C. elegans models of ASD-associated missense variants

Thesis by

Wan-Rong Wong

In Partial Fulfillment of the Requirements for the degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2021

Defended February 4, 2021

Wan-Rong Wong
ORCID: 0000-0002-9757-814

ACKNOWLEDGEMENTS

I would never have imagined this moment: I am writing my Ph.D. dissertation's acknowledgement on my birthday. I cannot imagine a better day to reflect and express my gratitude. I am extremely grateful for all the people and opportunities I encountered in the past seven years.

First and foremost, I want to thank my thesis committee: Professors Paul Sternberg, David Prober, Kai Zinn, and Betty Hong. Thank you for sparing the time from your busy schedule to give me scientific and life advice. You always caught the blind spots in my project and provided crucial scientific criticism, good ideas, and in-depth knowledge of the field. Special thanks to my advisor Paul. Doing *C. elegans* research was not my original plan, but you were brave to take me on-board. Your enthusiasm for science inspires me to challenge my knowledge every day. Your generosity taught me to be a better team player. You are always willing to share, from small things like dumplings in your lunch box to big things like reagents and projects. You showed me the true collaborative spirit in science.

A large part of my Ph.D. thesis project was done as group work. I would like to thank the fantastic autism team: Katie Brugman, Shayda Maher, Daniel Oh, Shahla Gharib, Mihoko Kato, and Kevin Howe. Thanks for being supportive and encouraging during the ups and downs of the project. Working with you guys has been an amazing experience. I would also like to thank my former students: Katie Wendover, Judy Kim, Greg Grigorian, and Altyn Rymbek. They explored the functions of variant mutations with me and pointed out the best way to take them. I learned so much from being a mentor.

I am fortunate to work with a group of talented Ph.D. students and postdocs in the Sternberg lab. I want to thank Han Wang for helping me set up the CRISPR pipeline and Chun-Hao Chen for clarifying the BRAF project's scientific goals. I would also like to thank Pei Shih for being my mentor during my rotation and a good friend throughout the Ph.D. program. I want

to thank Hillel Schwartz for the expertise in genetics and critical revision of manuscripts. I would also like to thank Barb Perry, Sarah Torres, Dan Leighton, Sarah Cohen, and Wilber Palma for keeping the lab organized and safe. To my bench mates: Ryoji Shinya, Chris Cronin, James Lee, Ravi Nath, Ali Akagi, Heenam Park, Mandy Tan, Stephanie Nava, and Mark Zhang; thank you for providing a great social vibe that makes coming to work enjoyable. To my lunch buddies: Wen Chen, Mengyi Cao, James Tan, Yan Jin, and Olivia Wan; I miss our entertaining conversations at the lounge before COVID. To my coffee break partners: David Angeles, Jon Liu, Porfi Quintero, Carmie Puckett-Robinson, Liz Holman, and Sharan Prakash; thank you for all the inspiring conversations during coffee breaks.

I am lucky to have a support group outside of the lab. Cynthia Chai, thank you for being my biggest support in the first few years of graduate school. I cannot ask for a better roommate and collaborator. My fellow neurobiology classmates Zeynep Turan, Zhannetta Gugel, Bing Yang, Sangjun Lee, and Dhruv Zocchi are like brothers and sisters to me, and you always will be. My mid-night snack buddies: Sophia Cheng and Ruohan Wang, I am blessed to have you girls in life. I want to thank the Caltech Taiwanese Association for making the life transition into a new country much easier. I would also like to thank Charisma Bartlett, Aunty Shelly, Aunty Lily, and Mrs. Han for being my extended family in California and showing your love for me. Many thanks to my life-long friends worldwide: Hsiao-Wen Wu, Min-Ru Hsu, Tracy Liu, Becky Chou, and Melody Kao; you girls are my biggest cheerleaders.

Last but not least, I would like to thank my family. Mom and Dad, thanks for continually had so much faith in me and pushed me towards success every step of the way. My little sister, Wan-Ling, you are my muse for the journey of studying psychiatric disorders. Grandpa and Grandma, I wish to share this degree with you, wherever you are. Lastly, to my number one fan, Albert Han, your love and support helped me flourished. Thanks for always pushing me out of my comfort zone and accompany me with your unique sense of humor.

ABSTRACT

The evolving next-generation sequencing technology accelerates the identification of disease-associated genetic variants. However, interpretation of these variants remains challenging, especially variants with subtle effects such as missense variants. Missense variants account for a large proportion of genetic variants in human diseases, including autism spectrum disorder (ASD). The causal relationship of most missense variants in the pathogenesis of ASD has not yet been demonstrated, and an experimental method systematically prioritizing missense alleles can gain crucial insight into the molecular basis for disease pathology. Therefore, I developed an *in vivo* multi-cellular system using *Caenorhabditis elegans* to systematically evaluate the functional consequences of disease-associated missense variants. I identified highly conserved human ASD-associated missense variants in their *C. elegans* orthologs, used a CRISPR/Cas9-mediated homology-directed knock-in strategy to generate missense mutants, and analyzed their impact on behaviors and development via several broad-spectrum assays. Overall, I generated 60 ASD-associated missense variants and characterized these missense mutant strains using a fecundity assay, an automated locomotor tracking system, and a chemotaxis assay. I found that 19% of the human disease-associated alleles have conserved loci in their *C. elegans* orthologs. Among the genes I tested, 64-70% of the missense variants predicted to perturb protein function showed detectable phenotypic changes in morphology, locomotion, or fecundity. Our results also revealed that missense mutants in different gene networks displayed distinct phenotypic profiles. Moreover, I focused on studying the genetic properties of missense variants on two ASD risk genes. I discovered the developmental defects in the ALDH1A3 C174Y missense mutation involved in the retinoic acid signaling pathway. I also identified a conserved missense residue lin-45(K565N), orthologous to human BRAF(K499N), which displayed a hypersensitive non-dominant phenotype in the diacetyl chemotaxis assay that was capable of being inhibited by RNAi. The finding suggests a potential gain-of-function allele in *BRAF*,

especially in the sensory function. To sum up, I established a working pipeline to systematically identify and generate evolutionarily conserved ASD-associated missense mutants in *C. elegans*. This approach will help assess the impact of a single missense mutation in the whole organism and prioritize consequential missense variants for further intensive analysis in vertebrate models and human cells.

PUBLISHED CONTENT AND CONTRIBUTIONS

Wong, W. R., Brugman, K. I., Maher, S., Oh, J. Y., Howe, K., Kato, M., & Sternberg, P. W. (2019). Autism-associated missense genetic variants impact locomotion and neurodevelopment in Caenorhabditis elegans. *Human Molecular Genetics*, *28*(13), 2271–2281. https://doi.org/10.1093/hmg/ddz051

W.R. Wong participated in designing research, performing experiments, contributing new reagents, analyzing data, and writing the manuscript.

Wong, W. R., Maher, S., Oh, J. Y., Brugman, K. I., Gharib, S., & Sternberg, P. W. (2021). Conserved missense variant in ALDH1A3 ortholog impairs fecundity in C . elegans. *MicroPubl Biol*, https://doi.org/10.17912/micropub.biology.000357.

W.R. Wong participated in designing research, performing experiments, contributing new reagents, analyzing data, and writing the manuscript.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS AND TABLES

NOMENCLATURE

Autism spectrum disorder (ASD). A neurodevelopmental disorder characterized by the impaired social ability and repetitive behaviors.

***Caenorhabditis elegans* (*C. elegans*).** A free-moving transparent nematode commonly used to genetic study.

Missense variants. A type of non-synonymous mutation where the nucleic acid change results in an amino acid difference in the protein sequence.

Clustered regularly interspaced short palindromic repeats (CRISPR). A gene-editing system adapted from the bacterial defense system that can be programmed to introduce double-strand breaks at precise DNA locations in the genome.

Chapter 1

GENERAL INTRODUCTION

1.1 What is autism spectrum disorder?

Autism was first described as a trait of "enclosure in one's self" by Doctor Leo Kanner in 1943. In his paper, Dr. Kanner documented a cohort of 11 children with "extreme desire for aloneness and sameness and lack of direct affective contact" (Kanner, 1943). Although the clinical description of these 11 children varied, they all shared a combination of symptoms: severe social deficits, language dysfunction, and the presence of repetitive restrictive behaviors. Over the past decades, broader diagnosis criteria have been adapted to include those with normal intelligence and language (Asperger syndrome) and those almost meeting the strict diagnostic criteria in all three domains (pervasive development disorder – not otherwise specified, PDD-NOS) as autism spectrum disorders (ASD) (American Psychiatric Association, 2013). Ever since the first documentation of autism, intense efforts have been attempted to study the physiopathology underlying this complex disorder. However, current diagnoses still rely solely on expert observation and assessment of behavior and cognition. Due to the high heritability shown in ASD, more evidence point to the understanding of genetic risk variants can help untangle the heterogeneity of ASD.

1.1.1 Impact of ASD

ASD is one of the most prevalent neurodevelopmental disorders. The reported prevalence of autism has increased over the past decades, partly due to better knowledge of the disease variability, broader diagnosis with a higher degree of public and professional awareness of the disorder. Currently, ASD affects 1 in 54 children in the US, and it is 4.3 times more common

among boys than girls (Maenner et al., 2020). According the Surveillance Summaries from Centers for Disease Control and Prevention (CDC), ASD prevalence varies slightly by each state but it impacts identically on all ethnic groups, except for the Hispanic group. Hispanic group shows lower ASD prevalence (approximately 15.4) compared to the prevalence of 18.5 per 1,000 children in the general population. Globally, ASD affects 52 million people and accounts for 7.7 million disability-adjusted life years (DALYs), which indicates the number of years lost due to the disability (Baxter et al., 2015). It is estimated that about 30% of children with ASD remain minimally verbal throughout life, even after receiving years of interventions and a range of educational opportunities (Tager-Flusberg & Kasari, 2013). The burden of ASD commences in infancy and persists through their lifespans.

There are multiple challenges in identifying and quantifying ASD cases. First, it is difficult to detect ASD in young children without verbal capability. Only half of the children with ASD have their first evaluation by the age of 36 months, and the overall median age of ASD diagnosis is higher in black children (Maenner et al., 2020). Interestingly, higher ASD prevalence is reported among more socioeconomically advantaged groups, which is unique among developmental disorders. The second challenge comes from the high comorbidity of ASD and other psychiatric disorders. As many as 85% of individuals with ASD have other comorbid psychiatric diagnoses. The most common diagnosed comorbidities are seizure, attention-deficit/ hyperactivity disorder (ADHD), anxiety, and depression (Doshi-Velez et al., 2014). In addition, 33% of individuals with ASD also display intellectual disabilities (ID, intelligence quotient [IQ] ≤70) (Maenner et al., 2020), indicating a sharing disease mechanism between ASD and other mental disorders.

1.1.2 Symptoms of ASD

According to the latest version of Diagnostic and Statistical Manual of Mental Disorders (DSM-5), ASD is characterized by two core features: (1) persistent deficits in social communication and social interaction across multiple contexts; and (2) restricted, repetitive patterns of behavior, interests, or activities (American Psychiatric Association, 2013). In addition to the core domains, individuals with ASD also display a wide range of symptoms, such as sleep disruption (80%), gastrointestinal problems (70%), motor abnormalities (79%), intellectual disability (45%), or epilepsy (30%) (Lai et al., 2014). Additionally, language delays are frequently co-occurring with ASD and were even included in the DSM-IV diagnostic criteria (Rylaarsdam & Guemez-Gamboa, 2019). These strikingly clinical heterogeneities were reflected in the DSM-5 by introducing the idea of a "spectrum" of symptoms and severity (Chaste et al., 2017).

To date, diagnosis of ASD relies mostly on clinical behavioral assessments, including administration of the Autism Diagnostic Observation Schedule 2 (ADOS-2) (Catherine Lord et al., 2012) and the Autism Diagnostic Interview-Revised (ADI-R)(C Lord et al., 1994). Although these behavioral assessments show high sensitivity and specificity in diagnosing ASD cases as young as 12 months old (Charman & Gotham, 2013), they require well-trained examiners to carry out the interviews, which can be subjective and time-consuming. Moreover, research showed that the presence of shared symptoms of ADHD and sensory processing in children with ASD could delay the ASD diagnosis for three years (Miodovnik et al., 2015). Therefore, we are in need of more reliable biomarkers for earlier and more accurate diagnoses.

Over the past few decades, countless efforts have been devoted to finding a reliable biomarker for ASD. The first set of evidence came from the analyses of postmortem ASD brain tissues. Studies on ASD postmortem brains revealed both global and cellular changes, including disorganized gray and white matter, increased number of neurons, decreased volume

of neuronal soma, and increased neuropil, as well as changes in densities of dendritic spines and cerebral vasculature (Varghese et al., 2017). Given the fruitful findings from postmortem studies, the consistency in conclusions is dampened, and studies with larger sample size are desired to validate the findings. Fortunately, recent advances in magnetic resonance imaging (MRI) provided more lines of evidence with the information of the developmental timeline in the pathogenesis of ASD. An MRI study detected an enlargement of cortical surface area in infants as early as 6-12 months of age, and the authors conclude that the brain volume overgrowth accurately align with the autistic social deficits in later life (Hazlett et al., 2017). In addition, functional imaging studies highlighted the altered activation patterns in brain regions involved in processing social cues and regulating emotions, such as fusiform gyrus and anterior cingulate cortex, during cognitive tasks in individuals with ASD (Minshew & Keller, 2010). Moreover, a meta-analysis of resting-state functional magnetic resonance imaging (rs-fMRI) revealed reduced local connections in brain regions guarding attention and sensory integration (e.g., dorsal posterior cingulate cortex and right medial paracentral lobule (Lau et al., 2019). Given the fruitful findings from imaging studies, the consistency in conclusions is limited by technical challenges, such as low temporal resolution and the need for higher functioning ASD individuals as subjects, and thus studies with larger sample size are desired (Varghese et al., 2017).

1.1.3 Etiology of ASD

Despite the considerable efforts to study the pathogenesis of ASD, the majority of ASD cases belong to idiopathic ASD, which refers to cases without known specific causes. Scientists proposed the contribution of both genetic factors (nature) and environmental factors (nurture) to the etiology of ASD, and twin and family studies provided clues to the "nature or nurture" debate. Early twin studies reported a 70-90% concordance rate in ASD diagnosis in

monozygotic twins compared to the 10% concordance rate shared in dizygotic twins, suggesting a high heritability of about 90% (Bailey et al., 1995; Steffenburg et al., 1989). A higher sibling recurrence rate of 7-20% is also documented in the siblings of ASD individuals, depending on study design and whether half or full siblings were considered (Ozonoff et al., 2011; Palmer et al., 2017). Recent studies calculated a more precise estimation of ASD heritability decreased to 83%, meaning that 83% of the variation in autistic traits in a population can be explained by genetic factors (Sandin et al., 2017). Overall, the high heritability implicated in ASD highlights the strong components of genetic variants, yet it also indicated the involvement of environmental factors in the pathogenesis of ASD.

Multiple environmental risk factors for ASD have been proposed, including parental ages, pregnancy-related factors, maternal medications and diseases, nutrition, environmental toxins, etc. Recent studies suggested that environmental risk factors may determine up to 40–50% of variances in ASD liability. However, investigations on ASD environmental risk factors often produce inconsistent results. In order to resolve the inconsistency in previous studies, Modabbernia et al. performed a systematic review and meta-analyses of 663 studies, and they found strong links of advanced parental ages and birth complications (e.g., birth trauma or ischemia and hypoxia showed) to ASD. They also discovered modest (but significant) associations between pregnancy-related risk factors (e.g., maternal obesity, maternal diabetes, and caesarian section) and prenatal exposure to heavy metals with ASD. Other environmental risk factors such as nutrition elements (e.g., folic acid, omega 3, and vitamin D), maternal immune activation, and maternal nicotine and thimerosal exposures showed inconsistent results across studies (Modabbernia et al., 2017). These environmental risk factors potentially influence neurodevelopment through epigenetic mechanisms. Indeed, a histone acetylome-wide association study (HAWAS) using chromatin immunoprecipitation sequencing (ChIP-seq) discovered that 68% of syndromic and idiopathic ASD cases shared a common acetylome

signature in patients with both rare mutations and common “epi-mutations”, suggesting a convergence into common signaling pathways in ASD (W. Sun et al., 2016). Therefore, defining the molecular pathologies underpinning ASD will allow for a more precise establishment of subgroups that are consistently linked to a specific biomarker, enabling the discovery of new therapeutic targets (McCammon & Sive, 2015).

The known genetic factors contribute to about 20-30% of ASD cases (De Rubeis & Buxbaum, 2015; Pinto et al., 2014). The earliest understanding of ASD genetic origin came from truncated mutations identified in individuals with Mendelian genetic syndromes (e.g., Fragile X, Rett, and Angelman Syndromes, tuberous sclerosis). These ASD-liked mental disorders share some autistic characteristics and have a high comorbid rate with ASD. The familial studies of these Mendelian genetic syndromes enabled identifying large truncated protein on single genes, such as FMR1, MECP2, UBE3A, TSC1, and TSC2 (Chung et al., 2014; Y. S. Kim & Leventhal, 2015). It is estimated that these inherited mutations account for 5% of ASD cases (Varghese et al., 2017).

With rapidly evolving next-generation sequencing (NGS) technologies, there has been an explosion in the discovery of smaller pieces of ASD genetic variants. One such category is copy number variations (CNVs), which refers to the repeated DNA segments with kilobase to several mega-bases in range present in various numbers of copies between individuals. These CNVs can result in gain-of-function (duplication and insertional transposition), loss-of-function (deletion), and complex chromosome rearrangement (inversions and chromosome translocation). The application of NGS and chromosome microarray (CMA) has identified numerous CNVs and cytogenetically visible chromosomal anomalies (e.g., 15q11-13, 22q11.2) that affect almost all chromosomes (Chung et al., 2014). Since the first publications of CNV studies (Iafrate et al., 2004; Sebat et al., 2007, 2004), it has helped identify more than 100

disease genes and over 40 genomic loci among patients with ASD. Overall, CNVs comprise approximately 10% of ASD cases (Varghese et al., 2017).

The breakthrough on whole-exome sequencing (WES) in the last decade facilitates the discovery of polygenetic single nucleotide variants (SNVs) that existed in individuals with ASD. Scientists can distinguish the origin of the SNVs using samples from simplex families (i.e., sequencing one affected individual and both of his/her parents) or quartet families (i.e., a simplex family plus one unaffected sibling). Some of the SNVs are inherited, whereas others are *de novo* mutations (i.e., mutations arising in the germline that are not present in the parental somatic genome). A serial of studies using WES to investigate the ASD trios in the Simons Simplex Collection documented a vast increase of gene-disruptive *de novo* mutations (nonsense, splice-site, or frameshift) and a smaller increase of *de novo* missense variants and transmitted nonsense mutations in ASD probands (Iossifov et al., 2012; Neale et al., 2012; O'Roak et al., 2012; Sanders et al., 2012) as well as patients with other psychiatric conditions (Kenny et al., 2014). Another WES study identified that 16% of ASD probands carry a *de novo* loss-of-function mutation, whereas 9% of their unaffected siblings also carry such mutations (De Rubeis et al., 2014). Interestingly, an analysis using simplex cases in the Simons Simplex Collection revealed a negative association between *de novo* loss-of-function (LOF) mutation rate and intellectual quotient (IQ), meaning that individuals with ASD and IQ<100 harbor excess of *de novo* loss-of-function mutations in comparison to their higher functioning counterparts. The study also pointed out that individuals with ASD and high IQ were more likely to have family histories of psychiatric disease and thus a more significant familial burden (Robinson et al., 2014).

Overall, these WES studies consistently showed that affected ASD subjects have roughly twice as many gene-disruptive *de novo* loss-of-function variants than would be expected by chance, and at least half of the *de novo* loss-of-function have some impact on ASD liability (Chaste et

al., 2017). Noticeably, these *de novo* mutations displayed strong origins with paternal germline mosaicism, and the rate of *de novo* mutations was positively correlated with paternal ages (Iossifov et al., 2012; O'Roak et al., 2012; Sanders et al., 2012). This gave an example of how environmental risk factors impacted genetic components and shape the ASD pathogenesis trajectory.

1.2 Missense variants (MS) in ASD

1.2.1 Types of ASD genetic variants

In addition to their origins sizes, genetic variants can be categorized by their allele frequency in a given population. Common variants (population frequency >1%), also known as single nucleotide polymorphism (SNPs), contribute to the pathogenesis of ASD. Compared to rare variants, common variants are under strong selection pressure and are less likely to have a strong negative effect on early survival or fecundity (reproductive fitness) (Geschwind & State, 2015). Overall, these common variants accounted for 52.4% of ASD etiology when they were additively considered (Gaugler et al., 2014). However, early genome-wide association studies (GWAS) detected few genome-wide significant risk loci, and these studies were of low statistical power due to the low effect size (Anney et al., 2010; Ma et al., 2009; Weiss et al., 2009). Subsequent large-scale GWAS study performed a meta-analysis with over 15,000 samples and revealed 93 significant genome-wide markers, many of which were replicated in independent cohorts (Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium, 2017; Grove et al., 2019). Some SNPs implicated in these studies overlapped with schizophrenia and other psychiatric disorders, suggesting shared genetic underpinnings among psychiatric disorders.

Most *de novo* mutations belong to rare (population frequency <1%) or very rare (population frequency < 0.01) genetic variants. The advance of NGS arouses the interest in finding *de novo* mutations associated with ASD. Several studies identified thousands of ASD-associated *de novo* missense variants in hundreds of genes utilizing the WES data from the Simon Simplex Collection, which consists of a set of 2,517 simplex families (Fischbach & Lord, 2010). Each of these genes harbors two or more dnLoF mutations identified in separate probands, corresponding to a statistically significant false discovery rate (FDR) of 0.1 (Iossifov et al., 2012; Neale et al., 2012; O'Roak et al., 2012; Sanders et al., 2012). A complete list of missense variants can be found on the Simon Foundation Autism Research Initiative (SFARI) website (https://gene.sfari.org/database/human-gene/) (Fischbach & Lord, 2010).

1.2.2 Contribution of missense variants to ASD

A recent project done by the Autism Sequencing Consortium (ASC) examined 6430 individuals with ASD and revealed 7131 *de novo* variants in protein-coding exons, among which 4503 (63.1%) were missense variants and 972 (13.6%) were protein-truncating variants (Satterstrom et al., 2020). Indeed, individuals with ASD showed a slightly higher *de novo* mutation rate than the $1.18*10^{8}$ rate identified in the general population (Alonso-Gonzalez et al., 2018; Conrad et al., 2011). The enrichment of gene-disrupting *de novo* loss-of-function mutations was consistently found in individuals with ASD. Heterozygous loss-of-function *de novo* mutations presented in 20% of ASD probands but only in 10% of unaffected siblings (Iossifov et al., 2012; Ronemus et al., 2014; Sanders et al., 2012). In particular, *de novo* missense mutations, especially the most damaging types, were more enriched in ASD probands than in their unaffected siblings, and these missense mutations contributed to at least 10% of ASD diagnoses (Iossifov et al., 2014).

1.2.3 Gene network and pathway analysis

Over decades, large-scale exome sequencing studies have uncovered hundreds of genes associated with ASD (Coe et al., 2019; Iossifov et al., 2012; Kaplanis et al., 2020; Neale et al., 2012; O'Roak et al., 2012; Sanders et al., 2012; Satterstrom et al., 2020). Given the heterogeneity of ASD, it would be valuable to find some molecular convergence in ASD risk genes. Indeed, studies combining gene expression, protein-protein interactions (PPIs), and other systematic gene annotation resources suggested that ASD-associated genes clustered in two major gene network: those regulated gene expression during early fetal development and those involved in synaptic functions during late fetal development (Ben-David & Shifman, 2013; Chaste et al., 2017; Gilman et al., 2011; Sakai et al., 2011; Voineagu et al., 2011).

To identify the genetic network, few groups discovered enrichments of ASD risk genes in chromatin remodeling complexes. These chromatin modifier genes encode proteins, such as histone demethylase and fragile X mental retardation protein (FMRP), transcriptionally regulated gene expression during early fetal development (De Rubeis et al., 2014; Parikshak et al., 2013). Notably, these chromatin modifier genes harboring missense and protein disrupting *de novo* mutations were dosage-sensitive, reflecting the heterozygous state observed in ASD probands (Parikshak et al., 2013). In addition, some highly-confidence ASD-associated genes converged in the development of layer 5/6 glutamatergic cortical neuron projection (Parikshak et al., 2013; Willsey et al., 2013). These disrupted cortical cytoarchitecture may lead to behavioral abnormalities implicated in ASD models (G. B. Choi et al., 2016). Indeed, the disorganized cortical microstructure was reported in the postmortem brain samples from children with ASD (Stoner et al., 2014).

Transcriptome analysis in ASD brains revealed a down-regulation in genes of extended length (>100 kb) (King et al., 2013). These long genes, which often encode synaptic proteins and ion channels, are more susceptible to changes in the transcriptional machinery (Sullivan et al.,

2019). For example, Zhao et al. identified a set of long genes containing broad enhancer-like-chromatin domains (BELDs) associated with high transcriptional activities and frequent chromatin interactions. These BELDs genes are enriched in synaptic genes during late fetal and early postnatal development (Parikshak et al., 2013; Zhao et al., 2018). Moreover, Wen et al. created a tool called Gene Set Enrichment Analysis (GSEA), which combined the gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) database, to classify significantly enriched or depleted ASD candidate genes. They found a substantial overlap of GO and KEGG terms in genes involved in the calcium and MAPK signaling pathways (Wen et al., 2016).

1.3 Modeling disease-associated missense mutations

1.3.1 Challenges for studying the functional consequences of missense variants

Decades of GWAS and NGS studies produced a vast amount of genetic variation implicated in human diseases with complex traits. On average, a healthy individual carries 250-300 disease-associated variants, and each variant can be found as heterozygous or homozygous state (Cooper et al., 2013; The 1000 Genomes Project Consortium, 2010). Overall, 4.6 million missense variants have been identified in about 140,000 individuals yet only 2% of the missense variants have a clinical annotation in the ClinVar database (Landrum et al., 2014; Lek et al., 2016). Unfortunately, over half of the clinical interpreted variants showed unknown or controversial clinical outcomes, so-called variants of uncertain significance (VUS) (Starita et al., 2017). As a result, spotting the subtle functional deteriorating variants from the immense pool of missense variants has been an enormous challenge.

1.3.2 Prioritizing missense variants using computational inference tools

Since the completion of Human Genome Project (HGP) at the beginning of the 21st century, scientists have dedicated countless efforts to decipher the genomic codes and distinguish the disease-causing mutations in the human genome. The early studies focused on the allele frequency and how these haplotype blocks changed the human population structures (Belmont et al., 2005; The 1000 Genomes Project Consortium, 2012). Recent approaches analyzed the data from the Exome Sequencing Project (ESP) and calculated a "residual variation intolerance score" (RVIS) by comparing the number of common missense and truncated variants observed against the number of all variants observed, including the synonymous and rare variants (Petrovski et al., 2013). This approach used the excess of rare versus common missense variation within the human genome to highlight functional significant intolerant genes. Samocha et al. captured the functional deteriorating missense variants by comparing the rare missense variants with the enumerating set of all possible variants in each gene (Samocha et al., 2014). The revolutionary concept of selective "constraint" proved to be a powerful tool for identifying the region in a gene where *de novo* variants are more likely to be functional impactful (Samocha et al., 2017).

In addition to identifying the functional deteriorating genes and regions (Samocha et al., 2017; Silk et al., 2019), several computational inference tools are available for predicting the pathogenicity of a given missense variant. One of the popular features is sequence conservation. These methods, such as SIFT, compare protein multiple alignments and prioritize residues based on the assumption that amino acid substitutions in highly conserved regions are more likely to be damaging (Ng & Henikoff, 2003). Algorithms in this category have reasonable predictive values for loss-of-function missense variants but they are not accurate in predicting gain-of-function or edgetic variants (Yi et al., 2017). Another type of interference tool integrates the three-dimension protein structure with sequence information. These algorithms (e.g., PolyPhen-2) assume missense mutations located in the protein active sites or

specific structural regions are more likely to affect the protein functions (Adzhubei et al., 2010). Most algorithms in this category predict mutational effect only for mutations in the protein-coding genes (Yi et al., 2017). The third type of computational inference tools explore the function of missense variants in regulatory regions and their effect on target genes' interaction. Adapting from the experimental data in the Encyclopedia of DNA Elements (ENCODE) project, these tools, such as CADD, consider the genetic and epigenetic regulation to prioritize the pathogenic missense variants (Dunham et al., 2012; Kircher et al., 2014). Furthermore, the fourth type of computational inference tools takes the gene network and pathway features into consideration. These algorithms, such as SuSPect and Eigen, integrate sequence and structural information with gene ontology similarity and highlight genes with high topological centrality (Ionita-Laza et al., 2016; Yates et al., 2014).

Despite considerable efforts, computational inference tools' performance is limited by the existing experimental training data (Grimm et al., 2015; Miosge et al., 2015). Few studies evaluated the performance inference algorithms and estimated a 30% false prediction rate due to the proximity with common SNPs, sequencing errors, or lack of evidence of pathogenicity (Bell et al., 2011; S. Sun et al., 2016). In most cases, *in silico* methods tend to overpredict pathogenic missense changes and give low specificity (Y. Choi et al., 2012), and the accuracy is even less reliable in missense variants with milder effects (Richards et al., 2015). The poor performance and lack of consistency hampered these computational inference tools' credibility and presented a need for direct evidence using functional assays in the biological systems (Richards et al., 2015; Starita et al., 2017).

1.3.3 Characterizing missense variants in system-level experimental platforms

A couple of high throughput functional assays use cell-based systems to study the function of variants in a comprehensive manner. For example, Gespirini et al. established a

multiplexed functional assays for variant effect (MAVEs) using yeast to directly study each missense variant to its impact on protein function, regulation, splicing, and RNA stability (Gasperini et al., 2016). Chen et al. detected protein interaction-disrupting *de novo* missense variants using the yeast two-hybrid system. They discovered that the ASD-associated missense variants were more likely to affect proteins with many interactions (i.e., hub genes), and their interacting partners tend to be ASD risk genes (Chen et al., 2018). Generally speaking, the *in vitro* approach enables high throughput screening at the scale of 10^4 -10^6 variants per experiment, and the results are readily comparable to each other (Gasperini et al., 2016). However, these cell-based functional assays take proteins out of their endogenous genomic and cellular contexts, and this may lead to neglection of the effect caused by cell-type-specific chromatin states (Ernst & Kellis, 2015; Starita et al., 2017).

The advance in gene-editing technology enables disease-associated genetic variants to be generated and tested in their endogenous genomic and cellular contexts. For example, Miosge et al. generated missense mutant mice using the N-ethyl-N-nitrosourea (ENU) mutagenesis method and compared the predicting and actual deleterious effects in immune functions (Miosge et al., 2015). Recently, few groups utilized the lentiviral library to introduce mutations in disease-associated genes and studied the cell-type effects *in vivo*. With this in mind, Jin et al. introduced frameshift mutations in 35 ASD risk genes *in utero* and performed postnatal single-cell transcriptomic analysis (Perturb-Seq) in the offspring. They described cell-type-specific functional impact in neural development consistent with single-cell data from ASD patients (Jin et al., 2020). Similarly, Wertz et al. conducted the first unbiased genome-wide knock-down screening in the mammalian central nervous system by injecting short hairpin RNA (shRNA) directly to the striatum of the mouse brains, and they discovered genes and cellular processes essentials in neurons and associated novel druggable candidate genes in neurological disorders (Wertz et al., 2020).

The genome-wide screening in rodent models allows studying the disease candidate gene in its endogenous context. However, each missense variant can be found in the heterozygous or homozygous state, and there are effectively infinite combinations between missense variants. Generating all disease-associated missense changes in rodent models is not practical. As a result, many research groups, including our laboratory, examined the functional consequences of human variants by knocking-in the equivalent amino acid substitution into an orthologous gene in *Caenorhabditis elegans (C. elegans)*. These studies analyzed the phenotype and genetic properties of candidate genes implicated in human diseases (Bai et al., 2020; S. Kim et al., 2017a; Levitan et al., 1996), especially in ASD (Buddell et al., 2019; Gonzalez-cavazos et al., 2019; Lipstein et al., 2017; McDiarmid et al., 2018, 2020; Post et al., 2020; Wong et al., 2019). In particular, Wong et al. systematically evaluated the effects of 20 missense alleles that were predicted to be phenotype altering and found that only 70% of them displayed phenotypic changes in morphology, locomotion, and fecundity (Wong et al., 2019). McDiarmid et al. quantified tactile sensitivity and habituation learning phenotypes in ASD-associated missense mutants and found the phenotypic defect can be rescued by adult neuroligin re-expression (McDiarmid et al., 2020).

1.3.4 The advantages of studying missense variants in C. elegans

C. elegans has proven to be an advantageous system to model disease-associated human genetic variants. It enhances our understandings in disease gene functions, the consequences of genetic variants, and embedding genes into functional pathways (www.wormbase.org; www.wormbook.org). Compared to other model organisms, the *C. elegans* nervous system is highly amenable to cellular level resolution analysis of neural circuits and behavior given its numerically simple nervous system, known connectome, and relatively weak dependence on the nervous system for organism viability. Additionally, *C. elegans* share 60-80% orthologous genes

as the human genome and functional conservation in essential signaling pathways (Kaletta & Hengartner, 2006; Shaye & Greenwald, 2011). Over the decades, *C. elegans* research has facilitated our understanding of many human diseases, including the discovery of presenilin and its druggable targets in human Alzheimer's disease (Levitan et al., 1996; Levitan & Greenwaldh, 1995; Strooper et al., 1999; Sundaram & Greenwald, 1993), the mechanism of action of the antidepressant fluoxetine (Ranganathan et al., 2001), and negative regulators of the insulin signaling pathway in diabetes research (Ogg et al., 1997). Moreover, *C. elegans* also accelerates the screening of functional variance of unknown significance (VUSs) with the readily accessible gene editing techniques and the relatively short life cycle (Brenner, 1974; Engleman et al., 2016; S. Kim et al., 2017b). Phenotypic analyses of transgenic animals are routinely done to determine the mechanism of disease-associated mutations. Recent advances in gene editing technology, like CRISPR/Cas9 system, further expedite the causality studies of variants with large quantities, such as *de novo* missense variants. The precise gene editing by CRISPR in *C. elegans* genome avoids the confounding effects caused by mosaic overexpression of extrachromosomal arrays or specialized genetic backgrounds in the traditional methods (Chiu et al., 2013). Therefore, the *C. elegans* model presents an opportunity to determine subtle effects of missense variants and impede the use of these transgenic strains in large-scale drug screens (Kaletta & Hengartner, 2006).

1.4 Thesis chapter summary

I joined the Ph.D. program at Caltech with the intention to study the molecular mechanisms underlying ASD. I was fortunate to meet Paul, who happened to have a grant matching my research interest. We came up with a plan to study ASD genetics using the *C. elegans* model. I aim to investigate the functional consequences of ASD-associated missense

variants in different physiological functions, especially in sensory modulation. In this Ph.D. dissertation, I present my findings in several papers and manuscripts, and these papers are summarized as follow.

In Chapter 1, I review the history, symptoms, and etiology of ASD. I summarize the current progress in *in silico*, *in vitro*, and *in vivo* methods for prioritizing ASD-associated genetic variants. The growing number of genetic variants implicated in ASD proposes a niche for a more accurate and efficient method to identify missense variants with functional consequences and prioritize them for future functional studies in other models. *C. elegans*, with its short lifespan and readily accessible genome, offers a powerful tool to model human disease and facilitate the screening process.

Chapter 2 presents a paper published in Human Molecular Genetics in 2019. This study describes our working pipeline for screening functional impactful missense alleles implicated in ASD. We identified highly conserved human ASD-associated missense variants in their *C. elegans* orthologs, used a CRISPR/Cas9-mediated homology-directed knock-in strategy to generate 20 missense mutants, and analyzed their impact on behaviors and development via several broad-spectrum assays. Additionally, we compared our screening results to existing computational inference tools and found only 70% of them showed detectable phenotypic changes in morphology, locomotion, or fecundity. This paper points out a need for *in vivo* screening platform and prioritizes 14 phenotypic changing missense variants in 11 human genes.

In Chapter 3, I present a manuscript on the second half of the screening project. This paper expands on the previous collection and characterizes 28 more ASD-associated missense variants, focusing on genes involved in synaptic functions and gene regulation. We use morphology, fecundity, and an automated locomotor tracking system to examine the changes in neurodevelopment. We identify 18 of the 28 (64%) missense variants that exhibit phenotypic

changes. Specifically, missense variants in the gene regulatory cluster are more likely to influence fecundity whereas missense variants in the synaptic gene cluster impact morphology and locomotor patterns. This study proves that missense mutants in different gene clusters display distinct phenotype profiles and lay the foundation for future systematic characterization of disease-associated variants.

Chapter 4 presents a micropublication on one missense variant of an ASD risk gene, ALDH1A3. The ALDH1A3 is the key enzyme in the retinoic acid signaling pathway, and missense mutations in *ALDH1A3* have been identified in family studies of ASD and other developmental disorders. However, there has been no evidence from animal models that verify the functional consequence of missense mutations in *ALDH1A3*. Therefore, we introduce the equivalent of the *ALDH1A3* C174Y variant into the *C. elegans* ortholog, *alh-1*, at the corresponding locus. Mutant animals with this missense mutation exhibit decreased fecundity by 50% compared to wild-type animals, indicating disrupted protein function. To our knowledge, this is the first ALDH1A3 C174Y missense model, which may be used to elucidate the effects of ALDH1A3 C174Y missense mutation in the retinoic acid signaling pathway during development. This manuscript is accepted and published on microPublication Biology in January 2021.

In Chapter 5, I present a manuscript on the functional characterization of *BRAF*, one of the ASD risk genes. BRAF is a critical kinase linking canonical NMDAR signaling to the MEK-ERK cascade at the synapse. The role of BRAF in sensory modulation has been demonstrated in both human and animal studies. In this study, we inspect the sensory function of three ASD-associated missense variants in BRAF. We use a previously established platform to identify conserved missense variants and introduced them in the orthologous gene in *C. elegans*. We then characterize the functional consequences of the missense variants using the chemotaxis assay in *C. elegans*. Our results reveal a conserved missense residue orthologous to human BRAF K499N

displays a hypersensitive phenotype in the diacetyl chemotaxis assay. This phenotype is not dominant but can be inhibited by *lin-45* RNAi. We also check the expression of the lin-45 protein and confirm the strong expression of *lin-45* in the RIC interneurons. Overall, this study suggested a potential gain-of-function allele in the orthologous gene of *BRAF*, especially in the sensory function.

