HEK293T cells using the 102 hcASD risk genes identified in Satterstrom et al., 2020 as a starting point. Our ASD-PPI network identified 1074 unique proteins that are connected to 100 hcASD risk genes by 1881 interactions, 87% of which are novel (**Figure 2.1**). We also demonstrated that our hcASD protein interactome generated in HEK293T cells are ASD-relevant by showing that the interacting proteins (preys) identified are 1) expressed in the human brain and enriched in the prenatal period, consistent with the expression pattern of the hcASD risk genes and are 2) enriched for ASD genetic risk (**Figure 2.4**). Through a protein complex and gene ontology enrichment analysis, we showed that hcASD risk genes shared protein interactors that are either of the same protein complexes or involved in similar biological fuctions (**Figure 2.3**). This systematic mapping of protein interactome in ASD can be used to uncovering new biological processes and molecular mechanisms underlying ASD etiology.

In addition to uncovering many new proteins that are enriched in ASD genetic risk, our ASD-PPI also highlighted many direct interactions between hcASD risk genes. One such interaction that we focused on was DYRK1A and KIAA0232. DYRK1A is a conserved kinase that

plays an important role in brain development and its haploinsufficiency has been implicated in microcephaly in patients with intellectual disabilities and reduced brain size in model organisms (Courcet et al., 2012; van Bon et al., 1993; H. R. Willsey et al., 2020). Contrary to the well-studied DYRK1A, little is known about the uncharacterized protein - KIAA0232, which interacts with DYRK1A in our ASD-PPI (**Figure 2.5**). We confirmed this direct interaction via a co-immunoprecipitation followed by mass spectrometry (**Figure 2.6**). Since interacting proteins are likely to share the same function, we hypothesized that, like DYRK1A, KIAA0232 is involved in brain development and that the haploinsufficiency of *KIAA0232* will lead to reduced brain size. We found that in the model organism *Xenopus tropicalis,* knock-down of both *dyrk1a* and *kiaa0232* resulted in reduced telencephalon size, which recapitulates prior findings for DYRK1A and unveils a novel function for KIAA0232 (**Figure 2.7**). In addition to the direct interaction with KIAA0232, we showed that DYRK1A forms a protein complex with DCAF7, CCDC8 and ATAD3B (**Figure 2.5**, **Figure 2.6**). Further studies should investigate the role of this protein complex in neurodevelopment, including whether haploinsufficiency of the complex members phenocopy haploinsufficiency of *DYRK1A* or *KIAA0232* in reduced brain size.

ASD is generally considered to be caused by impaired brain development, which is tightly regulated by a series of events such as cellular proliferation, migration, and differentiation (Hardwick et al., n.d.). The missense mutations found in high confidence ASD risk genes are predicted to have deleterious effects (Satterstrom et al., 2020). However, the phenotype of these mutants in the neuronal differentiation process, and the mechanisms that lead to these phenotypes remain elusive. In this study, we systematically assessed the impact of ASD patient-derived *de novo* damaging missense mutations in disrupting the protein-protein interaction landscape of ASD. We generated PPI data for 64 missense mutations in 35 hcASD risk genes and found that of the 929 high-confident interactors in the mutant ASD protein-protein interaction network (ASD_mut-PPI), 133 interacting proteins showed higher affinity for the mutant proteins and 152 for the wild-type proteins (**Figure 3.1**). Our hierarchical clustering analysis of the high-

confident interactors based on their log2 fold-change levels of intensities show that there is a common pattern of changes in interactors among the ASD mutants of the same gene (**Figure S3.2**, **S3.3**). Because of this, we posit that there is a convergent molecular function among the different mutants of the same gene and that the protein interactors commonly lost or gained in mutant proteins can be exploited to provide insights into the functional pathways that are disrupted in ASD.

One such interaction loss that we studied was the loss of interaction with FOXP4 for the patient-derived mutations in FOXP1 and FOXP2. We confirmed the loss of FOXP4 interaction by Western blot assay (**Figure 3.5**). We postulated that this loss of interaction that was observed in HEK293T would be recapitulated in neurons and would likely impact neuronal differentiation. Indeed, our data showed that when mutant FOXP1 was overexpressed in neural progenitor cells (NPCs), it formed nuclear condensates (**Figure 3.6**), and when NPCs overexpressing mutant FOXP1 were differentiated into cortical glutaminergic neurons, they demonstrated decreased neurite length (**Figure 3.7**). We posit that mutant FOXP1 loses its ability to interact with FOXP4, which impairs its ability to bind to DNA; the resulting transcriptional dysregulation proper dendritic development.

