Mapping the Protein Interactome of ASD-associated Genes by using BioID2

BY

Anran (Annie) Cheng, B.H.Sc. (Honours)

A Thesis

Submitted to the School of Graduate Studies

In Partial Fulfillment of the Requirements

For the Degree

Master of Science

McMaster University

Descriptive Note

MASTER OF SCIENCE (2019) (Biochemistry)	McMaster University Hamilton, Ontario
TITLE:	Mapping the Protein Interactome of ASD-associated Genes by using BioID2
AUTHOR:	Anran (Annie) Cheng, B.H.Sc.(Honours)
SUPERVISOR:	Dr. Bradley Doble, Dr. Karun Singh
NUMBER OF PAGES:	*xii*, 79

Lay Abstract

Autism, or autism spectrum disorder includes a range of conditions characterized by an affected individual's challenges with social skills, repetitive behaviours, and communication (verbal and nonverbal). There are many types of autism, which are caused by genetic and environmental influences. Children are usually diagnosed between 18 months to 3 years of age. Earlier diagnosis results in earlier intervention, which improves outcomes for the child. However, diagnosis can be difficult, as there are no biological tests to diagnose the disorders. Trained doctors look at the child's behavior and development to make a diagnosis, which is prone to human error. Many genes are linked to ASD, but not much is known about how they work. My project examines the interactions between proteins expressed from genes that are associated with ASD and other proteins that are in close proximity. By looking at the function of genes identified in putative protein-protein interactions, we hope to generate functional interaction maps and pinpoint common pathways that interacting proteins may be involved in. This will allow us to have a better understanding of ASD that may lead to therapeutic strategies.

Abstract

Autism spectrum disorder (ASD) is an array of neurodevelopmental disorders that vary in the severity of symptoms presented by affected individuals. In Canada, ASD is estimated to affect 1 in 66 children[1]. Though severity can vary, the clinical diagnostic criteria define deficits in social interaction, presence of repetitive behaviors and restricted interests, as characteristic manifestations[2]. The disorder not only reduces the quality of life for those affected, but also places an immense emotional and financial burden on their families and an economic burden on society. ASD is a heterogeneous disorder, with hundreds of implicated loci and genes. However, only a small number of implicated genes have well-defined neurobiological functions. Defining the pathogenesis and etiology of ASD in genetic and molecular terms is essential for expanding our understanding of the disorder, uncovering protein biomarkers for diagnosis, and identifying targets for therapeutic development. My project used a chemico-genetic proximity labeling approach, termed BioID2, to map the protein interactome of ASD-associated gene products. I employed the genome editing CRISPR-Cas9 technique to engineer mouse embryonic stem cells with a BioID2 expression cassette linked to an ASD-associated gene of interest. A mouse model will be generated from the edited mouse embryonic stem cells. This approach will permit the tracking of protein-protein interactions on an endogenous level, which provides a more accurate representation of the

interacting partners. I also initiated a second approach in which lentiviral constructs harboring BioID2-tagged ASD-associated genes of interest will be used to overexpress resultant BioID2 fusion proteins in primary cortical neurons. This method will allow us to propose protein networks for our genes of interest and to dissect the molecular function of these genes. Studying protein-protein interactions of candidate ASD gene products is critical for understanding the molecular etiology of ASD. The elucidation of ASD-associated protein networks will be indispensable to the development of novel therapeutic agents, by informing target selection. This project will expand our current understanding of ASD, identify novel protein markers, and contribute to development of therapies. It can also serve as proof-of-concept that can be adapted for various other neurological disorders.

Acknowledgements

I would like to express my sincere gratitude to my supervisors, Dr. Karun Singh, and Dr. Bradley Doble. Their doors are always open for me, and always welcome any concerns or problems that I present them with. I truly enjoyed being co-supervised by two of the best PIs and mentors. They've shown me what true collaboration looks like and how to communicate different ideas in a constructive manner to produce a synergistic outcome. My experience in both labs allowed me to dabble in different scientific fields, and a wide variety of people, accelerating my academic and personal growth. I will carry this forward in my future endeavors.

I would also like to thank my supervisory committee member, Dr. Yu Lu. Without his help, this project would not have been possible. His knowledge, patience, and support with troubleshooting the mass spectrometry is indispensable to configuring the BioID2 experiment process.

Special thanks to my fellow lab members for their emotional support and experimental contributions. I know not everyone is lucky enough to be in a lab filled with people they enjoy working with every day. They've been with me through ups and downs, and I feel grateful to have them as my mentors, colleagues, and life-long friends.

Thanks to all my friends and family for their unwavering support through all my years of study. Their support and confidence in me kept me going when things were difficult. I could not have done this without them.

Table of Contents

List of Abbreviations

	Abbreviation	Full Term
A	AP-MS	Affinity-purification mass spectrometry
	ASD	Autism spectrum disorder
B	bp	Base pairs
	BPL	Biotin protein ligase
C	CRAPome	Contaminant repository for affinity purification–mass spectrometry data
	CRISPR	Clustered regularly interspaced short palindromic repeats
D	DAPI	4’,6-diamidino-2-phenylindole
	DIV	Days *in vitro*
	DSB	Double stranded break
	DSM-5	Diagnostic and Statistical Manual of Mental Disorders 5th Edition
E	ESC	Embryonic stem cell
F	FDR	False discovery rate
G	GFP	Green fluorescent protein
H	HA	Homology arm
I	iPSC	Induced pluripotent stem cells
K	kb	Kilobases
L	LTR	Long terminal repeat
M	MAP2	Microtubule-associated protein 2
	mESC	Mouse embryonic stem cell
	MOI	Multiplicity of infection
	MS	Mass spectrometry
N	NDD	Neurodevelopmental disorder
	NeuN	Neuronal Nuclei
	NHEJ	Non-homologous end joining
P	PCR	Polymerase Chain Reaction
	PPI	Protein protein interaction
S	SAINT	Significance Analysis of INTeractome
	SFARI	Simons Foundation Autism Research Initiative
	sgRNA	Single guide RNA
	STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
T	tGFP	Turbo green fluorescent protein
	Tuj1	Beta III tubulin

W	WT	Wildtype

Declaration of Academic Achievement

This thesis was completed primarily by Anran (Annie) Cheng, with the following contributions from other individuals:

i. Brianna Unda from Dr. Karun Singh's lab for making the pLV-hSYN-tGFP-BioID2 control plasmid.
ii. Nadeem Murtaza from Dr. Karun Singh's lab for making the pLV-hSYN-tGFP-Luciferase plasmid
iii. Dr. Lu for mass spectrometry for the BioID2 experiments

CHAPTER 1: INTRODUCTION

1. Autism Spectrum Disorder

Autism spectrum disorder (ASD) encompasses an array of neurodevelopmental disorders (NDD) that vary in the severity of symptoms displayed in affected individuals. The clinical diagnostic criteria for ASD as described in Diagnostic and Statistical Manual of Mental Disorders 5th Edition (DSM-5) defines impairments in social interaction and communication, and presence of repetitive and restricted behaviors as characteristic manifestations[2,3]. Secondary symptoms commonly observed with the core phenotype include hyperactivity, hypoactivity, impulsivity, aggressiveness, and/or anxiety[4]. ASD is often comorbid with other conditions such as attention-deficit hyperactivity disorder, bipolar disorder, or fragile X syndrome, which makes diagnosis more difficult[5]. While some ASD individuals can live independently with minimal support, others require life-long care and support[6,7]. Studies have estimated the prevalence of ASD to have plateaued at 1 in 66 children in Canada. Globally, ASD is thought to affect approximately 1% of the world's population[1,8].

There has been an increase in the observed prevalence of ASD that may be reflective of enhanced clinical tests and diagnostic tools, as well as changes to diagnostic criteria (DSM-V)[5,9,10]. A reliable diagnosis can be made by age 2, but many patients receive a full diagnosis at an older age[11]. This results in delayed treatment, which is detrimental because earlier

intervention has been shown to improve an affected child's development more significantly than later intervention[12]. ASD not only reduces the quality of life for those affected, but also places an immense emotional and financial burden on families and an economic burden on society[13].

The consensus view is that ASD is a result of complex gene-environment interactions with strong genetic influences. Advances in genetic technologies, large cohort studies, and widespread database sharing have contributed to the discovery and validation of causative genes in ASD[14]. The recent identification of genetic risk factors for ASD has revealed many genes that are likely to be involved in these debilitating disorders. Even with the new genetic information, most of the identified genes do not cluster into any well-defined cellular pathways and may fall into incomplete networks that are unique to the human brain. In addition, due to the heterogeneity of this neurodevelopmental disorder, with great variations observed in behavioral manifestations and cognitive profiles, determination of a single most important genetic risk factor is extremely difficult[12].

There are currently no medications that can cure or treat the core symptoms of ASD. Medications that target secondary symptoms are available, however, with varied efficacy accompanied by negative side effects[6]. The limited understanding of the molecular basis of ASD is reflected in the lack of therapeutics and biochemical diagnostic criteria.

Therefore, it is crucial to study the pathogenesis and etiology of ASD in genetic and molecular terms[12].

To elucidate the genetic and molecular foundations of ASD, my project will focus on mapping the protein interactome of ASD-associated gene products through a BioID2 screen. This will expand our understanding of the disorder, possibly identify novel biomarkers for diagnosis, reveal new targets for development of novel therapeutics, and expand current ASD databases from a proteomic perspective.

1.1 Genetics of Autism Spectrum Disorder

The exact cause of ASD is currently unknown, but studies have highlighted both environmental and genetic factors that contribute to the disorder's etiology. While environmental factors have not been clearly defined, numerous studies have indicated a strong genetic component in the origin of ASD[14–16]. A review of twin and family studies strongly supports cases of genetic predisposition with heritability of up to 90%, and an estimated recurrence risk among siblings of autistic children to range from 3% to 18%[16,17]. However, high heritability does not necessarily imply a specific genetic model, as ASD is a highly heterogeneous disorder, and many cases of ASD occur with no evidence of family history[18]. Genetic studies in the past decade have identified hundreds of inherited genetic variations and *de novo* mutations related to autism or autistic features. Yet,

few of the implicated genes have well-defined neurobiological functions, which further complicates our understanding of ASD pathobiology[19,20].

1.1.1 Identification of genetic variants

Genetic variants associated with autism spectrum disorder can be single-nucleotide variants (SNVs) or structural variants (SVs)[15]. Technologies such as next-generation sequencing have allowed whole human genome and exome sequencing, which allows the identification of variants. SNVs are mutations in a single nucleotide in a specific location of the genome. They can occur in both coding and noncoding regions, with some variants leading to the development of diseases[21]. Natural variations occur within the genome, and SNVs were initially identified through genome-wide association studies[22]. Structural variants include insertions or deletions (indels), commonly referred to as copy number variants (CNVs), that are long stretches of DNA ranging from 1 kilobase to a few megabases[19]. As a result of their size, CNVs often include multiple genes. Variants can be classified as a common inherited, rare inherited, or *de novo* variations[22].

1.1.2 Common Inherited, Rare Inherited, and De Novo Variations

Genome wide-association studies (GWAS) have examined the association between common variants and ASD diagnosis by using family-based

cohorts[22]. Common variants refer to genetic variation from the reference genome, which is present in >1% of the population[12]. Common variants typically have small effects but are predicted to act synergistically in the development of ASD. Recent studies estimate that 15-40% of the genetic risk of ASD is tagged by common variants. Though common variants are estimated to be a large driving factor for the disorder, the effect size of individual common variants is small. This observation has led to the search for rare variants, which may exhibit larger individual effect sizes.

Rare variants are found in less than 1% of the population but tend to have stronger effects. *De novo* variants originate from mutations in the parental germ line or early somatic cells of the developing individual[22]. One can deduce that *de novo* genetic variants may play an important role due to the large mutational target sized, reduced fertility and high prevalence of correlated ASD[23]. An individual's genome will contain a few *de novo* copy number variants (CNVs) and single nucleotide variants (SNVs). Thus, if a recurrent CNV or SNV is observed within a particular gene, it is likely involved in ASD susceptibility because of the very small likelihood of this event occurring randomly. The search for *de novo* causal variants is simpler than the search for inherited variants because of their relative rarity[12,22].

1.1.3 Syndromic and Non-Syndromic Autism

ASD is categorized as syndromic or non-syndromic. Syndromic autism harbors a set of phenotypes that can be fully attributed to a mutation in a particular gene or set of genes. Non-syndromic autism is not linked to other neurological diseases but is also heritable[24].

2. Identifying Protein-Protein Interactions of ASD Associated Proteins through BioID2

2.1 Proteomics of ASD:

Proteomics is the study of protein expression in cells and tissues and is a powerful way to provide insight into the disease conditions. It is advantageous to study proteins, rather than only mRNA transcripts, as they represent the primary functioning units in cells that execute information held within the genetic code. Proteomic technology has the ability to identify and quantify proteins, post-translational modifications, signaling pathways, and protein interaction networks of a tissue, cell types or subcellular regions across various genetic and metabolic states. Unlike other approaches, such as genomic studies, there have been few proteomic studies done in the context of ASD[25]. To date, all major papers have used APEX or BioID to study 2-3 genes for protein identification, with only two major labs having studied neuronal genes in neuronal models.

Our lab wants to study ASD-risk genes and their protein functions to identify protein networks that have known and unknown synaptic functions. To further this research, we want to study the impact of disease-associated mutations on protein interactions. By integrating proteomic data with a growing repository of bioinformatic tools, it is possible to construct dynamic functional schema for observed patterns of expression. This discovery may elucidate useful drug targets or diagnostic markers.

2.2 BioID2*:*

Identification of protein-protein interactions (PPIs) is essential for proteomics and for uncovering biological mechanisms. Proximity-dependent biotin identification, named BioID, is a method used to identify protein-protein interactions and proximal proteins for a gene of interest within a living cell. BioID utilizes a promiscuous bacterial biotin protein ligase (BPL), which is fused to the protein product of a gene of interest[26]. The BPL fused to the protein of interest will then biotinylate interacting and proximal proteins, that can be isolated by using biotin-streptavidin affinity capture[27]. The isolated samples can then be subjected to proteomic analysis to identify the interacting or proximal proteins *(Figure 1)*.

The original BioID method is based on a 35-kDa DNA-binding BPL isolated from *Escherichia coli* (*E.coli)* known as *BirA*[28]. Biotin is a vitamin that is present in small amounts in all living cells and is critical for a number

of biological processes including cell growth and the citric acid cycle[29]. BPL is a ubiquitous enzyme that regulates the process of biotinylation, a rare post-translational modification that attaches biotin to a protein, nucleic acid or other molecules[30]. Biotinylation is catalyzed through a two-step reaction where biotin is activated by adenosine triphosphate (ATP) to form biotinyl-5'-AMP. BPL can then catalyze the covalent attachment (amide linkage) of the carboxyl group of biotin to the epsilon-amino group of a lysine residue in its substrate protein, carboxylase, to complete biotinylation of the protein[30].

Biotin protein ligase and biotin domains within different organisms have been found to be relatively conserved and functionally homologous which allows interspecies biotinylation to occur[27,29,30]. The limiting factor of wild type *BirA* is its stringent selectivity for its endogenous substrate. To overcome the selectivity of *BirA*, the Roux Lab, from Sanford Research, has employed a *BirA* mutant (R118G, *BirA**) that prematurely releases highly reactive, yet labile bioAMP. The mutant displays a two-order magnitude decreased affinity for BioAMP, and *BirA** expression results in promiscuous protein biotinylation, as free bioAMP will readily react with primary amines.

More significantly, it has been demonstrated *in vitro* that *BirA** will promiscuously biotinylate proteins in a proximity-dependent fashion[26]. Biotinylated proteins can be isolated by affinity capture and identified by mass spectrometry. The protein avidin and related family members and

derivatives have strong affinity for biotin. The avidin-biotin associated is one of the strongest non-covalent interactions between protein and ligand. This non-covalent bond is unaffected by pH, temperature, organic solvents, and other denaturing agents[31].

The Roux lab has improved the BioID system and has named the new technique BioID2. Unlike the *E. coli BirA* in BioID, BioID2 uses the R40G mutant variant of a biotin ligase from *Aquifex aeolicus.* The BPL is smaller, may be used under thermophilic conditions, and requires less biotin. In BioID2, the biotinylation range was also increased using a 13xLinker consisting of 13 tandem repeats of GGGGS sequence, which increases the proximate biotinylation range from ~10nm to ~35nm[32].