In Chapter 6, I give a general discussion of my Ph.D. project. I summarize the essential findings and significance of these studies. I point out the advantages and limitations of using *C. elegans* as a model to study ASD. Finally, I reflect on my Ph.D. journey and gives out some advice for my past self and the next-generation researchers who plan to embark on this ASD project or projects with a similar approach.

1.5 References

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., Kondrashov, A. S., & Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nature Methods*, *7*(4), 248–249. https://doi.org/10.1038/nmeth0410-248

Alonso-Gonzalez, A., Rodriguez-Fontenla, C., & Carracedo, A. (2018). De novo mutations (DNMs) in autism spectrum disorder (ASD): Pathway and network analysis. *Frontiers in Genetics*, *9*(SEP). https://doi.org/10.3389/fgene.2018.00406

American Psychiatric Association. (2013). *Dianostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Publishing.

Anney, R., Klei, L., Pinto, D., Regan, R., Conroy, J., Magalhaes, T. R., Correia, C., Abrahams, B. S., Sykes, N., Pagnamenta, A. T., Almeida, J., Bacchelli, E., Bailey, A. J., Baird, G., Battaglia, A., Berney, T., Bolshakova, N., Bölte, S., Bolton, P. F., … Hallmayer, J. (2010). A genome-wide scan for common alleles affecting risk for autism. *Human Molecular*

Genetics, *19*(20), 4072–4082. https://doi.org/10.1093/hmg/ddq307

Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium. (2017). Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Molecular Autism*, *8*, 21. https://doi.org/10.1186/s13229-017-0137-9

Bai, X., Bouffard, J., Lord, A., Brugman, K., Sternberg, P. W., Cram, E. J., & Golden, A. (2020). Caenorhabditis elegans piezo channel coordinates multiple reproductive tissues to govern ovulation. *ELife*, *9*, 1–35. https://doi.org/10.7554/eLife.53603

Bailey, A., Le Couteur, A., Gottesman, I., Bolton, P., Simonoff, E., Yuzda, E., & Rutter, M. (1995). Autism as a strongly genetic disorder: evidence from a British twin study. *Psychological Medicine*, *25*, 63–77.

Baxter, A. J., Brugha, T. S., Erskine, H. E., Scheurer, R. W., Vos, T., & Scott, J. G. (2015). The epidemiology and global burden of autism spectrum disorders. *Psychological Medicine*, *45*(3), 601–613. https://doi.org/10.1017/S003329171400172X

Bell, C. J., Dinwiddie, D. L., Miller, N. A., Hateley, S. L., Ganusova, E. E., Mudge, J., Langley, R. J., Zhang, L., Lee, C. C., Schilkey, F. D., Sheth, V., Woodward, J. E., Peckham, H. E., Schroth, G. P., Kim, R. W., & Kingsmore, S. F. (2011). Carrier testing for severe childhood recessive diseases by next-generation sequencing. *Science Translational Medicine*, *3*(65). https://doi.org/10.1126/scitranslmed.3001756

Belmont, J. W., Boudreau, A., Leal, S. M., Hardenbol, P., Pasternak, S., Wheeler, D. A., Willis, T. D., Yu, F., Yang, H., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., … Stewart, J. (2005). A haplotype map of the human genome. *Nature*, *437*(7063), 1299–1320. https://doi.org/10.1038/nature04226

Ben-David, E., & Shifman, S. (2013). Combined analysis of exome sequencing points toward a major role for transcription regulation during brain development in autism. *Molecular*

Psychiatry, *18*(10), 1054–1056. https://doi.org/10.1038/mp.2012.148

Brenner, S. (1974). The genetics of Caenorhabditis elegans. *Genetics*, *77*(May), 71–94. https://doi.org/10.1111/j.1749-6632.1999.tb07894.x

Buddell, T., Friedman, V., Drozd, C. J., & Quinn, C. C. (2019). An autism-causing calcium channel variant functions with selective autophagy to alter axon targeting and behavior. *PLoS Genetics*, *15*(12), 1–20. https://doi.org/10.1371/journal.pgen.1008488

Charman, T., & Gotham, K. (2013). Measurement Issues: Screening and diagnostic instruments for autism spectrum disorders - lessons from research and practise. *Child and Adolescent Mental Health*, *18*(1), 52–63. https://doi.org/10.1111/j.1475-3588.2012.00664.x

Chaste, P., Roeder, K., & Devlin, B. (2017). The Yin and Yang of Autism Genetics: How Rare de Novo and Common Variations Affect Liability. *Annual Review of Genomics and Human Genetics*, *18*, 167–187. https://doi.org/10.1146/annurev-genom-083115-022647

Chen, S., Fragoza, R., Klei, L., Liu, Y., Wang, J., Roeder, K., Devlin, B., & Yu, H. (2018). An interactome perturbation framework prioritizes damaging missense mutations for developmental disorders. *Nature Genetics*, *50*, 1032–1040. https://doi.org/10.1038/s41588-018-0130-z

Chiu, H., Schwartz, H. T., Antoshechkin, I., & Sternberg, P. W. (2013). Transgene-free genome editing in Caenorhabditis elegans using CRISPR-Cas. *Genetics*, *195*(3), 1167–1171. https://doi.org/10.1534/genetics.113.155879

Choi, G. B., Yim, Y. S., Wong, H., Kim, S., Kim, H., Kim, S. V., Hoeffer, C. A., Littman, D. R., & Huh, J. R. (2016). The maternal interleukin-17a pathway in mice promotes autism-like phenotypes in offspring. *Science*, *351*(6276), 933–939. https://doi.org/10.1126/science.aad0314

Choi, Y., Sims, G. E., Murphy, S., Miller, J. R., & Chan, A. P. (2012). Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLoS ONE*, *7*(10).

https://doi.org/10.1371/journal.pone.0046688

Chung, B. H. Y., Tao, V. Q., & Tso, W. W. Y. (2014). Copy number variation and autism: New insights and clinical implications. *Journal of the Formosan Medical Association*, *113*(7), 400–408. https://doi.org/10.1016/j.jfma.2013.01.005

Coe, B. P., Stessman, H. A. F., Sulovari, A., Geisheker, M. R., Bakken, T. E., Lake, A. M., Dougherty, J. D., Lein, E. S., Hormozdiari, F., Bernier, R. A., & Eichler, E. E. (2019). Neurodevelopmental disease genes implicated by de novo mutation and copy number variation morbidity. *Nature Genetics*, *51*(1), 106–116. https://doi.org/10.1038/s41588-018-0288-4

Conrad, D. F., Keebler, J. E. M., Depristo, M. A., Lindsay, S. J., Zhang, Y., Casals, F., Idaghdour, Y., Hartl, C. L., Torroja, C., Garimella, K. V., Zilversmit, M., Cartwright, R., Rouleau, G. A., Daly, M., Stone, E. A., Hurles, M. E., & Awadalla, P. (2011). Variation in genome-wide mutation rates within and between human families. *Nature Genetics*, *43*(7), 712–714. https://doi.org/10.1038/ng.862

Cooper, D. N., Krawczak, M., Polychronakos, C., Tyler-Smith, C., & Kehrer-Sawatzki, H. (2013). Where genotype is not predictive of phenotype: Towards an understanding of the molecular basis of reduced penetrance in human inherited disease. In *Human Genetics* (Vol. 132, Issue 10). https://doi.org/10.1007/s00439-013-1331-2

De Rubeis, S., & Buxbaum, J. D. (2015). Genetics and genomics of autism spectrum disorder: Embracing complexity. *Human Molecular Genetics*, *24*(R1), R24–R31. https://doi.org/10.1093/hmg/ddv273

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

https://doi.org/10.1038/nature13772

Doshi-Velez, F., Ge, Y., & Kohane, I. (2014). Comorbidity clusters in autism spectrum disorders: An electronic health record time-series analysis. *Pediatrics*, *133*(1). https://doi.org/10.1542/peds.2013-0819

Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., Doyle, F., Epstein, C. B., Frietze, S., Harrow, J., Kaul, R., Khatun, J., Lajoie, B. R., Landt, S. G., Lee, B. K., Pauli, F., Rosenbloom, K. R., Sabo, P., Safi, A., Sanyal, A., … Lochovsky, L. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, *489*(7414), 57–74. https://doi.org/10.1038/nature11247

Engleman, E. A., Katner, S. N., & Neal-beliveau, B. S. (2016). Caenorhabditis elegans as a model to study the molecular and genetic mechanisms of drug addiction. *Prog Mol Biol Transl Sci*, *137*, 229–252. https://doi.org/10.1016/bs.pmbts.2015.10.019.Caenorhabditis

Ernst, J., & Kellis, M. (2015). Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. *Nature Biotechnology*, *33*(4), 364–376. https://doi.org/10.1038/nbt.3157

Fischbach, G. D., & Lord, C. (2010). The simons simplex collection: A resource for identification of autism genetic risk factors. *Neuron*, *68*(2), 192–195. https://doi.org/10.1016/j.neuron.2010.10.006

Gasperini, M., Starita, L., & Shendure, J. (2016). The power of multiplexed functional analysis of genetic variants. *Nature Protocols*, *11*(10), 1782–1787. https://doi.org/10.1038/nprot.2016.135

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

https://doi.org/10.1038/ng.3039

Geschwind, D. H., & State, M. W. (2015). Gene hunting in autism spectrum disorder: on the path to precision medicine. *Lancet Neurol*, *14*(11), 1109–1120. https://doi.org/10.1016/S1474-4422(15)00044-7.Gene

Gilman, S. R., Iossifov, I., Levy, D., Ronemus, M., Wigler, M., & Vitkup, D. (2011). Rare De Novo Variants Associated with Autism Implicate a Large Functional Network of Genes Involved in Formation and Function of Synapses. *Neuron*, *70*(5), 898–907. https://doi.org/10.1016/j.neuron.2011.05.021

Gonzalez-cavazos, C., Cao, M., Wong, W., Chai, C., & Sternberg, P. W. (2019). Effects of ASD-associated daf-18/PTEN missense variants on C. elegans dauer development. *MicroPubl Biol*, *10*, 17912.

Grimm, D. G., Azencott, C. A., Aicheler, F., Gieraths, U., Macarthur, D. G., Samocha, K. E., Cooper, D. N., Stenson, P. D., Daly, M. J., Smoller, J. W., Duncan, L. E., & Borgwardt, K. M. (2015). The evaluation of tools used to predict the impact of missense variants is hindered by two types of circularity. *Human Mutation*, *36*(5), 513–523. https://doi.org/10.1002/humu.22768

Grove, J., Ripke, S., Als, T. D., Mattheisen, M., Walters, R. K., Won, H., Pallesen, J., Agerbo, E., Andreassen, O. A., Anney, R., Awashti, S., Belliveau, R., Bettella, F., Buxbaum, J. D., Bybjerg-Grauholm, J., Bækvad-Hansen, M., Cerrato, F., Chambert, K., Christensen, J. H., … Børglum, A. D. (2019). Identification of common genetic risk variants for autism spectrum disorder. *Nature Genetics*, *51*(3), 431–444. https://doi.org/10.1038/s41588-019-0344-8

Hazlett, H. C., Gu, H., Munsell, B. C., Kim, S. H., Styner, M., Wolff, J. J., Elison, J. T., Swanson, M. R., Zhu, H., Botteron, K. N., Collins, D. L., Constantino, J. N., Dager, S. R., Estes, A. M., Evans, A. C., Fonov, V. S., Gerig, G., Kostopoulos, P., McKinstry, R. C., … Piven, J.

(2017). Early brain development in infants at high risk for autism spectrum disorder. *Nature*, *542*(7641), 348–351. https://doi.org/10.1038/nature21369

Iafrate, A. J., Feuk, L., Rivera, M. N., Listewnik, M. L., Donahoe, P. K., Qi, Y., Scherer, S. W., & Lee, C. (2004). Detection of large-scale variation in the human genome. *Nature Genetics*, *36*(9), 949–951. https://doi.org/10.1038/ng1416

Ionita-Laza, I., Mccallum, K., Xu, B., & Buxbaum, J. D. (2016). A spectral approach integrating functional genomic annotations for coding and noncoding variants. *Nature Genetics*, *48*(2), 214–220. https://doi.org/10.1038/ng.3477

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

Iossifov, I., Ronemus, M., Levy, D., Wang, Z., Hakker, I., Rosenbaum, J., Yamrom, B., Lee, Y. H., Narzisi, G., Leotta, A., Kendall, J., Grabowska, E., Ma, B., Marks, S., Rodgers, L., Stepansky, A., Troge, J., Andrews, P., Bekritsky, M., … Wigler, M. (2012). De novo gene disruptions in children on the autistic spectrum. *Neuron*, *74*(2), 285–299. https://doi.org/10.1016/j.neuron.2012.04.009

Jin, X., Simmons, S. K., Guo, A. X., Shetty, A. S., Ko, M., Nguyen, L., Robinson, E., Oyler, P., Curry, N., Deangeli, G., Lodato, S., Levin, J. Z., Regev, A., Zhang, F., & Arlotta, P. (2020). In vivo Perturb-Seq reveals neuronal and glial abnormalities associated with Autism risk genes. *Science*, *370*, eaaz6063. https://doi.org/10.1101/791525

Kaletta, T., & Hengartner, M. O. (2006). Finding function in novel targets: C. elegans as a model organism. *Nature Reviews Drug Discovery*, *5*(5), 387–399. https://doi.org/10.1038/nrd2031

Kanner, L. (1943). Autistic disturbances of affective contact. *Nerv. Child*, *2*, 217–250.

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

Kenny, E. M., Cormican, P., Furlong, S., Heron, E., Kenny, G., Fahey, C., Kelleher, E., Ennis, S., Tropea, D., Anney, R., Corvin, A. P., Donohoe, G., Gallagher, L., Gill, M., & Morris, D. W. (2014). Excess of rare novel loss-of-function variants in synaptic genes in schizophrenia and autism spectrum disorders. *Molecular Psychiatry*, *19*(8), 872–879. https://doi.org/10.1038/mp.2013.127

Kim, S., Twigg, S. R. F., Scanlon, V. A., Chandra, A., Hansen, T. J., Alsubait, A., Fenwick, A. L., McGowan, S. J., Lord, H., Lester, T., Sweeney, E., Weber, A., Cox, H., Wilkie, A. O. M., Golden, A., & Corsi, A. K. (2017a). Localized TWIST1 and TWIST2 basic domain substitutions cause four distinct human diseases that can be modeled in Caenorhabditis elegans. *Human Molecular Genetics*, *26*(11), 2118–2132. https://doi.org/10.1093/hmg/ddx107

Kim, S., Twigg, S. R. F., Scanlon, V. A., Chandra, A., Hansen, T. J., Alsubait, A., Fenwick, A. L., McGowan, S. J., Lord, H., Lester, T., Sweeney, E., Weber, A., Cox, H., Wilkie, A. O. M., Golden, A., & Corsi, A. K. (2017b). Localized TWIST1 and TWIST2 basic domain substitutions cause four distinct human diseases that can be modeled in Caenorhabditis elegans. *Human Molecular Genetics*, *26*(11), 2118–2132. https://doi.org/10.1093/hmg/ddx107

Kim, Y. S., & Leventhal, B. L. (2015). Genetic Epidemiology and Insights into Interactive Genetic and Environmental Effects in Autism Spectrum Disorders. *Biological Psychiatry*, *77*(1), 66–74. https://doi.org/10.1016/j.biopsych.2014.11.001

King, I. F., Yandava, C. N., Mabb, A. M., Hsiao, J. S., Huang, H. S., Pearson, B. L., Calabrese, J. M., Starmer, J., Parker, J. S., Magnuson, T., Chamberlain, S. J., Philpot, B. D., & Zylka, M. J. (2013). Topoisomerases facilitate transcription of long genes linked to autism. *Nature*, *501*(7465), 58–62. https://doi.org/10.1038/nature12504

Kircher, M., Witten, D. M., Jain, P., Roak, B. J. O., Cooper, G. M., & Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics*, *46*(3), 310–315. https://doi.org/10.1038/ng.2892

Lai, M. C., Lombardo, M. V., & Baron-Cohen, S. (2014). Autism. *The Lancet*, *383*(9920), 896–910. https://doi.org/10.1016/S0140-6736(13)61539-1

Landrum, M. J., Lee, J. M., Riley, G. R., Jang, W., Rubinstein, W. S., Church, D. M., & Maglott, D. R. (2014). ClinVar: Public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Research*, *42*(D1), 980–985. https://doi.org/10.1093/nar/gkt1113

Lau, W. K. W., Leung, M. K., & Lau, B. W. M. (2019). Resting-state abnormalities in Autism Spectrum Disorders: A meta-analysis. *Scientific Reports*, *9*(1), 1–8. https://doi.org/10.1038/s41598-019-40427-7

Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., O'Donnell-Luria, A. H., Ware, J. S., Hill, A. J., Cummings, B. B., Tukiainen, T., Birnbaum, D. P., Kosmicki, J. A., Duncan, L. E., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J., … Williams, A. L. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, *536*(7616), 285–291. https://doi.org/10.1038/nature19057

Levitan, D., Doyle, T. G., Brousseau, D., Lee, M. K., Thinakaran, G., Slunt, H. H., Sisodia, S. S., & Greenwald, I. (1996). Assessment of normal and mutant human presenilin function in Caenorhabditis elegans. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(25), 14940–14944. https://doi.org/10.1073/pnas.93.25.14940

Levitan, D., & Greenwaldh, I. (1995). Facilitation of lin-12-mediated signalling by sel-12, a

Caenorhabditis elegans S182 Alzheimer's disease gene. *Nature*, *377*, 351–354.

Lipstein, N., Verhoeven-Duif, N. M., Michelassi, F. E., Calloway, N., Van Hasselt, P. M., Pienkowska, K., Van Haaften, G., Van Haelst, M. M., Van Empelen, R., Cuppen, I., Van Teeseling, H. C., Evelein, A. M. V., Vorstman, J. A., Thoms, S., Jahn, O., Duran, K. J., Monroe, G. R., Ryan, T. A., Taschenberger, H., … Brose, N. (2017). Synaptic UNC13A protein variant causes increased neurotransmission and dyskinetic movement disorder. *Journal of Clinical Investigation*, *127*(3), 1005–1018. https://doi.org/10.1172/JCI90259

Lord, C, Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord*, *24*(5), 659–685.

Lord, Catherine, Rutter, M., DiLavore, P. C., & Risi, S. (2012). *Autism diagnostic observation schedule* (2nd ed.). Western Psychological Corporation.

Ma, D., Salyakina, D., Jaworski, J. M., Konidari, I., Whitehead, P. L., Andersen, A. N., Hoffman, J. D., Slifer, S. H., Hedges, D. J., Cukier, H. N., Griswold, A. J., McCauley, J. L., Beecham, G. W., Wright, H. H., Abramson, R. K., Martin, E. R., Hussman, J. P., Gilbert, J. R., Cuccaro, M. L., … Pericak-Vance, M. a. (2009). A genome-wide association study of autism reveals a common novel risk locus at 5p14.1. *Annals of Human Genetics*, *73*(3), 263–273. https://doi.org/10.1111/j.1469-1809.2009.00523.x

Maenner, M., Shaw, K., Baio, J., Washington, A., Patrick, M., DiRienzo, M., Christensen, D. L., Wiggins, L. D., Pettygrove, S., Andrews, J. G., Maya Lopez, M., Hudson, A., Baroud, T., Schwenk, Y., White, T., Rosenberg, C. R., Lee, L.-C., Harrington, R. A., Huston, M., … Dietz, P. M. (2020). Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2016. *MMWR Surveill Summ*, *69*(No. SS-4), 1–12.

McCammon, J. M., & Sive, H. (2015). Challenges in understanding psychiatric disorders and

developing therapeutics: A role for zebrafish. *Disease Models and Mechanisms*, *8*, 647–656. https://doi.org/10.1242/dmm.019620

McDiarmid, T. A., Au, V., Loewen, A. D., Liang, J., Mizumoto, K., Moerman, D. G., & Rankin, C. H. (2018). CRISPR-Cas9 human gene replacement and phenomic characterization in *Caenorhabditis elegans* to understand the functional conservation of human genes and decipher variants of uncertain significance. *Disease Models & Mechanisms*, dmm.036517. https://doi.org/10.1242/dmm.036517

McDiarmid, T. A., Belmadani, M., Liang, J., Meili, F., Mathews, E. A., Mullen, G. P., Hendi, A., Wong, W. R., Rand, J. B., Mizumoto, K., Haas, K., Pavlidis, P., & Rankin, C. H. (2020). Systematic phenomics analysis of autism-associated genes reveals parallel networks underlying reversible impairments in habituation. *P Natl Acad Sci USA*, *117*(1), 656–667. https://doi.org/10.1073/pnas.1912049116

Minshew, N. J., & Keller, T. A. (2010). The nature of brain dysfunction in autism: Functional brain imaging studies. *Current Opinion in Neurology*, *23*(2), 124–130. https://doi.org/10.1097/WCO.0b013e32833782d4

Miodovnik, A., Harstad, E., Sideridis, G., & Huntington, N. (2015). Timing of the diagnosis of attention-deficit/hyperactivity disorder and autism spectrum disorder. *Pediatrics*, *136*(4), e830–e837. https://doi.org/10.1542/peds.2015-1502

Miosge, L. A., Field, M. A., Sontani, Y., Cho, V., Johnson, S., Palkova, A., Balakishnan, B., Liang, R., Zhang, Y., Lyon, S., Beutler, B., Whittle, B., Bertram, E. M., Enders, A., Goodnow, C. C., & Andrews, T. D. (2015). Comparison of predicted and actual consequences of missense mutations. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(37), E5189–E5198. https://doi.org/10.1073/pnas.1511585112

Modabbernia, A., Velthorst, E., & Reichenberg, A. (2017). Environmental risk factors for

autism: an evidence-based review of systematic reviews and meta-analyses. *Molecular Autism*, *8*(1), 1–16. https://doi.org/10.1186/s13229-017-0121-4

Neale, B. M., Kou, Y., Liu, L., Ma'ayan, A., Samocha, K. E., Sabo, A., Lin, C.-F., Stevens, C., Wang, L.-S., Makarov, V., Polak, P., Yoon, S., Maguire, J., Crawford, E. L., Campbell, N. G., Geller, E. T., Valladares, O., Schafer, C., Liu, H., … Daly, M. J. (2012). Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature*, *485*(7397), 242–245. https://doi.org/10.1038/nature11011

Ng, P. C., & Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Research*, *31*(13), 3812–3814. https://doi.org/10.1093/nar/gkg509

O'Roak, B. J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B. P., Levy, R., Ko, A., Lee, C., Smith, J. D., Turner, E. H., Stanaway, I. B., Vernot, B., Malig, M., Baker, C., Reilly, B., Akey, J. M., Borenstein, E., Rieder, M. J., … Eichler, E. E. (2012). Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature*, *485*(7397), 246–250. https://doi.org/10.1038/nature10989

Ogg, S., Paradis, S., Gottlieb, S., Patterson, G. I., Lee, L., Tissenbaum, H. A., & Ruvkun, G. (1997). The fork head transcription factor DAF-16 transduces insulin-like metabolic and longevity signals in C. elegans. *Nature*, *389*(6654), 994–999. https://doi.org/10.1038/40194

Ozonoff, S., Young, G. S., Carter, A., Messinger, D., Yirmiya, N., Zwaigenbaum, L., Bryson, S., Carver, L. J., Constantino, J. N., Dobkins, K., Hutman, T., Iverson, J. M., Landa, R., Rogers, S. J., Sigman, M., & Stone, W. L. (2011). Recurrence risk for autism spectrum disorders: A baby siblings research consortium study. *Pediatrics*, *128*(3). https://doi.org/10.1542/peds.2010-2825

Palmer, N., Beam, A., Agniel, D., Eran, A., Manrai, A., Spettell, C., Steinberg, G., Mandl, K., Fox, K., Nelson, S. F., & Kohane, I. (2017). Association of sex with recurrence of autism spectrum disorder among siblings. *JAMA Pediatrics*, *171*(11), 1107–1112.

https://doi.org/10.1001/jamapediatrics.2017.2832

Parikshak, N. N., Luo, R., Zhang, A., Won, H., Lowe, J. K., Chandran, V., Horvath, S., & Geschwind, D. H. (2013). Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell*, *155*(5), 1008. https://doi.org/10.1016/j.cell.2013.10.031

Petrovski, S., Wang, Q., Heinzen, E. L., Allen, A. S., & Goldstein, D. B. (2013). Genic Intolerance to Functional Variation and the Interpretation of Personal Genomes. *PLoS Genetics*, *9*(8). https://doi.org/10.1371/journal.pgen.1003709

Pinto, D., Delaby, E., Merico, D., Barbosa, M., Merikangas, A., Klei, L., Thiruvahindrapuram, B., Xu, X., Ziman, R., Wang, Z., Vorstman, J. A. S., Thompson, A., Regan, R., Pilorge, M., Pellecchia, G., Pagnamenta, A. T., Oliveira, B., Marshall, C. R., Magalhaes, T. R., … Scherer, S. W. (2014). Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *American Journal of Human Genetics*, *94*(5), 677–694. https://doi.org/10.1016/j.ajhg.2014.03.018

Post, K. L., Belmadani, M., Ganguly, P., Meili, F., Dingwall, R., McDiarmid, T. A., Meyers, W. M., Herrington, C., Young, B. P., Callaghan, D. B., Rogic, S., Edwards, M., Niciforovic, A., Cau, A., Rankin, C. H., O'Connor, T. P., Bamji, S. X., Loewen, C. J. R., Allan, D. W., … Haas, K. (2020). Multi-model functionalization of disease-associated PTEN missense mutations identifies multiple molecular mechanisms underlying protein dysfunction. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-15943-0

Ranganathan, R., Sawin, E. R., Trent, C., & Horvitz, H. R. (2001). Mutations in the Caenorhabditis elegans serotonin reuptake transporter MOD-5 reveal serotonin-dependent and -independent activities of fluoxetine. *Journal of Neuroscience*, *21*(16), 5871–5884. https://doi.org/10.1523/jneurosci.21-16-05871.2001

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M.,

Lyon, E., Spector, E., Voelkerding, K., & Rehm, H. L. (2015). Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*, *17*(5), 405–424. https://doi.org/10.1038/gim.2015.30

Robinson, E. B., Samocha, K. E., Kosmicki, J. A., McGrath, L., Neale, B. M., Perlis, R. H., & Daly, M. J. (2014). Autism spectrum disorder severity reflects the average contribution of de novo and familial influences. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(42), 15161–15165. https://doi.org/10.1073/pnas.1409204111

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

Rylaarsdam, L., & Guemez-Gamboa, A. (2019). Genetic Causes and Modifiers of Autism Spectrum Disorder. *Frontiers in Cellular Neuroscience*, *13*(August), 1–15. https://doi.org/10.3389/fncel.2019.00385

Sakai, Y., Shaw, C. A., Dawson, B. C., Dugas, D. V., Al-Mohtaseb, Z., Hill, D. E., & Zoghbi, H. Y. (2011). Protein interactome reveals converging molecular pathways among autism disorders. *Science Translational Medicine*, *3*(86). https://doi.org/10.1126/scitranslmed.3002166

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

Samocha, K. E., Robinson, E. B., Sanders, S. J., Stevens, C., Sabo, A., McGrath, L. M., Kosmicki, J. A., Rehnström, K., Mallick, S., Kirby, A., Wall, D. P., MacArthur, D. G., Gabriel, S. B., DePristo, M., Purcell, S. M., Palotie, A., Boerwinkle, E., Buxbaum, J. D.,

Cook, E. H., … Daly, M. J. (2014). A framework for the interpretation of de novo mutation in human disease. *Nature Genetics*, *46*(9), 944–950. https://doi.org/10.1038/ng.3050

Sanders, S. J., Murtha, M. T., Gupta, A. R., Murdoch, J. D., Raubeson, M. J., Willsey, A. J., Ercan-Sencicek, A. G., DiLullo, N. M., Parikshak, N. N., Stein, J. L., Walker, M. F., Ober, G. T., Teran, N. a., Song, Y., El-Fishawy, P., Murtha, R. C., Choi, M., Overton, J. D., Bjornson, R. D., … State, M. W. (2012). De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature*, *485*(7397), 237–241. https://doi.org/10.1038/nature10945

Sandin, S., Lichtenstein, P., Kuja-Halkola, R., Hultman, C., Larsson, H., & Reichenberg, A. (2017). The Heritability of Autism Spectrum Disorder. *Jama*, *318*(12), 1182–1184.

Satterstrom, F. K., Kosmicki, J. A., Wang, J., Breen, M. S., De Rubeis, S., An, J. Y., Peng, M., Collins, R., Grove, J., Klei, L., Stevens, C., Reichert, J., Mulhern, M. S., Artomov, M., Gerges, S., Sheppard, B., Xu, X., Bhaduri, A., Norman, U., … Buxbaum, J. D. (2020). Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell*, *180*(3), 568-584.e23. https://doi.org/10.1016/j.cell.2019.12.036

Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-martin, C., Walsh, T., Yamrom, B., Yoon, S., Krasnitz, A., Kendall, J., Leotta, A., Pai, D., Zhang, R., Lee, Y., Hicks, J., Spence, S. J., Lee, A. T., Puura, K., Lehtimäki, T., … Warburton, D. (2007). Strong association of de novo copy number mutations with autism. *Science*, *316*(2005), 445–449.

Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Månér, S., Massa, H., Walker, M., Chi, M., Navin, N., Lucito, R., Healy, J., Hicks, J., Ye, K., Reiner, A., Gilliam, T. C., Trask, B., Patterson, N., … Wigler, M. (2004). Large-scale copy number polymorphism in the human genome. *Science*, *305*(5683), 525–528. https://doi.org/10.1126/science.1098918

Shaye, D. D., & Greenwald, I. (2011). Ortholist: A compendium of C. elegans genes with human orthologs. *PLoS ONE*, *6*(5). https://doi.org/10.1371/journal.pone.0020085

Silk, M., Petrovski, S., & Ascher, D. B. (2019). MTR-Viewer: identifying regions within genes under purifying selection. *Nucleic Acids Research*, *47*(W1), W121–W126. https://doi.org/10.1093/nar/gkz457

Starita, L. M., Ahituv, N., Dunham, M. J., Kitzman, J. O., Roth, F. P., Seelig, G., Shendure, J., & Fowler, D. M. (2017). Variant interpretation: functional assays to the rescue. *American Journal of Human Genetics*, *101*(3), 315–325. https://doi.org/10.1016/j.ajhg.2017.07.014

Steffenburg, S., Gillberg, C., Hellgren, L., Andersson, L., Gillberg, I. C., Jakobsson, G., & Bohman, M. (1989). A twin study of autism in Denmark, Finland, Iceland, Norway, and Sweden. *Journal of Child Psychology and Psychiatry*, *30*(3), 405–416.

Stoner, R., Chow, M. L., Boyle, M. P., Sunkin, S. M., Mouton, P. R., Roy, S., Wynshaw-Boris, A., Colamarino, S. A., Lein, E. S., & Courchesne, E. (2014). Patches of Disorganization in the Neocortex of Children with Autism. *New England Journal of Medicine*, *370*(13), 1209–1219. https://doi.org/10.1056/nejmoa1307491

Strooper, B. De, Annaert, W., Cupers, P., Saftig, P., Craessaerts, K., Mumm, J. S., Schroeter, E. H., Schrijvers, V., Wolfe, M. S., Ray, W. J., Goate, A., & Kopan, R. (1999). A presenilin-1-dependent gamma-secretase-like protease mediates release of Notch intracellular domain. *Nature*, *398*, 518–522.

Sullivan, J. M., De Rubeis, S., & Schaefer, A. (2019). Convergence of spectrums: neuronal gene network states in autism spectrum disorder. *Current Opinion in Neurobiology*, *59*, 102–111. https://doi.org/10.1016/j.conb.2019.04.011

Sun, S., Yang, F., Tan, G., Costanzo, M., Oughtred, R., Hirschman, J., Theesfeld, C. L., Bansal, P., Sahni, N., Yi, S., Yu, A., Tyagi, T., Tie, C., Hill, D. E., Vidal, M., Andrews, B. J., Boone, C., Dolinski, K., & Roth, F. P. (2016). An extended set of yeast-based functional assays

accurately identifies human disease mutations. *Genome Research*, *26*(5), 670–680. https://doi.org/10.1101/gr.192526.115

Sun, W., Poschmann, J., Cruz-Herrera del Rosario, R., Parikshak, N. N., Hajan, H. S., Kumar, V., Ramasamy, R., Belgard, T. G., Elanggovan, B., Wong, C. C. Y., Mill, J., Geschwind, D. H., & Prabhakar, S. (2016). Histone Acetylome-wide Association Study of Autism Spectrum Disorder. *Cell*, *167*(5), 1385-1397.e11. https://doi.org/10.1016/j.cell.2016.10.031

Sundaram, M., & Greenwald, I. (1993). Suppressors of a lin-12 hypomorph define genes that interact with both lin-12 and glp-1 in Caenorhabditis elegans. *Genetics*, *135*, 765–783.

Tager-Flusberg, H., & Kasari, C. (2013). Minimally verbal school-aged children with autism spectrum disorder: The neglected end of the spectrum. *Autism Research*, *6*(6), 468–478. https://doi.org/10.1002/aur.1329

The 1000 Genomes Project Consortium. (2010). A map of human genome variation from population-scale sequencing. *Nature*, *467*(7319), 1061–1073. https://doi.org/10.1038/nature09534

The 1000 Genomes Project Consortium. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, *491*(7422), 56–65. https://doi.org/10.1038/nature11632

Varghese, M., Keshav, N., Jacot-Descombes, S., Warda, T., Wicinski, B., Dickstein, D. L., Harony-Nicolas, H., Rubeis, S. De, Drapeau, E., Buxbaum, J. D., & Hof, P. R. (2017). Autism spectrum disorder: neuropathology and animal models. *Acta Neuropathologica*, *134*(4), 537–566. https://doi.org/10.1016/j.physbeh.2017.03.040

Voineagu, I., Wang, X., Johnston, P., Lowe, J. K., Tian, Y., Horvath, S., Mill, J., Cantor, R. M., Blencowe, B. J., & Geschwind, D. H. (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature*, *474*(7351), 380–386. https://doi.org/10.1038/nature10110

Weiss, L. a, Arking, D. E., Daly, M. J., & Chakravarti, A. (2009). A genome-wide linkage and

association scan reveals novel loci for autism. *Nature*, *461*(7265), 802–808. https://doi.org/10.1038/nature08490

Wen, Y., Alshikho, M. J., & Herbert, M. R. (2016). Pathway Network Analyses for Autism Reveal Multisystem Involvement , Major Overlaps with Other Diseases and Convergence upon MAPK and Calcium Signaling. *PloS One*, *11*(4), e0153329. https://doi.org/10.1371/journal.pone.0153329

Wertz, M. H., Mitchem, M. R., Pineda, S. S., Hachigian, L. J., Lee, H., Lau, V., Powers, A., Kulicke, R., Madan, G. K., Colic, M., Therrien, M., Vernon, A., Beja-Glasser, V. F., Hegde, M., Gao, F., Kellis, M., Hart, T., Doench, J. G., & Heiman, M. (2020). Genome-wide In Vivo CNS Screening Identifies Genes that Modify CNS Neuronal Survival and mHTT Toxicity. *Neuron*, *106*(1), 76-89.e8. https://doi.org/10.1016/j.neuron.2020.01.004

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

Wong, W. R., Brugman, K. I., Maher, S., Oh, J. Y., Howe, K., Kato, M., & Sternberg, P. W. (2019). Autism-associated missense genetic variants impact locomotion and neurodevelopment in Caenorhabditis elegans. *Human Molecular Genetics*, *28*(13), 2271–2281. https://doi.org/10.1093/hmg/ddz051

Yates, C. M., Filippis, I., Kelley, L. A., & Sternberg, M. J. E. (2014). SuSPect: Enhanced prediction of single amino acid variant (SAV) phenotype using network features. *Journal of Molecular Biology*, *426*(14), 2692–2701. https://doi.org/10.1016/j.jmb.2014.04.026

Yi, S., Lin, S., Li, Y., Zhao, W., Mills, G. B., & Sahni, N. (2017). Functional variomics and network perturbation: Connecting genotype to phenotype in cancer. *Nature Reviews*

Genetics, *18*(7), 395–410. https://doi.org/10.1038/nrg.2017.8

Zhao, Y. T., Kwon, D. Y., Johnson, B. S., Fasolino, M., Lamonica, J. M., Kim, Y. J., Zhao, B. S., He, C., Vahedi, G., Kim, T. H., & Zhou, Z. (2018). Long genes linked to autism spectrum disorders harbor broad enhancer-like chromatin domains. *Genome Research*, *28*(7), 933–942. https://doi.org/10.1101/gr.233775.117

Chapter 2

AUTISM-ASSOCIATED MISSENSE GENETIC VARIANTS IMPACT LOCOMOTION AND NEURODEVELOPMENT IN *CAENORHABDITIS ELEGANS*

Wong, W. R., Brugman, K. I., Maher, S., Oh, J. Y., Howe, K., Kato, M., & Sternberg, P. W. (2019). Autism-associated missense genetic variants impact locomotion and neurodevelopment in Caenorhabditis elegans. *Human Molecular Genetics*, *28*(13), 2271–2281. https://doi.org/10.1093/hmg/ddz051

2.1 Abstract

Autism spectrum disorder (ASD) involves thousands of alleles in over 850 genes but the current functional inference tools are not sufficient to predict phenotypic changes. As a result, the causal relationship of most of these genetic variants in the pathogenesis of ASD has not yet been demonstrated, and an experimental method prioritizing missense alleles for further intensive analysis is crucial. For this purpose, we have designed a pipeline that uses *C. elegans* as a genetic model to screen for phenotype-changing missense alleles inferred from human ASD studies. We identified highly conserved human ASD-associated missense variants in their *C. elegans* orthologs, used a CRISPR/Cas9-mediated homology-directed knock-in strategy to generate missense mutants, and analyzed their impact on behaviors and development via several broad-spectrum assays. All tested missense alleles were predicted to perturb protein function, but we found only 70% of them showed detectable phenotypic changes in morphology, locomotion, or fecundity. Our findings indicate that certain missense variants in the *C. elegans* orthologs of

human *CACNA1D*, *CHD7*, *CHD8*, *CUL3, DLG4, GLRA2*, *NAA15*, *PTEN*, *SYNGAP1*, and *TPH2* impact neurodevelopment and movement functions, elevating these genes as candidates for future study into ASD. Our approach will help prioritize functionally important missense variants for detailed studies in vertebrate models and human cells.