In summary, our study provides the most comprehensive protein interactome for ASD which enable us to identify convergent functional pathways through which ASD risk genes may participate in brain development, a connection necessary to bridge gene discovery and translational research. Importantly, we showed that building a protein-protein interaction map in HEK293T cells to study ASD is both scalable and ASD-relevant. This systematic approach of building a hypothesis-free protein-protein interaction network can reveal interactors that are critical for the molecular pathways and cellular processes dysregulated by ASD variants and can be used as a resource for the identification of biomarkers and therapeutic targets. Furthermore, this framework can be adapted and expanded to study other genetically heterogeneous disorders.

REFERENCES

1. Abrahams, B. S., Arking, D. E., Campbell, D. B., Mefford, H. C., Morrow, E. M., Weiss, L. A., Menashe, I., Wadkins, T., Banerjee-Basu, S., & Packer, A. (2013). SFARI Gene 2.0: A community-driven knowledgebase for the autism spectrum disorders (ASDs). *Molecular Autism*, *4*(1), 36. https://doi.org/10.1186/2040-2392-4-36

2. Aguet, F., Brown, A. A., Castel, S. E., Davis, J. R., He, Y., Jo, B., Mohammadi, P., Park, Y., Parsana, P., Segrè, A. V., Strober, B. J., Zappala, Z., Cummings, B. B., Gelfand, E. T., Hadley, K., Huang, K. H., Lek, M., Li, X., Nedzel, J. L., … Biospecimen Collection Source Site—NDRI. (2017). Genetic effects on gene expression across human tissues. *Nature*, *550*(7675), Article 7675. https://doi.org/10.1038/nature24277

3. American Psychological Association (APA). (2013). Diagnostic and Statistical Manual of Mental Disorders: Depressive Disorders. In *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition*. American Psychiatric Publishing, Inc. https://doi.org/10.1176/appi.books.9780890425596.dsm04

4. Bacon, C., & Rappold, G. A. (2012). The distinct and overlapping phenotypic spectra of FOXP1 and FOXP2 in cognitive disorders. *Human Genetics*, *131*(11), 1687–1698. https://doi.org/10.1007/s00439-012-1193-z

5. Banerjee-Basu, S., & Packer, A. (2010). SFARI Gene: An evolving database for the autism research community. *Disease Models & Mechanisms*, *3*(3–4), 133–135. https://doi.org/10.1242/dmm.005439

6. Bekker-Jensen, D. B., Kelstrup, C. D., Batth, T. S., Larsen, S. C., Haldrup, C., Bramsen, J. B., Sørensen, K. D., Høyer, S., Ørntoft, T. F., Andersen, C. L., Nielsen, M. L., & Olsen, J. V. (2017). An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes. *Cell Systems*, *4*(6), 587-599.e4. https://doi.org/10.1016/j.cels.2017.05.009

7. Ben-David, E., & Shifman, S. (2013). Combined analysis of exome sequencing points toward a major role for transcription regulation during brain development in autism. *Molecular Psychiatry*, *18*(10), Article 10. https://doi.org/10.1038/mp.2012.148

8. Bouhaddou, M., Eckhardt, M., Chi Naing, Z. Z., Kim, M., Ideker, T., & Krogan, N. J. (2019). Mapping the protein–protein and genetic interactions of cancer to guide precision medicine. *Current Opinion in Genetics & Development*, *54*, 110–117. https://doi.org/10.1016/j.gde.2019.04.005

9. Buxbaum, J. D., Daly, M. J., Devlin, B., Lehner, T., Roeder, K., & State, M. W. (2012). The Autism Sequencing Consortium: Large scale, high throughput sequencing in autism spectrum disorders. *Neuron*, *76*(6), 10.1016/j.neuron.2012.12.008. https://doi.org/10.1016/j.neuron.2012.12.008

10. Cassa, C. A., Weghorn, D., Balick, D. J., Jordan, D. M., Nusinow, D., Samocha, K. E., O'Donnell-Luria, A., MacArthur, D. G., Daly, M. J., Beier, D. R., & Sunyaev, S. R. (2017). Estimating the selective effects of heterozygous protein-truncating variants from human exome data. *Nature Genetics*, *49*(5), 806–810. https://doi.org/10.1038/ng.3831

11. CDC. (2022, March 2). *Data and Statistics on Autism Spectrum Disorder | CDC*. Centers for Disease Control and Prevention. https://www.cdc.gov/ncbddd/autism/data.html

12. Cho, N. H., Cheveralls, K. C., Brunner, A.-D., Kim, K., Michaelis, A. C., Raghavan, P., Kobayashi, H., Savy, L., Li, J. Y., Canaj, H., Kim, J. Y. S., Stewart, E. M., Gnann, C., McCarthy, F., Cabrera, J. P., Brunetti, R. M., Chhun, B. B., Dingle, G., Hein, M. Y., … Leonetti, M. D. (2022). OpenCell: Endogenous tagging for the cartography of human cellular organization. *Science*, *375*(6585), eabi6983. https://doi.org/10.1126/science.abi6983