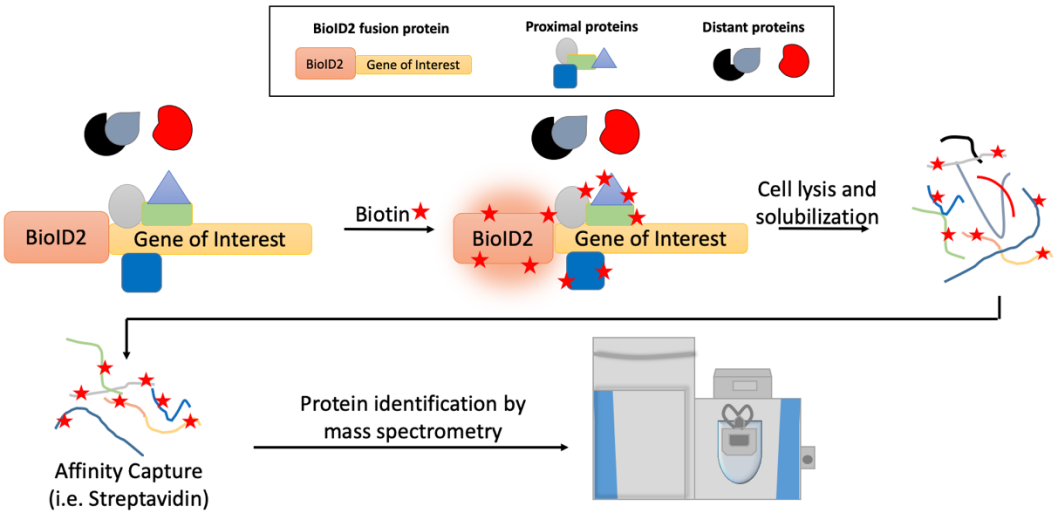

Figure 1: Overview of the BioID2 system: Proximity-dependent biotin identification, named BioID2, is a method used to identify protein-protein interactions and proximal proteins for the gene of interest in a cellular

environment. BioID2 utilizes a promiscuous bacterial biotin ligase (BPL), which is fused to a protein product of a gene of interest. The BPL fused to the protein of interest will biotinylate interacting and proximal proteins, that can be isolated by using biotin-streptavidin affinity capture. The isolated samples can then be subjected to proteomic analysis to identify the interacting or proximal proteins.

3. ASD Genes Selection Criteria

3.1 Endogenous Approach:

My project first focused on the postsynaptic density protein 95 (PSD95), SH3 and multiple ankyrin repeat domains 3 (SHANK3), and OTU Deubiquitinase 7A (OTUD7A). PSD95 and SHANK3 are widely accepted as genes associated with ASD. PSD95 protein has a vital role in regulating the density and activity of glutamate receptors. Altered levels of PSD95 protein changes the excitatory to inhibitory synapses ratio in hippocampal neurons (SFARI). Research has revealed that SHANK3 proteins have a role in synapse formation and dendritic spine maturation, as they are multidomain scaffold proteins of the postsynaptic density connecting neurotransmitter receptors, ion channels, and other membrane proteins to the actin cytoskeleton and G-protein-coupled signaling pathways. These two genes will serve as controls for the BioID technique, via confirmation of known interactions, and will also provide an opportunity to discover novel interactors. In preliminary studies done by Karun Singh's lab, characterizing a 15q13.3 microdeletion mouse model, discovered significant defects in the forebrain development signaling pathways and abnormal synaptic

morphology. The OTUD7A gene is being studied as the driver gene for these phenotypes.

3.1.1 DLG4/PSD95 and Autism Spectrum Disorder:

The *DLG4* gene encodes for postsynaptic density-95 (PSD95) protein, also known as synapse-associated protein (SAP90). The gene is located on chromosome 17p13.1 (human), chromosome 11 (mouse), and the encoded protein is predominantly expressed in the brain where it localizes in the postsynaptic density of neuronal cells[33]. PSD95 contains 3 PDZ domains, a SH3 domain, and a guanylate kinase (GUK) domain, which is characteristic of the membrane-associated guanylate kinase (MAGUK) family[34]. Three major isoforms for PSD95 exists, PSD95α, PSD95β, and truncated PSD95, with PSD95α being the dominant isoform[33]. However, all isoforms of PSD95 interact with the cytoplasmic tail of NMDA receptors, shaker-type potassium channels, and nitric oxide synthase through the PDZ domains, to induce the clustering of these molecules at the PSD. This gene is important for the concentration of receptors and channels at the PSD, as overexpression or depletion leads to an imbalance in the ratio of excitatory to inhibitory synapses.[35] Studies have shown PSD95 levels are diminished in aging and ASD patients and individuals with other neurodevelopmental disorders[36]. *DLG4* is a defined *de novo* gene for ASD[37].

3.1.2 SHANK3 and Autism Spectrum Disorder:

SH3 and multiple ankyrin repeat domains 3 (*SHANK3*) protein, is also known as proline-rich synapse-associated protein 2 (ProSAP2). SHANK3 is encoded by the *SHANK3* gene located on chromosome 22q13.33 (human), chromosome 15 (mouse), and localizes to the postsynaptic density, presynaptic density and axons[38]. It is a member of the SHANK (or ProSAP) family of proteins, which includes: SHANK1, SHANK2, and SHANK3. SHANK family members share five main domain regions: N-terminal ankyrin repeats, a SH3 domain, a PDZ domain, a proline-rich region, and a C-terminal SAM domain[39]. Through these functional domains, SHANK interacts with many postsynaptic density proteins[40]. Both *de novo* and inherited SNVs and CNVs involving SHANK3 have been identified in individuals with ASD in multiple reports[41,42].

3.1.3 OTUD7A and Autism Spectrum Disorder:

A microdeletion approximately 1.5 Mb from break points 4 to 5 on chromosome 15q13.3 (human), chromosome 7 (mouse), has been linked to an increased risk of intellectual disability, epilepsy, schizophrenia and ASD. This deletion encompasses 7 genes and results in high penetrance with a heterogeneous phenotype and variable expression in patients[43]. One of the genes in this region, ovarian tumor (OTU) domain containing protein 7A (*OTUD7A*), is the suspected driver gene for the neurodevelopmental

disorder symptoms[44]. It is a deubiquitinating enzyme that belongs to a family of 15 other OTU deubiquitinases.

Deubiquitinating enzymes are proteases that remove ubiquitin chains from proteins. They are responsible for the activation of ubiquitin precursor proteins, recycling of ubiquitin trapped in the ubiquitination of proteins, and reverse ubiquitination of target proteins[45]. Studies have shown that *OTUD7A* is the only gene in the 15q13.3 region that regulates dendritic and spine morphology and is critical for protein-signaling networks that localize to the post synaptic density (dendritic spines)[44]. Although *OTUD7A* appears to be an important gene in NDD, current knowledge of the role of *OTUD7A* is limited, and protein-protein interactions in neurons need to be identified.

3.2 Overexpression Approach

We have selected a total of 63 ASD-associated genes to perform overexpression BioID2 experiments (*Table 1*). The list of genes to study was curated from the Simons Foundation Autism Research Initiative (SFARI) Gene Database as well as current literature.

Gene Symbol	Size (aa)
NCKAP1	1128
GFAP	1296
ERBIN	1412
PTCHD1	888
ADNP	1102
CTNND2	1225
CDKL5	1030
PHIP	1821
GNAI1	1062
GIGYF1	3132
DYRK1A	763
PPP2R5D	602
TLK2	772
NTRK2	822
MET	1390
PTEN	403
DHCR7	475
BCKDK	412
LEO1	666
AFF2/FMR2	1311
CNOT3	753

Gene Symbol	Size (aa)
DDX3X	662
RHEB	184
SPAST	616
TM9SF4	1926
LNPK	428
CUX1	1505
DHX30	1194
AP2S1	426
SRPRA	1914
CUL3	768
TBL1XR1	514
UBE3A	973
USP15	981
USP7	1102
MECP2	486
SLC6A1	599
CACNA2D3	1091
GRIP1	1128
GABRB3	473
GRIA1	906
KCNQ2	872
GRIN2B	1484
KCNQ3	2616
CACNA1E	6939

Gene Symbol	Size (aa)
NLGN3	848
CNTN4	1026
PCDH19	1148
PCDH11X	1347
SHANK2	1470
NRXN1	1477
CNTNAP2	1331
SYNGAP1	1343
PSD95	724
GPHN	736
NLGN4	816
SHANK3	1731
NRXN1	4431
SHANK2	4410
STXBP1	1782
DSCAM	6036
DPYSL2	1716
PRR12	3645
KIAA0232	4188

LEGEND
Cytoskeleton
Metabolism/Enzymatic
Intracellular Signalling
Kinase/Phosphatase Activity
mRNA binding
Nucleotide Hydrolase
Organelle Function
Protein Degradation
Receptor/Transporter/Ion Channel
Synaptic Protein
Axon Guidance
Unknown

Table 1: Table and Legend of the ASD genes chosen for the BioID2 experiments. The list of genes to study was curated from the Simons Foundation Autism Research Initiative (SFARI) Gene Database as well as current literature.

The SFARI database is a research program establish in 2005 by the Simons Foundation that focuses on all aspects of autism research. SFARI Gene is a well-recognized ASD database that provides a publicly available web portal for ongoing collection, curations and visualization of genes linked to autism disorders. A crucial part of this curation is to establish criteria for assessing the strength of evidence linking candidate genes to ASD through the assessment of relative sample size, statistical significance, replication, functional evidence, and other kinds of data that exist for each gene[46].

The ASD-associated genes that we picked belong to category 1-2 which are high confidence/strong candidate ASD genes, after considering rigorous genome-wide statistical comparisons between cases and controls, with independent replication. Any genes that have evidence implicating them in idiopathic autism will also contain the letter S for syndromic disorder with the number.

In addition, we selected genes from the following papers: Yuen et al. 2017. Nature Neuroscience paper, which contains 18 New ASD Genes[47]; Hoang, Buchanan, and Scherer. 2018. NPJ Genomic Medicine paper that contains a curated list of the 16 Top ASD genes[48]; and, Satterstrom et al. 2019. BioRxivs paper that lists a subset of neural communication genes[49]. Other considerations when choosing the genes included the size of the cDNA, as the packaging capacity of lentiviruses is limited, and the ability to readily obtain cDNAs for candidate genes (*Figure 2*).

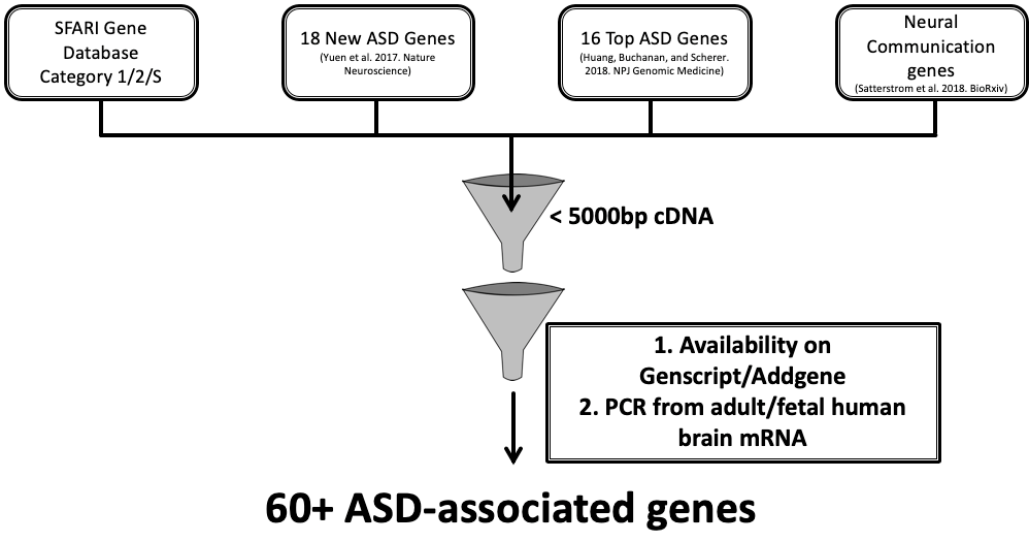

Figure 2: Schematic of ASD-associated gene selection criteria.

4. Project Rationale

The long-term goal of this research project is to identify unifying characteristics among risk genes/proteins. There have been numerous important advances in understanding aspects of Autism Spectrum Disorder (ASD), including insights into its pathogenetic processes and diverse possible etiologies in recent years. Many of the promising recent findings have focused on the genomic aspects of ASD and have identified hundreds of ASD-associated genes. However, the etiology is unclear, as there is little known about these genes and how they contribute to ASD. To study these genes, we will study their protein interaction networks through BioID2 tagging followed by affinity purification mass spectrometry (AP-MS).

The experiments and future work outlined in this thesis will provide insights into previously known protein networks and will elucidate uncharacterized novel networks in the brain through proteomics. A better understanding of protein interactions in the cells of normal and ASD brain tissue will help to unveil new therapeutic targets for the treatment of cure for ASD.

5. Hypotheses

1. **Tagging of BioID2 to *DLG4*, *OTUD7A,* and *SHANK3* will reveal novel protein-protein interactions in the post synaptic density**
2. **Overexpression of BioID2 fusion proteins in primary cortical neurons will recapitulate known protein interactions and reveal novel protein-protein interactions**

CHAPTER 2: MATERIALS AND METHODS

1. Cell Culture

All cell lines were maintained at 37°C, 5% CO_2 in a humidified incubator.

E14 mouse ES cells and JM8A3.N1 agouti mouse ES cell lines were maintained on 0.1% gelatin-coated tissue culture plates with mESC media (82% High glucose DMEM, 15% Fetal Bovine Serum, 1mM sodium pyruvate, 2mM Glutamax, 100uM Non-essential amino acids, 0.1 mM 2-mercaptoethanol stock; all from Thermo Fisher Scientific) supplemented with 1000 U/mL mLIF (Miltenyi Biotec). For routine maintenance of adherent cells, monolayers at 90% confluency were gently rinsed with 10 mL Dulbecco's phosphate buffered saline (DPBS) and detached by incubation with ACCUTASE™ at 37°C, 5% CO_2 for 5 minutes. Cell were split by 1:6 to 1:8 dilution factors every other day.

HEK293FT cells were obtained from Dr. Kristen Hope (McMaster University – SCCRI). Cells were maintained on tissue culture plates with HEK media (87% DMEM, 10% Fetal Bovine Serum, 2mM Glutamax, 100uM Non-essential amino acid, 1mM sodium pyruvate; all from Thermo Fisher Scientific). For routine maintenance of adherent cells, monolayers at 80-90% confluency were rinsed gently with 5 mL Dulbecco's phosphate buffered saline (DPBS) and detached by incubation with 3mL trypsin (0.25%) at 37°C,

5% CO_2 for 5 minutes. Cell were split by 1:6 to 1:8 dilution factors every other day.

Lenti-X 293T cells were obtained from Takara. Low passage cells were maintained on tissue culture plates with Lenti-X media (90% High glucose DMEM, 10% Fetal Bovine Serum, 4mM Glutamax, 1mM sodium pyruvate; all from Thermo Fisher Scientific). For routine maintenance of adherent cells, monolayers at 80-90% confluence were rinsed gently with 5 mL Dulbecco's phosphate buffered saline (DPBS) and detached by incubation with 3mL trypsin (0.25%) at 37°C, 5% CO_2 for 5 minutes. Cell were split by 1:6 to 1:8 dilution factors every other day.

iPS cells were maintained on Matrigel coated tissue culture plates with mTeSR™1.

2. sgRNA design

The single guide RNA (sgRNA) sequences were identified through Benchling using the Doench, Fusi et al algorithm. The gRNAs were designed to target the C-terminus of *Mus musculus* genes. Oligos were ordered from Integrated DNA Technologies.

3. Plasmids

The sgRNA expression vectors were constructed by insertion of annealed oligonucleotides into *BbsI*-digested pSpCas9(BB)-2A-Puro (PX459 V2.0) plasmid (Addgene plasmid #62988).