2.2 Introduction

Many psychiatric disorders such as autism spectrum disorder (ASD, OMIM: 209850) have been linked to genetic variants that disrupt but do not necessarily eliminate protein functions. Missense variants in particular account for approximately half of the genetic changes known to cause disease (Andrews et al., 2013) but most studies focus on identifying likely gene-disruptive (LGD) mutations (e.g., nonsense, frameshift, or splice-site) instead of missense variants. The severity of ASD is thought to be correlated with the average contribution of familial influences and *de novo* mutations (Robinson et al., 2014); individuals with ASD are more likely to carry a *de novo* missense mutation (Iossifov et al., 2014). Missense mutations account for a large number of variants of uncertain significance (VUS), which are genomic variants that have unclear effect on protein function and clinical significance due to inadequate or conflicting information (Han, 2013; Petrucelli et al., 2002). Given that some missense alleles have been validated, one challenge is to identify the subset of ASD-associated mutations that are deleterious.

Because missense variants are numerous, functional inference tools are widely used to predict the damaging effects of specific missense variants. Most current software relies heavily on sequence conservation to predict the potency of missense variants as conserved regions are considered more likely to be affected by purifying selection (Alfoldi & Lindblad-toh, 2013) but only 27% of missense mutations predicted by sequence conservation showed disrupted protein function in a recent rodent study (Miosge et al., 2015). Given that any gene carries a certain

chance of containing a missense mutation and every individual will have a different subset of missense mutations in their genome, computational analyses are insufficient for predicting the functional importance of such mutations (Miosge et al., 2015). Additionally, interpretation of these data is inadequate due to variable penetrance, dosage sensitivity, and functional redundancy of mutated proteins, and can result in a high false-positive rate of prediction (Andrews et al., 2013; Tennessen et al., 2012). On the other hand, variants that scored as neutral/benign may impact other physiological functions which were not hypothesized/expected (Billack & Monteiro, 2004). Therefore, the functional inference tools used to predict damaging effects are not accurate enough to be used as the sole basis for a conclusion, and a test of broad biological phenotypes is necessary to understand the nature of missense variants with uncertain significance.

Evaluation of missense variants *in vivo* is essential to accurately interpret available data as only 13% of identified *de novo* missense variants are suspected to contribute to the risk of ASD (Iossifov et al., 2014, 2012). Due to the large number of missense variants, an efficient pipeline is needed to evaluate the functional consequence of all residues *in vivo.* There have only been a few studies conducted to validate the functional consequence of missense variants *in vivo*. Chen et al. evaluated the disruptiveness of a mutation exclusively on its capacity to disrupt protein interactions using the yeast two-hybrid method (Chen et al., 2018). Another study by Miosge et al. compared the deleterious effects predicted computationally to the actual ENU mutant rodent models (Miosge et al., 2015). Despite such exciting findings, a comprehensively targeted screen to test whether ASD-associated missense mutations are function-disrupting in a multi-cellular model organism has not yet been done. The short life cycle and easily accessible genome in *Caenorhabditis elegans* (*C. elegans*) make it an useful tool to rapidly evaluate whether a particular disease-associated missense variant results in phenotypic consequences (Brenner, 1974; Engleman et al., 2016; Kim et al., 2017).

In this study, we established a pipeline for identifying ASD-associated protein-disrupting missense residues in the orthologous *C. elegans* proteins (Figure 1A). First, the *C. elegans* residues corresponding to human missense variants were identified based on sequence conservation. The *C. elegans* equivalents of human missense mutants were generated using Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-Cas9 and homology-directed genome editing ("knock-in"). We then analyzed the effects of these autism-associated missense alleles by comparing observable phenotypes from these missense mutants to the wild-type and to known loss-of-function mutant controls. Missense mutants with phenotypic changes reflect alteration in protein function, indicating the importance of these alleles. We found that 19% of the ASD-associated missense variants are conserved in *C. elegans*. We evaluated the effects of 20 missense alleles that were predicted to be phenotype altering and found that only 70% of them displayed phenotypic changes in morphology, locomotion, and fecundity. Our method demonstrates our ability to screen for subtle phenotypic changes and in doing so, illustrate the functional importance of the effect of missense mutations on human disease.

2.3 Materials and Methods

Mapping locations of human residues to the C. elegans *genome*

ASD missense variants were obtained from the SFARI Gene–Human Gene Module (Fischbach & Lord, 2010) (Table S1). We used the comparative genomics resources provided by Ensembl (release 90), which integrates in-house annotation for nearly 100 vertebrate genomes (e.g., human, mouse, and zebrafish) with reference annotation for selected invertebrate model organisms (e.g., *C. elegans*, with genome and annotation provided by WormBase). Ensembl provides a protein multiple alignment and evolutionary trees for each gene family, and asserts orthology and paralogy relationships between pairs of genes (Herrero et al., 2016). These data

were organized with a custom automated pipeline (Yates et al., 2015): for a given human genome coordinate, (a) identify which human protein-coding gene (if any) coincided with the provided coordinate; (b) obtain the amino acid coordinates in that protein; (c) check if the human gene has a *C. elegans* ortholog; (d) if so, use the multiple alignment associated with the orthology assertion to identify the orthologous amino acid in the *C. elegans* protein; and (e) from the protein coordinates, obtain the corresponding position in the *C. elegans* reference genome.

Strains

The Bristol N2 *C. elegans* strain was used as the wild-type control and background for all CRISPR experiments(Brenner, 1974). The control strains for functional assays were obtained from lab stock, the Caenorhabditis Genetics Center (CGC), and the National BioResource Project- *C. elegans* (NBRP). Loss-of-function mutant controls were JD105 *avr-15(ad1051)* (Dent et al., 1997), FX17094 *chd-7(tm6139)*, PS3071 *egl-19(n2368sd)* (Lee et al., 1997), SD464 *mpk-1(ga117)* (Lackner et al., 1994), and PS3156 *tph-1(mg280)* (Sze et al., 2000). All strains were maintained on nematode growth medium (NGM) agar plates seeded with *Escherichia coli* OP50 at room temperature (20-22 °C).

Generation of missense mutant strains

The Cas9 protein-based CRISPR knock-in protocol was adapted from Paix et al. (Paix et al., 2015). The sgRNA sequences were selected using the *C. elegans* CRISPR guide RNA tool (Au et al., 2018). Single-stranded donor oligonucleotides contained 35 bp of flanking homology on both sides of the mutated region. An online tool for restriction analysis, WatCut, was used to assist in designing restriction sites that did not affect protein sequence. The crRNA, tracrRNA and donor oligonucleotides were commercially synthesized and dissolved in Nuclease Free Duplex Buffer (Integrated DNA Technologies Inc., Coralville, IA). Purified Cas9 protein was a

kind gift from Dr. Tsui-Fen Chou (LA BioMed). gRNA duplexes were generated by mixing crRNA and tracrRNA at 1:1 ratio and incubating at 94°C for 2 minutes. The Cas9 protein (25 μM final concentration) and gRNA duplex (27 μM final concentration) were mixed and incubated at room temperature for 5 minutes before adding donor oligonucleotides (0.6 μM final concentration). To facilitate screening, *dpy-10(cn64)* or *unc-58(e665)* was used as a co-conversion marker and made up part of the crRNA and donor oligo used (Arribere et al., 2014) (Figure 1B). A crRNA ratio [marker: target gene] of 1:4 and 2:3 were used for *dpy-10* and *unc-58,* respectively. A donor ratio [marker: target gene] of 1:2 was used for both *dpy-10* and *unc-58*.

The F1 offspring displaying the co-conversion phenotype were genotyped as follows: About 5 worms were picked into 10 μL lysis buffer (10 mM Tris, 50 mM KCl, 2 mM MgCl2, pH 8.0) with proteinase K (500 ng/mL; Invitrogen, Carlsbad, CA) and incubated at 65°C for an hour to extract genomic DNA. The genomic prep was amplified in a PCR reaction and then treated with a restriction enzyme (NEB, Ipswich, MA) to check the presence of targeted missense mutation. Mutants with the correct length were confirmed by sequencing (Laragen, Culver City, CA). When available, we saved two independent missense mutant lines. While there are little to no off-targets effects of Cas9 (Chiu et al., 2013), *C. elegans* N2 suffers approximately one mutation per generation so it was useful to have more than one strain for each locus.

Fecundity assay

Well-fed *C. elegans* were synchronized at the L4 stage. Individual L4 hermaphrodites were placed on separate NGM plates seeded with OP50 and these animals were subsequently transferred to a new plate every day. The number of newly hatched larvae progeny was counted for every plate 1 day after the adult was transferred. The total fecundity consisted of the sum of progeny produced for three days per animal.

Locomotion tracking

Well-fed L4 hermaphrodites were picked at about 16 hours before the experiment to provide synchronized young adults. On the day of the experiment, 8 young adults were picked onto NGM plates freshly seeded with a 50-μL drop of a saturation-phase culture of OP50. The worms were given 30 min for habituation and then tracked for 4 min. Strains were tracked between 1 p.m. and 6 p.m. across several days. WormLab (MBF Bioscience, Williston, VT) equipment and software were used for tracking and analyses. The camera was a Nikon AF Micro 60/2.8D with zoom magnification. A 2456×2052–resolution, 7.5-f.p.s. camera with a magnification that results in 8.2 μm per pixel and an FOV of roughly 2×2 cm^2 was used. Approximately 8-10 plates were tracked per experimental strain. The mean of each plate was first calculated and then the total mean of all plates of the same genotype was computed.

Statistical analysis

The fecundity assay was analyzed using a non-parametric bootstrap analysis (D. Angeles-Albores & P.W.S., unpublished). Initially, the two datasets were mixed, samples were selected at random with replacement from the mixed population into two new datasets, and then the difference in the averages of these new datasets were calculated; this process was iterated 10^6 times. We reported the *p*-value as the probability when the difference in the average of simulated datasets was greater than the difference in the average of the original datasets. If $p < 0.01/(\text{total testing number})$, we rejected the null hypothesis that the average values of the two datasets were not equal to each other. Morphology and locomotion were analyzed by one-way ANOVA using GraphPad Prism version 6 (GraphPad, La Jolla, CA). Dunnett multiple comparisons were performed between wild-type and mutant strains. The significant level was defined as $p < 0.01$.

2.4 Results

Identifying C. elegans analogs of ASD-associated missense mutations

In order to identify the functionally important missense variants implicated in complex human diseases, we established a pipeline to screen for functional changes in orthologous proteins in *C. elegans* (Figure 2-1A). Of 1811 human ASD-associated missense variants from 423 human genes, 778 alleles (43%) from 221 human genes were identified in *C. elegans* orthologs. Most of the human genes were aligned to one *C. elegans* ortholog, but approximately 20% of the genes (47 of 221) had more than one orthologous protein in *C. elegans* (Figure 2-2A). In some cases, human genes from the same family share the same *C. elegans* orthologous protein (e.g., both human *CHD7* and *CHD8* genes share the same C. elegans orthologs *chd-7*). Our goal was to identify each orthologous protein and corresponding equivalent residue based on sequence conservation. To achieve this, our software utilized comparative genomics and multiple alignments to ensure that the detected residue reflects conservation across the evolutionary tree and gene family. We found that 345 (19%) of missense loci from 157 human genes not only have orthologs in *C. elegans* but also had at least one conserved amino acid residue between human and *C. elegans* (Figure 2-2A). Sometimes, one human residue could be matched to multiple orthologs in *C. elegans* (37 of the 345 conserved residues). For example, *GLRA2* has multiple orthologous proteins in *C. elegans*, *glc-1*, *glc-2*, *glc-3*, *glc-4*, *avr-14*, and *avr-15*. In these cases, we picked the worm residue candidate that had the sgRNA sequence most likely to produce an efficient CRISPR-Cas9 double strand break based on an online sgRNA prediction tool (Au et al., 2018). For each allele, we identified the corresponding *C. elegans* ortholog, assessed the residues affected by missense mutations for evolutionary conservation, and selected genes with a known phenotype for their loss-of-function mutation in *C. elegans* (from existing mutants or RNAi). To prioritize genes for functional screening, we focused on those genes with multiple missense variants as the chance of one causing a phenotypic defect

increases when multiple missense mutations are observed in a single gene (Geisheker et al., 2017). We also prioritized genes involved in multiple biological pathways (Krumm et al., 2014) or genes with other mutations resulting in a stop codon.

To capture the impacts of missense mutations in diverse physiological functions, we sampled 20 ASD-associated missense changes in residues conserved in the *C. elegans* orthologs of 11 human genes (Table 2-1; Figure S1). These ASD-associated missense mutations were identified in genes that were known to have a role in synaptic function (i.e., *DLG4*, *SYNGAP1, CACNA1D,* and *GLRA2*), gene expression regulation (i.e., *CHD7*, *CHD8, and CUL3)* or neuronal signaling and cytoskeleton functions (i.e., *PTEN*, *MAPK3*, *TPH2*, and *NAA15*). Multiple aspects of physiological functions were examined, including morphology, locomotion, and fecundity. These well-established quantitative assays enabled us to detect subtle changes in morphology, movement, and coordination, as well as reproduction and completion of embryonic development (Bono & Villu Maricq, 2005; Engleman et al., 2016).

Morphology of missense mutants

To examine changes in morphology, we utilized a quantitative tracking system to measure the length, width, and body area of these missense mutants under freely moving condition. Alterations in size were detected in, *avr-15/GLR2*, *chd-7/CHD7 or CHD8*, *cul-3/CUL3*, *daf-18/PTEN*, *gap-2/SYNGAP1*, *egl-19/CACNA1D*, *hpo-29/NAA15*, and *tph-1/TPH2* (Table 2-2). Every *chd-7* mutant tested showed a significant decrease in body width and area. A null mutant, *chd-7(sy956),* displayed the most severe defects. Other missense alleles, *chd-7(L1220P)*, *chd-7(L1487R)*, *chd-7(G1225S)*, and *chd-7(P253L)* showed milder degree of defects. One of the *egl-19* missense mutants, *egl-19(Y333S),* displayed a smaller decrease in body length, width and areas compared to the semidominant allele, *egl-19(n2368)* (Lee et al., 1997). Another *egl-19* mutant, *egl-19(V331M)*, showed a similar body size as the N2 wild-type

strain. Similarly, the *tph-1(R259Q)* mutant showed a decrease in body length and area, and the change was milder in the missense mutant as compared to the null mutant *tph-1(mg280)* (Sze et al., 2000). One missense mutation in *avr-15, avr-15(R364Q),* caused a decrease in body length, width, and area, similar to its null mutant, *avr-15(ad1051)* (Dent et al., 1997). Another *avr-15* missense mutant, *avr-15(N347S)*, showed no morphological changes. Missense mutant *hpo-29(L575S)* exhibited shorter body length. Missense mutants *cul-3(H728R)*, *daf-18(H168Q)*, and *gap-2(C417Y)* displayed increased body width and area. Missense mutants of *dlg-1/DLG4* and *mpk-1/MAPK3* did not show morphological changes.

Movement and coordination of missense mutants

To examine movement and coordination in these missense mutants, a quantitative tracking system was used to measure moving speed, reversal rate, and sinusoidal wavelength and amplitude (Supplemental Movie). Locomotion defects were found in missense mutants of *chd-7/CHD7* or *CHD8*, *daf-18/PTEN*, *gap-2/GLRA2*, and *hpo-29/NAA15* (Figure 2-3; Table S2). Less severe than null mutant, all missense mutants in *chd-7*, except *chd-7(P253L)*, exhibited decreased speed. Missense mutant *hpo-29(L575S)* also showed a significant decrease in speed. In terms of reversal rate, most *chd-7* mutants, except *chd-7(P253L)*, showed a significant reduction in turns per minute. Missense mutants *daf-18(H138R)* and *gap-2(C417Y)* displayed an increased reversal rate. Missense mutations in *avr-15/GLRA2*, *cul-3/CUL3*, *dlg-1/DLG4*, *egl-19/CACNA1D*, *mpk-1/MAPK3*, and *tph-1/TPH2* did not result in differences in speed and reversal rate.

Locomotion in *C. elegans* is typically expressed as the wavelength and amplitude of a sinusoidal wave (Cronin et al., 2005). Motor coordination defects have been associated with ASD (Fournier et al., 2010) and were found in missense mutants of *avr-15/GLRA2*, *chd-7/CHD7* or *CHD8*, *daf-18/PTEN*, and *tph-1/TPH2* (Figure 2-3; Table S2-2). The *tph-1(R259Q)*

mutant exhibited significantly lower wavelength and higher amplitude, indicating a more curvy sinusoidal wave similar to but less severely than its null mutant (Figure S2-2). One of the *avr-15* missense mutants, *avr-15(R364Q)*, showed a decrease in wavelength, slightly milder than the null mutant, *avr-15(ad1051)* (Dent et al., 1997). Another *avr-15* mutant, *avr-15(N347S)*, showed normal sinusoidal shape. All missense mutants in *chd-7*, except *chd-7(P253L)*, displayed a decreased wavelength and/or amplitude. The mutation in *chd-7(L1220P)* resulted in a decrease in both wavelength and amplitude. Mutations in *chd-7(G1225S)* and *chd-7(L1487R)* led to a decrease in wavelength and amplitude respectively. A null mutant of *chd-7* also displayed a decrease wavelength. Two of the *daf-18* missense mutants, *daf-18(H138R)* and *daf-18(H168Q)*, exhibited an increase in amplitude and wavelength respectively. Missense mutations in *cul-3/CUL3*, *dlg-1/DLG4*, *egl-19/CACNA1D*, *gap-2/SYNGAP1, hpo-29/NAA15*, and *mpk-1/MAPK3* did not lead to differences in the sinusoidal wave.

Fecundity of missense mutants

We used the fecundity assay to examine larvae viability in genes with reported sterile or lethal phenotypes in null mutants. Fecundity defects were found in missense mutants of *chd-7/CHD7 or CHD8*, *cul-3/CUL3*, and *dlg-1/DLG4* (Figure 2-4). Three of four missense mutations in the chromatin modifier gene *chd-7* displayed a reduced fecundity phenotype compared to the wild-type control strain N2. Specifically, the *chd-7(L1220P)* allele had a median fecundity of 119 ($p < 10^{-6}$); *chd-7(G1225S)* had a median fecundity of 176 ($p = 1.2*10^{-5}$); *chd-7(L1487R)* had a median fecundity of 168 ($p = 4.4*10^{-5}$); and *chd-7(P253L)* had a median fecundity of 254.5 compared to a median fecundity of 228 for N2 control. These missense alleles showed weaker fecundity defects compared to its deletion (*chd-7(tm6139)*) or frameshift (*chd-7(sy956)*) controls, which had median fecundity of 38 and 47, respectively ($p < 10^{-6}$). Missense variants in the DNA replication gene, *cul-3(H728R)*, also displayed a

decreased fecundity of 168.5 ($p = 10^{-6}$). In addition, the missense variant *dlg-1(V964I)* showed a reduced median fecundity of 154.5 ($p < 10^{-6}$), which is slightly less severe than the 67% reduction in a previous RNAi study (Pilipiuk et al., 2009). We did not observe changes in fecundity in missense mutants in other genes, namely *avr-15/GLRA2*, *daf-18/PTEN*, *egl-19/CACNA1D*, *gap-2/SYNGAP1*, *hpo-29/NAA15*, *mpk-1/MAPK3*, and *tph-1/TPH2*.

Comparison with phenotype-predicting software

To examine the accuracy of our biological platform, we compared our results to the existing prediction software Sorting Intolerant From Tolerant (SIFT) and Polymorphism Phenotyping v.2 (PolyPhen-2) (Table 2-3). SIFT emphasizes sequence conservation and the physical properties of amino acids (Ng & Henikoff, 2003) whereas PolyPhen-2 considers both the analysis of multiple sequence alignments and protein 3D-structures (Adzhubei et al., 2010). Both software programs are commonly used to predict the effects of non-synonymous amino acid changes. For SIFT, all the alleles tested were predicted to be damaging due to having a similar approach of analyzing sequence conservation as our software. Our phenotypic assays identified six residues (among the 20 predictions) that did not align with the prediction. As compared to PolyPhen-2's prediction, 35% (7/20) of the phenotypic results do not agree with the predictions. Among the seven strains that did not match, five were predicted to have damaging effects but had no phenotypic change in our functional assays (false positive), and two were predicted to be benign but displayed phenotypic changes (false negative). Overall, our results demonstrated that 70% (14 of 20) missense alleles predicted to be damaging by at least one functional inference tool actually showed detectable phenotypic changes in morphology, locomotion, and fecundity.

2.5 Discussion

In this study, we have developed a fast and tractable pipeline to comprehensively screen for ASD-associated missense mutations. Our analysis finds that 43% of the human disease-associated alleles have an ortholog in the genome of *C. elegans*, which is consistent with previous estimates (Markaki & Tavernarakis, 2010). Among the 19% conserved loci, we evaluated 20 missense alleles that were predicted to be damaging, and found 70% of them actually cause detectable phenotypic changes. We have successfully prioritized 14 missense variants that are functionally significant in *C. elegans* orthologs of human genes. These are the first animal models with deliberately engineered missense mutations in these loci. Our approach is useful for characterizing novel missense alleles that are potentially relevant to human disease and be used as a tool to identify functionally consequential alleles.

Compared to null mutants, most of the phenotypically altered missense alleles displayed milder phenotypes, indicating that our assays can detect relatively subtle changes in protein functions. For example, the *chd-7* missense mutants and *tph-1(R259Q)* displayed hypomorphic phenotypes less severe than their null mutants (Yemini et al., 2013). The *cul-3(H728R)* and *dlg-1(V964I)* missense mutants displayed a smaller reduction in fecundity compared to previous RNAi studies (Maeda et al., 2001; Pilipiuk et al., 2009; Sonnichsen et al., 2005). *avr-15(R364Q)* showed defects in morphology and locomotion similar to its null mutant, *avr-15(ad1051)*, even though it did not recapitulate the spontaneous reversal rate defect documented in a RNAi study (Cook et al., 2006). The functional consequences of missense alleles can vary in different assays. For instance, the *egl-19(Y333S)* showed milder morphological changes similar to its null mutant (Yemini et al., 2013) but displayed normal functions in locomotion and fecundity. Missense mutants *hpo-29(L575S)* and *gap-2(C417Y)* displayed defects in morphology and locomotion, but they did not show the fecundity defects reported in RNAi studies (Maeda et al., 2001; Rual et al., 2004). The *daf-18(H138R)* and *daf-*

18(H168Q) missense mutants showed defects in morphology and locomotion, which were not documented before, suggesting a role for our biological screening platforms to detect subtle phenotypic changes in different physiological functions.

As pointed out in previous literature, computational inference tend to have a higher false-positive rate of identifying protein function-disrupting missense alleles (Andrews et al., 2013). This study demonstrated that predictions based solely on sequence conservation did not effectively distinguish missense mutations that cause phenotypic changes from ones that exhibit no observable phenotype. Only 70% of our behavioral results agreed with the predictions from two commonly used computational programs, PolyPhen-2 and SIFT. Most of the discrepancies are false positive predictions. Absence of a phenotype *in vivo* may occur due to genetic redundancy and robust gene networks compensating for the inhibition of a single component, especially in tightly regulated cellular networks involving in signaling, metabolic, and transcriptional pathways (El-Brolosy & Stainier, 2017), or, that we simply did not observe every possible phenotype. More pointedly, our study showed that two missense alleles, predicted by PolyPhen-2 as benign, presented phenotypes. The fecundity defect found in *dlg-1(V964I)* can be recapitulated by RNAi whereas the *tph-1(R259Q)* displayed hypomorphic phenotypes similar to its null mutant (Pilipiuk et al., 2009). The false negatives predicted by the software indicate a void in current prediction algorithms, suggesting a need for a screening platform in a multicellular model organism such as our own. Our *in vivo* screening platform not only selects for genes that display sequence conservation across evolution, it also reflects the complex nature in biological system, such as redundancy and compensation. Our phenotypic results can also provide feedback to improve the accuracy of prediction algorithm.

In contrast with previous studies on the phenotypic consequences of missense mutations, our platform examines gene functions in its endogenous multicellular context. Compared to a previous study using yeast two-hybrid to verify the effects of missense mutations in protein

interaction experimentally and computationally (Chen et al., 2018), our strategy captures the overall readout of mutation effects and intercellular interaction. Furthermore, our use of endogenous proteins allow all other molecular interactions to remain intact, and thus avoids potential confounding factors, such as intron disruption and isoform imbalance (Reble et al., 2018; Robison, 2014). As a result, our approach may more accurately reflect the consequence of a variant than does a "humanized" model organism (Baruah et al., 2017; McDiarmid et al., 2018; Walsh et al., 2017), and CRISPR knocking-in a DNA missense template is more efficient. Our high throughput screening strategy occupies an unusual niche in primary screening for the consequence of missense mutations *in vivo*.

Using *C. elegans* as a model for psychiatric disorders has some limitations, including a lack of highly complex behaviors and some neurotransmitter systems (e.g., norepinephrine). However, *C. elegans* and humans share essential physiological pathways (e.g., insulin signaling, Ras/Notch signaling, p53, and many miRNAs), neurotransmitter systems, and receptor pharmacology (Engleman et al., 2016; Markaki & Tavernarakis, 2010). The transparency and easy access genetic tools make *C. elegans* a powerful model for dissecting the mechanisms of pathological conditions and drug target identification. The short generation time of *C. elegans* enable high-throughput screening for numerous targets (such as missense variants) before embarking on less efficient and more costly animal models (Markaki & Tavernarakis, 2010). In addition, with the tissue-specific promoters and conditional knockout techniques available in *C. elegans*, it is possible to decipher the effects of these disease-associated missense mutations spatially and temporally (Hubbard, 2014; Shen et al., 2014; Voutev & Hubbard, 2008). For genetic candidates that show correlated expression, our platform also can be used to investigate the interaction between missense variants by generating double/multiple missense mutations model.

The discovery of novel genetic variants associated with human diseases has accelerated

due to technical improvements and decreasing costs of next-generation sequencing. However, it is difficult to assess the impact of single missense mutations due to the complexity of human genetic backgrounds. One solution is to test variants in a model organism with an isogenic background to quickly identify variants producing changes in protein function. Here, we developed an experimental pipeline to investigate the functional consequences of ASD-associated missense variants in *C. elegans*. Our approach will help prioritize consequential missense variants for detailed studies in vertebrate models or human cells. This pipeline will serve as a stepping stone for defining molecular mechanisms in complex human diseases such as ASD.

2.6 Acknowledgements

We thank WormBase for genome information. We thank CGC (funded by NIH Office of Research Infrastructure Programs, P40 OD010440) and NBRP for providing strains. We thank Tsui-Fen Chou for Cas9 protein. We thank Shahla Gharib for the assistance in strain generation. We thank David Angeles-Albores for sharing his data analysis software. We thank Hillel Schwartz and Han Wang for comments on manuscript. This work was supported by Simons Foundation (SFARI award # 367560) to PWS. KB was supported by National Institute of Health (NIH) pre-doctoral training grant T32GM007616. PWS was an investigator with the Howard Hughes Medical Institute during part of this study.

Conflicts of Interest

The authors have no conflicts of interest to disclose.

Web Resources

SFARI Gene–Human Gene Module, https://gene.sfari.org/database/human-gene/

Ensembl, www.ensembl.org

WormBase, www.wormbase.org

C. elegans CRISPR guide RNA tool, http://genome.sfu.ca/crispr/

WatCut, http://watcut.uwaterloo.ca

PolyPhen-2, http://genetics.bwh.harvard.edu/pph2/

SIFT, http://sift.bii.a-star.edu.sg

2.7 References

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., Kondrashov, A. S., & Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nature Methods*, *7*(4), 248–249. https://doi.org/10.1038/nmeth0410-248

Alfoldi, J., & Lindblad-toh, K. (2013). Comparative genomics as a tool to understand evolution and disease. *Genome Research*, *23*, 1063–1068. https://doi.org/10.1101/gr.157503.113.Freely

Andrews, T. D., Sjollema, G., & Goodnow, C. C. (2013). Understanding the immunological impact of the human mutation explosion. *Trends in Immunology*, *34*(3), 99–106. https://doi.org/10.1016/j.it.2012.12.001

Arribere, J., Bell, R., Fu, B., Artiles, K., Hartman, P., & Fire, A. (2014). Efficient marker-free recovery of custom genetic modifications with CRISPR/Cas9 in Caenorhabditis elegans. *Genetics*, *198*(3), 837–846. https://doi.org/10.1534/genetics.114.169730

Au, V., Li-Leger, E., Raymant, G., Flibotte, S., Chen, G., Martin, K., Fernando, L., Doell, C.,

Rosell, F., Wang, S., Edgley, M., Rougvie, A., Hutter, H., & Moerman, D. (2018). Optimizing guide RNA selection and CRISPR/Cas9 methodology for efficient generation of deletions in C. elegans. *BioRxiv*, 359588. https://doi.org/10.1101/359588

Baruah, P. S., Beauchemin, M., Parker, J. A., & Bertrand, R. (2017). Expression of human Bcl-xL (Ser49) and (Ser62) mutants in Caenorhabditis elegans causes germline defects and aneuploidy. *PLoS ONE*, *12*(5), e0177413. https://doi.org/10.1371/journal.pone.0177413

Billack, B., & Monteiro, A. N. A. (2004). Methods to classify BRCA1 variants of uncertain clinical significance: The more the merrier. *Cancer Biology and Therapy*, *3*(5), 458–459. https://doi.org/10.4161/cbt.3.5.831

Bono, M. de, & Villu Maricq, A. (2005). Neuronal substrates of complex behaviors in C. elegans. *Annual Review of Neuroscience*, *28*(1), 451–501. https://doi.org/10.1146/annurev.neuro.27.070203.144259

Brenner, S. (1974). The genetics of Caenorhabditis elegans. *Genetics*, *77*(May), 71–94. https://doi.org/10.1111/j.1749-6632.1999.tb07894.x

Chen, S., Fragoza, R., Klei, L., Liu, Y., Wang, J., Roeder, K., Devlin, B., & Yu, H. (2018). An interactome perturbation framework prioritizes damaging missense mutations for developmental disorders. *Nature Genetics*, *50*, 1032–1040. https://doi.org/10.1038/s41588-018-0130-z

Chiu, H., Schwartz, H. T., Antoshechkin, I., & Sternberg, P. W. (2013). Transgene-free genome editing in Caenorhabditis elegans using CRISPR-Cas. *Genetics*, *195*(3), 1167–1171. https://doi.org/10.1534/genetics.113.155879

Cook, A., Aptel, N., Portillo, V., Siney, E., Sihota, R., Holden-Dye, L., & Wolstenholme, A. (2006). Caenorhabditis elegans ivermectin receptors regulate locomotor behaviour and are functional orthologues of Haemonchus contortus receptors. *Molecular and Biochemical Parasitology*, *147*(1), 118–125. https://doi.org/10.1016/j.molbiopara.2006.02.003

Cronin, C. J., Mendel, J. E., Mukhtar, S., Kim, Y. M., Stirbl, R. C., Bruck, J., & Sternberg, P. W. (2005). An automated system for measuring parameters of nematode sinusoidal movement. *BMC Genetics*, *6*, 1–19. https://doi.org/10.1186/1471-2156-6-5

Dent, J. A., Davis, M. W., & Avery, L. (1997). avr-15 encodes a chloride channel subunit that mediates inhibitory glutamatergic neurotransmission and ivermectin sensitivity in Caenorhabditis elegans. *The EMBO Journal*, *16*(19), 5867–5879.

El-Brolosy, M. A., & Stainier, D. Y. R. (2017). Genetic compensation : A phenomenon in search of mechanisms. *PLoS Genetics*, *13*(7), e1006780. https://doi.org/10.1371/journal.pgen.1006780

Engleman, E. A., Katner, S. N., & Neal-beliveau, B. S. (2016). Caenorhabditis elegans as a model to study the molecular and genetic mechanisms of drug addiction. *Prog Mol Biol Transl Sci*, *137*, 229–252. https://doi.org/10.1016/bs.pmbts.2015.10.019.Caenorhabditis

Fischbach, G. D., & Lord, C. (2010). The simons simplex collection: A resource for identification of autism genetic risk factors. *Neuron*, *68*(2), 192–195. https://doi.org/10.1016/j.neuron.2010.10.006

Fournier, K. A., Hass, C. J., Naik, S. K., Lodha, N., & Cauraugh, J. H. (2010). Motor coordination in autism spectrum disorders: A synthesis and meta-analysis. *Journal of Autism and Developmental Disorders*, *40*(10), 1227–1240. https://doi.org/10.1007/s10803-010-0981-3

Geisheker, M. R., Heymann, G., Wang, T., Coe, B. P., Turner, T. N., Stessman, H. A. F., Hoekzema, K., Kvarnung, M., Shaw, M., Friend, K., Liebelt, J., Barnett, C., Thompson, E. M., Haan, E., Guo, H., Anderlid, B. M., Nordgren, A., Lindstrand, A., Vandeweyer, G., … Eichler, E. E. (2017). Hotspots of missense mutation identify neurodevelopmental disorder genes and functional domains. *Nature Neuroscience*, *20*(8), 1043–1051. https://doi.org/10.1038/nn.4589

Han, P. K. J. (2013). Conceptual, Methodological, and Ethical Problems in Communicating

Uncertainty in Clinical Evidence. *Med Care Res Rev*, *70*(10), 14S-36S. https://doi.org/10.1177/1077558712459361.Conceptual

Herrero, J., Muffato, M., Beal, K., Fitzgerald, S., Gordon, L., Pignatelli, M., Vilella, A. J., Searle, S. M. J., Amode, R., Brent, S., Spooner, W., Kulesha, E., Yates, A., & Flicek, P. (2016). Ensembl comparative genomics resources. *Database*, *2016*(Feb), bav096. https://doi.org/10.1093/database/bav096

Hubbard, E. J. A. (2014). FLP/FRT and Cre/lox recombination technology in C. elegans. *Methods*, *68*(3), 417–424. https://doi.org/10.1016/j.ymeth.2014.05.007

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

Iossifov, I., Ronemus, M., Levy, D., Wang, Z., Hakker, I., Rosenbaum, J., Yamrom, B., Lee, Y. H., Narzisi, G., Leotta, A., Kendall, J., Grabowska, E., Ma, B., Marks, S., Rodgers, L., Stepansky, A., Troge, J., Andrews, P., Bekritsky, M., … Wigler, M. (2012). De novo gene disruptions in children on the autistic spectrum. *Neuron*, *74*(2), 285–299. https://doi.org/10.1016/j.neuron.2012.04.009

Kim, S., Twigg, S. R. F., Scanlon, V. A., Chandra, A., Hansen, T. J., Alsubait, A., Fenwick, A. L., McGowan, S. J., Lord, H., Lester, T., Sweeney, E., Weber, A., Cox, H., Wilkie, A. O. M., Golden, A., & Corsi, A. K. (2017). Localized TWIST1 and TWIST2 basic domain substitutions cause four distinct human diseases that can be modeled in Caenorhabditis elegans. *Human Molecular Genetics*, *26*(11), 2118–2132. https://doi.org/10.1093/hmg/ddx107

Krumm, N., O'Roak, B. J., Shendure, J., & Eichler, E. E. (2014). A de novo convergence of

autism genetics and molecular neuroscience. *Trends in Neurosciences*, *37*(2), 95–105. https://doi.org/10.1016/j.tins.2013.11.005

Lackner, M. R., Kornfeld, K., Miller, L. M., Robert Horvitz, H., & Kim, S. K. (1994). A MAP kinase homolog, mpk-1, is involved in ras-mediated induction of vulval cell fates in Caenorhabditis elegans. *Genes and Development*, *8*(2), 160–173. https://doi.org/10.1101/gad.8.2.160

Lee, R. Y., Lobel, L., Hengartner, M., Horvitz, H. R., & Avery, L. (1997). Mutations in the alpha1 subunit of an L-type voltage-activated Ca2+ channel cause myotonia in Caenorhabditis elegans. *Embo J*, *16*(20), 6066–6076.