13. Choi, M., Chang, C.-Y., Clough, T., Broudy, D., Killeen, T., MacLean, B., & Vitek, O. (2014). MSstats: An R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics*, *30*(17), 2524–2526. https://doi.org/10.1093/bioinformatics/btu305

14. Co, M., Anderson, A. G., & Konopka, G. (2020). FOXP transcription factors in vertebrate brain development, function, and disorders. *Wiley Interdisciplinary Reviews: Developmental Biology*. https://doi.org/10.1002/wdev.375

15. Courcet, J.-B., Faivre, L., Malzac, P., Masurel-Paulet, A., Lopez, E., Callier, P., Lambert, L., Lemesle, M., Thevenon, J., Gigot, N., Duplomb, L., Ragon, C., Marle, N., Mosca-Boidron, A.-L., Huet, F., Philippe, C., Moncla, A., & Thauvin-Robinet, C. (2012). The DYRK1A gene is a cause of syndromic intellectual disability with severe microcephaly and epilepsy. *Journal of Medical Genetics*, *49*(12), 731–736. https://doi.org/10.1136/jmedgenet-2012-101251

16. Cox, J., & Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*, *26*(12), 1367–1372. https://doi.org/10.1038/nbt.1511

17. Dang, T., Duan, W. Y., Yu, B., Tong, D. L., Cheng, C., Zhang, Y. F., Wu, W., Ye, K., Zhang, W. X., Wu, M., Wu, B. B., An, Y., Qiu, Z. L., & Wu, B. L. (2018). Autism-associated Dyrk1a truncation mutants impair neuronal dendritic and spine growth and interfere with postnatal cortical development. *Molecular Psychiatry*, *23*(3), 747–758. https://doi.org/10.1038/mp.2016.253

18. De Rubeis, S., He, X., Goldberg, A. P., Poultney, C. S., Samocha, K., Cicek, A. E., Kou, Y., Liu, L., Fromer, M., Walker, S., Singh, T., Klei, L., Kosmicki, J., Shih-Chen, F., Aleksic, B., Biscaldi, M., Bolton, P. F., Brownfeld, J. M., Cai, J., … Buxbaum, J. D. (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature*, *515*(7526), 209–215. https://doi.org/10.1038/nature13772

19. Dougherty, J. D., Marrus, N., Maloney, S. E., Yip, B., Sandin, S., Turner, T. N., Selmanovic, D., Kroll, K. L., Gutmann, D. H., Constantino, J. N., & Weiss, L. A. (2022). Can the “female protective effect” liability threshold model explain sex differences in autism spectrum disorder? *Neuron*. https://doi.org/10.1016/j.neuron.2022.06.020

20. Fischbach, G. D., & Lord, C. (2010). The Simons Simplex Collection: A Resource for Identification of Autism Genetic Risk Factors. *Neuron*, *68*(2), 192–195. https://doi.org/10.1016/j.neuron.2010.10.006

21. Fu, J. M., Satterstrom, F. K., Peng, M., Brand, H., Collins, R. L., Dong, S., Wamsley, B., Klei, L., Wang, L., Hao, S. P., Stevens, C. R., Cusick, C., Babadi, M., Banks, E., Collins, B., Dodge, S., Gabriel, S. B., Gauthier, L., Lee, S. K., … Talkowski, M. E. (2022). Rare coding variation provides insight into the genetic architecture and phenotypic context of autism. *Nature Genetics*, 1–12. https://doi.org/10.1038/s41588-022-01104-0

22. Gaugler, T., Klei, L., Sanders, S. J., Bodea, C. A., Goldberg, A. P., Lee, A. B., Mahajan, M., Manaa, D., Pawitan, Y., Reichert, J., Ripke, S., Sandin, S., Sklar, P., Svantesson, O., Reichenberg, A., Hultman, C. M., Devlin, B., Roeder, K., & Buxbaum, J. D. (2014). Most genetic risk for autism resides with common variation. *Nature Genetics*, *46*(8), 881–885. https://doi.org/10.1038/ng.3039

23. Giurgiu, M., Reinhard, J., Brauner, B., Dunger-Kaltenbach, I., Fobo, G., Frishman, G., Montrone, C., & Ruepp, A. (2019a). CORUM: the comprehensive resource of mammalian protein complexes—2019. *Nucleic Acids Research*, *47*(D1), D559–D563. https://doi.org/10.1093/nar/gky973