Donor DNA plasmid vectors were constructed using In-Fusion ® HD Cloning. Left and right homology arms were amplified by PCR using genomic DNA extract from E14 mESCs as template.

BioID2 plasmid was ordered from Genscript. It contains a 13x linker sequence, a mouse codon optimized BioID2 sequence, and a 3xFLAG sequence.

Transfer plasmid backbone was constructed by replacing the RFP sequence from pLV-hSYN-RFP with GFP-2A from pCW57-GFP-2A-MCS (Addgene #71783) and BioID2 through InFusion. *NheI* cut sites were designed to put between the 2A sequence and BioID2 to allow for cDNA insertion. The final construct is pLV-hSYN-GFP-2A-cDNA-BioID2. All other transfer plasmids were constructed by PCR amplification of cDNA and cloned into the backbone through InFusion.

4. Transient Transfection

For transfection of JM8A3.N1 agouti mouse line, cells were taken at a confluence of 70-80% in a 10cm tissue culture dish. 1 ug of donor DNA plasmid vector and 2 ug of its respective gRNA expression vector were incubated with Opti-MEM to a total of 100uL, and then added dropwise to a mixture of 2uL Lipofectamine 2000 with 98 uL of Opti-MEM to a final volume of 200 uL. Transfection mixture was incubated at room temperature for 20 minutes. During the incubation, the cells were detached using ACCUTASE ™, collected, spun down, and counted. 2 million cells were resuspended in the transfection mixture and topped up to final volume of 1mL. Transfected cells were plated at clonal density of 50k and 100k in 10cm tissue culture plates. The cells were maintained in mESC media.

For transfection of HEK293FT, cells were 70-90% confluent at time of transfection in a 6-well tissue culture plate. HEK media was changed 1 hour before transfection. 4ug of transfer plasmid was incubated with Opti-MEM to a total of 250u L, and then added dropwise to a mixture of 5uL Lipofectamine 2000 and 245 uL of Opti-MEM.

5. Generation of mESC endogenous lines

Two million JM8A3.N1 agouti mESC were transfected with 1ug of donor DNA plasmid vector and 2 ug of its respective gRNA expression vector. mESC media was changed supplemented with 250ug/mL of G418 every 48 hours for 9-10 days. Once the colonies reached a large enough size, they were isolated, expanded, double plated, and one was collected for PCR screening while the other was maintained on mESC media. Positive colonies were further expanded and the loxP-neo-loxP site was removed through transfection with Cre.

6. Antibodies

The following primary antibodies were used for Western blot and/or immunofluorescent staining: mouse anti-FLAG (1:2000 Sigma Ca. F1804), rabbit anti-β-actin (1:1000, Cell Signaling Ca. 4970S), mouse anti- β-actin (1:5000, Sigma Ca. A5441), rabbit anti-turboGFP (1:1000, Thermo Fisher Scientific Ca. PA5-22688), chicken anti-MAP2 (1:1000, Cedarlane Ca. CLN182), rabbit anti-NeuN (1:1000, NEB Ca. 24307), mouse anti-Tubulin β 3 Tuj1 (1:1000, BioLegend ca. 801202). The following secondary antibodies were used for Western blot and/or immunofluorescent staining. Chicken Alex Fluor 488 (1:500, Jackson ImmunoResearch Ca. 703-545-155), Mouse Cy3, Jackson ImmunoResearch Ca. 715-165-151), Rabbit Cy5 (1:500, Jackson ImmunoResearch Ca. 711-175-152), Alexa Fluor® 647

Streptavidin (1:500, Jackson ImmunoResearch Ca. 016-600-084), Amersham ECL Mouse IgG (1:5000, GE Healthcare Ca. NA931-1ML), Amersham ECL Rabbit IgG (1:5000, GE Healthcare Ca. NA934-1ML).

7. Protein Lysate Preparation

48 hours after transfection, cells were washed once with 1x PBS, then incubated with NP40 lysis buffer with protease inhibitor on ice. Lysed cells were collected and gently mixed on a tube rotator for 10-15 minutes in 4°C. Lysed cells were then sedimented by centrifugation for 10-15 minutes. Lysed cells were then sedimented by centrifugation for 10 minutes at 12,000 rpm at 4°C to isolate the supernatant. The supernatant is transferred to a new tube. Protein concentration was determined using the DC Protein Assay (Bio-Rad) according to the manufacturer's protocol. Protein lysates were diluted to a final concentration of 3.75 ug/uL in 4x Laemilli sample buffer with 2-mercaptoethanol, then heated at 95°C for 5 minutes prior to western blot.

8. Western Blot Analysis

For western blotting analysis, 50 ug of protein per lane were separated through 6% Bis-Tris gels (100V for 15 minutes, followed by 140V for 0.5-1.5 hrs). Proteins were transferred to polyvinylidene fluoride (PVDF) membranes using a Bio-Rad Mini TransBlot System (200mA, 2 hrs) and

transfer buffer (144 g glycine, 30.2 g TRIS, 1L H_2O). Membranes were blocked with 3% skim milk/tris-buffered saline with 1% Tween (TBS-T) solution for 45min-1hr on shaker at room temperature. Blocked membranes were incubated in primary antibody diluted in 3% skim milk/TBS-T solution at 4°C overnight. After watching 3x8minutes with 3%milk/TBS-T solution, membranes were incubated in secondary antibody for 1-1.5hours at room temperature. Membranes were then washed for 3x8 minutes with TBS-T, and incubate with Amersham ECL Western Blotting Detection Reagent (GE Healthcare) for 1 minute. Blots were visualized with the ChemiDoc MP Imaging System (Bio-Rad)

9. Immunostaining

Immunofluorescence assay was done on mouse primary neuronal cultures after infection and mESCs after neuronal differentiation Cells were rinsed with 1x PBS once and then fixed with 10% Formalin for 10 min at room temperature. After three cycles of washing with 1x PBS, cells were permeabilized and blocked in 0.1% triton and 2% donkey serum for 30 min. Cells were then incubated with primary antibodies: mouse anti-FLAG (dilution, Sigma F1804) and chicken anti-GFP (dilution, Abcam ab13970), at 4°C overnight. After three cycles of washing with 1x PBS, cells were incubated with donkey anti-chicken Alexa488 or anti-mouse Cy3 (Thermo Fisher) secondary antibodies at 1:500, for 1 hr at room temperature. After

three cycles of washing with 1x PBS, cells were mounted in mounting medium containing DAPI and imaged with a confocal microscope (Zeiss 880) on the following days.

10. Flow Cytometry and Virus Titer Determination

Titers were determined by transducing 600,000 DIV3 primary cortical neurons in one well of a 12-well plate in 0.5 mL of culturing medium at 1:100, 1:333, and 1:1000 dilutions. After 24 hours, fresh medium was added to a final volume of 2.5mL. After 48 hours the number of cells expressing GFP was determined using flow cytometry.

The tGFP-positive rate of primary cortical neurons was determined using the MACS Quant flow cytometer (Miltenyi Biotec). Twenty thousand events were counted for each sample. The analytical gates were set such that the GFP-positive rate of uninfected cell sample is ≤ 1%. The GFP-positive rate was determined for each viral dilution using the FlowJo software (FlowJo).

We used the formula $\frac{\#tGFPcellscounted}{\#livecellscounts} * dilution * \#cellsplates \div volume\ of\ media$ to obtain the titer for each dilution. The three was then combined to obtain the average titer. The volume required to obtain our desired MOI was calculated using $desired\ MOI * \#cells\ plated \div average\ titer * 1000$.

11. Primary Cell Culture

2% B-27 supplement (Life Technology, Ca. 17504-044), 1x Gluta-MAX supplement (Life Technology, Ca. 35050-061), and 1x penicillin-streptomycin (Life Technology, Ca. 15140-122), were added to Neurobasal Medium (Thermo Fisher, Ca. 21103-049) to make complete Neurobasal Medium for primary cortical neuronal culture. Primary culture of cortex neurons was prepared from embryonic day 16 CD-1 mouse embryos. The mouse was sacrificed, and the cortices were removed from the embryos and dissected and freed of meninges. The cortices from each mouse were placed into 500 mL complete Neurobasal Medium with 10% Fetal Bovine Serum (Thermo Fisher Scientific, Ca. 10439016) in an eppendoff tube, on a 37°C heat block, shaking at 1,000 rpm. After dissection was performed for all mice, tubes were taken out of the shaker to let the tissue settle. Next, supernatants were removed, and tissues were allowed to incubate in 500 mL preheated papain solution from the Papain kit (Worthington Biochemical Corp, Ca. LK003150) on the 37°C shaker for 45 min. Then, supernatants were removed and

tissues were triturated rigorously in preheated 500 mL complete Neurobasal Medium with 1 mL pipette, until no clumps of cell were visible (about 25 times of pipetting up and down). The number of cells was quantified, and 600,000 cells were plated in complete Neurobasal Medium with 10% Fetal Bovine Serum (FBS) in 12-well plates. Cultures were maintained at 37°C,

5% CO_2 and half media were changed with complete Neurobasal medium without DBS on the following day, and then was changed every 3 days until infection for titering or BioID2 experiments.

13. Generation of Lentivirus

1.0 x10^7 Lenti-X 293T cells per T150 flask were plated 24 hours before transfection. Lentivirus was produced through co-transfection of Lenti-X cells with 13 ug packaging plasmid psPAX2 (Addgene #12260), envelope plasmid 9 ug pMD2.G (Addgene #12259), and 23 ug transfer plasmid using Lipofectamine 2000. One hour before transfection, medium was aspirated, and cells were replenished with 15 mL of fresh medium. 5mL of media was removed from each flask after an hour and received a solution containing 6mL of transfection solution for a total of 16 mL. Media was replaced after 6 hours. Viral supernatants were collected 48 and 96 hours after transfection and filtered through a 0.45 µm strainer. Supernatant at each time point was pooled and concentrated through ultracentrifugation. Virus was resuspended in 1x PBS and kept at -80°C for storage.

14. Ultracentrifugation

Thermo Sci Centrifu Bench TUBE PP THINWALL 36ML ultracentrifuge tubes (Cat. 03-141) was sterilized with 70% ethanol. The viral media is pelleted for 5 mins at 100 x g, and then filtered through a Steriflip-HV Sterile

Centrifuge Tube Top Filter Unit (0.45 µm pore size) (Ca. SE1M003M00). For each round of ultracentrifugation, 32 mL of viral supernatant was centrifuged at 25,000 rpm for 2 hours at 4°C in a Thermo Scientific™ Sorvall™ WX ultracentrifuge with a SureSpin 630 Swinging Bucket Rotor. Following centrifugation, medium was carefully decanted into a bleach-filled container. Pellets were gently resuspended in 40uL 1xPBS and pipetting 50 times with a P200 pipette and aliquoted in cryotubes and kept in -80°C.

15. Affinity Capture of Biotinylate Proteins and Sample Preparation for MS

For each BioID2 overexpression condition, two biological replicates were performed and analysed via MS. pLV-hSYN-tGFP-P2A-BioID2 virus was used as control for the first batch of experiments. A full twelve well plate per condition were infected with virus at MOI 0.7. After three days, 50uM Biotin was added to each well. Eighteen to twenty-four hours after addition of biotin, cells were washed twice with PBS incubated with RIPA (1% NP40 or Triton x-100, 50 mM Tris-HCl pH=8, 150 mM NaCl, 0.1% SDS, 0.5% deoxycholic acid) and protease inhibitor cocktail tablet. The lysates were collected and benzonase nuclease is added. The cells were sonicated for (Diagenode Bioruptor® Standard Sonicator, three cycles at high intensity, 10 s per cycle).

After sonication, the cells are spun in an ultracentrifuge and the supernatant is incubated with 70uL of streptavidin sepharose beads (GE Ca.17-5113-01). To allow binding between the biotinylated proteins and beads, the mixture is incubated for 3 hours at 4 °C on an end-over-end rotator. Beads were washed 3 times in RIPA buffer at room temperature prior to incubation. After incubation, the sample is centrifuged, and the supernatant is removed. The beads are transferred to a tube with RIPA. The beads are then washed six times with 50mM ammonium bicarbonate. The beads are resuspended in 200 uL 50mM ammonium bicarbonate and 2ug of sequencing grade trypsin is added to cleave the proteins into peptides. This is incubated overnight at 37 °C on an end-over-end rotator. 1ug of trypsin is added after ~16 hours and incubated for 2hours at 37 °C. The samples are centrifuged and wash two times with 50mM ammonium bicarbonate. The samples are pelleted, and supernatant is collected. Samples are then dried in the speed vacuum for 3-3.5 hours. Sample can then be store at -80 °C.

16. Liquid chromatography/MS analysis

Liquid chromatography was conducted using a home-made trap-column (5cm x 200 mm inner diameter; Thermo POROS 10R2 C18 resin) and a home-made analytical column (50cm x 50mm inner diameter; Dr. Maisch ReproSil-Pur 120C18-AQ 5mm resin), running a 120min reversed-phase gradient at 70nl/min on a Thermo Fisher Ultimate 3000 RSLCNano UPLC

system coupled to a Thermo QExactive HF quadrupole-Orbitrap mass spectrometer. A parent ion scan was performed using a resolving power of 120,000 and then up to the 20 most intense peaks were selected for MS/MS (minimum ion count of 1000 for activation), using higher energy collision induced dissociation (HCD) fragmentation. Dynamic exclusion was activated such that MS/MS of the same m/z (within a range of 10ppm; exclusion list size=500) detected twice within 5s were excluded from analysis for 30s.

Data were analyzed using Thermo Proteome Discoverer 2.1. For protein identification, search was against the SwissProt mouse proteome database (25,020 protein entries), while the search parameters specified a parent ion mass tolerance of 10ppm, and an MS/MS fragment ion tolerance of 0.02Da, with up to two missed cleavages allowed for trypsin. Dynamic modification of +16@M was allowed. Proteins identified with an FDR rate less than 0.01 were analyzed with SAINT to filter for significant hits that are uniquely enriched with the bait proteins.

17. Protein Identification by Mass Spectrometry

Files are received in excel format. The file is uploaded into CRAPome (contaminant repository for affinity purification) to filter out contaminants in the AP-MS experiment. After the initial filter, the file is put through SAINT

(Significance Analysis of INTeractome) analysis, a computational tool that assigns confidence scores to each protein protein interaction (PPI). PPIs with scores above 0.8 (BFDR. 0.01) were inputted into the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) Database to visualize known and unknown PPIs. We selected to visualize PPIs that showed interaction through text mining, experiments, databases, and co-expression data.

Extra Training

18. Generation and Maturation of Induced Neurons from mESCs

18.1 Induced neurons from Ngn2

mESCs were plated at 500k per well in a 6-well plate coated in gelatin in mESC media. 2 ug of Ngn2 and 2 ug of rtTA plasmid were used to transfect the cells and kept in mESC media for 24 hours (Day 0). Cell are replenished with iNPC media with dox and puromycin (puro) for the next 48 hours (Day 1-2). Then the cells are switched to iNi media with BDNF, GDNF, and laminin (Day 3). Half media changes are done every other day until Day 9 where it is switched to iNi media with BDNF, GDNF, and laminin, and FBS.

18.2 Neural differentiation of mouse embryonic stem cells in serum-free monolayer culture

mESCs were plated at 500k per well in a 6-well plate coated in gelatin in mESC media. To initiate neural differentiation, media was changed to N2B27 medium (50% DMEM/F12, 50% Neurobasal, modified N2 (contains insulin, apotransferrin, progesterone, putrescine, and sodium selenite) plus BSA fraction V, B27, Glutamax, Pen/Strep). Medium is changed every other day. Under this monolayer condition and without LIF, mESCs exit the pluripotent state and predominately commit to a neural fate over a 4-5 day period.