Maeda, I., Kohara, Y., Yamamoto, M., & Sugimoto, A. (2001). Large-scale analysis of gene function in Caenorhabditis elegans by high-throughput RNAi. *Current Biology*, *11*, 171–176. http://www.wormbase.org/db/misc/paper?name=WBPaper00004651

Markaki, M., & Tavernarakis, N. (2010). Modeling human diseases in Caenorhabditis elegans. *Biotechnology Journal*, *5*(12), 1261–1276. https://doi.org/10.1002/biot.201000183

McDiarmid, T. A., Au, V., Loewen, A. D., Liang, J., Mizumoto, K., Moerman, D. G., & Rankin, C. H. (2018). CRISPR-Cas9 human gene replacement and phenomic characterization in *Caenorhabditis elegans* to understand the functional conservation of human genes and decipher variants of uncertain significance. *Disease Models & Mechanisms*, dmm.036517. https://doi.org/10.1242/dmm.036517

Miosge, L. A., Field, M. A., Sontani, Y., Cho, V., Johnson, S., Palkova, A., Balakishnan, B., Liang, R., Zhang, Y., Lyon, S., Beutler, B., Whittle, B., Bertram, E. M., Enders, A., Goodnow, C. C., & Andrews, T. D. (2015). Comparison of predicted and actual consequences of missense mutations. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(37), E5189–E5198. https://doi.org/10.1073/pnas.1511585112

Ng, P. C., & Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Research*, *31*(13), 3812–3814. https://doi.org/10.1093/nar/gkg509

Paix, A., Folkmann, A., Rasoloson, D., & Seydoux, G. (2015). High efficiency, homology-directed genome editing in Caenorhabditis elegans using CRISPR/Cas9 ribonucleoprotein complexes. *Genetics*, *201*(September), 47–54. https://doi.org/10.1534/genetics.115.179382

Petrucelli, N., Lazebnik, N., Huelsman, K. M., & Lazebnik, R. S. (2002). Clinical interpretation and recommendations for patients with a variant of uncertain significance in BRCA1 or BRCA2: A survey of genetic counseling practice. *Genetic Test*, *6*(2), 107–113. https://doi.org/10.1089/10906570260199357

Pilipiuk, J., Lefebvre, C., Wiesenfahrt, T., Legouis, R., & Bossinger, O. (2009). Increased IP3/Ca2+ signaling compensates depletion of LET-413/DLG-1 in C. elegans epithelial junction assembly. *Developmental Biology*, *327*(1), 34–47. https://doi.org/10.1016/j.ydbio.2008.11.025

Reble, E., Dineen, A., & Barr, C. L. (2018). The contribution of alternative splicing to genetic risk for psychiatric disorders. *Genes, Brain and Behavior*, *17*(3), 1–12. https://doi.org/10.1111/gbb.12430

Robinson, E. B., Samocha, K. E., Kosmicki, J. A., McGrath, L., Neale, B. M., Perlis, R. H., & Daly, M. J. (2014). Autism spectrum disorder severity reflects the average contribution of de novo and familial influences. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(42), 15161–15165. https://doi.org/10.1073/pnas.1409204111

Robison, A. J. (2014). Emerging role of CaMKII in neuropsychiatric disease. *Trends in Neurosciences*, *37*(11), 653–662. https://doi.org/10.1016/j.tins.2014.07.001

Rual, J.-F., Ceron, J., Koreth, J., Hao, T., Nicot, A., Hirozane-kishikawa, T., Vandenhaute, J., Orkin, S. H., Hill, D. E., & Vidal, M. (2004). Toward improving Caenorhabditis elegans

phenome mapping with an ORFeome-based RNAi library. *Genome Research*, *14*, 2162–2168. https://doi.org/10.1101/gr.2505604.7

Shen, Z., Zhang, X., Chai, Y., Zhu, Z., Yi, P., Feng, G., Li, W., & Ou, G. (2014). Conditional knockouts generated by engineered CRISPR-Cas9 endonuclease reveal the roles of coronin in c.elegans neural development. *Developmental Cell*, *30*(5), 625–636. https://doi.org/10.1016/j.devcel.2014.07.017

Sonnichsen, B., Koski, L. B., Walsh, A., Marschall, P., Neumann, B., Brehm, M., Alleaume, A.-M. M., Artelt, J., Bettencourt, P., Cassin, E., Sönnichsen, B., Koski, L. B., Walsh, A., Marschall, P., Neumann, B., Brehm, M., Alleaume, A.-M. M., Artelt, J., Bettencourt, P., … Echeverri, C. J. (2005). Full-genome RNAi profiling of early embryogenesis in Caenorhabditis elegans. *Nature*, *434*(7032), 462–469. https://doi.org/10.1038/nature03353

Sze, J. Y., Victor, M., Loer, C., Shi, Y., & Ruvkun, G. (2000). Food and metabolic signalling defects in a Caenorhabditis elegans serotonin-synthesis mutant. *Nature*, *403*(6769), 560–564. https://doi.org/10.1038/35000609

Tennessen, J. A., Bigham, A. W., O'Connor, T. D., Fu, W., Kenny, E. E., Gravel, S., McGee, S., Do, R., Liu, X., Jun, G., Kang, H. M., Jordan, D., Leal, S. M., Gabriel, S., Rieder, M. J., Abecasis, G., Altshuler, D., Nickerson, D. A., Boerwinkle, E., … on behalf of the NHLBI Exome Sequencing Project. (2012). Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science*, *337*(6090), 64–69. https://doi.org/10.1126/science.1219240

Voutev, R., & Hubbard, E. J. A. (2008). A "FLP-out" system for controlled gene expression in Caenorhabditis elegans. *Genetics*, *180*(1), 103–119. https://doi.org/10.1534/genetics.108.090274

Walsh, N., Kenney, L., Jangalwe, S., Aryee, K., Greiner, D. L., Brehm, M. A., Shultz, L. D., & Harbor, B. (2017). Humanized mouse models of clinical disease. *Annual Review of*

Pathology, *24*(12), 187–215. https://doi.org/10.1146/annurev-pathol-052016-100332.Humanized

Yates, A., Beal, K., Keenan, S., McLaren, W., Pignatelli, M., Ritchie, G. R. S., Ruffier, M., Taylor, K., Vullo, A., & Flicek, P. (2015). The Ensembl REST API: Ensembl Data for Any Language. *Bioinformatics*, *31*(1), 143–145. https://doi.org/10.1093/bioinformatics/btu613

Yemini, E., Jucikas, T., Grundy, L. J., Brown, A. E. X., & Schafer, W. R. (2013). A database of Caenorhabditis elegans behavioral phenotypes. *Nature Methods*, *10*(9), 877–879. https://doi.org/10.1038/nmeth.2560

2.8 Tables and figures

Table 2-1. Strain information.

20 ASD-associated missense alleles and their corresponding mutant strains in *C. elegans.*

Human gene	Human cDNA change*	Human protein change	Inheritance pattern	*C. elegans* gene (allele)	*C. elegans* protein change	Strain name
CACNA1D	c.1105G>A	V369M	unknown	*egl-19(sy849)*	V331M	PS7085
CACNA1D	c.1112A>C	Y371S	unknown	*egl-19(sy850)*	Y333S	PS7156
CHD7	c.2986G>A	G996S	*de novo*	*chd-7(sy861)*	G1225S	PS7293
CHD7	c.3770T>G	L1257R	*de novo*	*chd-7(sy855)*	L1487R	PS7317
CHD8	c.2501T>C	L834P	*de novo*	*chd-7(sy859)*	L1220P	PS7318
CHD8	c.494C>T	P165L	unknown	*chd-7(sy1049)*	P253L	PS7267
CUL3	c.2156A>G	H719R	de novo	*cul-3(sy874)*	H728R	PS7387
DLG4	c.2281G>A	V761I	unknown	*dlg-1(sy872)*	V964I	PS7343
GLRA2	c.407A>G	N136S	*de novo*	*avr-15(sy873)*	N347S	PS7384
GLRA2	c.458G>A	R153Q	*de novo*	*avr-15(sy851)*	R364Q	PS7257
MAPK3	c.833G>A	R278Q	*de novo*	*mpk-1(sy870)*	R332Q	PS7382
NAA15	c.1319T>C	L440S	familial	*hpo-29(sy877)*	L575S	PS7394
PTEN	c.66C>G	D22E	familial	*daf-18(sy879)*	D66E	PS7439
PTEN	c.208C>G	L70V	unknown	*daf-18(sy887)*	L115V	PS7432
PTEN	c.278A>G	H93R	*de novo*	*daf-18(sy881)*	H138R	PS7436
PTEN	c.369C>G	H123Q	unknown	*daf-18(sy885)*	H168Q	PS7430
PTEN	c.392C>T	T131I	*de novo*	*daf-18(sy882)*	T176I	PS7434
SYNGAP1	c.698G>A	C233Y	*de novo*	*gap-2 (sy889)*	C417Y	PS7433
SYNGAP1	c.1288C>T	L430F	familial	*gap-2(sy886)*	L660F	PS7457
TPH2	c.674G>A	R225Q	familial	*tph-1(sy878)*	R259Q	PS7395

***** The virtual cDNA was provided by the SFARI database.

Table 2-2. Morphological phenotypes of missense alleles and their controls

Gene	Length (μm)	Width (μm)	Area (μm^2)
N2	1105 ± 5	87.8 ± 0.7	98798 ± 1147
avr-15(N347S)	1101 ± 7	87.9 ± 1.5	98214 ± 1273
avr-15(R364Q)	989 ± 7 *	77.3 ± 0.9 *	77860 ± 1457 *
avr-15(ad1051)	1033 ± 10 *	78.5 ± 0.8 *	82468 ± 1410 *
chd-7(P253L)	1110 ± 9	78.9 ± 0.9 *	86572 ± 1815 *
chd-7(L1220P)	978 ± 10 *	76.2 ± 1.0 *	75909 ± 1432 *
chd-7(G1225S)	1038 ± 11 *	80.8 ± 1.3 *	85400 ± 1977 *
chd-7(L1487R)	1033 ± 13 *	81.7 ± 1.3 *	86060 ± 2277 *
chd-7(sy956)	954 ± 6 *	75.4 ± 0.9 *	73201 ± 1254 *
cul-3(H728R)	1111 ± 9	96.1 ± 2.8 *	108771 ± 3865 *
daf-18(D66E)	1125 ± 14	91.5 ± 1.1	104804 ± 2496
daf-18(L115V)	1128 ± 9	94.2 ± 1.1	108208 ± 1902
daf-18(H138R)	1130 ± 6	94.6 ± 1.2	108706 ± 1366
daf-18(H168Q)	1148 ± 6	103.3 ± 2.7 *	120936 ± 3535 *
daf-18(T176I)	1135 ± 9	93.1 ± 1.4	107395 ± 1711
dlg-1(V964I)	1129 ± 4	83.16 ± 0.9	95502 ± 1210
egl-19(V331M)	1082 ± 4	83.1 ± 1.1	91504 ± 1448
egl-19(Y333S)	1052 ± 8 *	81.2 ± 0.7 *	86896 ± 1310 *
egl-19(n2368sd)	639 ± 10 *	68.1 ± 0.8 *	44402 ± 1104 *
gap-2(C417Y)	1128 ± 14	98.7 ± 2.6 *	113351 ± 3713 *
gap-2(L660F)	1120 ± 13	89.0 ± 1.7	101634 ± 3020
hpo-29(L575S)	1044 ± 12 *	94.0 ± 2.4	99849 ± 3111
mpk-1(R332Q)	1145 ± 7	88.0 ± 1.7	104188 ± 1889
tph-1(R259Q)	1026 ± 9 *	84.5 ± 1.1	88238 ± 1151 *
tph-1(mg280)	993 ± 20 *	80.9 ± 1.3 *	81918 ± 2765 *

All values are presented as mean ± SEM.

*$p < 0.01$ via one-way ANOVA and multiple comparison

Table 2-3. Comparison of behavioral results to software prediction

Human gene	*C. elegans* gene	PolyPhen-2[a]	(1-SIFT)[a]	Phenotype
CACNA1D(V369M)	*egl-19(V331M)*	0.995	0.99	no
CACNA1D(Y371S)	*egl-19(Y333S)*	1	1	morphology changes
CHD7(G996S)	*chd-7(G1225S)*	0.998	1	morphology changes, locomotion variants, reduced fecundity
CHD7(L1257R)	*chd-7(L1487R)*	1	1	morphology changes, locomotion variants, reduced fecundity
CHD8(L834P)	*chd-7(L1220P)*	1	1	morphology changes, locomotion variants, reduced fecundity
CHD8(P165L)	*chd-7(P253L)*	0.996	0.86	morphology changes
CUL3(H719R)	*cul-3(H728R)*	1	0.96	morphology changes, reduced fecundity
DLG4(V761I)	*dlg-1(V964I)*	0.001	0.84	reduced fecundity
GLRA2(N136S)	*avr-15(N347S)*	0.979	1	locomotion variants
GLRA2(R153Q)	*avr-15(R364Q)*	0.997	1	locomotion variants
MAPK3(R278Q)	*mpk-1(R332Q)*	0.997	1	no
NAA15(L440S)	*hpo-29(L575S)*	0.999	0.96	morphology changes, locomotion variants
PTEN(D22E)	*daf-18(D66E)*	0.297	0.93	no
PTEN(L70V)	*daf-18(L115V)*	0.999	1	no
PTEN(H93R)	*daf-18(H138R)*	1	0.97	locomotion variants
PTEN(H123Q)	*daf-18(H168Q)*	1	1	morphology changes, locomotion variants
PTEN(T131I)	*daf-18(T176I)*	1	0.82	no
SYNGAP1(C233Y)	*gap-2(C417Y)*	0.94	1	morphology changes, locomotion variants
SYNGAP1(L430F)	*gap-2(L660F)*	1	1	no

TPH2(R225Q)	*tph-1(R259Q)*	0.162	0.92	morphology changes, locomotion variants

[a] PolyPhen-2 and SIFT prediction score were based on human sequence (*: 1=probably damaging)

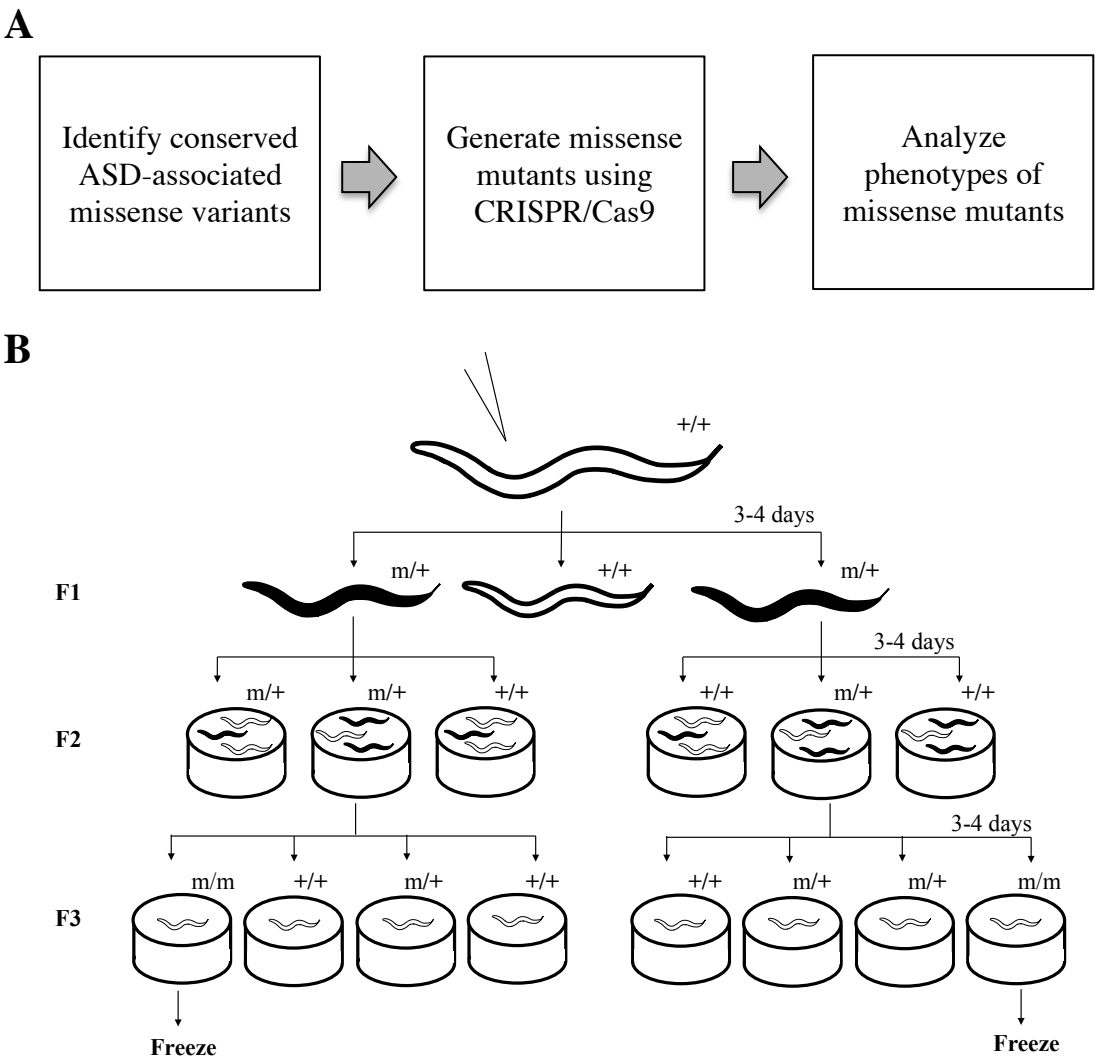

Figure 2-1. Generation of ASD-associated missense variants in *C. elegans*.
(A) Experimental pipeline. First, we used bioinformatics to identify human ASD-associated missense variants conserved in *C. elegans* protein sequences. We generated the *C. elegans* missense mutants using CRISPR/Cas9 and homology-directed knock-in technique. We then analyzed the effects of these autism-associated alleles by comparing observable phenotypes from these alleles to both wild-type and known loss-of-function mutant controls. (B) Mutant strain screening process. Three days after injection (P0), F1 offspring expressing co-conversion marker, such as dumpy (black), were selected into individual plates. We genotyped the F2 offspring to identify the heterozygous plates. For each heterozygous plate containing the successful knock-in target missense residue, we randomly selected 16-20 F2 offspring and separated them into individual plates (F3). A second round of genotyping process was then conducted to identify homozygous plates, which were cryo-preserved for future studies.

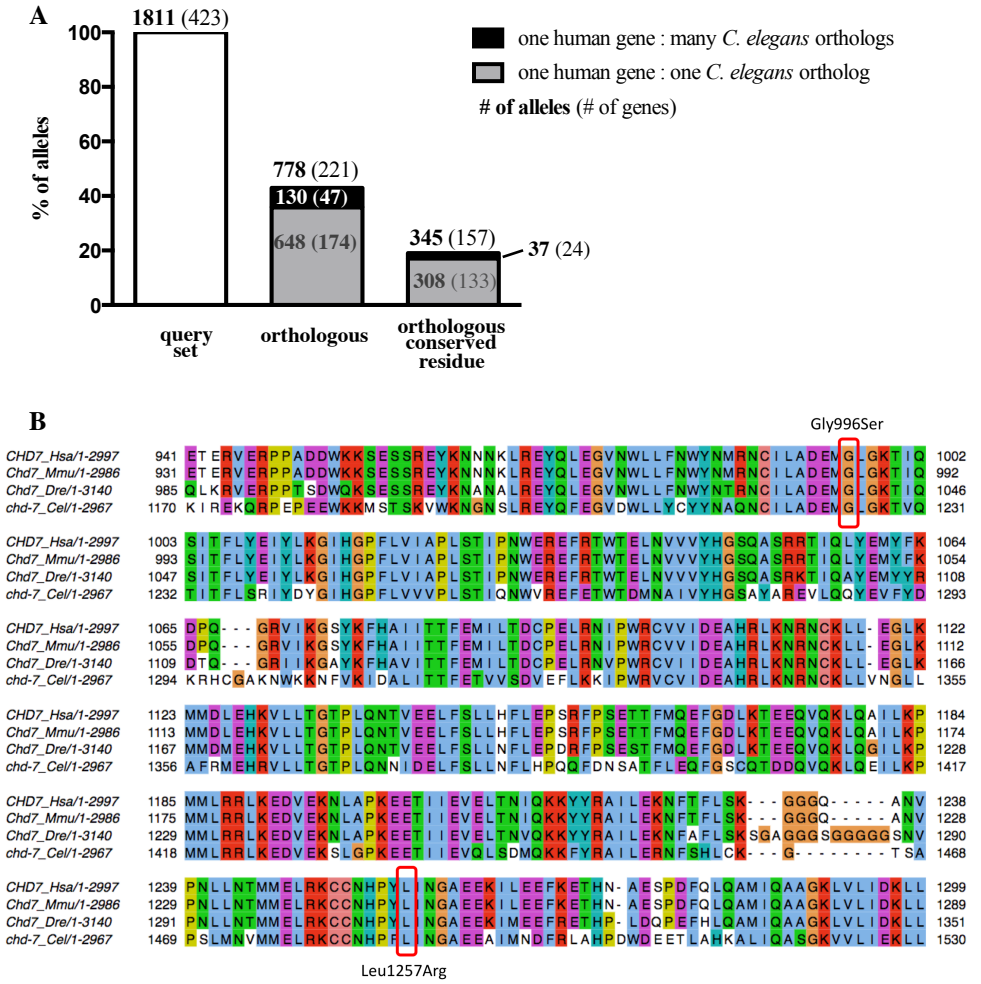

Figure 2-2. Detection of conserved ASD-associated missense residues in *C. elegans*.
(A) For each missense allele, the *C. elegans* ortholog of the corresponding human gene was identified using the Ensembl Compara method. Each Ensembl ortholog pair was underpinned by a protein multiple-sequence alignment, which can be used to identify the putative orthologous *C. elegans* amino acid for a given human amino acid. 43% of the human missense variants had at least one *C. elegans* ortholog. 130 of the 778 orthologous missense variants had more than one *C. elegans* ortholog (black bar). Overall, only 19% of the missense residues were orthologous and conserved in *C. elegans*. (B) Section of the protein multiple alignment for human (*Hsa*) *CHD7* and its orthologs in mouse (*Mmu*), zebrafish (*Dre*) and *C. elegans (Cel)*. Residues in the alignment have been colored by JalView (PMID:19151095) using the Clustal X coloring scheme. Circled are two ASD-linked missense variants in *CHD7* that occur in a highly conserved region across all the species.

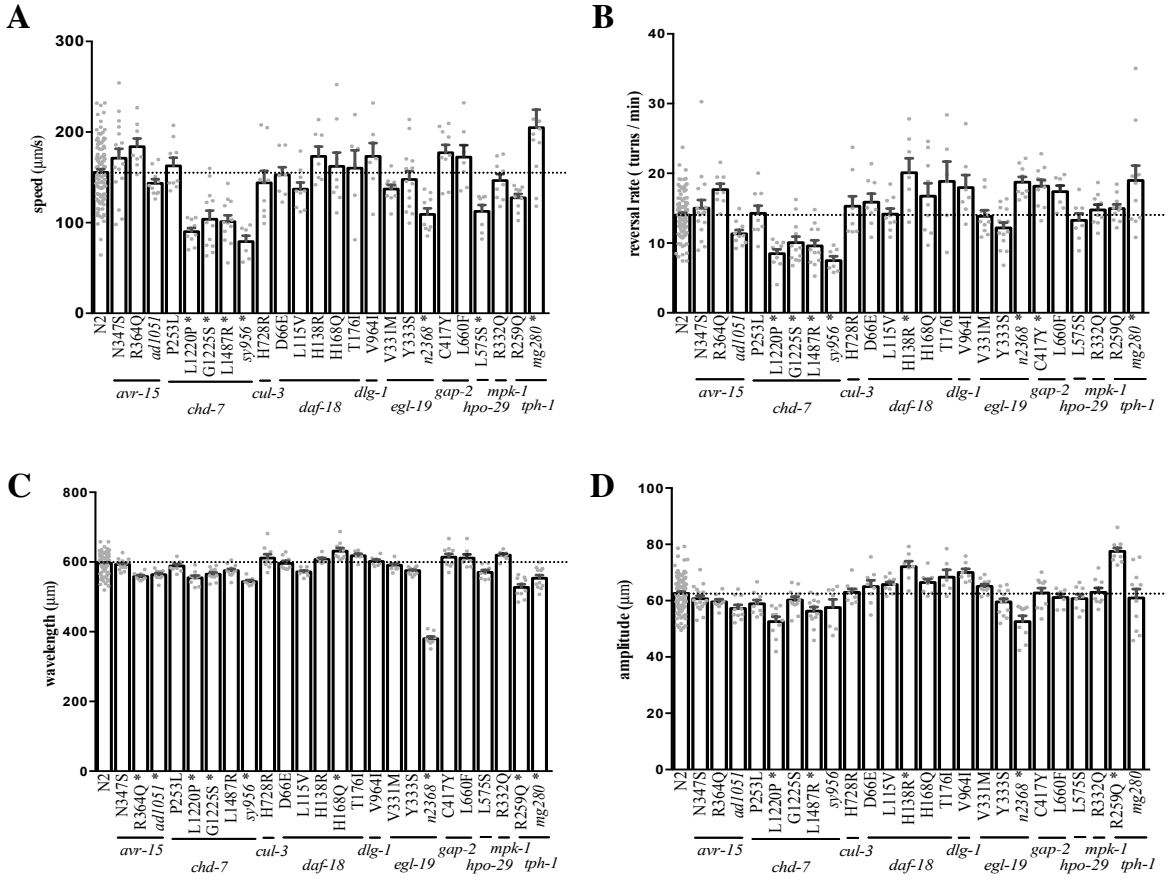

Figure 2-3. Locomotion phenotypes of missense alleles and their controls.
(A) Speed and (B) reversal rate are measurement of locomotion while (C) wavelength and (D) amplitude represent sinusoidal shape of movement. *avr-15(R364Q)* mutant showed decreased wavelength. Most *chd-7* missense mutants displayed a weaker version of the null mutant phenotypes in all measurements. *daf-18(H138R)* and *daf-18(H168Q)* exhibited larger sinusoidal wavelength and amplitude. *gap-2(C417Y)* showed higher reversal rate. *hpo-29(L575S)* displayed slower speed. *tph-1(R259Q)* exhibited changes in sinusoidal movement. Bars were presented as the mean ± SEM. Each dot represented the average of one plate containing 8-10 worms. *: $p <$ 0.01 via one-way ANOVA and multiple comparison to wild-type. Horizontal line indicated the mean of wild-type.

Figure 2-4. Fecundity phenotype of missense alleles and their controls.
Most *chd-7* mutants displayed decreased fecundity. The defects were more subtle than its deletion*(tm6139)* or frameshift*(sy956)* null controls. *cul-3(H728R)* and *dlg-1(V964I)* also showed decrease in fecundity. Each dot represented total number of living larvae from one animal. Approximately 20 animals were tested in each strain. Wild-type (N2) and its median value (dotted vertical line) was shown in blue. Mutants significantly different (sig) from wild-type were shown in red. $p < 0.01$/(total test number) via non-parametric bootstrap analysis.

Chapter 3

FUNCTIONAL SCREENING OF ASD-ASSOCIATED MISSENSE VARIANTS IN C. ELEGANS

3.1 Abstract

Evolving next-generation sequencing technology has accelerated the identification of disease-associated genetic variants. However, interpretation of these variants remains challenging, especially for genetic variants with subtle effects such as missense variants. There is a dire need for more comprehensive and efficient methods to test these variants' functional consequences *in vivo*. Therefore, we utilized a previously developed *in vivo* screening platform using *Caenorhabditis elegans* to systematically evaluate the functional consequences of disease-associated missense variants. This study expanded on the previous collections and generated 28 more autism-associated missense variants in their *C. elegans* orthologous genes, focusing on genes in synaptic function and gene regulatory networks. We characterized these missense mutant strains on their morphology, fecundity, and locomotor functions. Our results demonstrated that 18 of the 28 missense variants displayed phenotypic changes in one of the functions. Missense mutant strains in different gene clusters also showed distinct phenotypic profiles. Specifically, missense variants in the gene regulatory cluster were more likely to influence fecundity whereas missense variants in the synaptic gene cluster had more of an impact on morphology and locomotor patterns. Overall, we prioritized 18 functional changing missense variants for future studies on disease mechanisms and point out the general phenotypic assays for detecting gene network-specific functional impacts.

3.2 Introduction

The advance of next-generation sequencing has facilitated the rapid discovery of disease-associated genetic variants and adequate interpretation of sequence results has become a new challenge. A series of studies conducted whole-exome sequencing on the samples of Simons Simplex Collection and revealed a vast amount of gene-disruptive *de novo* mutations (nonsense, splice-site, or frameshift), *de novo* missense variants, and transmitted nonsense mutations in individuals with autism spectrum disorder (ASD) (Iossifov et al., 2012; Neale et al., 2012; O'Roak et al., 2012; Sanders et al., 2012). Another recent project done by the Autism Sequencing Consortium (ASC) also discovered 7,131 *de novo* variants in protein-coding regions, among which 63.1% were missense variants (Satterstrom et al., 2020). Indeed, *de novo* missense mutations, particularly the most damaging types, were more enriched in ASD probands than in their unaffected siblings, and these missense variants contribute to at least 10% of ASD diagnoses (Iossifov et al., 2014). Despite the rapid discovery of missense variants in genetic testing, only 2% of the missense variants have a clinical annotation in the NCBI ClinVar database (Landrum et al., 2014; Lek et al., 2016). Even for those missense variants annotated in the database, over half of them showed unknown or controversial clinical outcomes, so-called variants of uncertain significance (VUS) (Starita et al., 2017). Therefore, there is a dire necessity for a better system to distinguish missense variants with functional impacts.

Hundreds of risk genes have been implicated in ASD (Coe et al., 2019; Iossifov et al., 2012; Kaplanis et al., 2020; Neale et al., 2012; O'Roak et al., 2012; Sanders et al., 2012; Satterstrom et al., 2020). Most of these genes are involved in multiple physiological functions (Lord et al., 2020), but they converge to two major gene networks: genes in gene expression regulation and genes in synaptic functions (Ben-David & Shifman, 2013; Chaste et al., 2017; Gilman et al., 2011; Sakai et al., 2011; Voineagu et al., 2011). Chromatin modification controls cell fates and function, and misregulation, particularly during early fetal development, may

impair neurodevelopment and lead to cognitive deficits (van Bokhoven, 2011; Jakovcevski and Akbarian, 2013; Ronan et al., 2013). Dozens of chromatin remodeling genes have been implicated in ASD, such as histone demethylase and fragile X mental retardation protein (FMRP) (De Rubeis et al., 2014; Parikshak et al., 2013). Notably, these chromatin modifier genes harboring missense and protein disrupting *de novo* mutations in a dosage-sensitive manner, reflecting the heterozygous state observed in ASD probands (Parikshak et al., 2013). In addition, many ASD risk genes have been characterized as synaptic genes. The complex and dynamic synaptic structures contain thousands of proteins, such as NRX1, SHANK3, MAPK3, etc. These proteins regulate the formation and maturation of synapse, spine structural remodeling and morphogenesis, and the formation of neural circuits. Malfunction of synaptic proteins may result in autistic phenotypes, and disruption in the functional molecular network has been linked to many neurodevelopmental disorders. Furthermore, Zhao et al. identified a set of synaptic-enriched genes containing broad enhancer-like-chromatin domains (BELDs), and these BELDs genes were susceptible to the disruption of ASD-associated chromatin regulators (Zhao et al., 2018). Together, this evidence pointed to the combinatory effect of gene regulatory and synaptic functions in ASD.

To better predict the pathogenicity of disease-associated missense variants, several computational inference tools have been developed. These *in silico* tools integrate a variety of predicting features, such as evolutionary sequence conservation, protein structure, regulatory effect, and gene ontology, to predict the likelihood of a given missense variant as deteriorating (disease-causing) or benign (tolerant). Despite considerable efforts in developing various computational tools, these prediction algorithms were only 60-82% accurate, for their accuracy is affected by the proximity with common SNPs, sequencing errors, or lack of evidence of pathogenicity (Bell et al., 2011; Miosge et al., 2015; Sun et al., 2016; Thusberg et al., 2011; Wong et al., 2019). In most cases, *in silico* methods tend to overpredict pathogenic missense

changes and give low specificity (Choi et al., 2012). The *in silico* prediction tool is even less accurate in variants with milder effects, such as missense variants (Richards et al., 2015).

To overcome the omission and false predictions from computational inference tools, both *in vitro* and *in vivo* platforms have been developed to characterize functional consequences of missense variants. Most high throughput functional assays study the impacts of missense variants using cell-based systems. For example, Gasperini et al. and Chen et al. utilized yeast to detect missense variants disrupting protein biochemical functions, protein-protein interactions, splicing, and RNA stability (Chen et al., 2018; Gasperini et al., 2016). These *in vitro* pipelines screen for functionally significant missense variants at the scale of 10^4 -10^6 variants per experiment and prioritized interactome-disrupting genes (i.e., risk genes with many interaction partners, so-called hub genes) (Chen et al., 2018; Gasperini et al., 2016). However, these cell-based functional assays take proteins out of their endogenous genomic and cellular contexts, which can neglect the effects caused by cell-type-specific chromatin states (Ernst & Kellis, 2015; Starita et al., 2017). Therefore, multiple studies have attempted to evaluate the functional consequences of missense variants *in vivo*. Recently, Jin et al. utilized lentivirus to introduce frameshift mutations in 35 ASD risk genes *in utero* and observed the cell-type-specific functional impacts in the neural development of offspring is consistent with single-cell data from ASD patients (Jin et al., 2020). Likewise, Wertz et al. conducted the first unbiased genome-wide knock-down screening in the mammalian central nervous system using a pooled lentiviral library with short hairpin RNA (shRNA) and CRISPR, which led to the discovery of essential neuronal genes, cellular processes, and novel druggable candidate genes in neurological disorders (Wertz et al., 2020). In addition, several groups introduced the ASD-associated missense variant into its orthologous gene in invertebrate model organisms, such as *Caenorhabditis elegans* (*C. elegans*), and prioritized variants that impact sensory and neurodevelopmental phenotypes (Gonzalez-cavazos et al., 2019; McDiarmid et al., 2018, 2020; Post et al., 2020; Wong et al., 2019).

In a previous study, we developed an *in vivo* pipeline to screen conserved ASD-associated missense variants for functionally impactful residue changes (Wong et al., 2019). This study expands on previous research (Gonzalez-cavazos et al., 2019; McDiarmid et al., 2020; Wong et al., 2019, 2021) and examined 28 more ASD-associated missense variants in their corresponding loci in *C. elegans* orthologs. We evaluated the functional consequences of the animals bearing ASD-associated missense mutations at the system level (whole body change, involving many assays and functions). We found that 18 of the 28 missense mutations resulted in phenotypic changes in the fecundity, morphology and locomotion. In particular, missense mutations in the chromatin remodeling and nucleotide cycle gene clusters were more likely to affect embryonic germline development, as indicated by the fecundity assay. Comparatively, missense mutations in the synaptic development and regulation gene subsets impacted both survival (fecundity, morphology) and movement (locomotion). In addition, whole-animal functional evaluation revealed 8 missense variants displayed phenotypic changes in multiple functional assays, indicating impacts on broader circuit and tissue types. Overall, we identified 18 novel functional changing missense variants, and 8 of them caused system-level changes. We prioritized these phenotypic changing missense variants for future studies on functional validation and disease mechanisms.

3.3 Materials and methods

Strains

The Bristol N2 *C. elegans* strain was used as the wild-type control and background for all CRISPR experiments (Brenner, 1974). The control strains for functional assays were obtained from lab stock, the Caenorhabditis Genetics Center (CGC), and the National BioResource Project- *C. elegans* (NBRP). All strains were maintained on nematode growth medium (NGM) agar plates seeded with *Escherichia coli* OP50 at room temperature (21 ± 1◦C).

Generation of missense mutant strains

ASD-associated missense variants were extracted from the SFARI database–Human Gene Module (Fischbach & Lord, 2010). Ensembl compara were used to generate protein multiple alignment and determine the conserved residues. Targeted missense variants were introduced into the *C. elegans* genome using CRISPR/Cas9 and homology-directed repair as described before (Wong et al., 2019). Additional restriction enzyme sites were also introduced via synonymous mutations to facilitate the screening process. Mutant genotypes were determined by the correct PCR and digested products length, and then confirmed by commercial Sanger sequencing company (Laragen, Culver City, CA). When available, we saved two independent missense mutant lines. While there are little to no off-targets effects of Cas9 (Chiu et al., 2013), *C. elegans* N2 suffers approximately one mutation per generation so it was useful to have more than one strain for each locus.

Functional assays

In the fecundity assay, well-fed *C. elegans* were synchronized at the L4 stage. Individual L4 hermaphrodites were placed on separate NGM plates seeded with *E. coli* OP50, and these animals were subsequently transferred to a new plate every day. The number of newly hatched larvae progeny was counted for every plate one day after the hermaphrodite was transferred. The total fecundity consisted of the sum of living progeny produced for four days per hermaphrodite. In the morphology and locomotion tracking assay, 6 young adults were picked on NGM plates freshly seeded with a 50-µL drop of a saturation-phase culture of *E. coli* OP50. The animals were given 30 min for habituation and then tracked locomotor activity for 4 min using the automated WormLab tracking system (MBF Bioscience, Williston, VT). The camera was a Nikon AF Micro

60/2.8D with zoom magnification. A 2456×2052–resolution, 7.5-f.p.s. camera with a magnification that results in 8.2 μm per pixel and an FOV of roughly 2 × 2 cm^2 was used.

Statistical analysis

The fecundity assay was analyzed using a non-parametric bootstrap analysis. Initially, the two datasets were mixed, samples were selected at random with replacement from the mixed population into two new datasets, and then the difference in the averages of these new datasets was calculated; this process was iterated 10^6 times. We reported the *p*-value as the probability when the difference in the average of simulated datasets was greater than the difference in the average of the original datasets. If $p < 0.01$/(total testing number), we rejected the null hypothesis that the average values of the two datasets were not equal to each other.

Morphology and locomotion were analyzed by one-way ANOVA analysis in the scipy module of the Python 3.7 program. Tukey-paired multiple comparisons were performed between the wild-type group and mutant strains. The significant level was defined as $p < 0.01$. Approximately 10 plates were tracked per strain. The mean of each plate was first calculated, and the total mean of all plates of the same genotype was computed. Due to the laboratory safety practices during COVID-19, a second batch of experiments were tested in a different room, with 1-2°C increase in room temperature. This temperature difference has resulted in variations in some movement-related parameters, namely speed, reversal rate, and maximum amplitude. Therefore, for these three parameters, we compared the values of mutant strain with the values of wild-type animals recorded in the same batch (experimental conditions).

3.4 Results

Hundreds of genes have been implicated in ASD. These genes participate in diverse physiological functions and can be categorized into two major functions: gene regulation and

synaptic function (Satterstrom et al., 2020). In a previous study in our laboratory, we established a working pipeline to identify conserved human disease-associated missense variants and screen for functionally impactful variants (Wong et al., 2019). As can be seen in Figure 1, we identified 345 conserved missense residues from thousands of missense variants documented in Simon Foundation Autism Research Initiative (SFARI) database. We sampled 126 of the 345 conserved residues and generated missense mutations into the genome of *C. elegans* using CRISPR/Cas9 and homology-directed repair. We characterized these ASD-associated missense variants in a series of assays, targeting their neurodevelopmental functions. This study expanded on the previous collections (Gonzalez-cavazos et al., 2019; McDiarmid et al., 2020; Wong et al., 2019, 2021) and studied more ASD-associated missense mutants in genes involved in synaptic and gene regulatory functions.

In this study, we sampled 28 ASD-associated missense variants in 19 genes and studied their impacts on embryonic survival and locomotor activities. We classified these genes in two clusters: gene regulation and expression as well as synaptic function (Figure 2). The human genes (listed along with their *C. elegans* orthologs) involved in gene expression and regulation are: *TBR1(tbx-8), PAX6(vab-3)* in transcriptional regulation; *BCL11A(bcl-11), ELAVL3(exc-7), EP400(ssl-1), KDM6B(utx-1),* and *KMT2C(set-16)* for chromatin remodeling; *ADSL(adsl-1), AMPD1(ampd-1)* in nucleotide cycle. The genes affecting synaptic functions include *MAPK3(mpk-1), P4HA2(dpy-18), SPARCL1(ost-1), TRIO(unc-73)* in synaptic transmission; *SHANK3(shn-1), NRXN1(nrx-1), SLC6A1(snf-11), EFR3A(efr-3)* in synaptic development and regulation; and *ATP2B2(mca-3)* and *GNAS(gsa-1)* in calcium signaling. The details of the 29 missense variants and their corresponding loci in its *C. elegans* orthologs can be found in Table 1.