24. Giurgiu, M., Reinhard, J., Brauner, B., Dunger-Kaltenbach, I., Fobo, G., Frishman, G., Montrone, C., & Ruepp, A. (2019b). CORUM: the comprehensive resource of mammalian protein

complexes—2019. *Nucleic Acids Research*, *47*(D1), D559–D563. https://doi.org/10.1093/nar/gky973

25. Go, C. D., Knight, J. D. R., Rajasekharan, A., Rathod, B., Hesketh, G. G., Abe, K. T., Youn, J.-Y., Samavarchi-Tehrani, P., Zhang, H., Zhu, L. Y., Popiel, E., Lambert, J.-P., Coyaud, É., Cheung, S. W. T., Rajendran, D., Wong, C. J., Antonicka, H., Pelletier, L., Palazzo, A. F., … Gingras, A.-C. (2021). A proximity-dependent biotinylation map of a human cell. *Nature*, *595*(7865), Article 7865. https://doi.org/10.1038/s41586-021-03592-2

26. Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., O'Meara, M. J., Rezelj, V. V., Guo, J. Z., Swaney, D. L., Tummino, T. A., Huettenhain, R., Kaake, R. M., Richards, A. L., Tutuncuoglu, B., Foussard, H., Batra, J., Haas, K., Modak, M., … Krogan, N. J. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*, 1–13. https://doi.org/10.1038/s41586-020-2286-9

27. Grove, J., Ripke, S., Als, T. D., Mattheisen, M., Walters, R., Won, H., Pallesen, J., Agerbo, E., Andreassen, O. A., Anney, R., Belliveau, R., Bettella, F., Buxbaum, J. D., Bybjerg-Grauholm, J., Baekved-Hansen, M., Cerrato, F., Chambert, K., Christensen, J. H., Churchhouse, C., … Børglum, A. D. (2017). *Common risk variants identified in autism spectrum disorder* (Vol. 33, p. 42).

28. Hardwick, L. J. A., Ali, F. R., Azzarelli, R., & Philpott, A. (2015). Cell cycle regulation of proliferation versus differentiation in the central nervous system. *Cell and Tissue Research*, *359*(1), 187–200. https://doi.org/10.1007/s00441-014-1895-8

29. Hein, M. Y., Hubner, N. C., Poser, I., Cox, J., Nagaraj, N., Toyoda, Y., Gak, I. A., Weisswange, I., Mansfeld, J., Buchholz, F., Hyman, A. A., & Mann, M. (2015). A Human Interactome in Three Quantitative Dimensions Organized by Stoichiometries and Abundances. *Cell*, *163*(3), 712–723. https://doi.org/10.1016/j.cell.2015.09.053

30. Huttlin, E. L., Ting, L., Bruckner, R. J., Gebreab, F., Gygi, M. P., Szpyt, J., Tam, S., Zarraga, G., Colby, G., Baltier, K., Dong, R., Guarani, V., Vaites, L. P., Ordureau, A., Rad, R., Erickson, B. K., Wühr, M., Chick, J., Zhai, B., … Gygi, S. P. (2015). The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell*, *162*(2), 425–440. https://doi.org/10.1016/J.CELL.2015.06.043

31. Hyman, S. L., Levy, S. E., Myers, S. M., & COUNCIL ON CHILDREN WITH DISABILITIES, SECTION ON DEVELOPMENTAL AND BEHAVIORAL PEDIATRICS. (2020). Identification, Evaluation, and Management of Children With Autism Spectrum Disorder. *Pediatrics*, *145*(1), e20193447. https://doi.org/10.1542/peds.2019-3447

32. Iossifov, I., O'Roak, B. J., Sanders, S. J., Ronemus, M., Krumm, N., Levy, D., Stessman, H. A., Witherspoon, K. T., Vives, L., Patterson, K. E., Smith, J. D., Paeper, B., Nickerson, D. A., Dea, J., Dong, S., Gonzalez, L. E., Mandell, J. D., Mane, S. M., Murtha, M. T., … Wigler, M. (2014). The contribution of de novo coding mutations to autism spectrum disorder. *Nature*, *515*(7526), 216–221. https://doi.org/10.1038/nature13908

33. Jimenez-Morales, D., Campos, A. R., Dollen, J. V., Krogan, N., & Swaney, D. (2022). *ArtMS: Analytical R tools for Mass Spectrometry*. http://artms.org

34. Johnson, T. B., Mechels, K., Anderson, R. H., Cain, J. T., Sturdevant, D. A., Braddock, S., Pinz, H., Wilson, M. A., Landsverk, M., Roux, K. J., & Weimer, J. M. (2018). Characterization of a recurrent missense mutation in the forkhead DNA-binding domain of FOXP1. *Scientific Reports*, *8*(1), 16161. https://doi.org/10.1038/s41598-018-34437-0