19. Cell Reprogramming

Patient blood samples were extracted and used for blood cell reprogramming into induced pluripotent stem cells (iPSCs). CytoTune-iPS 2.0 Sendai Reprogramming Kit (Thermo Fisher Scientific) was used. The peripheral blood mononuclear cells (PBMCs) were cultured in in complete PBMC media (StemPro 34, 2mM L-glut, 20ng/mL IL-3, 20ng/mL IL-6, 100ng/mL SCF, 100ng/mL Flt3). 100k cells in 1 well of a 24 well non-adherent plate was infected according to manufacturer's instructions. Cells were transferred onto irradiated mouse embryonic fibroblasts with PBMC media without cytokines. Half media was changed every day. Once the colonies reached a confluent size, individual 'iPSC' colony was picked from

each well and added to separate wells of 12 well plates. Colonies will undergo validation to test their pluripotency (TRA-60 staining, nanog & oct4 staining). True iPSCs can be maintained on Matrigel coated plates with iPSC media. Vials can be frozen down in cryovials and stored in liquid nitrogen.

20. iPSC Cell Culture

iPSCs are thawed in 37°C until a small ice pellet remains. It is then washed with DMEM/F12 and resuspended in mTeSR media Y-27632 ROCK inhibitor (iROCK) (Sigma-Aldrich Ca. SCM075) taking care not to singularize the cells. Cells are plated onto one well of a 6-well plate with Matrigel coating. Cells are passaged when confluent (~3-4 days) using ReLeSR. Subsequent passages do not require iROCK. Cells can be frozen down in cold iPSC freezing media (10% DMSO, 50% KOSR, 40% mTeSR media) and stored in liquid nitrogen.

21. iPSC Neural Differentiation

iPSC cells are thawed in mTeSR with iROCK and penicillin-streptomycin (pen/strep) and expanded in mTeSR on Matrigel coated plates. Once cells are 70% confluent, the cells are singularized with Accutase and plated at 250k per well in a 6-well plate with mTeSR with iROCK and pen/strep with 6 wells with double Matrigel coating. After 24 hours, cells are infected with

1 vial of Ngn2 and 1 vial of rtTA virus per 3 wells of a 6 well plate in mTeSR, iROCK, and polybrene with Matrigel coating. Viral media is changed 24 hours after infection and mTeSR media is added. Once confluent, cells can be induced into neurons or frozen down. For neuron induction, iPSC cells are singularized with Accutase and filtered through a sterile 40 µm and plated at 500k cells per well in a 6 well plate with mTeSR media with iROCK (Day -1). Cells are doxycycline (dox) induced 24 hours after with iNPC media (Day 0). Cell are replenished with iNPC media with dox and puromycin (puro) for the next 48 hours (Day 1-2). Then the cells are switched to iNi media with BDNF, GDNF, and laminin (Day 3). The cells are re-plated 24 hours later (Day 4) and glia is added 24 hours later (Day 5). Half media changes are done every other day until Day 10 where it is switched to iNi media with BDNF, GDNF, and laminin, and FBS.

CHAPTER 3: RESULTS

1. Generate tools to study the protein-protein interactions of PSD95 and SHANK3 endogenously in mouse embryonic stem cells

1.1 Creating the 13xLinker-BioID2-Linker-3xFLAG-* construct

The BioID2 gene sequence was extracted from Addgene plasmid #74224 deposited by the Roux Lab[32]. We have modified the sequence to contain 13xLinker-BioID2-Linker-3xFLAG-* (stop) designed for C-terminus tagging of our genes of interest. As previously mentioned, the extensions of linkers will allow for a broader range in the biotinylation of proximal proteins. This will also allow the biotinylation of larger proteins that could not be previously be biotinylated due to the limited range of a short linker. A longer linker will also decrease the chance of the BioID2 protein interfering with the folding of our proteins of interest[32]. A 3xFLAG was added for detection and purification purposes. Fusion proteins containing 3xFLAG have enhanced detection up to 200-fold compared to other systems e.g. HA. The 3xFLAG tag is hydrophilic, contains an enterokinase cleavage site and is relatively small. Therefore, the risk of altering protein function, blocking other epitopes or decreasing solubility of the tagged protein is minimized. It can be used for immunoprecipitation, western blots and immunocytochemistry/ immunofluorescence[50]. The high sensitivity of antibodies recognizing the 3xFLAG epitope is ideal, as our genes of interest could have low-expression levels *(Figure 3)*. The modified BioID2 was designed on Benchling

(https://benchling.com) and produced by the science research company Genscript ®.

Figure 3: BioID2 plasmid: Design of BioID2 plasmid.

1.2 Designing and creating gRNA-CRISPR/Cas9 plasmids

To construct the CRISPR/Cas9 plasmids, 2-3 gRNAs for each gene were designed to target the C-terminus of *Mus musculus* (mouse) genes *Otud7a, Dlg4, and Shank3*. The mouse *Otud7a, Dlg4,* and *Shank3* gRNAs were designed to target near the stop codon of exons 14, 20, and 22, respectively. These sequences were identified through Benchling using the Doench, Fusi et al. algorithm *(Table 2).*

mESC gRNA Sequences	
OTUD7A_gRNA_1_F	CACCGAAGCAGTCCTGGACGGCTCA
OTUD7A_gRNA_1_R	AAACTGAGCCGTCCAGGACTGCTTC
OTUD7A_gRNA_2_F	CACCGAAAGCAGTCCTGGACGGCTC
OTUD7A_gRNA_2_R	AAACGAGCCGTCCAGGACTGCTTTC
PSD_gRNA_1_F	CACCGATCAGAGTCTCTCTCGGGCT
PSD_gRNA_1_R	AAACAGCCCGAGAGAGACTCTGATC
PSD_gRNA_2_F	CACCGAATCAGAGTCTCTCTCGGGC
PSD_gRNA_2_R	AAACGCCCGAGAGAGACTCTGATTC
Shank_gRNA_1_F	CACCGTGAACATCGAGCGTGCGCTC

Shank_gRNA_1_R	AAACGAGCGCACGCTCGATGTTCAC
Shank_gRNA_2_F	CACCGCGAGCGTGCGCTCAGGCAGC
Shank_gRNA_2_R	AAACGCTGCCTGAGCGCACGCTCGC
Shank_gRNA_3_F	CACCGCGTGCGCTCAGGCAGCTGGA
Shank_gRNA_3_R	AAACTCCAGCTGCCTGAGCGCACGC

Table 2: sgRNA Designs: Single guide RNAs (sgRNAs) containing complementary sequences to the strand of the target template DNA were designed via a gRNA design tool (Doench, Fusi et al. algorithm).

The gRNAs were ligated into the pSpCas9(BB)-2A-Puro (pX459) (Addgene #48139) plasmid. To compare the efficiency of the mouse gRNAs, a surveyor nuclease assay (T7 endonuclease assay) was performed on E14 mESCs. Results showed ~50% cleavage efficiency for all 7 gRNAs. To date, 7 CRISPR/Cas9 gRNAs have been generated for mESC targeting (Figure 4).

Figure 4: mESC CRISPR/Cas9 gRNA cleavage efficiency. Result of T7 cleavage assay to determine CRISPR/Cas9 gRNA efficiency through % gene modification. The % gene modification was calculated according to the formula "% gene modification = 100 x (1 – (1- fraction cleaved)^{½})". The fraction cleaved was determined by the image analyzer ImageJ.

1.3 Creating homology templates for C-terminal BioID2 integration

Targeting constructs were generated using the PGKneoF2L2DTA (PGKneo) plasmid as their backbone. For our purposes, PGKneo contains a Cre-lox system where a neomycin cassette is flanked by loxP sites. The final constructs contain 5' and 3' homology arms, the loxP neomycin cassette within intronic regions, and BioID2 directly before the stop codon of the gene. The construct designed for targeting *Otud7a* in mESCs contains **500HA-BioID2-200bp (after the end of the last exon)-loxP-neo-loxP-500HA** *(Figure 5A)***.** The constructs designed for *Dlg4* and *Shank3* contain **500HA-**

loxP-neo-loxP-200bp (before the start of the last exon)-BioID2-500HA *(Figure 5B & 5C)*.

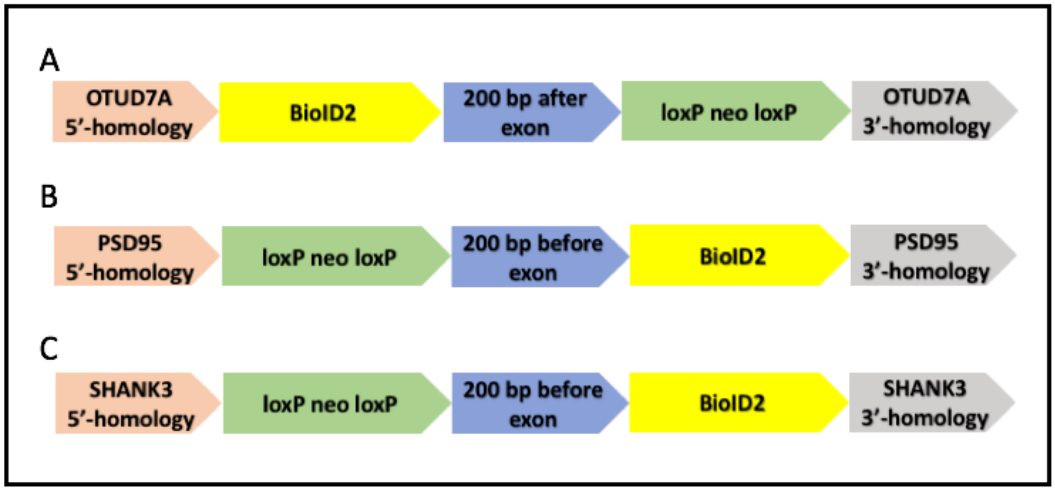

Figure 1. (A) Homology template design for OTUD7A targeting. (B) Homology template design for PSD95 targeting. (C) Homology template design for SHANK3 targeting

2. Generate a cell line using the tools designed for endogenous tagging with BioID2

The JM8A3.N1 agouti mESC cell line was used as a practice line to test the integration efficiency after co-transfection of CRISPR/Cas9 gRNAs with corresponding homology templates. After confirming the efficiency of the proposed knock-in strategy, cell lines for further BioID2 experiments will be produced in the JM8.N4 black mESC line. JM8A3.N1 mESCs were co-transfected with either *Otud7a, Dlg4,* or *Shank3* homology templates along with corresponding CRISPR/Cas9 gRNA plasmids to induce cellular repair of CRISPR-mediated DSBs via homology-directed repair (HDR). As

puromycin at the predetermined kill curve concentration was too toxic for the transfected cells, puromycin selection was forfeited and only neomycin selection at a concentration of 250 ug/mL as determined by a kill curve performed on JM8A3.A mESCs *(Figure 6)*. After selection, picking, and expansion, *Otud7a* C-terminus BioID2 tagging was not successful, but *Dlg4* and *Shank3* C-terminus taggings were both successful.

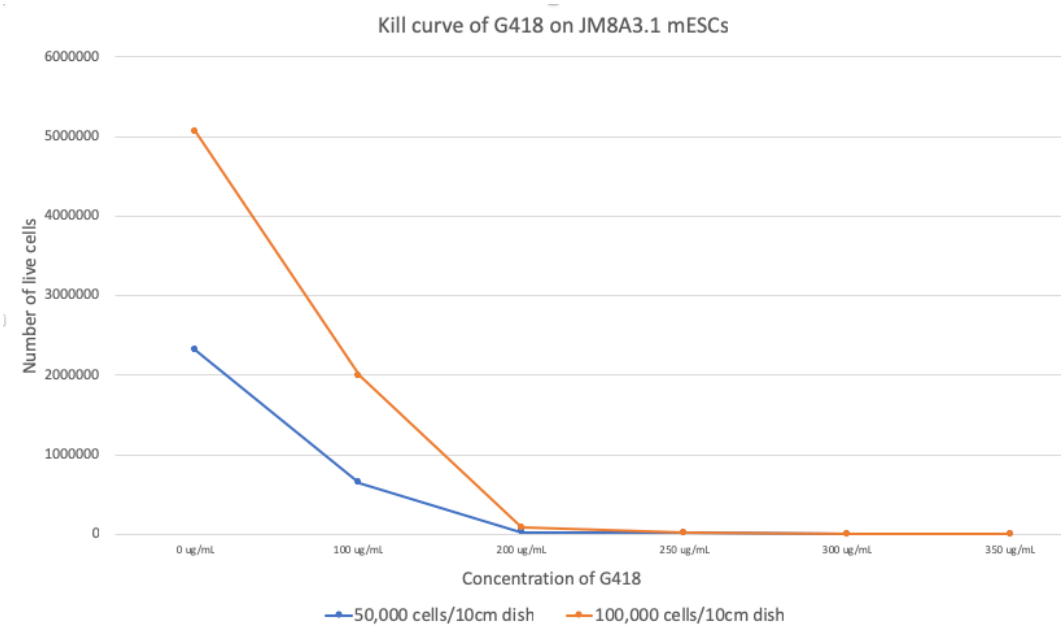

Figure 6: Kill curve of G418 on JM8A3.A mESCs. Cells were plated at a density 50k/10cm dish and 100k/10cm dish and allowed for recovery for 48 hours. G418 was added at different concentrations for 7 days. On day 7 of G418 treatment, cells were collected and counted for number of live cells.

2.1 Preliminary screening

Preliminary clone screening was performed by using PCR to detect the presence of BioID2 integration in each of the genes. A preliminary screen confirms the integration of the homology template but does not

indicate the area of integration in the genome. For *Otud7a*, the variation between the PCR product sizes of WT (~900bp) and the targeting construct (~2kb) was used to detect the presence of the *Otud7a* BioID2 targeting construct. For PSD95, the variation between the PCR product sizes of WT (~600bp) and the targeting construct (~1.5kb) was used to detect the presence of *Dlg4* BioID2 targeting construct. For *Shank3*, the primers used were able to allow preliminary screening and determination of successful 5' integration at the same time, as no WT PCR product will be produced and the PCR product for integration will be ~2.2kb.

Otud7a

I was able to pick 8 colonies putatively tagged with BioID2 at the C-terminus of *Otud7a*, all of which were screened to be negative. In my second attempt at adding BioID2 to the C-terminus of *Otud7a*, transfections with the *Otud7a* homology template DNA and the corresponding CRISPR/Cas9 gRNA plasmid used twice the amount of transfected plasmids. 28 colonies were picked, but none were properly targeted, as determined through PCR analysis *(Figure 7-A1).*

Dlg4

96 colonies were picked and a pooled PCR screen combining genomic DNA isolated from three different wells was performed to determine putative positive clones. Out of the 32 pooled PCR reactions, 15 had amplification products, and 9 showed a band at ~1.5kb *(Figure 7-B1)*.

Eight of the preliminary pools that yielded positive screening products were selected for PCR screening of gDNA from individual colonies. Of the eight colonies that produced a single band at ~1.5kb, three were selected for further screening.

Shank3

68 colonies were picked and a pooled PCR screen combining genomic DNA isolated from three different wells was performed to determine putative positive clones. Out of the 23 pooled PCR reactions, 17 had amplification products, and 16 showed a band at ~2.2kb *(Figure 7-C1)*. Three putative positive colonies were selected for further screening.

2.2 Screening for 5' integration

PCR was performed to confirm 5' construct integration in the putative positive mESC clones. The forward primer is located in the genomic region upstream of the homology template, and the reverse primer is located within the homology template. For *Dlg4*, the variation between the PCR product sizes of WT (no detectable product) and the targeting construct (~2.4kb) was used to detect the presence of the *Dlg4* BioID2 targeting construct. For *Shank3*, screening for 5' integration was confirmed during the preliminary screen.

Dlg4

Of the 3 putative positive clones, 1 showed a strong positive band at ~2.4kb *(Figure 7-B2).*

2.3 Screening for 3' integration

PCR was performed to confirm proper homology-directed integration of the 3' end of the targeting construct in putative positive mESC clones. The forward primer is located in the genomic region upstream of the homology template, and the reverse primer is located within the homology template. For *Dlg4*, the variation between the PCR product sizes of WT (no detectable product) and the targeting construct (~1.6kb) was used to detect the presence of *Dlg4* BioID2 targeting construct. For *Shank3*, the variation between the PCR product sizes of WT (~1.2kb) and the targeting construct (~2.2kb) was used to detect the presence of *Dlg4* BioID2 targeting construct.