To study the functional impacts of these ASD-associated missense mutations on embryonic/germline development, we conducted a fecundity assay, which measures the number

of viable progeny per hermaphrodite (Figure 3). The *C. elegans* germline development is a highly sensitive process, which reflects subtle changes in sensory perception, feeding behavior, and metabolism (Hubbard et al., 2013). We used the *C. elegans* fecundity assay as a screening tool to identify functional changing missense mutations, as defects in germline development will result in a decreased brood size. For the cluster of genes involved in gene regulation and expression, our results showed impaired fecundity in five missense mutants in the subset of genes in chromatin remodeling. Compared to N2 controls, missense mutant *bcl-11(C112F), exc-7(*Y85F), *ssl-1(G1301R)*, *set-16(P967R)*, and *utx-1(A976T)* showed decreased fecundity (Median: N2=241.0, *bcl-1(C112F)* = 61.0, *exc-7(Y85F)* = 162.5, *ssl-1(G1301R)* = 187.0, *set-16(P967R)* = 147.0, *utx-1(A976T)* = 173.0; $p < 0.0001$ for all mutants). Interestingly, missense mutants in two genes involved in the purine nucleotide cycle showed the opposite effect on fecundity. The orthologous missense mutation in the adenylosuccinate lyase (ADSL) gene, which participates in the purine biosynthesis, resulted in increased fecundity whereas one of the orthologous missense mutations in the adenosine monophosphate (AMP) deaminase, which recycles AMP, led to decreased fecundity (Camici et al., 2018) (Median: N2 = 241.0, *adsl-1(E76D)* = 252.0, *ampd-1(R566C)* = 181.5; $p < 0.0001$ for both mutants). As for genes in the synaptic gene cluster, all the fecundity phenotype changing missense mutations belonged to the synaptic development and regulation gene subset. Both missense mutations in the *C. elegans* ortholog of *SPARCL1* exhibited a reduction in fecundity (Median: N2=241.0, *ost-1(T142M)* = 139.0, *ost-1(W249R)* = 58.0 ; $p < 0.0001$ for both mutants). One of the missense mutants in the *TRIO* orthologous gene, *unc-73(V2062L)*, displayed a reduced fecundity phenotype whereas two other *unc-73* mutants we sampled showed normal fecundity function (Median: N2=241.0; *unc-73(V2062L)* = 177.0, $p=5*10^{-6}$; *unc-73(G1930S)* = 247.0; *unc-73(K2046R)* = 243.0). Additionally, one missense mutation in the *C. elegans* orthologous gene of *MAPK3*, showed an increase in fecundity (Median: *N2*=241.0, *mpk-1(P410S)* = 292.5; $p = 1.9x10^{-4}$).

To evaluate the effects of the missense mutations, an automated tracking system was used to evaluate changes in morphology and movement patterns. We first examined the morphology of the ASD-associated missense mutant strains using three parameters: length, width, and area (Figure 4, Table 2). We found the missense mutants in the gene expression and regulation cluster did not display any statistically significant differences from the morphology of the wild-type. In contrast, the gene group associated with synaptic function showed some morphological changes when compared to N2 wild-type animals. In particular, four missense mutant strains, *ost-1(W249R)*, *mpk-1(P410S)*, *dpy-18(Y413C),* and *mca-3(T763M)* in the synaptic development and calcium signaling clusters appeared to be more dumpy, indicated by decreases in length and area (Length: N2 = 1109 ± 5, *ost-1(W249R)* = 937 ± 17, *mpk-1(P410S)* = 991 ± 9, *dpy-18(Y413C)* = 914 ± 11, *mca-3(T763M)* = 1051 ± 7; $p = 0.001$ for all strains. Area: N2 = 100 301 ± 1152, *ost-1(W249R)* = 77 431 ± 2731, *mpk-1(P410S)* = 86 549 ± 822, *dpy-18(Y413C)* = 851 967 ± 2721, *mca-3(T763M)* = 91 434 ± 1761; $p = 0.001$ for all strains). Additionally, *efr-3(G256A)* displayed an increase in width (N2 = 88.8 ± 0.8, *efr-3(G256A)* = 98.1 ± 2.0; $p = 0.0299$) and area (N2 =100301 ± 1152, *efr-3(G256A)* = 114918 ± 3385; $p = 0.0058$). In conclusion, missense mutations in genes for synaptic function had significant impacts on the dimensional measurements of the organisms while those in the gene regulation and expression cluster had no effect on morphology.

To quantify coordination of the movement, we analyzed three parameters representing the sinusoidal waveform patterns of the worm, namely wavelength, mean amplitude, and max amplitude (Figure 4, Table 3). In the gene cluster of gene regulation and expression, four missense mutant strains, *tbx-8(K31E)*, *exc-7(Y85F)*, *ampd-1(R566C)*, and *ampd-1(P638S)*, displayed more curvy sinusoidal waveforms, showing as decreased wavelengths and increased mean amplitudes (Wavelength: N2 = 595.6 ± 4.5, *tbx-8(K31E)* = 558.1 ± 6.1, *exc-7(Y85F)* = 544.3 ± 9.4, *ampd-1(R566C)* = 553.6 ± 7.0, *ampd-1(P638S)* = 562.6 ± 4.1; $p < 0.001$ for all mutants. Mean amplitude: N2 = 60.0 ± 0.8, *tbx-8(K31E)* = 80.4 ± 1.8, *exc-7(Y85F)* = 81.9 ± 2.3,

ampd-1(R566C) = 91.8 ± 2.6, *ampd-1(P638S)* = 82.8 ± 0.9; p < 0.001 for all mutants). No significant difference was detected in maximum amplitude in any gene regulation-related mutants. With regards to the gene group that controls synaptic function, two missense mutants, *efr-3(D497G)* and *mca-3(T763M)*, showed more curvy sinusoidal waveforms (Wavelength: N2 = 595.6 ± 4.5, *efr-3(D497G)* = 561.3 ± 2.0, *mca-3(T763M)* = 533.9 ± 3.7; mean amplitude: N2 = 60.0 ± 0.8, *efr-3(D497G)* = 82.2 ± 1.3, *mca-3(T763M)* = 72.1 ± 1.7; $p = 0.0001$ for all). The *C. elegans* ortholog of *TRIO*, *unc-73(G1930S)*, showed an increase in mean amplitude with respect to wild-type (N2 =60.0 ± 0.8, *unc-73(G1930S)* = 85.3 ± 1.3; $p = 0.001$). Other than those genes, four missense mutant strains exhibited changes in the waveform parameters that were consistent with their morphological changes. For example, the decreased wavelength and maximum amplitude phenotypes observed in the *dpy-18(Y413C)* and *ost-1(W249R)* missense mutants may reflect their dumpy morphology (Wavelength: N2 = 595.6 ± 4.5, *dpy-18(Y413C)* = 522.0 ± 7.7, *ost-1(W249R)* = 493.6 ± 9.1; maximum amplitude: N2 = 204.9 ± 3.0, *dpy-18(Y413C)* = 167.9 ± 3.0, *ost-1(W249R)* = 171.2 ± 2.5; $p = 0.001$ for all). Similarly, *efr-3(G256A)* and *mpk-1(P410S)* exhibited changes in amplitude, which may also be dependent upon their morphological changes (Mean amplitude: N2 = 60.0 ± 0.8, *efr-3(G256A)* = 89.1 ± 1.8; $p = 0.001$. Maximum amplitude: N2 = 204.9 ± 3.0, *mpk-1(P410S)* = 184.3 ± 3.1, $p = 0.0082$). In general, we detected six missense mutant strains in both gene regulation and synaptic function clusters displayed more curvy sinusoidal waveform patterns. We also observed some small changes in waveform patterns that were independent of the morphology.

To study the effects of missense mutations on locomotor activity, we measured speed and reversal rate, which record the continuity of a worm's movement (Figure 4, Table 3). In the measurement of speed, we identified three missense mutants in the synaptic function cluster, *efr-3(D497G)*, *ost-1(W249R)*, and *mca-3(T763M),* with decreased speed phenotypes. (N2 of the first batch = 56.4 ± 2.3; *ost-1(W249R)* = 30.4 ± 4.4, $p = 0.0102$; *mca-3(T763M)* = 40.6 ± 2.7, $p =$

0.0375. N2 of the second batch = 169.9 ± 8.2; *efr-3(D497G)* = 126.9 ± 8.2, $p = 0.0138$. All strains were compared to the wild-type controls of the same experimental condition.) Missense mutants in the gene expression and regulation clusters did not show any difference in speed. As for the measurement of reversal rate, which we defined as the frequency of omega turns within the tracking duration, we detected higher reversal rates in two missense mutant strains, *tbx-8(K31E)* and *dpy-18(Y413C)* (N2 of the first batch = 14.4 ± 0.3, *tbx-8(K31E)* = 19.4 ± 1.1, $p = 0.001$; N2 of the second batch = 8.9 ± 0.5, *dpy-18(Y413C)* = 15.8 ± 2.1, p=0.001). Overall, these data suggested five ASD-associated missense mutant strains with movement phenotypes, most of whom belonged to the synaptic function gene cluster.

3.5 Discussion

In this study, we evaluated the functional consequences of 28 ASD-associated missense variants in their *C. elegans* orthologs. Our results revealed that 18 of the 28 (64%) conserved missense variants exhibited functional impacts on fecundity, morphology, or locomotor patterns (Table 4). The phenotypic changing rate presented in this study was similar to the rate (70%) in a previous study with the same approach (Wong et al., 2019). In addition, we also discovered that different physiological functions were more susceptible to mutations in different gene clusters. For example, most of the fecundity changing variants belonged to genes in chromatin remodeling, nucleotide cycle, and synaptic development and regulation. Morphological differences were only detected in missense mutant strains in the synaptic gene cluster. Sinusoidal waveform changes were found in both gene regulatory and synaptic function clusters whereas movement phenotypes such as speed were only displayed in the missense mutants in the synaptic function cluster. Notably, our phenotypic assay also reflected the direction of changes. Missense mutants in genes involved in opposite physiological functions showed opposite functional changes. For instance, missense variant in the adenylosuccinate lyase (ADSL) gene, which

converts inosine monophosphate (IMP) to adenosine monophosphate (AMP), and missense variant in the adenosine monophosphate deaminase 1 (AMPD1) gene, which recycles AMP to IMP, displayed the opposite effect on fecundity (Camici et al., 2018). Overall, we identified 18 functional changing missense loci, especially 8 of them caused multi-system functional changes. We prioritized these phenotypic changing missense variants for future studies on functional validation and disease mechanisms.

ASD is a heterogeneous disease with functionally diverse genes and various clinical descriptions. The relationship between genotype and functional outcomes can be defined on the basis of phenotype profiles of multiple genes across a standard battery of behavioral assays (Iakoucheva et al., 2019). Indeed, the genotype-phenotype correlation was used to distinguish the clinical subtype in common diseases (Luo et al., 2019), and factoring the analysis of phenotype data was proposed to further categorize clinical subtypes of ASD (Georgiades et al., 2013). With this in mind, recent studies applied a standard battery of assays to characterize ASD-associated genetic variants in various model systems in a comprehensive manner (Chen et al., 2018, 2020; Deneault et al., 2018; McDiarmid et al., 2020; Wong et al., 2019). Taking a similar approach, we conducted a set of behavioral assays to missense mutants in two gene clusters. We discovered that missense mutations in different gene clusters contributed to distinct phenotypic profiles. Moreover, we identified a proportion of missense variants causing phenotypic changes in multiple assays, indicating physiological changes at the system level. According to the genomic analysis in humans, the most highly constrained variants showed higher expression levels and broader tissue expression (Ardlie et al., 2015; Lek et al., 2016). Our results suggested these missense loci may locate in the constraint region that were less tolerant to sequence changes.

We used two standard functional assays to screen the phenotypic changing missense variants. In the fecundity assay, most phenotypic changing missense variants belonged to the gene expression and regulation cluster. This result indicated that changes in genes involved in chromatin remodeling and nucleotide cycle impacted developmental functions significantly. In fact, some of the genes have been documented to affect organism survival. For instance, histone demethylase activity of Utx was shown to impact viability in *Drosophila* (Copur & Müller, 2018). Similar to our finding of reduced fecundity in the *exc-7* missense mutant strain, the *Drosophila* fne^{null} mutant, one of the paralogs in ELAV family, also displayed reduced fecundity (Zanini et al., 2012). In addition, we examined the contribution of missense variants in the synaptic function cluster. We discovered that genes in the synaptic function cluster have more significant impacts on morphology and locomotor patterns. The effect in size was consistent with previous large-scale analyses of inactivating mutations in constrained genes in the humans genome and ASD-associated models in *C. elegans* (Ganna et al., 2018; McDiarmid et al., 2020). We detected a change in one of the constraint genes, *unc-73*, which is the *C. elegans* ortholog of human *TRIO.* Consistently, loss-of-function of *TRIO* impaired motor coordination and synaptic function in rodents and individuals with mild intellectual disability and (Ba et al., 2016; Katrancha et al., 2019). It is proposed that a single genetic variant can be widespread through inter-connecting gene networks and have emergent effects on specific phenotypes (Iakoucheva et al., 2019). Our findings proved this concept and highlight the importance of phenotypic specificity in different gene networks and fine-grained dimensional phenotyping.

The problem of a large amount of genetic variants with unknown significance presents a massive challenge for ASD studies. We attempted to solve this problem with a multi-cellular screening platform in *C. elegans*. We introduced the disease-associated missense variants in the endogenous *C. elegans* protein. This approach allows us to inspect the functional consequence of missense variants in its original genomic and cellular content, eliminating the effect caused by

cell-type-specific chromatin states (Ernst & Kellis, 2015; Starita et al., 2017). Besides, using the endogenous *C. elegans* protein preserves the original content, such as protein-protein interactions, intronic regulation, and isoform balance (Reble et al., 2018; Robison, 2014). The endogenous protein approach can avoid possible confounding effects such as mismatched heteromeric complexes presented in the "humanized" models as well (Bend et al., 2016; Prior et al., 2017) (Baruah et al., 2017; McDiarmid et al., 2018; Walsh et al., 2017). Furthermore, the short lifespan and well-developed genetic tools allow rapid generation of targeted mutation and double mutants before embarking on less efficient and more costly animal models (Markaki & Tavernarakis, 2010). Overall, our approach and model organism facilitates the dissection of the mechanisms of pathological conditions and drug target identification (Schmeisser et al., 2017).

Using *C. elegans* as a model to study psychiatric disorders has some limitations. Unlike rodent models, the *C. elegans* model lacks complicated behaviors. However, *C. elegans* and humans share essential physiological pathways (e.g., insulin signaling, Ras/Notch signaling, p53, and many miRNAs), neurotransmitter systems, and receptor pharmacology (Engleman et al., 2016; Markaki & Tavernarakis, 2010). The disruption of protein functions can be reflected in multiple well-developed behavioral assays, such as locomotion and habituation learning. Another limitation is that the evolutionary distance between humans and *C. elegans* restricts the subject to *C. elegans* orthologous genes. Fortunately, *C. elegans* and humans share 60% of gene orthologs (Kaletta & Hengartner, 2006; Shaye & Greenwald, 2011) and disease-associated genes are particularly conserved throughout evolution (López-Bigas & Ouzounis, 2004; Shpigler et al., 2017). Furthermore, for genetic candidates that show correlated expression, *C. elegans* models facilitate the study of interaction between missense variants by rapidly generating double or multiple missense mutation models.

With the accelerated identification of disease-associated variants, the challenge of interpreting their functional outcomes persists. To date, most preclinical research is biased toward experimentally well-accessible genes due to our limited knowledge of disease mechanisms (Stoeger et al., 2018). With the advance in data science, broad-scale phonemic analysis will be the future trend. Our study proved that missense mutants in different gene clusters displayed distinct phenotypic profiles and laid the foundation for future systematic characterization of disease-associated variants.

3.6 Acknowledgement

This work was supported by Simons Foundation (SFARI award # 367560 to P.W.S.). K.B. was supported by NIH pre-doctoral training grant T32GM007616. P.W.S. was an investigator with the Howard Hughes Medical Institute during part of this study.

3.7 References

Ardlie, K. G., DeLuca, D. S., Segrè, A. V., Sullivan, T. J., Young, T. R., Gelfand, E. T., Trowbridge, C. A., Maller, J. B., Tukiainen, T., Lek, M., Ward, L. D., Kheradpour, P., Iriarte, B., Meng, Y., Palmer, C. D., Esko, T., Winckler, W., Hirschhorn, J. N., Kellis, M., … Lockhart. (2015). The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science*, *348*(6235), 648–660. https://doi.org/10.1126/science.1262110

Ba, W., Yan, Y., Reijnders, M. R. F., Schuurs-Hoeijmakers, J. H. M., Feenstra, I., Bongers, E. M. H. F., Bosch, D. G. M., De Leeuw, N., Pfundt, R., Gilissen, C., De Vries, P. F., Veltman, J. A., Hoischen, A., Mefford, H. C., Eichler, E. E., Vissers, L. E. L. M., Kasri, N. N., & De Vries, B. B. A. (2016). TRIO loss of function is associated with mild intellectual disability and affects dendritic branching and synapse function. *Human Molecular Genetics*, *25*(5),

892–902. https://doi.org/10.1093/hmg/ddv618

Baruah, P. S., Beauchemin, M., Parker, J. A., & Bertrand, R. (2017). Expression of human Bcl-xL (Ser49) and (Ser62) mutants in Caenorhabditis elegans causes germline defects and aneuploidy. *PLoS ONE*, *12*(5), e0177413. https://doi.org/10.1371/journal.pone.0177413

Bell, C. J., Dinwiddie, D. L., Miller, N. A., Hateley, S. L., Ganusova, E. E., Mudge, J., Langley, R. J., Zhang, L., Lee, C. C., Schilkey, F. D., Sheth, V., Woodward, J. E., Peckham, H. E., Schroth, G. P., Kim, R. W., & Kingsmore, S. F. (2011). Carrier testing for severe childhood recessive diseases by next-generation sequencing. *Science Translational Medicine*, *3*(65). https://doi.org/10.1126/scitranslmed.3001756

Ben-David, E., & Shifman, S. (2013). Combined analysis of exome sequencing points toward a major role for transcription regulation during brain development in autism. *Molecular Psychiatry*, *18*(10), 1054–1056. https://doi.org/10.1038/mp.2012.148

Brenner, S. (1974). The genetics of Caenorhabditis elegans. *Genetics*, *77*(May), 71–94. https://doi.org/10.1111/j.1749-6632.1999.tb07894.x

Camici, M., Allegrini, S., & Tozzi, M. G. (2018). Interplay between adenylate metabolizing enzymes and AMP-activated protein kinase. *The FEBS Journal*, *285*, 3337–3352.

Chaste, P., Roeder, K., & Devlin, B. (2017). The Yin and Yang of Autism Genetics: How Rare de Novo and Common Variations Affect Liability. *Annual Review of Genomics and Human Genetics*, *18*, 167–187. https://doi.org/10.1146/annurev-genom-083115-022647

Chen, S., Fragoza, R., Klei, L., Liu, Y., Wang, J., Roeder, K., Devlin, B., & Yu, H. (2018). An interactome perturbation framework prioritizes damaging missense mutations for developmental disorders. *Nature Genetics*, *50*, 1032–1040. https://doi.org/10.1038/s41588-018-0130-z

Chen, S., Wang, J., Cicek, E., Roeder, K., Yu, H., & Devlin, B. (2020). De novo missense variants disrupting protein-protein interactions affect risk for autism through gene co-

expression and protein networks in neuronal cell types. *Molecular Autism*, *11*(1), 1–16. https://doi.org/10.1186/s13229-020-00386-7

Chiu, H., Schwartz, H. T., Antoshechkin, I., & Sternberg, P. W. (2013). Transgene-free genome editing in Caenorhabditis elegans using CRISPR-Cas. *Genetics*, *195*(3), 1167–1171. https://doi.org/10.1534/genetics.113.155879

Choi, Y., Sims, G. E., Murphy, S., Miller, J. R., & Chan, A. P. (2012). Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLoS ONE*, *7*(10). https://doi.org/10.1371/journal.pone.0046688

Coe, B. P., Stessman, H. A. F., Sulovari, A., Geisheker, M. R., Bakken, T. E., Lake, A. M., Dougherty, J. D., Lein, E. S., Hormozdiari, F., Bernier, R. A., & Eichler, E. E. (2019). Neurodevelopmental disease genes implicated by de novo mutation and copy number variation morbidity. *Nature Genetics*, *51*(1), 106–116. https://doi.org/10.1038/s41588-018-0288-4

Copur, Ö., & Müller, J. (2018). Histone demethylase activity of utx is essential for viability and regulation of HOX gene expression in drosophila. *Genetics*, *208*(2), 633–637. https://doi.org/10.1534/genetics.117.300421

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

Deneault, E., White, S. H., Rodrigues, D. C., Ross, P. J., Faheem, M., Zaslavsky, K., Wang, Z., Alexandrova, R., Pellecchia, G., Wei, W., Piekna, A., Kaur, G., Howe, J. L., Kwan, V., Thiruvahindrapuram, B., Walker, S., Lionel, A. C., Pasceri, P., Merico, D., … Scherer, S. W. (2018). Complete Disruption of Autism-Susceptibility Genes by Gene Editing

Predominantly Reduces Functional Connectivity of Isogenic Human Neurons. *Stem Cell Reports*, *11*(5), 1211–1225. https://doi.org/10.1016/j.stemcr.2018.10.003

Engleman, E. A., Katner, S. N., & Neal-beliveau, B. S. (2016). Caenorhabditis elegans as a model to study the molecular and genetic mechanisms of drug addiction. *Prog Mol Biol Transl Sci*, *137*, 229–252. https://doi.org/10.1016/bs.pmbts.2015.10.019.Caenorhabditis

Ernst, J., & Kellis, M. (2015). Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. *Nature Biotechnology*, *33*(4), 364–376. https://doi.org/10.1038/nbt.3157

Fischbach, G. D., & Lord, C. (2010). The simons simplex collection: A resource for identification of autism genetic risk factors. *Neuron*, *68*(2), 192–195. https://doi.org/10.1016/j.neuron.2010.10.006

Ganna, A., Satterstrom, F. K., Zekavat, S. M., Das, I., Kurki, M. I., Churchhouse, C., Alfoldi, J., Martin, A. R., Havulinna, A. S., Byrnes, A., Thompson, W. K., Nielsen, P. R., Karczewski, K. J., Saarentaus, E., Rivas, M. A., Gupta, N., Pietiläinen, O., Emdin, C. A., Lescai, F., … Neale, B. M. (2018). Quantifying the Impact of Rare and Ultra-rare Coding Variation across the Phenotypic Spectrum. *American Journal of Human Genetics*, *102*(6), 1204–1211. https://doi.org/10.1016/j.ajhg.2018.05.002

Gasperini, M., Starita, L., & Shendure, J. (2016). The power of multiplexed functional analysis of genetic variants. *Nature Protocols*, *11*(10), 1782–1787. https://doi.org/10.1038/nprot.2016.135

Georgiades, S., Szatmari, P., Boyle, M., Hanna, S., Duku, E., Zwaigenbaum, L., Bryson, S., Fombonne, E., Volden, J., Mirenda, P., Smith, I., Roberts, W., Vaillancourt, T., Waddell, C., Bennett, T., & Thompson, A. (2013). Investigating phenotypic heterogeneity in children with autism spectrum disorder: A factor mixture modeling approach. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *54*(2), 206–215.

https://doi.org/10.1111/j.1469-7610.2012.02588.x

Gilman, S. R., Iossifov, I., Levy, D., Ronemus, M., Wigler, M., & Vitkup, D. (2011). Rare De Novo Variants Associated with Autism Implicate a Large Functional Network of Genes Involved in Formation and Function of Synapses. *Neuron*, *70*(5), 898–907. https://doi.org/10.1016/j.neuron.2011.05.021

Gonzalez-cavazos, C., Cao, M., Wong, W., Chai, C., & Sternberg, P. W. (2019). Effects of ASD-associated daf-18/PTEN missense variants on C. elegans dauer development. *MicroPubl Biol*, *10*, 17912.

Hubbard, E. J. A., Korta, D. Z., & Dalfó, D. (2013). Physiological control of germline development. In *Advances in Experimental Medicine and Biology* (Vol. 757). https://doi.org/10.1007/978-1-4614-4015-4-5

Iakoucheva, L. M., Muotri, A. R., & Sebat, J. (2019). Getting to the Cores of Autism. *Cell*, *178*(6), 1287–1298. https://doi.org/10.1016/j.cell.2019.07.037

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

Iossifov, I., Ronemus, M., Levy, D., Wang, Z., Hakker, I., Rosenbaum, J., Yamrom, B., Lee, Y. H., Narzisi, G., Leotta, A., Kendall, J., Grabowska, E., Ma, B., Marks, S., Rodgers, L., Stepansky, A., Troge, J., Andrews, P., Bekritsky, M., … Wigler, M. (2012). De novo gene disruptions in children on the autistic spectrum. *Neuron*, *74*(2), 285–299. https://doi.org/10.1016/j.neuron.2012.04.009

Jin, X., Simmons, S. K., Guo, A. X., Shetty, A. S., Ko, M., Nguyen, L., Robinson, E., Oyler, P., Curry, N., Deangeli, G., Lodato, S., Levin, J. Z., Regev, A., Zhang, F., & Arlotta, P. (2020).

In vivo Perturb-Seq reveals neuronal and glial abnormalities associated with Autism risk genes. *Science*, *370*, eaaz6063. https://doi.org/10.1101/791525

Kaletta, T., & Hengartner, M. O. (2006). Finding function in novel targets: C. elegans as a model organism. *Nature Reviews Drug Discovery*, *5*(5), 387–399. https://doi.org/10.1038/nrd2031

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

Katrancha, S. M., Shaw, J. E., Zhao, A. Y., Myers, S. A., Cocco, A. R., Jeng, A. T., Zhu, M., Pittenger, C., Greer, C. A., Carr, S. A., Xiao, X., & Koleske, A. J. (2019). Trio Haploinsufficiency Causes Neurodevelopmental Disease– Associated Deficits. *Cell Reports*, *26*(10), 2805–2817. https://doi.org/10.1016/j.celrep.2019.02.022.Trio

King, I. F., Yandava, C. N., Mabb, A. M., Hsiao, J. S., Huang, H. S., Pearson, B. L., Calabrese, J. M., Starmer, J., Parker, J. S., Magnuson, T., Chamberlain, S. J., Philpot, B. D., & Zylka, M. J. (2013). Topoisomerases facilitate transcription of long genes linked to autism. *Nature*, *501*(7465), 58–62. https://doi.org/10.1038/nature12504

Landrum, M. J., Lee, J. M., Riley, G. R., Jang, W., Rubinstein, W. S., Church, D. M., & Maglott, D. R. (2014). ClinVar: Public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Research*, *42*(D1), 980–985. https://doi.org/10.1093/nar/gkt1113

Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., O'Donnell-Luria, A. H., Ware, J. S., Hill, A. J., Cummings, B. B., Tukiainen, T., Birnbaum, D. P., Kosmicki, J. A., Duncan, L. E., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J., … Williams, A. L. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, *536*(7616), 285–291. https://doi.org/10.1038/nature19057

López-Bigas, N., & Ouzounis, C. A. (2004). Genome-wide identification of genes likely to be involved in human genetic disease. *Nucleic Acids Research*, *32*(10), 3108–3114. https://doi.org/10.1093/nar/gkh605

Lord, C., Brugha, T. S., Charman, T., Cusack, J., Dumas, G., Frazier, T., Jones, E. J. H., Jones, R. M., Pickles, A., State, M. W., Taylor, J. L., & Veenstra-VanderWeele, J. (2020). Autism spectrum disorder. *Nature Reviews Disease Primers*, *6*(1). https://doi.org/10.1038/s41572-019-0138-4

Luo, Y., Mao, C., Yang, Y., Wang, F., Ahmad, F. S., Arnett, D., Irvin, M. R., & Shah, S. J. (2019). Integrating hypertension phenotype and genotype with hybrid non-negative matrix factorization. *Bioinformatics*, *35*(8), 1395–1403. https://doi.org/10.1093/bioinformatics/bty804

Markaki, M., & Tavernarakis, N. (2010). Modeling human diseases in Caenorhabditis elegans. *Biotechnology Journal*, *5*(12), 1261–1276. https://doi.org/10.1002/biot.201000183

McDiarmid, T. A., Au, V., Loewen, A. D., Liang, J., Mizumoto, K., Moerman, D. G., & Rankin, C. H. (2018). CRISPR-Cas9 human gene replacement and phenomic characterization in *Caenorhabditis elegans* to understand the functional conservation of human genes and decipher variants of uncertain significance. *Disease Models & Mechanisms*, dmm.036517. https://doi.org/10.1242/dmm.036517

McDiarmid, T. A., Belmadani, M., Liang, J., Meili, F., Mathews, E. A., Mullen, G. P., Hendi, A., Wong, W. R., Rand, J. B., Mizumoto, K., Haas, K., Pavlidis, P., & Rankin, C. H. (2020). Systematic phenomics analysis of autism-associated genes reveals parallel networks underlying reversible impairments in habituation. *Proc Natl Acad Sci USA*, *117*(1), 656–667. https://doi.org/10.1073/pnas.1912049116

Miosge, L. A., Field, M. A., Sontani, Y., Cho, V., Johnson, S., Palkova, A., Balakishnan, B., Liang, R., Zhang, Y., Lyon, S., Beutler, B., Whittle, B., Bertram, E. M., Enders, A.,

Goodnow, C. C., & Andrews, T. D. (2015). Comparison of predicted and actual consequences of missense mutations. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(37), E5189–E5198. https://doi.org/10.1073/pnas.1511585112

Neale, B. M., Kou, Y., Liu, L., Ma'ayan, A., Samocha, K. E., Sabo, A., Lin, C.-F., Stevens, C., Wang, L.-S., Makarov, V., Polak, P., Yoon, S., Maguire, J., Crawford, E. L., Campbell, N. G., Geller, E. T., Valladares, O., Schafer, C., Liu, H., … Daly, M. J. (2012). Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature*, *485*(7397), 242–245. https://doi.org/10.1038/nature11011

O'Roak, B. J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B. P., Levy, R., Ko, A., Lee, C., Smith, J. D., Turner, E. H., Stanaway, I. B., Vernot, B., Malig, M., Baker, C., Reilly, B., Akey, J. M., Borenstein, E., Rieder, M. J., … Eichler, E. E. (2012). Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature*, *485*(7397), 246–250. https://doi.org/10.1038/nature10989

Parikshak, N. N., Luo, R., Zhang, A., Won, H., Lowe, J. K., Chandran, V., Horvath, S., & Geschwind, D. H. (2013). Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell*, *155*(5), 1008. https://doi.org/10.1016/j.cell.2013.10.031

Post, K. L., Belmadani, M., Ganguly, P., Meili, F., Dingwall, R., McDiarmid, T. A., Meyers, W. M., Herrington, C., Young, B. P., Callaghan, D. B., Rogic, S., Edwards, M., Niciforovic, A., Cau, A., Rankin, C. H., O'Connor, T. P., Bamji, S. X., Loewen, C. J. R., Allan, D. W., … Haas, K. (2020). Multi-model functionalization of disease-associated PTEN missense mutations identifies multiple molecular mechanisms underlying protein dysfunction. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-15943-0

Reble, E., Dineen, A., & Barr, C. L. (2018). The contribution of alternative splicing to genetic

risk for psychiatric disorders. *Genes, Brain and Behavior*, *17*(3), 1–12. https://doi.org/10.1111/gbb.12430

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., & Rehm, H. L. (2015). Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*, *17*(5), 405–424. https://doi.org/10.1038/gim.2015.30

Robison, A. J. (2014). Emerging role of CaMKII in neuropsychiatric disease. *Trends in Neurosciences*, *37*(11), 653–662. https://doi.org/10.1016/j.tins.2014.07.001

Sakai, Y., Shaw, C. A., Dawson, B. C., Dugas, D. V., Al-Mohtaseb, Z., Hill, D. E., & Zoghbi, H. Y. (2011). Protein interactome reveals converging molecular pathways among autism disorders. *Science Translational Medicine*, *3*(86). https://doi.org/10.1126/scitranslmed.3002166

Sanders, S. J., Murtha, M. T., Gupta, A. R., Murdoch, J. D., Raubeson, M. J., Willsey, A. J., Ercan-Sencicek, A. G., DiLullo, N. M., Parikshak, N. N., Stein, J. L., Walker, M. F., Ober, G. T., Teran, N. a., Song, Y., El-Fishawy, P., Murtha, R. C., Choi, M., Overton, J. D., Bjornson, R. D., … State, M. W. (2012). De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature*, *485*(7397), 237–241. https://doi.org/10.1038/nature10945

Satterstrom, F. K., Kosmicki, J. A., Wang, J., Breen, M. S., De Rubeis, S., An, J. Y., Peng, M., Collins, R., Grove, J., Klei, L., Stevens, C., Reichert, J., Mulhern, M. S., Artomov, M., Gerges, S., Sheppard, B., Xu, X., Bhaduri, A., Norman, U., … Buxbaum, J. D. (2020). Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell*, *180*(3), 568-584.e23. https://doi.org/10.1016/j.cell.2019.12.036

Schmeisser, K., Fardghassemi, Y., & Parker, J. A. (2017). A rapid chemical-genetic screen utilizing impaired movement phenotypes in C. elegans: Input into genetics of neurodevelopmental disorders. *Experimental Neurology*, *293*, 101–114. https://doi.org/10.1016/j.expneurol.2017.03.022

Shaye, D. D., & Greenwald, I. (2011). Ortholist: A compendium of C. elegans genes with human orthologs. *PLoS ONE*, *6*(5). https://doi.org/10.1371/journal.pone.0020085

Shpigler, H. Y., Saul, M. C., Corona, F., Block, L., Ahmed, A. C., Zhao, S. D., & Robinson, G. E. (2017). Deep evolutionary conservation of autism-related genes. *P Natl Acad Sci USA*, *114*(36), 9653–9658. https://doi.org/10.1073/pnas.1913223116

Starita, L. M., Ahituv, N., Dunham, M. J., Kitzman, J. O., Roth, F. P., Seelig, G., Shendure, J., & Fowler, D. M. (2017). Variant interpretation: functional assays to the rescue. *American Journal of Human Genetics*, *101*(3), 315–325. https://doi.org/10.1016/j.ajhg.2017.07.014

Stoeger, T., Gerlach, M., Morimoto, R. I., & Nunes Amaral, L. A. (2018). Large-scale investigation of the reasons why potentially important genes are ignored. *PLoS Biology*, *16*(9), 1–25. https://doi.org/10.1371/journal.pbio.2006643

Sullivan, J. M., De Rubeis, S., & Schaefer, A. (2019). Convergence of spectrums: neuronal gene network states in autism spectrum disorder. *Current Opinion in Neurobiology*, *59*, 102–111. https://doi.org/10.1016/j.conb.2019.04.011

Sun, S., Yang, F., Tan, G., Costanzo, M., Oughtred, R., Hirschman, J., Theesfeld, C. L., Bansal, P., Sahni, N., Yi, S., Yu, A., Tyagi, T., Tie, C., Hill, D. E., Vidal, M., Andrews, B. J., Boone, C., Dolinski, K., & Roth, F. P. (2016). An extended set of yeast-based functional assays accurately identifies human disease mutations. *Genome Research*, *26*(5), 670–680. https://doi.org/10.1101/gr.192526.115

Thusberg, J., Olatubosun, A., & Vihinen, M. (2011). Performance of mutation pathogenicity prediction methods on missense variants. *Human Mutation*, *32*(4), 358–368.

https://doi.org/10.1002/humu.21445

Voineagu, I., Wang, X., Johnston, P., Lowe, J. K., Tian, Y., Horvath, S., Mill, J., Cantor, R. M., Blencowe, B. J., & Geschwind, D. H. (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature*, *474*(7351), 380–386. https://doi.org/10.1038/nature10110

Walsh, N., Kenney, L., Jangalwe, S., Aryee, K., Greiner, D. L., Brehm, M. A., Shultz, L. D., & Harbor, B. (2017). Humanized mouse models of clinical disease. *Annual Review of Pathology*, *24*(12), 187–215. https://doi.org/10.1146/annurev-pathol-052016-100332.Humanized

Wertz, M. H., Mitchem, M. R., Pineda, S. S., Hachigian, L. J., Lee, H., Lau, V., Powers, A., Kulicke, R., Madan, G. K., Colic, M., Therrien, M., Vernon, A., Beja-Glasser, V. F., Hegde, M., Gao, F., Kellis, M., Hart, T., Doench, J. G., & Heiman, M. (2020). Genome-wide In Vivo CNS Screening Identifies Genes that Modify CNS Neuronal Survival and mHTT Toxicity. *Neuron*, *106*(1), 76-89.e8. https://doi.org/10.1016/j.neuron.2020.01.004

Wong, W. R., Brugman, K. I., Maher, S., Oh, J. Y., Howe, K., Kato, M., & Sternberg, P. W. (2019). Autism-associated missense genetic variants impact locomotion and neurodevelopment in Caenorhabditis elegans. *Human Molecular Genetics*, *28*(13), 2271–2281. https://doi.org/10.1093/hmg/ddz051

Wong, W. R., Maher, S., Oh, J. Y., Brugman, K. I., Gharib, S., & Sternberg, P. W. (2021). Conserved missense variant in ALDH1A3 ortholog impairs fecundity in C . elegans. *MicroPubl Biol*, https://doi.org/10.17912/micropub.biology.000357.

Zanini, D., Jallon, J. M., Rabinow, L., & Samson, M. L. (2012). Deletion of the Drosophila neuronal gene found in neurons disrupts brain anatomy and male courtship. *Genes, Brain and Behavior*, *11*(7), 819–827. https://doi.org/10.1111/j.1601-183X.2012.00817.x

Zhao, Y. T., Kwon, D. Y., Johnson, B. S., Fasolino, M., Lamonica, J. M., Kim, Y. J., Zhao, B.