35. Kang, H. J., Kawasawa, Y. I., Cheng, F., Zhu, Y., Xu, X., Li, M., Sousa, A. M. M., Pletikos, M., Meyer, K. A., Sedmak, G., Guennel, T., Shin, Y., Johnson, M. B., Krsnik, Ž., Mayer, S., Fertuzinhos, S., Umlauf, S., Lisgo, S. N., Vortmeyer, A., … Šestan, N. (2011). Spatio-temporal

transcriptome of the human brain. *Nature*, *478*(7370), 483–489. https://doi.org/10.1038/nature10523

36. Kaplanis, J., Ide, B., Sanghvi, R., Neville, M., Danecek, P., Coorens, T., Prigmore, E., Short, P., Gallone, G., McRae, J., Carmichael, J., Barnicoat, A., Firth, H., O'Brien, P., Rahbari, R., & Hurles, M. (2022). Genetic and chemotherapeutic influences on germline hypermutation. *Nature*, *605*(7910), Article 7910. https://doi.org/10.1038/s41586-022-04712-2

37. Kaplanis, J., Samocha, K. E., Wiel, L., Zhang, Z., Arvai, K. J., Eberhardt, R. Y., Gallone, G., Lelieveld, S. H., Martin, H. C., McRae, J. F., Short, P. J., Torene, R. I., de Boer, E., Danecek, P., Gardner, E. J., Huang, N., Lord, J., Martincorena, I., Pfundt, R., … Retterer, K. (2020). Evidence for 28 genetic disorders discovered by combining healthcare and research data. *Nature*, *586*(7831), 757–762. https://doi.org/10.1038/s41586-020-2832-5

38. Klei, L., Sanders, S. J., Murtha, M. T., Hus, V., Lowe, J. K., Willsey, A. J., Moreno-De-Luca, D., Yu, T. W., Fombonne, E., Geschwind, D., Grice, D. E., Ledbetter, D. H., Lord, C., Mane, S. M., Martin, C. L., Martin, D. M., Morrow, E. M., Walsh, C. A., Melhem, N. M., … Devlin, B. (2012). Common genetic variants, acting additively, are a major source of risk for autism. *Molecular Autism*, *3*(1), 9. https://doi.org/10.1186/2040-2392-3-9

39. Kosmicki, J. A., Samocha, K. E., Howrigan, D. P., Sanders, S. J., Slowikowski, K., Lek, M., Karczewski, K. J., Cutler, D. J., Devlin, B., Roeder, K., Buxbaum, J. D., Neale, B. M., MacArthur, D. G., Wall, D. P., Robinson, E. B., & Daly, M. J. (2017). Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. *Nature Genetics*, *49*(4), Article 4. https://doi.org/10.1038/ng.3789

40. Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., O'Donnell-Luria, A. H., Ware, J. S., Hill, A. J., Cummings, B. B., Tukiainen, T., Birnbaum, D. P., Kosmicki, J. A., Duncan, L. E., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J., … Williams, A.

L. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, *536*(7616), 285–291. https://doi.org/10.1038/nature19057

41. Li, M., Santpere, G., Imamura Kawasawa, Y., Evgrafov, O. V., Gulden, F. O., Pochareddy, S., Sunkin, S. M., Li, Z., Shin, Y., Zhu, Y., Sousa, A. M. M., Werling, D. M., Kitchen, R. R., Kang, H. J., Pletikos, M., Choi, J., Muchnik, S., Xu, X., Wang, D., … Sestan, N. (2018). Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science*, *362*(6420), eaat7615. https://doi.org/10.1126/science.aat7615

42. Li, S., Weidenfeld, J., & Morrisey, E. E. (2004). Transcriptional and DNA Binding Activity of the Foxp1/2/4 Family Is Modulated by Heterotypic and Homotypic Protein Interactions. *Molecular and Cellular Biology*, *24*(2), 809–822. https://doi.org/10.1128/MCB.24.2.809-822.2004

43. Li, T., Wernersson, R., Hansen, R. B., Horn, H., Mercer, J., Slodkowicz, G., Workman, C. T., Rigina, O., Rapacki, K., Stærfeldt, H. H., Brunak, S., Jensen, T. S., & Lage, K. (2017). A scored human protein–protein interaction network to catalyze genomic interpretation. *Nature Methods*, *14*(1), 61–64. https://doi.org/10.1038/nmeth.4083

44. Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., Foster, B., Moser, M., Karasik, E., Gillard, B., Ramsey, K., Sullivan, S., Bridge, J., Magazine, H., Syron, J., … Moore, H. F. (2013a). The Genotype-Tissue Expression (GTEx) project. *Nature Genetics*, *45*(6), Article 6. https://doi.org/10.1038/ng.2653

45. Maenner, M. J. (2021). Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years—Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2018. *MMWR. Surveillance Summaries*, *70*. https://doi.org/10.15585/mmwr.ss7011a1

46. Nieuwkoop, P. D., Faber, J., Gerhart, J., & Kirschner, M. (1994). *Normal table of Xenopus laevis (Daudin): A systematical and chronological survey of the development from the fertilized egg till the end of metamorphosis*. Garland Science.