Dlg4

From the 3 putative positive clones, 2 showed a strong positive band at ~1.6kb *(Figure 7-B3).*

Shank3

From the 3 putative positive clones, 3 showed a strong positive band at ~2.2kb *(Figure 7-C2).*

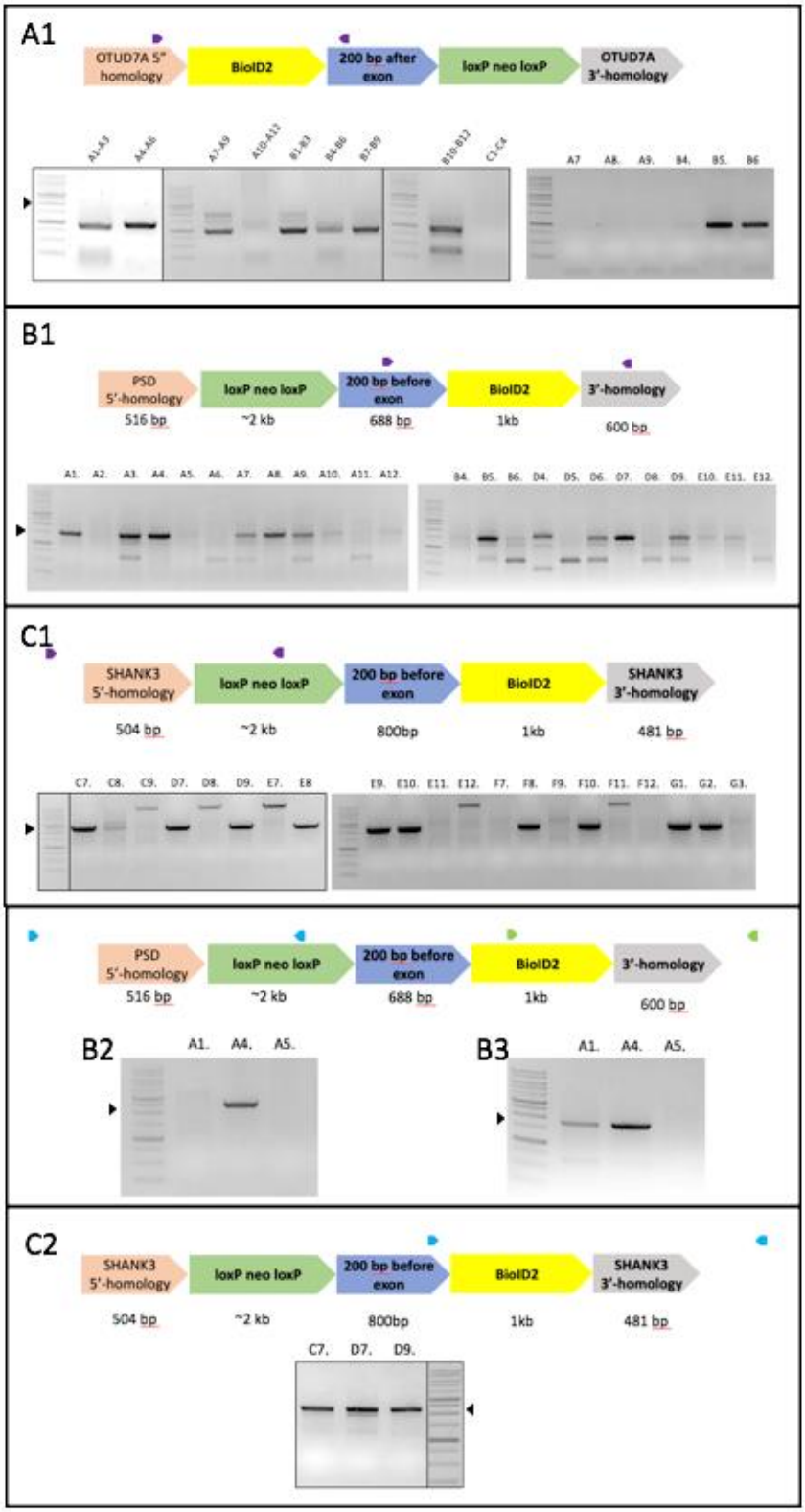
A1
OTUD7A 5" homology
BioID2
200 bp after exon
loxP neo loxP
OTUD7A 3'-homology
B1
PSD 5'-homology
516 bp
loxP neo loxP
~2 kb
200 bp before exon
688 bp
BioID2
1kb
3'-homology
600 bp
C1
SHANK3 5'-homology
504 bp
loxP neo loxP
~2 kb
200 bp before exon
800bp
BioID2
1kb
SHANK3 3'-homology
481 bp
B2
B3
C2

Figure 7: Screening Results for BioID2 tagging of OTUD7A, PSD95, and SHANK3. A1: Result of preliminary screening for OTUD7A positive clones. Positive clones will produce a band at ~2kb, whereas negative clones will produce a band at ~900bp. B1: Result of preliminary screening for PSD95 positive clones. Positive clones will produce a band at ~1.5kb, negative clones will produce a band at ~600 bp. C1: Result of preliminary 5' screening for SHANK3 positive clones. Positive clones will produce a band at ~2.2kb, whereas negative clones will not produce a detectable band. B2: Result of 5' homology screening for pSD95 positive clones. Positive clones will produce a band at ~2.4kb, whereas negative clones will not produce a detectable band. B3: Result of 3' homology screening for PSD95 positive clones. Positive clones will produce a band at ~1.6kb, whereas negative clones will not produce a detectable band. C2: Result of 3' homology screening for SHANK3 positive clones. Positive clones will produce a band at ~2.2k, whereas negative clones will produce a band at ~1.2kb.

3. Obtaining neuronal samples from mESCs for mass spectrometric analysis

As the genes being studied are ASD-related genes that are expressed in the brain, healthy neurons are required to study endogenous activity. The first approach was to generate neurons from mESCs. The K. Singh Lab uses a protocol by Zhang et al. that can convert human embryonic stem cells and iPSCs into functional induced neuronal (iN) cells with nearly 100% pure yield in less than two weeks, through expression of the transcription factor neurogenin-2 (Ngn2). By using the same transcription factor, mESCs were transfected with a plasmid that utilized a Tet-On system that contains Ngn2 *(Figure 8)*. The protocol for neural differentiation of mouse embryonic stem cells in serum-free monolayer culture by Wongpaiboonwattana et al. was tested in parallel *(Figure 9)*. Both were able to differentiate mESCs into neural progenitor cells to a certain degree.

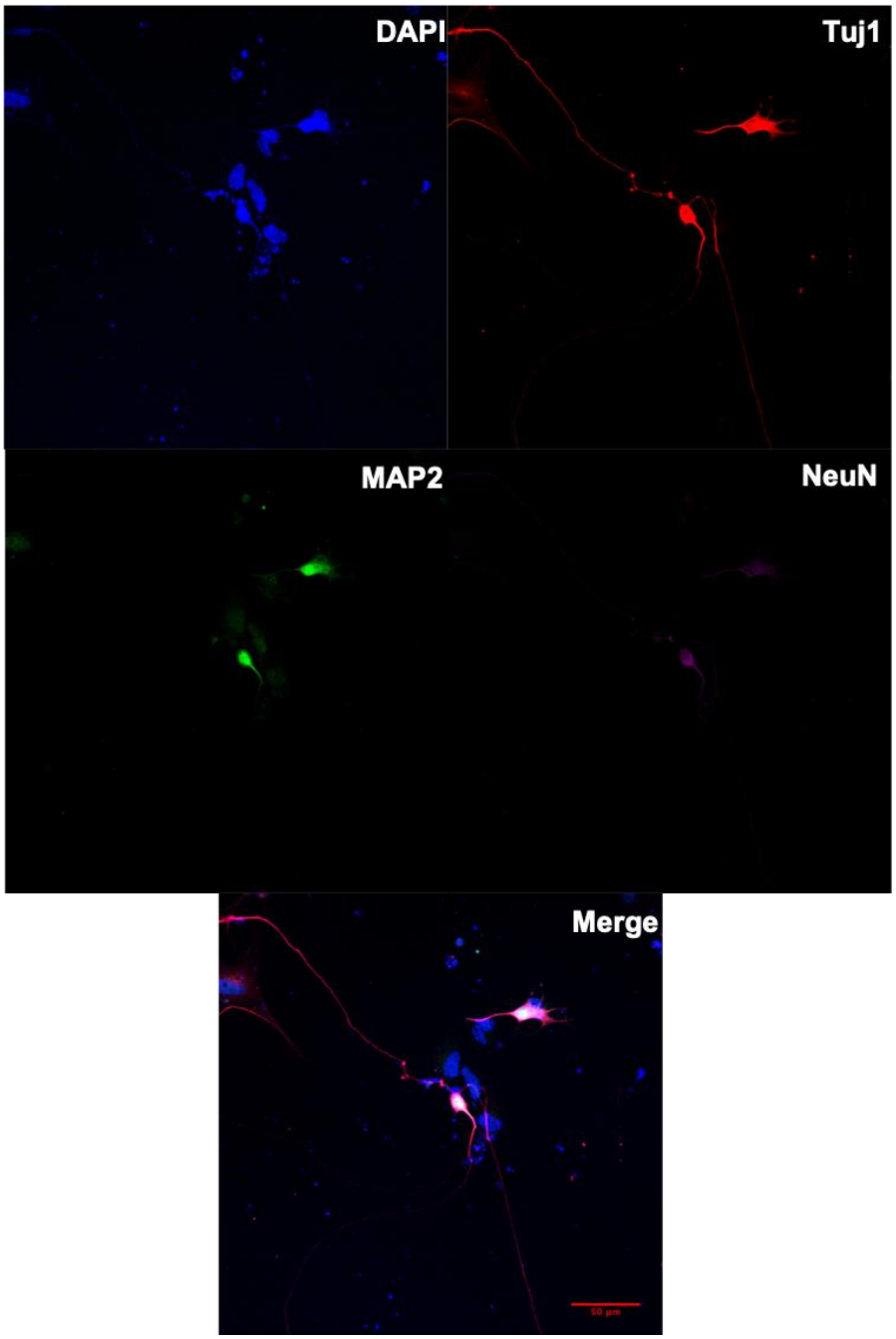

Figure 8: Immunofluorescent staining of induced neurons. mESCs were induced into neurons following a revised Ngn2 protocol. Cells were fixed on Day 9 and stained the subsequent days.

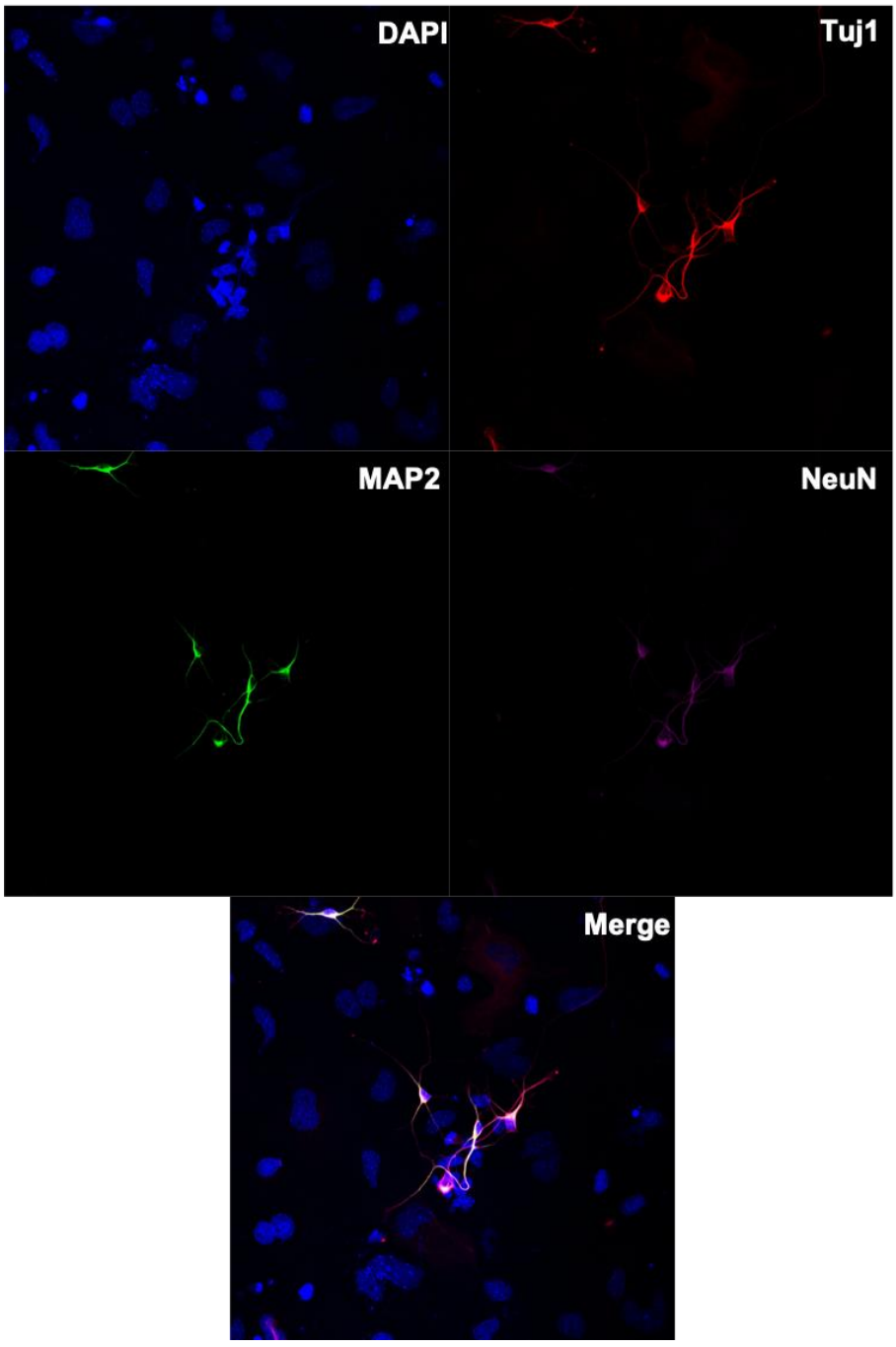

Figure 9: Immunofluorescent staining of induced neurons. mESCs were induced into neurons following the monolayer neural differentiation protocol. Cells were fixed on Day 11 and stained the subsequent days.

4. Construction of lentivirus plasmid for overexpression

I selected the pLV-hSYN-RFP viral plasmid – (Addgene #22900) as my backbone to express BioID2-fusion proteins in mouse embryonic cortical neuron cultures. This vector has been modified to contain the human *synapsin I* promoter, a turboGFP (tGFP) cassette, a 2A self-cleaving peptide sequence, NheI cut sites, and 13xL-BioID2-3xFLAG. This vector utilizes the human *synapsin I* (hSyn) promoter that has shown specificity for neuron-specific gene expression. This vector expresses green fluorescent protein, tGFP, to allow for convenient verification of its expression with live cell imaging, staining, and flow cytometry. As we will be co-expressing our BioID2-tagged gene of interest and the reporter gene, tGFP will enable visualization of transduced cells. A 2A self-cleaving peptide (2A) sequence is located between the tGFP and gene of interest-BioID2 sequence to allow expression of multiple proteins from a single mRNA transcript in approximately equimolar ratios*. NheI* cut sites were inserted between the P2A and BioID2 sequence to allow easy insertion of cDNA. The tGFP-P2A sequence was PCR amplified from pCW57-GFP-2A-MCS (Addgene #71783) and BioID2 was PCR amplified from our custom plasmid and cloned into the lentiviral backbone through In-Fusion™. The final template was termed pLV-hSYN-tGFP-P2A-cDNA-BioID2 *(Figure 10).*

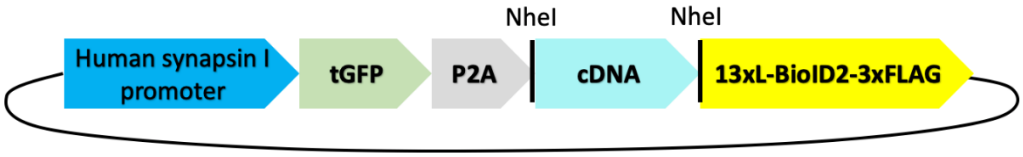

Figure 10: Lentiviral transfer plasmid design.