S., He, C., Vahedi, G., Kim, T. H., & Zhou, Z. (2018). Long genes linked to autism spectrum disorders harbor broad enhancer-like chromatin domains. *Genome Research*, *28*(7), 933–942. https://doi.org/10.1101/gr.233775.117

3.8 Tables and figures

Table 3-1. List of missense variants tested and strain information

Human gene	Human cDNA change [a]	Human protein change	Inheritance pattern	*C. elegans* gene	*C. elegans* protein change	Strain number
ADSL	242A>C	E80D	familial	*adsl-1(sy1067)*	E76D	PS7765
AMPD1	1498C>T	R500C	familial	*ampd-1(sy925)*	R566C	PS7572
AMPD1	1714C>T	P572S	*de novo*	*ampd-1(sy939)*	P638S	PS7613
ATP2B2	2453C>T	T818M	*de novo*	*mca-3(sy930)*	T763M	PS7600
BCL11A	143G>T	C48F	*de novo*	*bcl-11(sy932)*	C112F	PS7603
EFR3A	728G>C	G243A	familial	*efr-3(sy917)*	G256A	PS7550
EFR3A	1511A>G	D504G	unknown	*efr-3(sy915)*	D497G	PS7548
ELAVL3	245A>T	Y82F	*de novo*	*exc-7(sy914)*	Y85F	PS7547
EP400	2131C>T	R711C	unknown	*ssl-1(sy955)*	R212C	PS7707
EP400	5923G>A	G1975R	*de novo*	*ssl-1(sy1064)*	G1301R	PS7760
GNAS	2254G>A	A752T	*de novo*	*gsa-1(sy1106)*	A93T	PS7890
KDM6B	4420G>A	A1474T	unknown	*utx-1(sy348)*	A976T	PS7623
KMT2C	5006C>G	P1669R	familial	*set-16(sy1102)*	P967R	PS7886
KMT2C	14621G>A	R1487Q	familial	*set-16(sy1098)*	R2438Q	PS7879
MAPK3	934C>T	P312S	unknown	*mpk-1(sy871)*	P410S	PS7883
NRXN1	53T>A	L18Q	-	*nrx-1(sy869)*	L16Q	PS7330
P4HA2	458G>A	G153E	*de novo*	*dpy-18(sy344)*	G150E	PS7618
P4HA2	1262A>G	Y421C	*de novo*	*dpy-18(sy946)*	Y413C	PS7996
PAX6	136C>G	L46V	familial	*vab-3(sy894)*	L47V	PS8064
SHANK3	203T>C	L68P	familial	*shn-1(sy865)*	L67P	PS7259
SLC6A1	1078G>A	G360S	*de novo*	*snf-11(sy1073)*	G369S	PS7773
SPARCL1	G/A	T391M	unknown	*ost-1(sy922)*	T142M	PS7569
SPARCL1	A>G	W522R	familial	*ost-1(sy921)*	W249R	PS7568
TBR1	682A>G	K228E	*de novo*	*tbx-8(sy890)*	K31E	PS7454
TBR1	813G>T	W271C	*de novo*	*tbx-8(sy891)*	W74C	PS7459
TRIO	502G>A	G168S	familial	*unc-73(sy892)*	G1930S	PS7475
TRIO	5673A>G	K1891R	unknown	*unc-73(sy898)*	K2046R	ps7479
TRIO	6658G>C	V2220L	*de novo*	*unc-73(sy896)*	V2062L	PS7477

[a] The virtual cDNA number was provided by SFARI database.

Table 3-2. Morphology of missense mutant strains

Strain	Length (μm)	Width (μm)	Area (μm²)
N2	1109 ± 5	88.8 ± 0.8	100301 ± 1152
Gene regulation and expression genes			
tbx-8(K31E)	1090 ± 6	93.0 ± 2.2	103174 ± 2539
tbx-8(W74C)	1116 ± 14	94.6 ± 2.3	107393 ± 2539
vab-3(n3721)	1106 ± 4	89.7 ± 0.9	100969 ± 1245
bcl-11(C112F)	1099 ± 9	96.1 ± 2.6	107526 ± 3159
exc-7(Y85F)	1087 ± 14	87.4 ± 3.8	96996 ± 5041
ssl-1(G1301R)	1067 ± 23	87.9 ± 1.8	95761 ± 3919
ssl-1(R212C)	1137 ± 11	87.2 ± 2.2	101313 ± 3396
set-16(P967R)	1075 ± 12	85.9 ± 2.4	94131 ± 3321
set-16(R2438Q)	1098 ± 7	90.6 ± 2.1	101323 ± 2736
utx-1(A976T)	1075 ± 7	87.2 ± 1.4	95383 ± 1981
adsl-1(E76D)	1101 ± 8	85.5 ± 1.4	95628 ± 1042
ampd-1(P638S)	1099 ± 4	86.4 ± 0.7	96573 ± 1014
ampd-1(R566C)	1119 ± 12	87.5 ± 0.8	99641 ± 1685
Synaptic function genes			
efr-3(G256A)	11478 ±10	98.1 ± 2.0*	114918 ± 3385*
efr-3(D497G)	1145 ± 5	90.6 ± 0.7	105545 ± 995
nrx-1(L16Q)	1119 ± 5	84.5 ± 1.7	96139 ± 2115
shn-1(L67P)	1095 ± 10	82.8 ± 0.8	92172 ± 837
snf-11(G369S)	1088 ± 7	87.7 ± 1.0	97211 ± 1514
dpy-18(G150E)	1107 ± 17	95.1 ± 3.4	107340 ± 4888
dpy-18(Y413C)	914 ± 11*	91.6 ± 2.8	851967 ± 2721*
mpk-1(P410S)	991 ± 9*	85.9 ± 0.8	86549 ± 822*
ost-1(T142M)	1141 ± 25	90.9 ± 4.4	105983 ± 6774
ost-1(W249R)	937 ± 17*	80.9 ± 2.1	77431 ± 2731*
unc-73(G1930S)	1123 ± 6	87.4 ± 0.9	99793 ± 1253
unc-73(K2046R)	1099 ± 11	85.8 ± 0.8	95960 ± 1561
unc-73(V2062L)	1143 ± 7	87.2 ± 0.9	101396 ± 1470
gsa-1(A93T)	1092 ± 12	90.6 ± 2.4	100818 ± 3408
mca-3(T763M)	1051 ± 7*	85.4 ± 1.2	91434 ± 1761*

All values were presented as mean± SEM.

*: $p<0.05$ via one-way ANOVA and Tukey multiple comparison to N2 wild-type controls.

Table 3-3. Movement and coordination phenotypes in the missense mutant strains

	Sinusoidal waveform pattern			Movement	
Strain	Wavelength (μm)	Max Amplitude (μm)	Mean Amplitude (μm)	Reversal Rate (turns/min)	Speed (μm/s)
First batch					
N2	593.8 ± 3.9	195.4 ± 1.8	62.5 ± 0.7	14.4 ± 0.3	56.4 ± 2.3
Gene regulation and expression genes					
tbx-8(K31E)	558.1 ± 6.1*	206.6 ± 1.9	80.4 ± 1.8*	19.4 ± 1.1*	47.0 ± 2.6
tbx-8(W74C)	608.2 ± 8.9	201.3 ± 4.8	62.0 ± 1.6	16.8 ± 0.9	53.9 ± 2.7
vab-3(n3721)	594.4 ± 3.8	193.3 ± 8.2	64.2 ± 1.9	15.0 ± 1.4	47.7 ± 5.0
bcl-11(C112F)	603.4 ± 7.2	192.5 ± 6.4	64.7 ± 1.7	17.5 ± 1.2	48.4 ± 5.0
exc-7(Y85F)	544.3 ± 9.4*	201.1 ± 2.9	81.9 ± 2.3*	17.7 ± 1.2	46.5 ± 4.9
ssl-1(G1301R)	578.2 ± 12.3	188.3 ± 3.9	62.3 ± 1.3	13.8 ± 1.1	49.3 ± 7.4
ssl-1(R212C)	613.6 ± 7.1	201.4 ± 3.3	64.6 ± 1.4	15.1 ± 0.7	63.9 ± 5.4
set-16(P967R)	579.0 ± 7.4	179.6 ± 8.1	60.5 ± 2.9	12.8 ± 1.4	47.6 ± 6.4
adsl-1(E76D)	595.3 ± 2.9	199.4 ± 3.0	65.2 ± 1.3	16.6 ± 0.7	50.1 ± 4.0
ampd-1(P638S)	562.6 ± 4.1*	204.9 ± 2.3	82.8 ± 0.9*	16.1 ± 1.0	58.2 ± 5.1
ampd-1(R566C)	553.6 ± 7.0*	211.2 ± 3.0	91.8 ± 2.6*	14.4 ± 1.7	39.3 ± 6.3
Synaptic function genes					
efr-3(G256A)	577.6 ± 8.2	209.1 ± 2.6	89.1 ± 1.8*	16.9 ± 1.6	46.4 ± 7.8
nrx-1(L16Q)	591.3 ± 5.1	204.0 ± 4.8	65.6 ± 2.3	13.3 ± 0.6	41.2 ± 4.4
shn-1(L67P)	580.9 ± 5.7	195.6 ± 3.9	66.7 ± 0.9	14.9 ± 1.3	43.9 ± 5.2
snf-11(G369S)	582.5 ± 3.7	186.6 ± 3.2	57.2 ± 1.2	14.6 ± 0.8	54.6 ± 3.6
ost-1(T142M)	614.5 ± 16.8	199.3 ± 4.3	67.2 ± 2.6	15.0 ± 1.1	53.8 ± 6.6
ost-1(W249R)	493.6 ± 9.1*	171.2 ± 2.5*	57.0 ± 1.8	13.6 ± 1.3	30.4 ± 4.4*
unc-73(G1930S)	562.2 ± 3.4	207.5 ± 3.1	85.3 ± 1.3*	14.4 ± 1.0	51.3 ± 7.4
unc-73(K2046R)	603.2 ± 6.8	208.6 ± 1.9	66.2 ± 0.9	15.8 ± 0.8	73.8 ± 6.7
unc-73(V2062L)	620.9 ± 4.4	204.9 ± 4.5	64.6 ± 0.7	13.5 ± 1.1	58.6 ± 7.1
mca-3(T763M)	533.9 ± 3.7*	192.4 ± 3.3	72.1 ± 1.7*	16.1 ± 0.9	40.6 ± 2.7*
Second batch					
N2	595.6 ± 4.5	204.9 ± 3.0	60.0 ± 0.8	8.9 ± 0.5	169.9 ± 8.2
Gene regulation and expression genes					
set-16(R2438Q)	596.0 ± 4.0	205.4 ± 5.3	58.7 ± 1.7	9.6 ± 0.6	172.3 ± 5.6
utx-1(A976T)	578.1 ± 9.3	203.7 ± 2.4	61.7 ± 1.2	10.9 ± 1.6	163.2 ± 15.3
Synaptic function genes					
efr-3(D497G)	561.3 ± 2.0*	221.8 ± 2.7*	82.2 ± 1.3*	9.0 ± 0.7	126.9 ± 8.2*
dpy-18(G150E)	605.2 ± 14.7	204.7 ± 3.7	64.0 ± 1.3	11.4 ± 1.0	184.8 ± 15.6
dpy-18(Y413C)	522.0 ± 7.7*	167.9 ± 3.0*	53.6 ± 1.4	15.8 ± 2.1*	161.7 ± 10.3
mpk-1(P410S)	572.0 ± 4.5	184.3 ± 3.1*	64.6 ± 1.8	12.0 ± 0.8	152.6 ± 9.1
gsa-1(A93T)	598.1 ± 6.6	200.7 ± 5.4	59.4 ± 1.9	9.7 ± 0.9	183.5 ± 8.0

All values were presented as mean± SEM.

*: $p<0.05$ via one-way ANOVA and Tukey multiple comparison to N2 wild-type controls of the same batch. : the 2nd batch after room shift.

Table 3-4. Results summary

Human locus	*C. elegans* locus	Phenotypes			
Gene regulation and expression genes		Fecundity	Morphology	Waveform	Movement
Transcriptional regulation					
TBR1(K228E)	*tbx-8(K31E)**			+	+
TBR1(W271C)	*tbx-8(W74C)*				
PAX6(L46V)	*vab-3(L47V)*				
Chromatin remodeling					
BCL11A(C48F)	*bcl-11(C112F)*	+			
ELAVL3(Y82F)	*exc-7(Y85F)**	+		+	
EP400(R711C)	*ssl-1(R212C)*				
EP400(G1975R)	*ssl-1(G1301R)*	+			
KMT2C(P1669R)	*set-16(P967R)*	+			
KMT2C(R1487Q)	*set-16(R2438Q)*				
KDM6B(A1474T)	*utx-1(A976T)*	+			
Nucleotide cycle					
ADSL(E80D)	*adsl-1(E76D)*	+			
AMPD1(R500C)	*ampd-1(P638S)*			+	
AMPD1(P572S)	*ampd-1(R566C)**	+		+	
Synaptic function genes					
Synaptic transmission					
EFR3A(G243A)	*efr-3(G256A)*		+	Δ	
EFR3A(D504G)	*efr-3(D497G)**			+	+
NRXN1(L18Q)	*nrx-1(L16Q)*				
SHANK3(L68P)	*shn-1(L67P)*				
SLC6A1(G360S)	*snf-11(G369S)*				
Synaptic development & regulation					
P4HA2(G153E)	*dpy-18(G150E)*				
P4HA2(Y421C)	*dpy-18(Y413C)**		+	Δ	+
MAPK3(P312S)	*mpk-1(P410S)**	+	+	Δ	
SPARCL1(T391M)	*ost-1(T142M)*	+			
SPARCL1(W522R)	*ost-1(W249R)**	+	+	Δ	+
TRIO(G168S)	*unc-73(G1930S)*			+	
TRIO(K1891R)	*unc-73(K2046R)*				
TRIO(V2220L)	*unc-73(V2062L)*	+			
Calcium signaling					
GNAS(A752T)	*gsa-1(A93T)*				
ATP2B2(T818M)	*mca-3(T763M)**		+	+	+

*: missense variants that resulted in functional changes in more than one assay independently.
+: independent phenotypic changes.
Δ: phenotypic changes that may be resulted from morphology differences.

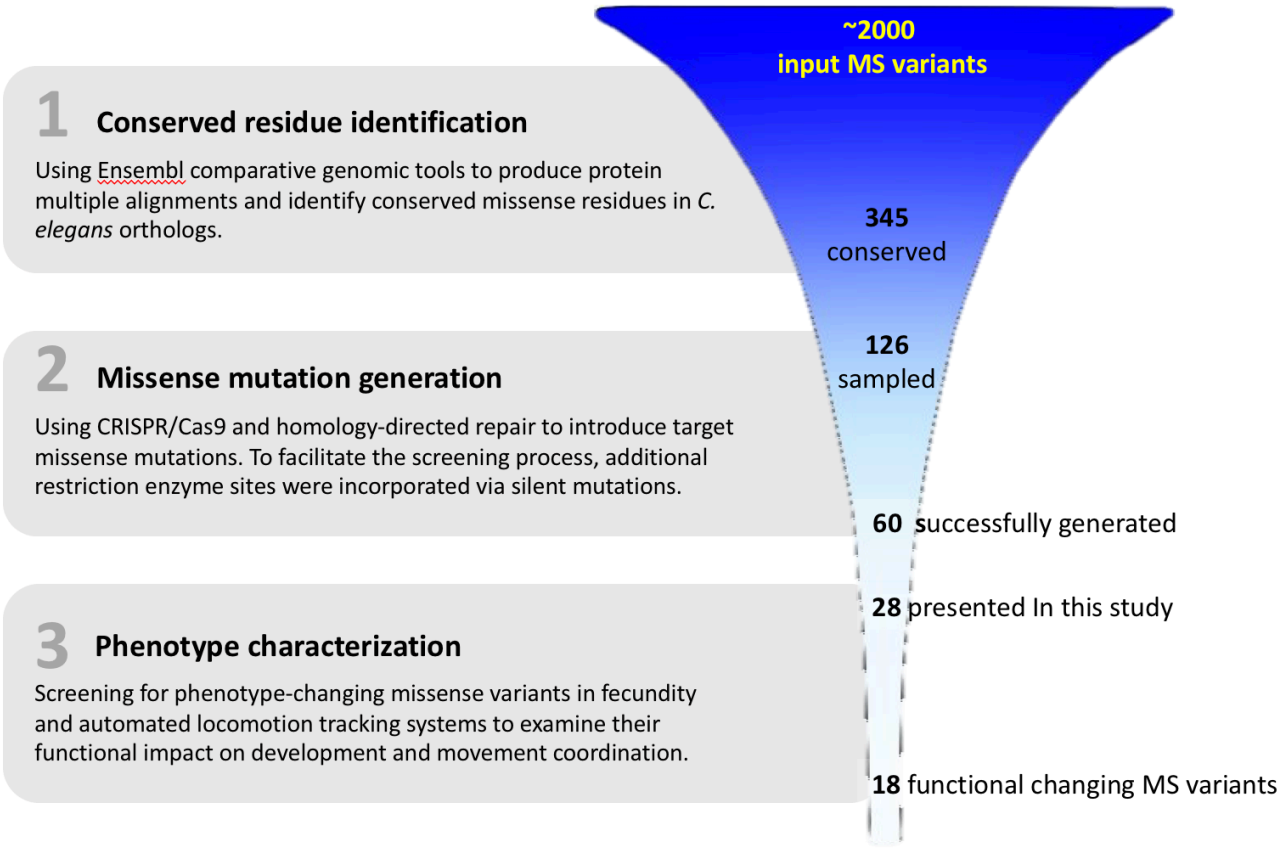

Figure 3-1. Experimental pipeline.
Thousands of missense (MS) variants implicated in ASD were obtained from the SFARI database. First, we utilized Ensembl comparative genomic tools to identify the conserved missense residues in their *C. elegans* orthologs. Among the 345 conserved missense residues, we sampled 126 loci in genes involved in gene regulation or synaptic functions. We used CRISPR/Cas9 and homology-directed repair (HDR) to introduce ASD-associated missense mutations in their *C. elegans* orthologous genes and successfully generated 60 missense mutant strains. This study presented the phenotype characterization of 28 missense mutant strains in fecundity and automated locomotor tracking assays. We found that 18 of the 28 missense mutants displayed phenotypic changes in fecundity, morphology, and movement coordination functions.

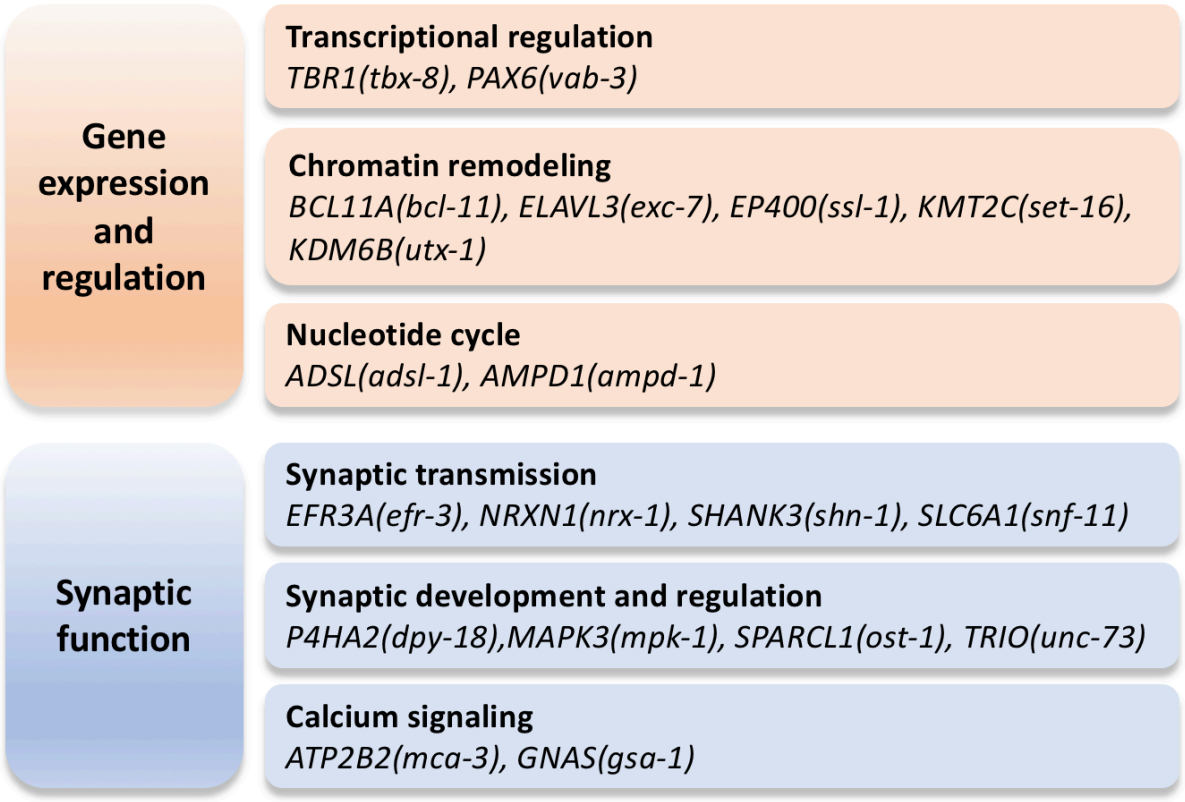

Figure 3-2. Gene clusters.
We sampled 19 ASD-associated genes from two major functional categories: (1) gene expression and regulation (2) synaptic function. Nine genes involved in gene expression and regulation take part in transcriptional regulation, chromatin remodeling, and nucleotide cycle. Ten genes related to synaptic function participate in synaptic transmission, synaptic development and regulation, and calcium signaling. For each human gene of interest, the *C. elegans* orthologs were presented in the parenthesis.

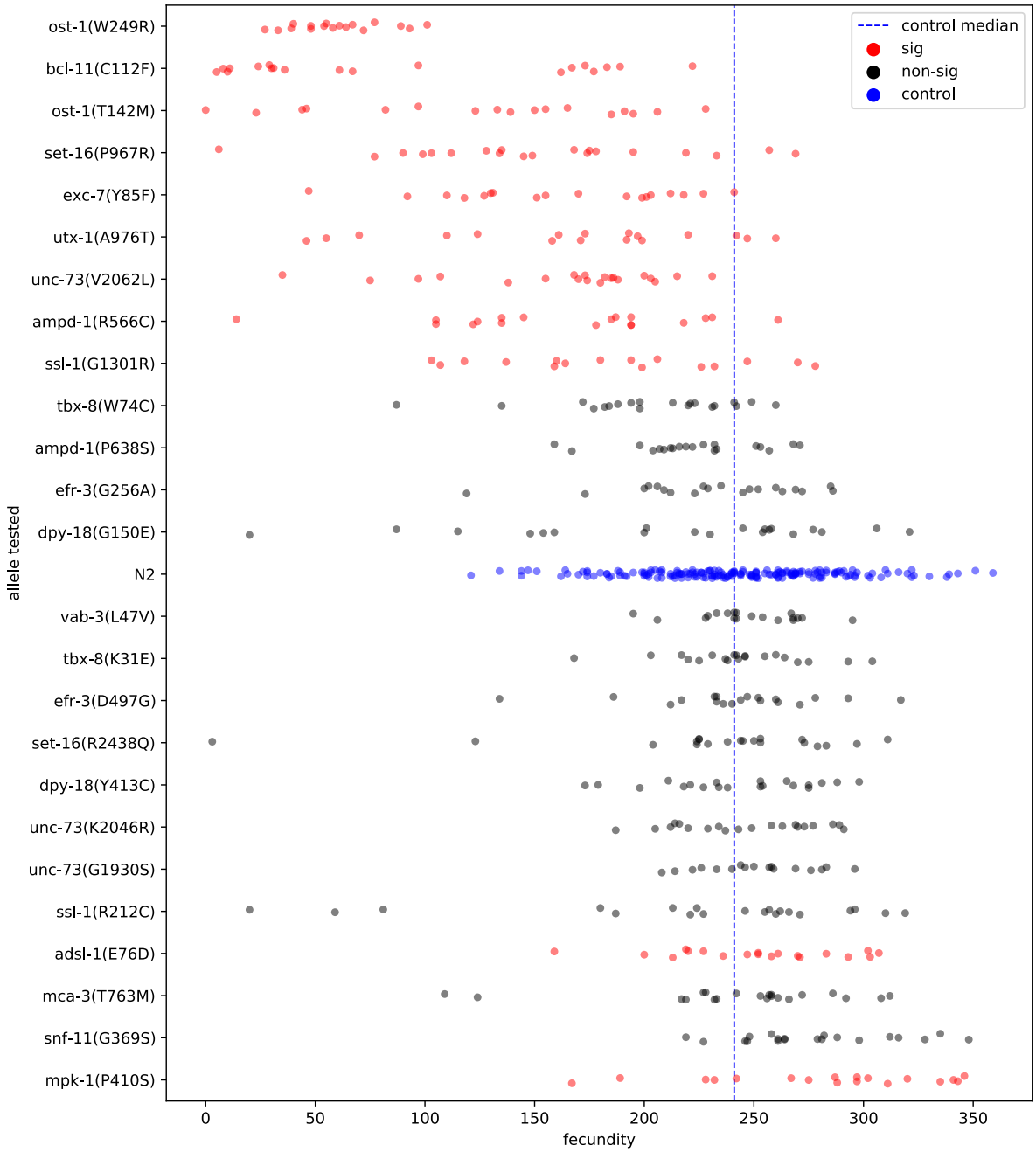

Figure 3-3. Fecundity phenotype in missense mutant strains.
In the gene regulation and expression cluster, six missense mutant strains dampened the fecundity in *C. elegans*. One missense mutant, *adsl-1(E76D)*, displayed an increase in fecundity. As for the synaptic gene cluster, the missense mutants exhibiting fecundity phenotypes all belong to the synaptic development and regulation subset. Approximately 20 animals were tested in each strain. Wild-type (N2) and its median values (dotted vertical line) were shown in blue. Mutants significantly different (sig) from wild type were shown in red. $p <$ 0.01/(total test number) via non-parametric bootstrap analysis.

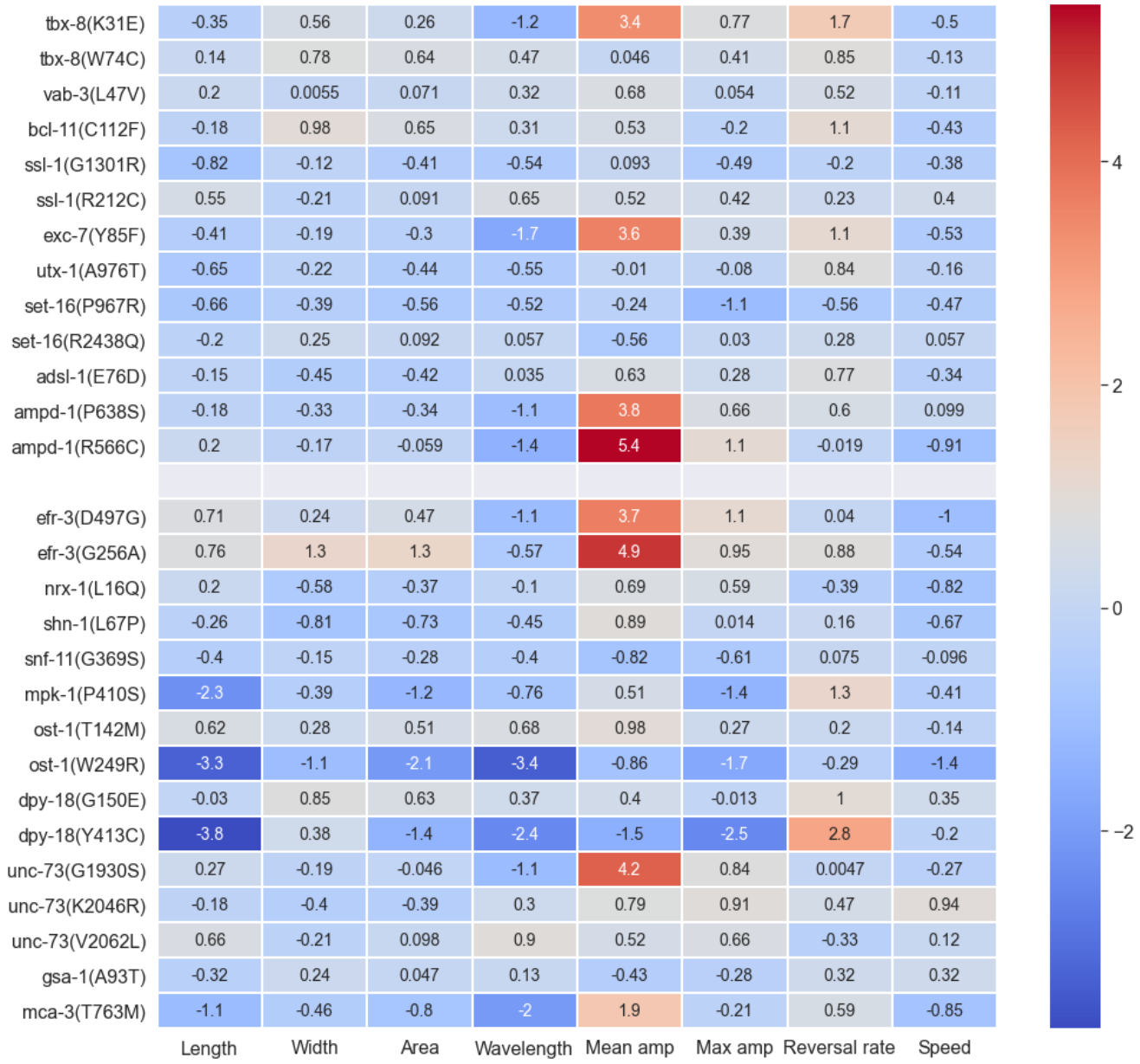

Figure 3-4. Morphological and locomotor phenotypes in missense mutant strains.
We quantify changes in morphology (length, width, and area), sinusoidal waveform pattern (wavelength, mean amplitude, and max amplitude), and movement (reversal rate and speed). Missense mutations in the gene regulation and expression cluster (top panel) did not affect morphology, and they had fewer effects on movement coordination. In contrast, missense mutations in synaptic function genes cluster (bottom panel) had some influences on morphology and much more impacts on the movement and coordination phenotypes. The number in each cell presented the z-score of the parameter in the missense mutant. The movement-related phenotypes were compared to the wild-type controls of the same batch. The more saturated color showed more deviation from the values in wild-type animals.

Chapter 4

CONSERVED MISSENSE VARIANT IN ALDH1A3 ORTHOLOG IMPAIRS FECUNDITY IN C. ELEGANS

Wong, W. R., Maher, S., Oh, J. Y., Brugman, K. I., Gharib, S., & Sternberg, P. W. (2021). Conserved missense variant in ALDH1A3 ortholog impairs fecundity in C . elegans. *MicroPubl Biol*, https://doi.org/10.17912/micropub.biology.000357.

4.1 Abstract

Accumulating evidence demonstrates that mutations in *ALDH1A3* (the aldehyde dehydrogenase 1 family, member A3) are associated with developmental defects. The ALDH1A3 enzyme catalyzes retinoic acid biosynthesis and is essential to patterning and neuronal differentiation in the development of embryonic nervous system. Several missense mutations in *ALDH1A3* have been identified in family studies of autosomal recessive microphthalmia, autism spectrum disorder, and other neurological disorders. However, there has been no evidence from animal models that verify the functional consequence of missense mutations in *ALDH1A3*. Here, we introduced the equivalent of the *ALDH1A3* C174Y variant into the *Caenorhabditis elegans* ortholog, *alh-1*, at the corresponding locus. Mutant animals with this missense mutation exhibited decreased fecundity by 50% compared to wild-type animals, indicating disrupted protein function. To our knowledge, this is the first ALDH1A3 C174Y missense model, which might be used to elucidate the effects of ALDH1A3 C174Y missense mutation in the retinoic acid signaling pathway during development.

4.2 Description

Retinoic acid (RA), the active metabolite of vitamin A, broadly regulates gene expression. The RA signaling pathway plays an essential role in embryonic development, including the development of the body axis, eye, brain, and heart (Ghyselinck & Duester, 2019).

One of the key enzymes in the biosynthesis of RA is aldehyde dehydrogenase 1 family member A3 (ALDH1A3). ALDH1A3 converts retinaldehyde to retinoic acid and it is expressed early in forebrain development (McCaffery & Drager, 1994). Several mutations in *ALDH1A3* have been implicated in patients with autosomal recessive microphthalmia and other neurological disorders (Fares-Taie et al., 2013; Roos et al., 2014). However, current animal models of ALDH1A3 have large truncations of the protein. Direct evidence of the effects of missense variants on ALDH1A3 protein activity has not yet been obtained.

In this study, we aimed to determine the functional consequence of *ALDH1A3(C174Y)* missense variants implicated in patients (Roos et al., 2014). Patients with the Cys174Tyr missense variant presented with autosomal recessive microphthalmia, autistic symptoms, and intellectual disabilities. The human ALDH1A3 C174Y missense residue is located at the amino-terminal nicotinamide adenine dinucleotide (NAD)-binding domain (Moretti et al., 2016), which is evolutionarily conserved between human and commonly used model organisms, such as *C. elegans*, *Drosophila,* zebrafish, and mouse. The *ALDH1A3* orthologs in these model organisms have similar amino acids at >75% of positions of the human ALDH1A3 protein. We believe the conservation of protein sequences among various species indicates the essential role of this domain in protein functions. Here, we use *C. elegans* as a prime candidate for studying the causality of missense variants in *ALDH1A3*. We generated the targeted *ALDH1A3 C174Y* missense variant in its *C. elegans* orthologous gene *alh-1* (Figure 4-1A).

The short lifespan and genetically modifiable nature of *C. elegans* allow rapid screening of functional impactful missense mutation implicated in human diseases. We previously established a working pipeline to introduce autism-associated missense mutations into the genome of *C. elegans* using CRISPR/Cas9 and homology-directed repair (Wong et al., 2019). Here, we used this pipeline to generate a missense mutant, *alh-1 (C172Y)*, that matches the human *ALDH1A3(C174Y)* variant. We then characterized this *C. elegans* missense mutant in a

fecundity assay to quantify the robustness of germline development. The *C. elegans* germline development is a highly sensitive process, which reflects subtle changes in sensory perception, feeding behavior, and metabolism (Hubbard et al., 2013). As a result, we used the *C. elegans* fecundity assay as a screening tool to identify functional changing missense mutations, as defects in germline development will result in a decreased brood size.

We evaluated the impact of C172Y missense mutation on development using a fecundity assay, which measures the number of viable progeny per hermaphrodite. The *alh-1(C172Y)* missense mutant nematode, *alh-1(sy899)*, showed only 54% of wild-type fecundity (Figure 4-1B. Wild-type: 234 ± 7; *alh-1(C172Y)*: 127 ± 8. $p < 0.0001$). We observed that two of the eighteen mutant animals had very low brood size and a lot of unhatched eggs (17 unhatched eggs to a brood size of 34, and 4 unhatched eggs to a brood size of 38). The other animals had essentially no unhatched eggs or dead larvae. This result indicated a malfunction in the aldehyde dehydrogenase protein, resulting in partial sterile and germline developmental defects. Our result is consistent with the observation that the *C. elegans alh-1(tm5823)* putative null mutant strain produces some dead animals and a highly expressed sterile phenotype (personal communication with Dr. Shohei Mitani, NBPJ). The *alh-1(tm5823)* mutant harbors a 618 bp deletion, located outside of our C172Y missense site and removes part of the NAD-binding and catalytic domains. Our conserved C172Y missense residue is also located in the NAD-binding domain. In addition, the developmental phenotype is also displayed in other animal models of the aldehyde dehydrogenase family. For example, *Aldh* null larvae and adults are less viable in the presence of ethanol in *Drosophila* (Fry & Saweikis, 2006). Moreover, mouse *Aldh3*$^{-/-}$ homozygotes were impaired in early forebrain development, possibly through failure to induce genes needed to establish regionalization (Molotkova et al., 2007). The *Aldh1a3* null mouse showed perinatal lethality that could be rescued by maternal RA treatment (Dupé et al., 2003).

The developmental defects caused by mutations in *ALDH1A3* are likely to act through alteration in the RA signaling pathway. RA binds to its nuclear RA receptor (RAR), which forms a heterodimer with retinoid X receptors (RXRs) and RA response elements (RARE). The RAR/RXR/RARE complex regulates gene transcription at specific locations during various developmental stages. In addition, RA is known to promote the differentiation of neurons prenatally; RA concentration, indicated indirectly by ALDH1A3 expression, influences the maturation of selected parts of the cerebral cortex postnatally (Wagner et al., 2006). This could be a possible explanation of a connection between *ALDH1A3* mutations and the neurological symptoms displayed in patients with intellectual disabilities and autism spectrum disorder. Our result illustrated that ALH-1(C172Y) residue is necessary for wild-type ALH-1 protein function in *C. elegans*. Since the ALH-1(C172Y) residue is conserved between *C. elegans* and humans, it is likely that this residue also has a functionally significant role in human ALDH1A3 protein. Further study in other model organisms is needed to validate this and potentially elucidate the mechanistic basis of the phenotypic consequences of the *ALDH1A3(C174Y)* variation.

4.3 Method & Reagents

Strains

The Bristol N2 *C. elegans* strain was used as the wild-type control and background for the CRISPR experiments (Brenner, 1974). The *alh-1(C172Y)* missense strain, PS7481 *alh-1(sy899)*, was generated using the protocol described previously (Wong et al., 2019). Detection primers sequence: forward primer 5'-TACAGTTATTACGCCGGATGG-3'; reverse primer 5'-CATGTGCGACGAAATAGCTTG -3'. The expected PCR product length for the wild-type is 370 bp; PCR product from mutant strain can be further digested by HaeIII (New England Biolabs,

Ipswich, MA) into two fragments of 104 and 266 bp. All strains were maintained on nematode growth medium (NGM) agar plates seeded with *Escherichia coli* OP50 at room temperature (21 ± 1◦C).

Fecundity assay

Well-fed *C. elegans* were synchronized at the L4 stage. In the fecundity assay, individual L4 hermaphrodites were placed on separate NGM plates seeded with OP50, and these animals were subsequently transferred to a new plate every day. The number of newly hatched larvae progeny was counted for every plate 1 day after the adult was transferred. The total fecundity comprises the sum of progeny produced for four days per animal.

Statistical analysis

The fecundity assay was analyzed by Mann-Whitney test using GraphPad Prism version 6 (GraphPad, La Jolla, CA).