47. O'Roak, B. J., & State, M. W. (2008). Autism genetics: Strategies, challenges, and opportunities. *Autism Research: Official Journal of the International Society for Autism Research*, *1*(1), 4–17. https://doi.org/10.1002/aur.3

48. Oughtred, R., Stark, C., Breitkreutz, B.-J., Rust, J., Boucher, L., Chang, C., Kolas, N., O'donnell, L., Leung, G., Mcadam, R., Zhang, F., Dolma, S., Willems, A., Coulombe-Huntington, J., Chatr-Aryamontri, A., Dolinski, K., & Tyers, M. (2018). The BioGRID interaction database: 2019 update. *Nucleic Acids Research*, *47*, 529–541. https://doi.org/10.1093/nar/gky1079

49. Parikshak, N. N., Luo, R., Zhang, A., Won, H., Lowe, J. K., Chandran, V., Horvath, S., & Geschwind, D. H. (2013). Integrative Functional Genomic Analyses Implicate Specific Molecular Pathways and Circuits in Autism. *Cell*, *155*(5), 1008–1021. https://doi.org/10.1016/j.cell.2013.10.031

50. Perica, T., Mathy, C. J. P., Xu, J., Jang, G. M., Zhang, Y., Kaake, R., Ollikainen, N., Braberg, H., Swaney, D. L., Lambright, D. G., Kelly, M. J. S., Krogan, N. J., & Kortemme, T. (2021). Systems-level effects of allosteric perturbations to a model molecular switch. *Nature*, *599*(7883), Article 7883. https://doi.org/10.1038/s41586-021-03982-6

51. Petryshen, T., Rennert, O. M., Kennedy, E., National, S., Werling, D. M., Rodriguez-Fontenla, C., Alonso-Gonzalez, A., & Carracedo, A. (2018). *De novo Mutations (DNMs) in Autism Spectrum Disorder (ASD): Pathway and Network Analysis*. https://doi.org/10.3389/fgene.2018.00406

52. Qi, Y., Zhang, X.-J., Renier, N., Wu, Z., Atkin, T., Sun, Z., Ozair, M. Z., Tchieu, J., Zimmer, B., Fattahi, F., Ganat, Y., Azevedo, R., Zeltner, N., Brivanlou, A. H., Karayiorgou, M., Gogos, J., Tomishima, M., Tessier-Lavigne, M., Shi, S.-H., & Studer, L. (2017). Combined small-molecule

inhibition accelerates the derivation of functional, early-born, cortical neurons from human pluripotent stem cells. *Nature Biotechnology*, *35*(2), 154–163. https://doi.org/10.1038/nbt.3777

53. Roewenstrunk, J., Di Vona, C., Chen, J., Borras, E., Dong, C., Arató, K., Sabidó, E., Huen, M. S. Y., & de la Luna, S. (2019). A comprehensive proteomics-based interaction screen that links DYRK1A to RNF169 and to the DNA damage response. *Scientific Reports*, *9*(1), 1–14. https://doi.org/10.1038/s41598-019-42445-x

54. Ronemus, M., Iossifov, I., Levy, D., & Wigler, M. (2014). The role of de novo mutations in the genetics of autism spectrum disorders. *Nature Reviews Genetics*, *15*(2), 133–141. https://doi.org/10.1038/nrg3585

55. Samocha, K. E., Kosmicki, J. A., Karczewski, K. J., O'Donnell-Luria, A. H., Pierce-Hoffman, E., MacArthur, D. G., Neale, B. M., & Daly, M. J. (2017). *Regional missense constraint improves variant deleteriousness prediction* (p. 148353). bioRxiv. https://doi.org/10.1101/148353

56. Sanders, S. J., He, X., Willsey, A. J., Ercan-Sencicek, A. G., Samocha, K. E., Cicek, A. E., Murtha, M. T., Bal, V. H., Bishop, S. L., Dong, S., Goldberg, A. P., Jinlu, C., Keaney, J. F., Klei, L., Mandell, J. D., Moreno-De-Luca, D., Poultney, C. S., Robinson, E. B., Smith, L., … State, M. W. (2015). Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron*, *87*(6), 1215–1233. https://doi.org/10.1016/j.neuron.2015.09.016