As lentiviral plasmids contain highly repetitive sequences (e.g. Long terminal repeats (LTR) of the lentiviral vector), each new lentiviral transfer plasmid was transformed into Stbl3 or NEBStbl bacterial cells as they keep the integrity of the plasmid over a longer period of time. Full-length Sanger sequencing was used to confirm that no mutations occurred during the cloning process of coding sequences, and LTR regions were sequenced intermittently to confirm their integrity. Subsequently, human embryonic kidney cells expressing a mutated version of the SV40 large T antigen (HEK293T) cells were transfected with each of the transfer plasmids to confirm protein expression and P2A cleavage efficiency. To date, 26 genes have been cloned, and are undergoing the aforementioned validation checkpoints *(Table 3).* Western blots have been performed on certain genes. All bands are at predicted sizes *(Figure 11).*

Cloned?	End Sequencing Good?	Full Sequencing Good?	Western Blot Verified	Gene Symbol	Size (aa)
Y	**Y**	**Y**	**Y**	**PTEN**	**403**
Y	**Y**	**Y**	**Y**	**PSD95**	**724**
Y	Y	Y	Y	GPHN	736
Y	**Y**	**Y**	**Y**	**GRIA1**	**906**
Y	Y	Y	Y	CDKL5	1030
Y	Y	Y	Y	AFF2	1311
Y	Y	Y	Y	CNTNAP2	1331
Y	Y	Y	Y	GRIN2B	1484
Y	Y	Y		MECP2	486
Y	Y	Y		NLGN4	816
Y	Y	Y		KCNQ2	872
Y	Y	Y		CACNA2D3	1091
Y	Y			DHCR7	475
Y	Y			USP15	981
Y	Y			CNTN4	1026
Y	Y			USP7	1102
Y	Y			NCKAP1	1128
Y	Y			DHX30	1194
Y	Y			MET	1390
Y				GABRB3	473
Y				SPAST	616
Y				NTRK2	822
Y				UBE3A	973
Y				SYNGAP1	1343
Y				SHANK3	1731
Y				PHIP	1821

Table 3: Current cloning progress on BioID2 experiment genes.

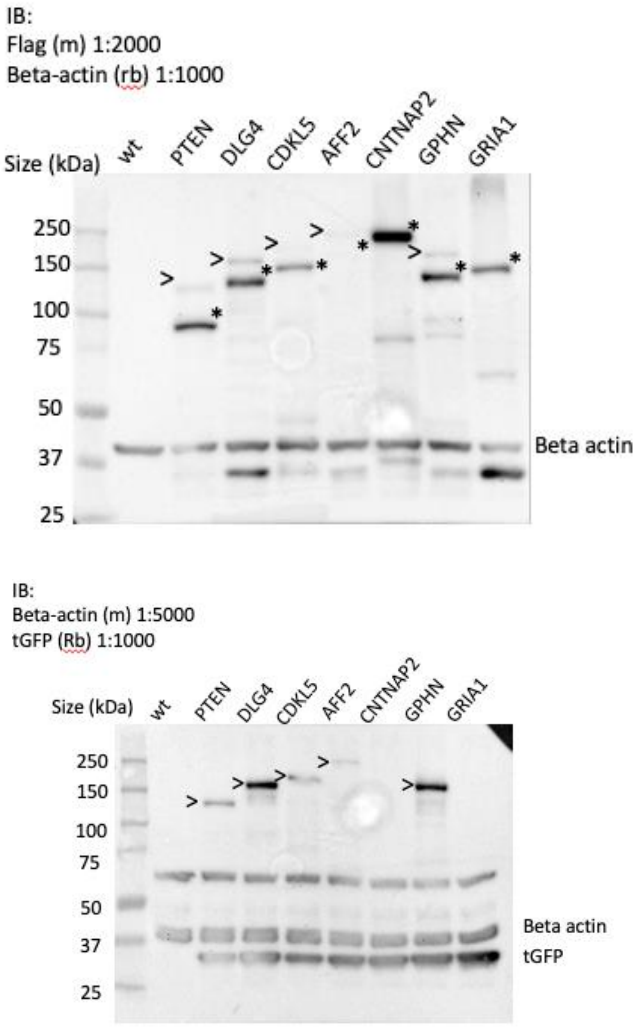

Figure 11: Western blots for to validate lentiviral plasmids for overexpression studies. (A) Blot was probed with beta-actin and tGFP antibody. (B) Blot was probed with FLAG and beta-actin antibody * corresponds to the correct protein size. > corresponds to uncleaved GFP product.

5. Virus Production

We are constantly producing virus for our experiments. The ones I have focused on are the BioID2 control plasmids and the PSD95 virus for the BioID2 experiments. We tested the virus on primary cortical neurons to visualize the localization and biotinylation ability of the transfer plasmid. PSD95 is a well-studied protein that localizes to the post synaptic density. Confocal image visualizing FLAG, which is directly conjugated to PSD95-BioID2, detected many punctate spots which is consistent with PSD95 localization to the post synaptic density. The GFP signal is less punctate, as expected, as it should be cleaved from the protein of interest and should be diffusely dispersed throughout the cells *(Figure 12)*.

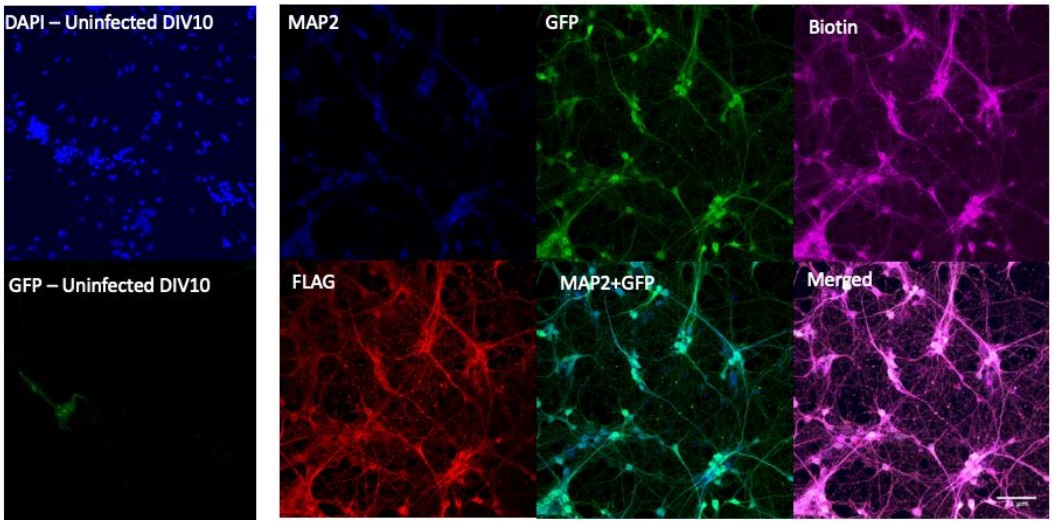

Figure 12: Transduction of pLV-hSYN-GFP-P2A-PSD95-BioID2 in mouse primary cortical neurons. Cells were stained on DIV9 after transduction with 2 uL of virus. MAP2 and GFP staining were merged to compare the efficiency of the virus.

6. Virus Production Optimization for transduction of primary neuron cultures

Generation of high titer lentiviral stocks and efficient virus concentration are central to maximizing the utility of lentiviral technology. After evaluating published protocols for lentivirus production, we have optimized the lentivirus production protocol and lentivirus titering for our purposes. To improve the virus production protocol, we tested 3 key parameters: (1) The cell type used, (2) harvest time, (3) amount of DNA used for transfection. The success of each trial was determined by titers obtained by using flow cytometry. As a starting point for the optimization of virus production, we evaluated virus titers generated from HEK293FT cells and Lenti-X 293T cells (Takara Bio USA, Inc.) under similar conditions. Previous studies claim that Lenti-X 293T cells can produce 6-fold more virus than HEK 293FT cells, and 30-fold more virus than HEK293 cells. Our titer results showed that Lenti-X cells and HEK293FT cells produced similar amount of BioID2 lentivirus, but Lenti-X cells were superior for producing larger viruses such as PSD95 virus *(Figure 13).* We next compared harvesting virus after an the initial 48-hours of production and after a second collection after an additional 48-hours of production to determine if enough virus is produced during the last two days (hours 48-96) harvest to justify a 2nd harvest. Comparing flow data for pLV-hSYN-tGFP-P2A-BioID2, pLV-hSYN-tGFP-P2A-Luciferase-P2A-BioID2, pLV-hSYN-tGFP-P2A-ANKG-BioID2 for the first harvest vs the

second harvest at the 96-hour time point, we found the titer to be reduced by half or 1/3 in the 2nd fraction collected for larger plasmids (ANKG), or plasmids with double P2A (Luciferase), and was about that same in small plasmids such as BioID2 *(Figure 14)*. Moving forward, it would be cost-effective to do a double harvest for viruses that produce high titer virus. We will not do a double harvest for viruses that produce lower titer virus as the resulting amount will not be usable. Additionally, increasing the total DNA amount in the transfection step did not increase resulting virus titer (data not shown). Thus, we will use a consistent amount of DNA for transfection regardless of transfer plasmid size.

Another factor that we consider while producing virus is the growth of the Lenti-X HEK cells. Culture cells grow in a semi-logarithmic fashion. Immediately after cells are seeded/reseeded, the cells are in a period of slow growth termed lag phase. As the cells are more adapted to the culture environment, the cells will propel to the log phase, where the cells proliferate exponentially and consume the nutrients in the growth medium. Once the medium is spent, or the cells have reach 100% confluence, mammalian cells enter the stationary phase where proliferation is greatly reduced or ceases entirely. Once cells reach 100% confluency, it takes longer to recover when reseeded. Thus, to maintain reproducible and consistent cell growth, we passage the cells in a 1/8 fraction once the plate reaches 85-90%. An issue we had when plating the cells in T150 flasks for transfection

was uneven distribution of adherent cells within the flask. We suspect that this is due to the relatively large size of the flask and incubator shelves are slightly off-leveled. Due to this combination of factors, one side would have cells at 20% confluency while the other side will be at 100% confluency. This leave only a fraction of the cells at optimal virus producing growth phase. To achieve more even plating of Lenti-X cells in T150 flasks, we have started to rotate the flasks 180 degrees after 30 minutes in the incubator so that the cells are distributed in a more uniform pattern. The cells are plated at a density of 10 million cells / flask to reach 85-90% confluency after 24 hours. If the majority of cells have not reached confluency within a 3-hour window of the 24-hour time point after plating, the cells will be discarded.

	Virus from HEK293FT Cells		**Virus from Lenti-X 293T Cells**	
Virus	*Average Titer*	*Volume for 0.7 MOI*	*Average Titer*	*Volume for 0.7 MOI*
BioID2	2.91E+08	1.44 uL	3.75E+08	1.12 uL
PSD95	3.96E+07	10.6 uL	1.96E+08	2.14 uL

Figure 13: Virus titer comparison between HEK293FT cells and Lenti-X293T cells

	Virus from 1-48 hrs		**Virus from 48-96 hrs**	
Virus	*Average Titer*	*Volume for 0.7 MOI*	*Average Titer*	*Volume for 0.7 MOI*
BioID2	3.75E+08	1.12 uL	2.70E+08	1.56 uL
Luciferase	8.07E+07	5.2 uL	2.93E+07	14.36 uL
ANKG	1.01E+08	4.17 uL	4.07E+07	10.33 uL

Figure 14: Virus titer comparison between virus harvested between 1-48 hours and 48-96 hours.

7. Developing protocol for primary cortical neuron culture titering

Traditionally in the K. Singh lab, titering has been done by using HEK293FT cells or through confocal imaging. The first titering attempt was to transduce primary cortical neurons at 4 days *in vitro* (DIV) with the pLV-hSYN-tGFP-P2A-PSD95-BioID2 and pLV-hSYN-tGFP-P2A-SHANK3-BioID2 viruses, followed by staining at 9 DIV to determine the multiplicity of infection (MOI) of each batch of virus. The MOI is the ratio between the number of virus particles to the number of cells present in a defined area and will be used to determine the amount of virus to use in optimized BioID2 experiments. This method derives the MOI by comparing the percentage of overlap between areas that are GFP-positive and MAP2-positive. After confocal imaging, it appears that pLV-hSYN-tGFP-P2A-PSD95-BioID2 was effective at 2 µL *(Figure 15),* and pLV-hSYN-tGFP-P2A-SHANK3-BioID2 was effective at 5 µL *(Figure 16)*.

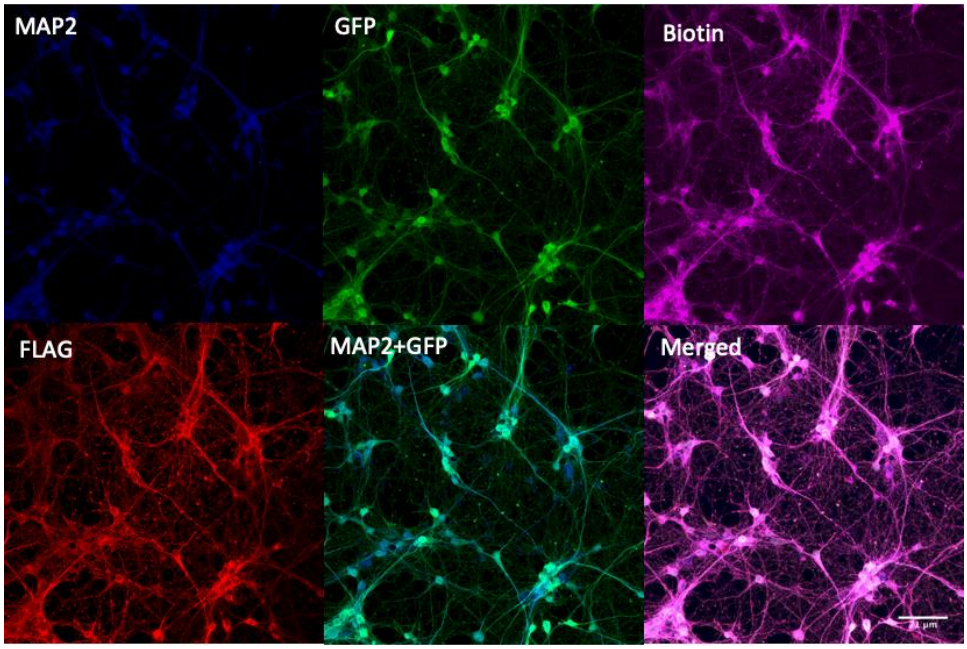

Figure 15: Cells were stained at DIV9 after transduction with 2 µL of pLV-hSYN-tGFP-P2A-PSD95-BioID2 virus*. MAP2 and GFP staining were merged to compare the efficiencies of the viral preparations.*

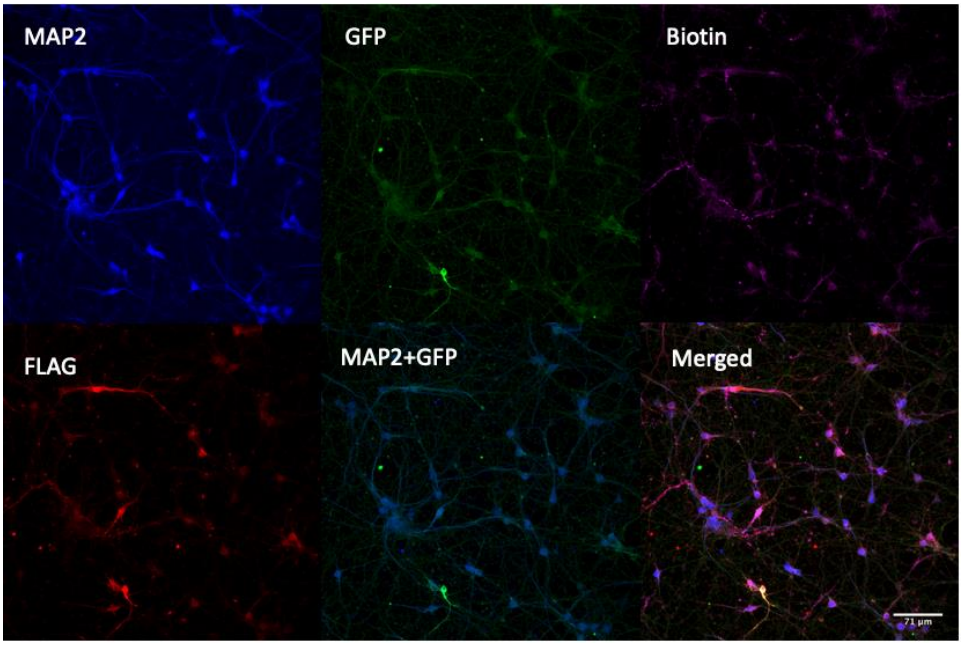

Figure 16: Cells were stained at DIV9 after transduction with 2 µL of pLV-hSYN-tGFP-P2A-SHANK3-BioID2 virus. *MAP2 and GFP staining were merged to compare the efficiency of the viral preparations.*

Upon further discussion with committee members, it was determined that the staining method to determine MOI is crude and subject to bias. Thus, we developed a model to more accurately titer the virus produced from the overexpression transfer plasmids. The new protocol employs DIV 3 primary cortical neurons that are infected with virus for 48 hours at 3 different dilutions. Flow cytometry is then performed on these cells to determine the percentage of GFP+ cells, which is subsequently used to determine the MOI.