4.4 Acknowledgement

This work was supported by Simons Foundation (SFARI award # 367560 to P.W.S.). P.W.S. was an investigator with the Howard Hughes Medical Institute during the initial part of this study.

4.5 References

Brenner, S. (1974). The genetics of Caenorhabditis elegans. *Genetics*, *77*(May), 71–94. https://doi.org/10.1111/j.1749-6632.1999.tb07894.x

Dupé, V., Matt, N., Garnier, J. M., Chambon, P., Mark, M., & Ghyselinck, N. B. (2003). A newborn lethal defect due to inactivation of retinaldehyde dehydrogenase type 3 is prevented by maternal retinoic acid treatment. *Proceedings of the National Academy of*

Sciences of the United States of America, *100*(SUPPL. 2), 14036–14041. https://doi.org/10.1073/pnas.2336223100

Fares-Taie, L., Gerber, S., Chassaing, N., Clayton-Smith, J., Hanein, S., Silva, E., Serey, M., Serre, V., Gérard, X., Baumann, C., Plessis, G., Demeer, B., Brétillon, L., Bole, C., Nitschke, P., Munnich, A., Lyonnet, S., Calvas, P., Kaplan, J., … Rozet, J. M. (2013). ALDH1A3 mutations cause recessive anophthalmia and microphthalmia. *American Journal of Human Genetics*, *92*(2), 265–270. https://doi.org/10.1016/j.ajhg.2012.12.003

Fry, J. D., & Saweikis, M. (2006). Aldehyde dehydrogenase is essential for both adult and larval ethanol resistance in Drosophila melanogaster. *Genetical Research*, *87*(2), 87–92. https://doi.org/10.1017/S0016672306008032

Ghyselinck, N. B., & Duester, G. (2019). Retinoic acid signaling pathways. *Development (Cambridge)*, *146*(13), 1–7. https://doi.org/10.1242/dev.167502

Hubbard, E. J. A., Korta, D. Z., & Dalfó, D. (2013). Physiological control of germline development. In *Advances in Experimental Medicine and Biology* (Vol. 757). https://doi.org/10.1007/978-1-4614-4015-4-5

McCaffery, P., & Drager, U. C. (1994). High levels of a retinoic acid-generating dehydrogenase in the meso- telencephalic dopamine system. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(16), 7772–7776. https://doi.org/10.1073/pnas.91.16.7772

Molotkova, N., Molotkov, A., & Duester, G. (2007). Role of retinoic acid during forebrain development begins late when Raldh3 generates retinoic acid in the ventral subventricular zone. *Developmental Biology*, *303*(2), 601–610. https://doi.org/10.1016/j.ydbio.2006.11.035

Moretti, A., Li, J., Donini, S., Sobol, R. W., Rizzi, M., & Garavaglia, S. (2016). Crystal structure of human aldehyde dehydrogenase 1A3 complexed with NAD+ and retinoic acid. *Scientific*

Reports, *6*(October), 1–12. https://doi.org/10.1038/srep35710

Roos, L., Fang, M., Dali, C., Jensen, H., Christoffersen, N., Wu, B., Zhang, J., Xu, R., Harris, P., Xu, X., Grønskov, K., & Tümer, Z. (2014). A homozygous mutation in a consanguineous family consolidates the role of ALDH1A3 in autosomal recessive microphthalmia. *Clinical Genetics*, *86*(3), 276–281. https://doi.org/10.1111/cge.12277

Wagner, E., Luo, T., Sakai, Y., Parada, L. F., & Dräger, U. C. (2006). Retinoic acid delineates the topography of neuronal plasticity in postnatal cerebral cortex. *European Journal of Neuroscience*, *24*(2), 329–340. https://doi.org/10.1111/j.1460-9568.2006.04934.x

Wong, W. R., Brugman, K. I., Maher, S., Oh, J. Y., Howe, K., Kato, M., & Sternberg, P. W. (2019). Autism-Associated missense genetic variants impact locomotion and neurodevelopment in Caenorhabditis elegans. *Human Molecular Genetics*, *28*(13), 2271–2281. https://doi.org/10.1093/hmg/ddz051

4.6 Figure

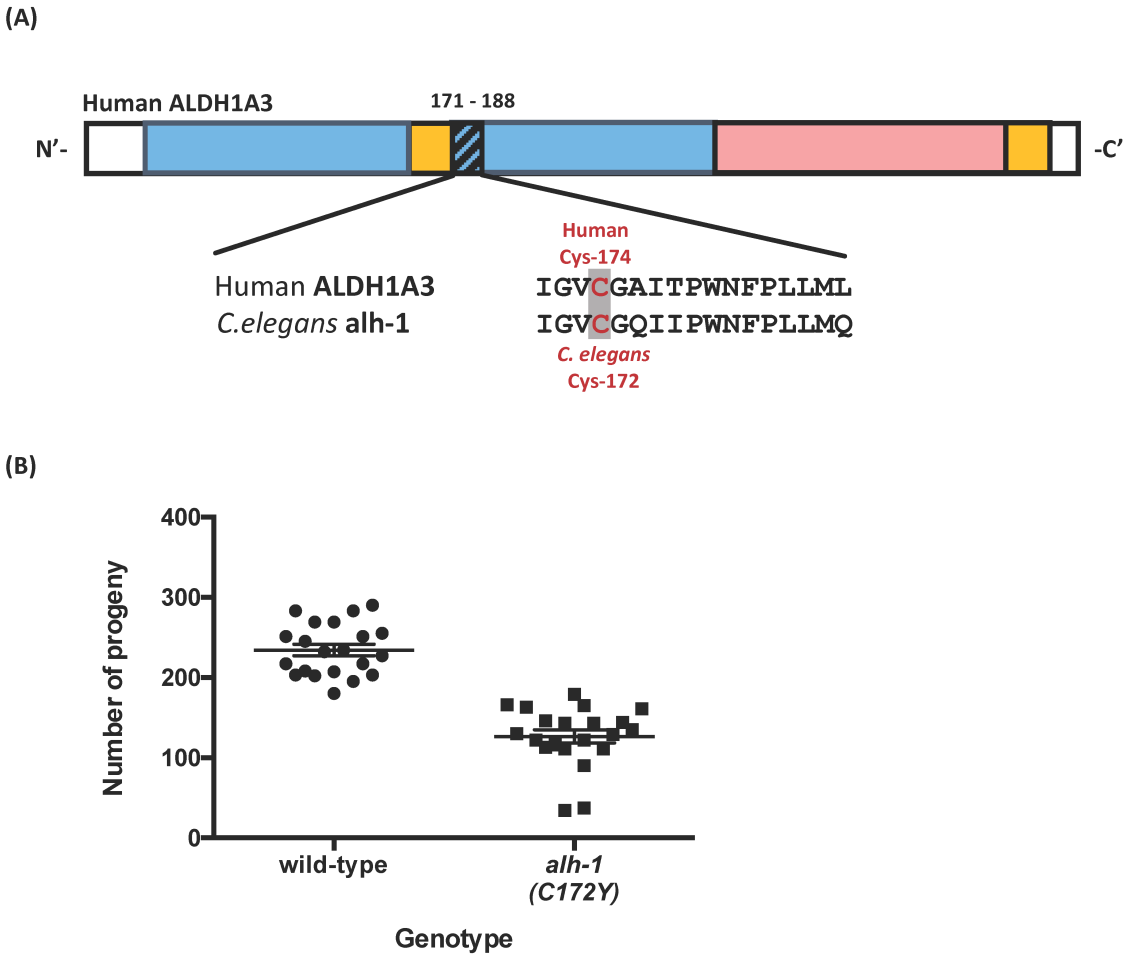

Figure 4-1. The ALDH1A3(C174Y) variant and its *C. elegans* ortholog, alh-1(C172Y). (A) Domains and sequence conservation between human *ALDH1A3* and its *C. elegans* ortholog, *alh-1*. The three domains in human ALDH1A3 are colored differently: the NAD binding domain is shown in blue, the catalytic domain in pink, and the oligomerization domain in yellow. Overall, the human ALDH1A3 is conserved with *C. elegans* ALH-1 at 77% of amino acid positions. The protein sequence alignment around the ALDH1A3 C174Y residue is highlighted in the bottom row. (B) In our *C. elegans* model, the *alh-1 (C172Y)* missense mutant displayed a significantly reduced fecundity. Approximately 20 animals were tested in each strain. Each dot represents one hermaphrodite. $P < 0.0001$ via non-parametric Mann-Whitney test.

Chapter 5

ASD-ASSOCIATED BRAF MISSENSE VARIANTS AND THEIR ROLES IN CHEMOSENSATION IN C. ELEGANS

5.1 Abstract

Sensory symptoms have been implicated in many neurodevelopmental disorders, including autism spectrum disorder (ASD) and cardio-facio-cutaneous (CFC) syndrome. One gene shared by these two neurodevelopmental disorders is *BRAF*, a critical kinase linking canonical NMDAR signaling to the MEK-ERK cascade at the synapse, which has been implicated in sensory modulation in both human and animal studies. In this study, we generated three ASD-associated missense variants of *BRAF* in *C. elegans* via CRISPR and investigated the sensory function of these missense variants via a chemotaxis assay. Our results identified that a conserved missense residue lin-45(K565N), orthologous to human BRAF K499N, displayed a hypersensitive non-dominant phenotype in the diacetyl chemotaxis assay that was capable of being inhibited by RNAi. Investigation into the expression pattern of the *lin-45* gene confirmed the expression in the RIC interneurons. This suggests a potential gain-of-function allele in BRAF, especially in the sensory function.

5.2 Introduction

A healthy human being receives hundreds of thousands of sensory inputs per second and the ability to filter relevant sensory inputs at a lower-level and integrate sensory information into higher cognitive processing is essential for mental functions. However, some individuals exhibit abnormalities in sensory processing such as hyporesponsiveness, sensory over-responsivity, and sensory-seeking behaviors, which have been implicated in many pervasive developmental disorders like the autism spectrum disorders (ASD) (Hazen et al., 2014; Tomchek & Dunn,

2007). Indeed, patients with ASD have a 69% prevalence of exhibiting sensory symptoms (Baranek et al., 2006) and the sensory symptoms occur as early as six months of age in babies later diagnosed with ASD (Estes et al., 2015). In addition, the parents and siblings of individuals with ASD also self-reported more sensory traits than the general population, suggesting a strong genetic component underlying these sensory traits (Donaldson et al., 2017). Similarly, sensory processing abnormalities have been identified in other developmental disorders, such as cardio-facio-cutaneous (CFC) syndrome (Boyd et al., 2010; Green et al., 2012; Hazen et al., 2014; Tomchek & Dunn, 2007). About 85% of children with CFC syndrome reported sensory modulation problems (Pierpont & Wolford, 2016), indicating that the disruption in essential signaling pathways results in many neurodevelopmental disorders.

ASD shares some common traits with the single-gene neurodevelopmental disorders known as RASopathies, which include CFC syndrome, Costello syndrome, Legius syndrome, Neurofibromatosis type 1, Noonan syndrome, and LEOPARD syndrome. This set of syndromes in RASopathies contain mutations in the Ras/Raf/MAPK signaling pathway (Adviento et al., 2014; Pinto et al., 2010), dysregulation of which accounts for the known genetic causes of ASD (Alfieri et al., 2014; Comings et al., 1996; Packer, 2012). While our understanding of the Ras/Raf/MAPK pathway originally came from the studies of somatic mutations in cancers, recently, the Ras/Raf/MAPK pathway has also been shown to be involved in the cell cycle, differentiation, axon projections, and dendritic spine density during development (Adviento et al., 2014; Kang & Lee, 2019; Samuels et al., 2009; Sarkozy et al., 2009). As a result, the Ras/Raf/MAPK pathway is a possible convergence point for ASD candidate genes (Levitt & Campbell, 2009) and studying its regulation could shed light on the molecular mechanism of ASD and other neurodevelopmental disorders.

A key component of the Ras/Raf/MAPK pathway is Raf (Rapidly accelerated fibrosarcoma), the enzyme linking canonical NMDAR-CaMKII-SynGap-Ras signaling to the

MEK-ERK cascade at the synapse (Lim et al., 2017). Mammalian cells contain three Raf isoforms: ARaf, BRaf, and CRaf (Raf-1). These Raf isoforms vary in their cell-specific expression and subcellular locations, yet BRaf is the only Raf kinase expressed at a high level at synapses (Morice et al., 1999). In addition, approximately 70% of individuals with CFC syndrome carry a missense mutation in *BRAF* (Niihori et al., 2006; Rodriguez-Viciana et al., 2006), and potentially active forms of BRAF arise from germline mutations in patients with RASopathies (Sarkozy et al., 2009). Many of these mutations (e.g., V600E and K499N) affect the BRaf kinase domain, as shown in the biochemical examination of kinase activity *in vitro* (Niihori et al., 2006; Rodriguez-Viciana et al., 2006). These results point to the influential role of the kinase domain in the Braf function.

The role of Braf in sensory modulation has been demonstrated in both human and animal studies. In cancer patients treated with BRAF inhibitors, such as vemurafenib, 30% or more developed aversive effects on arthralgia, fatigue, and photosensitivity reaction (Larkin et al., 2014). Another case reported acute motor and sensory axonal neuropathy after treatment with BRAF and MEK inhibitors (dabrafenib and trametinib) for six weeks (Taha et al., 2017). The role of *Braf* in the sensory system is further verified in mouse models where conditional elimination of *Braf* and *Raf1* in the dorsal root ganglion strongly reduced neurotrophin-dependent proprioceptive axon growth *in vivo* (Zhong et al., 2007) and constitutively active *Braf* expressed in neurons gated by Nav1.8 sodium channel enhanced expression of itch-sensing genes (Zhao et al., 2013). Furthermore, Raf mediates NGF-stimulated sensory axon growth (Markus et al., 2002). The effect of Braf in sensory modulation is also demonstrated in invertebrates. In the *Caenorhabditis elegans* (*C. elegans)* ortholog of BRAF, *lin-45*, loss-of-function mutant, *lin-45(sy96),* displayed defects in AWA- and AWC- chemotaxis (Hirotsu et al., 2000). Another lin-45 mutant strain, *lin-45(dx84)*, impaired the nose touch response, and Raf activation was regulated by PKC through the Raf kinase inhibitory protein (RKIP) (Hyde et al., 2011).

Additionally, CaMKI/IV and Raf pathway influenced the variability of temperature memory in AFD, and lower variability in response to temperature was found in *lin-45(dx84)* and *mpk-1* mutant strains (Kobayashi et al., 2016). However, compared to the extensive studies of BRaf in cancer, especially melanoma, how BRAF causes the sensory adaptation effect remains unclear.

To study the effects of *BRAF* mutations on sensory modulation, we modeled the human variants onto its orthologous *C. elegans* gene. With the easily accessible gene-editing techniques and the relatively short life cycle, *C. elegans* facilitates the screening of variance of the unknown significance (VUS) (Brenner, 1974; Engleman et al., 2016; Kim et al., 2017a). It is particularly useful for variants with large quantities, such as *de novo* missense mutations. In addition, compared to other model organisms, the *C. elegans* nervous system is highly amenable to cellular level resolution analysis of neural circuits and behavior given its numerically simple nervous system, known connectome, and relatively weak dependence on the nervous system for organism viability. Therefore, *C. elegans* models have been widely used to analyze the phenotype and genetic properties of candidate genes implicated in human diseases (Bai et al., 2020; Kim et al., 2017b; Levitan et al., 1996), especially in ASD (Buddell et al., 2019; Gonzalez-cavazos et al., 2019; Lipstein et al., 2017; McDiarmid et al., 2018, 2020; Post et al., 2020; Wong et al., 2019, 2021).

In this study, we modeled the functional consequences of human BRAF variants by knocking-in the equivalent amino acid substitution into its orthologous *C. elegans* gene, lin-45. We examined the genetic properties of ASD-associated missense variants on *lin-45* and examined the olfactory sensory behaviors in *C. elegans* using chemotaxis assays (Bargmann et al., 1993; Bargmann & Horvitz, 1991). A previous study demonstrated defects in diacetyl chemotaxis sensation in the *lin-45* loss-of-function mutants (Hirotsu et al., 2000). Our results have shown that a conserved missense residue orthologous to human BRAF K499N results in hypersensitivity to diacetyl chemosensation in a manner that is not dominant but can be inhibited

by *lin-45* RNAi. We have also confirmed high expression of the *lin-45* gene in the RIC interneurons. Overall, this study suggested a potential gain-of-function allele in BRAF, especially in the sensory function.

5.3 Materials and methods

Strain

The Bristol N2 *C. elegans* strain was used as the wild-type control and background for all CRISPR experiments (Brenner, 1974). The control strains for functional assays were obtained from lab stock, the Caenorhabditis Genetics Center (CGC). The *lin-45* loss-of-function mutant strains used in this study were: PS427 *lin-45(sy96)* and EJ521 *lin-45(dx19)*. All strains were maintained on nematode growth medium (NGM) agar plates seeded with *Escherichia coli* OP50 at room temperature (21 ± 1◦C).

ASD-associated conserved residue

ASD-associated missense variants were extracted from the SFARI database–Human Gene Module (Fischbach & Lord, 2010), and the conserved loci were determined using the Ensembl compara genomic method described previously (Wong et al., 2019). Targeted missense variants were introduced into the *C. elegans* genome using CRISPR/Cas9 and homology-directed repair. Additional restriction enzyme sites were also introduced via synonymous mutations to facilitate the screening process. Mutant genotypes were determined by the correct PCR and digested products length, and then confirmed by commercial Sanger sequencing company (Laragen, Culver City, CA).

Chemotaxis assays

Chemotaxis assays were adapted from previous studies (Bargmann et al., 1993). In our assay, chemotaxis was performed using 10 cm diameter chemotaxis plates prepared the night before an experiment. Chemotaxis plates contained 1.6% agar, 1 mM $MgSO_4$, 1 mM $CaCl_2$, 5 mM KPO_4 [pH 6.0]. The chemical mixture was then diluted in ethanol at the concentrations indicated. 1 µl of diluted chemical was pipetted on one side of the plate, while 1 µl of ethanol was dropped on the other side. 1 µl of 1 M sodium azide was also added on both sides to anaesthetize animals that reached odor or ethanol sources.

C. elegans strains were synchronized within 6hrs of egg-laying time. The offspring worms, the future test subjects, were then raised in 20°C. On the test day, these young adult hermaphrodites were washed three times in a S-basal buffer and once in ddH_2O. Approximately 100 animals were placed at the upper midline of chemotaxis plates with equal distance to both sides. Extra wash buffer was removed using Kimwipes and the plate was covered with a lid. The experiment was terminated after 1 hour and the number of worms within each region of interest was counted. The chemotaxis index (CI) is the number of worms on odor side minus the number of worms on diluent side divided by the sum of both sides. Assays for each condition were repeated on at least three different days.

Transgenetic GFP expression

The transcriptional reporters were built using fusion PCR (Hobert, 2002). Primers used to amplify the promoter regions and the amplified promoter sizes were as follows:

lin-45 promoter, 3047bp

forward: 5'- AAGCTTGCATGCGGCCGGCCagtcagccgagtcttgttgtc -3'

reverse: 5'- CGGGGATCCGGCGCGCCTTGAGATGATGTGACAGGTGACTT- 3'

tbh-1 promoter, 2913 bp

forward: 5'- AAGCTTGCATGCGGCCGGCCACA –3'

reverse: 5'- ACGGCACTTCTCATTTTTCTGAAATCGTAT -3'

The promoter region of *lin-45* was fused to PSM GFP plasmid and the tbh-1 promoter region was fused to a mCherry plasmid using Gibson system. Injection mixture was prepared at a concentration of 100 ng/µL Plin-45::GFP reporter construct, 30 ng/µL Ptbh-1::mCherry construct, and 15 ng/µL 1-kb DNA ladder. The transgenic strain was obtained by microinjecting the mixtures into the gonads of wildtype animals. After maintaining a few generations for stable reporter gene expression, images of adult animals were taken with a Zeiss LSM 710 Inverted confocal microscope with a ×63 Plan- APOCHROMAT objective and ZEN acquisition software.

RNA interference

The RNAi experiment was performed by feeding nematodes dsRNA-producing bacteria using standard procedures (Timmons et al., 2001) and modified as described before (Ghosh & Sternberg, 2014). Overnight starter cultures were grown with 1 ml LB supplemented with 25 mg/ml carbenicillin and 12.5 mg/ml tetracycline inoculated with a bacterial colony containing a plasmid producing dsRNA targeting a gene of interest. Starter cultures were diluted 1:80 the next day, using 100 uL of starter culture in 8 ml LB containing 25 ug/mL carbenicillin. Cultures were grown between 6 and 8 h to OD_{600} ~0.5. 6 cm Petri plates containing NGM agar that had been dried for at least three days at room temperature were prepared by using sterile glass beads to spread 50 ml of 25 mM carbenicillin, and 1 mM IPTG in M9 on each plate. RNAi cultures were then transferred to 1.5 ml microcentrifuge tubes (1 tube/plate) and then spun at 5000 rpm for 5 min. The majority of the supernatant was removed leaving about 50 µl of liquid plus bacterial pellets. Pellets were then resuspended in this solution, and then spread using sterile glass beads on the NGM agar + carbenicillin + IPTG Petri plates described in the preceding step. Plates were grown at room temperature overnight, and, if not used immediately, were stored at 4 ◦C for no more than a week. On the day of the experiment, plates were pipetted

with 50 ml of a 1:5 solution of 1 M IPTG:M9. Plates were then dried for 10 min near a Bunsen burner. Eggs were bleached onto RNAi plates and allowed to hatch and develop. Worms at L4 stage were synchronized for the chemotaxis assay. To enhance the efficiency of RNAi in neuronal cells, the RNAi experiment was done in the *rrf-3(pk1426)* mutant background.

Statistical analyses

Chemotaxis assays were analyzed by one-way ANOVA analysis in the scipy module of the Python 3.7 program. Tukey-paired multiple comparisons were performed between the wild-type group and mutant strains. The significance level was defined as $p < 0.01$.

5.4 Results

Several missense alleles have been identified from individuals with sensory symptoms in ASD and CFC syndrome (Niihori et al., 2006; Nyström et al., 2008; Rodriguez-Viciana et al., 2006; Schulz et al., 2008). This study used the Ensembl comparative genomic tools to identify the evolutionarily conserved residues between human BRAF and *C. elegans* LIN-45 (Wong et al., 2019). We found three conserved ASD-associated missense variants for *BRAF* (Figure 5-1): BRAF S467A corresponding to *C. elegans* LIN-45 S533A, BRAF K499N corresponding to *C. elegans* LIN-45 K565N, and BRAF W531C corresponding to *C. elegans* LIN-45 W597C. All three of these missense variants were located in the kinase domain, which was the most highly conserved region among the Raf protein and thus likely to be physiologically relevant. A previous study in our laboratory revealed a 58% amino acid conservation between the kinase domains of human BRAF and *C. elegans* orthologous gene, *lin-45* (Han et al., 1993). Next, we used the Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-Cas 9 system and homology-directed genome editing to introduce ("knock-in") these ASD-associated missense variants into their corresponding loci in the *C. elegans lin-*

45 gene. We aimed to characterize these *lin-45* missense alleles' functions in the diacetyl chemotaxis assay with the N2 wild-type and two known *lin-45* loss-of-function mutant controls, *sy96* and *dx19*. Both *lin-45* loss-of-function mutants were predicted to have truncated proteins before the kinase domain (Figure 5-2).

C. elegans lacks visual and auditory senses but its ability to sense chemicals is strong and the circuits are well-mapped, so we examined sensory ability in *C. elegans* using a standardized chemotaxis assay (Bargmann et al., 1993; Bargmann & Horvitz, 1991). Impaired chemo sensation to diacetyl has been documented in the loss-of-function mutant of *lin-45*, the *C. elegans* orthologous gene of *BRAF* (Hirotsu et al., 2000). For the purpose of this study, we evaluated the diacetyl chemotaxis response in *C. elegans lin-45* loss-of-function and missense mutant strains. We first determined the dose-response curve in diacetyl chemotaxis using the N2 wild-type control, a diacetyl chemosensory defected mutant strain *ord-10(ky225)*, and a *lin-45* loss-of-function strain *lin-45(sy96)*. While the attractiveness to diacetyl in the N2 animal was visible at the dilution constant of 10^3 and 10^4, *lin-45* impairment in diacetyl chemotaxis was only detectable at the concentration of 10^4 dilution. Therefore, we decided to use a less concentrated diacetyl odor in our chemotaxis assay. At the concentration of 10^4 dilution (Figure 5-3), wild-type N2 animals showed a slight attraction to diacetyl, presented as a 0.40 ± 0.03 chemotaxis index. Both *lin-45* loss-of-function mutant strains, *lin-45(sy96)* and *lin-45(dx19)*, exhibited impaired diacetyl chemotaxis to the same level as the diacetyl chemotaxis null mutant, *odr-10(ky225)* (CI: *lin-45(sy96)*$= 0.01 \pm 0.07$, *lin-45(dx19)*$= 0.07 \pm 0.08$, *odr-10(ky225)* $= 0.04 \pm 0.05$; $p > 0.05$). Among the three missense mutant strains, only the *lin-45(K565N)* missense mutant strain displayed hypersensitivity to diacetyl while the other two missense strains, *lin-45(S533A)* and *lin-45(W597C)*, did not show any differences in chemotaxis index to the wild-type controls (CI: *lin-45(K565N)*$= 0.57 \pm 0.04$, $p = 0.03$; *lin-45(S533A)*$= 0.40 \pm 0.08$, $p > 0.05$; *lin-45(W597C)*$= 0.49$

± 0.03; $p > 0.05$ compared to N2). These results pointed out a potential gain-of-function missense variants in the *C. elegans* ortholog of *BRAF* in diacetyl chemotaxis response.

Many of the gain-of-function variants typically demonstrate dominant gene properties. Naturally, we were interested in finding out whether our *lin-45(K565N)* locus had a dominance effect (Figure 5-4). To study the *lin-45* missense allele's dominance, we used a balancer tmC25, which carried a venus green fluorescent reporter, to distinguish heterozygous animals from homozygous ones (Dejima et al., 2018). As reported in the previous figure, the homozygous *lin-45(K565N)* missense mutant showed higher chemotaxis index than wild-type controls, indicating that more lin-45(K565N) mutant worms were able to distinguish the diacetyl odor (CI: N2= 0.39 ± 0.04; homozygous *lin-45(K565N)* = 0.57 ± 0.04; $p = 0.01$ compared to N2). However, this effect did not show in the heterozygous *lin-45(K565N)* animals. Both wildtype and *lin-45(K565N)* heterozygous animals displayed similar chemotaxis indexes as the wild-type homozygous animals (CI: heterozygous wildtype = 0.28 ± 0.11; heterozygous *lin-45(K565N)* = 0.39 ± 0.05; $p > 0.05$ compared to N2). This finding suggests that the *lin-45(K565N)* missense variant is not a dominant allele.

To ensure the hypersensitivity phenotype truly resulted from the variant changes in the *lin-45* gene, we applied an RNA interference (RNAi) experiment to test the necessity of *lin-45* in diacetyl chemosensory behaviors (Figure 5-5). In order to enhance the RNAi efficiency in neuronal cells, we crossed both *lin-45(wild-type)* and *lin-45(K565N)* alleles into the RNAi sensitive background strain, *rrf-3(pk1426)*. We also included the non-IPTG-activated RNAi groups in the *rrf-3(pk1426)* background for both genotypes. Our results showed that the *lin-45(K565N)* missense mutant treated with anti-lin45 RNAi presented a lower chemotaxis index than the vector-treated *lin-45(K565N)* animals (CI: RNAi treated *lin-45(K565N)* = 0.25 ± 0.04; vector-treated *lin-45(K565N)* = 0.56 ± 0.04; $p = 0.001$). In addition, the anti-*lin45* RNAi also impaired the diacetyl chemotaxis behaviors in wild-type animals (CI: RNAi treated wild-type

= 0.28 ± 0.05; vector-treated wild-type = 0.49 ± 0.04; $p = 0.0056$). Our results showed that the function of *lin-45(wild-type)* and *lin-45(K565N)* alleles can be dampened by the anti-*lin-45* RNAi treatments, demonstrating the necessity of *lin-45* in the diacetyl chemotaxis.

Knowing the effect of lin-45(K565N) allele on diacetyl chemotaxis response, we are interested in the site-of-action. To verify the expression of the *lin-45* gene, we injected an extrachromosomal array containing a 3kb *lin-45* promoter region fused with the GFP gene such that cells expressing the *lin-45* gene would also express the *GFP* reporter gene. We found that the *lin-45* gene was expressed in several head neurons as well as some cells in the vulva and tail regions of young adult hermaphrodites. Based on the location and morphology, we suspected one pair of the *lin-45* expressing neurons near the pharynx were the RIC interneurons. Therefore, we co-injected the *lin-45* promoter-driven GFP reporter with RIC-specific *tbh-1* promoter-driven mCherry reporters to confirm the identity of these neurons. As shown in Figure 5-6, we have identified one pair of the head neurons as the RIC interneurons, the major class of octopaminergic interneurons in *C. elegans*.

5.5 Discussion

In this study, we examined the genetic properties of ASD-associated missense variants on *lin-45* and examined the olfactory sensory behaviors in *C. elegans* using chemotaxis assays (Bargmann et al., 1993; Bargmann & Horvitz, 1991). Consistent with a previous study, *lin*-45 loss-of-function mutant strains displayed an impaired ability to distinguish the diacetyl odor in a chemotaxis assay (Hirotsu et al., 2000). More importantly, our assay revealed that one conserved missense residue orthologous to human *BRAF* K499N showed a hypersensitive phenotype in the diacetyl chemotaxis assay. In addition, the RNA interference experiment demonstrated the necessity of *lin-45* in the hypersensitive response to diacetyl. And lastly, we confirmed the expression of *lin-45* in the RIC interneurons via the transgenic experiment. Overall, these results

suggested a potential gain-of-function allele in the *C. elegans* ortholog of *BRAF*, especially in the sensory function.

Consistent with our gain-of-function phenotype in diacetyl chemotaxis assay, enhanced enzyme activities in the *BRAF(K499N)* locus have been documented in multiple cell lines. For example, an examination of the activation of ELK transcriptional factor in several BRAF missense mutant cell lines found a two- to four-fold increase in relative luciferase activity in cells transfected with the *BRAF(K499N)* mutation (Niihori et al., 2006). Rodriguez-Viciana et al. also revealed a two-fold increase in kinase activity in human embryonic kidney 293T cells transfected with the *BRAF(K499N)* mutations (Rodriguez-Viciana et al., 2006). In addition, the crystal structure of the BRAF kinase domain showed that the *BRAF(K499N)* mutation was located in the interface of the ATP binding cleft, suggesting that the K499N mutation may alter the catalytic activity of the kinase domain (Niihori et al., 2006). The CFC-associated BRAF mutants with elevated kinase activity, such as K499N, have been shown to induce MEK and ERK phosphorylation, activating the downstream signaling pathway.

According to the RNA-seq data on WormBase, *lin-45* was expressed broadly at low levels. Our GFP reporter gene results showed high expression of *lin-45* in the RIC interneuron, which is distantly connected to the pair of sensory neurons, AWAL and AWAR, responsible for the diacetyl chemotaxis response. AWA sensory neurons (AWAL and AWAR are collectively called AWA) form synaptic connections onto the AIZ and AIY interneurons. The AIY synapses to the backward motor command neuron AVE and RMD. The AVE interneuron connects to the backward command neuron AVA, to which RIC is distantly connected. It remains unclear at which level the highly expressed *lin-45* in RIC neurons mediate the diacetyl chemosensory behaviors. However, the RIC interneuron has been shown to provide a neurohumoral feedback loop to primary sensory neuron AWB via SER-6 signaling to regulate pharyngeal pumping behaviors (Liu et al., 2019).

Evidence of neuroimaging studies suggested that autistic sensory traits originated at low-level sensory processing (Robertson & Baron-Cohen, 2017). Yet the lack of directly comparable behavioral paradigms restricted the translational research on sensory behaviors. Most rodent studies use the startle response or pre-pulse inhibition assay to evaluate "hypersensitivity" in the disease animal models, and they often showed mixed results (Kohl et al., 2013; Madsen et al., 2014; McAlonan et al., 2002). Here, we presented a unique approach of using *C. elegans* to study the effect of ASD-associated missense mutations in olfactory sensory behaviors. Our findings revealed the genetic property of *lin-45(K565N)*, the orthologous locus of *BRAF(K499N)*, and prioritized this missense variant for future studies in sensory modulation.

5.6 References

Adviento, B., Corbin, I. L., Widjaja, F., Desachy, G., Enrique, N., Rosser, T., Risi, S., Marco, E. J., Hendren, R. L., Bearden, C. E., Rauen, K. A., & Weiss, L. A. (2014). Autism traits in the RASopathies. *Journal of Medical Genetics*, *51*(1), 10–20. https://doi.org/10.1136/jmedgenet-2013-101951

Alfieri, P., Piccini, G., Caciolo, C., Perrino, F., Gambardella, M. L., Mallardi, M., Cesarini, L., Leoni, C., Leone, D., Fossati, C., Selicorni, A., Digilio, M. C., Tartaglia, M., Mercuri, E., Zampino, G., & Vicari, S. (2014). Behavioral profile in RASopathies. *American Journal of Medical Genetics, Part A*, *164*(4), 934–942. https://doi.org/10.1002/ajmg.a.36374

Bai, X., Bouffard, J., Lord, A., Brugman, K., Sternberg, P. W., Cram, E. J., & Golden, A. (2020). Caenorhabditis elegans piezo channel coordinates multiple reproductive tissues to govern ovulation. *ELife*, *9*, 1–35. https://doi.org/10.7554/eLife.53603

Baranek, G. T., David, F. J., Poe, M. D., Stone, W. L., & Watson, L. R. (2006). Sensory Experiences Questionnaire: Discriminating sensory features in young children with autism,

developmental delays, and typical development. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *47*(6), 591–601. https://doi.org/10.1111/j.1469-7610.2005.01546.x

Bargmann, C. I., Hartwieg, E., & Horvitz, H. R. (1993). Odorant- selective genes and neurons mediate olfation in C. elegans. *Cell*, *74*, 515–527.

Bargmann, C. I., & Horvitz, H. R. (1991). Chemosensory neurons with overlapping functions direct chemotaxis to multiple chemicals in C. elegans. *Neuron*, *7*, 729–742. http://www.wormbase.org/db/misc/paper?name=WBPaper00001481

Boyd, B. A., Baranek, G. T., Sideris, J., Poe, M. D., Watson, L. R., Patten, E., & Miller, H. (2010). Sensory features and repetitive behaviors in children with autism and developmental delays. *Autism Research*, *3*(2), 78–87. https://doi.org/10.1002/aur.124.Sensory

Brenner, S. (1974). The genetics of Caenorhabditis elegans. *Genetics*, *77*(May), 71–94. https://doi.org/10.1111/j.1749-6632.1999.tb07894.x

Buddell, T., Friedman, V., Drozd, C. J., & Quinn, C. C. (2019). An autism-causing calcium channel variant functions with selective autophagy to alter axon targeting and behavior. *PLoS Genetics*, *15*(12), 1–20. https://doi.org/10.1371/journal.pgen.1008488

Comings, D. E., Wu, S., Chiu, C., Muhleman, D., & Sverd, J. (1996). Studies of the c-Harvey-Ras gene in psychiatric disorders. *Psychiatry Research*, *63*(1), 25–32. https://doi.org/10.1016/0165-1781(96)02829-6

Dejima, K., Hori, S., Iwata, S., Suehiro, Y., Yoshina, S., Motohashi, T., & Mitani, S. (2018). An Aneuploidy-Free and Structurally Defined Balancer Chromosome Toolkit for Caenorhabditis elegans. *Cell Reports*, *22*(1), 232–241. https://doi.org/10.1016/j.celrep.2017.12.024

Donaldson, C. K., Stauder, J. E. A., & Donkers, F. C. L. (2017). Increased Sensory Processing Atypicalities in Parents of Multiplex ASD Families Versus Typically Developing and

Simplex ASD Families. *Journal of Autism and Developmental Disorders*, *47*(3), 535–548. https://doi.org/10.1007/s10803-016-2888-0

Engleman, E. A., Katner, S. N., & Neal-beliveau, B. S. (2016). Caenorhabditis elegans as a model to study the molecular and genetic mechanisms of drug addiction. *Prog Mol Biol Transl Sci*, *137*, 229–252. https://doi.org/10.1016/bs.pmbts.2015.10.019.Caenorhabditis

Estes, A., Zwaigenbaum, L., Gu, H., St. John, T., Paterson, S., Elison, J. T., Hazlett, H., Botteron, K., Dager, S. R., Schultz, R. T., Kostopoulos, P., Evans, A., Dawson, G., Eliason, J., Alvarez, S., & Piven, J. (2015). Behavioral, cognitive, and adaptive development in infants with autism spectrum disorder in the first 2 years of life. *Journal of Neurodevelopmental Disorders*, *7*(1), 1–10. https://doi.org/10.1186/s11689-015-9117-6

Fischbach, G. D., & Lord, C. (2010). The simons simplex collection: A resource for identification of autism genetic risk factors. *Neuron*, *68*(2), 192–195. https://doi.org/10.1016/j.neuron.2010.10.006

Ghosh, S., & Sternberg, P. W. (2014). Spatial and molecular cues for cell outgrowth during C. elegans uterine development. *Developmental Biology*, *396*(1), 121–135. https://doi.org/10.1016/j.ydbio.2014.09.028

Gonzalez-cavazos, C., Cao, M., Wong, W., Chai, C., & Sternberg, P. W. (2019). Effects of ASD-associated daf-18/PTEN missense variants on C. elegans dauer development. *MicroPubl Biol*, *10*, 17912.

Green, S. A., Ben-Sasson, A., Soto, T. W., & Carter, A. S. (2012). Anxiety and sensory over-responsivity in toddlers with autism spectrum disorders: Bidirectional effects across time. *Journal of Autism and Developmental Disorders*, *42*(6), 1112–1119. https://doi.org/10.1007/s10803-011-1361-3

Han, M., Goldenh, A., Han, Y., & Sternberg, P. W. (1993). C. elegans lin-45 raf gene participates in let-60 ras-stimulated vulval differentiation. *Nature*, *363*, 133–140.