57. Sandin, S., Lichtenstein, P., Kuja-Halkola, R., Hultman, C., Larsson, H., & Reichenberg, A. (2017). The Heritability of Autism Spectrum Disorder. *JAMA*, *318*(12), 1182–1184. https://doi.org/10.1001/jama.2017.12141

58. Satterstrom, F. K., Kosmicki, J. A., Wang, J., Breen, M. S., De Rubeis, S., An, J. Y., Peng, M., Collins, R., Grove, J., Klei, L., Stevens, C., Reichert, J., Mulhern, M. S., Artomov, M., Gerges, S., Sheppard, B., Xu, X., Bhaduri, A., Norman, U., … Buxbaum, J. D. (2020). Large-Scale Exome

Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell*, *180*(3), 568-584.e23. https://doi.org/10.1016/j.cell.2019.12.036

59. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., & others. (2012). Fiji: An open-source platform for biological-image analysis. *Nature Methods*, *9*(7), 676–682.

60. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research*, *13*(11), 2498–2504. https://doi.org/10.1101/gr.1239303

61. Sin, C., Li, H., & Crawford, D. A. (2015). Transcriptional Regulation by FOXP1, FOXP2, and FOXP4 Dimerization. *Journal of Molecular Neuroscience*, *55*(2), 437–448. https://doi.org/10.1007/S12031-014-0359-7/FIGURES/3

62. Singh, T., Poterba, T., Curtis, D., Akil, H., Al Eissa, M., Barchas, J. D., Bass, N., Bigdeli, T. B., Breen, G., Bromet, E. J., Buckley, P. F., Bunney, W. E., Bybjerg-Grauholm, J., Byerley, W. F., Chapman, S. B., Chen, W. J., Churchhouse, C., Craddock, N., Cusick, C. M., … Daly, M. J. (2022). Rare coding variants in ten genes confer substantial risk for schizophrenia. *Nature 2022 604:7906*, *604*(7906), 509–516. https://doi.org/10.1038/s41586-022-04556-w

63. Sive, H. L., Grainger, R. M., & Harland, R. M. (2000). *Early development of Xenopus laevis: A laboratory manual*. CSHL Press.

64. Sollis, E., Deriziotis, P., Saitsu, H., Miyake, N., Matsumoto, N., Hoffer, M. J. V., Ruivenkamp, C. A. L., Alders, M., Okamoto, N., Bijlsma, E. K., Plomp, A. S., & Fisher, S. E. (2017). Equivalent missense variant in the FOXP2 and FOXP1 transcription factors causes distinct neurodevelopmental disorders. *Human Mutation*, *38*(11), 1542–1554. https://doi.org/10.1002/humu.23303

65. Sowa, M. E., Bennett, E. J., Gygi, S. P., & Harper, J. W. (2009). Defining the Human Deubiquitinating Enzyme Interaction Landscape. *Cell*, *138*(2), 389–403. https://doi.org/10.1016/j.cell.2009.04.042

66. Swaney, D. L., Ramms, D. J., Wang, Z., Park, J., Goto, Y., Soucheray, M., Bhola, N., Kim, K., Zheng, F., Zeng, Y., McGregor, M., Herrington, K. A., O'Keefe, R., Jin, N., VanLandingham, N. K., Foussard, H., Von Dollen, J., Bouhaddou, M., Jimenez-Morales, D., … Krogan, N. J. (2021). A protein network map of head and neck cancer reveals PIK3CA mutant drug sensitivity. *Science*, *374*(6563). https://doi.org/10.1126/SCIENCE.ABF2911/SUPPL_FILE/SCIENCE.ABF2911_DATA_S1_TO_S7.ZIP

67. Teo, G., Liu, G., Zhang, J., Nesvizhskii, A. I., & Choi, H. (2014). SAINTexpress: Improvements and additional features in Significance Analysis of INTeractome software. *Journal of Proteomics*, *100*, 37–43. https://doi.org/10.1016/J.JPROT.2013.10.023

68. Tick, B., Bolton, P., Happé, F., Rutter, M., & Rijsdijk, F. (2016). Heritability of autism spectrum disorders: A meta-analysis of twin studies. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *57*(5). https://doi.org/10.1111/jcpp.12499

69. Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Szigyarto, C. A.-K., Odeberg, J., Djureinovic, D., Takanen, J. O., Hober, S., … Pontén, F. (2015). Proteomics. Tissue-based map of the human proteome. *Science (New York, N.Y.)*, *347*(6220), 1260419. https://doi.org/10.1126/science.1260419