8. Identification of direct and proximal protein-protein interactions with BioID2-fusion proteins

PSD95 is the first protein that has undergone the BioID2 analysis. The data were collected from 2 mass spectrometry runs; primary cortical neurons were infected at a MOI of 0.7 at DIV14 and pulled-down at DIV18. pLV-hSYN-tGFP-P2A-BioID2 was used as the control. After receiving the mass spectrometry data in an excel file, the file was put through the CRAPome program to filter out nonspecific interactions, as it is a contaminant repository for affinity purification. Once the initial filter was processed, the new file was processed with SAINT (Significance Analysis of INTeractome), a computational tool that assigns confidence scores to protein-protein interaction data generated by using affinity-purification coupled with mass spectrometry. Out of over two thousand hits, 143 hits had a SAINT score of >0.8 (SAINT FDR ≤ 0.1). Hits included proteins that displayed high to

very high confidence of protein-protein interaction. 60 of these proteins showed overlap with Soderling's findings[51]. Using STRING, a database of known and predicted protein-protein interactions, we were able to visualize direct and indirect associations of the PPIs *(Figure 17).*

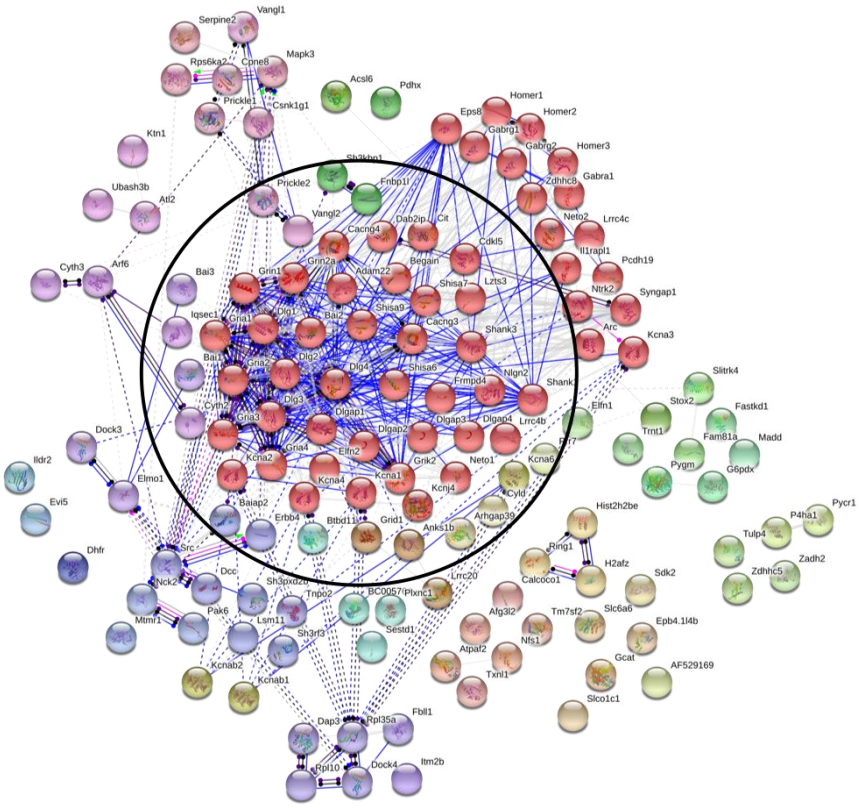

Figure 17: Protein-protein network analysis. Visual representation constructed using the STRING database. Interactions between candidate interaction proteins with PSD95. Proteins within the black circle represents overlapping PPIs with Soderling's PSD95 PPI findings.

CHAPTER 4: DISCUSSION

1. Summary of Findings

1.1 Endogenous BioID2 Tagging of PSD95, SHANK3, and OTUD7A

sgRNAs for targeting the C-terminus of each gene were designed by using the Doesch, Fusi et al algorithm to design sgRNAs. Special care was taken to select gRNAs that had high on-target scores, low off-target scores, and predicted off-target sites in non-coding regions. The gRNAs were integrated into the pX459 plasmid. Each CRISPR/Cas9 gRNA plasmid resulted in efficient cleavage efficiency after transfection.

The plasmids that delivered the homology templates were designed to target the C-termini of OTUD7A, SHANK3, and PSD95. Homology regions were amplified from E14Tg2a mESC genomic DNA. Sanger sequencing was used to validate all constructs.

The tools used to add a BioID2 tag to PSD95 and SHANK3 were designed, created and validated for use in mESCs, with the ultimate goal of generating transgenic mouse lines. However, there is a difference in the ability of certain regions of the genome to undergo homology directed repair, as suggested by the lack of success in tagging OTUD7A with BioID2.

We tested two protocols for generating neurons from mESCs. One using the transcription factor neurogenin-2 and the other utilizing media

changes for spontaneous neural differentiation. Both protocols were insufficient for mass induced neuron generation.

1.2 BioID2 Tagging of ASD Genes Overexpression Studies

The more than 60 putative ASD-associated genes that we selected are at different stages of the BioID2 cloning process. A full BioID2 experiment was done with pLV-hSYN-tGFP-P2A-mPSD95-BioID2, with pLV-hSYN-tGFP-P2A-BioID2 as its corresponding control. A paper published by Soderling et al. used an AAV-based in vivo BioID approach to find proteins in the synaptic complexes in mouse brain. They virally expressed inhibitory or excitatory PSD proteins, gephyrin and post synaptic density protein 95, respectively, fused to BioID to capture interacting proteins. These were pulled down and sent for mass spectrometry and analyzed. Through this method, they were able recapitulate known post synaptic density proteins, and they also identified many proteins previously unknown to exist in the inhibitory post synaptic density. Our approach used a lentivirus approach with BioID2 fused to PSD95 to transduce primary mouse cortical neurons and recapitulated many known proteins found in the postsynaptic density. 60 of the found proteins overlapped with Soderling's findings. The ability to recapitulate some but not all hits is to be expected as the systems used were different. This highlights the differences that can arise when using different models, and scientists have to be aware of the advantages and

caveats of each system. At the same time, the ability for our approach to capture many known proteins gives us confidence in the feasibility and potential impact of this experimental design.

2. Potential Pitfalls and Alternative Approaches

2.1 Endogenous BioID2 Tagging of PSD95, SHANK3, and OTUD7A

2.1.1 Caveats and alternative approaches: gRNA design and efficiency

While the CRISPR/Cas9 system has revolutionized the field of genome engineering, there are still disadvantages and limitations to using this system. One of the most important considerations of targeted genome editing is its specificity. Ideally, the tool used will introduce the desired modifications solely at the target site. However, the CRISPR/Cas9 system does not have perfect DNA recognition specificity, resulting in double stranded breaks in other parts of the genome.

Without a homology template, double-stranded breaks caused by the CRISPR/Cas9 system will undergo non-homologous end joining (NHEJ). DNA sites repaired through NHEJ will often undergo undesired changes such as insertions and deletions (indels) in the sequences of the genome, resulting in unpredictable consequences for the cell[52]. Currently, scientists are lacking an efficient and feasible method to screen for off-target effects. To mitigate the possible outcomes resulting from off-target effects, proactive measures were taken when selecting gRNAs. Only gRNAs that had minimal predicted off-targets that were mostly contained within introns were chosen. As the ultimate goal is to make transgenic mice using the cell lines produced via CRISPR/Cas9, the effects from off-target events will be minimal and

resulting mice will be cross-bred with wild type mice to dilute any undesired mutations.

2.1.2 Caveats and alternative approaches: Homology template design

As the homology template contains 3 areas of homology (5' homology, intronic homology, and 3' homology), there is a chance of partial integration of the homology template where desired recombination events do not occur within the intended regions. In some situations, only the loxP-neo-loxP cassette or the BioID2 sequence will be integrated. Careful screening employing neomycin selection, PCR amplification of targeted genomic DNA and Sanger sequencing allow us to select for clones with integration along the entire homology template.

2.1.3 Caveats and alternative approaches: Random integration of neomycin resistance cassette

After transfecting a practice cell line, we decided to select for positive clones after neomycin resistance selection (using the drug, G418) alone instead of combinatorial selection with puromycin followed by G418. Neomycin resistance selection began 48 hrs post-transfection to allow cells to integrate the homology template containing the neomycin resistance cassette into their genome. This approach is prone to false positives, as some cells under neomycin pressure will have only integrated the neomycin

resistance cassette randomly into their genome and the intended homologous recombination may not have occurred. This problem can be overcome by picking more G418-resistant colonies to increase the chance of selecting desired clones that have undergone successful homology-directed repair.

2.1.4 Caveats and alternative approaches: Difficulty of Otud7a-BioID2 integration

After various attempts at creating the *Otud7a*-BioID2 cell line, we concluded that the genomic locus seems to be very difficult to target. From a technical perspective, this could be due to inefficient cutting of the CRISPR/Cas9 gRNA or a poorly designed homology construct. The T7 endonuclease assay showed that the CRISPR/Cas9 gRNAs designed for targeting the C-terminus of *Otud7a* had approximately 50% efficiency, which is comparable to the cutting efficiency of the CRISPR/Cas9 gRNAs designed for *Shank3* and *Dlg4*. However, high cleavage activity does not necessarily correlate with efficient HDR-induced genome editing. The *Otud7a* homology construct design is similar to the *Shank3* and *Dlg4* homology templates in that they all have ~500 bp homology arms at the 5' and 3' ends, which should be sufficient for proper genome integration. However, there isn't a good method to truly test the performance of the homology template.

Technical factors aside, there are a few theories as to why BioID2 has been difficult to integrate into the C-terminus of the *Otud7a* gene.

There is currently limited information on OTUD7A and its full activity. Though it has not been found to be expressed in mESCs, it may play an essential role in cell survival. Thus, by introducing double stranded breaks in the gene region, cell viability may be adversely affected. Alternatively, the CRISPR/Cas9 gRNAs may be introducing breaks at off-target regions that are required for cell survival. However, this theory is less likely, as the majority of cells survived during the T7 endonuclease assay.

Secondly, certain genes in certain cell lines are more difficult to genetically engineer than others, though the underlying mechanism is unclear. This may be due differing abilities of cells to repair DSBs using non-homologous end joining versus homology directed repair with differential gene specificity.

While the HDR pathway has not yet been fully elucidated, research has shown that HDR is restricted to late S and G2 phases, when DNA replication is completed. Thus, synchronizing cells to those phases should increase HDR efficiency. A few studies have found that the addition of nocodazole, a G2/M phase synchronizer, increases HDR efficiency in various cell types. One paper found that the addition of both nocodazole and CCDN1, a cyclin that induces cell cycle transition from G0/G1 to S-

phase, had a synergistic effect, increasing HDR efficiency by 80-100% in human iPSCs, while decreasing NHEJ events[52]. Using nocodazole and CCDN1 could be used to increase HDR efficiency with the OTUD7A homology template.

As OTUD7A is not expressed in mESCs, the region may not be accessible for homology directed repair. Recently, a student from the Doble Lab found a correlation between Wnt-signaling and OTUD7A expression. I hypothesize that the activation of the Wnt-signaling pathway may lead to expression of the *Otud7a* gene. The synthetic molecule CHIR99021 could be used to examine its effect on OTUD7A expression levels. CHIR99021 is a GSK3 inhibitor that activates the WNT/β-catenin pathway via stabilization of β-catenin. If CHIR99021 upregulates OTUD7A expression levels, I could employ it in *Otud7a* targeting strategies, as the genomic locus would likely be more accessible for the integration of the homology template via HDR.

2.1.5 Caveats and alternative approaches: Producing neurons from BioID2 tagged mESC lines

We attempted to derive neurons from mESCs using the Ngn2 method or spontaneous differentiation. Though both methods were able to generate neural progenitor cells, the yield was very low and very heterogeneous. These methods were not sufficient to generate an adequate number of homogeneous neurons for meaningful BioID2 experiments.

To overcome this challenge, a mouse model could be generated from successfully targeted low-passage mESCs. Mouse brain tissue could then be extracted from the resulting mice for BioID2 experiments.

2.2 BioID2 Tagging of ASD Genes: Overexpression Studies

2.2.1 Caveats and alternative approaches: Lentiviral Plasmid Design

The lentiviral vectors that we have designed are multicistronic vectors that contain a “self-cleaving” 2A peptide to produce equimolar levels of multiple genes from a single mRNA template. The mechanism of 2A-mediated “self-cleavage” involves ribosomes skipping the formation of a glycyl-propyl peptide bond at the C-terminus of the 2A sequence, which results in translation of two cleaved proteins from one mRNA sequence: the protein upstream of the 2A sequence is attached to the majority of the 2A peptide, and the protein downstream of the 2A is attached to an extra proline at the N-terminus[53]. Neither should interfere with the function of the resulting proteins.

However, as this system is not perfect, there are two other possibilities that occur: (1) Ribosomal skipping is successful, but the ribosome falls off prematurely, inducing the termination of translation, which results in translation of only the protein upstream of 2A. (2) Unsuccessful skipping with translation resulting in an unintended fusion protein[53]. This phenomenon can be seen in western blots that revealed both properly

cleaved and uncleaved products. Additionally, in western blots stained with anti-tGFP, multiple bands were observed in samples from cells transfected with transfer plasmids that could not be due to cleaved product. These could arise from the production of truncated proteins via improper ribosome skipping resulting in premature stop codon(s). This phenomenon has not been described explicitly in the literature but has been experienced among other labs and Dr. Doble's lab members. However, this should not affect the final experiment greatly, as both proteins are produced, and the truncated version does not contain the BioID2 tag.

Moving forward, it may be beneficial to insert two consecutive 2A sequences to maximize cleavage efficiency and to minimize the possibility of the production of truncated proteins[53].

2.2.2 Caveats and alternative approaches: Virus Production

There are many variables to consider when making high titre virus. These range from the health of the cells to the ratios of plasmids used in transfections. We have tried many different conditions to optimize our virus production protocol. Our current protocol is efficient at producing virus from small transfer plasmids (< ~3kb cDNA insert), we will continue to optimize the protocol so that we are able to produce efficient amounts of virus for large transfer plasmids.

One variable that is difficult to control is the fetal bovine serum (FBS) used in our cell culture media. FBS is an integral component of our cell culture medium that influences cell viability, as it delivers nutrients, growth factors and attachment proteins to cells. It also protects the cells from oxidative damage and apoptosis. Each batch of FBS varies, is largely uncharacterized, and may affect the titre of the virus. Moving forward we can obtain highly qualified FBS batches.

Other factors influencing the ability to produce high titer virus include the quality of the plasmids used. Using an endotoxin-free maxiprep kit, which greatly reduces endotoxin levels in plasmid preps, could be beneficial. As larger genes tend to be less stable, we could also try decreasing the overall speed as well as the acceleration speed for the step involving ultracentrifugation, which is used to concentrate viral preparations[54]. The health of primary neuronal cultures at the time of transduction must also be considered for accurate titering of the virus as well as the BioID2 experiment. Culturing by senior members of the lab to maintain quality and consistency of primary cultures may serve to improve the success of viral transduction. Employing dedicated viral production personnel would also help to improve batch to batch consistency.