Hazen, E. P., Stornelli, J. L., O'Rourke, J. A., Koesterer, K., & McDougle, C. J. (2014). Sensory symptoms in autism spectrum disorders. *Harvard Review of Psychiatry*, *22*(2), 112–124. https://doi.org/10.1097/01.HRP.0000445143.08773.58

Hirotsu, T., Saeki, S., Yamamoto, M., & Iino, Y. (2000). The Ras-MAPK pathway is important for olfaction in Caenorhabditis elegans. *Nature*, *404*(6775), 289–293. https://doi.org/10.1038/nature03169

Hobert, O. (2002). PCR fusion-based approach to create reporter Gene constructs for expression analysis in transgenic C. elegans. *BioTechniques*, *32*(4), 728–730. https://doi.org/10.2144/02324bm01

Hyde, R., Corkins, M. E., Somers, G. A., & Hart, A. C. (2011). PKC-1 acts with the ERK MAPK signaling pathway to regulate Caenorhabditis elegans mechanosensory response. *Genes, Brain and Behavior*, *10*(3), 286–298. https://doi.org/10.1111/j.1601-183X.2010.00667.x

Kang, M., & Lee, Y. S. (2019). The impact of RASopathy-associated mutations on CNS development in mice and humans. *Molecular Brain*, *12*(1), 1–17. https://doi.org/10.1186/s13041-019-0517-5

Kim, S., Twigg, S. R. F., Scanlon, V. A., Chandra, A., Hansen, T. J., Alsubait, A., Fenwick, A. L., McGowan, S. J., Lord, H., Lester, T., Sweeney, E., Weber, A., Cox, H., Wilkie, A. O. M., Golden, A., & Corsi, A. K. (2017a). Localized TWIST1 and TWIST2 basic domain substitutions cause four distinct human diseases that can be modeled in Caenorhabditis elegans. *Human Molecular Genetics*, *26*(11), 2118–2132. https://doi.org/10.1093/hmg/ddx107

Kim, S., Twigg, S. R. F., Scanlon, V. A., Chandra, A., Hansen, T. J., Alsubait, A., Fenwick, A. L., McGowan, S. J., Lord, H., Lester, T., Sweeney, E., Weber, A., Cox, H., Wilkie, A. O. M., Golden, A., & Corsi, A. K. (2017b). Localized TWIST1 and TWIST2 basic domain substitutions cause four distinct human diseases that can be modeled in Caenorhabditis

elegans. *Human Molecular Genetics*, *26*(11), 2118–2132. https://doi.org/10.1093/hmg/ddx107

Kobayashi, K., Nakano, S., Amano, M., Tsuboi, D., Nishioka, T., Ikeda, S., Yokoyama, G., Kaibuchi, K., & Mori, I. (2016). Single-Cell Memory Regulates a Neural Circuit for Sensory Behavior. *Cell Reports*, *14*(1), 11–21. https://doi.org/10.1016/j.celrep.2015.11.064

Kohl, S., Heekeren, K., Klosterkötter, J., & Kuhn, J. (2013). Prepulse inhibition in psychiatric disorders--apart from schizophrenia. *Journal of Psychiatric Research*, *47*(4), 445–452. https://doi.org/10.1016/j.jpsychires.2012.11.018

Larkin, J., Del Vecchio, M., Ascierto, P. A., Krajsova, I., Schachter, J., Neyns, B., Espinosa, E., Garbe, C., Sileni, V. C., Gogas, H., Miller, W. H., Mandalà, M., Hospers, G. A. P., Arance, A., Queirolo, P., Hauschild, A., Brown, M. P., Mitchell, L., Veronese, L., & Blank, C. U. (2014). Vemurafenib in patients with BRAFV600 mutated metastatic melanoma: An open-label, multicentre, safety study. *The Lancet Oncology*, *15*(4), 436–444. https://doi.org/10.1016/S1470-2045(14)70051-8

Levitan, D., Doyle, T. G., Brousseau, D., Lee, M. K., Thinakaran, G., Slunt, H. H., Sisodia, S. S., & Greenwald, I. (1996). Assessment of normal and mutant human presenilin function in Caenorhabditis elegans. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(25), 14940–14944. https://doi.org/10.1073/pnas.93.25.14940

Levitt, P., & Campbell, D. B. (2009). The genetic and neurobiologic compass points toward common signaling dysfunctions in autism spectrum disorders. *The Journal of Clinical Investigation*, *119*(4), 747–754. https://doi.org/10.1172/JCI37934.The

Lim, C. S., Kang, X., Mirabella, V., Zhang, H., Bu, Q., Araki, Y., Hoang, E. T., Wang, S., Shen, Y., Choi, S., Kaang, B. K., Chang, Q., Pang, Z. P., Huganir, R. L., & Zhu, J. J. (2017). BRaf signaling principles unveiled by large-scale human mutation analysis with a rapid lentivirus-based gene replacement method. *Genes & Development*, *31*(6), 537–552.

https://doi.org/10.1101/gad.294413.116

Lipstein, N., Verhoeven-Duif, N. M., Michelassi, F. E., Calloway, N., Van Hasselt, P. M., Pienkowska, K., Van Haaften, G., Van Haelst, M. M., Van Empelen, R., Cuppen, I., Van Teeseling, H. C., Evelein, A. M. V., Vorstman, J. A., Thoms, S., Jahn, O., Duran, K. J., Monroe, G. R., Ryan, T. A., Taschenberger, H., … Brose, N. (2017). Synaptic UNC13A protein variant causes increased neurotransmission and dyskinetic movement disorder. *Journal of Clinical Investigation*, *127*(3), 1005–1018. https://doi.org/10.1172/JCI90259

Liu, H., Qin, L. W., Li, R., Zhang, C., Al-Sheikh, U., & Wu, Z. X. (2019). Reciprocal modulation of 5-HT and octopamine regulates pumping via feedforward and feedback circuits in C. elegans(Proceedings of the National Academy of Sciences of the United States of America(2019)116(7107–7112)Doi:10.1073/pnas.1819261116). *Proc. Natl. Acad. Sci. U.S.A.*, *116*(21), 10598. https://doi.org/10.1073/pnas.1906704116

Madsen, G. F., Bilenberg, N., Cantio, C., & Oranje, B. (2014). Increased prepulse inhibition and sensitization of the startle reflex in autistic children. *Autism Research*, *7*(1), 94–103. https://doi.org/10.1002/aur.1337

Markus, A., Zhong, J., & Snider, W. D. (2002). Raf and Akt mediate distinct aspects of sensory axon growth. *Neuron*, *35*(1), 65–76. https://doi.org/10.1016/S0896-6273(02)00752-3

McAlonan, G. M., Daly, E., Kumari, V., Critchley, H. D., Van Amelsvoort, T., Suckling, J., Simmons, A., Sigmundsson, T., Greenwood, K., Russell, A., Schmitz, N., Happe, F., Howlin, P., & Murphy, D. G. M. (2002). Brain anatomy and sensorimotor gating in Asperger's syndrome. *Brain*, *125*(7), 1594–1606. https://doi.org/10.1093/brain/awf150

McDiarmid, T. A., Au, V., Loewen, A. D., Liang, J., Mizumoto, K., Moerman, D. G., & Rankin, C. H. (2018). CRISPR-Cas9 human gene replacement and phenomic characterization in *Caenorhabditis elegans* to understand the functional conservation of human genes and decipher variants of uncertain significance. *Disease Models & Mechanisms*, dmm.036517.

https://doi.org/10.1242/dmm.036517

McDiarmid, T. A., Belmadani, M., Liang, J., Meili, F., Mathews, E. A., Mullen, G. P., Hendi, A., Wong, W. R., Rand, J. B., Mizumoto, K., Haas, K., Pavlidis, P., & Rankin, C. H. (2020). Systematic phenomics analysis of autism-associated genes reveals parallel networks underlying reversible impairments in habituation. *P Natl Acad Sci USA*, *117*(1), 656–667. https://doi.org/10.1073/pnas.1912049116

Morice, C., Nothias, F., König, S., Vernier, P., Baccarini, M., Vincent, J. D., & Barnier, J. V. (1999). Raf-1 and B-Raf proteins have similar regional distributions but differential subcellular localization in adult rat brain. *European Journal of Neuroscience*, *11*(6), 1995–2006. https://doi.org/10.1046/j.1460-9568.1999.00609.x

Niihori, T., Aoki, Y., Narumi, Y., Neri, G., Cavé, H., Verloes, A., Okamoto, N., Hennekam, R. C. M., Gillessen-Kaesbach, G., Wieczorek, D., Kavamura, M. I., Kurosawa, K., Ohashi, H., Wilson, L., Heron, D., Bonneau, D., Corona, G., Kaname, T., Naritomi, K., … Matsubara, Y. (2006). Germline KRAS and BRAF mutations in cardio-facio-cutaneous syndrome. *Nature Genetics*, *38*(3), 294–296. https://doi.org/10.1038/ng1749

Nyström, A. M., Ekvall, S., Berglund, E., Björkqvist, M., Braathen, G., Duchen, K., Enell, H., Holmberg, E., Holmlund, U., Olsson-Engman, M., Annerén, G., & Bondeson, M. L. (2008). Noonan and cardio-facio-cutaneous syndromes: Two clinically and genetically overlapping disorders. *Journal of Medical Genetics*, *45*(8), 500–506. https://doi.org/10.1136/jmg.2008.057653

Packer, A. (2012). *RAS Pathway, a potentially unifying theory of autism*. *March*, 1–7.

Pierpont, E. I., & Wolford, M. (2016). Behavioral functioning in cardiofaciocutaneous syndrome: Risk factors and impact on parenting experience. *American Journal of Medical Genetics, Part A*, *170*(8), 1974–1988. https://doi.org/10.1002/ajmg.a.37725

Pinto, D., Pagnamenta, A. T., Klei, L., Anney, R., Merico, D., Regan, R., Conroy, J., Magalhaes,

T. R., Correia, C., Abrahams, B. S., Almeida, J., Bacchelli, E., Bader, G. D., Bailey, A. J., Baird, G., Battaglia, A., Berney, T., Bolshakova, N., Bölte, S., … Betancur, C. (2010). Functional impact of global rare copy number variation in autism spectrum disorders. *Nature*, *466*(7304), 368–372. https://doi.org/10.1038/nature09146

Post, K. L., Belmadani, M., Ganguly, P., Meili, F., Dingwall, R., McDiarmid, T. A., Meyers, W. M., Herrington, C., Young, B. P., Callaghan, D. B., Rogic, S., Edwards, M., Niciforovic, A., Cau, A., Rankin, C. H., O'Connor, T. P., Bamji, S. X., Loewen, C. J. R., Allan, D. W., … Haas, K. (2020). Multi-model functionalization of disease-associated PTEN missense mutations identifies multiple molecular mechanisms underlying protein dysfunction. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-15943-0

Robertson, C. E., & Baron-Cohen, S. (2017). Sensory perception in autism. *Nature Reviews Neuroscience*, *18*(11), 671–684. https://doi.org/10.1038/nrn.2017.112

Rodriguez-Viciana, P., Tetsu, O., Tidyman, W. E., Estep, A. L., Conger, B. A., Cruz, M. S., McCormick, F., & Rauen, K. A. (2006). Germline mutations in genes within the MAPK pathway cause cardio-facio-cutaneous syndrome. *Science*, *311*(5765), 1287–1290. https://doi.org/10.1126/science.1124642

Samuels, I. S., Saitta, S. C., & Landreth, G. E. (2009). Mapping CNS development and cognition: an ERKsome process. *Neuron*, *61*(2), 160–167. https://doi.org/10.1038/jid.2014.371

Sarkozy, A., Carta, C., Moretti, S., Zampino, G., Digilio, M. C., Pantaleoni, F., Scioletti, A. P., Esposito, G., Cordeddu, V., Lepri, F., Petrangeli, V., Dentici, M. L., Mancini, G. M. S., Selicorni, A., Rossi, C., Mazzanti, L., Marino, B., Ferrero, G. B., Silengo, M. C., … Tartaglia, M. (2009). Germline BRAF mutations in Noonan, LEOPARD, and cardiofaciocutaneous syndromes: molecular diversity and associated phenotypic spectrum. *Human Mutation*, *30*(4), 695–702. https://doi.org/10.1002/humu.20955.Germline

Schulz, A. L., Albrecht, B., Arici, C., van der Burgt, I., Buske, A., Gillessen-Kaesbach, G.,

Heller, R., Horn, D., Hübner, C. A., Korenke, G. C., König, R., Kress, W., Krüger, K., Meinecke, P., Mücke, J., Plecko, B., Rossier, E., Schinzel, P., Schulze, A., … Zenker, W. (2008). Mutation and phenotypic spectrum in patients with cardio-facio-cutaneous and Costello syndrome. *Clinical Genetics*, *73*(1), 62–70. https://doi.org/10.1111/j.1399-0004.2007.00931.x

Timmons, L., Court, D. L., & Fire, A. (2001). Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in Caenorhabditis elegans. *Gene*, *263*, 103–112. www.elsevier.com/locate/gene

Tomchek, S. D., & Dunn, W. (2007). Sensory processing in children with and without autism: a comparative study using the short sensory profile. *American Journal of Occupational Therapy*, *61*(2), 190–200. https://doi.org/10.5014/ajot.61.2.190

Wong, W. R., Brugman, K. I., Maher, S., Oh, J. Y., Howe, K., Kato, M., & Sternberg, P. W. (2019). Autism-associated missense genetic variants impact locomotion and neurodevelopment in Caenorhabditis elegans. *Human Molecular Genetics*, *28*(13), 2271–2281. https://doi.org/10.1093/hmg/ddz051

Wong, W. R., Maher, S., Oh, J. Y., Brugman, K. I., Gharib, S., & Sternberg, P. W. (2021). Conserved missense variant in ALDH1A3 ortholog impairs fecundity in C . elegans. *MicroPubl Biol*, https://doi.org/10.17912/micropub.biology.000357.

Zhao, Z. Q., Huo, F. Q., Jeffry, J., Hampton, L., Demehri, S., Kim, S., Liu, X. Y., Barry, D. M., Wan, L., Liu, Z. C., Li, H., Turkoz, A., Ma, K., Cornelius, L. A., Kopan, R., Battey, J. F., Zhong, J., & Chen, Z. F. (2013). Chronic itch development in sensory neurons requires BRAF signaling pathways. *Journal of Clinical Investigation*, *123*(11), 4769–4780. https://doi.org/10.1172/JCI70528

Zhong, J., Li, X., McNamee, C., Chen, A. P., Baccarini, M., & Snider, W. D. (2007). Raf kinase signaling functions in sensory neuron differentiation and axon growth in vivo. *Nat Neurosci*,

10(5), 598–607. https://doi.org/10.1038/1898

5.7 Tables and figures

Figure 5-1. Protein sequence alignment between human BRAF and *C. elegans* LIN-45 in the kinase domain.
The human BRAF is conserved with *C. elegans* LIN-45 at 58% of amino acid positions in the kinase domain. Using the Ensembl comparative method described in a previous study (see method), we identified three ASD-associated conserve missense variants on BRAF, S467A, K499N, and W531C. Their corresponding residues in the *C. elegans* orthologs were S533A, K565N, and W597C. All variants were located in the kinase domain (indicated by the blue arrow).

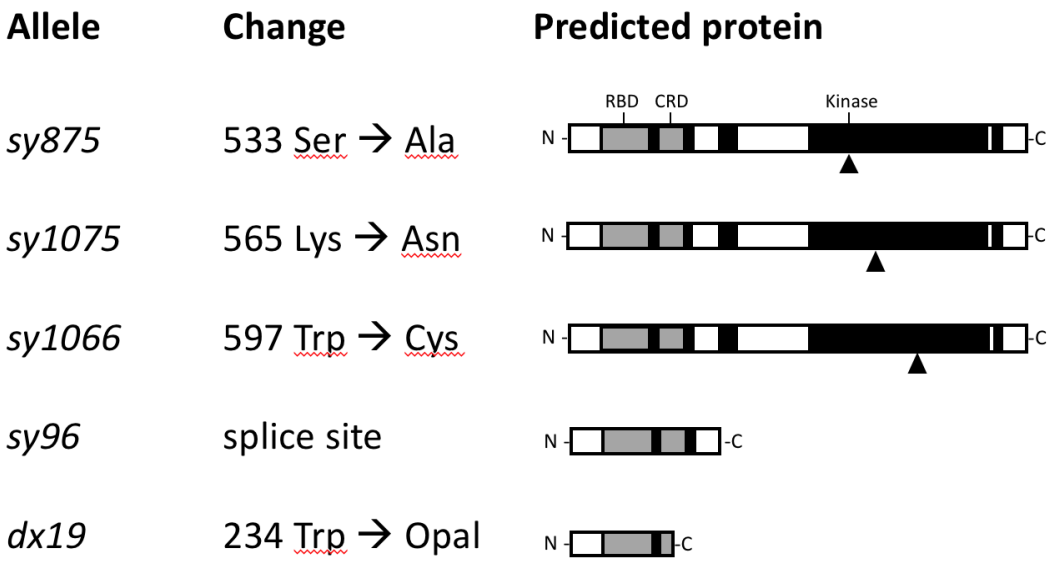

Figure 5-2. The relative position of the *lin-45* alleles and their predicted protein products.
The *C. elegans lin-45* gene contains three conserved regions. Conserved region 1 (CR1) includes a Cysteine-rich domain (CRD) and a Ras-binding domain (RBD). CRD is a zinc-finger-like domain that presumably negatively regulates the intrinsic kinase activity of the Raf protein. The inhibition is relieved upon binding to a regulatory factor. Conserved region 2 (CR2) consists of a serine/threonine-rich region that regulates phosphorylation sites. Conserved region 3 (CR3) is the kinase domain, the most highly conserved region among the Raf protein. The *lin-45* mutant alleles used in this study and their protein products are shown. The triangles indicate the relative position of missense variant change. Figure adapted from (Hsu et al., 2002).

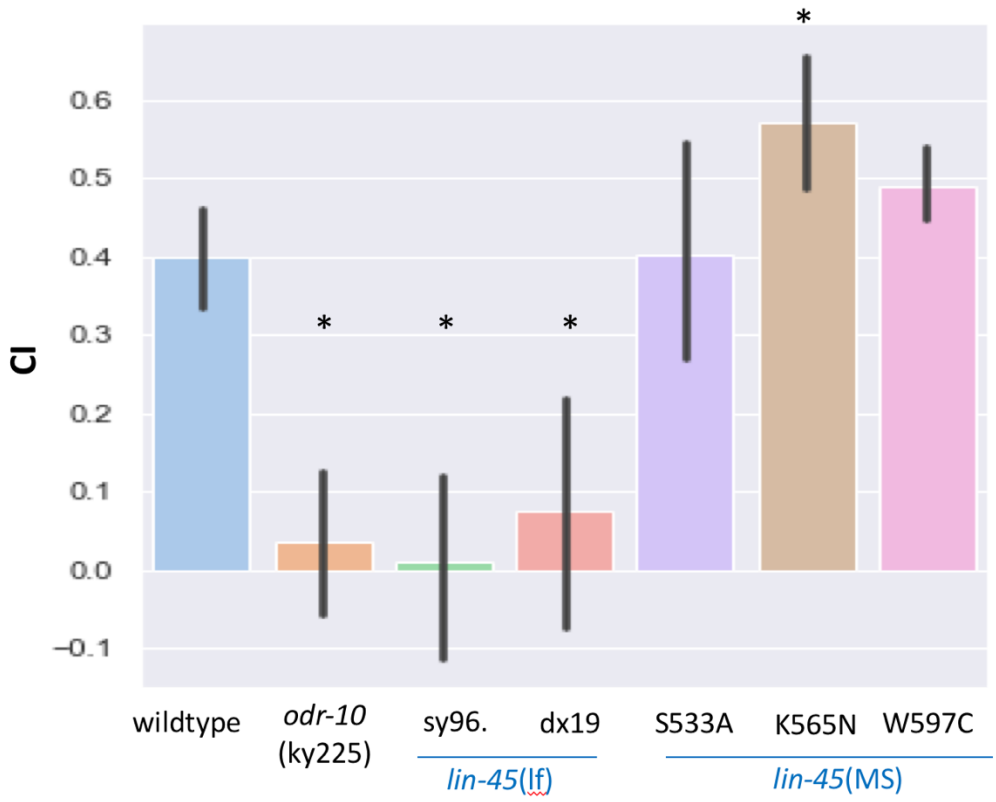

Figure 5-3. The chemotaxis phenotype in *lin-45* missense and null mutant strains.
We tested two *lin-45* loss-of-function(lf) strains and three *lin-45* missense (MS) mutant strains with the 1:10000 diluted diacetyl. Compared to the N2 wild-type controls, both *lin-45(lf)* strains exhibited impaired diacetyl chemotaxis. Among the missense mutant strains, only *lin-45(K565N)* missense mutant strain displayed hypersensitive to diacetyl while the other two missense strains showed similar chemotaxis index (CI) as the wild-type animals. The known diacetyl chemotaxis defected strain, odr-10(ky225), was also included as a positive control. Each dot represented one plate of ~100 animals. The height of box indicated the mean and standard deviation of the strain. Error bars represented the maximum and minimum values. All data were analyzed using one-way ANOVA and Tukey multiple comparison. * indicated significantly different with comparison to the wild-type animals ($p < 0.01$).

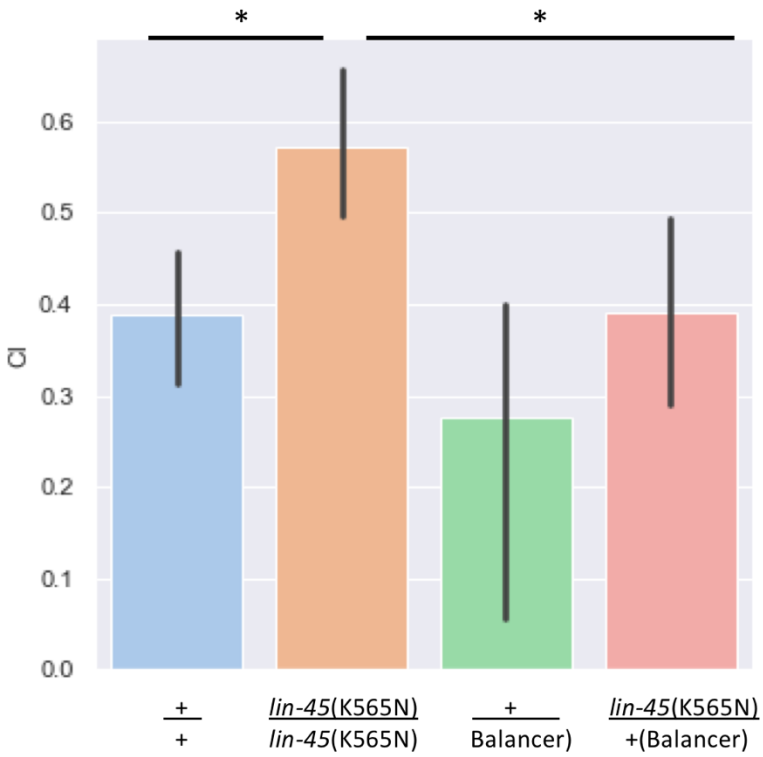

Figure 5-4. Chemotaxis phenotype in the lin-45(K565N) heterozygous strains.
Both wild-type and missense mutant *lin-45* strains were crossed with a fluorescent balancer reporter (tmC25) to distinguish heterozygous and homozygous states. The heterozygous *lin-45(K565N)* strain (red bar) displayed similar chemotaxis index (CI) as the wildtype homozygous (blue bar) or heterozygous (green bar) animals. Each dot represented one plate of ~150 animals. The height of box indicated the mean and standard deviation of the strain. Error bars represented the maximum and minimum values. All data were analyzed using one-way ANOVA and Tukey multiple comparison. * indicated significantly different with comparison to the wild-type animals ($p < 0.01$).

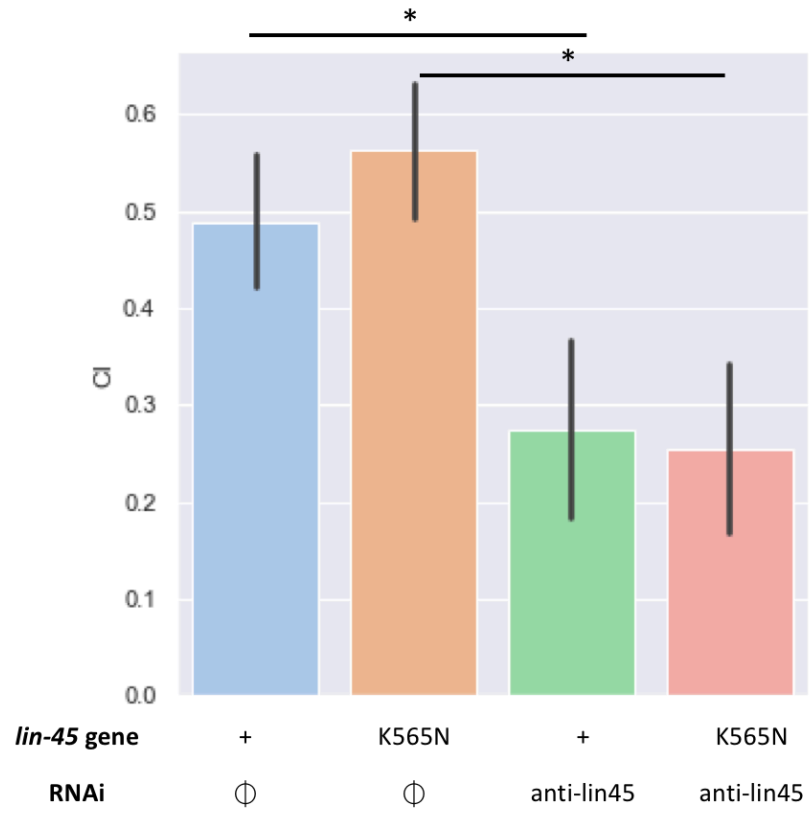

Figure 5-5. Chemotaxis phenotype in the *lin-45(K565N)* strains treated with *lin-45* RNAi.

The lin-45 RNAi treatment impaired diacetyl chemotaxis behaviors in both wild-type and lin-45(K565N) missense mutant strains. + indicated wild-type copies of the *lin-45* gene. ⌽ indicted feeding with lin-45 RNAi plasmid without the activation of IPTG. The height of box indicated the mean and standard deviation of the strain. Error bars represented the maximum and minimum values. All data were analyzed using one-way ANOVA and Tukey multiple comparison. * indicated significantly different with comparison to the wild-type animals ($p < 0.01$).

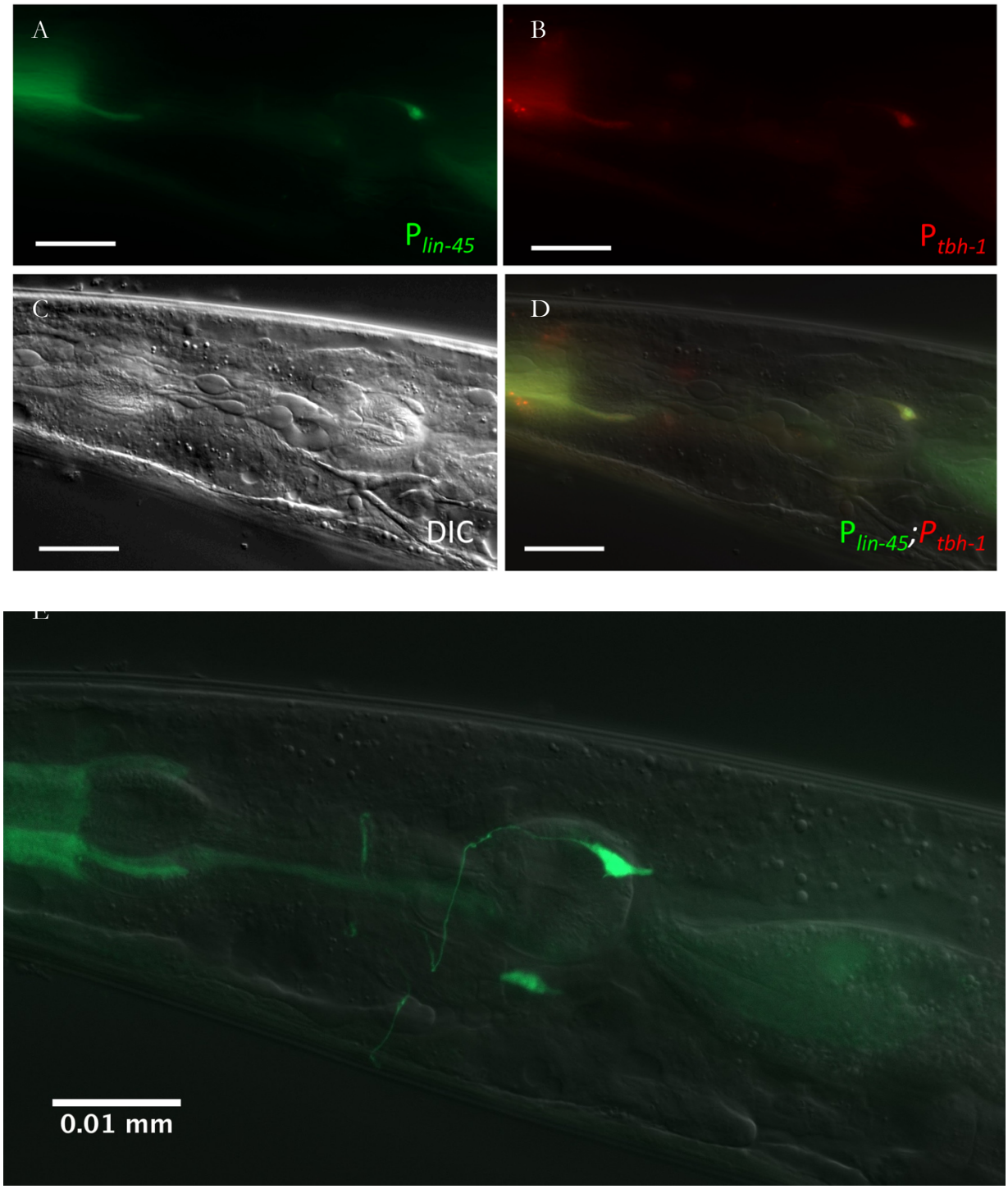

Figure 5-6. The expression pattern of *lin-45*.
(A)-(D) Co-expression of *lin-45* GFP and the RIC neuron-specific marker *tbh-1* mCherry. According to the DIC image (C), the pair of neurons co-expressing of *lin-45* and *tbh-1* were located near the pharyngeal. Using the location, co-expression pattern, and morphology (E), we identified these lin-45 expressing neurons as RIC interneurons. Scale bar: 10 μm.

Chapter 6

GENERAL DISCUSSION

6.1 Key findings and significance

Over the past seven years, I established a working pipeline to identify and generate evolutionarily conserved ASD-associated missense mutants in *C. elegans*. My colleagues and I used this pipeline to screen ASD-associated missense variants with functional consequences. Since these missense residues are conserved between *C. elegans* and humans, it is likely that they also have functionally significant roles in the human proteins. Overall, we identified 345 (19%) conserved residues from the ASD SFARI database and successfully generated 60 missense mutants using the CRISPR/Cas9 system and homology-directed repair. We then evaluated the neurodevelopmental phenotypes in these missense mutants using fecundity and an automated locomotion tracking system. In the first batch (Chapter 2), we examined 20 missense mutants predicted to have damaging effects and found only 70% of them displayed phenotypic changes in our behavioral tests battery. This study prioritized the phenotypic changing missense variants and provided feedback to improve the existing prediction algorithms' accuracy. In the second batch (Chapter 3), we investigated the functional outcomes of 28 ASD-associated missense variants on two gene clusters: synaptic function genes and gene expression and regulatory genes. This study revealed that missense mutants in different gene clusters displayed distinct phenotype profiles and lay the foundation for future systematic characterization of disease-associated variants.

In addition to the systematic approach, I worked on the ASD-associated missense variants on two specific genes. I found that missense mutation in the ALDH1A3, the key enzyme of

retinoic acid biosynthesis, impaired the fecundity in the orthologous *C. elegans* model. This study elucidated the effects of ALDH1A3 C174Y missense mutation in the retinoic acid signaling pathway during development (Chapter 4). In addition, I studied the diacetyl chemotaxis response in multiple genetic variants of *BRAF*, another ASD risk gene linking synaptic receptors to the intracellular signaling cascade. I discovered that the orthologous C. elegans model of *BRAF* K499N mutation displays a hypersensitive phenotype in the diacetyl chemotaxis assay. This study suggested a potential gain-of-function allele in the orthologous gene of *BRAF*, especially in the sensory function (Chapter 5).

6.2 Advantages and limitations of using C. elegans to model human diseases

Traditionally, rodent models, particularly mice (Mus musculus), are commonly used in pre-clinical research. Mice share the orthologs of 99% of human genes. The similarity in genome sequence leads to conserved physiological processes between mice and humans. In ASD research, the neural development defects can be modeled by social interaction and preference tests, repetitive behaviors (marble burying and nest shredding), and sensory and motor function (Crawley, 2012; Homberg et al., 2016). Although genetically modified mice (knockout, knockin, transgenics) with ASD risk gene mutations are widely available, the generation of a mouse model takes at least six months to a year (Kazdoba et al., 2016). The high maintenance costs and long time required for experimentation also restricted rodent models for genetic screens. On the other hand, an emerging amount of studies model ASD with human induced pluripotent stem cells (iPSCs), deprogrammed immortal stem cells that resemble embryonic stem cells (Takahashi & Yamanaka, 2006). Although iPSCs can reiterate the neuronal lineage formation, it lacks integration in functional circuits (Schmeisser & Parker, 2018). As a result, the *C. elegans* model offers a powerful tool for rapidly screening missense variants and studying functional changes

across tissues. Our approach and other *C. elegans* models (McDiarmid et al., 2018) open the new door for the systematic study of missense variants *in vivo*.

Using *C. elegans* to model psychiatric disorders has some limitations, including a lack of highly complex behaviors, brain structure, and some neurotransmitter systems (e.g., norepinephrine). However, *C. elegans* and humans share essential physiological pathways, such as insulin signaling, Ras/Notch signaling, p53, and many miRNAs. Although it is difficult to model complex behavioral traits (e.g., social interaction) in *C. elegans*, the simplicity of behaviors provides a more precise map to navigate the relationship between gene and physiological functions. The *C. elegans* are cost-effectiveness models for the initial screening for the subtle genetic variants or identifying novel bioactive drug compounds (Schmeisser et al., 2017).

6.3 A message to my past self

Knowing what I know now, I would take a slightly different approach to study the ASD-associated missense variants. We kicked start the autism project by generating missense mutants, making the pipeline going smoothly among several individuals. Looking back, I focused too much on the number of mutants generated, and I did not spend enough time thinking about the big picture. If I have planned the project in a result-oriented way, I would begin the project by screening the existing mutants of ASD risk genes. Identifying the mutants of interest would probably require some collaboration with bioinformaticians. Also, I would dedicate more time figuring out a stable and more ASD-relevant phenotype and use only one assay for the whole screening process. When I first switched from mouse models to the *C. elegans* model, the ability to CRISPR a targeted mutant excited me. Little did I consider the opportunity cost when generating a missense mutant without a clear hypothesis in mind. Although *C. elegans* allows rapid generation of mutant strains, it still takes tons of time and effort for genotyping and

managing the pipeline. Therefore, I would choose the missense variants in the following principles: (1) generating missense variants on a few genes whose loss-of-function mutants show phenotypes, (2) generating missense variants covering multiple domains of a gene, (3) focusing on genes within the same genetic network, for they are more likely to have interactions.

One of the lessons I learned from my Ph.D. is the growth mindset. I used to value accomplishment more than progress. I spent a lot of time trying to look smart, and I was afraid to step out of my comfort zone. With a growth mindset, I would encourage my past self to present and publish more, even though the manuscript is not ready. I would organize committee meetings more often and apply for more fellowships. I would be more willing to ask for help and go talk to people about my research so I can figure out what I did not know. Another lesson I learned about myself is that I enjoy working in a team environment. I like being able to talk about the excitement and frustration with a group of people who actually care about the project. I like the weekly subgroup meeting so we can keep each other on track. Working on the autism group project has inspired me to pursue a career in the pharmaceutical industry.

Perhaps, I had an unrealistic expectation for myself when I first started graduate school. During the process, I was disappointed by how trivial my discovery seemed to be. It was not until I came across a preprint on bioRxiv and recognized one sentence describing my research. At that moment, I realized I did contribute something to the scientific community.

6.4 References

Crawley, J. N. (2012). Translational animal models of autism and neurodevelopmental disorders. *Dialogues Clin Neurosci*, *14*(3), 293–305. https://doi.org/10.1038/jid.2015.269

Homberg, J. R., Kyzar, E. J., Scattoni, M. L., Norton, W. H., Pittman, J., Gaikwad, S., Nguyen, M., Poudel, M. K., Ullmann, J. F. P., Diamond, D. M., Kaluyeva, A. A., Parker, M. O., Brown, R. E., Song, C., Gainetdinov, R. R., Gottesman, I. I., & Kalueff, A. V. (2016).

Genetic and environmental modulation of neurodevelopmental disorders: Translational insights from labs to beds. *Brain Research Bulletin*, *125*, 79–91. https://doi.org/10.1016/j.brainresbull.2016.04.015

Kazdoba, T. M., Leach, P. T., Yang, M., Silverman, J. L., Solomon, M., & Crawley, J. N. (2016). Translational mouse models of autism: advancing toward pharmacological therapeutics. *Curr Topics Behav Neurosci*, *28*, 1–52. https://doi.org/10.1007/7854

McDiarmid, T. A., Au, V., Loewen, A. D., Liang, J., Mizumoto, K., Moerman, D. G., & Rankin, C. H. (2018). CRISPR-Cas9 human gene replacement and phenomic characterization in *Caenorhabditis elegans* to understand the functional conservation of human genes and decipher variants of uncertain significance. *Disease Models & Mechanisms*, dmm.036517. https://doi.org/10.1242/dmm.036517

Schmeisser, K., Fardghassemi, Y., & Parker, J. A. (2017). A rapid chemical-genetic screen utilizing impaired movement phenotypes in C. elegans: Input into genetics of neurodevelopmental disorders. *Experimental Neurology*, *293*, 101–114. https://doi.org/10.1016/j.expneurol.2017.03.022

Schmeisser, K., & Parker, J. A. (2018). Worms on the spectrum - C. elegans models in autism research. *Experimental Neurology*, *299*, 199–206. https://doi.org/10.1016/j.expneurol.2017.04.007

Takahashi, K., & Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*, *126*(4), 663–676. https://doi.org/10.1016/j.cell.2006.07.024

INDEX