70. van Bon, B. W., Coe, B. P., de Vries, B. B., & Eichler, E. E. (1993). DYRK1A Syndrome. In M. P. Adam, D. B. Everman, G. M. Mirzaa, R. A. Pagon, S. E. Wallace, L. J. Bean, K. W. Gripp,

& A. Amemiya (Eds.), *GeneReviews®*. University of Washington, Seattle. http://www.ncbi.nlm.nih.gov/books/NBK333438/

71. Vu, D. H., Koster, R. A., Wessels, A. M. A., Greijdanus, B., Alffenaar, J. W. C., & Uges, D. R. A. (2013). Troubleshooting carry-over of LC-MS/MS method for rifampicin, clarithromycin and metabolites in human plasma. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences*, *917–918*, 1–4. https://doi.org/10.1016/j.jchromb.2012.12.023

72. Werling, D. M., & Geschwind, D. H. (2013). Sex differences in autism spectrum disorders. *Current Opinion in Neurology*, *26*(2), 146–153. https://doi.org/10.1097/WCO.0b013e32835ee548

73. Willsey, A. J., Morris, M. T., Wang, S., Willsey, H. R., Sun, N., Teerikorpi, N., Baum, T. B., Cagney, G., Bender, K. J., Desai, T. A., Srivastava, D., Davis, G. W., Doudna, J., Chang, E., Sohal, V., Lowenstein, D. H., Li, H., Agard, D., Keiser, M. J., … Krogan, N. J. (2018). The Psychiatric Cell Map Initiative: A Convergent Systems Biological Approach to Illuminating Key Molecular Pathways in Neuropsychiatric Disorders. *Cell*, *174*(3), 505–520. https://doi.org/10.1016/j.cell.2018.06.016

74. Willsey, A. J., Sanders, S. J., Li, M., Dong, S., Tebbenkamp, A. T., Muhle, R. A., Reilly, S. K., Lin, L., Fertuzinhos, S., Miller, J. A., Murtha, M. T., Bichsel, C., Niu, W., Cotney, J., Ercan-Sencicek, A. G., Gockley, J., Gupta, A. R., Han, W., He, X., … State, M. W. (2013). Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell*, *155*(5), 997–1007. https://doi.org/10.1016/j.cell.2013.10.020

75. Willsey, H. R., Exner, C. R. T., Xu, Y., Everitt, A., Sun, N., Wang, B., Dea, J., Schmunk, G., Zaltsman, Y., Teerikorpi, N., Kim, A., Anderson, A. S., Shin, D., Seyler, M., Nowakowski, T. J., Harland, R. M., Willsey, A. J., & State, M. W. (2021). Parallel in vivo analysis of large-effect autism genes implicates cortical neurogenesis and estrogen in risk and resilience. *Neuron*, *109*(5), 788-804.e8. https://doi.org/10.1016/J.NEURON.2021.01.002

76. Willsey, H. R., Walentek, P., Exner, C. R., Xu, Y., Lane, A. B., Harland, R. M., Heald, R., & Santama, N. (2018). Katanin-like protein Katnal2 is required for ciliogenesis and brain development in Xenopus embryos. *Developmental Biology*, *442*(2), 276–287.

77. Willsey, H. R., Xu, Y., Everitt, A., Dea, J., Exner, C. R. T., Willsey, A. J., State, M. W., & Harland, R. M. (2020). The neurodevelopmental disorder risk gene DYRK1A is required for ciliogenesis and control of brain size in Xenopus embryos. *Development*, *147*(21), dev189290. https://doi.org/10.1242/dev.189290

78. Yamagaki, T., & Yamazaki, T. (2019). Troubleshooting Carry-Over in the LC-MS Analysis of Biomolecules: The Case of Neuropeptide Y. *Mass Spectrometry*, *8*(2), S0083. https://doi.org/10.5702/massspectrometry.S0083

79. Yu, G., Wang, L.-G., Han, Y., & He, Q.-Y. (2012). clusterProfiler: An R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A Journal of Integrative Biology*, *16*(5), 284–287. https://doi.org/10.1089/omi.2011.0118

80. Zhou, X., Feliciano, P., Shu, C., Wang, T., Astrovskaya, I., Hall, J. B., Obiajulu, J. U., Wright, J. R., Murali, S. C., Xu, S. X., Brueggeman, L., Thomas, T. R., Marchenko, O., Fleisch, C., Barns, S. D., Snyder, L. G., Han, B., Chang, T. S., Turner, T. N., … Chung, W. K. (2022). Integrating de novo and inherited variants in 42,607 autism cases identifies mutations in new moderate-risk genes. *Nature Genetics*, *54*(9), Article 9. https://doi.org/10.1038/s41588-022-01148-2

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
Eun Ear Chi Naing
CDBE13BA7672435...
Author Signature

12/12/2022
Date