2.2.3. Caveats and alternative approaches: Variability between mass spectrometry runs

In the first set of experiments with pLV-hSYN-tGFP-P2A-PSD95-BioID2 and BioID2 control, we utilized a label-free method for quantitation. In label-free experiments, sample is measured in a separate mass spectrometry run. This increases the machine time as well as variability between samples due to differences in run conditions (e.g. temperature, experimenter, column conditions, etc.)[55].

To overcome this variability, we could perform a labelled experiment. In labelled experiments, samples of each condition are combined prior to the MS run. We could use a TMT labeling method in which it is incubated with our samples before combining them[56,57]. As all the samples are measured in the same MS run, there will not be differences due to run conditions.

2.2.4 Caveats and alternative approaches: BioID2 control

To determine viral titer, we infect primary cortical neurons at DIV3 and perform flow cytometry 48 hours following infection. As every virus is titered the same way, it should infect the same amount of cells when adjusted for MOI.

For the actual BioID2 experiments, the cells are infected at a later time point for 96 hours for a more accurate representation of protein

interactions in more mature neurons that have formed better networks. However, as determined through western blotting of the infected neurons of pLV-hSYN-tGFP-P2A-PSD95-BioID2 and pLV-hSYN-tGFP-P2A-BIOID2, there is a higher abundance of the BioID2 protein produced relative to the PSD95-BioID2 protein. This could lead to inaccurate analysis of our mass spectrometry data when using the BioID2 as control.

We hypothesize that this may be due to the size of the proteins. The cost of gene expression to a cell consists of investment of building blocks and energy and allocation of cellular resources, such as ribosomes and tRNAs. Due to a cell's limited energy and resources, smaller proteins usually take less time to translate than larger ones. Furthermore, the cell has to consider the accuracy of the translated protein so that it will be error free with the proper sequence and folding. The chances of an error occurring during translation increases the longer a sequence is. Errors can result in the formation of non-functional proteins and/or protein degradation[58].

I propose that a solution could be to insert a non-mammalian sequence before the BioID2 sequence in the control plasmid so that the BioID2 will be produced at a slower pace that more evenly matches the amount of proteins produced by other overexpression constructs. I propose pLV-hSYN-tGFP-P2A-Luciferase-BioID2 plasmid as an improved control plasmid, as luciferase is not found in mouse or human cells,

3. Future Directions

3.1 Endogenous BioID2 Tagging of PSD95, SHANK3, and OTUD7A

In the future transgenic mice can be generated to obtain ample amounts of neurons from the BioID2 tagged mESC lines I generated. Mice act as a good analogue for human biological processes, as both genomes share up to 99% similarity. From a practical perspective, mice also have a short reproductive cycle, large litters, and relatively low maintenance costs. Thus, genetically modified mice are a powerful tool to study human disease and gene functions *in vivo*.

Currently, there are a few conventional methods for generating chimeric mice. These methods include microinjection, well sandwich aggregation, or coculture of embryos with ES cells. There are pros and cons of each method such as the cost associated and technical difficulty. We will be using the microinjection method to produce our desired transgenic mice. The microinjection method produces good and reliable results but is hard to practice in the laboratory[59]. Therefore, this task will be outsourced to a commercial company.

Following successful screening, confirmed mESC clones will be sent for blastocyst microinjection. The positive mESCs lines were obtained by targeting the JM8.N4 mESC parental line derived from mice with black coat colour. They will be injected into white coat albino blastocysts before being

transferred to foster mothers. Successful integration of mESCs into the inner cell mass (ICM) will results in black and white chimeric pups. These pups will be genotyped to identify ones containing the BioID2 tag and further cross-bred to produce a homozygous BioID2 tagged mice. The brains of these mice will be collected for further mass spectrometric analyses.

3.2 BioID2 Tagging of ASD Genes: Overexpression Studies

For overexpression studies of BioID2 tagged ASD genes, the K. Singh lab will work to complete cloning and validation for each of the genes on the curated list in the next 3 months. The first batch of virus will consist of GRIA1 and PTEN, as there are more published studies describing these genes and they are small in size. We will also be testing out the pLV-hSYN-tGFP-P2A-Luciferase-BioID2 control to see if it will produce a more similar level of protein relative to other experimental constructs. From there, we will decide whether to continue with the Luciferase-BioID2 control or BioID2 only control. TMT-labeling will be done with cells infected with the next batch of virus and results will be compared with non-labelled experiments to inform regarding the best route forward. The long-term goal is to produce virus for each of the genes and to perform the BioID2 experiments on a large scale. We also want to perform some BioID2 experiments on clinically relevant ASD associated mutants to compare the PPIs of a normally functioning protein and its mutated counterpart. Through the analyses, we

hope to find common proteins that are shared between networks that can be used as targets in future experiments.

3.3 Concluding remarks

The findings in this thesis outlines and efficient way to generate larger knock-ins into mouse embryonic stem cells through CRISPR/Cas9 engineering. We have also developed a streamlined method to perform overexpression studies of ASD genes in a neuronal context. Our data outlines an optimized method to produce virus as well as validation of our BioID2 experimental design through PSD95 experimental runs. Overall, the experiments outlined in this thesis has the potential to elucidate common protein networks in ASD genes in a relevant context that may lead to therapeutics and biomarkers.

References

1. Autism Spectrum Disorder among Children and Youth in Canada 2018 - Canada.ca. Available at: https://www.canada.ca/en/public-health/services/publications/diseases-conditions/autism-spectrum-disorder-children-youth-canada-2018.html. (Accessed: 2nd August 2018)
2. Dilip Jeste, P. V *et al.* American Psychiatric Association Board of Trustees. (2012).
3. Faras, H., Al Ateeqi, N. & Tidmarsh, L. Autism spectrum disorders. *Ann. Saudi Med.* **30**, 295–300 (2010).
4. NIMH » Autism Spectrum Disorder. Available at: https://www.nimh.nih.gov/health/topics/autism-spectrum-disorders-asd/index.shtml. (Accessed: 6th February 2018)
5. Zafeiriou, D. I., Ververi, A. & Vargiami, E. Childhood autism and associated comorbidities. *Brain Dev.* **29**, 257–272 (2007).
6. Moyal, W. N., Lord, C. & Walkup, J. T. Quality of Life in Children and Adolescents with Autism Spectrum Disorders: What Is Known About the Effects of Pharmacotherapy? *Pediatr. Drugs* **16**, 123–128 (2014).
7. van Heijst, B. F. & Geurts, H. M. Quality of life in autism across the lifespan: A meta-analysis. *Autism* **19**, 158–167 (2015).
8. Vos, T. *et al.* Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: A systematic analysis for the Global Burden of Disease Study 2016. *Lancet* **390**, 1211–1259 (2017).
9. Fakhoury, M. Autistic spectrum disorders: A review of clinical features, theories and diagnosis. *Int. J. Dev. Neurosci.* **43**, 70–77 (2015).
10. Hansen, S. N., Schendel, D. E. & Parner, E. T. Explaining the increase in the prevalence of autism spectrum disorders: The proportion attributable to changes in reporting practices. *JAMA Pediatr.* **169**, 56–62 (2015).
11. Lord, C. Autism From 2 to 9 Years of Age. **63**, (2006).
12. Yoo, H. Genetics of Autism Spectrum Disorder: Current Status and Possible Clinical Applications. *Exp. Neurobiol.* **24**, 257 (2015).
13. Economic Burden of Childhood Autism Spectrum Disorders.
14. Newschaffer, C. J. *et al.* The Epidemiology of Autism Spectrum Disorders. *Annu. Rev. Public Health* **28**, 235–258 (2007).
15. Lyall, K. *et al.* The Changing Epidemiology of Autism Spectrum Disorders. *Annu. Rev. Public Health* **38**, 81–102 (2017).
16. Sandin, S. *et al.* The Heritability of Autism Spectrum Disorder. *Jama* **318**, 1182 (2017).
17. Tick, B., Bolton, P., Happé, F., Rutter, M. & Rijsdijk, F. Heritability of autism spectrum disorders: a meta-analysis of twin studies. *J. Child*

Psychol. Psychiatry. **57**, 585–95 (2016).
18. Jonathan Chang, Sarah R Gilman, Andrew H Chiang, Stephan J Sanders, & D. V. *et al.* Genotype to phenotype relationships in autism spectrum disorders. *Nat. Neurosci.* **18**, 191–8 (2015).
19. Betancur, C. Etiological heterogeneity in autism spectrum disorders: More than 100 genetic and genomic disorders and still counting. *Brain Res.* **1380**, 42–77 (2011).
20. Hazelbaker, D. Z. *et al.* A Scaled Framework for CRISPR Editing of Human Pluripotent Stem Cells to Study Psychiatric Disease. *Stem Cell Reports* **9**, 1315–1327 (2017).
21. Nachman, M. W. Single nucleotide polymorphisms and recombination rate in humans. *Trends Genet.* **17**, 481–485 (2001).
22. Chen, J. A., Peñagarikano, O., Belgard, T. G., Swarup, V. & Geschwind, D. H. The Emerging Picture of Autism Spectrum Disorder: Genetics and Pathology. *Annu. Rev. Pathol. Mech. Dis.* **10**, 111–144 (2015).
23. Vissers, L. E. L. M. *et al.* A de novo paradigm for mental retardation. *Nat. Genet.* **42**, 1109–1112 (2010).
24. Singh, S. K. & Eroglu, C. Neuroligins Provide Molecular Links Between Syndromic and Non-Syndromic Autism. *Sci. Signal.* **6**, re4
25. Szoko, N., McShane, A. J. & Natowicz, M. R. Proteomic explorations of autism spectrum disorder. *Autism Res.* **10**, 1460–1469 (2017).
26. Roux, K. J., Kim, D. I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J. Cell Biol.* **196**, 801–810 (2012).
27. Chapman-Smith, A. & Cronan, J. E. The enzymatic biotinylation of proteins: A post-translational modification of exceptional specificity. *Trends Biochem. Sci.* **24**, 359–363 (1999).
28. Barker, D. F. & Campbell, A. M. The birA gene of Escherichia coli encodes a biotin holoenzyme synthetase. *J. Mol. Biol.* **146**, 451–467 (1981).
29. Kwon, K. & Beckett, D. Function of a conserved sequence motif in biotin holoenzyme synthetases. *Protein Sci.* **9**, 1530–1539 (2000).
30. Chapman-Smith, a & Cronan, J. E. Molecular biology of biotin attachment to proteins. *J. Nutr.* **129**, 477S-484S (1999).
31. Barat, B. & Wu, A. M. Metabolic biotinylation of recombinant antibody by biotin ligase retained in the endoplasmic reticulum. *Biomol. Eng.* **24**, 283–291 (2007).
32. Kim, D. I. *et al.* An improved smaller biotin ligase for BioID proximity labeling. *Mol. Biol. Cell* **27**, 1188–1196 (2016).
33. Stathakis, D. G., Hoover, K. B., You, Z. & Bryant, P. J. Human postsynaptic density-95 (PSD95): Location of the gene (DLG4) and possible function in nonneural as well as in neural tissues. *Genomics* **44**, 71–82 (1997).

34. Chen, X. *et al.* PSD-95 family MAGUKs are essential for anchoring AMPA and NMDA receptor complexes at the postsynaptic density. *Proc. Natl. Acad. Sci.* **112**, E6983–E6992 (2015).
35. Prange, O., Wong, T. P., Gerrow, K., Wang, Y. T. & El-Husseini, A. A balance between excitatory and inhibitory synapses is controlled by PSD-95 and neuroligin. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 13915–20 (2004).
36. Bustos, F. J. *et al.* Epigenetic editing of the Dlg4/PSD95 gene improves cognition in aged and Alzheimer's disease mice. *Brain* **140**, 3252–3268 (2017).
37. Stessman, H. A. F. *et al.* Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. *Nat. Genet.* **49**, 515–526 (2017).
38. Halbedl, S., Schoen, M., Feiler, M. S., Boeckers, T. M. & Schmeisser, M. J. Shank3 is localized in axons and presynaptic specializations of developing hippocampal neurons and involved in the modulation of NMDA receptor levels at axon terminals. *J. Neurochem.* **137**, 26–32 (2016).
39. Boeckers, T. M., Bockmann, J., Kreutz, M. R. & Gundelfinger, E. D. ProSAP/Shank proteins - a family of higher order organizing molecules of the postsynaptic density with an emerging role in human neurological disease. *J. Neurochem.* **81**, 903–10 (2002).
40. Park, E. *et al.* The Shank family of postsynaptic density proteins interacts with and promotes synaptic accumulation of the beta PIX guanine nucleotide exchange factor for Rac1 and Cdc42. *J. Biol. Chem.* **278**, 19220–9 (2003).
41. Durand, C. M. *et al.* Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nat. Genet.* **39**, 25–7 (2007).
42. Gauthier, J. *et al.* Novel de novo SHANK3 mutation in autistic patients. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **150B**, 421–424 (2009).
43. Forsingdal, A., Fejgin, K., Nielsen, V., Werge, T. & Nielsen, J. 15Q13.3 Homozygous Knockout Mouse Model Display Epilepsy-, Autism- and Schizophrenia-Related Phenotypes. *Transl. Psychiatry* **6**, e860 (2016).
44. Uddin, M. *et al.* OTUD7A Regulates Neurodevelopmental Phenotypes in the 15q13 . 3 Microdeletion Syndrome. *Am. J. Hum. Genet.* 278–295 (2018). doi:10.1016/j.ajhg.2018.01.006
45. Amerik, A. Y. & Hochstrasser, M. Mechanism and function of deubiquitinating enzymes. *Biochim. Biophys. Acta - Mol. Cell Res.* **1695**, 189–207 (2004).
46. Abrahams, B. S. *et al.* SFARI Gene 2.0: A community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol.*

Autism **4**, 1 (2013).
47. Ryan KC Yuen, Daniele Merico, Matt Bookman, Jennifer L Howe, Bhooma Thiruvahindrapuram, Rohan V Patel, Joe Whitney, Nicole Deflaux, Jonathan Bingham, Zhuozhi Wang, Giovanna Pellecchia, Janet A Buchanan, Susan Walker, Christian R Marshall Mohammed Uddin, M, B. J. F. *et al.* Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nat. Neurosci.* **20**, 602–611 (2017).
48. Hoang, N., Buchanan, J. A. & Scherer, S. W. Heterogeneity in clinical sequencing tests marketed for autism spectrum disorders. *npj Genomic Med.* **3**, 1–4 (2018).
49. Satterstrom, F. K., Kosmicki, J. A., Wang, J. & Breen, M. S. Novel genes for autism implicate both excitatory and inhibitory cell lineages in risk. 1–52 (2018).
50. Aldrich Corp, S. CLONING AND EXPRESSION FLAG ® and 3xFLAG ® Overview A Proven System for Detection and Purification of Proteins Ideal Epitope Tags: Small, Hydrophilic and Cleavable.
51. Uezu, A. *et al.* Identification of an elaborate complex mediating postsynaptic inhibition. *Science (80-.).* **353**, 1123–1129 (2016).
52. Zhang, J.-P. *et al.* Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. *Genome Biol.* **18**, 35 (2017).
53. Kim, J. H. *et al.* High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PLoS One* **6**, 1–8 (2011).
54. Yacoub, N. al, Romanowska, M., Haritonova, N. & Foerster, J. Optimized production and concentration of lentiviral vectors containing large inserts. *J. Gene Med.* **10**, 610–618 (2008).
55. Wang, M., You, J., Bemis, K. G., Tegeler, T. J. & Brown, D. P. G. Label-free mass spectrometry-based protein quantification technologies in proteomic analysis. *Briefings Funct. Genomics Proteomics* **7**, 329–339 (2008).
56. Tandem Mass Tag Systems - CA.
57. Zhang, L. & Elias, J. E. Relative Protein Quantification Using Tandem Mass Tag Mass Spectrometry. in 185–198 (Humana Press, New York, NY, 2017). doi:10.1007/978-1-4939-6747-6_14
58. Gingold, H. & Pilpel, Y. Determinants of translation efficiency and accuracy. *Mol. Syst. Biol.* **7**, 1–13 (2011).
59. Lee, K.-H. Methods to Generate Chimeric Mice from Embryonic Stem Cells. in *Embryonic Stem Cells - Basic Biology to Bioengineering* (InTech, 2011). doi:10.5772/23420